CENTRO LATINOAMERICANO DE DEMOGRAFIA
CELADE

Stgo, Chile

LECTURAS SELECCIONADAS DE

D E M O G R A F I A   E C O N O M I C A

Compilado por Angel Fucaraccio
para las Conferencias del
Profesor George Tapinos
22 al 29 de Agosto de
1985

68

Stgo, Chile

INDICE

Página

LOS DETERMINANTES ECONÒMICOS DE LA FECUNDIDAD Y MORTALIDAD

G.S. Becker, 'A theory of the allocation of Time' THE ECONOMIC JOURNAL, 1965, páginas 493-517 1

A. C. DeSerpa 'A Theory of the Economics of Time' THE ECONOMIC JOURNAL, No. 324 Dec.1971, vol.81 14

G.S. Becker, 'An Economic Analysis of Fertility', in NBER. DEMOGRAPHIC AND ECONOMIC CHANGE IN DEVELOPED COUNTRIES, Princeton University Press, 1960 33

J. Blake, Are Babies Consumer Durables? POPULATION STUDIES, March 1968 51

W. Sanderson, or On Two Schools of the Economic of Fertility, POPULATION AND DEVELOPMENT REVIEW, September 1976, pp. 469-477. 72

P.A. Samuelson, An Economist's Non Linear Model of Self Generated Fertility Waves, POPULATION STUDIES, July 1976. 77

J. Caldwell, 'Toward A Restatement of Demographic Transition Theory' POPULATION AND DEVELOPMENT REVIEW, September 1976, pp.321-366. 82

R. Pollock, R.A. Easterlin and M. Wachter, Toward a More General Theory of Fertility Determination, in ECONOMIC AND DEMOGRAPHIC CHANGE IN LESS DEVELOPED COUNTRIES, Easterlin (ed.), 1980. 106

D.P. Parson, On the Economics of the Intergenerational Control, POPULATION AND DEVELOPMENT REVIEW, March 1984, pp. 41-54. 141

J. Ermisch, The Relevance of the Easterlin Hypothesis and the New Home Economics to Fertility Movements in Great-Britain. POPULATION STUDIES, vol. 33, No. 1, pp. 39-58. 155

S.H. Preston, Causes and Consequences of Mortality Declines in Less Developed Countries During the Twentieth Century, in POPULATION AND ECONOMIC CHANGE IN DEVELOPING COUNTRIES, R.A. Easterlin Ed., NBER, The University of Chicago Press 1980. See pp. 289-315. 175

CRECIMIENTO DEMOGRAFICO, ACUMULACION DE CAPITAL Y CAMBIO TECNOLOGICO

G. Ranis and J.C. Fei, A Theory of Economic Development, AMERICAN ECONOMIC REVIEW, September 1961. 212

F.T. Denton and B.C. Spencer, Household and Population Effects on Aggregate Consumption. THE REVIEW OF ECONOMICS AND STATISTICS, No. 58, 1976, pp. 86-95. 245

Allen C. Kelley, 'Population Growth, the Dependency rate, and the Pace of Economic Development' POPULATION STUDIES, Vol. XXVII, 1973 255

T.J. Espenshade, The Impact of Children on Household Saving: Age Effects versus Family Size. POPULATION STUDIES, March 1973. 265

D. Eilenstine, J.P. Cunningham, Projected Consumption Patterns for a Stationary Population. POPULATION STUDIES, No. 26, 2, July 1972, pp. 223-231. 267

T.J. Espenshade, How A Trend Toward Stationary Population Affects Consumer Demand. POP.STUDIES, vol.32 No. 1 276

R.J. Willis, The Direction of Intergenerational Transfers and Demographic Transition: The Caldwell Hypothsis Reexamined, in Y. Ben Porath (ed.) Demographic Variables and Income Distribution. POPULATION AND DEVELOPMENT REVIEW, 1982. 288

R.A. Easterlin, What will 1984 be Like? Socioeconomic Implications of Recent Twists in Age Structure. DEMOGRAPHY. November 1978. See pp. 397-421. 316

F. Welch, Effects of Cohort Size on Earnings: The Baby Boom Babies Financial Bust. J.P.E., vol. 87, No. 5, 1976. 341

R.B. Freeman, The Effects of Demographic Factors on Age-Earnings Profiles, THE JOURNAL OF HUMAN RESOURCES, Summer 1979, pp. 289-318. 358

S. Enke, Economic Consequences of Rapid Population Growth. THE ECONOMIC JOURNAL, December 1971, p. 800-811. 374

S. Enke, Reducing Fertility to Accelerate Development, THE ECONOMIC JOURNAL, June 1974, pp. 349-366. 386

G. Myrdal, A Note on Inadequate Approaches to the Effects of Population Changes, AN ASIAN DRAMA, vol. 3, appendix 7. 404

1

THE ECONOMIC JOURNAL

SEPTEMBER 1965

A THEORY OF THE ALLOCATION OF TIME

I. Introduction

THROUGHOUT history the amount of time spent at work has never consistently been much greater than that spent at other activities. Even a work week of fourteen hours a day for six days still leaves half the total time for sleeping, eating and other activities. Economic development has led to a large secular decline in the work week, so that whatever may have been true of the past, to-day it is below fifty hours in most countries, less than a third of the total time available. Consequently the allocation and efficiency of non-working time may now be more important to economic welfare than that of working time; yet the attention paid by economists to the latter dwarfs any paid to the former.

Fortunately, there is a movement under way to redress the balance. The time spent at work declined secularly, partly because young persons increasingly delayed entering the labour market by lengthening their period of schooling. In recent years many economists have stressed that the time of students is one of the inputs into the educational process, that this time could be used to participate more fully in the labour market and therefore that one of the costs of education is the forgone earnings of students. Indeed, various estimates clearly indicate that forgone earnings is the dominant private and an important social cost of both high-school and college education in the United States.[1] The increased awareness of the importance of forgone earnings has resulted in several attempts to economise on students' time, as manifested, say, by the spread of the quarterly and tri-mester systems.[2]

Most economists have now fully grasped the importance of forgone earnings in the educational process and, more generally, in all investments in human capital, and criticise educationalists and others for neglecting them. In the light of this it is perhaps surprising that economists have not been

[1] See T. W. Schultz, "The Formation of Human Capital by Education," *Journal of Political Economy* (December 1960), and my *Human Capital* (Columbia University Press for the N.B.E.R., 1964), Chapter IV. I argue there that the importance of forgone earnings can be directly seen, *e.g.*, from the failure of free tuition to eliminate impediments to college attendance or the increased enrolments that sometimes occur in depressed areas or time periods.

[2] On the cause of the secular trend towards an increased school year see my comments, *ibid.*, p. 103.

equally sophisticated about other non-working uses of time. For example, the cost of a service like the theatre or a good like meat is generally simply said to equal their market prices, yet everyone would agree that the theatre and even dining take time, just as schooling does, time that often could have been used productively. If so, the full costs of these activities would equal the sum of market prices and the forgone value of the time used up. In other words, indirect costs should be treated on the same footing when discussing all non-work uses of time, as they are now in discussions of schooling.

In the last few years a group of us at Columbia University have been occupied, perhaps initially independently but then increasingly less so, with introducing the cost of time systematically into decisions about non-work activities. J. Mincer has shown with several empirical examples how estimates of the income elasticity of demand for different commodities are biased when the cost of time is ignored;[1] J. Owen has analysed how the demand for leisure can be affected;[2] E. Dean has considered the allocation of time between subsistence work and market participation in some African economies;[3] while, as already mentioned, I have been concerned with the use of time in education, training and other kinds of human capital. Here I attempt to develop a general treatment of the allocation of time in all other non-work activities. Although under my name alone, much of any credit it merits belongs to the stimulus received from Mincer, Owen, Dean and other past and present participants in the Labor Workshop at Columbia.[4]

The plan of the discussion is as follows. The first section sets out a basic theoretical analysis of choice that includes the cost of time on the same footing as the cost of market goods, while the remaining sections treat various empirical implications of the theory. These include a new approach to changes in hours of work and "leisure," the full integration of so-called "productive" consumption into economic analysis, a new analysis of the effect of income on the quantity and "quality" of commodities consumed, some suggestions on the measurement of productivity, an economic analysis of queues and a few others as well. Although I refer to relevant empirical

[1] See his "Market Prices, Opportunity Costs, and Income Effects," in *Measurement in Economics: Studies in Mathematical Economics and Econometrics in Memory of Yehuda Grunfeld* (Stanford University Press, 1963). In his well-known earlier study Mincer considered the allocation of married women between "housework" and labour force participation. (See his "Labor Force Participation of Married Women," in *Aspects of Labor Economics* (Princeton University Press, 1962).)

[2] See his *The Supply of Labor and the Demand for Recreation* (unpublished Ph.D. dissertation, Columbia University, 1964).

[3] See his *Economic Analysis and African Response to Price* (unpublished Ph.D. dissertation, Columbia University, 1963).

[4] Let me emphasise, however, that I alone am responsible for any errors.

I would also like to express my appreciation for the comments received when presenting these ideas to seminars at the Universities of California (Los Angeles), Chicago, Pittsburgh, Rochester and Yale, and to a session at the 1963 Meetings of the Econometric Society. Extremely helpful comments on an earlier draft were provided by Milton Friedman and by Gregory C. Chow; the latter also assisted in the mathematical formulation. Linda Kee provided useful research assistance.

work that has come to my attention, little systematic testing of the theory has been attempted.

II. A Revised Theory of Choice

According to traditional theory, households maximise utility functions of the form

$$U = U(y_1, y_2, \ldots, y_n) \qquad (1)$$

subject to the resource constraint

$$\sum p_i' y_i = I = W + V \qquad (2)$$

where y_i are goods purchased on the market, p'_i are their prices, I is money income, W is earnings and V is other income. As the introduction suggests, the point of departure here is the systematic incorporation of non-working time. Households will be assumed to combine time and market goods to produce more basic commodities that directly enter their utility functions. One such commodity is the seeing of a play, which depends on the input of actors, script, theatre and the playgoer's time; another is sleeping, which depends on the input of a bed, house (pills?) and time. These commodities will be called Z_i and written as

$$Z_i = f_i(x_i, T_i) \qquad (3)$$

where x_i is a vector of market goods and T_i a vector of time inputs used in producing the ith commodity.[1] Note that, when capital goods such as refrigerators or automobiles are used, x refers to the services yielded by the goods. Also note that T_i is a vector because, *e.g.*, the hours used during the day or on weekdays may be distinguished from those used at night or on week-ends. Each dimension of T_i refers to a different aspect of time. Generally, the partial derivatives of Z_i with respect to both x_i and T_i are non-negative.[2]

In this formulation households are both producing units and utility maximisers. They combine time and market goods via the "production functions" f_i to produce the basic commodities Z_i, and they choose the best combination of these commodities in the conventional way by maximising a utility function

$$U = U(Z_i, \ldots Z_m) \equiv U(f_1, \ldots f_m) \equiv U(x_1, \ldots x_m;\ T_1, \ldots T_m) \qquad (4)$$

[1] There are several empirical as well as conceptual advantages in assuming that households combine goods and time to produce commodities instead of simply assuming that the amount of time used at an activity is a direct function of the amount of goods consumed. For example, a change in the cost of goods relative to time could cause a significant substitution away from the one rising in relative cost. This, as well as other applications, are treated in the following sections.

[2] If a good or time period was used in producing several commodities I assume that these "joint costs" could be fully and uniquely allocated among the commodities. The problems here are no different from those usually arising in the analysis of multi-product firms.

subject to a budget constraint

$$g(Z_i, \ldots Z_m) = Z \qquad \ldots \ldots \quad (5)$$

where g is an expenditure function of Z_i and Z is the bound on resources. The integration of production and consumption is at odds with the tendency for economists to separate them sharply, production occurring in firms and consumption in households. It should be pointed out, however, that in recent years economists increasingly recognise that a household is truly a "small factory":[1] it combines capital goods, raw materials and labour to clean, feed, procreate and otherwise produce useful commodities. Undoubtedly the fundamental reason for the traditional separation is that firms are usually given control over working time in exchange for market goods, while "discretionary" control over market goods and consumption time is retained by households as they create their own utility. If (presumably different) firms were also given control over market goods and consumption time in exchange for providing utility the separation would quickly fade away in analysis as well as in fact.

The basic goal of the analysis is to find measures of g and Z which facilitate the development of empirical implications. The most direct approach is to assume that the utility function in equation (4) is maximised subject to separate constraints on the expenditure of market goods and time, and to the production functions in equation (3). The goods constraint can be written as

$$\sum_1^m p_i x_i = I = V + T_w \bar{w} \qquad \ldots \ldots \quad (6)$$

where p_i is a vector giving the unit prices of x_i, T_w is a vector giving the hours spent at work and $\bar{w}$ is a vector giving the earnings per unit of T_w. The time constraints can be written as

$$\sum_1^m T_i = T_c = T - T_w \qquad \ldots \ldots \quad (7)$$

where T_c is a vector giving the total time spent at consumption and T is a vector giving the total time available. The production functions (3) can be written in the equivalent form

$$\left.\begin{aligned} T_i &\equiv t_i Z_i \\ x_i &\equiv b_i Z_i \end{aligned}\right\} \qquad \ldots \ldots \quad (8)$$

where t_i is a vector giving the input of time per unit of Z_i and b_i is a similar vector for market goods.

The problem would appear to be to maximise the utility function (4) subject to the multiple constraints (6) and (7) and to the production relations (8). There is, however, really only one basic constraint: (6) is not independent of (7) because time can be converted into goods by using less time

[1] See, *e.g.*, A. K. Cairncross, "Economic Schizophrenia," *Scottish Journal of Political Economy* (February 1958).

at consumption and more at work. Thus, substituting for T_w in (6) its equivalent in (7) gives the single constraint [1]

$$\sum p_i x_i + \sum T_i \bar{w} = V + T\bar{w} \quad . \quad . \quad . \quad . \quad . \quad (9)$$

By using (8), (9) can be written as

$$\sum (p_i b_i + t_i \bar{w}) Z_i = V + T\bar{w} \quad . \quad . \quad . \quad . \quad (10)$$

with

$$\left.\begin{aligned} \pi_i &\equiv p_i b_i + t_i \bar{w} \\ S' &\equiv V + T\bar{w} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

The full price of a unit of Z_i (π_i) is the sum of the prices of the goods and of the time used per unit of Z_i. That is, the full price of consumption is the sum of direct and indirect prices in the same way that the full cost of investing in human capital is the sum of direct and indirect costs.[2] These direct and indirect prices are symmetrical determinants of total price, and there is no analytical reason to stress one rather than the other.

The resource constraint on the right side of equation (10), S', is easy to interpret if $\bar{w}$ were a constant, independent of the Z_i. For then S' gives the money income achieved if all the time available were devoted to work. This achievable income is "spent" on the commodities Z_i either directly through expenditures on goods, $\sum p_i b_i Z_i$, or indirectly through the forgoing of income, $\sum t_i \bar{w} Z_i$, *i.e.*, by using time at consumption rather than at work. As long as $\bar{w}$ were constant, and if there were constant returns in producing Z_i so that b_i and t_i were fixed for given p_i and $\bar{w}$ the equilibrium condition resulting from maximising (4) subject to (10) takes a very simple form:

$$U_i = \frac{\partial U}{\partial Z_i} = \lambda \pi_i \qquad i = 1, \ldots m \quad . \quad . \quad . \quad (12)$$

where λ is the marginal utility of money income. If $\bar{w}$ were not constant the resource constraint in equation (10) would not have any particularly useful interpretation: $S' = V + T\bar{w}$ would overstate the money income achievable as long as marginal wage-rates were below average ones. Moreover, the equilibrium conditions would become more complicated than (12) because marginal would have to replace average prices.

The total resource constraint could be given the sensible interpretation of the maximum money income achievable only in the special and unlikely case when average earnings were constant. This suggests dropping the approach based on explicitly considering separate goods and time constraints and substituting one in which the total resource constraint necessarily equalled the maximum money income achievable, which will be simply called "full income." [3] This income could in general be obtained by devoting all the

[1] The dependency among constraints distinguishes this problem from many other multiple-constraint situations in economic analysis, such as those arising in the usual theory of rationing (see J. Tobin, "A Survey of the Theory of Rationing," *Econometrica* (October, 1952)). Rationing would reduce to a formally identical single-constraint situation if rations were saleable and fully convertible into money income. ω

[2] See my *Human Capital*, *op. cit.*

[3] This term emerged from a conversation with Milton Friedman.

time and other resources of a household to earning income, with no regard for consumption. Of course, all the time would not usually be spent "at" a job: sleep, food, even leisure are required for efficiency, and some time (and other resources) would have to be spent on these activities in order to maximise money income. The amount spent would, however, be determined solely by the effect on income and not by any effect on utility. Slaves, for example, might be permitted time "off" from work only in so far as that maximised their output, or free persons in poor environments might have to maximise money income simply to survive.[1]

Households in richer countries do, however, forfeit money income in order to obtain additional utility, *i.e.*, they exchange money income for a greater amount of psychic income. For example, they might increase their leisure time, take a pleasant job in preference to a better-paying unpleasant one, employ unproductive nephews or eat more than is warranted by considerations of productivity. In these and other situations the amount of money income forfeited measures the cost of obtaining additional utility.

Thus the full income approach provides a meaningful resource constraint and one firmly based on the fact that goods and time can be combined into a single overall constraint because time can be converted into goods through money income. It also incorporates a unified treatment of all substitutions of non-pecuniary for pecuniary income, regardless of their nature or whether they occur on the job or in the household. The advantages of this will become clear as the analysis proceeds.

If full income is denoted by S, and if the total earnings forgone or "lost" by the interest in utility is denoted by L, the identity relating L to S and I is simply

$$L(Z_1, \ldots, Z_m) \equiv S - I(Z_1, \ldots, Z_m) \quad . \quad . \quad . \quad (13)$$

I and L are functions of the Z_i because how much is earned or forgone depends on the consumption set chosen; for example, up to a point, the less leisure chosen the larger the money income and the smaller the amount forgone.[2] Using equations (6) and (8), equation (13) can be written as

$$\sum p_i b_i Z_i + L(Z_1, \ldots, Z_m) \equiv S \quad . \quad . \quad . \quad . \quad (14)$$

[1] Any utility received would only be an incidental by-product of the pursuit of money income. Perhaps this explains why utility analysis was not clearly formulated and accepted until economic development had raised incomes well above the subsistence level.

[2] Full income is achieved by maximising the earnings function

$$W = W(Z_1, \ldots Z_m) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1')$$

subject to the expenditure constraint in equation (6), to the inequality

$$\sum_1^m T_1 \leq T \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2')$$

and to the restrictions in (8). I assume for simplicity that the amount of each dimension of time

This basic resource constraint states that full income is spent either directly on market goods or indirectly through the forgoing of money income. Unfortunately, there is no simple expression for the average price of Z_i as there is in equation (10). However, marginal, not average, prices are relevant for behaviour, and these would be identical for the constraint in (10) only when average earnings, $\bar{w}$, was constant. But, if so, the expression for the loss function simplifies to

$$L = \bar{w}T_c = \bar{w}\sum t_i Z_i \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

and (14) reduces to (10). Moreover, even in the general case the total marginal prices resulting from (14) can always be divided into direct and indirect components: the equilibrium conditions resulting from maximising the utility function subject to (14) [1] are

$$U_i = T(p_i b_i + L_i), \qquad i = 1, \ldots, m \quad . \quad . \quad . \quad (16)$$

where $p_i b_i$ is the direct and L_i the indirect component of the total marginal price $p_i b_i + L_i$.[2]

Behind the division into direct and indirect costs is the allocation of time and goods between work-orientated and consumption-orientated activities. This suggests an alternative division of costs; namely, into those resulting from the allocation of goods and those resulting from the allocation of time. Write $L_i = \partial L/\partial Z_i$ as

$$L_i = \frac{\partial L}{\partial T_i}\frac{\partial T_i}{\partial Z_i} + \frac{\partial L}{\partial x_i}\frac{\partial x_i}{\partial Z_i} \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

$$= l_i t_i + c_i b_i \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

where $l_i = \dfrac{\partial L}{\partial T_i}$ and $c_i = \dfrac{\partial L}{\partial x_i}$ are the marginal forgone earnings of using more time and goods respectively on Z_i. Equation (16) can then be written as

$$U_i = T[b_i(p_i + c_i) + t_i l_i] \quad . \quad . \quad . \quad . \quad . \quad (19)$$

The total marginal cost of Z_i is the sum of $b_i(p_i + c_i)$, the marginal cost of using goods in producing Z_i, and $t_i l_i$, the marginal cost of using time. This division would be equivalent to that between direct and indirect costs only if $c_i = 0$ or if there were no indirect costs of using goods.

---

difficult to incorporate this constraint. Maximising (1′) subject to (6) and (8) yields the following conditions

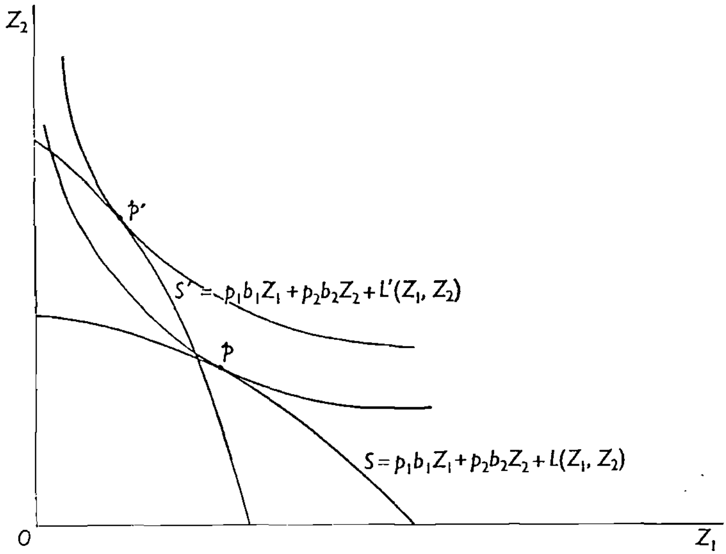
$$\frac{\partial W}{\partial Z_i} = \frac{p_i b_i \sigma}{1 + \sigma} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3')$$

where σ is the marginal productivity of money income. Since the loss function $L = (S - V) - W$, the equilibrium conditions to minimise the loss is the same as (3′) except for a change in sign.

[1] Households maximise their utility subject only to the single total resource constraint given by (14), for once the full income constraint is satisfied, there is no other restriction on the set of Z_i that can be chosen. By introducing the concept of full income the problem of maximising utility subject to the time and goods constraints is solved in two stages: first, full income is determined from the goods and time constraints, and then utility is maximised subject only to the constraint imposed by full income.

[2] It can easily be shown that the equilibrium conditions of (16) are in fact precisely the same as

The accompanying figure shows the equilibrium given by equation (16) for a two-commodity world. In equilibrium the slope of the full income



opportunity curve, which equals the ratio of marginal prices, would equal the slope of an indifference curve, which equals the ratio of marginal utilities. Equilibrium occurs at p and p' for the opportunity curves S and S' respectively.

The rest of the paper is concerned with developing numerous empirical implications of this theory, starting with determinants of hours worked and concluding with an economic interpretation of various queueing systems. To simplify the presentation, it is assumed that the distinction between direct and indirect costs is equivalent to that between goods and time costs; in other words, the marginal forgone cost of the use of goods, c_i, is set equal to zero. The discussion would not be much changed, but would be more cumbersome were this not assumed.[1] Finally, until Section IV goods and time are assumed to be used in fixed proportions in producing commodities; that is, the coefficients b_i and t_i in equation (8) are treated as constants.

III. Applications

(a) *Hours of Work*

If the effects of various changes on the time used on consumption, T_c, could be determined their effects on hours worked, T_w, could be found residually from equation (7). This section considers, among other things, the effects of changes in income, earnings and market prices on T_c, and thus on T_w,

[1] Elsewhere I have discussed some effects of the allocation of goods on productivity (see my "Investment in Human Capital: A Theoretical Analysis," *Journal of Political Economy*, special supplement (October 1962), Section 2); essentially the same discussion can be found in *Human Capital, op. cit.*, Chapter II.

using as the major tool of analysis differences among commodities in the importance of forgone earnings.

The relative marginal importance of forgone earnings is defined as

$$\alpha_i = \frac{l_i t_i}{p_i b_i + l_i t_i} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

The importance of forgone earnings would be greater the larger l_i and t_i, the forgone earnings per hour of time and the number of hours used per unit of Z_i respectively, while it would be smaller the larger p_i and b_i, the market price of goods and the number of goods used per unit of Z_i respectively. Similarly, the relative marginal importance of time is defined as

$$\gamma_i = \frac{t_i}{p_i b_i + l_i t_i} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

If full income increased solely because of an increase in V (other money income) there would simply be a parallel shift of the opportunity curve to the right with no change in relative commodity prices. The consumption of most commodities would have to increase; if all did, hours worked would decrease, for the total time spent on consumption must increase if the output of all commodities did, and by equation (7) the time spent at work is inversely related to that spent on consumption. Hours worked could increase only if relatively time intensive commodities, those with large γ, were sufficiently inferior.[1]

A uniform percentage increase in earnings for all allocations of time would increase the cost per hour used in consumption by the same percentage for all commodities.[2] The relative prices of different commodities would, however, change as long as forgone earnings were not equally important for all; in particular, the prices of commodities having relatively important forgone earnings would rise more. Now the fundamental theorem of

[1] The problem is: under what conditions would

$$\frac{-\partial T_w}{\partial V} = \frac{\partial T_c}{\partial V} = \Sigma t_i \frac{\partial Z_i}{\partial V} < 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1')$$

when

$$\Sigma(p_i b_i + l_i t_i) \frac{\partial Z_i}{\partial V} = 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2')$$

If the analysis were limited to a two-commodity world where Z_1 was more time intensive, then it can easily be shown that (1′) would hold if, and only if,

$$\frac{\partial Z_1}{\partial V} < \frac{-\gamma_2}{(\gamma_1 - \gamma_2)(p_1 b_1 + l_1 t_1)} < 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3')$$

[2] By a uniform change of β is meant

$$W_1 = (1 + \beta) W_0(Z_1, \ldots Z_n)$$

where W_0 represents the earnings function before the change and W_1 represents it afterwards. Since the loss function is defined as

$$\begin{aligned} L &= S - W - V \\ &= W(\hat{Z}) - W(Z), \end{aligned}$$

then

$$\begin{aligned} L_1 &= W_1(\hat{Z}) - W_1(Z) \\ &= (1 + \beta)\,[W_0(\hat{Z}) - W_0(Z)] = (1 + \beta) L_0 \end{aligned}$$

Consequently, all opportunities costs also change by β.

demand theory states that a compensated change in relative prices would induce households to consume less of commodities rising in price. The figure shows the effect of a rise in earnings fully compensated by a decline in other income: the opportunity curve would be rotated clockwise through the initial position p if Z_1 were the more earnings-intensive commodity. In the figure the new equilibrium p' must be to the left and above p, or less Z_1 and more Z_2 would be consumed.

Therefore a compensated uniform rise in earnings would lead to a shift away from earnings-intensive commodities and towards goods-intensive ones. Since earnings and time intensiveness tend to be positively correlated,[1] consumption would be shifted from time-intensive commodities. A shift away from such commodities would, however, result in a reduction in the total time spent in consumption, and thus an increase in the time spent at work.[2]

The effect of an uncompensated increase in earnings on hours worked would depend on the relative strength of the substitution and income effects. The former would increase hours, the latter reduce them; which dominates cannot be determined *a priori.*

The conclusion that a pure rise in earnings increases and a pure rise in income reduces hours of work must sound very familiar, for they are traditional results of the well-known labour–leisure analysis. What, then, is the relation between our analysis, which treats all commodities symmetrically and stresses only their differences in relative time and earning intensities, and the usual analysis, which distinguishes a commodity having special properties called "leisure" from other more commonplace commodities? It is easily shown that the usual labour–leisure analysis can be looked upon as a special case of ours in which the cost of the commodity called leisure consists entirely of forgone earnings and the cost of other commodities entirely of goods.[3]

[1] According to the definitions of earning and time intensity in equations (20) and (21), they would be positively correlated unless l_i and t_i were sufficiently negatively correlated. See the further discussion later on.

[2] Let it be stressed that this conclusion usually holds, even when households are irrational; sophisticated calculations about the value of time at work or in consumption, or substantial knowledge about the amount of time used by different commodities is not required. Changes in the hours of work, even of non-maximising, impulsive, habitual, etc., households would tend to be positively related to compensated changes in earnings because demand curves tend to be negatively inclined even for such households (see G. S. Becker, "Irrational Behavior and Economic Theory," *Journal of Political Economy* (February 1962)).

[3] Suppose there were two commodities Z_1 and Z_2, where the cost of Z_1 depended only on the cost of market goods, while the cost of Z_2 depended only on the cost of time. The goods-budget constraint would then simply be

$$p_1 b_1 Z_1 = I = V + T_w \bar{w}$$

and the constraint on time would be

$$t_2 Z_2 = T - T_w$$

This is essentially the algebra of the analysis presented by Henderson and Quandt, and their treat-

As a description of reality such an approach, of course, is not tenable, since virtually all activities use both time and goods. Perhaps it would be defended either as an analytically necessary or extremely insightful approximation to reality. Yet the usual substitution and income effects of a change in resources on hours worked have easily been derived from a more general analysis which stresses only that the relative importance of time varies among commodities. The rest of the paper tries to go further and demonstrate that the traditional approach, with its stress on the demand for "leisure," apparently has seriously impeded the development of insights about the economy, since the more direct and general approach presented here naturally leads to a variety of implications never yet obtained.

The two determinants of the importance of forgone earnings are the amount of time used per dollar of goods and the cost per unit of time. Reading a book, taking a haircut or commuting use more time per dollar of goods than eating dinner, frequenting a night-club or sending children to private summer camps. Other things the same, forgone earnings would be more important for the former set of commodities than the latter.

The importance of forgone earnings would be determined solely by time intensity only if the cost of time was the same for all commodities. Presumably, however, it varies considerably among commodities and at different periods. For example, the cost of time is often less on week-ends and in the evenings because many firms are closed then,[1] which explains why a famous liner intentionally includes a week-end in each voyage between the United States and Europe.[2] The cost of time would also tend to be less for commodities that contribute to productive effort, traditionally called "productive consumption." A considerable amount of sleep, food and even "play" fall under this heading. The opportunity cost of the time is less because these commodities indirectly contribute to earnings. Productive consumption has had a long but bandit-like existence in economic thought; our analysis does systematically incorporate it into household decision-making.

Although the formal specification of leisure in economic models has ignored expenditures on goods, cannot one argue that a more correct specification would simply associate leisure with relatively important forgone earnings? Most conceptions of leisure do imply that it is time intensive and does not indirectly contribute to earnings,[3] two of the important

equilibrium condition that the rate of substitution between goods and leisure equals the real wage-rate is just a special case of our equation (19) (see *Microeconomic Theory* (McGraw-Hill, 1958), p. 23).

[1] For workers receiving premium pay on the week-ends and in the evenings, however, the cost of time may be considerably greater then.

[2] See the advertisement by United States Lines in various issues of the *New Yorker* magazine: "The S.S. *United States* regularly includes a week-end in its 5 days to Europe, saving [economic] time for businessmen" (my insertion).

[3] For example, *Webster's Collegiate Dictionary* defines leisurely as "characterized by leisure, taking *abundant time*" (my italics); or S. de Grazia, in his recent *Of Time, Work and Leisure*, says, "Leisure is a state of being in which activity is performed for its own sake or as its own end" (New York:

characteristics of earnings-intensive commodities. On the other hand, not all of what are usually considered leisure activities do have relatively important forgone earnings: night-clubbing is generally considered leisure, and yet, at least in its more expensive forms, has a large expenditure component. Conversely, some activities have relatively large forgone earnings and are not considered leisure: haircuts or child care are examples. Consequently, the distinction between earnings-intensive and other commodities corresponds only partly to the usual distinction between leisure and other commodities. Since it has been shown that the relative importance of forgone earnings rather than any concept of leisure is more relevant for economic analysis, less attention should be paid to the latter. Indeed, although the social philosopher might have to define precisely the concept of leisure,[1] the economist can reach all his traditional results as well as many more without introducing it at all!

Not only is it difficult to distinguish leisure from other non-work [2] but also even work from non-work. Is commuting work, non-work or both? How about a business lunch, a good diet or relaxation? Indeed, the notion of productive consumption was introduced precisely to cover those commodities that contribute to work as well as to consumption. Cannot pure work then be considered simply as a limiting commodity of such joint commodities in which the contribution to consumption was nil? Similarly, pure consumption would be a limiting commodity in the opposite direction in which the contribution to work was nil, and intermediate commodities would contribute to both consumption and work. The more important the contribution to work relative to consumption, the smaller would tend to be the relative importance of forgone earnings. Consequently, the effects of changes in earnings, other income, etc., on hours worked then become assimiliated to and essentially a special case of their effects on the consumption of less earnings-intensive commodities. For example, a pure rise in earnings would reduce the relative price, and thus increase the time spent on these commodities, *including the time spent at work*; similarly, for changes in income and other variables. The generalisation wrought by our approach is even greater than may have appeared at first.

Before concluding this section a few other relevant implications of our

[1] S. de Grazia has recently entertainingly shown the many difficulties in even reaching a reliable definition, and *a fortiori*, in quantitatively estimating the amount of leisure. See *ibid.*, Chapters III and IV; also see W. Moore, *Man, Time and Society* (New York: Wiley, 1963), Chapter II; J. N. Morgan, M. H. David, W. J. Cohen and H. E. Brazer, *Income and Welfare in the United States* (New York: McGraw-Hill, 1962), p. 322, and Owen, *op. cit.*, Chapter II.

[2] Sometimes true leisure is defined as the amount of discretionary time available (see Moore, *op. cit.*, p. 18). It is always difficult to attach a rigorous meaning to the word "discretionary" when referring to economic resources. One might say that in the short run consumption time is and working time is not discretionary, because the latter is partially subject to the authoritarian control of employers. (Even this distinction would vanish if households gave certain firms authoritarian control over their consumption time; see the discussion in Section II.) In the long run this definition of discretionary time is suspect too because the availability of alternative sources of employment would make working time also discretionary.

theory might be briefly mentioned. Just as a (compensated) rise in earnings would increase the prices of commodities with relatively large forgone earnings, induce a substitution away from them and increase the hours worked, so a (compensated) fall in market prices would also induce a substitution away from them and increase the hours worked: the effects of changes in direct and indirect costs are symmetrical. Indeed, Owen presents some evidence indicating that hours of work in the United States fell somewhat more in the first thirty years of this century than in the second thirty years, not because wages rose more during the first period, but because the market prices of recreation commodities fell more then.[1]

A well-known result of the traditional labour–leisure approach is that a rise in the income tax induces at least a substitution effect away from work and towards "leisure." Our approach reaches the same result only via a substitution towards time-intensive consumption rather than leisure. A simple additional implication of our approach, however, is that if a rise in the income tax were combined with an appropriate excise on the goods used in time-intensive commodities or subsidy to the goods used in other commodities there need be no change in full relative prices, and thus no substitution away from work. The traditional approach has recently reached the same conclusion, although in a much more involved way.[2]

There is no exception in the traditional approach to the rule that a pure rise in earnings would not induce a decrease in hours worked. An exception does occur in ours, for if the time and earnings intensities (*i.e.*, $l_i t_i$ and t_i) were negatively correlated a pure rise in earnings would induce a substitution towards time-intensive commodities, and thus away from work.[3] Although this exception does illustrate the greater power of our approach, there is no reason to believe that it is any more important empirically than the exception to the rule on income effects.

(b) *The Productivity of Time*

Most of the large secular increase in earnings, which stimulated the development of the labour–leisure analysis, resulted from an increase in the productivity of working time due to the growth in human and physical capital, technological progress and other factors. Since a rise in earnings resulting from an increase in productivity has both income and substitution

[1] See *op. cit.*, Chapter VIII. Recreation commodities presumably have relatively large forgone earnings.

[2] See W. J. Corbett and D. C. Hague, "Complementarity and the Excess Burden of Taxation," *Review of Economic Studies*, Vol. XXI (1953–54); also A. C. Harberger, "Taxation, Resource Allocation and Welfare," in the *Role of Direct and Indirect Taxes in the Federal Revenue System* (Princeton University Press, 1964).

[3] The effect on earnings is more difficult to determine because, by assumption, time intensive commodities have smaller costs per unit time than other commodities. A shift towards the former would, therefore, raise hourly earnings, which would partially and perhaps more than entirely offset the reduction in hours worked. Incidentally, this illustrates how the productivity of hours worked is influenced by the consumption set chosen.

effects, the secular decline in hours worked appeared to be evidence that the income effect was sufficiently strong to swamp the substitution effect.

The secular growth in capital and technology also improved the productivity of consumption time: supermarkets, automobiles, sleeping pills, safety and electric razors, and telephones are a few familiar and important examples of such developments. An improvement in the productivity of consumption time would change relative commodity prices and increase full income, which in turn would produce substitution and income effects. The interesting point is that a very different interpretation of the observed decline in hours of work is suggested because these effects are precisely the opposite of those produced by improvements in the productivity of working time.

Assume a uniform increase only in the productivity of consumption time, which is taken to mean a decline in all t_i, time required to produce a unit of Z_i, by a common percentage. The relative prices of commodities with large forgone earnings would fall, and substitution would be induced towards these and away from other commodities, causing hours of work also to fall. Since the increase in productivity would also produce an income effect,[1] the demand for commodities would increase, which, in turn, would induce an increased demand for goods. But since the productivity of working time is assumed not to change, more goods could be obtained only by an increase in work. That is, the higher real income resulting from an advance in the productivity of consumption time would cause hours of work to *increase*.

Consequently, an emphasis on the secular increase in the productivity of consumption time would lead to a very different interpretation of the secular decline in hours worked. Instead of claiming that a powerful income effect swamped a weaker substitution effect, the claim would have to be that a powerful substitution effect swamped a weaker income effect.

Of course, the productivity of both working and consumption time increased secularly, and the true interpretation is somewhere between these extremes. If both increased at the same rate there would be no change in relative prices, and thus no substitution effect, because the rise in l_i induced by one would exactly offset the decline in t_i induced by the other, marginal forgone earnings (l_it_i) remaining unchanged. Although the income effects would tend to offset each other too, they would do so completely only if the income elasticity of demand for time-intensive commodities was equal to unity. Hours worked would decline if it was above and increase if it was below unity.[2] Since these commodities have probably on

[1] Full money income would be unaffected if it were achieved by using all time at pure work activities. If other uses of time were also required it would tend to increase. Even if full money income were unaffected, however, full real income would increase because prices of the Z_i would fall.

[2] So the "Knight" view that an increase in income would increase "leisure" is not necessarily true, even if leisure were a superior good and even aside from Robbins' emphasis on the substitution effect (see L. Robbins, "On the Elasticity of Demand for Income in Terms of Effort," *Economica*

the whole been luxuries, such an increase in income would tend to reduce hours worked.

The productivity of working time has probably advanced more than that of consumption time, if only because of familiar reasons associated with the division of labour and economies of scale.[1] Consequently, there probably has been the traditional substitution effect towards and income effect away from work, as well as an income effect away from work because time-intensive commodities were luxuries. The secular decline in hours worked would only imply therefore that the combined income effects swamped the substitution effect, not that the income effect of an advance in the productivity of working time alone swamped its substitution effect.

Cross-sectionally, the hours worked of males have generally declined less as incomes increased than they have over time. Some of the difference between these relations is explained by the distinction between relevant and reported incomes, or by interdependencies among the hours worked by different employees;[2] some is probably also explained by the distinction between working and consumption productivity. There is a presumption that persons distinguished cross-sectionally by money incomes or earnings differ more in working than consumption productivity because they are essentially distinguished by the former. This argument does not apply to time series because persons are distinguished there by calendar time, which in principle is neutral between these productivities. Consequently, the traditional substitution effect towards work is apt to be greater cross-sectionally, which would help to explain why the relation between the income and hours worked of men is less negatively sloped there, and be additional evidence that the substitution effect for men is not weak.[3]

Productivity in the service sector in the United States appears to have advanced more slowly, at least since 1929, than productivity in the goods sector.[4] Service industries like retailing, transportation, education and health, use a good deal of the time of households that never enter into input, output and price series, or therefore into measures of productivity. Incorporation of such time into the series and consideration of changes in its productivity would contribute, I believe, to an understanding of the apparent differences in productivity advance between these sectors.

An excellent example can be found in a recent study of productivity

[1] Wesley Mitchell's justly famous essay "The Backward Art of Spending Money" spells out some of these reasons (see the first essay in the collection, *The Backward Art of Spending Money and Other Essays* (New York: McGraw-Hill, 1932)).

[2] A. Finnegan does find steeper cross-sectional relations when the average incomes and hours of different occupations are used (*see* his "A Cross-Sectional Analysis of Hours of Work," *Journal of Political Economy* (October, 1962)).

[3] Note that Mincer has found a very strong substitution effect for women (see his "Labor Force Participation of Married Women," *op. cit.*).

[4] See the essay by Victor Fuchs, "Productivity Trends in the Goods and Service Sectors, 1929–61: A Preliminary Survey," N.B.E.R. Occasional Paper, October 1964.

trends in the barbering industry in the United States.[1] Conventional productivity measures show relatively little advance in barbers' shops since 1929, yet a revolution has occurred in the activities performed by these shops. In the 1920s shaves still accounted for an important part of their sales, but declined to a negligible part by the 1950s because of the spread of home safety and electric razors. Instead of travelling to a shop, waiting in line, receiving a shave and continuing to another destination, men now shave themselves at home, saving travelling, waiting and even some shaving time. This considerable advance in the productivity of shaving nowhere enters measures for barbers' shops. If, however, a productivity measure for general barbering activities, including shaving, was constructed, I suspect that it would show an advance since 1929 comparable to most goods.[2]

(c) *Income Elasticities*

Income elasticities of demand are often estimated cross-sectionally from the behaviour of families or other units with different incomes. When these units buy in the same market-place it is natural to assume that they face the same prices of goods. If, however, incomes differ because earnings do, and cross-sectional income differences are usually dominated by earnings differences, commodities prices would differ systematically. All commodities prices would be higher to higher-income units because their forgone earnings would be higher (which means, incidentally, that differences in real income would be less than those in money income), and the prices of earnings-intensive commodities would be unusually so.

Cross-sectional relations between consumption and income would not therefore measure the effect of income alone, because they would be affected by differences in relative prices as well as in incomes.[3] The effect of income would be underestimated for earnings-intensive and overestimated for other commodities, because the higher relative prices of the former would cause a substitution away from them and towards the latter. Accordingly, the income elasticities of demand for " leisure," unproductive and time-intensive commodities would be under-stated, and for " work," productive and other goods-intensive commodities over-stated by cross-sectional estimates. Low apparent income elasticities of earnings-intensive commodities and high apparent elasticities of other commodities may simply be illusions resulting from substitution effects.[4]

[1] See J. Wilburn, " Productivity Trends in Barber and Beauty Shops," mimeographed report, N.B.E.R., September 1964.

[2] The movement of shaving from barbers' shops to households illustrates how and why even in urban areas households have become " small factories." Under the impetus of a general growth in the value of time they have been encouraged to find ways of saving on travelling and waiting time by performing more activities themselves.

[3] More appropriate income elasticities for several commodities are estimated in Mincer, " Market Prices . . .," *op. cit.*

[4] In this connection note that cross-sectional data are often preferred to time-series data in estimating income elasticities precisely because they are supposed to be largely free of co-linearity

Moreover, according to our theory demand depends also on the importance of earnings as a source of income. For if total income were held constant an increase in earnings would create only substitution effects: away from earnings-intensive and towards goods-intensive commodities. So one unusual implication of the analysis that can and should be tested with available budget data is that the source of income may have a significant effect on consumption patterns. An important special case is found in comparisons of the consumption of employed and unemployed workers. Unemployed workers not only have lower incomes but also lower forgone costs, and thus lower relative prices of time and other earnings-intensive commodities. The propensity of unemployed workers to go fishing, watch television, attend school and so on are simply vivid illustrations of the incentives they have to substitute such commodities for others.

One interesting application of the analysis is to the relation between family size and income.[1] The traditional view, based usually on simple correlations, has been that an increase in income leads to a reduction in the number of children per family. If, however, birth-control knowledge and other variables were held constant economic theory suggests a positive relation between family size and income, and therefore that the traditional negative correlation resulted from positive correlations between income, knowledge and some other variables. The data I put together supported this interpretation, as did those found in several subsequent studies.[2]

Although positive, the elasticity of family size with respect to income is apparently quite low, even when birth-control knowledge is held constant. Some persons have interpreted this (and other evidence) to indicate that family-size formation cannot usefully be fitted into traditional economic analysis.[3] It was pointed out, however, that the small elasticity found for children is not so inconsistent with what is found for goods as soon as quantity and quality income elasticities are distinguished.[4] Increased expenditures on many goods largely take the form of increased quality–expenditure per pound, per car, etc.—and the increase in quantity is modest. Similarly, increased expenditures on children largely take the form of increased expenditures per child, while the increase in number of children is very modest.

---

between prices and incomes (see, *e.g.*, J. Tobin, "A Statistical Demand Function for Food in the U.S.A.," *Journal of the Royal Statistical Society*, Series A (1950)).

[1] Biases in cross-sectional estimates of the demand for work and leisure were considered in the last section.

[2] See G. S. Becker, "An Economic Analysis of Fertility," *Demographic and Economic Change in Developed Countries* (N.B.E.R. Conference Volume, 1960); R. A. Easterlin, "The American Baby Boom in Historical Perspective," *American Economic Review* (December 1961); I. Adelman, "An Econometric Analysis of Population Growth," *American Economic Review* (June 1963); R. Weintraub, "The Birth Rate and Economic Development: An Empirical Study," *Econometrica* (October 1962); Morris Silver, *Birth Rates, Marriages, and Business Cycles* (unpublished Ph.D. dissertation, Columbia University, 1964); and several other studies; for an apparent exception, see the note by D. Freedman, "The Relation of Economic Status to Fertility," *American Economic Review* (June 1963).

[3] See, for example, Duesenberry's comment on Becker, *op. cit.*

[4] See Becker, *op. cit.*

Nevertheless, the elasticity of demand for number of children does seem somewhat smaller than the quantity elasticities found for many goods. Perhaps the explanation is simply the shape of indifference curves; one other factor that may be more important, however, is the increase in forgone costs with income.[1] Child care would seem to be a time-intensive activity that is not "productive" (in terms of earnings) and uses many hours that could be used at work. Consequently, it would be an earnings-intensive activity, and our analysis predicts that its relative price would be higher to higher-income families.[2] There is already some evidence suggesting that the positive relation between forgone costs and income explains why the apparent quantity income elasticity of demand for children is relatively small. Mincer found that cross-sectional differences in the forgone price of children have an important effect on the number of children.[3]

(d) *Transportation*

Transportation is one of the few activities where the cost of time has been explicitly incorporated into economic discussions. In most benefit-cost evaluations of new transportation networks the value of the savings in transportation time has tended to overshadow other benefits.[4] The importance of the value placed on time has encouraged experiment with different methods of determination: from the simple view that the value of an hour equals average hourly earnings to sophisticated considerations of the distinction between standard and overtime hours, the internal and external margins, etc.

The transport field offers considerable opportunity to estimate the marginal productivity or value of time from actual behaviour. One could, for example, relate the ratio of the number of persons travelling by aeroplane to those travelling by slower mediums to the distance travelled (and, of course, also to market prices and incomes). Since relatively more people use faster mediums for longer distances, presumably largely because of the greater importance of the saving in time, one should be able to estimate a marginal value of time from the relation between medium and distance travelled.[5]

[1] In *Ibid.*, p. 214 fn. 8, the relation between forgone costs and income was mentioned but not elaborated.

[2] Other arguments suggesting that higher-income families face a higher price of children have generally confused price with quality (see *ibid.*, pp. 214–15).

[3] See Mincer, "Market Prices . . .," *op. cit.* He measures the price of children by the wife's potential wage-rate, and fits regressions to various cross-sectional data, where number of children is the dependent variable, and family income and the wife's potential wage-rate are among the independent variables.

[4] See, for example, H. Mohring, "Land Values and the Measurement of Highway Benefits," *Journal of Political Economy* (June 1961).

[5] The only quantitative estimate of the marginal value of time that I am familiar with uses the relation between the value of land and its commuting distance from employment (see *ibid.*). With many assumptions I have estimated the marginal value of time of those commuting at about 40% of [illegible] hourly earnings. It is not clear whether this value is so low because of errors in these

Another transportation problem extensively studied is the length and mode of commuting to work.[1] It is usually assumed that direct commuting costs, such as train fare, vary positively and that living costs, such as space, vary negatively with the distance commuted. These assumptions alone would imply that a rise in incomes would result in longer commutes as long as space (" housing ") were a superior good.[2]

A rise in income resulting at least in part from a rise in earnings would, however, increase the cost of commuting a given distance because the forgone value of the time involved would increase. This increase in commuting costs would discourage commuting in the same way that the increased demand for space would encourage it. The outcome depends on the relative strengths of these conflicting forces: one can show with a few assumptions that the distance commuted would increase as income increased if, and only if, space had an income elasticity greater than unity.

For let Z_1 refer to the commuting commodity, Z_2 to other commodities, and let

$$Z_1 = f_1(x, t) \qquad (22)$$

where t is the time spent commuting and x is the quantity of space used. Commuting costs are assumed to have the simple form $a + l_1 t$, where a is a constant and l_1 is the marginal forgone cost per hour spent commuting. In other words, the cost of time is the only variable commuting cost. The cost per unit of space is $p(t)$, where by assumption $p' < 0$. The problem is to maximise the utility function

$$U = U(x, t, Z_2) \qquad (23)$$

subject to the resource constraint

$$a + l_1 t + px + h(Z_2) = S \qquad (24)$$

If it were assumed that $U_t = 0$—commuting was neither enjoyable nor irksome—the main equilibrium condition would reduce to

$$l_1 + p'x = 0 \text{ [3]} \qquad (25)$$

which would be the equilibrium condition if households simply attempt to minimise the sum of transportation and space costs.[4] If $l_1 = kS$, where k

[1] See L. N. Moses and H. F. Williamson, " Value of Time, Choice of Mode, and the Subsidy Issue in Urban Transportation," *Journal of Political Economy* (June 1963), R. Muth, " Economic Change and Rural–Urban Conversion," *Econometrica* (January 1961), and J. F. Kain, *Commuting and the Residential Decisions of Chicago and Detroit Central Business District Workers* (April 1963).

[2] See Muth, *op. cit.*

[3] If $U_t \neq 0$, the main equilibrium condition would be

$$\frac{U_t}{U_x} = \frac{l_1 + p'x}{p}$$

Probably the most plausible assumption is that $U_t < 0$, which would imply that $l_1 + p'x < 0$.

10

is a constant, the effect of a change in full income on the time spent commuting can be found by differentiating equation (25) to be

$$\frac{\partial t}{\partial S} = \frac{k(\epsilon_x - 1)}{p''x} \qquad (26)$$

where ϵ_x is the income elasticity of demand for space. Since stability requires that $p'' > 0$, an increase in income increases the time spent commuting if, and only if, $\epsilon_x > 1$.

In metropolitan areas of the United States higher-income families tend to live further from the central city,[1] which contradicts our analysis if one accepts the traditional view that the income elasticity of demand for housing is less than unity. In a definitive study of the demand for housing in the United States, however, Margaret Reid found income elasticities greater than unity.[2] Moreover, the analysis of distance commuted incorporates only a few dimensions of the demand for housing; principally the demand for outdoor space. The evidence on distances commuted would then only imply that outdoor space is a "luxury," which is rather plausible [3] and not even inconsistent with the traditional view about the total elasticity of demand for housing.

(e) *The Division of Labour Within Families*

Space is too limited to do more than summarise the main implications of the theory concerning the division of labour among members of the same household. Instead of simply allocating time efficiently among commodities, multi-person households also allocate the time of different members. Members who are relatively more efficient at market activities would use less of their time at consumption activities than would other members. Moreover, an increase in the relative market efficiency of any member would effect a reallocation of the time of all other members towards consumption activities in order to permit the former to spend more time at market activities. In short, the allocation of the time of any member is greatly influenced by the opportunities open to other members.

IV. Substitution Between Time and Goods

Although time and goods have been assumed to be used in fixed proportions in producing commodities, substitution could take place because different commodities used them in different proportions. The assumption of fixed proportions is now dropped in order to include many additional implications of the theory.

It is well known from the theory of variable proportions that households

[1] For a discussion, including many qualifications, of this proposition see L. F. Schnore, "The Socio-Economic Status of Cities and Suburbs," *American Sociological Review* (February 1963).

[2] See her *Housing and Income* (University of Chicago Press, 1962), p. 6 and *passim*.

[3] According to Reid, the elasticity of demand for indoor space is less than unity (*ibid.*, Chapter 12). If her total elasticity is accepted this suggests that outdoor space has an elasticity exceeding unity.

would minimise costs by setting the ratio of the marginal product of goods to that of time equal to the ratio of their marginal costs.[1] A rise in the cost of time relative to goods would induce a reduction in the amount of time and an increase in the amount of goods used per unit of each commodity. Thus, not only would a rise in earnings induce a substitution away from earnings-intensive commodities but also a substitution away from time and towards goods in the production of each commodity. Only the first is (implicitly) recognised in the labour–leisure analysis, although the second may well be of considerable importance. It increases one's confidence that the substitution effect of a rise in earnings is more important than is commonly believed.

The change in the input coefficients of time and goods resulting from a change in their relative costs is defined by the elasticity of substitution between them, which presumably varies from commodity to commodity. The only empirical study of this elasticity assumes that recreation goods and " leisure " time are used to produce a recreation commodity.[2] Definite evidence of substitution is found, since the ratio of leisure time to recreation goods is negatively related to the ratio of their prices. The elasticity of substitution appears to be less than unity, however, since the share of leisure in total factor costs is apparently positively related to its relative price.

The incentive to economise on time as its relative cost increases goes a long way towards explaining certain broad aspects of behaviour that have puzzled and often disturbed observers of contemporary life. Since hours worked have declined secularly in most advanced countries, and so-called " leisure " has presumably increased, a natural expectation has been that " free " time would become more abundant, and be used more " leisurely " and " luxuriously." Yet, if anything, time is used more carefully to-day than a century ago.[3] If there was a secular increase in the productivity of working time relative to consumption time (see Section III (*b*)) there would be an increasing incentive to economise on the latter because of its greater expense (our theory emphatically cautions against calling such time " free "). Not surprisingly, therefore, it is now kept track of and used more carefully than in the past.

Americans are supposed to be much more wasteful of food and other

[1] The cost of producing a given amount of commodity Z_i would be minimised if

$$\frac{\partial f_i/\partial x_i}{\partial f_i/\partial T_i} = \frac{P_i}{\partial L/\partial T_i}$$

If utility were considered an indirect function of goods and time rather than simply a direct function of commodities the following conditions, among others, would be required to maximise utility:

$$\frac{\partial U/\partial x_i}{\partial U/\partial T_i} = \frac{\partial Z_i/\partial x_i}{\partial Z_i/\partial T_i} = \frac{p_i}{\partial L/\partial T}$$

which are exactly the same conditions as above. The ratio of the marginal utility of x_i to that of T_i depends only on f_i, x_i and T_i, and is thus independent of other production functions, goods and time. In other words, the indirect utility function is what has been called " weakly separable " (see R. Muth, " Household Production and Consumer Demand Functions," unpublished manuscript).

[2] See Owen, *op. cit.*, Chapter X.

[3] See, for example, de Grazia, *op. cit.*, Chapter IV.

goods than persons in poorer countries, and much more conscious of time: they keep track of it continuously, make (and keep) appointments for specific minutes, rush about more, cook steaks and chops rather than time-consuming stews and so forth.[1] They are simultaneously supposed to be wasteful—of material goods—and overly economical—of immaterial time. Yet both allegations may be correct and not simply indicative of a strange American temperament because the market value of time is higher relative to the price of goods there than elsewhere. That is, the tendency to be economical about time and lavish about goods may be no paradox, but in part simply a reaction to a difference in relative costs.

The substitution towards goods induced by an increase in the relative cost of time would often include a substitution towards more expensive goods. For example, an increase in the value of a mother's time may induce her to enter the labour force and spend less time cooking by using pre-cooked foods and less time on child-care by using nurseries, camps or baby-sitters. Or barbers' shops in wealthier sections of town charge more and provide quicker service than those in poorer sections, because waiting by barbers is substituted for waiting by customers. These examples illustrate that a change in the quality of goods [2] resulting from a change in the relative cost of goods may simply reflect a change in the methods used to produce given commodities, and not any corresponding change in *their* quality.

Consequently, a rise in income due to a rise in earnings would increase the quality of goods purchased not only because of the effect of income on quality but also because of a substitution of goods for time; a rise in income due to a rise in property income would not cause any substitution, and should have less effect on the quality of goods. Put more dramatically, with total income held constant, a rise in earnings should increase while a rise in property income should decrease the quality chosen. Once again, the composition of income is important and provides testable implications of the theory.

One analytically interesting application of these conclusions is to the recent study by Margaret Reid of the substitution between store-bought and home-delivered milk.[3] According to our approach, the cost of inputs into the commodity " milk consumption at home " is either the sum of the price of milk in the store and the forgone value of the time used to carry it home or simply the price of delivered milk. A reduction in the price of store relative to delivered milk, the value of time remaining constant, would reduce the cost of the first method relatively to the second, and shift production towards the first. For the same reason a reduction in the value of time, market prices

[1] For a comparison of the American concept of time with others see Edward T. Hall, *The Silent Language* (New York: Doubleday, 1959), Chapter 9.

[2] Quality is usually defined empirically by the amount spent per physical unit, such as pound of food, car or child. See especially S. J. Prais and H. Houthakker, *The Analysis of Family Budgets* (Cambridge, 1955); also my " An Economic Analysis of Fertility," *op. cit.*

[3] See her " Consumer Response to the Relative Price of Store versus Delivered Milk," *Journal*

of milk remaining constant, would also shift production towards the first method.

Reid's finding of a very large negative relation between the ratio of store to delivered milk and the ratio of their prices, income and some other variables held constant, would be evidence both that milk costs are a large part of total production costs and that there is easy substitution between these alternative methods of production. The large, but not quite as large, negative relation with income simply confirms the easy substitution between methods, and indicates that the cost of time is less important than the cost of milk. In other words, instead of conveying separate information, her price and income elasticities both measure substitution between the two methods of producing the same commoditity, and are consistent and plausible.

The importance of forgone earnings and the substitution between time and goods may be quite relevant in interpreting observed price elasticities. A given percentage increase in the price of goods would be less of an increase in commodity prices the more important forgone earnings are. Consequently, even if all commodities had the same true price elasticity, those having relatively important forgone earnings would show lower apparent elasticities in the typical analysis that relates quantities and prices of goods alone.

The importance of forgone earnings differs not only among commodities but also among households for a given commodity because of differences in income. Its importance would change in the same or opposite direction as income, depending on whether the elasticity of substitution between time and goods was less or greater than unity. Thus, even when the true price elasticity of a commodity did not vary with income, the observed price elasticity of goods would be negatively or positively related to income as the elasticity of substitution was less or greater than unity.

The importance of substitution between time and goods can be illustrated in a still different way. Suppose, for simplicity, that only good x and no time was initially required to produce commodity Z. A price ceiling is placed on x, it nominally becomes a free good, and the production of x is subsidised sufficiently to maintain the same output. The increased quantity of x and Z demanded due to the decline in the price of x has to be rationed because the output of x has not increased. Suppose that the system of rationing made the quantity obtained a positive function of the time and effort expended. For example, the quantity of price-controlled bread or medical attention obtained might depend on the time spent in a queue outside a bakery or in a physician's office. Or if an appointment system were used a literal queue would be replaced by a figurative one, in which the waiting was done at "home," as in the Broadway theatre, admissions to hospitals or air travel during peak seasons. Again, even in depressed times the likelihood of obtaining a job is positively related to the time put into job hunting.

Although x became nominally a free good, Z would not be free, because would be greater than the supply (fixed by assumption) if the cost of this time was less than the equilibrium price of Z before the price control. The scrambling by households for the limited supply would increase the time required to get a unit of Z, and thus its cost. Both would continue to increase until the average cost of time tended to the equilibrium price before price control. At that point equilibrium would be achieved because the supply and demand for Z would be equal.

Equilibrium would take different forms depending on the method of rationing. With a literal " first come first served " system the size of the queue (say outside the bakery or in the doctor's office) would grow until the expected cost of standing in line discouraged any excess demand;[1] with the figurative queues of appointment systems, the " waiting " time (say to see a play) would grow until demand was sufficiently curtailed. If the system of rationing was less formal, as in the labour market during recessions, the expected time required to ferret out a scarce job would grow until the demand for jobs was curtailed to the limited supply.

Therefore, price control of x combined with a subsidy that kept its amount constant would not change the average private equilibrium price of Z,[2] but would substitute indirect time costs for direct goods costs.[3] Since, however, indirect costs are positively related to income, the price of Z would be raised to higher-income persons and reduced to lower-income ones, thereby redistributing consumption from the former to the latter. That is, women, the poor, children, the unemployed, etc., would be more willing to spend their time in a queue or otherwise ferreting out rationed goods than would high-earning males.

V. Summary and Conclusions

This paper has presented a theory of the allocation of time between different activities. At the heart of the theory is an assumption that households are producers as well as consumers; they produce commodities by combining inputs of goods and time according to the cost-minimisation rules of the traditional theory of the firm. Commodities are produced in quantities determined by maximising a utility function of the commodity set subject to prices and a constraint on resources. Resources are measured by what is called full income, which is the sum of money income and that forgone or " lost " by the use of time and goods to obtain utility, while commodity prices are measured by the sum of the costs of their goods and time inputs.

[1] In queueing language the cost of waiting in line is a " discouragement " factor that stabilises the queueing scheme (see, for example, D. R. Cox and W. L. Smith, *Queues* (New York: Wiley 1961)).

[2] The social price, on the other hand, would double, for it is the sum of private indirect costs and subsidised direct costs.

[3] Time costs can be criticised from a Pareto optimality point of view because they often result in external diseconomies: *e.g.*, a person joining a queue would impose costs on subsequent joiners. The diseconomies are real, not simply pecuniary, because time is a cost to demanders, but is not revenue to suppliers.

The effect of changes in earnings, other income, goods prices and the productivity of working and consumption time on the allocation of time and the commodity set produced has been analysed. For example, a rise in earnings, compensated by a decline in other income so that full income would be unchanged, would induce a decline in the amount of time used at consumption activities, because time would become more expensive. Partly goods would be substituted for the more expensive time in the production of each commodity, and partly goods-intensive commodities would be substituted for the more expensive time-intensive ones. Both substitutions require less time to be used at consumption, and permit more to be used at work. Since the reallocation of time involves simultaneously a reallocation of goods and commodities, all three decisions become intimately related.

The theory has many interesting and even novel interpretations of, and implications about, empirical phenomena. A few will be summarised here.

A traditional " economic " interpretation of the secular decline in hours worked has stressed the growth in productivity of working time and the resulting income and substitution effects, with the former supposedly dominating. Ours stresses that the substitution effects of the growth in productivity of working and consumption time tended to offset each other, and that hours worked declined secularly primarily because time-intensive commodities have been luxuries. A contributing influence has been the secular decline in the relative prices of goods used in time-intensive commodities.

Since an increase in income partly due to an increase in earnings would raise the relative cost of time and of time-intensive commodities, traditional cross-sectional estimates of income elasticities do not hold either factor or commodity prices constant. Consequently, they would, among other things, be biased downward for time-intensive commodities, and give a misleading impression of the effect of income on the quality of commodities consumed. The composition of income also affects demand, for an increase in earnings, total income held constant, would shift demand away from time-intensive commodities and input combinations.

Rough estimates suggest that forgone earnings are quantitatively important and therefore that full income is substantially above money income. Since forgone earnings are primarily determined by the use of time, considerably more attention should be paid to its efficiency and allocation. In particular, agencies that collect information on the expenditure of money income might simultaneously collect information on the " expenditure " of time. The resulting time budgets, which have not been seriously investigated in most countries, including the United States and Great Britain, should be integrated with the money budgets in order to give a more accurate picture of the size and allocation of full income.

Gary S. Becker

Columbia University.

14

The Economic Journal n° 324 Dec. 1971 Vol. 81

A THEORY OF THE ECONOMICS OF TIME[1]

THIS study represents a theory of consumer behaviour, specifically designed to handle economic problems wherein a time dimension is relevant. Over the years there have been a number of attempts to modify neoclassical consumer theory to deal with problems of this nature, but none of these works achieved the level of sophistication of the traditional approach, whereby testable properties of demand functions are deduced. In part, this has been intentional. These studies were more concerned with specific problems such as the decision to work more or fewer hours [19], the effect of foregone earnings upon consumer choice [1] and the valuation of travel time [10, 17], rather than with the general properties of demand functions. More importantly, however, the restrictions on demand functions derived from neoclassical theory cannot be derived from existing theories of the time dimension in consumer choice. This is due not to any property intrinsic to the time dimension but to the fact that these theories are improperly specified.

The essential features of the model presented in this paper are: (1) utility is a function not only of commodities but also of the time allocated to them; (2) the individual's decision is subject to two resource constraints, a money constraint and a time constraint; and (3) the decision to consume a specified amount of any commodity requires that some minimum amount of time be allocated to it, but the individual may spend more time in that activity if he so desires. Under these specifications, all the implications of neoclassical theory are preserved and many additional results, applicable to situations involving a time dimension, are generated.

I. THE MODEL

Neoclassical consumer theory analyses individual preferences among alternative commodity bundles, $X = (X_1, \ldots, X_n)$, within an attainable set defined by the individual's income and a set of parametric money prices. Once a time dimension is introduced, the field of choice expands considerably: commodities might be consumed one at a time, or concurrently, or pure time might be consumed independently of consumer goods, etc. For simplicity, we shall consider only the case in which goods are consumed one at a time and all the time available to the individual is spent in the consumption of some commodity.

[1] Assistant Professor of Economics, Navy Management Systems Center, U.S. Naval Postgraduate School. This paper is adapted from the author's doctoral dissertation submitted to the University of California, Santa Barbara. The author is indebted to Professor M. Bruce Johnson, W. Douglas Morgan and Lloyd J. Mercer of U.C.S.B. and to Professor William Eadington of the University of Nevada, Reno, for their many valuable comments. However, the author is fully responsible for any errors of omission or commission.

14

Consider a set of commodity bundles, $X = (X_1, \ldots, X_n, T_1, \ldots, T_n)$, in the positive orthant of a Euclidean space of dimension $2n$. The variable, X_i, denotes some quantity of the ith consumption good, while T_i denotes the amount of time allocated to the ith good. Assume the individual possesses a complete, consistent preference ordering among alternative commodity bundles and exhibits rational behaviour. That is, from an attainable set of commodity bundles, $Z(Z \in X)$, the individual will select a bundle, X^* $(X^* \in Z)$, such that $X^* RX^i$ for all $X^i \in Z$, where R is a binary relation which reads "is preferred or indifferent to." The attainable set is defined by the constraints, to be specified below. The individual's preferences are assumed to be representable by a continuous, twice-differentiable real valued utility function, $U(X)$.

Following neoclassical consumer theory, we assume that the consumer receives a fixed positive amount of money income (Y), and chooses to spend all of it on consumption goods during the time period under consideration. Thus, total expenditures on consumption goods must be equal to the money endowment:

$$Y = \sum_{i=1}^{n} P_i X_i, \qquad (1.1)$$

where $P_i \geqslant 0$ is the money price of the ith consumption good.

Similarly, the individual receives a fixed time endowment (T°) equal to the length of the decision period.[1] Since money income is a flow concept, the time endowment must be consistent with that of income. The time constraint requires that the amounts of time allocated to each specific use add up to the time available:

$$T^\circ = \sum_{i=1}^{n} T_i \qquad (1.2)$$

This relationship follows directly from our assumption that goods are consumed one at a time and that all available time is allocated to consumption. It is important to note that (1.1) and (1.2), as specified, are independent of each other. Each represents a resource constraint, but unlike earlier attempts to describe the consumer's time allocation problem, parametric time prices of the goods are absent from the time constraint. This is in keeping

[1] The assumption of fixed endowments of money and time might appear to preclude the ability of the individual to exchange time for money through the work decision. This interpretation would reflect an institutionally determined work week, wherein T° is redefined as the time endowment net of working hours $(T^\circ = \bar{T} - \bar{W})$ and Y redefined to include wage and non-wage income $(Y = I + \bar{w}W$, where $\bar{w}$ is the individual's wage rate).

However, this need not be the case. The model could be modified to include the work decision, either by defining one of the commodities (X_w) as work $(P_w < 0)$, where $P_w X_w$ augments income and $T_w = a_w X_w$ diminishes the time endowment, or by including a pure time commodity, W, in the utility function and in the two resource constraints. In neither case are the qualitative properties of the model altered. The modifications affect the solution vector only indirectly through the structure of the H matrix, defined below. In short, any attention to the work decision tends to obscure the model's more important properties without adding anything substantial. See [6], pp. 130–4, 139–40.

15

16

with our specification of the T_i's as decision variables distinct from the X_i's. The use of time prices would reduce the number of decision variables from $2n$ to n, for the choice of any X_i would determine, by means of the time price, the corresponding T_i. Time prices are excessively rigid and unnecessary.

In reality, the amount of time allocated to the consumption of any commodity is partly a matter of choice and partly a matter of necessity. The fact that the consumption of goods generally requires the expenditure of time as well as money does not place an upper bound upon the amount of time an individual may spend consuming the chosen amount of the good. Thus we shall assume that the choice of a positive amount of any X_i places only a lower bound upon the amount T_i consumed. For simplicity, we shall also assume these relationships to be linear. Mathematically, they take the form of the inequalities

$$T_i \geqslant a_i X_i, \; i = 1, \ldots, n, \qquad (1.3)$$

where a_i may be interpreted as a technologically or institutionally determined minimum amount of time required to consume one unit of X_i. For simplicity, we shall assume that the vector of a_i's, like the vector of prices, is known with certainty.

Henceforth, we shall refer to the n inequalities of (1.3) as the *time consumption constraints*, as distinguished from the *time resource constraint*, equation (1.2). They are specified as inequalities because the individual is free to allocate more than the required amount of time to any activity. Whether the equality is binding or not is a matter of individual preference, although common experience suggests that the constraint will be binding for nearly all individuals in certain activities, due to the nature of these activities. Examples of a binding technological or physical constraint are a round of golf, movies, meals, road congestion, reading a book, etc. Examples of the institutional type of constraint are speed limits, rigid work weeks, banquets, etc. These constraints, be they physical or institutional, must be made explicit in the formal maximisation model; along with (1.1) and (1.2), they define the attainable set of commodity bundles.

A. *First-order Conditions*

The individual's problem of efficiently allocating his time and money resources may be expressed as the maximisation of the Lagrange function,

$$L = U(X_1, \ldots, X_n, T_1, \ldots, T_n) + \lambda(Y - \sum_{i=1}^{n} P_i X_i) + \mu(T^\circ - \sum_{i=1}^{n} T_i) + \sum_{i=1}^{n} K_i(T_i - a_i X_i)$$

where $K_i \geqslant 0$, $i = 1, \ldots, n$, and $\mu, \lambda, > 0$. For simplicity, we shall assume that all commodities and usages of time are consumed in positive amounts, although the model could easily be modified to allow for corner solutions.

16

Under these specifications, the following conditions, plus (1.1) and (1.2), are necessary for maximisation:

$$U_i = \lambda P_i + K_i a_i, \quad i = 1, \ldots, n, \qquad (1.4)$$

$$U_{n+i} = \mu - K_i, \quad i = 1, \ldots, n, \qquad (1.5)$$

$$K_i(T_i - a_i X_i) = 0, \quad i = 1, \ldots, n; \qquad (1.6)$$

that is, either $T_i = a_i X_i$, or $K_i = 0$, $i = 1, \ldots, n$.

The subscripts i and $n + i$ denote the partial derivatives of the utility function with respect to X_i and T_i, respectively. The Lagrangian multipliers, λ and μ, are shadow variables representing the marginal utility of money and the marginal utility of time, respectively. The ratio, μ/λ, the marginal rate of substitution between time and money, may be interpreted as the value of time. The generation of this value as an equilibrium condition is an important feature of the model,[1] but at this juncture the effect of the time consumption constraints upon the equilibrium conditions warrants our full attention. If the time consumption constraint is binding, the first-order conditions of traditional consumer theory cease to apply. In goods space, the marginal rate of substitution between two goods is not equal to their price ratios. Consider diagrammatically a two-good case. For expository purposes, we shall assume the individual's preferences are such that the time consumption constraint is binding between X_2 and T_2, but ineffective between X_1 and T_1. Unfortunately, even the two-good case requires a four-dimensional picture. As an alternative, the interaction among the constraints is illustrated in Figs. 1(*a*), (*b*), (*c*) and (*d*) by means of four cross-sectional diagrams. In Fig. 1(*a*), the budget constraint is truncated at point A, the intersection of the budget line and the line, $X_1 = (T^\circ/a_1) - (a_2X_2/a_1)$.[1] That portion of the budget constraint lying above A represents combinations of X_1 and X_2 which are incompatible with the time constraint. The requirement that X_2 and T_2 be consumed in fixed proportions, illustrated in Fig. 1(*b*), causes non-tangency solutions in both Fig. 1(*a*) (goods space) and Fig. 1(*d*) (time space). In equilibrium, the rate of substitution between X_1 and X_2 is less than the price ratio, while the rate of substitution between T_1 and T_2 exceeds unity (the absolute value of the slope of the time constraint). Although a partial equilibrium analysis of Figs. 1(*a*) and 1(*d*) indicates that the consumer could improve his position by substituting X_2 for X_1 and substituting T_1 for T_2, the two substitutions cannot be made concurrently, because of the time consumption constraint. The total equilibrium generates the shadow variables, λ, μ, K_1 and K_2.

The economic interpretation of this equilibrium position reveals some rather interesting aspects of the problem of time valuation. Dividing through (1.5) by λ, we get

$$\frac{U_{n+i}}{\lambda} = \frac{\mu}{\lambda} - \frac{K_i}{\lambda} \qquad (1.7)$$

[1] Johnson [10] and Oort [17] drew the same conclusion from their respective models.

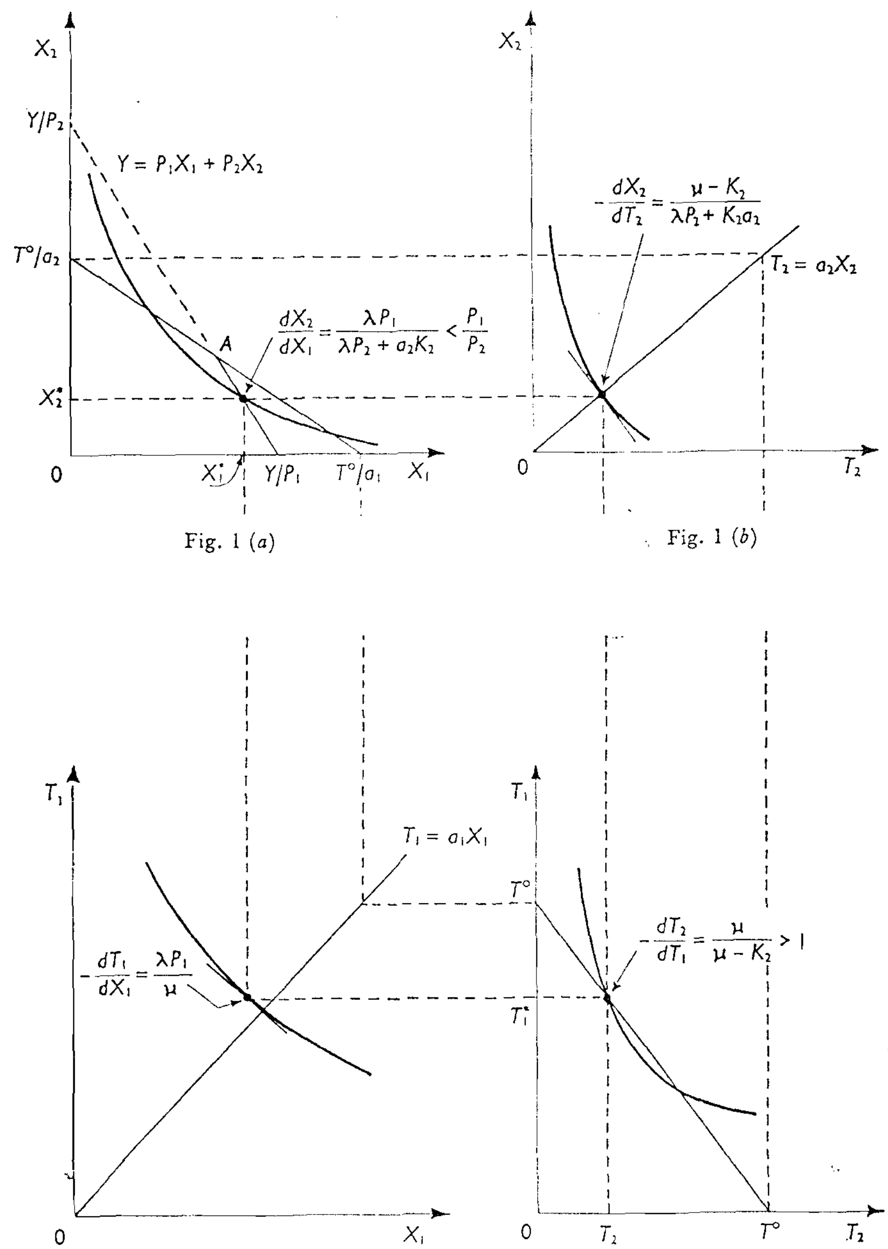


Fig. 1 (a) Fig. 1 (b)

Fig. 1 (c) Fig. 1 (d)

19

The term, U_{n+i}/λ, the marginal rate of substitution of T_i for money, represents the value of time allocated to the consumption of X_i. It is the value of time as a *commodity*, not as a resource. The distinction between the two is crucial. They are equal if, and only if, $K_i = 0$. This condition will prevail if the individual elects to spend more than the required amount of time consuming X_i.[1]

B. *The Value of Time and the Value of Saving Time*

Since each of the time consumption constraints represents the requirement to spend time consuming some commodity, relaxing the ith time consumption constraint is equivalent to saving time in the ith consumption activity. Therefore, K_i may be interpreted as the marginal utility of saving time and the ratio, K_i/λ, the value of saving time. This interpretation acquires greater intuitive appeal in light of the distinction made between the value of time as a resource and the value of time as a commodity. Attributing a positive value to saving time from any activity presupposes that the time saved can be transferred to some alternative usage of greater value. The algebraic difference between the value of time in alternative uses and the value of time in any particular use (U_{n+i}/λ) determines the value of saving time from that activity:

$$\text{Value of Saving Time Consuming } X_i = \frac{\mu}{\lambda} - \frac{U_{n+i}}{\lambda} = \frac{K_i}{\lambda} \tag{1.8}$$

Only this latter value, the value of saving time, has any empirical content. The value of time as a resource derives from its scarcity; it is an opportunity cost which is positive because the time resource is available in a limited amount. However, it makes little sense to consider the value associated with acquiring more of it, for such an increment is not possible, either under the specifications of this model or in reality, in any meaningful sense. Time saving, on the other hand, is both possible and observable in many facets of human activity. Thus the formal distinction between these two values is a most important feature of our model.

C. *Leisure and the Price of Travel Time*

An earlier study by M. Bruce Johnson [10] suggested that the marginal rate of substitution of time for money, μ/λ, is an appropriate theoretical representation of both the value of leisure and the value of travel time. Although this conclusion apparently contradicts ours, it can be demonstrated that Johnson's conclusion follows from a special case of our model. Let us

[1] The equilibrium conditions do not preclude the possibility that $K_i = 0$ and $T_i = a_i X_i$. Strictly speaking, then, the individual decision to spend more than the required amount of time consuming X_i is sufficient but not necessary for U_{n+i}/λ to equal μ/λ. That is, $T_i > a_i X_i$ implies that $K_i = 0$, but $K_i = 0$ does not imply that $T_i > a_i X_i$.

first consider the concept of leisure within the context of our model. Economists have traditionally defined leisure as the residual of work. This practice has proved useful in the theoretical analysis of work, but its usefulness in the analysis of other theoretical questions pertaining to time has more recently been challenged. Addressing the problem of travel time valuation specifically, David Tipping asks:

> How should one distinguish between, and measure the time which is spent in enjoying the consumption of goods and services . . . and the time which is spent in making such final consumption possible? Perhaps we need a category of " intermediate consumption " to cover such activities as travelling to the theatre, doing the shopping . . . [22, p. 848].

The distinction made by Tipping between " leisure " and " intermediate " goods can be formalised in terms of our general model. Leisure is conventionally defined as *free time* which may be used for rest, recreation, etc. This definition suggests a freedom from responsibility, a specific case of which is freedom from work (hence, the traditional economic definition of leisure as non-work). In terms of our model, the primary responsibility of the time resource is with respect to the consumption of goods. Given the solution vector of the n goods, $(X_1^*, \ldots, X_n^*)$, a minimum amount of time, specified by the parameters, $a_1, \ldots, a_n$, must be allocated to their consumption. Freedom from this responsibility, which inheres in the choice to allocate more time to any particular good than is required, thus constitutes leisure. Thus, those goods for which the time consumption constraint is ineffective, *i.e.*, $K_i = 0$, may be classified as leisure goods, while all other goods may be classified as intermediate goods: $L(i) = \{i \mid K_i = 0\}$, denotes leisure goods; $I(X) = \{i \mid K_i > 0\}$, denotes intermediate goods. All time allocated to leisure goods might thus be defined as leisure time:

$$L = \sum_{i \in L(X)} T_i \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1.9)$$

Thus the concept of leisure is developed from the model. The definition is meaningful in an economic sense and is also consistent with a more philosophical concept of leisure.

Returning to Johnson's results, we find that the equilibrium conditions of our model support the conclusion that the value of leisure time is equal to μ/λ. This follows trivially from (1.7) and (1.9). However, they do not in general support the conclusion that the value of commuting time is equal to μ/λ. This would be the case only if the journey to work were a leisure good, but, of course, it is not. However, Johnson was considering not the value of commuting time, but the value of *saving* commuting time, which we shall henceforth call the *price* of time. Under the specifications of his model, the marginal utility of commuting time must equal zero, for commuting time does not appear in the utility function (see [10], p. 138). Given that the

20

21

marginal utility of commuting time is equal to zero, (1.7) indicates that the price of commuting time would indeed equal μ/λ, as Johnson suggested. However, assigning a specific value to any partial derivative of the utility function can add nothing to the analysis. Neither the assumption nor its implication is empirically verifiable, for utility is not measurable in any meaningful sense and μ/λ cannot be related to any set of empirical data. On the other hand, a relationship can be derived between the measure, K_i/λ, and empirically observable data. This relationship is but one of many meaningful qualitative results that follow from our specification of the consumer's time allocation problem. It is to these results that we now turn.

II. Comparative Statics and the Demand Function

From our equilibrium conditions, (1.1), (1.2), (1.4), (1.5) and (1.6), the quantity demanded of each commodity may be derived as a function of the parameters of the system:

$$X_i = X_i(P_1, \ldots, P_n, a_1, \ldots, a_n, Y, T^\circ), \; i = 1, \ldots, n$$

In order to derive refutable qualitative properties of these demand functions, we shall assume, following neoclassical theory, that the consumer is in the neighbourhood of a constrained optimum and consider the effects upon quantity demanded of small changes in the parameters. Because we are considering only the properties of an optimal position, ineffective constraints may be ignored, whereby the total differential of the first-order conditions reduces to:

$$\sum_{j=1}^{n} U_{i\,j} dX_j + \sum_{j=1}^{n} U_{i\;n+j}\, dT_j - P_i d\lambda - a_i dK_i = \lambda dP_i + K_i da_i, \quad i = 1, \ldots, n, \quad . \quad (2.1)$$

$$\sum_{j=1}^{n} U_{n+i\;j} dX_j + \sum_{j=1}^{n} U_{n+i\;n+j} dT_j - d\mu + dK_i = 0, \; i = 1, \ldots, n, \quad . \quad (2.2)$$

$$dY = \sum_{i=1}^{n} P_i dX_i + \sum_{i=1}^{n} X_i dP_i, \quad . \; . \; . \; . \quad (2.3)$$

$$\sum_{i=1}^{n} dT_i = dT^\circ = 0, \quad . \; . \; . \; . \; . \quad (2.4)$$

$$-a_i dX_i + dT_i = X_i da_i, \;\; i \in I(X), \quad . \quad (2.5)$$

where $K_i = dK_i = 0$, for all $i \in L(X)$.[1] In the limiting case in which all

[1] The set of commodities, $\{X_i \mid K_i = 0 \textit{ and } T_i = a_i X_i\}$ is assumed to be empty and, further, it is assumed that small changes in the parameters do not alter the composition of the sets $I(X)$ and $L(X)$. This assumption does not appear to be excessively restrictive since, for most applications of the theory, the commodities under analysis could be clearly labelled as leisure or intermediate goods. Strictly speaking, the total differential of (1.6) is

$$-a_i K_i dX_i + K_i dT_i = K_i X_i da_i + (a_i X_i - T_i) dK_i, \; i = 1, \ldots, n. \quad . \quad (2.5')$$

In the case of leisure goods, every term in the equation is equal to zero; in the case of intermediate goods, $T_i - a_i X_i = 0$, whereupon the K_i's cancel out, leaving (2.5).

21

22

goods are intermediate goods, this system may be rewritten as $Hy = c$, where

$$H = \begin{bmatrix}
U_{11} \cdots\cdots U_{1n} & U_{1n+1} \cdots\cdots U_{1\,2n} & -P_1 & 0 & -a_1 \quad \ddots \quad 0 \\
\vdots \qquad\quad \vdots & \qquad\qquad\quad \vdots & \vdots & \vdots & \\
U_{n1} \cdots\cdots U_{nn} & U_{nn+1} \cdots\cdots U_{n\,2n} & -P_n & 0 & 0 \qquad -a_n \\
U_{n+11} \cdots\cdots U_{2n\,n} & U_{n+1n+1} \cdots\cdots U_{n+1\,2n} & 0 & -1 & 1 \qquad 0 \\
\vdots \qquad\quad \vdots & \vdots \qquad\quad \vdots & \vdots & \vdots & \ddots \\
U_{2n\,1} \cdots\cdots U_{n+1\,n} & U_{2n\,n+1} \cdots\cdots U_{2n\,2n} & 0 & -1 & 0 \qquad 1 \\
-P_1 \cdots\cdots -P_n & 0 \cdots\cdots 0 & 0 & \cdots\cdots & \cdots\cdots 0 \\
0 \cdots\cdots 0 & -1 \cdots\cdots -1 & \vdots & & \vdots \\
-a_1 \qquad 0 & 1 \qquad 0 & \vdots & & \vdots \\
\ddots & \ddots & \vdots & & \vdots \\
0 \qquad -a_n & 0 \qquad 1 & 0 & \cdots\cdots & \cdots\cdots 0
\end{bmatrix}$$

$$y = \begin{bmatrix} dX_1 \\ \vdots \\ dX_n \\ dT_1 \\ \vdots \\ dT_n \\ d\lambda \\ d\mu \\ dK_1 \\ \vdots \\ dK_n \end{bmatrix}_{3n+2 \times 1}, \text{ and } c = \begin{bmatrix} \lambda dP_1 + K_1 da_1 \\ \vdots \\ \lambda dP_n + K_n da_n \\ 0 \\ \vdots \\ 0 \\ -dY + \sum_{i=1}^{n} X_i dP_i \\ 0 \\ X_1\, da_1 \\ \vdots \\ X_n\, da_n \end{bmatrix}_{3n+2 \times 1}.$$

In this case, the rank of H is $3n + 2$, although there will normally be leisure goods as well as intermediate goods. In this general case, the $2n + 2 + i$th row and column of H as well as the $2n + 2 + i$th element y and c are deleted, for all $i \in L(X)$, and the rank of H reduces to $2n + 2 + h$, h representing the number of intermediate goods in the neighbourhood of solution.

22

23

Using Cramer's rule, this system may be solved for the differentials, dX_i, $i = 1, \ldots, n$, $d\lambda$, and dK_i, $i \in I(X)$.[1] They are as follows:

$$dX_i = \lambda \sum_{j=1}^{n} \frac{D_{ji}}{D} dP_j + \sum_{j=1}^{n} K_j \frac{D_{ji}}{D} da_j$$
$$+ (-dY + \sum_{j=1}^{n} X_j dP_j) \frac{D_{2n+1\,i}}{D}$$
$$+ \sum_{j \in I(X)} X_j \frac{D_{2n+2+j\,i}}{D} da_j,\ i = 1, \ldots, n \quad . \quad (2.6)$$

$$d\lambda = \sum_{j=1}^{n} (\lambda dP_j + K_j da_j) \frac{D_{j\,2n+1}}{D}$$
$$+ (-dY + \sum_{j=1}^{n} X_j dP_j) \frac{D_{2n+1\,2n+1}}{D}$$
$$+ \sum_{j \in I(X)} X_j \frac{D_{2n+2+j\,2n+1}}{D} da_j \quad . \quad . \quad . \quad . \quad . \quad (2.7)$$

$$dK_i = \sum_{j=1}^{n} (\lambda dP_j + K_j da_j) \frac{D_{j\,2n+2+i}}{D}$$
$$+ (-dY + \sum_{j=1}^{n} X_j dP_j) \frac{D_{2n+1\,2n+2+i}}{D}$$
$$+ \sum_{j \in I(X)} X_j \frac{D_{2n+2+j\,2n+2+i}}{D} da_j,\ i \in I(X) \quad . \quad (2.8)$$

where D is the determinant of H and D_{rk}, r, $k = 1, \ldots, 2n + 2 + h$ is the minor determinant of H, formed by deleting the rth row and the kth column. As a condition of optimisation, H must be negative semi-definite (see [5], or [13], p. 53 ff.). Thus

$$\frac{D_{rr}}{D} \leqslant 0,\ r = 1, \ldots, 2n \quad . \quad . \quad . \quad . \quad (2.9)$$

Moreover, since H is symmetrical,

$$\frac{D_{rk}}{D} = \frac{D_{kr}}{D},\ r,\ k = 1, \ldots, 2n + 2 + h \quad . \quad . \quad (2.10)$$

A. *Price and Income Effects, and the Substitution Theorem*

All of the meaningful theorems derived from the neoclassical theory of demand retain their validity despite the fact that a time dimension has been added to the consumer's decision.

It follows directly from (2.6) that

$$\frac{\partial X_i}{\partial P_j} = \lambda \frac{D_{ji}}{D} + X_j \frac{D_{2n+1\,i}}{D} \quad . \quad . \quad . \quad . \quad (2.11)$$

and

$$\frac{D_{2n+1\,i}}{D} = -\frac{\partial X_i}{\partial Y} \quad . \quad . \quad . \quad . \quad . \quad (2.12)$$

[1] The system may also be solved for the differentials, dT_i, $i = 1, \ldots, n$, and $d\mu$, but these variables are not used in the derivation of any theorems. They are, therefore, not included.

23

24

Substituting (2.12) into (2.11), and rearranging, we get

$$\frac{\partial X_i}{\partial P_j} + X_j \frac{\partial X_i}{\partial Y} = \lambda \frac{D_{ji}}{D} \quad . \quad . \quad . \quad . \quad (2.13)$$

In the special case where $i = j$, it follows from (2.9) that

$$\frac{\partial X_i}{\partial P_i} + X_i \frac{\partial X_i}{\partial Y} = \lambda \frac{D_{ii}}{D} \leqslant 0 \quad . \quad . \quad . \quad (2.14)$$

This inequality is the well-known Slutsky equation [21], and can be tested empirically. However, the left side of the inequality does not represent the substitution effect, except in the trivial case in which all goods are leisure goods.[1] In light of the fact that the consumer's equilibrium position is no longer characterised by a tangency between an indifference surface and a budget hyperplane, this result should not be surprising. Moreover, strictly from the standpoint of hypothetical restrictions on demand functions, it would not appear to be particularly important, since (2.14) is valid in any case. Nevertheless, it demonstrates the invalidity of applying the substitution theorem to situations in which the time dimension is relevant (see Becker [1], p. 304 ff.).

B. *Homogeneity and Symmetry*

Homogeneity of degree zero in prices and income, as well as the Slutsky symmetry conditions, follow from (2.11). From the theorem of expansion by alien co-factors, it follows that

$$\sum_{i=1}^{n} P_i \frac{D_{ij}}{D} = 0, j = 1, \ldots, n \quad . \quad . \quad . \quad (2.15)$$

By direct substitution of (2.13), this becomes

$$\sum_{i=1}^{n} P_i\left[\frac{\partial X_j}{\partial P_i} + X_i \frac{\partial X_j}{\partial Y}\right] = \sum_{i=1}^{n} P_i \frac{\partial X_j}{\partial P_i} + Y \frac{\partial X_i}{\partial Y} = 0 \qquad (2.16)$$

[1] Suppose the price change were accompanied by an income change which kept the individual at the same level of utility. Then

(i) $dU = \sum_{i=1}^{n} U_i dX_i + \sum_{i=1}^{n} U_{n+i} dT_i = 0.$

Substituting (1.4) and (1.5), we get

$$\text{(ii)} \quad dU = \sum_{i=1}^{n} (\lambda P_i + a_i K_i) dX_i + \sum_{i=1}^{n} (\mu - K_i) dT_i = 0$$
$$= \lambda \sum_{i=1}^{n} P_i dX_i + \sum_{i=1}^{n} a_i K_i dX_i + \mu \sum_{i=1}^{n} dT_i - \sum_{i=1}^{n} K_i dT_i.$$

Substituting (2.3), (2.4) and (2.5'), and dividing through by λ, we get

(iii) $-dY + \sum_{j=1}^{n} X_i dP_i = -\frac{1}{\lambda} \sum_{i=1}^{n} K_i X_i da_i.$

It follows from (2.6) and (2.14) that the term, $\lambda D_{ii}/D$, represents the substitution effect if and only if $-dY + \sum_{i=1}^{n} X_i dP_i$ goes to zero when utility is held constant. But it follows from (iii) that this condition holds only in the trivial case in which all goods are leisure goods.

24

which, by Euler's theorem on homogeneous functions, is equivalent to $X_i(P_1, \ldots, P_n, Y) = X_i(\beta P_1, \ldots, \beta P_n, \beta Y), i = 1, \ldots, n, \beta > 0$. Equation (2.16) also implies that the sum of the elasticities of any good with respect to each price is equal in absolute value, but opposite in sign to the income elasticity of demand for that good. Dividing through (2.16) by X_j yields this result, which adds another n restrictions on the demand functions.

The symmetry conditions follow from (2.13) and (2.10):

$$\frac{\partial X_i}{\partial P_j} + X_j \frac{\partial X_i}{\partial Y} = \frac{\partial X_j}{\partial P_i} + X_i \frac{\partial X_j}{\partial Y}, \; i, j = 1, \ldots, n \qquad (2.17)$$

C. *Demand as a Function of Time Costs*

Economic literature is full of examples of demand curves depicting quantity demanded as a function of per unit time costs. This type of demand function may be found in models of highway congestion in which an equilibrium level of road usage critically depends on anticipated time costs [9, 23]. In addition, queuing models have acknowledged that the distribution of demand over the course of a day responds not only to explicit price changes but also to the price changes implicit in the variation of waiting time [15]. This notion has been discussed in the context of the demand for telephone communication [15] and the effect of queuing costs upon the demand for air transport [7].

The type of demand function commonly specified is in terms of own price and own " time price." In our notation,

$$X_i = X_i(\bar{P}_1, \ldots, P_i, \ldots, \bar{P}_n, \bar{a}_1, \ldots, a_i, \ldots, \bar{a}_n, \bar{Y}, T^\circ)$$
$$= X_i(P_i, a_i)$$

Although our model admits of this type of demand function, X_i need not respond to changes in its own time consumption parameter (a_i). In the case of leisure goods, the own time consumption constraint is not binding and hence exerts no influence on the equilibrium solution. Consequently,

$$\frac{\partial X_i}{\partial a_i} = 0, \; i \in L(X) \qquad . \quad . \quad . \quad . \quad . \qquad (2.18)$$

More rigorously, it is demonstrated in the appendix that

$$\frac{\partial X_i}{\partial a_i} = K_i \frac{D_{ii}}{D} - X_i \frac{K_i}{\lambda} \frac{\partial X_i}{\partial Y} + X_i \left[\frac{\partial}{\partial P_i}\left(\frac{K_i}{\lambda}\right) + X_i \frac{\partial}{\partial Y}\left(\frac{K_i}{\lambda}\right) \right] \qquad (2.19)$$

Since K_i/λ represents the price of time in the ith activity, $\frac{\partial}{\partial P_i}\left(\frac{K_i}{\lambda}\right)$ and $\frac{\partial}{\partial Y}\left(\frac{K_i}{\lambda}\right)$ are the rates of change of that price with respect to P_i and Y, respectively. Since K_i is constant and equal to zero for all leisure goods, (2.18) immediately follows.

25

25

Despite the simplicity of (2.18), it is a rather important feature of our model. For a large majority of consumption activities, individuals do not consider time as part of the price of the commodity being purchased. The fact that our model generates a specific set of commodities whose demand functions are not " time elastic " assumes greater significance in light of this body of casual empirical evidence.[1] More importantly, what does our model enable us to say about the sign of $\partial X_i/\partial a_i$ in the case of intermediate goods, in which a_i may be interpreted as a parametric time price? Unfortunately, very little. The time effect, like the price effect, includes the strictly non-positive term, D_{ii}/D, which is enforced by an income effect so long as the good is not inferior. However, a positive income effect does not ensure a negative time effect, as it does a negative price effect. The term, $\frac{\partial X_i}{\partial a_i} + X_i \frac{K_i}{\lambda} \frac{\partial X_i}{\partial Y}$, might exceed zero due to the variability of the price of time. The model therefore provides little theoretical justification for the practice of drawing time-elastic demand curves with a downward slope, for the empirical significance of $\frac{\partial}{\partial P_i}\left(\frac{K_i}{\lambda}\right)$ and $\frac{\partial}{\partial Y}\left(\frac{K_i}{\lambda}\right)$ cannot be ignored. One study which explicitly considered variations in the price of travel time [14] measured relationships between the price of time and both income and time costs. These relationships were statistically significant at the 95% level.

III. Measurement of the Price of Travel Time

A large majority of empirical works dealing with time valuation use what might be termed a revealed preference approach or a " trade-off " approach. The essence of this method is to identify a situation in which the traveller reveals, usually via questionnaire, a preference between alternatives involving a trade-off between higher (lower) money costs and lower (higher) time costs. Situations which have been used for this purpose involve choices among alternative routes [4], alternative modes of travel [2, 18], and alternative speeds [16]. A preference for a slower, less expensive alternative is interpreted to mean that the respondent values his time at a rate no greater than the measured trade-off, and vice versa. However, the assumptions necessary to justify these conclusions are very restrictive. Non-economic factors which play such an important role in the foundations of demand theory, namely subjective preferences, play no role in this type of analysis, for the only choice criterion is the relative costs of the various alternatives. The importance of subjective preference in this type of decision is suggested in the results of Beesley's study wherein a significant number of " inconsistent " choices were found. Inconsistent choices are defined by Beesley

[1] Throughout this paper, we shall use the term " time-elastic," to indicate that demand is responsive to changes in time costs per unit (changes in the parameter, a_i). It does not refer to any numerical value of the elasticity of demand with respect to a_i.

26

as those in which the preferred journey required both more time and more money or more of one, the other cost being the same for each journey [11, p. 178, Table 1].

As an alternative, our model suggests that the price of travel time is measurable directly from the demand function for trips. Dividing (2.19) by (2.14) and rearranging terms yields the relation,

$$\frac{K_t}{\lambda} = \frac{\dfrac{\partial X_t}{\partial a_t} - X_t\left[\dfrac{\partial}{\partial P_t}\left(\dfrac{K_t}{\lambda}\right) + X_t\dfrac{\partial}{\partial Y}\left(\dfrac{K_t}{\lambda}\right)\right]}{\dfrac{\partial X_t}{\partial P_t}} \quad . \quad . \quad (3.1)$$

where X_t might represent the demand for trips on a particular route over a specified period of time. If X_t is specified as a function of time costs, own price and income, the terms, $\partial X_t/\partial a_t$ and $\partial X_t/\partial P_t$, are reflected in the regression coefficients of the demand estimating equation. Obtaining estimates of the rates of change of the price of time with respect to income and the money price of the trip is not quite so easy, however.

One way of handling these two terms would be to ignore them on grounds that they would tend to offset each other. Indeed, it would appear that they do act in opposite directions. One would expect, *ceteris paribus*, that an individual would be willing to pay more money for the purpose of saving time, the lower the price he is already paying for the journey and the higher his income. Nevertheless, however reasonable it appears to hypothesise opposite algebraic signs for these two terms, there is no reason to lead us to believe they would offset each other. Such an assumption would be no less arbitrary than assuming them both to be equal to zero.

As an alternative, one could attempt to infer the values of these terms from the demand estimating model. If, for example, the estimating equation were of the general log linear form,

$$\log X_t = \log \alpha_t + \beta_t \log P_t + \gamma_t \log a_t + \delta_t \log Y + u_t \quad . \quad (3.2)$$

the price of time, derived from (3.2), would be

$$\frac{K_t}{\lambda} = \frac{\dfrac{\gamma_t}{a_t} - \left[\dfrac{\partial}{\partial P_t}\left(\dfrac{K_t}{\lambda}\right) + X_t\dfrac{\partial}{\partial Y}\left(\dfrac{K_t}{\lambda}\right)\right]}{\dfrac{\beta_t}{P_t}} \quad . \quad . \quad (3.3)$$

Ignoring second-order effects, this equation may be solved directly for the relevant rates of change:

$$\frac{\partial}{\partial P_t}\left(\frac{K_t}{\lambda}\right) = \frac{\partial}{\partial P_t}\left[\frac{\gamma_t}{\beta_t}\frac{P_t}{a_t}\right] \quad . \quad . \quad . \quad (3.4)$$

$$\frac{\partial}{\partial Y}\left(\frac{K_t}{\lambda}\right) = \frac{\partial}{\partial Y}\left[\frac{\gamma_t}{\beta_t}\frac{P_t}{a_t}\right] \quad . \quad . \quad . \quad (3.5)$$

27

Aside from the significant estimating difficulties involved, this procedure suffers from the theoretical problem that it requires that $\frac{\partial^2}{\partial P_i^2}\left(\frac{K_i}{\lambda}\right)$ and $\frac{\partial^2}{\partial Y^2}\left(\frac{K_i}{\lambda}\right)$ be assumed equal to zero or, at least, negligible. Such an assumption is without empirical or theoretical foundation, although it might be avoided if an iterative means of solving for the relevant rates of change were adopted. Despite these difficulties, the procedure has considerable merit. Its prime advantages are threefold: (1) the estimated demand equation reflects, at least theoretically, the preferences of individuals as a whole. The important "non-economic" factors, such as comfort and convenience are therefore implicitly considered; (2) the validity of the aggregation technique implicit in the estimation of demand functions does not depend on any arbitrary assumptions about the individuals comprising the group; (3) most importantly, the measure itself is compatible with the hypothesis of utility maximisation. No other measure can make that claim.

This method might be contrasted with a recent attempt by Reuben Gronau to estimate the price of time from the demand function for air passenger travel [8]. The estimating equation, derived from a model similar to the one developed by Gary Becker [1], suggests that demand is a function of "full price" $(P_j + kW_iT_j)$ and income (Y_i):

$$\log X_{ij} = \beta_{0j} + \beta_{1j} \log (P_j + kW_iT_j) + \beta_{2j} \log Y_i + u_{ij} \qquad (3.6)$$

where the subscripts, i and j, denote the ith income group and align the jth destination, and u_{ij} denotes the stochastic disturbance term. In Gronau's model, time value estimation reduces to selecting the proper value of k (presumed to be constant), the ratio of the individual's price of time to his hourly wage (W_i). Gronau's T_i is equivalent to a_i in our model.

A number of problems arise in connection with this procedure. First, k cannot be estimated directly. As a means of determining k, Gronau selects the value that yields the highest adjusted R^2. Secondly, there is no reason to suspect that the price of time is a constant. Both our model and Gronau's suggest that this is not the case, since the price of time is the ratio of two shadow prices. Because of the variability of the price of time, Gronau's measure contains a theoretical bias. From (3.6), it may be demonstrated that

$$kW_i = \frac{\partial X_i}{\partial a_i} \bigg/ \frac{\partial X_i}{\partial P_i} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.7)$$

which is not the same as (3.1), the measure implicit in the individual's utility maximising decision. Thus, even granting the validity of the R^2 test, his conclusion that individuals behave as if they value time at their respective wage rates is suspect (see [8], pp. 52–3).

Moreover, it is suggested that any alleged relationship of this type between the price of time and the individual's wage is artificial. Though there is likely to be a high positive correlation between the two, it is unlikely that there exists any common proportionality constant for any significant proportion of the population. The fact that Gronau's estimating equations revealed virtually no difference among R^2 values would seem to bear this out (see [8], p. 48, Table 7).

IV. Summary and Conclusions

The model developed in this study is neither a theory of working time, nor a theory of leisure time, nor a theory of travel time, but a theory of time. Particular usages of time are not intrinsic to the role of time in affecting consumer decisions. The single feature which distinguishes this model from others dealing with this problem is the time consumption constraints, which allow for the fact that the amount of time spent in any activity is partly a matter of choice and partly a matter of necessity. When it becomes a matter of necessity, an additional constraint becomes binding upon the consumer's preferences and this constraint must be made explicit. When it is solely a matter of choice, the constraint is not effective and " time prices " have no effect upon the consumer's decision. The non-linear programming model is the only way to capture both features.

All of the implications of our model that relate to the time dimension follow directly from these constraints, although most of these results are not new. The distinction between the value of time and the value of time saving as well as the artificiality of defining leisure as non-work have been brought out by Tipping and others. The responsiveness of demand for certain activities to changes in time costs is the basic premise underlying the search for a socially optimal level of use of highways and airport facilities during periods of congestion. That demand is not responsive to changes in time costs in all situations has clearly been recognised. The fact that the timeless neoclassical theory has endured for over half a century is ample evidence of this. At the conclusion of his study, Johnson wrote: " As important as they are in the context of transportation, leisure, and work decisions, the time dimensions of activities may be irrelevant for many problems in neoclassical economic theory " [10, p. 143].

Yet, while these concepts are not new in themselves, they have never been explicitly derived from a formal model of consumer choice. The fact that all of these implications are already widely accepted is a tribute to the accuracy of the model's predictions. Moreover, the wide applicability of the timeless neoclassical theory lends greater import to another significant feature of our model. If all the time consumption constraints are assumed to be ineffective in the neighbourhood of solution, our model is qualitatively indistinguishable from the timeless one. Thus, our model applies in all those situations where neoclassical theory is applicable, as well as in a wide variety of other situations.

29

30

Finally, the most important new result generated by our model is the relationship established between the price of time and the demand function. With this type of measure, the important " non-economic " factors, which are largely ignored in the revealed preference approach to time value estimation, are implicit in the regression coefficients of the demand equation. The prospects of estimating the price of time in this manner at an acceptable level of statistical confidence may not be viewed with total optimism, however, for we have been unable to suggest any clearly acceptable method of estimating the *rates of change* of the price of time, which directly affect the price of time itself. Moreover, there is considerable doubt that the price of time, even if accurately measurable, is sufficiently stable (either with respect to time or with respect to parameter changes) to be used as a basis for solving any practical economic problem. Nevertheless, these difficulties are attributable not to any shortcoming of the theoretical analysis, but to the nature of the beast about which we have been theorising. The model has enabled us to define clearly many of the difficult methodological and conceptual problems that must be dealt with if ever we are to arrive at a useful measure of the price of time. It therefore establishes a firm analytical foundation by which future theoretical and empirical studies of this subject will most assuredly profit.

A. C. DeSerpa

Monterey, California.

Appendix: The Time Effect

From (2.6) in the text, it follows that

$$\frac{\partial X_i}{\partial a_i} = K_i \frac{D_{ii}}{D} + X_i \frac{D_{2n+2+i\,i}}{D} \qquad . \quad . \quad . \quad \text{(A.1)}$$

But, from (2.8) and (2.10), it follows that

$$\frac{D_{2n+2+i\,i}}{D} = \frac{D_{i\,2n+2+i}}{D} = \frac{1}{\lambda}\left[\frac{\partial K_i}{\partial P_i} + X_i \frac{\partial K_i}{\partial Y}\right] \qquad . \quad . \quad \text{(A.2)}$$

whereupon, by substitution, (A.1) becomes

$$\frac{\partial X_i}{\partial a_i} = K_i \frac{D_{ii}}{D} + \frac{X_i}{\lambda}\left[\frac{\partial K_i}{\partial P_i} + X_i \frac{\partial K_i}{\partial Y}\right] \qquad . \quad . \quad \text{(A.3)}$$

Differentiating K_i/λ with respect to P_i and Y, we get

$$\frac{\partial}{\partial P_i}\left(\frac{K_i}{\lambda}\right) = \frac{1}{\lambda}\frac{\partial K_i}{\partial P_i} - \frac{K_i}{\lambda^2}\frac{\partial \lambda}{\partial P_i} \qquad . \quad . \quad . \quad . \quad \text{(A.4)}$$

and

$$\frac{\partial}{\partial Y}\left(\frac{K_i}{\lambda}\right) = \frac{1}{\lambda}\frac{\partial K_i}{\partial Y} - \frac{K_i}{\lambda^2}\frac{\partial \lambda}{\partial Y} \qquad . \quad . \quad . \quad . \quad \text{(A.5)}$$

31

Multiplying through (A.5) by X_t, adding (A.4) to (A.5), and rearranging, we get

$$\frac{1}{\lambda}\left[\frac{\partial K_t}{\partial P_t}+X_t\frac{\partial K_t}{\partial Y}\right]=\frac{\partial}{\partial P_t}\left(\frac{K_t}{\lambda}\right)+X_t\frac{\partial}{\partial Y}\left(\frac{K_t}{\lambda}\right)+\frac{K_t}{\lambda^2}\left[\frac{\partial\lambda}{\partial P_t}+X_t\frac{\partial\lambda}{\partial Y}\right] \quad (A.6)$$

From (2.7), (2.10) and (2.12) in the text, it follows that

$$\frac{\partial\lambda}{\partial P_t}+X_t\frac{\partial\lambda}{\partial Y}=\lambda\frac{D_{t\,2n+1}}{D}=\lambda\frac{D_{2n+1\,t}}{D}=-\lambda\frac{\partial X_t}{\partial Y} \quad . \quad (A.7)$$

Combining (A.6) and (A.7) we find that

$$\frac{1}{\lambda}\left[\frac{\partial K_t}{\partial P_t}+X_t\frac{\partial K_t}{\partial Y}\right]=\frac{\partial}{\partial P_t}\left(\frac{K_t}{\lambda}\right)+X_t\frac{\partial}{\partial Y}\left(\frac{K_t}{\lambda}\right)-\frac{K_t}{\lambda}\frac{\partial X_t}{\partial Y} \quad . \quad (A.8)$$

Substituting (A.8) into (A.3) produces (2.19) in the text.

Bibliography

1. Becker, Gary S., "A Theory of the Allocation of Time," Economic Journal (September 1965), pp. 493–517.
2. Beesley, M. E., "The Value of Time Spent in Travelling: Some More Evidence," *Economica* (May 1965), pp. 174–85.
3. Chiswick, Barry R., "The Economic Value of Time and the Wage Rate: Comment," *Western Economic Journal* (March 1967), pp. 294–5.
4. Claffey, P. J., "Characteristics of Travel on Toll Roads and Comparable Free Roads for Highway User Benefit Studies," *Highway Research Board Bulletin* No. 306, Highway Research Board, (Washington, D.C., 1961).
5. Debreu, Gerard, "Definite and Semi-Definite Quadratic Forms," *Econometrica* (1952), pp. 295–300.
6. DeSerpa, Allan C., *A General Theory of the Economics of Time*, unpublished Ph.D. dissertation, University of California, (Santa Barbara, December 1970).
7. Eckert, Ross, and Minasian, Jora, "The Economics of Airport Use, Congestion and Safety," *California Management Review* (Spring, 1969), pp. 11–24.
8. Gronau, Reuben, *The Value of Time in Passenger Transportation: The Demand for Air Travel*, National Bureau of Economic Research, New York, Occasional Paper No. 109, (1970).
9. Johnson, M. Bruce, "On the Economics of Road Congestion," *Econometrica* (January 1964), pp. 137–50.
10. ——, "Travel Time and the Price of Leisure," *Western Economic Journal* (Spring, 1966), pp. 135–45.
11. ——, "Reply," *Western Economic Journal* (March 1967), p. 296.
12. Kuhn, H. W., and Tucker, A. W., "Nonlinear Programming," in Neyman, J. (ed.), *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pp. 481–92. Berkeley, California: (University of California Press, 1951).
13. Lancaster, Kelvin, *Mathematical Economics* (New York: The Macmillan Company, 1968).

31

32

14. Lee, N. and Dalvi, M. Q., "Variations in the Value of Travel Time," *Manchester School of Economics and Social Studies* (September 1969), pp. 213–36.
15. Minasian, Jora, "Ambiguities in the Theory of Peak-Load Pricing and Application of Theory of Queues," *Land Economics* (August 1966), pp. 355–62.
16. Mohring, Herbert, "Urban Highway Benefits," in Dorfman, Robert (ed.), *Measuring Benefits of Government Investments* (Brookings Institution, 1965).
17. Oort, C. J., "The Evaluation of Travelling Time," *Journal of Transport Economics and Policy* (September 1969).
18. Quarmby, D. A., "Choice of Travel Mode for the Journey to Work," *Journal of Transport Economics and Policy*, (September 1967) pp. 273–314.
19. Robbins, Lionel, "On the Elasticity of Demand for Income in Terms of Effort," *Economica* (June 1930), pp. 123–9.
20. Samuelson, Paul A., *Foundations of Economic Analysis* (New York: Atheneum, 1965).
21. Slutsky, Eugen, "Sulla Teoria del Bilancia del Consumatore," *Giornale degli Economisti* (1915), pp. 1–26.
22. Tipping, David G., "Time Savings in Transport Studies," ECONOMIC JOURNAL (December 1968), pp. 843–54.
23. Walters, Alan A., "The Theory and Measurement of Private and Social Cost of Highway Congestion," *Econometrica* (October 1961), pp. 676–99.
24. Zangwill, Willard I., *Non-linear Programming: A Unified Approach* (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1969).

32

33

Demographic and Economic Change in Developed Countries

A CONFERENCE OF THE
UNIVERSITIES-NATIONAL BUREAU COMMITTEE
FOR ECONOMIC RESEARCH

A REPORT OF THE
NATIONAL BUREAU OF ECONOMIC RESEARCH, NEW YORK, 1960

NATIONAL BUREAU OF ECONOMIC RESEARCH
New York

Distributed by COLUMBIA UNIVERSITY PRESS
New York and London

33

34

An Economic Analysis of Fertility

GARY S. BECKER

COLUMBIA UNIVERSITY

AND

NATIONAL BUREAU OF ECONOMIC RESEARCH

THE inability of demographers to predict western birth rates accurately in the postwar period has had a salutary influence on demographic research. Most predictions had been based either on simple extrapolations of past trends or on extrapolations that adjusted for changes in the age-sex-marital composition of the population. Socio-economic considerations are entirely absent from the former and are primitive and largely implicit in the latter. As long as even crude extrapolations continued to give fairly reliable predictions, as they did during the previous half century, there was little call for complicated analyses of the interrelation between socio-economic variables and fertility. However, the sharp decline in birth rates during the thirties coupled with the sharp rise in rates during the postwar period swept away confidence in the view that future rates could be predicted from a secularly declining function of population compositions.

Malthus could with some justification assume that fertility was determined primarily by two primitive variables, age at marriage and the frequency of coition during marriage. The development and spread of knowledge about contraceptives during the last century greatly widened the scope of family size decision-making, and contemporary researchers have been forced to pay greater attention to decision-making than either Malthus or the forecasters did. Psychologists have tried to place these decisions within a framework suggested by psychological theory; sociologists have tried one suggested by sociological theory, but most persons would admit that neither framework has been particularly successful in organizing the information on fertility.

Two considerations encouraged me to analyze family size decisions within an economic framework. The first is that Malthus' famous discussion was built upon a strongly economic framework; mine can be viewed as a generalization and development of his. Second, although no 34

Note: I am indebted to Richard A. Easterlin and Eugenia Scandrett for helpful comments, and to many others, especially Cornelius J. Dwyer, who commented on the draft prepared for the conference.

single variable in the Indianapolis survey[1] explained more than a small fraction of the variation in fertility, economic variables did better than others. Section I develops this framework and sets out some of its implications. Section II uses this framework to analyze the actual effects of income on fertility. Section III speculates about some further implications of the discussion in I and II.

I. The Economic Framework

GENERAL CONSIDERATIONS

In societies lacking knowledge of contraception, control over the number of births can be achieved either through abortion or abstinence, the latter taking the form of delayed marriage and reduced frequency of coition during marriage. Since each person maintains some control over these variables, there is room for decision-making even in such societies. Other things the same, couples desiring small families would marry later and have more abortions than the average couple. Yet the room for decision-making would be uncomfortably small, given the taboos against abortion, the strong social forces determining the age of marriage, and the relative inefficiency of reductions in the frequency of coition. Chance would bulk large in determining the distribution of births among families.[2]

The growth of knowledge about contraception has greatly widened the scope of decision-making, for it has separated the decision to control births from the decision to engage in coition. Presumably, such a widening of the scope of decision-making has increased the importance of environmental factors, but which of the numerous environmental factors are most important? To simplify the analysis of this problem I assume initially that each family has perfect control over both the number and spacing of its births.

For most parents, children are a source of psychic income or satisfaction, and, in the economist's terminology, children would be considered a consumption good. Children may sometimes provide money income and are then a production good as well. Moreover, neither the outlays on children nor the income yielded by them are fixed but vary in amount with the child's age, making children a durable consumption and production good. It may seem strained, artificial, and perhaps even immoral to classify children with cars, houses, and machinery. This classification does not imply, however, that the satisfactions or costs associated with

[1] *Social and Psychological Factors Affecting Fertility*, ed. by P. K. Whelpton and C. V. Kiser, Milbank Memorial Fund, Vols. 1–4.

[2] The effect of chance will be fully discussed in a subsequent paper.

children are morally the same as those associated with other durables. The satisfaction provided by housing, a "necessity," is often distinguished from that provided by cars, a "luxury," yet both are treated as consumer durables in demand analysis. Abstracting from the kind of satisfaction provided by children makes it possible to relate the "demand" for children to a well-developed body of economic theory. I will try to show that the theory of the demand for consumer durables is a useful framework in analyzing the demand for children.

TASTES

As consumer durables, children are assumed to provide "utility." The utility from children is compared with that from other goods via a utility function or a set of indifference curves. The shape of the indifference curves is determined by the relative preference for children, or, in other words, by "tastes." These tastes may, in turn, be determined by a family's religion, race, age, and the like. This framework permits, although it does not predict, fertility differences that are unrelated to "economic" factors.

QUALITY OF CHILDREN

A family must determine not only how many children it has but also the amount spent on them—whether it should provide separate bedrooms, send them to nursery school and private colleges, give them dance or music lessons, and so forth. I will call more expensive children "higher quality" children, just as Cadillacs are called higher quality cars than Chevrolets. To avoid any misunderstanding, let me hasten to add that "higher quality" does not mean morally better. If more is voluntarily spent on one child than on another, it is because the parents obtain additional utility from the additional expenditure and it is this additional utility which we call higher "quality."

INCOME

An increase in income must increase the amount spent on the average good, but not necessarily that spent on each good. The major exceptions are goods that are inferior members of a broader class, as a Chevrolet is considered an inferior car, margarine an inferior spread, and black bread an inferior bread. Since children do not appear to be inferior members of any broader class, it is likely that a rise in long-run income would increase the amount spent on children.[3]

[3] This is also suggested by another line of reasoning. It is known that $\Sigma k_i n_i \equiv 1$, where k_i is the fraction of income spent on the ith commodity, and n_i is the income

For almost all other consumer durables, such as cars, houses, or refrigerators, families purchase more units as well as better quality units at higher income levels, with the quantity income elasticity usually being small compared to the quality elasticity.[4] If expenditures on children responded in a similar way, most of the increased expenditures on children would consist of an increase in the quality of children. Economic theory does not guarantee that the quantity of children would increase at all, although a decrease in quantity would be an exception to the usual case. Thus an increase in income should increase both the quantity and quality of children, but the quantity elasticity should be small compared to the quality elasticity.

Malthus, on the other hand, concluded that an increase in income would lead to a relatively large increase in family size. His argument has two major components. First, an increase in income would cause a decline in child mortality, enabling more children to survive childhood. If a decrease in births did not offset the decrease in child mortality, the number of children in the average family would increase. His second argument is less mechanical and takes greater account of motivation. An increase in income increases fertility by inducing people to marry earlier and abstain less while married.

My analysis has generalized that of Malthus by relating the quantity of children to the quality of children and by permitting small (even negative) quantity income elasticities as well as large ones. My conclusion that in modern society the quantity elasticity is probably positive but small differs from his for the following reasons. First, child mortality has fallen so low that the ordinary changes in income have little effect on the number of survivors out of a given birth cohort. Moreover, it is doubtful that even a large decline in child mortality would have much effect on family size, for parents are primarily interested in survivors, not in births per se. Therefore, a decline in child mortality would induce a corresponding decline in births.[5] Second, births can now be controlled without abstinence and this has greatly reduced the psychic costs of birth

---

elasticity of the amount spent on the ith commodity. Other things the same, the larger k_i is, the less likely it is that n_i is either very small or very large. In particular, the less likely it is that n_i is negative. In most families the fraction of income spent on children is quite large and this decreases the likelihood that the income elasticity for children is negative.

[4] Chow estimated the total income elasticity for automobiles at about +2. Cf. G. C. Chow, *Demand for Automobiles in the United States*, North Holland Publishing Co., Amsterdam, 1957; however, the quantity elasticity is only about +0.31. Cf. *Federal Reserve Bulletin*, August, 1956, p. 820.

[5] This will be discussed more fully in a future publication.

control. "Human nature" no longer guarantees that a growth in income appreciably above the subsistence level results in a large inadvertent increase in fertility.

COST

In principle the net cost of children can be easily computed. It equals the present value of expected outlays plus the imputed value of the parents' services, minus the present value of the expected money return plus the imputed value of the child's services. If net costs were positive, children would be on balance a consumer durable and it would be necessary to assume that psychic income or utility was received from them. If net costs were negative, children would be a producer durable and pecuniary income would be received from them. Children of many qualities are usually available, and the quality selected by any family is determined by tastes, income, and price. For most families in recent years the net expenditure on children has been very large.[6]

Real incomes per capita in the United States have increased more than threefold in the last 100 years, which must have increased the net expenditure on children. It is possible that in the mid-nineteenth century children were a net producer's good, providing rather than using income. However, the marginal cost of children must have been positive in families receiving marginal psychic income from children; otherwise, they would have had additional children. Even in 1850, the typical family in the United States was producing fewer children than was physically possible. Some more direct inferences can be drawn from the data on Negro slaves, an extreme example of a human producer's good. These data indicate a positive net expenditure on male slaves during their first eighteen years.[7] Slave raising was profitable because the high price that an eighteen-year-old could bring more than offset the net cost during the first eighteen years. Presumably, in most families expenditures on white children during their first eighteen years were greater than those on slaves. Moreover, after eighteen, white children became free agents and could decide

[6] See J. D. Tarver, "Costs of Rearing and Educating Farm Children," *Journal of Farm Economics*, February, 1956, pp. 144–153, and L. I. Dublin and A. J. Lotka, *The Money Value of a Man*, Ronald Press, 1946, ch. 4. Most studies consider only the costs and returns before age eighteen. It is possible that returns bulk larger than costs at later ages; but because these ages are heavily discounted and because costs are so large before age eighteen, there is little chance that a correction of this bias would substantially reduce the net cost of children.

[7] See A. H. Conrad and J. R. Meyer, "The Economics of Slavery in the Ante Bellum South," *Journal of Political Economy*, April, 1958, p. 108. At an 8 per cent discount rate (about the estimated rate of return on slaves), the present value of the net costs is + $35, or about one-third of the present value of gross costs. The data are subject to considerable error and are at best a rough indication of the magnitudes involved.

whether to keep their income or give it to their parents. The amount given to parents may have been larger than the costs before eighteen, but it is more likely that costs before eighteen dominated returns after eighteen. This conclusion does not imply that monetary returns from children were unimportant, and indeed, they are stressed at several points in this paper. It does imply, however, that a basic framework which treats children as a consumer's good is relevant not only for the present, but also for some time in the past.

A change in the cost of children is a change in the cost of children of *given quality*, perhaps due to a change in the price of food or education. It is well to dwell a little on this definition for it is widely misunderstood. One would not say that the price of cars has risen over time merely because more people now buy Cadillacs and other expensive cars. A change in price has to be estimated from indexes of the price of a given quality. Secular changes in real income and other variables have induced a secular increase in expenditures on children, often interpreted as a rise in the cost of children. The cost of children may well have risen (see pp. 227–28) but the increase in expenditure on children is no evidence of such rise since the quality of children has risen. Today children are better fed, housed, and clothed, and in increasing numbers are sent to nursery schools, camps, high schools, and colleges. For the same reason, the price of children to rich parents is the same as that to poor parents even though rich parents spend more on children.[8] The rich simply choose higher quality children as well as higher qualities of other goods.[9]

It is sometimes argued that social pressures "force" richer families to

[8] One qualification is needed because the rich may impute a higher value than the poor to the time spent on children. The same qualification is needed in analyzing the demand for other goods.

[9] As an example of how prevalent this error is, even among able economists, we refer to a recent discussion by H. Leibenstein in *Economic Backwardness and Economic Growth*, John Wiley, 1957, pp. 161–170. He tries to relate cost of children to level of income, arguing, among other things, that "The relation between the value of a child as a contributor to family income and changes in per capita income is fairly clear. As per capita income increases, there is less need to utilize children as sources of income. At the same time the level of education and the general quality of the population implied by a higher income per head mean that more time must be spent on child training, education, and development, and, therefore, less time is available to utilize the child as a productive agent. Therefore, the higher the income, the less the utility to be derived from a prospective child as a productive agent" and "The conventional costs of child maintenance increase as per capita income increases. The style in which a child is maintained depends on the position and income of the parents; therefore, we expect such costs to rise as incomes rise. . . ." (*ibid.*, pp. 163–164.)

By trying to relate cost to income Leibenstein confused cost and quality, and succeeded only in inadvertently relating quality to income. His technique would imply that the relative price of almost every group of goods rose over time because the quality chosen spend more on children, and that this increases the cost of children to the rich. This higher cost is supposed to explain why richer families have fewer children than others and why richer societies have fewer children than poorer ones. However, since the cost of different goods is given in the market place, social pressures cannot change this, but can only change the basket of goods selected. That is, social pressures influence behavior by affecting the indifference curve structure, not by affecting costs. To put this differently, social pressures may affect the income elasticity of demand for children by rich (and poor) families, but not the price elasticity of demand. Therefore, the well known negative relationship between cost (or price) and quantity purchased cannot explain why richer families have had relatively few children. Moreover, nothing in economic analysis implies that social pressures would make the quantity income elasticity of demand for children negative. Thus my conclusion that the quantity income elasticity is relatively small but positive and the quality elasticity relatively large is entirely consistent with an analysis which emphasizes social pressures.

Suppose there was an equal percentage decline in the price of all qualities of children, real income remaining constant. Although economic theory suggests that the "amount" of children consumed would increase, it does not say whether the amount would increase because of an increase in quantity, quality, or both—the last, however, being most likely. It also has little to say about the quantitative relationship between price and amount. There are no good substitutes for children, but there may be many poor ones.[10]

---

rose, an obvious impossibility. This flaw in his procedure greatly weakens his analysis of the secular decline in birth rates.

Bernard Okun also applied economic analysis to the population area, and explicitly assumed that the cost of children is higher to rich people because they spend more on children (see *A Rational Economic Model Approach to the Birth Rate*, Rand Corp. Series, P1458, August, 1958). His argument, like Leibenstein's, would imply that the cost of many (if not most) goods is greater to richer families than to poorer ones. Also see S. H. Coontz, *Population Theories and the Economic Interpretation*, Routledge, London, 1957, Part II.

[10] Let x be the quantity of children, p an expenditure measure of the quality of x, y an index of other goods, I money income, U a utility function, α a parameter shifting the cost of each quality of x by the same percentage, and π the price of y. A consumer maximizes $U(x, y, p)$ subject to the constraint $\alpha px + \pi y = I$. This leads to the equilibrium conditions

$$\frac{Ux}{\alpha p} = \frac{Up}{\alpha x} = \frac{Uy}{\pi}$$

The marginal utility from spending a dollar more on the quantity of children must equal the marginal utility from spending a dollar more on their quality.

After a draft of this paper was written I came across an article by H. Theil, "Qualities, Prices, and Budget Inquiries," *The Review of Economic Studies*, XIX, pp. 129–147, which

SUPPLY

By and large, children cannot be purchased on the open market but must be produced at home. Most families are no longer self-sufficient in any major commodity other than children. Because children are produced at home, each uncertainty in production is transferred into a corresponding uncertainty in consumption, even when there is no uncertainty for all families taken together. Although parents cannot accurately predict the sex, intelligence, and height of their children, the distribution of these qualities is relatively constant for the country as a whole. This uncertainty makes it necessary to distinguish between actual and expected utility. Thus suppose a group of parents received marginal utility equal to U_m from a male child and U_f from a female child. The expected utility from an additional child equals $EU = PU_m + (1 - P)U_f \cong \frac{U_m + U_f}{2}$, where P, the probability of a male is approximately equal to 1/2. They would have additional children whenever the expected utility per dollar of expected cost from an additional child were greater than that from expenditures elsewhere. The actual utility is either U_f or U_m, which differs from EU as long as $U_f \neq U_m$. In fact, if U_f (or U_m) were negative, some parents would receive negative utility.

A second important consequence of uniting consumption and production is that the number of children available to a family is determined not only by its income and prices but also by its ability to produce children. One family can desire three children and be unable to produce more than two, while another can desire three and be unable to produce fewer than five.[11] The average number of live births produced by married women in societies with little knowledge of contraception is very high. For example, in nineteenth-century Ireland, women marrying at ages 20–24 averaged more than 8 live births.[12] This suggests that the average family more frequently had excess rather than too few children.

---

treats the interaction of quality and quantity in an elegant manner. Also see, in the same issue, H. S. Houthhakker, "Compensated Changes in Quantities and Qualities Consumed," pp. 155–164. Theil differentiates equations like these and shows that a compensated decrease in the price of a good of given quality must increase either the quantity of goods or the quality, or both.

[11] There is some ambiguity in the last part of this sentence since abstinence enables a family to produce as few children as desired. The terms "unplanned," "excess," or "unwanted" children refer to children that would not be conceived if there were perfect mechanical control over conception. No children are unplanned in terms of the contraceptive knowledge and techniques actually known.

[12] See D. V. Glass and E. Grebenik, *The Trend and Pattern of Fertility in Great Britain*, Paper of the Royal Commission on Population, Vol. VI, p. 271.

Relatively effective contraceptive techniques have been available for at least the last 100 years, but knowledge of such techniques did not spread rapidly. Religious and other objections prevented the rapid spread of knowledge that is common to other technological innovations in advanced countries. Most families in the nineteenth century, even in advanced Western countries, did not have effective contraceptive information. This information spread slowly from upper socio-economic groups to lower ones.[13]

Each family tries to come as close as possible to its desired number of children. If three children are desired and no more than two are available, two are produced; if three are desired and no fewer than five are available, five are produced. The marginal equilibrium conditions would not be satisfied for children but would be satisfied for other goods, so the theory of consumer's choice is not basically affected.[14] Families with excess children consume less of other goods, especially of goods that are close substitutes for the quantity of children. Because quality seems like a relatively close substitute for quantity, families with excess children would spend less on each child than other families with equal income and tastes. Accordingly, an increase in contraceptive knowledge would raise the quality of children as well as reduce their quantity.

II. An Empirical Application

Having set out the formal analysis and framework suggested by economic theory, we now investigate its usefulness in the analysis of fertility patterns. It suggests that a rise in income would increase both the quality and quantity of children desired; the increase in quality being large and the increase in quantity small. The difficulties in separating expenditures on children from general family expenditures notwithstanding, it is evident that wealthier families and countries spend much more per child than do poorer families and countries. The implication with respect to quantity is not so readily confirmed by the raw data. Indeed,

[13] For evidence supporting the statements in this paragraph see the definitive work by N. A. Himes, *Medical History of Contraception*, The Williams and Wilkins Company, Baltimore, 1936.

[14] A consumer maximizes a utility function $U = u(x_1, \ldots x_n)$ (neglecting quality considerations) subject to the constraints $\sum_{i=1}^{n} p_i x_i \equiv Y$, and $x_1 \geq$ or $\leq c$, where p_i is the price of the ith commodity, Y is money income, and x_1 refers to children. If the second constraint were effective, x_1 would equal c. Then the consumer would maximize $U = U(c, x_2, \ldots x_n)$ subject only to $\sum_{i=2}^{n} p_i x_i \equiv Y' \equiv Y - p_1 c$, and this gives the usual marginal conditions for $x_2, \ldots, x_n$.

most data tend to show a negative relationship between income and fertility. This is true of the Census data for 1910, 1940, and 1950, where income is represented by father's occupation, mother's education, or monthly rental; the data from the Indianapolis survey, the data for nineteenth century Providence families, and several other studies as well.[15] It is tempting to conclude from this evidence either that tastes vary systematically with income, perhaps being related to relative income, or that the number of children is an inferior good. Ultimately, systematic variations in tastes may have to be recognized; but for the present it seems possible to explain the available data within the framework outlined in section I, without assuming that the number of children is an inferior good. First, it is well to point out that not all the raw evidence is one way. In some studies, the curve relating fertility and income flattens out and even rises at the higher income classes, while in other studies the curve is positive throughout.[16] Second, tastes are not the only variable that may have varied systematically with income, for there is a good deal of general evidence that contraceptive knowledge has been positively related to income. Himes, in his history of contraception, indicates that the upper classes acquired this knowledge relatively early.[17] If such knowledge spread gradually from the upper classes to the rest of society, fertility differentials between classes should have first increased and then narrowed. This was clearly the pattern in England and was probably the pattern in the United States.[18]

Such evidence does little more than suggest that differential knowledge of contraceptive techniques might explain the negative relationship between fertility and income. Fortunately, the Indianapolis survey makes it possible, at least for 1941, to assess its quantitative importance. Table 1 presents some data from this study. In column (1) the native-white Protestant couples in the sample are classified by the husband's income, and column (2) gives the number of children born per 100 couples in each income class. The lowest income class was most fertile (2.3 children per couple) and a relatively high class least fertile (1.5

[15] U.S. Bureau of the Census, Census of Population, 1940; *Differential Fertility 1910 and 1940*, Government Printing Office, Washington, 1945; U.S. Bureau of the Census, Census of Population, 1950; *Fertility*, Government Printing Office, Washington, 1955; *Social and Psychological Factors Affecting Fertility*, by P. K. Whelpton and C. V. Kiser, eds., Milbank Memorial Fund, 1951; A. J. Jaffe, "Differential Fertility in the White Population in Early America," *Journal of Heredity*, August, 1940, pp. 407–411.

[16] K. A. Edin and E. P. Hutchinson, *Studies of Differential Fertility*, London, 1935; W. H. Banks, "Differential Fertility in Madison County, New York, 1865," *Milbank Memorial Fund Quarterly*, Vol. 33, April, 1955, pp. 161–186.

[17] Himes, *op. cit.*

[18] See the papers by C. V. Kiser and G. Z. Johnson in this volume.

children per couple), but the highest class averaged slightly more children than the next highest. This relationship between economic level and fertility was about the same as that shown by the 1940 Census.[19] Sterility did not vary systematically with income, so column (3), which is restricted to relatively fecund families, differs only slightly from column (2).

TABLE 1

Children Ever Born per 100 Couples in Indianapolis Classified by Husband's Income and Planning Status (native-white Protestants)

Income (1)	*All Couples* (2)	*Relatively Fecund* (3)	*Number and Spacing Planners* (4)	*All Planners* (5)	*Desires of Relatively Fecund* (6)
$3,000+	159	180	149	175	171
2,000–2,999	149	176	182	161	170
1,600–1,999	163	194	91	126	153
1,200–1,599	189	229	97	144	175
1,200 and less	227	266	68	146	193

Source: *Social and Psychological Factors Affecting Fertility*, P. K. Whelpton and C. V. Kiser, eds., N.Y., Milbank Memorial Fund, 1951, Vol. 2, part 9. Columns (2) and (3) from Table 4; columns (4) and (5) computed from Figure 8; column (6) computed from Figures 8 and 21.

It is well known that rich families use contraception earlier and more frequently than poor families. It has been difficult to determine whether poor families are ignorant of contraceptive methods or whether they desire more children than richer ones. The Indianapolis survey tried to separate ignorance from tastes by classifying couples not only by use of contraception but also by control over births. Column (4) gives the average number of children for "number and spacing planning" couples, including only couples who had planned all their children. A positive pattern now emerges, with the richest families averaging more than twice as many children as the poorest families. The income elasticity is about +0.42. Column (5) presents data for "number planned" couples, including all couples that planned their last child. These data also show a positive pattern, with an elasticity of +0.09, lower than that for number and spacing planners.

Fecund couples having excess children were asked questions about the number of such children. Column (6) uses this information and that in column (5) to relate income to the number of children desired by all

[19] Whelpton and Kiser, eds., *op. cit.*, Vol. 2, p. 364.

fecund couples. The elasticity is negative, being about -0.07.[20] After an intensive study, however, Potter found evidence that the number of desired children was overestimated; his own estimates of desired fertility show a positive relationship with income.[21] Thus evidence from the Indianapolis survey indicates that differential knowledge of contraception does convert a positive relation between income and *desired* fertility into a negative relation between income and *actual* fertility.[22]

Several other surveys provide information on desired fertility. For example, in 1954 a group at Michigan asked Detroit area families; "In your opinion what would be the ideal number of children for a young couple to have, if their standard of living is about like yours?" There was a distinct positive relationship between the ideal number of children and income of the family head.[23]

If knowledge of contraceptive techniques did not vary with income, the relation between actual fertility and income would equal that between desired fertility and income. Contraceptive knowledge is said to be diffused among all income classes in Stockholm, and the fertility of Stockholm families from 1917–1930 was positively related to income.[24] Contraceptive knowledge was said to be very primitive in *all* income

[20] These elasticities are estimates of the slope of the regression of the logarithm of fertility on the logarithm of income. The mean of the open end income class is assumed to be $4,000, and the mean of the other classes is assumed to be at their mid-points.

[21] R. G. Potter, *The Influence of Primary Groups on Fertility*, unpublished Ph.D. dissertation, Department of Social Relations, Harvard University, 1955, Appendix A, pp. 277–304.

[22] This conclusion must be qualified to allow for the possibility that tastes and costs also varied with income. Since all couples lived in the same city the cost of children was presumably the same. Age, religion, color, and nativity were held constant in an attempt to limit the systematic variation in tastes. Education did vary with income, but for number and spacing planners it was possible to separate the effect of income from the effect of education. The simple correlation coefficient between fertility and income is $+0.24$ and between fertility and education $+0.17$, with both significant at the 1 per cent level. The partial correlation coefficient between fertility and income, holding education constant is $+0.23$, about the same as the simple coefficient, and is also significant at the 1 per cent level. The partial correlation between fertility and education is only $+0.04$, not significant even at the 10 per cent level. (For these correlations see Whelpton and Kiser, eds., *op. cit.*, Vol. 3.) Holding education constant has little effect on the relationship between income and fertility.

[23] See R. Freedman, D. Goldberg, and H. Sharp, " 'Ideals' about Family Size in the Detroit Metropolitan Area, 1954," *Milbank Memorial Fund Quarterly*, Vol. 33, April, 1955, pp. 187–197. An earlier survey asked about the ideal family size for the average American couple, and found a negative relationship between ideal size and income of the head. But ideal size should be related to the income *assumed* by a respondent, rather than to his own income; and there is no way to do this. R. G. Potter has criticized both surveys because of their tendency to show larger ideal than realized families. See his "A Critique of the Glass-Grebenik Model for Indirectly Estimating Desired Family Size," *Population Studies*, March, 1956, pp. 251–270. It is not possible to determine whether this bias is systematically related to income.

[24] See Edin and Hutchinson, *op. cit.*

classes of prewar China, and a positive relation between fertility and income also seemed to prevail there.[25] Graduates in the same college class are probably relatively homogeneous in contraceptive knowledge and values as well as in formal education. I have the impression that income and fertility of these graduates tend to be positively related, but I have been able to examine only one sample. Some graduates from Harvard and Yale were classified by occupation and "degree of success." Within each occupation, the more successful graduates usually had more children.[26]

Information has been obtained on the family income, education, earners, and dependent children of a sample of the subscribers to Consumers Union.[27] This sample is particularly valuable for our purposes since it primarily consists of families with a keen interest in rational, informed consumption. If my analysis is at all relevant, fertility and income should be more positively related in this group than in the U.S. population as a whole. Table 2 presents the average number of dependent

TABLE 2

Average Number of Dependent Children for Single Earner Families with Head Age 35–44 in a Sample of Subscribers to Consumers Union, April, 1958

	Average Number of Dependent Children by Education Class of Head			
Income Class	High School Graduate or Less	Some College	Graduate of Four Year College	Graduate Degree
Less than $3,000	2.43	1.61	2.50	2.17
$ 3,000–3,999	2.15	2.47	2.18	2.23
4,000–4,999	2.70	2.40	2.04	2.18
5,000–7,499	2.68	2.73	2.88	2.67
7,500–9,999	2.80	2.94	3.00	3.03
10,000–14,999	2.89	3.03	3.12	3.23
15,000–24,999	2.85	3.04	3.04	3.31
25,000 and over	3.12	3.23	3.28	3.60

Source: Unpublished data from consumer purchases study by Thomas Juster at National Bureau of Economic Research.

[25] See H. D. Lamson, "Differential Reproductivity in China," *The Quarterly Review of Biology*, Vol. 10, no. 3, September, 1933, pp. 308–321. Abstinence, which is equally available to lower and upper classes, is the major form of birth control when contraceptive knowledge is limited.

[26] See E. Huntington and L. F. Whitney, *The Builders of America*, New York, Morrow, 1927, ch. xv. Although they did not clearly define "success," it appears that income was a major factor in ranking persons within an occupation and a less important factor in ranking occupations. 40

[27] This is part of a study by Thomas Juster on buying plans, and I am indebted to him for making the data available to me.

children for single earner families with the head aged 35–44, each family classified by its income and by the education of the head. There is a substantial positive relationship between income and children within each educational class; education *per se* has relatively little effect on the number of children. The income elasticity is about 0.09 and 0.14 for graduates of a four year college and of a graduate school respectively. These data, then, are very consistent with my analysis, and indicate that well-informed families do have more children when their income increases.

Contraceptive knowledge in the United States spread rapidly during the War, largely fostered by the military in its effort to limit venereal disease and illegitimacy. We would expect this to have reduced the relative fertility of low income classes, and Census Bureau studies in 1952 and 1957 confirm this expectation. Table 3 presents the data for urban and rural nonfarm families for 1952 and all families for 1957 with column (1) giving husband's income, column (2) the age-standardized number

TABLE 3
Fertility by Husband's Income

Husband's Income (1)	*Children Under 5 per 100 Married Men 20–59* (age standardized) (2)	*Children Born per 100 Wives 15–44 Years Old* (age standardized) (3)	*Children Born per 100 Wives over 45* (4)
Part I: In Urban and Rural Nonfarm Areas in the United States in 1952			
$7 000+	53	189	194
6,000–6,999	52	188	210
5,000–5,999	50	188	210
4,000–4,999	52	177	217
3,000–3,999	52	184	240
2,000–2,999	51	189	256
1,000–1,999	40	181	279
1,000 and less	40	211	334
Part II: For the United States in 1957			
$7,000+	—	216	213
5,000–6,999	—	220	230
4,000–4,999	—	221	240
3,000–3,999	—	236	279
2,000–2,999	—	247	304
1,000–1,999	—	289	341
1,000 and less	—	—	383

Source I. U.S. Bureau of the Census, *Current Population Reports*, Wash., Government Printing Office, 1953, no. 46, p. 20.
II. U.S. Bureau of the Census, *Current Population Reports*, Wash., Government Printing Office, 1958, no. 84, p. 12.

of children under 5 per 100 men aged 20 to 59, column (3) the age-standardized number of children ever born per 100 wives aged 15 to 44, and column (4) the number ever born per 100 wives aged 45 and older. Columns (2) and (3) deal primarily with childbearing since 1940 and show a much weaker negative relationship between fertility and income than does column (4), which deals primarily with childbearing before 1940.

The relationship between fertility and income can be investigated not only with cross-sectional income differences but also with time series differences. Cyclical fluctuations in income have regularly occurred in Western nations, and, if our analysis is correct, a change in income would induce a change in fertility in the same direction. For our purpose cyclical fluctuations in fertility can be measured by the cyclical fluctuations in births (although see p. 227). Some earlier studies presented evidence that births do conform positively to the business cycle, even when adjusted for fluctuations in the marriage rate.[28]

I have related some annual figures since 1920 on first and higher order birth rates—brought forward one year—to the National Bureau annual business cycle dates. Column (3) of Table 4 gives the percentage change per year in first and higher order birth rates from the beginning of one phase to the beginning of the next phase. The strong secular decline in births before World War II makes most of these entries negative before that time and hence obscures the effect of cyclical fluctuations in economic conditions. If economic conditions affected births they should have declined more rapidly (or risen less rapidly) during a downswing than during an upswing. This can be detected from the first differences of the entries in column (3), which are shown in column (4). Aside from the wartime period, 1938–1948, second and higher order births conform perfectly in direction to the reference dates and first births conform almost as well. So reference cycle analysis strongly indicates that business conditions affect birth rates. This effect is not entirely dependent on cyclical fluctuations in the marriage rate since second and higher order births conform exceedingly well.

The next step is to relate the magnitude of the movement in births to that in general business, and to compare this with corresponding figures for other consumer durables. Time series giving net national product and purchases of consumer durables were analyzed in the same way as birth

[28] V. L. Galbraith and D. S. Thomas, "Birth Rates and the Interwar Business Cycles," and D. Kirk, "The Relation of Employment Levels to Births in Germany," both in *Demographic Analysis*, J. J. Spengler and O. D. Duncan, eds., Free Press, Glencoe, 1956.

TABLE 4
Reference Cycle Pattern of Birth Rates for U.S. Since 1920

REFERENCE CYCLE DATES[1]		*Birth Rates per 1,000 Women 15–44 Years of Age, Brought Forward One Year at Reference Cycle Dates*[2]		*Annual Percentage Change During a Business*		*Excess of Annual Percentage Change During Business Expansion Over*	
Peak	*Trough*	*At Peak*	*At Trough*	*Expansion*	*Contraction*	*Preceding Contraction*	*Succeeding Contraction*
(1)		(2)		(3)		(4)	
FIRST BIRTHS							
1920		39					
	1921		34		−12.82		
1923		34		0.00		+12.82	
	1924		34		0.00		0.00
1926		32		−2.94		−2.94	
	1927		30		−6.25		−3.31
1929		30		0.00		+6.25	
	1932		25		−5.57		−5.57
1937		31		4.80		+10.37	
	1938		31		0.00		−4.80
1944		29		−1.06		−1.06	
	1946		46		28.33		+29.39
1948		36		−10.87		−39.20	
	1949		33		−8.33		+2.54
1953		34		0.76		+9.09	
	1954		33		−2.94		−3.70
1957		33*		0.00		+2.94	
HIGHER ORDER BIRTHS							
1920		82					
	1921		78		−4.88		
1923		78		0.00		+4.88	
	1924		74		−5.13		−5.13
1926		68		−4.05		+1.08	
	1927		64		−5.88		−1.83
1929		60		−3.12		+2.76	
	1932		52		−4.45		−1.33
1937		48		−1.53		+2.92	
	1938		47		−2.08		−0.55
1944		57		3.46		+5.54	
	1946		67		8.47		+5.01
1948		71		2.98		−5.49	
	1949		73		2.81		−0.17
1953		84		3.77		+0.96	
	1954		85		1.19		−2.58
1957		88*		3.53		+2.34	

* Last figure is for 1956.
Source: [1] See National Bureau of Economic Research Standard Reference Dates for Business Cycles.
[2] See Dudley Kirk, Appendix to "The Influence of Business Cycles on Marriage and Birth Rates," this volume.

rates were. The figures for birth rates in column (4) of Table 4 and corresponding figures for purchases of consumer durables were divided by corresponding figures for national product to obtain cyclical income elasticities for births and consumer durables. These figures, shown in Table 5, are positive for almost all phases, and this indicates that cyclical

TABLE 5

Cyclical Income Elasticities for Births and Consumer Durable Purchases During Reference Cycle Phases

Reference Cycle Phases (1)		*First Births* (2)	*Higher Order Births* (3)	*Purchases of Consumer Durables* (4)
1920–1921	Down	0.81	0.31	2.48
1921–1923	Up	0.00	.58	2.96
1923–1924	Down	−1.55	.57	6.63
1924–1926	Up	.87	.48	5.26
1926–1927	Down	2.05	.90	4.05
1927–1929	Up	.37	.09	1.40
1929–1932	Down	.47	.13	1.51
1932–1937	Up	.26	.03	1.96
1937–1938	Down	− .09	.46	1.38
1938–1944	Up	4.26	.73	9.20
1944–1946	Down	3.89	.54	5.33
1946–1948	Up	− .44	.03	0.11
1948–1949	Down	.88	.09	0.01
1949–1953	Up	.78	.54	1.78
1953–1954	Down	1.19	.95	3.23
1954–1957	Up			
Simple Average excluding				
1938–1948		.56	.42	2.84
and negative figures		.77	.42	2.84

Source: Birth rates from column (4) of Table 4; similar figures were computed for consumer durable purchases and net national product. The durable figures were from Raymond W. Goldsmith, *A Study of Savings in the United States*, Vol. 1, Tables Q-6 and A-25 for 1920–1949 and from U.S. Dept. of Commerce, *Survey of Current Business*, July, 1958, Table 2, for 1949–1957. Net National Product figures were from Simon Kuznets, Technical Tables (mimeo), T-5, underlying series in *Supplement to Summary Volume on Capital Formation and Financing* for 1920–1955 and from U.S. Dept. of Commerce, *Survey of Current Business*, July, 1958, Table 4, for 1955–1957.

changes in births and purchases of consumer durables have been in the same direction as those in national output. The cyclical change in first births was usually greater than that in higher order births, and both were usually less than the change in output. Changes in first and higher order births were, however, far from insignificant, averaging 74 and 42 per cent of the corresponding change in output. 42

Cyclical changes in births are small compared to those in consumer durables. The latter averaged about 2.84 times the change in output, or about 4 and 7 times the change in first and higher order births respectively. This is consistent with our emphasis on inadequate knowledge of birth control; inadequate knowledge seems to explain much but not all of the difference between the average cyclical change in higher order births and in purchases of durables.[29] Some would be explained by the fact that the data for children include only fluctuations in numbers, while those for durables include both fluctuations in numbers and in quality. The rest may be explained by other differences between children and consumer durables.

For example, to purchase a consumer durable it is necessary to make a down payment with one's own resources and to finance the remainder either with one's own or with borrowed resources. The economic uncertainty generated by a depression increases the reluctance to use own or borrowed resources and induces creditors to raise standards and screen applicants more carefully.[30] Therefore some purchases of durables would be postponed until economic conditions improved. The "purchase" of children, however, is less apt to be postponed than the purchase of other durables. The initial cost of children (physician and delivery charges, nursery furniture, expenses, and so on) is a smaller fraction of its total cost than is the initial cost of most other durables because expenditures on children are more naturally spread over time. Hence children can be "purchased" with a smaller down payment and with less use of borrowed funds than can most other durables.

There is still another reason why the "purchase" of children is less apt to be postponed. *Ceteris paribus*, the demand for a good with a lengthy construction period is less sensitive to a temporary economic movement than the demand for more readily constructed goods, since delivery is likely to occur when this movement has passed. The construction and delivery period is very short for durables like cars and quite long for

[29] An estimate of the desired change in births of planned families can be readily obtained if we assume that the distribution of contraceptive knowledge among U.S. whites is the same as among families in the Indianapolis study, that for planned families the actual change in births equals the desired change, and that for other families the actual change is nil. Then the desired change equals the actual change (averaging 42 per cent of the change in output) divided by the fraction of all births in planned families (31 per cent), or about 136 per cent of output. This is about half of the change for consumer durables.

[30] For evidence relating credit conditions to cyclical fluctuations in the demand for housing, see J. Guttentag, *Some Studies of the Post-World War II Residential Construction and Mortgage Markets*, unpublished Ph.D. dissertation, Department of Economics, Columbia University, 1958.

children. It takes about 10 months on the average to produce a pregnancy and this period combined with a nine-month pregnancy period gives a total average construction period of nineteen months. This period is sufficiently long to reduce the impact on the demand for children of temporary movements in income.

There are also some reasons why the "purchase" of children is more apt to be postponed. For example, since children cannot be bought and sold they are a less "liquid" asset than ordinary durables, and the economic uncertainty accompanying a depression would increase the community's preference for liquid assets. A more complete analysis would also have to take account of other factors, such as the accelerator and the permanent income concept, which may have produced different cyclical responses in fertility and consumer durables. Our aim here, therefore, is not to present a definitive explanation of the relative cyclical movement in fertility but only to suggest that economic analysis can be useful in arriving at such an explanation.

Although the data on cyclical movements in fertility appears consistent with our analysis, another piece of time series data is in apparent conflict with it. Over time per capita incomes in the United States have risen while fertility has declined, suggesting a negative relationship between income and fertility. Of course, many other variables have changed drastically over time and this apparent conflict in the secular movements of fertility and income should not be taken too seriously until it can be demonstrated that these other changes were not responsible for the decline in fertility. Three changes seem especially important: a decline in child mortality; an increase in contraceptive knowledge; and a rise in the cost of children.

The number of children in the average completed urban white family declined by about 56 per cent from 1870 to 1940. The decline in child mortality explains about 14 percentage points or 25 per cent of this decline.[31] Some evidence already presented indicates that a large secular increase in contraceptive knowledge occurred in the United States. It is not possible, however, to estimate its magnitude precisely enough to compare it to the decline in fertility.

I have emphasized that the increase over time in expenditures on children is not evidence that the cost of children has increased since the quality of children has also increased. Changes in the relative cost of children have to be assessed from indexes of the relative cost of given 43

[31] Taken from my unpublished paper "Child Mortality, Fertility, and Population Growth."

quality children. There are several reasons why the relative cost of a given quality child may have changed over time. The decline in child mortality decreased the cost of a given quality child, although it may have only a small effect. The growth of legislation prohibiting child labor and requiring education may have raised the cost of children, but largely made compulsory only what was being done voluntarily by most parents.[32] This is another aspect of the increase in quality of children and does not imply any increase in their cost. If such legislation raised costs at all, it did so primarily for the poorest families since they would be less apt to give their children much education. Therefore, legislation may have been partly responsible for the narrowing of fertility differentials by income class in the last fifty years.[33] The movement from farm to urban communities raised the average cost of children to the population as a whole since it is cheaper to raise children on a farm, but did not appreciably affect the cost within urban communities. Because technological advance has probably been more rapid in the market place than in the home, the imputed cost of time and effort spent on children probably rose, perhaps by a substantial amount. This discussion suggests that there was a secular rise in the cost of children which also contributed to the secular decline in fertility.

Secular changes in educational attainment, religious attachment, discrimination against women, and so on, may also have decreased fertility, and presumably there were changes other than the growth of income which increased fertility. It would take a major study—and even that might be inconclusive—to determine whether the factors decreasing fertility were sufficiently strong to produce a secular decline in fertility in spite of the secular rise in income. At present, it seems that the negative correlation between the secular changes in fertility and income is not strong evidence against the hypothesis that an increase in income would cause an increase in fertility—tastes, costs, and knowledge remaining constant.

III. Some Further Implications

Section II tries to show that the economic analysis of section I is very useful in understanding the effect of income on fertility. This section sketches some additional implications. Our understanding of temporal

[32] See G. J. Stigler, *Employment and Compensation in Education,* National Bureau of Economic Research, Occasional Paper 33, 1950, Appendix B.

[33] This analysis casts doubt on the view that the sharp decline in British fertility during the 1870's and 1880's resulted from the introduction of compulsory education. The decline was greatest in the upper classes which were least affected by this legislation.

fluctuations in births would be deepened if it were more widely recognized that births are "flows" to the "stock" of children, just as new car purchases are flows to the stock of cars. Flows are determined not only by variables determining stocks, but also by depreciation rates, acceleration, savings, and, as shown in our discussion of cyclical movements in births, by considerations of timing. The recent work relating births to parity shows that demographers as well as economists are beginning to stress the interaction between stocks and flows.[34] This work needs to be extended in a systematic fashion.

The discussion in section I made it clear that the quantity and quality of children are intimately related. An increase in income or a decline in the cost of children would affect both the quantity and quality of children, usually increasing both. An increase in contraceptive knowledge would also affect both, but would increase quality while decreasing quantity. The quality of children is very important in its own right, for it determines the education, health, and motivation of the future labor force. It is a major contribution of an economic framework to bring out the mutual interaction of quantity and quality—an interaction that has been neglected all too often in writings both on population and on the quality of the labor force.

It is often said that farm families are larger than urban families because of a difference in tastes. Since farmers have a comparative cost advantage in raising children as well as in raising foodstuffs, they would tend to be more fertile even without any difference in tastes. The rural advantage may not be the same at all qualities and, indeed, presumably is less at higher qualities where child labor and food are less important. Over time, rural as well as urban families have moved to higher quality children, and this may have contributed to the narrowing of urban-rural fertility differentials in recent decades. The influence of differences in the cost of children deserves much more systematic study, for it may partly explain not only these urban-rural fertility differences but also the secular decline in fertility up to World War II and the apparent secular narrowing of fertility differentials among urban economic classes.

In the Western World, birth rates in the early postwar period were well above rates of the thirties. In some countries, including the United States and Canada, they have remained at about the early postwar level; in others, including Great Britain and Sweden, they have drifted down to about their 1940 level; in still others, including France, they have

[34] Both economists and demographers found that wartime effects on stocks had important consequences for postwar flows.

drifted down to a position intermediate between their immediate pre- and postwar levels. The analysis in this paper does not readily explain these differences, but it does explain why birth rates in all these countries are well above levels predicted from their secular trends. The secular decline in child mortality and the secular increase in contraceptive knowledge were important causes of the secular decline in births. By 1945 the level of child mortality was so low that little room remained for a further improvement. Although contraceptive knowledge was not well spread throughout every layer of society, the room for its further improvement was also more limited than it had been. With the weakening of these forces, much of the steam behind the secular decline in birth rates has been removed. Positive forces like the growth in income are now opposed by weaker negative forces, and it is not too surprising that fertility has ceased to decline and even has risen in some countries.

Several recent studies of consumption have used a measure of family size as an independent variable along with measures of income and price.[35] This procedure is justifiable if family size were a random variable or completely determined by "non-economic" factors.[36] If, on the other hand, family size were partly determined by economic factors, this procedure would result in misleading estimates of the regression coefficients for the other independent variables. Thus, suppose family size were positively related to income, and food consumption varied with income only because family size did. The regression coefficient between food consumption and income, holding family size constant, would be zero, an incorrect estimate of the long-run effect of an increase in income on food consumption. One would not estimate the effect of income on gasoline consumption by finding the regression coefficient between gasoline consumption and income, holding the number of cars constant. For gasoline consumption might increase with income largely because the number of cars does, just as food consumption might increase because family size does. This discussion, brief as it is, should be sufficient to demonstrate that students of consumption economics need to pay more attention to the determinants of family size than they have in the past.

[35] See, for example, Theil, *op. cit.*, S. J. Prais and H. S. Houthhakker, *The Analysis of Family Budgets*, Cambridge, Cambridge University Press, 1955. Measures of family size often include not only the inner core of parents and their children but also other relatives living in the same household. My discussion refers only to the inner core; a somewhat different discussion is required for "other relatives."

[36] Prais and Houthhakker appear to believe that family size is determined by non-economic factors when they say "It might be thought that since household size is, in a sense, a noneconomic factor. . . ." *ibid.*, p. 88.

IV. Summary

This paper employs an economic framework to analyze the factors determining fertility. Children are viewed as a durable good, primarily a consumer's durable, which yields income, primarily psychic income, to parents. Fertility is determined by income, child costs, knowledge, uncertainty, and tastes. An increase in income and a decline in price would increase the demand for children, although it is necessary to distinguish between the quantity and quality of children demanded. The quality of children is directly related to the amount spent on them.

Each family must produce its own children since children cannot be bought or sold in the market place. This is why every uncertainty in the production of children (such as their sex) creates a corresponding uncertainty in consumption. It is also why the number of children in a family depends not only on its demand but also on its ability to produce or supply them. Some families are unable to produce as many children as they desire and some have to produce more than they desire. Therefore, actual fertility may diverge considerably from desired fertility.

I briefly explored some implications of this theory. For example, it may largely explain the postwar rise in fertility in Western nations, the relatively small cyclical fluctuation in fertility compared to that in other durables, some observed relations between the quantity and quality of children, and why rural women are more fertile than urban women.

I tested in more detail one important implication, namely that the number of children desired is directly related to income. Crude cross-sectional data show a negative relationship with income, but the crude data do not hold contraceptive knowledge constant. When it is held constant, a positive relationship appears. This view is supported by the positive correspondence between cyclical movements in income and fertility. The secular decline in fertility may also be consistent with a positive relationship since the secular decline in child mortality and the secular rise in both contraceptive knowledge and child costs could easily have offset the secular rise in income.

COMMENT

JAMES S. DUESENBERRY, Harvard University

I. For many years economists have taken variations in rates of population growth, and in family size, as *data* which help to explain various economic phenomena but which cannot themselves be explained in terms of economic theory. Becker has done us a real service in bringing economic analysis to bear on the problem once more. He has not only worked out 45 the implications of traditional economic theory for demographic theory but has also gone some distance in testing those implications against the empirical data.

Becker argues that those couples with sufficient contraceptive knowledge to control births have to decide how many children to have. For most people, children produce certain satisfactions and have a net cost. In those circumstances we expect (with some qualifications) that the number of children per family will rise with income just as we expect the number of cars or chairs or cubic feet of housing space per family to rise with income. But just as in those cases we expect the quality of cars or chairs or houses to rise with income as well as the number, we also expect the *quality of children* to rise with income as well as the number. That is, we expect the children of the rich to be better housed, fed, and educated than those of the poor.

Becker then qualifies the argument by taking into account the fact that in some circumstances children may yield their parents a net income instead of having a net cost. In that case the theory of investment is relevant as well as the theory of consumption. He has brought in a number of other considerations which I need not review but which lead to only minor qualifications of his main arguments.

After reviewing the implications of economic theory, Becker then faces the fact that for many years the raw data on differential fertility have shown a fairly strong negative relationship between variations in income and variations in numbers of children per family. Moreover, until recently the average number of children per completed family has been declining although average family income has been rising secularly.

Becker maintains that the negative correlation between income and family size is due to the negative association between income and knowledge of contraceptive methods. I think that most of us would agree that differential knowledge does explain a large part of the apparent negative relation between income and family size.

The evidence of the Indianapolis study certainly supports that conclusion. Becker, however, tries to use the study to support his conclusion that there should be a positive association between income and family size. I must say that the evidence he cites did not strike me as exactly overwhelming.

The empirical evidence offers, I would say, rather ambiguous support for Becker's hypothesis. That may be because we have only a limited amount of the right kind of data but there are, I think, some reasons for thinking that Becker's theoretical case may not be so open and shut as appears. Those reasons have to do with the nature of the "cost" of children and with the limitations on the possibility of substitution between quantity and quality of children.

II. Becker has taken the occasion to correct the simple-minded who fail to distinguish between the cost of children of given quality and expenditure per child. Now, of course, it is correct to regard changes in prices (or relative prices) of a given quality of a good as changes in the cost of that good and changes in amount or quality of the good purchased (at a given price schedule) as changes in expenditure not involving changes in cost. But not all of those who say that the cost of children rises with income are so simple-minded as Becker suggests, though their language may not be exact. What Leibenstein, for example, appears to mean is that the expenditure per child which the parents consider to be necessary rises with income.

Questions of semantics aside, there is an important substantive difference between Becker's approach and that taken by economists whose approach is, if he will excuse the expression, more sociological.

I used to tell my students that the difference between economics and sociology is very simple. Economics is all about how people make choices. Sociology is all about why they don't have any choices to make.

Becker assumes that any couple considers itself free to choose any combination it wishes of numbers of children and expenditure per child (prices of particular goods and services being given). I submit that a sociologist would take the view that given the educational level, occupation, region, and a few other factors, most couples would consider that they have a very narrow range of choice. To take only one example, I suggest that there is no one in the room, not even Becker, who considers himself free to choose either two children who go to university or four children who stop their education after high school. It may be said that that still leaves lots of room for variation, but I think it can be said that no one in this room considers seriously having, say, four children who attend third-rate colleges at low cost per head or three who attend better ones.

For this audience I need not go through the whole routine about roles, goals, values, and so on. It will be sufficient to remark that there is no area in which the sociological limitations of freedom of choice apply more strongly than to behavior in regard to bringing up children.

Effective freedom of choice between quantity and quality of children is also limited by more mundane and mechanical considerations. The principle of substitution which is at the basis of Becker's argument 76 suggests that if the parents have low quality children, as he puts it, they can spend more of their income on something else. Quality of children means, in Becker's terminology, nothing more than expenditure per child (with a given price schedule). But in many respects the standard of living of the children is mechanically linked to that of the parents. Is it possible to have crowded housing conditions for the children and uncrowded conditions for the parents? As the father of four I am in a position to answer with an unqualified negative. Children may eat a different menu from their parents, but if so, it is because they *like* peanut-butter sandwiches. I could go on but I am sure it's unnecessary. A final point in this connection is the non-cash cost of improving quality in children. Becker has used the term quality as though it were just another expression for expenditure at constant prices. But in the more ordinary sense of the term, quality has to be bought with time as well as money. Most parents think (probably mistakenly) that their children are better off if the parents spend time with them. Now time can be bought in the sense that domestic help and appliances can be bought to free time for other things. But even if one had nothing else to do, the marginal disutility of Cub Scout and PTA meetings rises rapidly. These non-cash costs must certainly be of some importance in determining family size.

Becker will say that this is merely an aspect of the diminishing marginal utility of numbers of children. He is correct, of course, but the investment of time in children is not a matter of individual choice any more than the investment of money. The time which parents spend on children is largely determined by social conventions. Those conventions differ among social classes. Since social class is often associated with income, the non-cash costs I have mentioned will influence the apparent relation between income and family size in many cross-sections.

Those considerations lead me to the following conclusions: (1) the effect of income on family size which Becker expects will be greatly weakened by the tendency for the standard of living for children to advance more or less proportionately with that of the parents, and (2) standards of education and of expenditure of time on children will vary with social class. Social class in turn will be associated with income but not in a unique way. In some societies it may turn out that the "cost" of children rises faster than income, in others more slowly.

To put it more generally, economic factors are certainly likely to influence the number of children born to those who are able to plan births. But I do not feel that we are likely to find out much about their influence by simply drawing an analogy between children and durable goods.

BERNARD OKUN, Princeton University

Rupert Vance, in his Presidential address before the Population Association of America at Princeton in 1952, prescribed for demographers "a good stiff dosage of theory, adequately compounded."[1] Gary Becker has heeded Vance's prescription, and in addition has presented us with an interesting and challenging paper.

The essence of Becker's theory of the demand for children can be reduced to two propositions. First, as family income increases, parents will provide their children with a higher level of living. Second, as income increases, parents will probably increase the number of children which they demand. My comments will strongly concur with the first proposition, but will question the second.

In his economic theory, Becker treats children like consumer durables. His justification for treating children in this fashion is that, like automobiles, children are a source of utility and require a considerable outlay of expenditures. It is then contended that the nature of the demand for children is similar to that for automobiles—higher income families demand more automobiles and better quality automobiles; similarly, they demand more children and spend more per child.

In his analysis, Becker distinguishes two components in the expenditures on a child. The first relates to the size and nature of the basket of goods and services which the child consumes. The second relates to the prices of the goods and services which the child consumes. Expenditures per child can change either because of a change in the composition of the child's consumption basket, or because of a change in the prices of the components of the basket. Becker equates the additional utility received by the parents resulting from an increase in expenditures of the first kind, i.e., expenditures directed toward an improvement in the composition of the basket, with the increase in the "quality" of the child. Such expenditures shall be referred to as quality expenditures. Becker defines a change in expenditures per child resulting simply from a change in the prices of one or more components of the basket, that is, the second kind of change, as a change in the "*cost*" of a child. Thus, in Becker's framework, the concept of a change in the cost of a child is a very narrow one. It refers only to a change in expenditures per child, where the child continues to consume a fixed basket of goods and services.

A word of caution is necessary here. Becker has related quality expenditures to the amount of utility that children provide for their parents. This relation, however, is valid only within a given family unit. One

[1] Rupert B. Vance, "Is Theory for Demographers?", *Social Forces*, Vol. 31, 1952, p. 13.

cannot conclude that the family which purchases less for their child derives less total utility from him in comparison with the family which purchases more. Such a conclusion implies an unwarranted inter-personal comparison of utility. If the Jones boy is paying for dancing lessons while the Smith boy is reading a borrowed copy of Marshall's *Principles*, one cannot conclude that the Joneses are deriving more utility than the Smiths with regard to these alternative pursuits of their respective sons.

Becker has imputed much theoretical significance to his distinction between the concepts of quality expenditures and cost expenditures. For example, in his conclusion, it is suggested that a secular increase in the "cost" component contributed to a secular decline in fertility. Nothing is said about the effect of a secular change in the "quality" component. We shall return to this point shortly.

Becker's paper suggests a second distinction between the quality and cost components. The cost of a child, which depends on the prices of commodities and services, is determined by the market forces of supply and demand. Cost, therefore, is not a family decision variable, and is independent of family income. On the other hand, quality expenditures are a family decision variable and are positively related to income. Becker stresses that higher income families have higher quality children (spend more per child), and that this is a voluntary decision. According to Becker, "The rich simply choose higher quality children."

Is the quality expenditure component purely a decision variable? Do the rich really have a choice? To a large extent, I submit that they do not. It is almost impossible to conceive of a child who is raised at a much lower level of living than that of his parents. He lives where they live, tends to eat what they eat, and in general, as a matter of course, shares about the same standard of living that they do, by virtue of his living with them. Surely, the child cannot be sent to live in the slums of the Lower East Side while his parents dwell in a penthouse on Park Avenue. Thus, automatically, when parents raise their own level of living, their child's is also raised, and quality expenditures per child *must* rise.

When demographers, economists, or anybody, for that matter, speak of a decline in the birth rate, they are referring to a decline in the number of children born divided by either total population or some component of population (for example, women of childbearing age). The total number of children born is not weighted by a quality index—every child is given a weight of unity. Thus, if we are to turn to economic theory for an explanation of the decline in the birth rate, one must argue that *expenditures* (as opposed to "cost") per child have risen over time. It is theoretically irrelevant to distinguish between quality expenditures and "cost" expenditures. Becker's conclusion that an increase in "cost" expenditures per child contributed to the decline in the birth rate is useful only if he is referring to a decline in the birth rate of children of homogeneous quality. But since quality expenditures per child have increased over time, this is not the case. Therefore, in explaining birth-rate trends or fertility differences by income, where quality expenditures per child vary in a systematic way, although not reflected in the measurement of the birth rate, Becker's distinction between quality and "cost" expenditures is not useful. The relevant economic variable is simply expenditures per child.

It is for the reasons cited above that I wish to defend Harvey Leibenstein against Becker's criticism. I might add that I have a vested interest in doing so since an argument similar to Leibenstein's appears in one of my own writings.[2] Becker quotes Leibenstein as follows: "The conventional costs of child maintenance increase as per capita income increases. The style in which a child is maintained depends on the position and income of the parents; therefore, we expect such costs to rise as incomes rise. . . ."[3]

Although Becker is correct in noting that Leibenstein failed to make the statistical distinction between quality and "cost" expenditures, this is not the relevant distinction that should be made in a theory of the demand for children which attempts to link income and fertility. The relevant distinction, if any, is voluntary expenditures versus involuntary expenditures. To a large extent, the higher income and social position of the family *require* that it spend more per child. In this sense, a rise in income necessarily results in a relative increase in expenditures per child. This, economic theory suggests, would have a depressing effect on the quantity of children demanded. Consequently, the quantity income elasticity of demand for children is quite low. Indeed, for most of the income range, the quality income elasticity may be so high that it contributes to a negative quantity income elasticity of demand.

On the other hand, the quantity income elasticity for consumer durables is generally much higher. Several distinctions between children and consumer durables account for this. One distinction has already been alluded to in the preceding discussion. The quality of a child as a function of income is less of a decision variable than, for example, the

[2] See Bernard Okun, *Trends in Birth Rates in the United States Since 1870*, The Johns Hopkins Press, Baltimore, 1958, pp. 177–180.

[3] Harvey Leibenstein, *Economic Backwardness and Economic Growth*, John Wiley, 1957, pp. 163–164.

quality of an automobile. This may tend to cause the quality income elasticity of consumer goods to be less than that for children, and this would allow a higher quantity income elasticity for consumer goods.

For consumer goods, quantity appears to be a closer substitute for quality than in the case of children. Two lower-price cars may be considered equivalent to one high-priced car for the high income family. But is it just as likely that this family would be indifferent toward having two children who are untrained or not well-educated, or having one well-educated child? Probably not. In fact some parents may derive disutility if their children fall below their quality standards.

Probably a more common occurrence among higher income two-car families is that they will own one high-priced car, and also own a lower quality second car. Are they apt to follow a similar policy with regard to children—that is, we already have one son who is a Princeton graduate, so we can plan to finance our second son only through high school? I think not. This unwillingness to diminish the quality of successive children tends to diminish the quantity income elasticity of demand for children, relative to that for commodities.

For reasons cited above, it is suggested here that unlike the typical case for consumer durables, the quantity income elasticity of demand for children may well be negative, or if positive, be very low. Briefly restated, our main point is that as income increases, quality expenditures per child do—and in a large measure *must*—increase to such an extent that parents tend to reduce their demand for children. Note, however, an exception to this proposition—in the very high income families, where family size tends to be larger than in the middle income families, it seems evident that parents can satisfy their quality requirements without having to restrict the quantity of children by the same degree as the somewhat lower income families.

The weight of the empirical evidence presented at this National Bureau conference as well as that of other studies supports the proposition that, for most of the income range, fertility varies inversely with income. Becker contends that these data do not apply to his theory because lower income people have inadequate knowledge of birth control. If all families had perfect control over family size, Becker contends that the relationship would be reversed. In this fashion, he defends the position that the quantity income elasticity is positive.

Becker notes correctly that knowledge of birth control is ever-increasing. If the quantity income elasticity of the demand for children is positive, one would expect the inverse birth-rate differentials by income to be ever-narrowing as birth-control knowledge continuously spreads. While a narrowing trend has occurred, it has been far from persistent. For example, Clyde Kiser found, in comparing family size and income in the United States in 1952 and 1957 that "the apparent enlargement of the differentials by income was quite pronounced."[4] This is a finding which clearly weakens Becker's point that differentials in birth-control knowledge are the factor accounting for the inverse relation between income and family size.

There is also strong evidence to suggest that where social or economic forces prevail which tend to diminish the size of family desired, the lack of knowledge of modern birth-control techniques is not an obstacle in the path of declining family size. For example, according to Whelpton's figures, in the Southern United States, which was largely rural in 1800 as well as in 1870, the fertility rate declined by more than 50 per cent from 1800 to 1870—a period long before modern birth-control methods were known.[5] A similar experience occurred in France after 1800. This evidence tends to weaken the contention that an improvement in birth-control knowledge explained a significant share in the secular decline in fertility.

Becker also attempts to support his thesis that income and the demand for children are positively related by pointing to the positive conformity of the birth rate to fluctuation in the business cycle. I feel that during a business cycle, the time period may be too short for parents' views and standards regarding quality of children to change significantly as a result of a change in income. During the downswing, parents will strive to maintain their standard of living and the quality of their children. Faced with this economic pressure, they will postpone having more children. During the peak stages, income may be rising faster than child-quality standards, and couples can think in terms of having more children without encroaching on their accustomed level of living and their child-quality standards.

The well-known "making-up" theory may partly account for the positive association between fluctuations in income and fertility over the business cycle. This theory holds that the business cycle mainly affects the timing of the arrival of children, but has no or but negligible effect on completed family size. For all these reasons, a positive association between changes in income and the birth rate over the cycle is readily explained.

[4] Clyde V. Kiser, "Differential Fertility in the United States," in this volume.

[5] P. K. Whelpton, *Forecasts of the Population of the United States 1945–1975*, Bureau of the Census, 1947, p. 16.

49

In the long run, however, standards of living and child quality standards adjust to a secular rise in income. The secular rise in income causes an increase in the quality of children, and therefore expenditures per child rise. This tends to diminish the quantity of children demanded, and the well-known empirical inverse relation between income and the birth rate reasserts itself.

51

Population Studies Vol. XXII, nº 1 March. 1968

Are Babies Consumer Durables?

A Critique of the Economic Theory of Reproductive Motivation*

JUDITH BLAKE

Never before have couples been able to control so effectively the number of children they will have. Although involuntary factors still affect family size, continuing advances in contraceptive techniques make deliberate choice an ever more important determinant of fertility. But what factors determine the size of family people will choose ?[1] One type of answer advanced in recent years by Gary Becker views reproductive performance simply as economic behaviour.[2] Couples, he believes, desire fewer children when poor, more when rich.

Becker places family-size goals in the framework of economic theory by treating children as a consumption good analogous to cars, houses and refrigerators.[3] For almost all consumer durables, he says, there is income elasticity with respect to both quality and quantity. At higher income levels, families purchase both better and more units, and quality elasticity is greater than quantity elasticity. This framework suggests to him that '. . . a rise in income would increase both the quality and quantity of children desired . . .'[4] He does not believe that, according to economic analysis, the quantity income elasticity of demand for children can be expected to be negative.[5] In other words, according to Becker, the relationship of family-size desires and income should be positive – the more income (that is, the greater the purchasing power) the larger the family desired.

The theory underlying this view merits critical evaluation because this 'application' of economic analysis to demography has implications for both science and policy. Scientifically, it is hailed by some economists as an illustration of what an economic framework can do for population

* The research discussed is supported by a grant from The Equitable Life Assurance Society to International Population and Urban Research, Institute of International Studies, University of California, Berkeley. The author wishes to thank Kingsley Davis for his advice and criticism, and Valerie Caires, Katherine Carter and Barbara Heyns for their assistance in processing the studies involved in this analysis. The report is also indebted to General Research Support Grant of the National Institutes of Health (1501-TR-544104) for assistance to Statistical Services, School of Public Health.

[1] Modern conditions of low infant and child mortality, accompanied by a virtual disappearance of the value of children as production goods or as principal sources of security for parents, have eradicated some of the most cogent reasons for having large families. For a discussion of changing family-size motives in the framework of utilities and costs, see Harvey Leibenstein, *Economic Backwardness and Economic Growth* (New York, 1957), Chapter 10. A discussion of the small family goal in the context of modern familial and extra-familial roles and statuses may be found in Judith Blake, 'Demographic science and the redirection of population policy', in Mendel Shêps and Jeanne Clare Ridley (eds.), *Public Health and Population Change* (Pittsburgh, 1965), pp. 41–69.

[2] Gary Becker, 'An economic analysis of fertility', in National Bureau of Economic Research, *Demographic and Economic Change in Developed Countries* (Princeton, 1960), pp. 209–240.

[3] *Ibid.*, pp. 210–211.

[4] *Ibid.*, p. 217.

[5] *Ibid.*, p. 215.

studies. For example, in his presidential address before the American Economic Association, Spengler speaks of Becker's work as a 'notable contribution' to the economic analysis of fertility. Spengler sketches further how elaborations of Becker's thesis, using the consumer durables analogy, 'should be able to explain changes in gross reproduction'.[6] Demographers have been more cautious in their admiration, but their reticence has been based on lack of evidence for the thesis rather than on theoretical grounds.[7] With regard to policy, the thesis implies that the long-existing but nevertheless unfortunate pattern of differential fertility whereby the poor have more children and the rich have fewer – a pattern still found in the United States – will be reversed if only everyone has 'access' to contraception. Since this notion is one of the planks in the platform for large-scale, publicly supported family-planning programmes, considerable interest attaches to whether the actualization of family-size preferences will, in fact, be likely to transpose the present inverse relation of family size and finances. If Becker is wrong, if the inverse relation of family size and income is not because of lack of contraception among the poor, but is also due to the desire for larger families among them, then a costly family-planning programme based on assumptions such as he advances will not change the pattern of differential fertility. In fact, if one takes into account the attention to sub-fertility and infertility given by such programmes, the change in average family size among the poor may not be very great.[8]

In evaluating the Becker thesis as an example of economic reasoning about reproductive behaviour, we shall endeavour to answer two questions. First, do empirical data support the thesis of a positive relation between family-size desires and income in modern societies? And, second, if not, why not? – what fallacies and omissions inhere in this type of analysis as applied to reproductive motivation?

EVIDENCE CONCERNING INCOME AND FAMILY-SIZE DESIRES

Becker is aware that the preponderance of data on *actual* family size in relation to income suggests that the quantity income elasticity of 'demand' for children is negative.[9] Nonetheless, he tries to discount the inverse association between family size and purchasing power by arguing that the

[6] Joseph J. Spengler, 'The economist and the population question', *American Economic Review*, 56 (March 1966), p. 14. Spengler's enthusiasm is not shared by James Duesenberry, an economist specializing in problems of consumer behaviour. See Duesenberry's discussion of Becker's approach in the same volume as the Becker article, *op. cit.*, pp. 231–234.

[7] Demographers recognize that Becker's thesis is an elaboration of the notion long held in demographic thinking that income and family size will be positively related when everyone has access to contraception. See, for example, Amos Hawley, *Human Ecology* (New York, 1950), pp. 118–119. Ronald Freedman has gone so far as to suggest that perhaps the expectations themselves may be misguided. See, in particular, Clyde V. Kiser (ed.), *Research in Family Planning* (Princeton, 1962), pp. 221–223.

[8] For more detailed discussion of this point, see Judith Blake, *op. cit.*, and Kingsley Davis, 'Population policy: Will current programs succeed?', *Science* 158, 1967, pp. 730–739.

[9] Even where there sometimes seems to be a genuine change in the usually inverse relationship between family size and income, one typically finds that the apparently positive relation is due to some spurious factor, to some short-run happening, or to involuntary factors affecting fertility (the influence of which do not, of course, bear on a thesis concerning family-size desires). For example, in the United States, the simple measure of controlling for age rids one of the suggestion that family size may be higher where the husband has more income. An additional control for such a particularistically related variable as being Catholic or not is similarly helpful. See P. K. Whelpton,

52

intervening variable in this case is access to contraception. Greater access gives wealthier people more control than poorer people over the 'supply' of children. In his words, 'If knowledge of contraceptive techniques did not vary with income, the relation between actual fertility and income would equal that between desired fertility and income.[10]

It is clear that he assumes, without direct evidence, that wealthy people *desire* more children than do poorer ones. Not using direct data on reproductive preferences themselves, his only evidence for this assumption is the actual fertility of populations in which the contraceptive factor has been 'equalized'.

But such indirect reasoning is dubious because it is difficult to think of 'equalizing' contraceptive knowledge and practice apart from reproductive motivation itself – a point to be discussed later. In the meantime, even if we accept Becker's idea, we would still expect him to give systematic attention to populations where contraception has been 'equalized'. His actual procedure is more casual. He simply points out that there are exceptions to the inverse relation of family size and income, stating that 'not all evidence is one way.'[11] When we examine the few instances in which he claims that contraception has been 'equalized', we find that the positive relationships between family size and income, on which he lays such stress, result either from sample biases that he ignores, or from factors quite irrelevant to his analogy of children to consumer durables.[12] To show this, let us scrutinize the principal studies he mentions.

The Indianapolis Study. One case he cites is a special group in the Indianapolis Study – a sub-sample of couples all of whose children were 'number and spacing planned'.[13] Since the fertility

---

Arthur A. Campbell, and John E. Patterson, *Fertility and Family Planning in the United States* (Princeton, 1966), pp. 102–104. Data from the 1960 census on number of children ever born per 1,000 white wives of completed fertility who had born children show an inverse relation with husband's income up to the highest income class, where there is a rise of, at most, one-tenth of a child. It is not possible, however, to eliminate the effect of the fertility of upper-income Catholics from the data. See U.S. Bureau of the Census, *U.S. Census of Population: 1960. Subject Reports. Women by Number of Children Ever Born. Final Report PC (2)-3A*, (Washington, D.C., 1964), Table 37. Recent data for European countries have been discussed in a previous paper: Blake, 'Demographic science and the redirection of population policy', *loc. cit.* In that paper the thesis was advanced that recent European differentials (some of which may be positive) are not the result of some different motivational pattern in Europe, but of the different recovery rates among social classes as they moved out of the trauma of depression and war. It seems hazardous to regard differential European fertility of the past 25 years as if it represented the denouement of a long-run trend. It is particularly hazardous to interpret it entirely as the result of voluntary factors with respect to fertility, when one knows that there were, as well, changing marriage patterns and gross differences in health, medical care, nutrition and the like. To generalize from European fertility during depression and ravaging war makes as much sense as generalizing from the demographic situation during the plague. An analysis of recent European fertility differentials is currently being prepared by the author.

[10] Becker, *op. cit.*, p. 220.

[11] *Ibid.*, p. 218.

[12] The only direct data on family-size desires cited by Becker come from a Detroit area survey in which it is generally recognized that the wording of the question asked may well have induced lower-income respondents to lower their ideals. See Ronald Freedman, David Goldberg and Harry Sharp, ' "Ideals" about family size in the Detroit Metropolitan Area: 1954', *Milbank Memorial Fund Quarterly*, 33 (April 1955), pp. 187–197.

[13] See P. K. Whelpton and Clyde V. Kiser (eds.), IX, 'Fertility planning and fertility ratio by socio-economic status', *Social and Psychological Factors Affecting Fertility*, Vol. 2 (New York: Milbank Memorial Fund, 1950), pp. 359–415.

53

of these couples varies directly with the husband's income, Becker claims that they show what would be generally true if contraception were more widely available. However, he overlooks the fact that the various income groups are disproportionately represented in the 'number and spacing planned' category. Any conclusions based on couples in this category are, therefore, based on income classes whose probabilities of being included are very dissimilar. For example, whereas those in the highest income category constitute 8% of the total Indianapolis sample, they are 14% of the 'number and spacing planned' group. Those who are in the lowest income class constitute 19% of the total sample, but only 11% of the 'number and spacing planned' group. Moreover, if one looks at the fertility-planning status of the various income groups, one finds that efficiency of planning varies directly with income. 45% of the upper-income group have planned their families completely and 15% have experienced 'excess fertility'. The situation is just reversed for the lowest income group – 16% have families that are 'number and spacing planned' and 42% have 'excess fertility'.[14]

This statistical bias might not seem important were it not for the additional fact that being a member of the 'number and spacing planned' category is not a mysterious attribute unrelated to family-size motivation. Planning both the number and spacing of one's children requires high motivation, more so for a poorer person than a richer one. It thus seems likely that poorer people in the 'number and spacing planned' category will be those who are atypically motivated (for their class) to have families of a limited size. This point comes out clearly in the 1960 Growth of American Families study which does not substantiate the results from the previous Indianapolis Study. Among couples with completely planned fertility there are no statistically significant differences in the number of births expected or wanted among groups ranked according to husband's income. As the authors suggest, 'couples who are willing and able to exercise the care needed to control fertility so well share certain family-size values that cut across ordinary social and economic class lines. It is only within this relatively small group that the extinction of socio-economic differences in fertility often predicted for all couples appears to have occurred.'[15]

Reproduction in Pre-war China. For additional support, Becker goes to far-away pre-war China. On the basis of one article, published in 1933, he notes that contraceptive knowledge in China was 'said to be very primitive in *all* income classes . . . and a positive relation between fertility and income also seemed to prevail . . .'[16] What Becker overlooks is that reproductivity in traditional China bore little or no relation to a 'consumer-durables' approach to childbearing. The traditional Chinese case was a classic instance in which children – especially male children – functioned in both an economically and ceremonially productive sense. They were at the opposite extreme from 'consumer durables'. Under such conditions one expects to find a positive relation between fertility and income, other things being equal, but for quite different reasons from his.

[14] *Ibid.*, p. 384.

[15] Whelpton, Campbell and Patterson, *op. cit.*, pp. 240–241.

[16] Becker, *op. cit.*, p. 221. The reference is to Herbert D. Lamson, 'Differential reproductivity in China', *Quarterly Review of Biology*, 10 (September 1933), pp. 308–321.

Even so, if one looks up the one article he cites, one finds that even the statistical regularities are open to question.[17]

Subscribers to a Consumers' Union. Becker also utilizes some previously unpublished data on 'family income, education, earners, and dependent children of a sample of the subscribers to Consumers' Union'.[18] This group, he claims, 'is particularly valuable for our purposes since it primarily consists of families with a keen interest in rational, informed consumption. If my analysis is at all relevant, fertility and income should be more positively related in this group than in the U.S. population as a whole.'[19]

The figures he gives are reproduced in Table 1. From them Becker concludes that the 'income elasticity is about 0·09 and 1·14 for graduates of a four-year college and of a graduate school respectively. These data, then, are very consistent with my analysis, and indicate that well-informed families do have more children when their income increases.'[20]

TABLE 1. *Reproduction of Becker's table on average number of dependent children for single-earner families with head age 35–44 in a sample of subscribers to Consumers' Union, April 1958**

Income class	Average number of dependent children by education class of head			
	High school graduate – or less	Some college	Graduate of four-year college	Graduate degree
Less than $3,000	2·43	1·61	2·50	2·17
$3,000– 3,999	2·14	2·47	2·18	2·23
4,000– 4,999	2·70	2·40	2·04	2·18
5,000– 7,499	2·68	2·73	2·88	2·67
7,500– 9,999	2·80	2·94	3·00	3·03
10,000–14,999	2·89	3·03	3·12	3·23
15,000–24,999	2·85	3·04	3·04	3·31
25,000 and over	3·12	3·23	3·28	3·60

* Reproduced from Becker, *loc. cit.*, p. 221.

[17] Lamson's data on family size were obtained from students who reported on births and child mortality in their families. He himself admits that: 'It is quite likely that in these families a certain number of miscarriages, stillbirths, and abortions have been omitted through failure of the reporting students either to know or to state such facts concerning the puerperal history of their mothers' (pp. 308–309). Lamson assumes, however, that these errors are randomized over the different groups he is studying. But, this would only be true if the lower economic groups were no more subject to such risks than the upper economic groups. Data such as these cannot take into account the role of female infanticide, or the sale of infant girls into prostitution. Moreover, many of the findings are themselves highly suspect. For example, the highest fertility of all is found among the higher-educated Christian families, yet Lamson does not tell us whether these families were Catholic or not. If many of them were, their behaviour (as converts) would include neither contraception nor infanticide. In addition, although Lamson examines concubinage in relation to fertility (and finds that concubinal families do not have the highest reproduction), he does not bring out that concubinage was concentrated at the upper-income levels. He thus overlooks the fact that upper-income, non-Christian families had at their disposal a hedge against wifely infertility or sub-fecundity. The *average* fertility of such families could be kept high through the reproductive services of concubines. 55

[18] Becker, *op. cit.*, p. 221.

[19] *Ibid.*, p. 221.

[20] *Ibid.*, p. 222.

However, the data appear to suffer from statistical biases of some importance to his thesis. Noteworthy is the question of what makes the children in this sample eligible to be considered 'dependent'. Since Becker does not mention this problem, we may assume that dependency relates to an age limitation and/or to an economic condition. If dependent children are those under 18 (or under 21), then the presence or absence of such children in families whose head is aged 35–44 depends greatly upon when the head married and on the rate at which he formed his family. Since heads with lower incomes on the average, will have married and started their families earlier than those with higher incomes, the distribution of dependent children by income may well simply testify to the later marrying and wider child-spacing habits of individuals with higher incomes.[21] That this suspicion is not misplaced is suggested by the strong positive relationship between education and number of dependent children among those in the middle- and high-income brackets. In fact, it turns out that the highest number of dependent children of all is found among those with graduate degrees who occupy the $15,000–24,999 and $25,000-and-over income brackets! If 'dependency' is defined economically rather than chronologically, then the positive relationship between income and dependent children is doubly enhanced. Not only will the better-off group have married later, but they will have been more likely than individuals with low incomes, to tolerate economic dependency among offspring for longer periods of time. For example, the outstandingly 'fertile' wealthy, graduate-educated cases may simply be instances of families who are supporting their children through graduate and medical schools.[22]

Fertility of Stockholm Families. Citing data on Stockholm compiled by Edin and Hutchinson, Becker says: 'Contraceptive knowledge is said to be diffused among all income classes in Stockholm, and the fertility of Stockholm families from 1917–1930 was positively related to income.'[23] However, if we may judge from the account by David Glass of the Swedish birth-control movement as late as 1937, it is most doubtful whether contraceptive information of a modern sort was widely available in Sweden during the 1920's. In fact, even in Stockholm, Glass calls attention to the backwardness of the situation and the reluctance to have public discussion of contraception.[24] Hence, the actual fertility of Stockholm families of 1917–30 does not seem to be relevant as a basis for conclusions

[21] Actually, of course, early marriage and early childbearing may well exert an independently negative effect on future income. See Ronald Freedman and Lolagene Coombs, 'Childspacing and family economic position', *American Sociological Review*, **31** (October 1966), pp. 631–648.

[22] The apparent income elasticity of Becker's Consumers' Union example is further vitiated by the fact that a number of the 'single-earner' families may include as a 'head' a woman instead of a man. If so, this situation is more likely to occur at low-income levels than at high, both because of greater marital instability at lower levels and because women's incomes are lower. Women whose marriages have been dissolved are likely to have fewer children than others of their income class, to have had them at younger ages than male heads, and to have a smaller probability of child dependency in an economic sense after the children reach age 18. Furthermore, the inclusion of single-earner families alone leaves out families in which the wife may be working. Since the wife is more likely to be working among lower than among upper income families, and especially among those where there is a sense of economic stringency (perhaps because of the number of children involved), an additional bias is introduced. Finally, it is worth noting that in evaluating the magnitude of some of the figures on family size given, one should take into account the number of cases (not given by Becker). In the various educational categories at $25,000-and-over for 1958, the number may well be too small to have any significance.

[23] Becker, *loc. cit.*, p. 220. The data are from Karl Arvid Edin and Edward P. Hutchinson, *Studies of Differential Fertility in Sweden* (London, 1935), pp. 69–87.

[24] D. V. Glass, *Population Policies and Movements* (Oxford, 1940), pp. 319–322.

56

57

concerning *desired* family size. Furthermore, the data are only for *marital* fertility; they refer to births to couples who married during the four-year period 1917–20 and who were still living together in 1930. No account is taken of births prior to marriage. Yet during the period 1919–22, for example, 28% of all live births in Stockholm were illegitimate.[25] No data are available on the class distribution of the illegitimate births, but a reasonable assumption would be that they were skewed toward the lower-income brackets. If so, this would probably reverse the association between fertility and income.[26]

Perhaps most disturbing methodologically is the fact that the Stockholm material referred only to families living in the city at *two* censuses. Those families broken by death and divorce as well as those moving out of Stockholm were not included. It seems most likely, therefore, that lower-income families who had more than a very small number of children would of necessity move out of Stockholm as the family grew, since they could not compete with upper-income families for scarce housing. Typically, those poorer families would remain in the city when fertility was very low. Finally, it is not at all clear that the 'consumer durables' approach is applicable to upper-income European (including Swedish) families of more than 40 years ago. Such families were widely characterized by inheritance of occupation and by businesses and professional establishments based on kinship. If these establishments were to be carried on at all, they required some attention to reproduction regardless of the stringency of housing and so forth at the time.

In sum, the data on income and family size presented by Becker turn out to be either biased in favour of his thesis through sampling distortion, or irrelevant by virtue of being cases in which children had the status of being production goods rather than simply consumer goods.

Empirical Data on Family-size Preferences

Given the difficulty of estimating differential reproductive preferences from data on actual performance, direct evidence on the preferences themselves take on crucial importance in an evaluation of Becker's thesis. He did not utilize such evidence, but, in another paper, I have presented data on family-size ideals by income and economic status from 13 polls and surveys in the United States using national samples of the white population. The materials extend over a 30-year period from 1936 through 1966.[27]

These studies do not confirm Becker's expectation of a rise in reproductive desires with a rise in income. The mean family-size ideals for white men and women are reproduced in Table 2. The

[25] Edin and Hutchinson, *op. cit.*, p. 65. Professor D. V. Glass has kindly called my attention to the fact that he himself voiced this criticism of the Stockholm data some 32 years ago. Cf. *Eugenics Review*, **27**, 4 (January, 1936), pp. 297–301. Moreover, as Professor Glass notes in his review, R. A. Fisher made the same point when preliminary results of the Edin and Hutchinson work first appeared in 1929, *ibid.*, p. 300.

[26] Becker's interpretation also takes no account of the historical circumstances. The family as an institution was being traumatized in Stockholm at the time due to severe housing shortages. In this type of situation the advantage of extra income to obtain the bare necessity of housing for a family was quite considerable. It is notable, too, that the overall completed family size of even the largest families was extremely small. The highest average family size in the data referred to by Becker was 1·85 children (even among the wealthiest and best educated). A situation in which *all* couples experience severe housing difficulties – a particularistic one for Sweden – does not seem to be a good example of Becker's thesis, especially when the problem of illegitimacy is overlooked as well.

[27] Judith Blake, 'Income and reproductive motivation', *Population Studies*, **21**, 2 (November 1967), 185–206.

TABLE 2. *Mean number of children considered ideal by white males and females according to economic status or income, United States, selected years, 1936–1966*(1)

Date	Age range	Economic or income levels(2) 1	2	3	4	Total	(*N*)
		Females					
1936	21+	3·1	3·0	3·3	3·4	3·1	(527)
1941	21+	3·1		3·2	3·3	3·2	(918)
1943	20–34	2·9	2·7	2·8	2·6	2·7	(2,379)
1945	21+	3·4		3·5	3·6	3·5	(1,408)
1947	21+	3·0		3·2	3·3	3·3	(1,280)
1948	18–25	3·3	3·1	3·1	3·0	3·1	(771)
1948	40–55	3·3	3·3	3·4	3·7	3·4	(859)
1952	21+	3·3		3·4	3·3	3·3	(893)
1955(3)	18–39	3·3	3·2	3·3	3·5	3·3	(2,579)
1955(4)	18–39	3·5	3·4	3·5	3·7	3·5	(2,579)
1957	21+	3·3		3·3	3·5	3·4	(586)
1959	21+	3·5		3·5	3·7	3·6	(625)
1960(3)	18–39	3·3	3·3	3·5	3·3	3·4	(2,378)
1960(4)	18–39	3·4	3·5	3·6	3·5	3·5	(2,378)
1963	21+	3·5	3·4	3·7	3·6	3·5	(638)
1966	21+	3·4	3·4	3·6	3·6	3·4	(550)
		Males					
1936	21+	3·0	2·9	3·3	3·5	3·1	(1,236)
1941	21+	3·2		3·1	3·3	3·2	(1,870)
1945	21+	3·4		3·4	3·7	3·5	(1,221)
1947	21+	3·2		3·2	3·3	3·2	(1,236)
1948	18–25	2·9	2·9	2·9	2·9	2·9	(791)
1948	40–55	3·2	3·3	3·1	3·2	3·2	(854)
1952	21+	3·1		3·2	3·4	3·3	(880)
1955(5)	18–39	3·3	3·2	3·1	3·2	3·1	(1,827)
1955(6)	18–39	3·4	3·3	3·2	3·3	3·2	(1,827)
1957	21+	3·1		3·2	3·5	3·3	(543)
1959	21+	3·2		3·5	3·7	3·5	(588)
1960(6)	18–39	3·3	3·2	3·2	3·0	3·2	(2,191)
1963	21+	3·2	3·4	3·6	3·8	3·4	(595)
1966	21+	3·1	3·3	3·3	3·5	3·2	(528)

NOTES: (1) All the Gallup polls (dated 1936, 1941, 1945, 1947, 1952, 1957, 1959, 1963 and 1966) asked the following question: 'What do you consider is the ideal size of a family – a husband, wife, and how many children?' The Roper Poll of 1943 asked: 'How many children would you like to have, if you had your choice?', and that of 1948: 'How many children do you think makes the nicest size family?' The Growth of American Families Studies of 1955 and 1960 inquired concerning the ideal number of children for 'the *average* American family'. The *minimum* distribution arises from coding range answers (e.g. 'two to three') to the lowest figure, and the *maximum* distribution from coding them to the highest figure.

(2) Levels 1 to 4 range in order from 'high' to 'low'. For the years 1955, 1960, 1963 and 1966 the four levels of income for the husband (or the chief wage-earner) are: (1) $7,000 and over, (2) $5,000 to $6,999, (3) $3,000 to $4,999, and (4) under $3,000. For the remaining years the categories represent qualitative evaluations of the household's 'economic status' by interviewers. In 1943 and 1948 these categories are: (1–4) Prosperous, Upper Middle, Lower Middle, and Poor; in 1952, they are (1–2) Wealthy and Average-plus, (3) Average, and (4) Poor; and for 1957 and 1959 they are (1–2) Upper, (3) Middle, and (4) Lower.

(3) Minimum ideal.

(4) Maximum ideal.

(5) Minimum wanted as stated by wife (results from coding range answers to lowest figure). Question asked of wife: 'How many children does your husband want to have altogether?'

(6) Maximum wanted as stated by wife (results from coding range answers to highest figure). See footnote (5) for question asked of wife.

58

reader can see that there is relatively little variability by economic status. In addition, in so far as there is variation, the relationship is inverse or slightly U-shaped. If, moreover, one separates the Catholics from the non-Catholics (Tables 3 and 4), the ideals among the latter are either virtually identical or inverse among the various economic levels. This is true almost without exception among women, and except for the 1955 Growth of American Families Study in which wives reported on the number of children their husbands wanted, it is true among men as well. On the other hand, when one turns to the Catholics, the relationship of family-size ideals and income is

TABLE 3. *Mean number of children considered ideal by white Catholics, both sexes according to economic status or income, United States, selected years, 1943–1966*

<table>
<tr><th rowspan="2">Date</th><th rowspan="2">Age range</th><th colspan="4">Economic or income levels[1]</th><th rowspan="2">Total</th><th rowspan="2">(N)</th></tr>
<tr><th>1</th><th>2</th><th>3</th><th>4</th></tr>
<tr><td colspan="8">Females</td></tr>
<tr><td>1943</td><td>20–34</td><td>2·9</td><td>3·2</td><td>3·0</td><td>2·7</td><td>3·0</td><td>(510)</td></tr>
<tr><td>1948</td><td>18–25</td><td colspan="2">3·4</td><td>3·6</td><td>3·0</td><td>3·4</td><td>(175)</td></tr>
<tr><td>1948</td><td>40–55</td><td colspan="2">3·7</td><td>3·6</td><td>4·1</td><td>3·8</td><td>(144)</td></tr>
<tr><td>1952</td><td>21+</td><td colspan="2">*</td><td>3·7</td><td>3·5</td><td>3·5</td><td>(201)</td></tr>
<tr><td>1955[2]</td><td>18–39</td><td>3·7</td><td>3·4</td><td>3·5</td><td>3·5</td><td>3·5</td><td>(745)</td></tr>
<tr><td>1955[3]</td><td>18–39</td><td>3·9</td><td>3·6</td><td>3·6</td><td>3·7</td><td>3·7</td><td>(745)</td></tr>
<tr><td>1957</td><td>21+</td><td colspan="2">3·5</td><td>3·5</td><td>3·5</td><td>3·5</td><td>(161)</td></tr>
<tr><td>1959</td><td>21+</td><td colspan="2">3·5</td><td>3·8</td><td>3·8</td><td>3·7</td><td>(162)</td></tr>
<tr><td>1960[2]</td><td>18–39</td><td>3·8</td><td>3·6</td><td>3·8</td><td>3·9</td><td>3·7</td><td>(650)</td></tr>
<tr><td>1960[3]</td><td>18–39</td><td>3·9</td><td>3·8</td><td>3·9</td><td>4·0</td><td>3·9</td><td>(650)</td></tr>
<tr><td>1963</td><td>21+</td><td>4·2</td><td>3·7</td><td>4·2</td><td>3·9</td><td>4·0</td><td>(155)</td></tr>
<tr><td>1966</td><td>21+</td><td>3·7</td><td>3·7</td><td>3·9</td><td>*</td><td>3·7</td><td>(176)</td></tr>
<tr><td colspan="8">Males</td></tr>
<tr><td>1948</td><td>18–25</td><td colspan="2">3·2</td><td>3·2</td><td>3·2</td><td>3·2</td><td>(193)</td></tr>
<tr><td>1948</td><td>40–55</td><td colspan="2">4·0</td><td>3·8</td><td>3·4</td><td>3·7</td><td>(136)</td></tr>
<tr><td>1952</td><td>21+</td><td colspan="2">*</td><td>3·4</td><td>3·4</td><td>3·4</td><td>(194)</td></tr>
<tr><td>1955[4]</td><td>18–39</td><td>3·6</td><td>3·8</td><td>3·3</td><td>3·8</td><td>3·5</td><td>(483)</td></tr>
<tr><td>1955[5]</td><td>18–39</td><td>3·9</td><td>4·0</td><td>3·5</td><td>3·9</td><td>3·7</td><td>(483)</td></tr>
<tr><td>1957</td><td>21+</td><td colspan="2">3·6</td><td>3·0</td><td>3·5</td><td>3·4</td><td>(129)</td></tr>
<tr><td>1959</td><td>21+</td><td colspan="2">3·7</td><td>3·8</td><td>3·4</td><td>3·6</td><td>(151)</td></tr>
<tr><td>1960[5]</td><td>18–39</td><td>4·2</td><td>3·5</td><td>3·8</td><td>3·7</td><td>3·8</td><td>(548)</td></tr>
<tr><td>1963</td><td>21+</td><td>3·5</td><td>3·8</td><td>4·1</td><td>*</td><td>3·8</td><td>(155)</td></tr>
<tr><td>1966</td><td>21+</td><td>3·4</td><td>3·6</td><td>*</td><td>*</td><td>3·5</td><td>(129)</td></tr>
</table>

NOTES: (1) Levels 1 to 4 range in order from 'high' to 'low'. For the years 1955, 1960, 1963 and 1966 the four levels of income for the husband (or the chief wage-earner) are: (1) $7,000 and over, (2) $5,000 to $6,999, (3) $3,000 to $4,999, and (4) under $3000. For the remaining years the categories represent qualitative evaluations of the household's 'economic status' by interviewers. In 1943 and 1948 these categories are: (1–4) Prosperous, Upper Middle, Lower Middle, and Poor; in 1952 they are (1–2) Wealthy and Average-plus, (3) Average, and (4) Poor; and for 1957 and 1959 they are (1–2) Upper, (3) Middle, and (4) Lower.

(2) Minimum ideal (results from coding range answers to lowest figure).

(3) Maximum ideal (results from coding range answers to highest figure).

(4) Minimum wanted as stated by wife.

(5) Maximum wanted as stated by wife.

* Fewer than 25 cases.

59

frequently U-shaped with upper-income Catholics offering ideals higher than those at the next lower level, and sometimes even higher than those at the lowest level.

These results are analysed in detail in the paper mentioned. The principal point of interest here is that the relationship predicted by an *economic* interpretation of fertility – a rise in family-size preferences with rising income – is not actually found unless some powerful pro-natalist, *non*-economic influence, such as Catholicism, is at work.

TABLE 4. *Mean number of children considered ideal by white non-Catholics, both sexes, according to economic status or income, United States, selected years, 1943–1966*

<table>
<tr><th rowspan="2">Date</th><th rowspan="2">Age range</th><th colspan="4">Economic or income levels(1)</th><th rowspan="2">Total</th><th rowspan="2">(N)</th></tr>
<tr><th>1</th><th>2</th><th>3</th><th>4</th></tr>
<tr><td></td><td></td><td colspan="4">Females</td><td></td><td></td></tr>
<tr><td>1943</td><td>20–34</td><td>2·9</td><td>2·7</td><td>2·7</td><td>2·5</td><td>2·7</td><td>(1,869)</td></tr>
<tr><td>1948</td><td>18–25</td><td colspan="2">3·0</td><td>3·0</td><td>3·0</td><td>3·0</td><td>(499)</td></tr>
<tr><td>1948</td><td>40–55</td><td colspan="2">3·2</td><td>3·4</td><td>3·6</td><td>3·4</td><td>(600)</td></tr>
<tr><td>1952</td><td>21+</td><td colspan="2">3·3</td><td>3·3</td><td>3·3</td><td>3·3</td><td>(692)</td></tr>
<tr><td>1955(2)</td><td>18–39</td><td>3·2</td><td>3·1</td><td>3·2</td><td>3·5</td><td>3·3</td><td>(1,834)</td></tr>
<tr><td>1955(3)</td><td>18–39</td><td>3·4</td><td>3·3</td><td>3·4</td><td>3·7</td><td>3·4</td><td>(1,834)</td></tr>
<tr><td>1957</td><td>21+</td><td colspan="2">3·3</td><td>3·2</td><td>3·5</td><td>3·3</td><td>(425)</td></tr>
<tr><td>1959</td><td>21+</td><td colspan="2">3·5</td><td>3·5</td><td>3·6</td><td>3·5</td><td>(463)</td></tr>
<tr><td>1960(2)</td><td>18–39</td><td>3·1</td><td>3·2</td><td>3·3</td><td>3·2</td><td>3·2</td><td>(1,728)</td></tr>
<tr><td>1960(3)</td><td>18–39</td><td>3·2</td><td>3·3</td><td>3·5</td><td>3·4</td><td>3·4</td><td>(1,728)</td></tr>
<tr><td>1963</td><td>21+</td><td>3·3</td><td>3·3</td><td>3·5</td><td>3·5</td><td>3·4</td><td>(483)</td></tr>
<tr><td>1966</td><td>21+</td><td>3·2</td><td>3·2</td><td>3·4</td><td>3·7</td><td>3·3</td><td>(374)</td></tr>
<tr><td></td><td></td><td colspan="4">Males</td><td></td><td></td></tr>
<tr><td>1948</td><td>18–25</td><td colspan="2">2·8</td><td>2·9</td><td>2·8</td><td>2·9</td><td>(445)</td></tr>
<tr><td>1948</td><td>40–55</td><td colspan="2">3·2</td><td>3·1</td><td>3·3</td><td>3·2</td><td>(499)</td></tr>
<tr><td>1952</td><td>21+</td><td colspan="2">3·1</td><td>3·1</td><td>3·3</td><td>3·2</td><td>(686)</td></tr>
<tr><td>1955(4)</td><td>18–39</td><td>3·2</td><td>2·9</td><td>3·0</td><td>3·0</td><td>3·0</td><td>(1,344)</td></tr>
<tr><td>1955(5)</td><td>18–39</td><td>3·3</td><td>3·0</td><td>3·0</td><td>3·1</td><td>3·1</td><td>(1,344)</td></tr>
<tr><td>1957</td><td>21+</td><td colspan="2">3·0</td><td>3·3</td><td>3·5</td><td>3·2</td><td>(414)</td></tr>
<tr><td>1959</td><td>21+</td><td colspan="2">3·1</td><td>3·4</td><td>3·8</td><td>3·4</td><td>(437)</td></tr>
<tr><td>1960(5)</td><td>18–39</td><td>3·0</td><td>3·0</td><td>3·1</td><td>2·9</td><td>3·0</td><td>(1,643)</td></tr>
<tr><td>1963</td><td>21+</td><td>3·1</td><td>3·2</td><td>3·4</td><td>3·7</td><td>3·3</td><td>(440)</td></tr>
<tr><td>1966</td><td>21+</td><td>3·0</td><td>3·2</td><td>3·3</td><td>3·4</td><td>3·1</td><td>(399)</td></tr>
</table>

NOTES: (1) Levels 1 to 4 range in order from 'high' to 'low'. For the years 1955, 1960, 1963 and 1966 the four levels of income for the husband (or for the chief wage-earner) are: (1) $7,000 and over, (2) $5,000 to $6,999, (3) $3,000 to $4,999, and (4) under $3,000. For the remaining years the categories represent qualitative evaluations of the household's 'economic status' by interviewers. In 1943 and 1948 these categories are: (1–4) Prosperous, Upper Middle, Lower Middle, and Poor; in 1952 they are (1–2) Wealthy and Average-plus, (3) Average, and (4) Poor; and for 1957 and 1959 they are (1–2) Upper, (3) Middle, and (4) Lower.

(2) Minimum ideal (results from coding range answers to lowest figure).

(3) Maximum ideal (results from coding range answers to highest figure).

(4) Minimum wanted as stated by wife.

(5) Maximum wanted as stated by wife.

* Fewer than 25 cases.

Findings such as these, together with the doubtful status of Becker's own evidence, lead us to be sceptical about the power of the economic theory of demand for consumer durables to provide demographic insight. Since empirical evidence points the other way, it seems wise to re-examine the relevance of the framework itself for the analysis of reproductive motivation.

A CRITIQUE OF THE ECONOMIC FRAMEWORK

In trying to understand why Becker's expectations concerning income and family-size desires diverge so markedly from the available data, we must bear in mind an overall feature of his reasoning. Rather than simply trying to take economic factors *into account* in explaining family-size preferences, he has chosen to propound a *solely* economic analysis of fertility desires. In doing so he has ignored, or specifically attempted to invalidate, well-known sociological determinants of reproductive motivation. He thus ends up with a framework to explain non-existent facts, while he ignores or attempts to expunge explanations for existing ones.

Becker's neglect of the social context of reproduction is most evident in four features of his analysis: the analogy of children with consumer durables; the concentration on the 'consuming' as against the 'producing' role of parents with respect to children; the misapprehension of child costs; and the failure to analyse the utilities involved in having children.

The Limited Relevance of the Consumer Durables Analogy

Why are children like consumer durables? It is noteworthy that Becker arrives at the analogy by a back route. He points out that in modern societies children are no longer what economists call a 'production good'. The net costs of children are no longer negative, but rather are now positive. This places children in a residual category, 'consumption goods', because, since they are not good for production, it is necessary 'to assume that psychic income or utility is received from them'.[28]

There are numerous reasons, however, for regarding this analogy as implausible and misleading. For example, the assumed equivalence of 'demand' for consumer durables and 'desire' for a family of a particular size is unwarranted. 'Demand' for consumer durables bears a positive relation to income primarily because individuals acquire such goods in the context of direct economic constraint. Their acquisitive behaviour is limited by their credit, not by their choice. A theory of demand is thus not a theory of desires or wants. This point might not invalidate the consumer durables analogy as a model for reproductive analysis, if controls over the acquisition of children were as direct and severe as they are in the case of consumer durables. But, even if all unwanted pregnancies were eliminated, there is no direct control over the acquisition of *wanted* children, as there is over the acquisition of wanted cars, refrigerators and houses. In fact, one

[28] Becker, *loc. cit.*, p. 213. Becker says: 'For most parents, children are a source of psychic income or satisfaction, and, in the economist's terminology, children would be considered a consumption good' (p. 210).

must recognize that the sociology of the family is such that freedom to choose the number of children one wishes is sacrosanct. Not only are individuals under strong institutional pressure to marry and start a family, but the decision to do so, even in the face of financial difficulties, receives widespread moral (and, if necessary, tangible) encouragement. The 'consumption' of a family by individuals who cannot 'afford' one is regarded quite differently from their decision to purchase a consumer durable that they cannot afford. In fact, the right to have a family is widely extended to individuals who are impaired physically and mentally, as well as financially.[29] Consequently, unless we presume a 'means test' for the acquisition of children analogous to a 'credit rating' for the acquisition of consumer durables, the analogy between the demand for consumer durables and voluntary family size is far fetched. Clearly, therefore, a major assumption of the economic theory of demand for consumer durables does not hold for the acquisition of children.

The analogy also implies that offspring are a means or instrumentality for the 'consuming' parents.[30] Leaving aside for the moment a consideration of the types of goals for which children are a means (that is, what it is about children that gives 'psychic income' or satisfaction), let us ask whether they are instrumentalities in the same sense as other consumer durables? For instance, does the consumer of children have flexibility in arriving at an optimum equilibrium position? Can he engage in a dynamic reshuffling of his consumption behaviour so as to maximize his well-being by equalizing the marginal utilities per dollar that he spends on each item of consumption? Such an assumption of freedom to change the items one consumes – an assumption that underlies the economic theory of demand for consumer durables – is sociologically absurd when applied to children. At best, parents can only anticipate, not re-arrange, their equilibrium position with respect to offspring. But anticipation is highly unreliable, because there are many more uncertainties involved in the acquisition of children than in the purchase of ready-made and visible products. If the parents miscalculate and find that the marginal utility they actually derive from an additional child is less than they would have had from an expenditure on something else, they cannot, normally, adjust the situation. Since couples know about the normative irrevocability of becoming parents, this fact must be assumed to enter into their reproductive decisions. If such an assumption is granted, the model of their decision-making process with respect to children is substantially different from that relating to consumer durables.

What about the consumer's 'sovereignty' in choosing the initial quality and type of the item he is purchasing? With respect to consumer durables, he can avail himself of a market and choose

[29] One of the ideological bases of modern social and economic welfare policies is precisely that one does not question the validity of 'family values', one simply tries to maximize everyone's ability to achieve those family goals that exist. Thus, one should not question the 'right' of individuals to have more children than they can afford, nor should economic help to such individuals be made contingent on their demonstrated willingness to stop having children. Economic considerations *are supposed to be* subordinated to familial ones. Persons who have had more children than they can manage financially are referred to euphemistically as 'unfortunate'. In this sense, a modern welfare society takes up economically where the extended kinship system left off. See Kingsley Davis, 'Some demographic aspects of poverty in the United States', in Margaret S. Gordon (ed.), *Poverty in America*, (San Francisco, 1965), pp. 299–319.

[30] For a discussion of motivational aspects of the desire for consumption goods, see James S. Duesenberry, *Income, Saving and the Theory of Consumer Behavior* (Cambridge, Mass., 1959), Chapters II and III, pp. 6–46.

among visible products whose qualities he can, with some rational effort, ascertain. But the acquisition of children of a particular quality or type is, for any individual parent, beyond his control and quite unpredictable. The potential consumer is thus by way of acquiring an item that he can only *hope* will bring whatever utilities he seeks, but which actually may turn out to have biological defects or other qualities that he finds unattractive.

Finally, we may ask whether the consuming parent has relatively free use (or abuse) of this means? Can the means be overworked, allowed to fall into disrepair, become a victim of repeated accidents through the owner's negligence? Obviously not. Parents do not 'own' children, and, as guardians, they are legally required to keep them in minimum repair, not to abuse them physically or mentally, or, through negligence, allow them to be victimized by accidental violence and the like.

When one takes these factors into account, one realizes that children are not merely a means for consuming parents – like cars or refrigerators – but rather that their existence obliges parents to accept many onerous conditions and restrictions. Some of these are biological and some are sociological, but all place children at a far remove from monetarily acquired, readily disposable, and normatively indifferent 'consumer durables'.

In sum, although the demand for consumer durables is pegged to purchasing power, the 'demand' for children is not under such monetary control. In fact, by creating public support for the dominance of family 'values' over economic rationality, reproductive and social institutions are geared to *prevent* economic factors from inhibiting reproduction. When one takes into account as well that the *desire* for children will be influenced, among other things, by the social and biological constraints surrounding their acquisition and their 'use' – constraints that may be independent of income or may vary positively with it – one has little reason to believe that the demand for consumer durables constitutes a theoretically apt model for family-size preferences.

Parents as Producers of Children

Further insight into reproductive motivation comes from recognizing that parents are 'producers' as well as 'consumers' of children. Their desire for offspring in any particular quantity will, therefore, presumably be influenced by the production problems involved in having a family as well as by the utilities they expect to gain. If this is true, the theory of demand for consumer durables leaves out of account important influences on reproductive decisions.

To be sure, Becker recognizes the productive role of parents, but he treats it entirely as part of the costs parents pay for the utility they expect to gain. We shall discuss this type of cost later. There are, however, elements in the productive process – particularly relating to the structure of the productive unit and its articulation with the society at large – that cannot fruitfully be subsumed under the category of costs.

For example, in making decisions concerning family size and quality, parents must reckon that the interaction of children with one another is an element in producing a successful product – the socialized child. Of course, this interaction can be achieved in many different ways; it is not necessarily true that siblings are required for the purpose. But, the isolation of the nuclear family and its geographical mobility in modern society are such that parents find the substitution of non-siblings for siblings to be difficult. Hence, in facing up to the problem of providing adequate socialization, parents are typically highly motivated to avoid the 'only child' type of unit.[31]

Further constraint in the production of offspring comes from the societal surveillance over 'quality control'. Becker discusses the production of children of a particular 'quality' entirely in terms of an individual consumer's decision concerning the price of the child from whom *he* derives 'utility'. However, not only are 'tastes' in children more subject to social influence than Becker is willing to grant, but there are normative prescriptions to produce some minimum level of quality. For example, the child's behaviour must be within the law and not contrary to public policy. Further, in modern societies parents must comply with educational requirements up to virtually the end of childhood. In addition, as producers of children, parents are under some social pressure to recognize and validate a responsibility for the quality of their product throughout its life. The child can never go on the scrap heap without in some measure disgracing them. One must thus bear in mind that parental expenditure on children (or current savings for them) will often relate to a period when the parents may be dead. The present utility that they derive from such future-oriented provision for children inheres in the satisfaction gained through fulfilling kinship obligations.[32]

Finally, since a couple produce the children they 'consume', they must somehow adjust their desires regarding the family to the cyclical nature of family life. If the couple find utility in children, for how long is this utility to be made available? As the producers, they have to resolve the problem of how long the unit – a family with children – stays 'in business', so that, as consumers, the couple can enjoy the product. This point relates to the well-known problem of the 'empty nest', and to decisions concerning the number and spacing of children throughout the parents' lives.

In sum, the productive role of parents will have an influence on family-size desires. The nature of this influence is such that it seems unlikely to produce a positive relation between them and income under any and all conditions. In fact, one could claim that 'production' features introduce elements of *in*elasticity into the desire for children in relation to income. Poor parents as well as rich ones will view the only-child unit as a deprived one. Hence, two children, and not one, become the minimum for the avoidance of childlessness. Poor parents may also be concerned about the

[31] The Indianapolis Study found that a desire to avoid the 'only child' child-rearing situation (because of a belief that the only child is handicapped) was a major reason for couples having a second child. See Erwin S. Solomon, Jeanne E. Clare and Charles F. Westoff, 'Fear of childlessness, desire to avoid an only child, and children's desires for siblings', *Milbank Memorial Fund Quarterly*, **34** (April 1956), pp. 160–171.

[32] A considerable literature – written mostly by economists and historians – exists on the topic of inheritance and family size. By and large, this literature takes as 'given' the obligation experienced by parents to provide for the futures of their offspring and the 'problem' is seen as relating to the quantity of resources available, to their depreciation over time, to legal restraints, etc. One must point out, however, that of equal relevance is the set of controls that induces parents to take their children's adult futures into account at all. From a societal point of view, the parent's role as consumer is only instrumental to his role as producer and guardian. For a discussion of social control over parents' influence on age at marriage, see Judith Blake, 'Parental control, delayed marriage, and population policy', in *Proceedings*, United Nations, World Population Conference, 1965, Vol. II, pp. 132–136. Recognition of the changing problems experienced by parents in making provision for their children is given explicitly by J. A. Banks, *Prosperity and Parenthood* (London, 1954).

timing of the 'empty nest' – perhaps, as we shall see, they may be more concerned than rich parents. On the other hand, the social forces demanding quality control in children may impinge more heavily and immediately on the affluent than on the poor. The rich are, therefore, likely on these grounds alone to inhibit their family-size desires.

Misapprehension of the Costs of Children

Let us now turn to the costs of children, which Becker believes encourage a positive relation of family-size desires and income. He reaches this conclusion by ignoring *indirect* costs – alternative utilities on which parents could expend their resources, and by concentrating on *direct* costs – the resources actually expended on childbearing and rearing. Even with reference to the latter, however, he does not recognize, or denies the importance of, factors making direct costs heavier for the rich.

Direct Costs of Children. In determining the quantity one wishes to acquire of any item, at least two aspects of the direct cost would seem to be crucial – the timing of the necessary financial expenditure and the amount of such expenditure. In the case of children we have already pointed out that there is no formal purchase price or down payment. Little in the economic context of acquiring a child conduces to financial 'reality-testing' concerning its total potential costs. Since children are 'purchased' on the instalment plan, if the 'payments' are kept low, a major constraint on acquiring more is removed; if the payments are kept high, on the other hand, the impact of the costs involved may readily lead to a desire not to incur many more of a similar nature.

Are child costs likely to be relatively high for the poor and low for the affluent? Becker appears to believe that this is the case. He denies explicitly that there is any significant sociological determination of the 'quality' of child that people at different income levels will 'consume'. In particular he rejects the idea that wealthier parents are under significant social pressure to have 'expensive' children. He thus sees no 'cost effect' invalidating the expectation of a positive relation of family-size desires and income. He here overlooks two important points. First, parents find it difficult to separate significantly the level of living of their children from their own, since, after all, they normally all reside at the same address, eat the same types of food, etc. This criticism has already been made by Duesenberry.[33] Second, the way of life at a given social level puts its mark on *standards* of child quality as well. The fact is that the way of life of the poor leads them to accept low standards of child quality – standards that do not greatly transcend the actualities of the moment, whereas the pressures on the more well-to-do contribute to standards in children that will both conserve present advantages and secure added ones for the child.

Why do the poor not choose to have very few higher quality children, rather than more lower priced ones? The answer is, in part, that poorer people are not actively dissatisfied with low-priced children because they cannot transcend their own limitations to that extent. Low-quality children fit in with the way of life of the poor and, in an atmosphere of general scarcity and limitation, parents are not goaded into dissatisfaction with such children to the extent of making changes

[33] See Duesenberry's discussion of Becker's article, *op. cit.*, pp. 231–234.

in their own lives and objectives to rectify the situation. Since poorer children only rarely come into direct contact and competition with wealthier ones, poor parents are shielded from comprehending the overall effects of low price on comparative child performance. Thus, the poor parents' lack of perspective and knowledge concerning 'what it takes' to rear children effectively – a limitation that the situation of poverty both generates and leaves unshaken – conduces to a false sense of security about making low payments on children.[34]

By contrast, as one goes up the social scale, the standards of quality in children (like such standards with regard to other things) become more and more demanding. At issue here is not simply a predilection for luxury in children. Rather, at these levels, child-rearing is an exclusionary device relative to classes below.[35] Richer parents are thus motivated to invest in higher-quality children because, as producers of children, they feel obliged to give their issue the competitive advantages of their class.

Concern with child quality is intensified among the more affluent by the social mobility – both upward and downward – that is a persistent feature of life in middle and upper income strata. In this atmosphere of opportunity to rise and fall, standards of quality in children may become very inflated, causing parents to over-extend their resources in order to have offspring with the requisite qualifications. It is thus not uncommon for families in the wealthier strata of modern societies to feel subjectively under great economic pressure despite their objectively prosperous condition.

Finally, class-oriented standards of child-rearing do not relate merely to cash expenditure alone. As is well known, the more advantaged groups in the United States have been in the vanguard of parent-intensive child-rearing. Such large doses of personalized care – of non-cash parental

[34] A survey taken in 1959 by Roper questioned parents (with one or more children under 18 years of age but not in college) concerning their college intentions for their children and the financial preparations they were making to help the children. Among upper-economic level parents 97% expected their children to attend college, among the lowest level 44% had this expectation. Even upper and middle economic classes underestimated potential college costs by discounting the possibility of inflation by the time the children would be eligible for attendance. The lowest economic group was typically unable to make any such cost estimate at all. Questions concerning realistic financial preparations by parents revealed a minimal provision, if any, among even higher economic groups. The typical answer of the parents in the lowest economic group was that they 'hadn't had a chance to think about it yet.' Almost a fifth of lower economic level parents intended to pass the problem on to the children themselves, whereas this was true for only 3% of parents in the highest economic group. Elmo Roper, 'College ambitions and parental planning', *Public Opinion Quarterly*, **25**, 2 (Summer, 1961), pp. 159–166.

[35] Characteristics that are acquired by children within the family setting and during a long period of socialization such as accent, 'manners', and social facility and poise, are especially difficult for outsiders to acquire later in life. The literature on English class differences in speech is particularly graphic. See T. H. Pear, *English Social Differences* (London, 1955), Chapter 3; Nancy Mitford, 'The English aristocracy', *Encounter* (September 1955), pp. 5–12; and Alan S. C. Ross (ed.), *Noblesse Oblige* (London, 1956), *passim*. It is perhaps necessary to point out that we are not concerned here with the advantages to children in acquiring upper-class characteristics as these advantages might appear to an objective, and perhaps more knowledgeable, observer. For example, being reared as an aristocrat had certain drawbacks for a young Russian adult in 1917, just as attending an English public school to-day may result in certain trained incapacities for coping with life in the twentieth (and not the nineteenth) century. For a discussion of the persistence of the nineteenth-century gentleman as a goal of 'character building' in English public schools, see Ian Weinberg, *The English Public Schools* (New York, 1967), *passim*. We are merely concerned with the costs to parents of helping their children acquire expensive characteristics, regardless of the wisdom of having them do so.

67

inputs per child – have not been found among working-class parents.[36] Since the time and effort of even wealthy parents is limited, so is their propensity to have children.[37]

Indirect Costs of Children. Becker deals with the marginal utility to be gained out of some balance between quantity and quality in children, but does not consider that alternative utilities enter into family-size decisions. Such indirect, or opportunity, costs are of particular importance in societies having a wide range of consumption opportunities and many organizational axes in addition to kinship. One of the principal reasons for the decline in family size with rising income in the history of the West may well have been an expansion of competing items of expenditure in addition to a rise in direct costs.[38]

Even in the United States, where gross differences in consumption may no longer be widely evident, the way of life of upper-income groups is more competitive with children for time, effort and finances than is the life-style of those in lower-income brackets.[39] The former tend to be more active in political, civic and community affairs.[40] They also tend to be more wholly committed to

[36] For example, Daniel R. Miller and Guy E. Swanson, *The Changing American Parent* (New York, 1958); Murray A. Straus, 'Deferred gratification, social class, and the achievement syndrome', *American Sociological Review*, **27** (June 1962), pp. 326–335; and Glen H. Elder, Jr., and Charles E. Bowerman, 'Family structure and child-rearing patterns: The effect of family size and sex composition', *American Sociological Review*, **28** (December 1963), pp. 891–905.

[37] Duesenberry has noted that the problem of non-cash parental costs is particularly troublesome to parents. As he says, '. . . the marginal disutility of Cub Scouts and PTA meetings rises rapidly', *op. cit.*, p. 234. We might add that if well-to-do parents try to find surcease from the demands of child-rearing through the services of governesses, maids, boarding schools and the like, they soon begin to experience diminishing marginal utility in their parental roles. Assuming that a primary motive for having children is to 'enjoy them', to purchase more than can be attended personally and then turn over their 'consumption' to a third party is like hiring someone else to drive a car that one is oneself too busy to use. Only if one assumes that wealthier parents have motives for large families that go beyond the 'enjoyment-of-consumer-durables' syndrome can one make a strong case for their wanting more than a relatively few children.

[38] See Leibenstein, *op. cit.*, Chapter 10, and Banks, *op. cit.*, pp. 48–102. Banks attempts specifically to document the thesis of a widespread increase in the nineteenth century of the opportunity costs of children.

[39] Kurt Mayer has claimed that the decline in income differences in the United States has resulted in a great similarity of gross consumption patterns. 'By and large everybody in America wants to buy the same things everyone else buys. Americans exhibit a remarkable homogeneity of tastes, attitudes and buying habits, regardless of occupation' (Kurt Mayer, 'Diminishing class differentials in the United States', *Kyklos: International Review for Social Sciences*, **12** (1959), pp. 605–628). For a recent discussion by Riesman of the 'standard package', see 'Careers and consumer behaviour', in Norman W. Bell and Ezra F. Vogel, *A Modern Introduction to the Family* (Glencoe, 1960) pp. 143–162. Some of the European literature on this subject is cited by Richard F. Hamilton, 'Affluence and the worker: The West German case', *American Journal of Sociology*, **71** (September 1965), pp. 144–152. In general, the mass consumption thesis has a mass communications-social-psychological bias. It assumes that wants and goals are simply 'learned' or internalized either in the primary group or by means of mass persuasion. Thus, people are said to 'adopt' a middle-class style of life – as if the latter had some static meaning and as if the motives for 'adopting' were entirely unrelated to some dynamic calculus of interests, goals and pressures experienced by the individual and largely determined by his social roles and statuses. In this sense, the mass society approach attempts to turn Marx on his head with no theoretically valuable outcome whatever.

[40] For information concerning the voluntary association memberships of Americans by economic level and income, see Mirra Komarovsky, 'The voluntary associations of urban dwellers', *American Sociological Review* **11** (December 1946), pp. 686–698; Floyd Dotson, 'Patterns of voluntary association among urban working-class families', *American Sociological Review*, **16** (October 1961), pp. 687–693; Morris Axelrod. 'Urban structure and social participation', *American Sociological Review*, **21** (February 1956), pp. 13–18; Charles R. Wright and Herbert H. Hyman, 'Voluntary association memberships of American adults: Evidence from national sample surveys', *American Sociological Review*, **23** (June 1958), pp. 284–294; and Erich Goode, 'Social class and church participation', *American Journal of Sociology*, **72** (July 1966), pp. 102–111. In the Hyman–Wright study the proportions belonging to no associations drop precipitously at the 'above average' and 'very high' living levels, and the proportions belonging to two or more rise precipitously. For example, among five categories of living levels, the lowest three have 1, 5 and 17% belonging to two or more associations; but the highest two have 32 and 64% holding this many memberships.

the demands of work and of 'running things' in general.[41] They are stimulated and harassed by the social mobility that we have already discussed in relation to direct child costs.[42] And, they have more attractive and diversified consumption opportunities than have those of lesser income. An upper-income person is normally under some social pressure to take advantage of these opportunities. Unless he wishes to lead the life of an eccentric, he does not typically have the choice of consuming like an upper *or* a lower class person.

To summarize, not only must the affluent typically entertain higher standards of child quality than the poor, but affluence introduces opportunity costs into people's lives that do not exist where the range of choice is narrowed by poverty. Hence, even if wealthier couples feel that they can afford the *direct* costs of large families, they may not wish to sustain the *indirect* ones unless they are motivated by some powerful *non*-economic force such as, for example, Catholicism.

The Utilities of Children

From whence comes a sense of utility in children? By leaving this question untouched, Becker is free to make two assumptions necessary to his thesis of quantity income elasticity. The first is that there is no family-size threshold below which even poor couples will strongly resist falling, and the second is that there are no systematic social class differences in the relative utilities of children (and, hence, in 'taste' for children) which limit the family-size desires of the well-to-do. Neither of these assumptions is correct.

Children are sometimes said to be playthings, emotional objects and the like. But in view of the costs involved in their rearing and the restrictions on their 'use', an explanation of their desirability can be found only in terms of goals to which children are intrinsically related – goals that is, which can be achieved only through children. The creation and maintenance of such goals is a principal function of familial and kinship institutions in all societies. By exercising control over every step in the reproductive process, but principally by a ruthless exclusion of structured alternatives to and substitutes for family statuses, family satisfactions and kinship affiliations – alternatives that extend from prostitution and homosexuality, on the one hand, to celibacy and

[41] This situation contrasts greatly with that of static, *élite* societies in which a condition of 'non-effort' is the supreme achievement. See Kingsley Davis, 'The role of class mobility in economic development', *Population Review*, **6** (July 1962). Relevant here are data on lower-class as against upper-middle and upper-class marriages which indicate that the involvement of the *couple* in the man's work world varies greatly with level of living. In *Blue Collar Marriage* a dominant theme is the limitation of interaction between husband and wife extending not only to different patterns of leisure but to the participation of the wife in her husband's work world. Mirra Komarovsky, *Blue Collar Marriage* (New York, 1964), pp. 154–156, 311–312, and Chapter 14 (pp. 311–329) which deals extensively with limitations on blue collar social life. A contrast with this picture may be found in the families of corporation executives. See William H. Whyte, Jr., 'The wives of management' and 'The corporation and the wife', *Fortune*, **44** (October 1951), pp. 86–88, 204, 206–208, 210, 211, and **44** (November 1951), pp. 109–111, 150, 152, 155–156 and 158 respectively.

[42] Kurt Mayer grants that: '. . . there is some reason to believe that the emphasis on subtle status differences is now heightening as the leveling process increasingly blurs the income differentials between manual and white collar groups' (Mayer, *op. cit.*, p. 623). C. Wright Mills cites the 'status panic' of white collar groups in the face of rapidly diminishing indicators of separateness from blue collar workers. See C. Wright Mills, *White Collar* (Oxford, 1951), pp. 72–73. An interesting discussion of the differences between the Mayer and the Mills approach may be found in Richard Hamilton, 'The income differences between skilled and white collar workers', *British Journal of Sociology*, **14** (December 1963), pp. 363–373. The Mills thesis has been taken up popularly by Vance Packard in *The Status Seekers* (New York, 1959).

careers for women on the other – societies channel motivation in the direction of goals *that imply the advent and existence of children.* One can become a 'parent', 'have a family', be a 'mother' or 'father', only by acquiring children. That one should desire these statuses is the final result of complex institutional control, but, *given this desire,* children and only children can satisfy it. It is the societal support for the family that provides the strong desire for children and that makes it highly unlikely that poorer people will be willing either to remain childless, or to curtail their family size to the extent required for producing a direct relation of family size and income.[43]

By the same reasoning, one may expect some variability in the relative impact of familial goals on motivation, and hence some structured differences in 'taste' for children. In fact, in complex societies having highly developed institutions that compete with the family, such differences in taste may be expected to run counter to the Becker thesis. As we have noted, the upper classes are under greater pressure from non-familial demands than the lower and hence may find utility quite readily in small-to-moderate size families and disutility in larger ones.

Therefore, since familial institutions motivate almost everyone to have some children and non-familial institutions create a sense of disutility in large families among the affluent in particular, one has reason to doubt the Becker thesis on these grounds as well as on others.

CONCLUSION: THE FALSE TRAIL OF QUANTITY INCOME ELASTICITY

Clearly there are cogent reasons for concluding that the consumer durables model is inapplicable to children and hence cannot predict fertility differentials by income. The acquisition of consumer durables is externally limited by credit. The poor are prevented from over-extending themselves very far by the need to give evidence of ability to meet the purchase price. With respect to children, on the other hand, there is no purchase price. They are home-produced, and all strata have a right to produce them and to receive charity, if necessary, after they have produced them.

Looking at reproductive motivation rather than demand, we have found that the poor seem to share in certain society-wide *pro*-natalist motivational pressures, but do not share in many of the *anti*-natalist ones affecting the middle and upper income groups in modern societies. On the pro-natalist side, the 'utilities' involved in having children are built into the institutional structure of reproduction. These utilities, such as being a 'parent' and living in a family setting, are part of the generally unexamined assumptions of human societies. Little in the current social situation would lead poorer people to question the desirability of these utilities, and in fact they may find such goals to be relatively more satisfying than the rich when taken in the total context of all utilities available. Furthermore, since parents are producers of children – poor parents as well as rich ones – they are motivated by the demands of the structure of the family and its continuity, as well as by the prospective utilities. The poor may be expected to share with the rich an antipathy

[43] The widespread antipathy to childlessness is shown by the fact that this condition is considered preferable by a maximum of 1% of all female respondents in most surveys on family size preferences. See Judith Blake, 'Ideal family size among white Americans: A quarter of a century's evidence', *Demography*, 3, 1 (1966), pp. 154–173. This is true for respondents at all economic levels. See, by the same author, 'Income and reproductive motivation', *loc. cit.*

69

to the 'only child' type of unit, both because of the child-rearing limitations it entails and because it limits the period of one's life during which children can be enjoyed. Again, if poorer people experience proportionately more of their total utility in the family than do the affluent, they will have considerable resistance to accelerating the 'empty nest' period of their lives.

With respect to anti-natalist pressures, the quality requirements for children can be ignored or fended off more readily by lower- than by upper-income groups. Since poor children live with poor parents, such children share in the low overhead that this entails. In addition, the instalment nature of child-payments allows for self-delusion concerning child costs. The way of life of poverty does not impel parents to realize higher standards by restricting quantity. After all, the poor are not threatened by chasms of downward mobility opening up for children. In addition, the off-spring of the poor do not directly interact and compete with those of the rich. If one asks why poorer people overlook the 'cues' available to them in advantaged children, the answer is that selective inattention on their part protects them from unbearable awareness. If the well-to-do must struggle to rear three or four medium- to high-quality children, then the poor 'obviously' should have none at all. Were poor people economically rational and informed in their reproductive preferences, 'the rich would get richer, and the poor wouldn't even get children'.

Because of the simultaneously dynamic and systematic nature of the world in which middle- and lower-income groups live, they are surrounded both by pressures to watch over child quality and facilities for doing so effectively. Part of this quality entails personal effort (non-cash inputs) by parents who, in the nature of the case, are very limited in how much effective interaction, supervision and attention they can provide. But over and above these and other direct costs are the indirect ones that increase concomitantly with income. Affluent parents experience many alternative demands on their resources and many attractions that compete with children. Hence, they are unlikely to desire really large families. Beyond two children, their desires may be influenced by the wish to lengthen the family cycle, by sex preferences, by the woman's feeling that she wants to 'make something' out of the career of motherhood, and similar considerations.[44] But, beyond four children, such considerations are clearly replaced by others, and among a good share of couples the tolerance does not extend beyond three. Only if they are under strong ideological pressure (combined with organizational enforcement) to devote themselves to reproduction and to overlook and denigrate the costs involved, as is the case with Catholics, are such upper-income parents willing to think in terms of larger families than four.

It thus seems true that a theory of reproductive motivation is at the same time a theory of the family and society. Becker's framework fails to explain the direction of family-size ideals and income, because it fails to take into account important elements in the sociology of reproduction.

The results of our analysis have, in addition, long-run implications for population policy. By exposing the simplicity of the economic assumptions, they demonstrate that we cannot rely on want and poverty to provide the motives for fertility decrease, even if contraception were

[44] For example, the effects on family size of various combinations of sex preference in a family and the 'stopping rules' involved are discussed by Nathan Keyfitz in *Introduction to Mathematical Demography* (in press), Chapter 17.

'available' to everyone. Pro-natalist motives have helped societies survive thousands of years of want. The institutional context responsible for such motives is geared to combat the anti-natalist effect of poverty with desires that relentlessly override perceptions of current realities and demand the production of children in spite of everything. Regardless of child quality, or the toll it takes of individuals, this institutional complex concentrates on insuring the biological survival of the species. When we examine the anti-natalist pressures on the well-to-do, we find clues to potentially more widespread motives for fertility-control in the diminished utilities of family involvement, in the high direct costs of children (enforced by the prospect of virtually instantaneous downward mobility if the parents refuse to pay them), and in the competitors with the family and children for time, effort, finances and emotional involvement. But even among the higher income groups, family-size ideals are large enough to insure substantial population growth. We must recognize, therefore, that as yet we have no control over the social context of reproduction comparable to the control over consumer durables provided by the credit system.

72

Population and Development Review
Vol 2, No. 3 and 4 Sep/Dec. 1976

On Two Schools of the Economics of Fertility

WARREN C. SANDERSON

Since the beginning of the last decade, American economists interested in the determinants of fertility have been divided into three rival camps. Richard Easterlin and Gary Becker headed the two major factions, while those who agreed with neither side formed the third group. The following discussion reports on the current status of the conflict between the two main parties. The task I undertake is analogous to that of a war correspondent wiring from the front the news of how the battle is progressing.

The conflict between the groups, although never very heated, has run quite deep. The research produced by members of the contending factions is clearly separable. No ideas suggested by members of one group have been seriously entertained in the literature by members of the other. (To appreciate the utter lack of serious communication between the two contingents, one need only investigate the references to the work of Easterlin in the conference volume edited by Schultz,[1] a supporter of the Becker camp, and the references to the works of the Becker group in the papers of Easterlin on economic-demographic interactions.[2]) Indeed, the conflict has spread beyond the confines of fertility analysis with attacks recently being made by members of the Easterlin school on the economic foundations of the work of the Becker school.[3] To the non-specialist, the schism between the two schools of thought must certainly appear to be at least as wide now as it was a decade ago.

Within the past few years, however, the differences between the two groups of economists have considerably narrowed and on a few of the most important issues of substance there is already a rough consensus. This is not to say that there are no remaining areas of disagreement; it is only to say that the remaining divisive issues are mostly those of style rather than substance. A scholar who would begin work today on the economic determinants of fertility would find that there is a large common ground on which he could work in relative safety from snipers from either side.

The Origins of the Conflict: 1960–1966

In order to understand the extent of the current implicit consensus between the two rival parties, it is necessary to understand how the schism originated. In 1960, Becker published an article in which he claimed that variations in completed fertility could be understood within the same framework economists used for the analysis of the demand for durable goods. His argument rested on two traditional economic postulates:[4] (1) that the representative household behaves rationally on the basis of unchanging tastes, and (2) that the prices of commodities desired by the representative household are unaffected by that household's consumption decisions.[5]

In 1966, Easterlin challenged the early Becker formulation by showing that the movements in the age-specific fertility rates of young women over time were positively related to the movements in an index of "intergenerational relative income," or the ratio of the current income of young married couples to the income level they had experienced as adolescents in their parents' household.[6] The explanation Easterlin gave for this covariance explicitly rejected the notion, still held as sacrosanct by many economists, that tastes ought always to be treated as immutable and replaced it with a mechanism through which tastes change systematically according to one's upbringing. This introduction of the concept of socialization into economic model-building represented a radical departure from standard economic theory and still is not accepted by many economists who study fertility. In contrast with his novel treatment of tastes, Easterlin accepted the traditional assumption that the prices of commodities desired by the representative household are unaffected by the household's consumption decisions. Thus, Easterlin rejected one of the two basic assumptions upon which Becker's early analysis was based.

With the publication of the Becker article in 1960 and the Easterlin article in 1966, the distinction between explanations of fertility behavior based on pure economic theory and those based on the alloy of economic theory with sociology was clear-cut. The schism between the two groups of economists was based in part on the substantive details of their explanations and, unfortunately, in part on the issue of purity, as well.

The Situation at the End of 1973

In the period from 1966 to 1972, the division between the two camps of scholars was the deepest. It was not until 1973, with the publication of two articles from the Becker group, one by Becker and Lewis[7] and one by Willis,[8] that the gap between the parties closed appreciably. In essence, the Becker school of thought moved considerably closer to the views held by Easterlin and his followers.

When Becker and Lewis successfully analyzed the model of fertility suggested by Becker in a footnote in 1960,[9] they found that it also differed significantly from the traditional model of demand. Whereas Easterlin had modified the assumption of stable tastes and maintained the assumption of constant prices for desired commodities, the Becker model did just the reverse. It accepted the assumption of stable tastes, but maintained that the relative price of children and the relative price of goods consumed per child are not independent of household decisions.[10] Thus, by 1973, the two factions took exactly parallel stances with regard to the early Becker formulation. In brief, the Easterlin contingent discarded postulate 1 and accepted postulate 2, while the Becker contingent accepted postulate 1 and discarded postulate 2. The simple early Becker argument that the *observed* relation between income and fertility ought to be positive is now no longer maintained by either group and not even by Becker himself.[11]

The dependence of prices on household decisions in the Becker model is the result of a specification that may be appropriate for the analysis of fertility decisions, but whose applicability to the demand for consumer durables is dubious. The specification in question is the assumption that the family cares about its average level of expenditures per child, but not about its expenditures on each child taken separately.[12] Parents are assumed by the Becker group to desire three things: commodities for their own consumption, children, and commodities for their children's consumption. As the parents' incomes increase, they are assumed to want to spend more both on themselves and on each of their children. It is this positive relation between desired expenditures per child and parental income that causes children to be more expensive for wealthier parents than for poorer ones.[13]

If we consider houses instead of children, the unusual nature of this specification becomes clear. Suppose a couple who owned a $50,000 house decided to buy a $20,000 summer cottage in the mountains. The Becker model would tell us that the summer cottage would be a source of utility, but the fact that the couple spent less on their summer cottage than on their permanent residence would be a source of disutility because their average expenditure per house would have declined. But this is clearly unrealistic. People feel no qualms about buying a summer cottage that costs less than their main residence, nor about buying a second car that costs less than their first. The model when applied to children, however, seems more appropriate. It is plausible that parents could suffer some disutility from not being able to spend as much money on, say, a third child as the average amount they spent on the first two. The assumption that this is the case, however, is not derived from pure economic theory. Rather, it is, like Easterlin's assumption of changing tastes, a special assumption that is necessary to create a plausible model of fertility behavior.

As of 1973, not only did the models of the rival factions take symmetric positions with regard to Becker's early work, but also, the major operative forces in their models were remarkably similar. In both groups of models, the observed relationship between fertility and income may be either positive or negative. These ambiguous results arise from essentially the same circumstances in each model. Both schools assume that, keeping enough things constant, the underlying relation between fertility and income is positive, but they proceed to show that, when income changes, something else is likely to change that has an offsetting effect on fertility. The major source of disagreement between the two scholarly camps at that time was the nature, not the existence, of this offsetting force.

The Easterlin group held that the force which offsets the underlying positive income effect is related to parents' aspirations for their own material standard of living. Over time, both current income levels and aspiration levels rise, leaving the net effect of these two forces on fertility unclear. The Becker group held that the force which offsets the underlying positive income effect is related to parents' aspirations for their children's material standard of living. As parents' incomes rise, they want to increase their average expenditures per child, thus increasing the cost to them of an appropriately raised child. The increasing cost of children raised with higher standards of living would offset the effects of higher income and render the observed relation between fertility and income again unclear.[14] By 1973, the specification of the force that offsets the underlying positive income effect was the main territory contested by the two rival parties.

The Situation at the End of 1976

In 1976, members of both schools published important articles. Although the major question of the nature of the offsetting force was not resolved, a further narrowing of the differences between the rival groups was apparent. Easterlin explained the decline in the fertility of rural women in the United States in the nineteenth century by use of a model in which farm families desired to leave as a bequest to each child as much wealth in real terms as their own parents gave them.[15] The addition of a level of desired bequests per child to the Easterlin camp's formulation parallels the Becker camp's concept of a desired level of expenditures per child. The two specifications differ now only in that, *holding other things constant,* the Becker group expects the desired level of expenditures per child to be positively related to parental income, while the Easterlin group expects desired bequests (and expenditures) per child to be independent of parental income. It is likely that this point will be resolved soon by empirical tests.

In 1976, Becker and Tomes[16] added a new concept to the Becker school model, which brought it closer to Easterlin's formulation. This addition is based, in part, on a distinction between the sources of influence on what Becker has called "child quality." "Child quality" in the Becker-Tomes model depends on the level of expenditures per child and on a host of other influences over which the parents have little or no direct control. The contribution of the Becker-Tomes paper is in its explicit attention to the latter set of influences, which they call "child endowment." One advantage of this extension is that the Becker-Tomes model can be used to analyze fertility in an intergenerational context, a task that had previously only been addressed by members of the Easterlin camp. Indeed, Becker and Tomes do provide some intergenerational fertility analysis. In the context of a discussion of this aspect of their work, the authors make a remarkable statement that also constitutes the first reference to the work of Easterlin in a paper written by Becker. It reads:

> Our conclusions about the effects of economic growth on the number of children are similar to those reached by Richard Easterlin in his important work on fertility. . . . Both Easterlin's and our own analysis are based on changes in the economic position of children relative to their parents.[17]

Thus, yet another difference between the works of the two schools has been substantially reduced.

The two papers published in 1976 demonstrate clearly the emerging consensus. This is not to say that the two groups do not vigorously attack each other over the remaining differences. What is important is that the scope of these differences is now considerably narrower than it was before 1973 and that the area of agreement between the two sides now exceeds the area of disagreement.

The Future of the Conflict

75

The major difference between the two groups is now, as it was in 1973, the nature of the force that offsets the presumed positive income effect. But even this conflict between competing specifications of the offsetting force is unnecessary. It is possible to entertain both the hypothesis that parents' aspirations for their own standard of living and bequests for their children depend, in part, on their background and the hypothesis that parents' aspirations for their children's standard of living as well as their own depend, in part, on their current income. Indeed, other scholars who wrote before the rift had no difficulty in seeing that the two notions were closely interconnected. In 1893, John S. Billings, one of America's foremost nineteenth century demographers, wrote:

> [One] cause [of the fertility decline] is the great increase in the use of things which were formerly considered as luxuries, but which now have become almost necessities. The greater temptations to expenditure for the purpose of securing or maintaining social position, and the correspondingly greater cost of family life . . . lead to the desire to have fewer children in order that they may be each better provided for.[18]

To Billings it was natural to join the ideas subsequently developed separately by the Easterlin school and the Becker school in a single paragraph.

In 1954 J. A. Banks, in *Prosperity and Parenthood,* went even further, summarizing the basic ideas of the two schools of thought in a single sentence. He wrote, ". . . it cannot be denied that the attitude toward the *material comforts of modern existence and the growing expensiveness* of children and adolescents contributed their share to the acceleration in the fall of family size."[19] Although the Easterlin school concentrates on the "attitude toward the material comforts of modern existence" and the Becker school concentrates on "the growing expensiveness of children and adolescents," there is no particular reason why the two notions should not be considered as portions of a single explanation. Indeed, historically the types of arguments put forward by members of the Easterlin school and the Becker school were often considered together as components of a full explanation of fertility variations. Their division into competing hypotheses only occurred in the 1960s.

It is now fairly clear, I hope, that the basic ideas of the Easterlin school and the Becker school are similar and complementary. There is no inherent conflict between them. Economists interested in fertility have, for the most part, followed Easterlin's lead in discarding the simple economic model of fertility put forward by Becker in 1960. Out of the ashes of the old model have grown two more realistic models, one of which stresses the effect of taste formation on secular fertility trends, the other of which stresses the effect of differences in the cost of children on cross-sectional fertility differentials.

The Easterlin framework and the Becker framework can easily be merged into a single model that would formalize the words of Billings, Banks, and others. Indeed, I have already done so.[20] Today, a student of economic demography has at his disposal a number of closely related economic models. The choice between them should depend on the details of the particular study. Choosing between the Easterlin and Becker frameworks on any general grounds, however, is almost guaranteed to be counterproductive. It would be like removing one blade from a scissors or deleting either the concepts of demand or supply from an economist's tool kit. The initial Becker model was an attempt to break with the past and to emphasize the unique contributions of traditional economic theory. In more recent work, the situation has been reversed. The discipline of attempting to use economic theory to analyze fertility decisions has led to new developments in economic models that have made them more consistent with the views held by demographers and sociologists. This same discipline has also brought together economists whose views were initially quite disparate.

Thus, the news that I can report from the front lines is good news. The worst part of the battle is over. The remainder of the decade, I conjecture, will see the progressive reconciliation of two factions, and by 1980 their differences will largely be only of historical interest. Better yet, it is even possible to hope that our understanding of the determinants of fertility behavior will prove to have been strengthened by the conflict.

Notes

1. Theodore W. Schultz, ed., *Economics of the Family: Marriage, Children and Human Capital,* a conference report of the National Bureau of Economic Research (Chicago: University of Chicago Press, 1974). In this entire conference volume the substance of Easterlin's work is never mentioned and even the existence of his work is rarely noted.

2. For example, see Richard A. Easterlin, "On the relation of economic factors to recent and projected fertility changes," *Demography* 3, no. 1 (1966): 131–151; and Easterlin, "Relative economic status and the American fertility swing," in Eleanor Bernert Sheldon (ed.), *Family Economic Behavior: Problems and Prospects* (Philadelphia: J. B. Lippincott Company, 1973), pp. 170–223. Neither paper even acknowledges the existence of the Becker school.

3. Easterlin, "Does economic growth improve the human lot?", in Paul A. David and Melvin W. Reder (eds.), *Nations and Households in Economic Growth: Essays in Honor of Moses Abramovitz* (New York: Academic Press, 1974), pp. 89–125; Robert A. Pollak and Michael L. Wachter, "The relevance of the household production and its implications for the allocation of time," *Journal of Political Economy* 83, no. 2 (April 1975): 255–277; Robert A. Pollak, "Interdependent preferences," *American Economic Review* 66, no. 3 (June 1976): 309–320.

4. Gary S. Becker, "An economic analysis of fertility," in *Demographic and Economic Change in Developed Countries,* Universities-National Bureau Committee for Economic Research, Conference Series 11 (Princeton: Princeton University Press, 1960): 209–231. Footnote 10 on p. 215 contains the specification of a somewhat different model. After more than a decade had elapsed, Becker and Lewis discovered that the statements in the text 75

of Becker's paper could not be inferred from the model in footnote 10. The discussion in the text of Becker's article, however, is consistent with the two postulates presented here. For more detailed analyses of these points see Warren C. Sanderson, "Economic theories of fertility: What do they explain?", NBER Working Paper Series, No. 36 (New York: National Bureau of Economic Research, March 1974): 11–18; and Gary S. Becker and H. Gregg Lewis, "On the interaction between the quantity and quality of children," *Journal of Political Economy* 81, no. 2, part II (March/April 1973): S279–S288.

5. A translation of these two conditions into a language more familiar to economists would read, (1) the existence of a stable and well-behaved utility function for the representative household, and (2) a budget constraint for that household which is linear in the arguments of the utility function.

6. In Easterlin, "On the relation of economic factors to recent and projected fertility changes," intergenerational relative income was operationally defined as the ratio of mean money income of families with head 14–24 years old to the mean money income of families with head 35–44 years old five years earlier. All income data were converted to a 1959 base and calculated as three-year averages. In Easterlin, "Relative economic status and the American fertility swing," the intergenerational relative income relevant to fertility in year t was operationally defined as the ratio of a five-year average centered on year $t-5$ of annual mean money incomes of families with head 14–24 years old to a seven-year average centered on year $t-7$ of annual mean money incomes of families with head 45–54 years old. Again all income data were converted to real dollars.

7. Becker and Lewis, cited in note 4.

8. Robert J. Willis, "A new approach to the economic theory of fertility behavior," *Journal of Political Economy* 81, no. 2, part II (March/April 1973): S14-S64.

9. See note 4.

10. In the Becker model, the shadow price of children is πq and the shadow price of expenditures per child (called "child quality" by Becker) is πn, where π is a price index for goods consumed by children, q is expenditures per child, and n is the number of children. Since Becker assumed that n and q were arguments of the utility function, the shadow prices of children and "child quality" in his model are not independent of household consumption decisions.

11. See Becker and Lewis, cited in note 4, p. S28, footnote 1.

12. A variety of economic models of fertility behavior including two in which parents are assumed to obtain utility from expenditures on each child taken separately are presented in Sanderson, cited in note 4.

13. Becker has used the shorthand phrase "child quality" to refer to expenditures per child. Since the commonly understood notion of quality and expenditures need not be closely related (parents may make substantial expenditures on a terminally ill child), this terminological convention caused substantial misunderstanding. The broader definition of "child quality" given by Becker and Nigel Tomes in 1976 in "Child endowments and the quantity and quality of children," *Journal of Political Economy* 84, no. 4, part II (August 1976): S143-S162, accords more with its usual interpretation.

14. Willis, cited in note 8, suggested another possible offsetting force. Since childbearing takes time as well as money, the cost of a child raised at an appropriate standard of living depends not only on the quantity and price of goods consumed by children, but also on the quantity and price of time spent on children. Under a set of reasonably plausible conditions, Willis demonstrated that the relative cost of raising a child could increase as parents' income rose, even if parents did not desire to increase the standard of living of their children as their own standard of living increased. On a theoretical level, both

camps are essentially in agreement with this aspect of Willis's work.

15. Easterlin, "Population change and farm settlement in the northern United States," *Journal of Economic History* 36, no. 1 (March 1976): 45–75.

16. Gary S. Becker and Nigel Tomes, cited in note 13.

17. Becker and Tomes, cited in note 13, p. S151.

18. John S. Billings, "The diminished birth rate in the United States," *Forum* 15 (June 1893): 475. Portions of this article contain what, to my knowledge, is the most sophisticated analysis of the American fertility decline published in the nineteenth century. Readers interested in the context of the quotation in the text or more broadly in Billings' extraordinarily perceptive discussion can find a readily accessible version in the June 1976 issue of this *Review*, pp. 279–282.

19. Joseph Ambrose Banks, *Prosperity and Parenthood* (London: Routledge and Kegan Paul, Limited, 1954): p. 206.

20. See Sanderson, cited in note 4, pp. 63–67.

H
9

77

An Economist's Non-Linear Model of Self-Generated Fertility Waves

PAUL A. SAMUELSON*

Although the classical economists, such as Adam Smith, Robert Malthus, David Ricardo, and John Stuart Mill, had considered population analysis part of economics, by the early twentieth century economists had decided that demographic movements were largely exogenous to the economic system and were to be turned over to sociologists and other non-economists for scientific discussion. Richard Easterlin[1] was an exception to this trend, who endeavoured to explain by economic analysis the post-World-War II baby boom and, what is more notable, to explain the decline in fertility since 1957. The Easterlin theory, moreover, has possible interesting implications for prediction of a revival of fertility at some future date.

The Easterlin theory is therefore worthy of careful analysis for its own sake. It is also an interesting model in non-linear population analysis and one well calculated to illustrate how far wrong demographers can be if they insist on extrapolating to the non-linear domain the nice stable age distribution asymptotes of the linear Lotka–Bernardelli model.[2]

The purpose of this paper is to give an oversimplified version of the Easterlin theory, and subject it to rigorous analysis, both in the large and in the neighbourhood of equilibrium. Since the degree of oversimplification is excessive, readers will understand that the formulation is illustrative only, designed to bring out the logic of more complicated hypotheses.

* I owe thanks for financial aid to NIH Grant #1 R01 HD-09081-01 and Hiroaki Nagatani and Joel Yellin for research assistance that eliminated many errors.

[1] For classical economists' population dynamics, see an exposition such as that of W. J. Baumol, *Economic Dynamics* (New York: Macmillan, 1951). The original paper by Easterlin is R. A. Easterlin, 'The American Baby Boom in Historical Perspective', *American Economic Review*, 60 (1961) pp. 869–911, reproduced under the same title as *Occasional Paper No.* 79 of the National Bureau of Economic Research (1962). But see also the same author's 'Towards a Socioeconomic Theory of Fertility: A Survey of Recent Research on Economic Factors in American Fertility', in S. J. Behrman *et al.* (eds), *Fertility and Family Planning: A World View* (Ann Arbor: University of Michigan Press, 1969); 'On the Relation of Economic Factors to Recent and Projected Fertility Changes', *Demography*, 3 (1966), pp. 131–153; 'Relative Economic Status and the American Fertility Swing' (unpublished paper, 1975); and R. A. Easterlin and G. Condran, 'A Note on the Recent Fertility Swing in Australia, Canada, England and Wales and the United States', in *Migration, Foreign Capital and Economic Development: Essays in Honor of Brinley Thomas* (forthcoming). I am indebted to Nathan Keyfitz for the reference to N. Keyfitz, 'Population Waves', in T. N. E. Greville (ed.), *Population Dynamics* (New York: Academic Press, 1972), pp. 1–38, and to Ansley Coale for the valuable reference to R. D. Lee, 'Natural Fertility, Population Cycles and the Spectral Analysis of Series of Births and Marriages', *Journal of the American Statistical Association*, 70 (1975), pp. 295–304, which also cites as relevant T. R. Malthus, *An Essay on the Principle of Population, 1798*, Anthony Flew (ed.) (Baltimore: Penguin Books, 1970); G. U. Yule, 'Changes in the Marriage and Birth Rates in England and Wales during the Past Half Century', *Journal of the Royal Statistical Society*, 69 (1906), pp. 18–132; J. Grauman, 'Comment' in Universities–National Bureau of Economic Research, *Demographic and Economic Change in Developed Countries* (Princeton: Princeton University Press, 1960); R. D. Lee, 'The Formal Dynamics of Controlled Populations and the Echo, the Boom and the Bust', *Demography*, 11 (1974), pp. 563–585.

[2] The classical post-1911 work of A. J. Lotka, H. Bernardelli and P. H. Leslie is surveyed in N. Keyfitz, *Introduction to the Mathematics of Population* (Reading, Mass.: Addison–Wesley, 1968); A. J. Coale, *The Growth and Structure of Human Populations: A Mathematical Investigation* (Princeton: Princeton University Press, 1972); and J. H. Pollard, *Mathematical Models for the Growth of Human Populations* (Cambridge: Cambridge University Press, 1973). See J. Yellin and P. A. Samuelson, 'A Dynamical Model for Human Population', *Proceedings of the National Academy of Sciences*, 7 (1974), pp. 2813–2817, for a development of the Kendall marriage model that illustrates the pitfall of expecting a non-linear model *necessarily* to approach an exponential balanced growth. D. G. Kendall, 'Stochastic Processes and Population Growth', *Journal of the Royal Statistical Society*, B11 (1949), pp. 230–264. See P. Das Gupta, 'On Two-Sex Models Leading to Stable Populations', *Theoretical Population Biology*, 3 (1972), pp. 358–375, for invalid conjectures concerning the representation of non-linear solutions by infinite series of Lotka-type exponentials.

POPULATION STUDIES Vol. 30 nº 2 July 1976. Editors: D.V. GLASS and E. GREBENIK

In a nutshell, Easterlin suggests that families in the childbearing age groups will have more children if they are enjoying a standard of living and of economic security greater than their parents enjoyed when they were conceived and reared. Thus, those of parental age in the early 1950s were born in the 1930s when the Great Depression led to small numbers of births and when economic life was risky. Being so few in numbers in the 1950s, the usual laws of economic demand predispose their adult earnings to be high and relatively secure. They then in turn have many children. By contrast, those of parental age in the late 1960s and early 1970s were products of the war baby boom. They are relatively numerous, and the remorseless laws of the market place yield them relatively low and relatively insecure incomes.

Thus, the Easterlin hypothesis can explain fertility waves not unlike those actually experienced in the United States during the last 40 years.[3] The Easterlin theory is all the more valuable for its scarcity among economic theories, standing out in welcome relief from the rather sterile verbalizations by which economists have tended to describe fertility decisions in terms of the jargon of indifference curves, thereby tending to intimidate non-economists who have not mis-spent their youth in mastering the intricacies of modern utility theory.[4]

SIMPLIFIED EASTERLIN MODEL

Let there be two age groups: $N_1\ (t)$, the number of persons (or females) in the prime childbearing stage of life; $N_2\ (t)$ the number of those some 20 years older who are not yet quite past childbearing age but are nearing it. Let female births at t be written interchangeably as $B(t)$ or B_t, of whom $p_1 B_t$ survive to constitute $N_1\ (t+1)$ and $p_1 p_2 B_t$ survive to constitute $N_2\ (t+2)$. The p's of mortality can be assumed to be constants. Let (f_1, f_2) be the age-specific female fertility rates, as in the Lotka–Bernardelli–Leslie standard analysis. But, unlike the standard case, f_1 here will depend upon the ratio of numbers in the two adult age groups, namely on $N_1\ (t)/N_2\ (t)$. When this ratio is high, the relative incomes of those of prime childbearing ages may be expected to be low and insecure, so that f_1 is assumed by Easterlin to be a declining function of $N_1/N_2, f_1[N_1/N_2]$ with $f_1'\ [N_1/N_2] < 0$.

Now our autonomous dynamic system can be reduced to:

$$
\begin{aligned}
B(t) &= f_1 N_1(t) + f_2 N_2(t) = f_1 p_1 B(t-1) + f_2 p_1 p_2 B(t-2) \qquad (1)\\
&= p_1 f_1[p_1 B(t-1)/p_1 p_2 B(t-2)]B(t-1) + f_2 p_1 p_2 B(t-2)\\
&= a_1[B(t-1)/B(t-2)]B(t-1) + a_2 B(t-2)\\
a_1[B(t-1)/B(t-2)] &\equiv p_1 f_1[B(t-1)/p_2 B(t-2)]\\
a_1'[N_1/N_2] &< 0;\ a_2 \equiv f_2 p_1 p_2 \ \ a \text{ constant.}
\end{aligned}
$$

[3] The flavour of Easterlin's analysis is indicated by quoting some samples of his 1961 paper: '... the favorable impact ... of a swing in the rate of growth of demand – itself much larger than heretofore – was felt with much greater force [after 1939]... As a result, the rate of change of fertility reproduced the swing in labor demand in significant measure for the first time' (p. 894). 'As for prediction of the shorter-term future, the decade of the sixties, ... a relative weakening in the exceptional labor market engaged by young persons in the recent past is implied, and a consequent adverse response in the fertility rate (though not necessarily in the *number* of births)' (p. 899). 'The implications of the present analysis for the longer-term future of fertility change are in contrast with those likely to be suggested by the typical demographic discussion of our fertility history ... [namely] a resumption of the primary trend [of a long-term secular decline in fertility]. ... One might imagine a more-or-less self-generating mechanism, by which in one period a decline in the rate of labor market entry causes a concurrent rise in the rate of change of fertility, and this in turn leads with a lag of around two decades, to a rise in the rate of labor-market entry and a consequent decline in the rate of change of fertility' (p. 900). Obviously, I have not done justice to Dr Easterlin's careful qualifications.

[4] See H. Leibenstein, 'An Interpretation of the Economic Theory of Fertility', *Journal of Economic Literature*, 12 (1974), pp. 457–479, for a survey of economists' theories of fertility, including that of the Chicago School theorists, Gary Becker and T. W. Schultz, and others in the *Journal of Political Economy* (March/April 1973) supplement entitled *New Economic Approaches to Fertility*.

79

For prescribed initial $[N_1(0), N_2(0)] = [N_1^0, N_2^0]$ or $[B_{-1}, B_{-2}]$, the system forever after determines its own development,

$$B(t)=\beta(t;N_1^0, N_2^0) = b(t;B_{-1},B_{-2}),\ t\geq 0. \tag{2}$$

Those accustomed to the standard Lotka model will surmise that there exists an asymptotic exponential rate of increase with stable age distribution:

$$\begin{aligned} &\lim_{t\to\infty} B(t)\,(1+r)^{-t} = \text{constant dependent on } (B_{-1}, B_{-2}) \\ &\lim_{t\to\infty} [N_1(t)/B(t)] = p_1(1+r)^{-1} \\ &\lim_{t\to\infty} [N_2(t)/B(t)] = p_1p_2(1+r)^{-2} \end{aligned} \tag{3}$$

where $1+r = \lambda^*$ is the root of

$$\lambda^2 = a_1[\lambda]\lambda + a_2 \tag{4}$$

This is not a quadratic equation. However, if $|a_1'[\ ]|$ is small enough, there will certainly be such a real root, with $r < 0$ as $a_1[1+r]+a_2 < 1$ to a good approximation. (If a_1 were a constant as in the linear Lotka model, the quadratic equation would have a second negative root, λ_2, $|\lambda_2| < |\ \lambda_1 = \lambda^* = 1+r$, which provides transient oscillations in births.)

Readers of Easterlin will expect oscillations around the stable age distribution of an every-other generation type, oscillations which are *additional* to those incurred as a Lotka model irons out abnormalities in its initial age distribution. And they will not be disappointed. Indeed, if $a_1'[\lambda]$ is sufficiently negative near $\lambda^* = 1+r$, the exponential mode of growth will be unstable with respect to the slightest perturbation, which will send it into a 'limit-cycle' motion reminiscent of the famous cobweb model of the pig cycle.[5]

To see all this, solve the problem for $B(t)/B(t-1) = Y(t)$. Then (1) reduces to the non-linear first-order difference equation

$$\begin{aligned} Y(t+1) &= a_1[Y(t)]+a_2\,Y(t)^{-1} \\ &= F[Y(t)],\ Y(t)>0 \end{aligned} \tag{4}$$

With $a_1\,[Y]$ a positive function that is non-increasing in its argument, there will be exactly one positive root,[6] $Y^* = \lambda^* = 1+r$, to

$$Y = F[Y] = a_1[Y]+a_2\,Y^{-1},\ Y>0 \tag{5}$$

If $a_1[Y]$ were a constant,[7] Lotka and Leslie could easily prove that

[5] See M. Ezekiel, 'The Cob-Web Theorem', *Quarterly Journal of Economics*, 52 (1938), pp. 255–280, for a survey of the famous cobweb model, or the mathematical analysis and references in P. A. Samuelson, 'Dynamic Process Analysis', in H. S. Ellis (ed.), *A Survey of Contemporary Economics* (Homewood, Ill.: Richard D. Irwin for the American Economic Association, 1948–1952), Ch. 10, pp. 352–387, and 'Mathematics of Speculative Price', in R. H. Day and S. M. Robinson (eds), *Mathematical Topics in Economic Theory and Computation* (Philadelphia: Society for Industrial and Applied Mathematics, 1972). This 1971 John von Neumann Lecture is reprinted in *SIAM Review*, 15 (1973), pp. 1–42.

[6] There is no meaningful definition possible for any further root. However, there is a definable limit, for $B_{-1}/B_{-2} \neq Y^*$ and $|a_1'[\ ]|$ small, to

$$\lim_{t\to\infty} \frac{b(t+1;B_{-1}, B_{-2})\,(Y^*)^{-t-1}-b(t;B_{-1},B_{-2})\,(Y^*)^{-t}}{b(t;B_{-1},B_{-2})\,(Y^*)^{-t}-b(t-1;B_{-1},B_{-2})\,(Y^*)^{-t+1}} = \lambda_2 = +\,Y^*F'[Y^*]<0'$$

When $a_1\,[\ ]$ is a constant this is the second Lotka root. When $a_1'[\ ]$ is a small negative number, the non-linear λ_2 contains an Easterlin amplification of the Lotka transient component. When $|a_1'[\ ]|$ is large and Y^* unstable, the limit must be taken backward as $t \to -\infty$ and initial B_{-1}/B_{-2} is near Y^*.

[7] Keyfitz, *loc. cit* in footnote 1, gives a *linear* model that somewhat resembles my (1). In the absence of economic effects on fertility, he postulates (in my notation), $B(t) = R_0B(t-1)$, $0< R_0 <1$. Easterlin effects lead him to modify this to $B(t) = R_0B(t-1)+\gamma[R_0B(t-1)-B(t-2)]$, $r>0$, so that $B(t) = c_1R_0^t+c_2\,(-\gamma)^t$. Such a linear system is everywhere damped, or anti-damped, or conservative; it can never approach a limit cycle of definite

79

$$-1<F'[Y^*] = -a_2/(Y^*)^2 = \lambda_2/\lambda_1<0 \tag{6}$$

and that Y^* is a globally stable root, approached asymptotically by every solution $y[t;Y_0]$ that begins with initial positive Y_0 – namely

$$\lim_{t\to\infty} y[t;Y_0] = Y^*,\ Y_0>0 \tag{7}$$

However, with $a_1'[Y^*]$ a large enough negative number, we now have

$$F'[Y^*] = a_1'[Y^*]-a_2(Y^*)^{-2}<-1 \tag{8}$$

Hence, Y^* is locally unstable, and any initial growth at a uniform exponential rate proportional to $(Y^*)^t$ will, when perturbed ever so little, diverge from that steady path.

Indeed, if we keep $a_1[Y]$ bounded between m and $m+h = M$, but let $a_1'[Y^*]$ vary with a parameter $1/\varepsilon$, then for ε near enough to zero we can be sure that an every-other-generation limit cycle will be asymptotically approached. An example is provided by:

$$\begin{aligned} a_1[Y,\varepsilon] &\equiv M>m \quad \text{for } Y<1-\varepsilon^{-1} \\ a_1[Y,\varepsilon] &\equiv m>0 \quad \text{for } Y>1+\varepsilon^{-1} \\ a_1[Y,\varepsilon] &= \frac{1}{2}(M+m)+\frac{1}{2}\frac{(M-m)}{2\varepsilon}(Y-1),\ 1-\varepsilon^{-1}<Y<1+\varepsilon^{-1} \\ a_2 &= 1-\frac{1}{2}(M+m) \end{aligned} \tag{9}$$

Here a_2 is adjusted to make $r = 0 = Y^*-1$, so that the population happens to have a steady state with no growth. If we begin with $B_{-1} = B_{-2}$, the system will stay forever until disturbed, at $B(t) \equiv B_{-1}$, $Y(t) \equiv Y^* \equiv 1$. However, the slightest initial disturbance, $Y_0> Y^*$ will, when ε is small enough, lead to oscillations of $B(t)/B(t-1)$ every other generation, approaching a stable periodic motion of the type

$$\begin{aligned} B(2t)/B(2t-1) &= A^*> Y^* \\ B(2t+1)/B(2t) &= B^*< Y^* \end{aligned} \tag{10}$$

To solve for the exact (A^*, B^*) values we find the (A, B) roots of

$$B = F[A],\ Y^*<A = F[B],\ A = F[F[A]],\ B = F[F[B]] \tag{11}$$

eschewing the trivial root $(A,B) = (Y^*,Y^*)$. The example in (9) is designed so that, for ε small enough, there will be exactly one such pair of stationary roots to the iteration

$$Y_{t+2} = F[F[Y_t]] \equiv F_2[Y_t] \tag{12}$$

amplitude. Lee (1975), *loc. cit* in footnote 1, gives a non-linear system, similar to my (1) but involving so great a number of phase variables that its exact qualitative properties cannot be determined. However, in the neighbourhood of its exponential-growth mode, he does provide linear small-vibration approximations and even tests the implied spectrum against the historical 1900–1972 spectral periodogram: the observed 40-year cycle is reported not to be captured by the model. Belatedly, I have learned that Lee [1974, Figure 3] clearly anticipated this 'limit-cycle' finding, saying (p. 567): '. . . the equilibrium point [of a two-age non-linear model] may be unstable, in which case violent oscillations may lead to extinction . . . or a stable "limit cycle" may occur, with perpetual oscillation'. Also, I applaud Lee's recognition that, quite aside from cohorts' relative-security effects, environmental constraints and the law of diminishing returns cause Lotka–Leslie coefficients to be variable functions not constants, so that the long-run intrinsic rate of increase, $1+r^*$, must be unity in such models rather than *any* positive number as in the conventional linear models. Like reasoning led to my view: 'No economist could have devised the [standard, *conservative*] Lotka–Volterra [predator-prey] model, which ignores diminishing returns and the ratios of the species to fixed land.' See P. A. Samuelson, 'A Universal Cycle?', in R. Henn (ed.), *Methods of Operations Research* (Meisenheim: Verlag Anton Hain, 1967), pp. 307–320.

For ε very large, we will have

$$-1<\partial F[Y^*;\varepsilon]/\partial Y<0 \tag{13}$$

Therefore, as is probably realistic, *the population does settle down to its Lotka stable-growth state, but with a detectable extra tendency to resonate in doing so, with a damped every-other-generation Easterlin component.* However, when the Easterlin effect is very strong, we do have a stable limit cycle, as in (10).

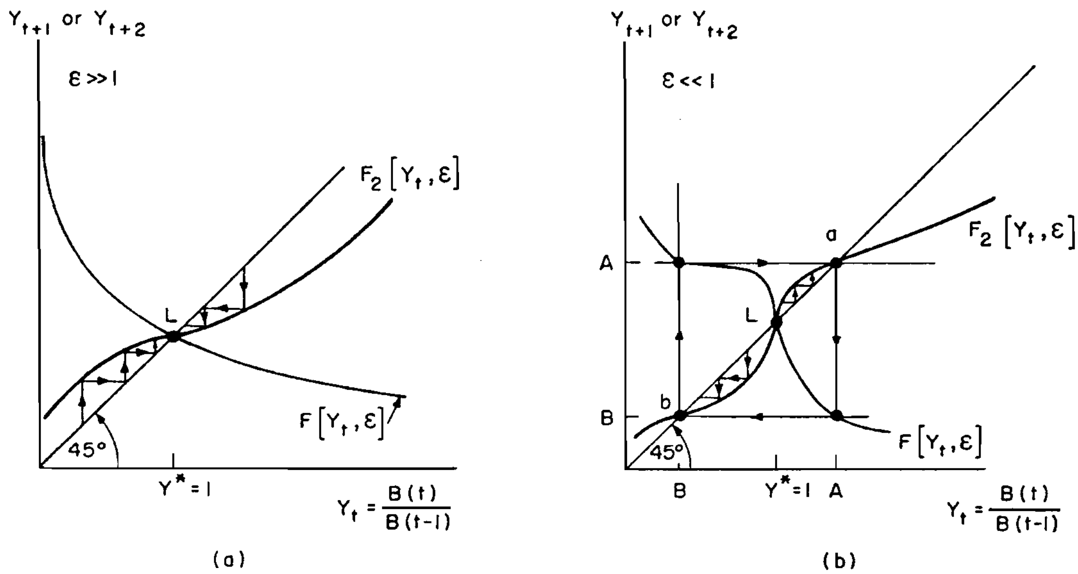


FIGURE 1. In (a) where the Easterlin fertility response to relative labour demand is weak, equilibrium at Y^* is locally and globally stable in the Lotka linear fashion – but with amplified every-other-generation transient component. In (b), where ε is small and the Easterlin effect large, the Lotka equilibrium at L is locally unstable – the slightest perturbation near L sends the system into asymptotic fertility oscillations that approach the limit cycle, $B(2t)/B(2t-1) \rightarrow A$, and $B(2t+1)/B(2t) \rightarrow B$, as $t \rightarrow \infty$.

Figure 1 illustrates the topology of the $F_2[Y_t]$ iteration. In 1(a), there is shown a damped system that returns to its Lotka exponential growth mode. In 1(b) is shown a stable limit cycle of the economic 'cobweb' type encountered in the corn-hog literature.

Data and Perspectives

Joseph E. Potter, Myriam Ordóñez G., Anthony R. Measham — 509 The Rapid Decline in Colombian Fertility

Archives

529 Inheritance, Marriage, and the Relative Size of Families

533 **Selected Publications**

549 **Abstracts**

559 **List of Contents, volumes 1 and 2**

Population and Development Review 82
Vol 2, N: 3 and 4 Sept/dec. 1976

Toward A Restatement of Demographic Transition Theory

JOHN C. CALDWELL

Our interpretation of past population movements and our expectations about future trends rest primarily on a body of observations and explanations known as "demographic transition theory." The conventional wisdom of this theory has had a deep impact and guides the work programs of international organizations, technical assistance decisions by governments, and popular analyses in the media.

The theory has changed little in the last 20 years. Indeed the period has seen a plethora of analyses of differentials in fertility, especially those found in contemporary American society, which have tended to obscure the all-important distinction between the origins of fertility decline and the subsequent demographic history of societies experiencing such decline.[1] This failure to update the theory is curious because the last two decades have provided researchers with far more experience of pretransitional and early transitional societies than they had previously been able to obtain.

It is also unfortunate because it has led to unnecessary misunderstandings, misinterpretations, and frustrations. It will be argued here that an inadequate understanding of the way in which birth levels first begin to fall has led both to premature gloom about the success of family planning programs and unnecessary hysteria about the likely long-term size of the human race, as well as to antagonisms at such forums as the Bucharest World Population Conference between countries at different stages of demographic transition.

Development and Testing of the Theory

The thrust of the paper is that there are only two types of fertility regime, with the exception of the situation at the time of transition: one where there is no economic gain to individuals from restricting fertility; and the second where there is often or eventually economic gain from such restriction. In both situations behavior is not only rational but economically rational. Another corollary is that there is not a whole range of economically rational levels of fertility in different societies, but instead only two situations, the first where the economically rational response is *an indefinitely large number* of children and the second where it is to be childless. It is admitted that in many societies at different times there is not a steep economic gradient between different levels of fertility; however, maximum and minimum family sizes in these societies are determined by personal, social, and physiological reasons, not economic ones. Further, it will be posited that the movement from a society characterized by economically unrestricted fertility to a society characterized by economically restricted fertility is essentially the product of social, rather than economic, change, *although with economic implications.* It will also be argued that the forces sustaining economically unrestricted fertility are frequently strengthened by economic modernization unaccompanied by specific types of social change and that this is the explanation for sustained high fertility in a situation in which "modernization"—urbanization, increase in the proportion of nonagricultural production, and so on—is demonstrably occurring. The social revolution—one of familial relationships and particularly of the direction of intrafamilial flows of wealth dictated by familial obligations—need not by its nature *accompany* economic modernization. However, it almost inevitably will occur either simultaneously with, or to a considerable degree preceding and perhaps hastening, economic modernization in the contemporary world. This is due largely to the phenomenon of Westernization, an essentially social process with a range of mechanisms for its spread (which have depended on economic advance in the West and to a more limited extent elsewhere, but which have not been dictated or formed by economic growth).[2]

The discussion will cover three types of society: (1) primitive societies where food gatherers, nomadic pastoralists or agriculturalists live in largely self-sufficient communities feeling little or no impact from a national state or a world religion; (2) traditional societies, predominantly agrarian, where the apparatus of a state government or the attitudes, and often the structure, of an organized religion make an impact on both community and individuals, especially in giving guarantee of safety or assistance; (3) transitional societies where rapid change in way of life towards that followed by people in lands with a "modern" economy usually in recent times has been catalyzed by outside contacts. It will be maintained that, at least in the contemporary world, the supports for unlimited fertility finally crumble in the transitional society, and that the analysis of this crumbling and of its preconditions is largely unrelated to the analysis of the frequently slow and sometimes vicissitudinous reduction in family size that subsequently occurs in transitional and modern societies. Much of the argument draws primarily on African examples, both because of my experience in Africa, and because all three types of society are well represented on the continent.

Demographic Transition Theory

By the end of the nineteenth century it was common knowledge that fertility levels were falling in many Western countries and there was a presumption that birth rates would stabilize at lower levels (although there was no agreement about what the new levels would mean in terms of natural increase). An attempt was made by Warren Thompson in 1929 to divide this transition into three phases and by C. P. Blacker in 1947 to distinguish five phases.[3] Neither could be said to be the father of demographic transition theory in that neither suggested an explanation for fertility change.

Modern demographic transition theory was born almost in mature form in a paper written by Frank Notestein in 1945. Notestein offered a twofold explanation for why fertility had begun to decline. Fertility in premodern countries had been kept, if not artificially high, then high only by the maintenance of a whole series of props: "religious doctrines, moral codes, laws, education, community customs, marriage habits and family organizations . . . all focused towards maintaining high fertility."[4] High fertility was necessary for survival because otherwise the very high mortality rate would have led to population decline and extinction. But eventually in country after country mortality began to decline, and the props were no longer needed or were not needed at their original strength. One could leave the explanation here and argue that the props would inevitably wither, as social adjustments were made in response to other changes. However, Notestein put forward the view that, in the West at least, more positive forces (arising out of the same process of modernization that had brought the death rates down) were at work destroying the props. Fundamental was "the growth of huge and mobile city populations," which tended to dissolve the largely corporate, family-based way of life of traditional society, replacing it with individualism marked above all by growing personal aspirations. Large families became 83 "a progressively difficult undertaking; expensive and difficult for a population increasingly freed from older taboos and increasingly willing to solve its problems rather than accept them."[5]

Again in 1953 Notestein pointed to the "urban industrial society" as the crucible of demographic transition and stated, "It is difficult to avoid the conclusion that the development of technology lies at the root of the matter." Once again he placed emphasis on the erosion of the traditional family, "particularly the extended family," and on the growth of individualism, but he also drew attention to other important social movements: "the development of a rational and secular point of view; the growing awareness of the world and modern techniques through popular education; improved health; and the appearance of alternatives to early marriage and childbearing as a means of livelihood and prestige for women." However, this time the description of pretransitional society was not drawn largely from the experience of the West but was generalized to include the developing world:

> The economic organization of relatively self-sufficient agrarian communities turns almost wholly upon the family, and the perpetuation of the family is the main guarantee of support and elemental security. When death rates are high the individual's life is relatively insecure and unimportant. The individual's status in life tends to be that to which he was born. There is, therefore, rather little striving for advancement. Education is brief, and children begin their economic contributions early in life. In such societies, moreover, there is scant opportunity for women to achieve either economic support or personal prestige outside the roles of wife and mother, and women's economic functions are organized in ways that are compatible with continuous childbearing.[6]

The mainstream arguments of the theory are that fertility is high in poor, traditional societies because of high mortality, the lack of opportunities for individual advancement, and the economic value of children. All these things change with modernization or urban industrialism, and individuals, once their viewpoints become reoriented to the changes that have taken place, can make use of the new opportunities.[7]

The argument appears at first clear and convincing, but it has elements and implications that are more complex or debatable and that have had an enormous effect on our way of looking at demographic change. The most fundamental issue is whether the theory actually deals with reactions and accommodations to material circumstances. There is a persistent strain in demographic transition theory writings that claims that rationality comes only with industrial, urban society, and a related strain that regards traditional agrarian societies as essentially brutish and superstitious. This arises in two distinct ways.

The first is from the references to pre-demographic-transition society.

The concept of the brutishness of the poor, and their inability and unwillingness to help themselves, is a fundamental proposition of Malthus. But the origin of the view in modern demographic transition theory is the argument that, in spite of the high mortality, insecurity, and lack of cost of children in pre-demographic-transition societies, all kinds of religious and social institutions and preserves were needed to keep fertility high. This is why demographic transition literature is full of references not to the behavior or reactions of such people but to *attitudes, beliefs, traditions,* and *irrationality.* Kingsley Davis wrote of the contrast between traditional societies and "the growing rationalism of modern life" and, again, describing sex and reproduction in the former, that "towards this aspect of life the woman has mainly a nonrational approach—religious, supersitious and incurious"[8]; George Stolnitz described "a shift in attitudes from the traditional fatalism of peasant societies"[9]; Eva Mueller observed that, "it is difficult to influence deep-seated attitudes"[10]; William Rich believed that "large-scale fertility declines cannot be expected until the living conditions of the majority of the population improve enough so that they no longer *consider* large families necessary for economic reasons"[11]; Stephen Enke deduced that, "many simple peoples understand very little about why reproduction occurs and how it can be prevented"[12]; Michael Endres has written recently, "people directed by tradition resist rational intervention and choice between behavioral patterns," and "to urge upon a traditional people a rational technical means of birth control is to challenge the tenacious hold of a hard-won culture to which choice and change are the enemy"[13]; while G. T. Trewartha indicted the irrationality of premodern society for causing not only high fertility but also maldistribution of settlement: "Indeed, much of the distribution does not appear to be particularly rational . . . Tradition, which is unusually strong among the tribal peoples of Negro Africa, plays a more than ordinary role."[14]

The second respect in which an implicit assumption of pretransition irrationality enters into the theory is through references to cultural lags in making fertility adjustments to the arrival of the new urban, industrial conditions. Such references are plausible in a way because a period of change is under consideration instead of an extended stable situation. Several of the quotations above do refer also to such lags, but the concept is both implicit and explicit in Notestein's 1945 paper. There he argued that the supports for high fertility "change only gradually and in response to the strongest stimulation" and described "a population increasingly freed from older taboos and increasingly willing to solve its problems rather than accept them."[15]

That the central tradition of demographic transition theory is still very much that of Notestein's 1945 and 1953 formulations and that the belief in increasing rationality with modernization is still an integral

84

element has been demonstrated vividly by the publication of the most recent United Nations *Population Studies,* which justifies the latest United Nations population projections. The argument is worth quoting at some length:

> The entire process of economic and social development . . . itself changes people's outlooks from traditions and fatalism towards modern concepts and rationalism. . . . The past record in the more developed countries demonstrates . . . that it [fertility decline] can . . . be expected to occur in the normal course of the modern development process . . . the deliberate regulation of fertility defies age-old custom. . . . A high frequency of childbirth . . . was necessary for the continuation and security of families and this found emphatically strong support in the prevailing values and customs. In many cultures it has also been considered that children provide a much needed insurance against destitution in old age. Associated with such cultural norms has been the regard for women in their seemingly principal function as bearers and rearers of children, limiting thereby their participation in economic and social roles held to be mainly the prerogative of men. Interwoven with such attitudes there can also be a fatalistic refusal, or even an abhorrence, to contemplate any regulatory interference with the reproductory process. It is not to be wondered at that such a traditional outlook on life can be highly resistant to change. But as shown by the earlier experience of the more developed regions . . . change is possible or eventually to be expected.[16]

Much of the argument for demographic transition concepts as they are now widely held turns on the definition of rational. The term "economically rational" is frequently substituted so as to avoid having to judge "social rationality" with the possibility of having to agree that a certain mode of behavior was rational in a given setting in that it met the ends of religious beliefs or of community obligations. Even so, the criteria employed are highly ethnocentric and are laden with Western values. It is assumed that it is rational for a man or a couple to maximize the expenditure on the individuals in his or their nuclear family; but there are any number of non-Western societies in which there is greater pleasure in spending on some relatives outside the nuclear family (adult brothers for instance) than on some within it, and in which children are happier to spend on parents than are parents on children. Obviously the fundamental choices are social ones and economic behavior is rational only insofar as it is rational within the framework established by social ends. What demographic transition theory has always regarded as rational are primarily Western social ends with economically logical steps to maximize satisfactions given those ends.

The underlying assumption of this study is that all societies are economically rational. The point is a simple one, but its acceptance is absolutely necessary if we are to arrive at an adequate theory of demographic transition, if we are to understand the contemporary population changes, and if we are going to make adequate predictions for planning purposes. It is, in fact, difficult to have a rigorous analysis on any other assumption. Social ends differ but can be largely explained on a rational basis—usually even in economic terms. Furthermore, change in social ends can often be observed, measured, explained, and predicted. The view that the fertility behavior of the Third World arises largely from ignorance and should be combatted with education and guidance is held strongly by many family planning movements and leads to friction and even confrontation; the same reaction arising out of much the same origins was witnessed writ a little larger at the Bucharest Conference. Indeed the view that peasants are usually mistaken in evaluating the effect of their fertility on their own economic well-being has recently been seriously argued in a paper by Mueller.[17]

A second implication of demographic transition theory, at least as originally conceived by Notestein, is that industrialization and concomitant urbanization are preconditions to development. Notestein placed stress on "urban industrial living" (in 1945) and later on "urban industrial society" (in 1953), as the context in which the social changes leading to fertility decline occur. Similarly Thompson (in 1946) referred to "industrialization" as the necessary condition. In the last 20 years such terms have largely been replaced by "modernization" or near synonyms like "the modern development process" as it became clear that great numbers of people in the Third World were unlikely to be living in industrial cities for generations. The demographic transition theory did allow for the possibility that the new way of life and the consequent new fertility behavior might be generated in the urban industrial setting and then be exported to nonurban and nonindustrial populations either by exporting some of its institutions (such as schools, women's rights legislation or a full market economy) or by simply exporting its attitudes or ideas. This tenet received historical support from the decline in fertility among rural populations in the West. The theory did not specify whether the urban industrial melting pot from which the changes were derived had to be in the same society or whether a global economy and society was beginning to operate that could export the necessary ideas and institutions from the economically developed countries to the commercial cities of Asia and Africa and on to the rural hinterlands. (Demonstrably this has long been happening with regard to governmental institutions and more recently in terms of schools and political ideology.) In any case the link with the emphasis on the props for high fertility is clear. If high fertility in developing countries were a wholly rational response to economic circumstances, then the small family pattern could never be exported; but, if the large family were to a considerable extent the product of beliefs and attitudes sustained largely by religion and shib- 85 boleth in order to compete with high mortality rather than to meet the needs of the economic system, then export was quite possible. Those who doubted the validity of a theory based only on the transmission of ideas but who were prepared to accept the possibility that the spread of small families could be achieved by the spread of institutions made little progress in identifying those institutions that were minimally necessary for fertility transition—schools? nonagricultural employment?

A third problem lurked in demographic transition theory but was not specifically identified. Was it primarily modernization that was being exported? Is there a specific form of social modernization that is a necessary adjunct to economic modernization? Or is the export Westernization, which by historical accident has been tailored to fit the world's first economic modernization and which is easily exportable partly because of the West's economic strength (clearly visible in its earlier ability to colonize) and partly because this tailoring makes it easily adaptable to modernizing economies? Notestein wrestled with problem areas in his 1953 paper and the whole question of Westernization almost arose: why had fertility fallen steeply between World Wars I and II in almost wholly agricultural Bulgaria while failing to do so during the 1950s in the larger urban areas of Egypt and the Far East?[18]

Suggested Modifications to the Theory Without actually saying as much, Davis argued in 1955 and again, with Judith Blake, in 1956 that the props were not needed. High fertility was a perfectly rational response to socioeconomic conditions in a traditional agrarian society: the extended family means that the cost and care of children are shared; children, once past infancy, may in fact pay for their costs, especially in conditions of cottage industry, but more generally in any farming situation; both husbands' and wives' families of origin may help establish the newly married couple, often on a farm of their own; large families may bring economic strength through political strength in the local decision-making organizations.[19]

Recently this aspect of the demographic transition debate has been summarized and evaluated by Thomas Burch and Murray Gendell, who demonstrated that research findings from India and Taiwan fail to show the predicted fertility contrasts between families residing as nuclear families and those living together in larger agglomerations of relatives.[20] The point is an important one, and, in order to clear the way for the subsequent argument in this paper, should be dealt with here. The research in India and Taiwan is almost certainly irrelevant for three reasons, of which the second is most important. The first is that survey or census data do not accurately measure even residential family size. The building materials, mud and stone in contrast to bamboo and thatch for instance, often determine whether considerable numbers of people can be housed in a single structure or alternatively in several smaller structures adjacent or close by. The second (a point to be elaborated later) is that family residence arrangements have little or nothing to do with the true extended family of mutual obligations, at least as long as residence outside the traditional community is not specified. It is the size and ramifications of this family of obligations that may well help to determine fertility. The third is that family residential patterns are often a function of the life cycle; in some societies nuclear residence is most likely to be found immediately after husband and wife (often with children of their own by this time) first move away from their parents to a farm or business of their own. What demographers should really be interested in are the families of this type who are unlikely to subsequently attract or retain many other relatives (except perhaps aged parents or nephews and nieces undergoing education) often because they have moved to a city or have been fairly highly educated and so have opted for a different way of life from their relatives.

Family sociologists added some riders to the picture. William Goode decided that the nuclear family's fundamental demographic characteristic was not that it leaned toward small size but that it was more flexible than the extended family in reacting to economic conditions favoring high or low fertility; thus at much the same time (eighteenth and early nineteenth centuries) European populations had chosen high fertility in frontier North America and moderate fertility in their homelands in Europe.[21] This had, of course, been a major contention of Malthus. Some, Colin Clark, for example, went further and identified nuclear families with advanced economies and extended families with nonindustrial societies—probably, as will be argued later, a fundamental mistake at least in terms of European history.[22]

Another attack on the props came from David Heer and Dean Smith who argued that the props had at every stage been wholly rational because of high mortality and had withered as the death rates fell.[23]

Recent Ideas An important contribution in the 1950s was that of the economists, especially Ansley Coale and Edgar Hoover in 1958 with a major analysis of India, together with Mexico. What is apt to be overlooked is that Coale and Hoover accepted as their starting point the existing demographic transition theory,[24] and that most of the subsequent economic analysis is independent of theories about when and if fertility is likely to fall. Coale and Hoover spelled out the economic implications of transition theory but they did not test its basic assumptions. Their analyses were essentially those of macroscopic data, and their main conclusion was that national economic growth is impaired if fertility levels too greatly exceed mortality levels. However, most nonspecialists received the message that they had shown convincingly that high fertility is economically disadvantageous for every size of population unit, and the view that high-fertility agrarian families were behaving irrationally was given a powerful boost.

It is possible to extrapolate part of the argument from national populations to individual families: to suggest that lower fertility will produce a family age structure with a higher ratio of potential adult producers to child consumers than will high fertility and that fewer children will allow mothers to participate more in economic activity.[25] For reasons analyzed below all these arguments ring somewhat hollowly in an actual agrarian society: children work at young ages; often the peasant's analysis is dynamic in contrast to the demographer's static one in that the peasant is thinking less of the present and more of safeguarding the future; and, in many societies, the peasant's wife already works long hours (freed from minding the product of her recent fertility by the child care being practiced by the product of her earlier fertility).

Two years before Coale and Hoover's study appeared, R. Nelson had produced his "low-level equilibrium trap model." Subsequently Harvey Liebenstein made the model more specifically demographic, suggesting that in "backward areas" people are merely caught by circumstances: they lack the inducement to save or invest and are unlikely to make quantum jumps in technology; as a result, per capita income remains static, mortality does not decline, and, hence, population does not grow.[26] The model does imply at least short-term rationality, although it could also be taken to mean that the society as a whole was incapable of planning its course to a better future. A more important limitation is that the model seems to have no real significance for social theory (except for historical studies) in a world where societies are no longer isolated from each other and where imported health technology means that population is growing increasingly fast, even in many societies with largely subsistence economies.

In 1974 Julian Simon summarized and assessed much of the research evidence available on fertility and stage of economic development, concluding that "fertility is everywhere subject to much rational control." He largely avoided the question of why—within this framework of rational decision—fertility decline sets in, contenting himself with pragmatically observing that "we may rely on the fact that, as education rises, fertility will fall" and that "if one wishes to reduce fertility, one should think about raising educational levels as well as aiding birth control."[27]

Since the 1950s, sociologists have contributed powerfully—not always intentionally—to the thesis of irrationality by apparently showing a substantial gap between desired and achieved fertility in the Third World (together with a smaller gap in developed economies). The origin of this formulation dated from the beginning of fertility studies, when the Indianapolis Survey of 1941 asked American respondents what they considered the ideal family size. The concept of "norms" had been one of the basic planks of modern sociology, and in the early 1960s Ronald Freedman applied it to fertility studies in a way that seemed to have implications not only for behavioral rationality but for behavioral economic rationality: "family size norms will tend to correspond to a number which maximizes the net utility to be derived from having children in the society or stratum." In developing countries, he concluded, "there may be a delicate balance of pressures towards higher fertility to ensure at least a certain minimum number of children and counter pressures to minimize or eliminate an intolerable surplus of children under difficult subsistence conditions."[28]

During the mid-1960s, knowledge, attitude, and practice (KAP) surveys were used to measure desired or "ideal family size" in the developing world using questions about the "best" or "ideal" number of children or the family size that would be desired if the respondent were to start her reproductive history all over again. Comparisons made in 1965 between "desired" and actual fertility prompted W. Parker Mauldin to state, "although it is not yet true that people in the developing areas share the small family ideal, it is true that most of them no longer want very large families,"[29] and Bernard Berelson to calculate that, while ideal family size in the United States was 97 percent of the achieved size, it ranged in a number of developing countries between 60 and 92 percent.[30]

The whole question of ideal family size is of the utmost importance for the discussion of demographic transition theory in this paper. It is not necessary to regard the gap between ideal and achieved size as evidence of irrational behavior; indeed Berelson regarded it as arising from "lack of information, services and supplies" and this was the most common position taken during the 1960s by technical aid organizations in the family planning field. Indeed the significant gap—that created by the props, according to demographic transition theory—is essentially that between the family size which would be dictated by economically rational behavior and ideal family size. In fact there is little relationship between the demographic transition concern with the attainment of economic rationality and the KAP study attention to ideal family size; KAP studies essentially attempt to measure potential consumer demand, and in this they ignore the issue of rationality except to the extent that it seems reasonable for a person to do what he wants to do. Some researchers appear to take it for granted, however, that a movement in ideals is almost inevitably a movement toward rationality and, hence, evidence of the decay of the props.

There are three fundamental questions.

The first is whether there are "norms" at all in the high-fertility situation. It will be argued here that economically there is no ceiling in primitive and traditional societies to the number of children who would be economically beneficial; the actual number is kept down because physiological and social problems arise from too frequent childbirth and the failure to cease childbearing at a certain stage. Achieved fertility is a product of this conflict and can hardly be described as approximating a norm.[31]

The second question is whether fertility behavior must be regarded as mainly economically motivated, or whether social motivations are also important or even dominant—whether norms, if they exist, and fertility behavior can be taken as an approximate measure of the individual's reaction to economic circumstances. Simon argues that fertility can be taken to be primarily economically motivated and justifies "an important omission [from his study] . . . , social norms and values. The reason . . . is that in the context of long-run analysis, culture and values do *not* have independent lives of their own."[32] This, it will be noted, is a direct assault on the props. This proposition differs from that put forward in this paper in that the argument here is that fertility is economically rational only between certain limits that are set by noneconomic factors; that there are two types of society, one in which it is economically rational for fertility to be ever lower, but in which a floor is interposed by noneconomic considerations, and the other, in which it is rational for it to be ever higher, restrained only by a noneconomic ceiling.

The third question is whether fertility can be used as a measure of desired behavior. The apparent demonstration by the KAP surveys that there is a wide gulf between what Third World people want to do and what they succeed in doing introduced a large element of chance (and not random chance at that) into the whole matter. It is perhaps impossible to study the motivation behind fertility decline if the populations of the Third World habitually exhibit fertility well above what both economic rationality and the attitudes molded by the props dictate. I suggest that this apparent gap is partly the product of the present unusual circumstances, but largely an artifact of the method of investigation. Change is at present so rapid in many societies that there is a fast increase in the number of people who will economically benefit from lower fertility. However, the "ideal family" questions ultimately fail to measure likely fertility behavior even under conditions of adequate access to contraception because they are imported almost undigested from Western society and contain a range of assumptions about non-Western societies that will not bear up under examination. The fundamental problem is the questioning of a woman about the "best" number of children, as if the chief cultural thrust were optimization of family size instead of a range of other concerns such as meeting the expectations of husband and other relatives, conforming with peer group behavior, and so on. In many surveys most respondents probably do not fully understand the question. They know what the words mean, but they also know that they are being asked to define "best" in a modernizing sense by interviewers (and, behind them, some institution) who interpret "best" in a futuristic sense or in the sense of the elites. The "politeness response" is only a small part of the reaction.[33] The "ideal family" question was shaped by Western, middle-class researchers, living in conjugal families in which husbands and wives consult each other over matters of reproduction and sex, and it achieves its greatest reliability among such people. In this paper it will be taken that achieved fertility everywhere comes close to being a rational response to the circumstances of the society.

In 1965 the publication of a United Nations study directed the attention of researchers to the prime importance of the changing conditions that lead to fertility decline at a point identified as the "threshold." The analysis distinguished six levels of fertility, in what was essentially a cross-sectional and not an historical analysis, but for further analysis combined the levels into two groups, one in which relatively low fertility had been achieved and the other in which it had not. Every Asian and African population, except Japan, was in the high-fertility group, while, with the exception of Albania, every European population in Europe, North America, and Oceania was in the low-fertility group. In Latin America, only Argentina and Uruguay were among the low-fertility countries. The United Nations recognized that it was "perhaps no coincidence that most of the countries where fertility is low . . . are in Europe and European-settled regions," concluding that "fertility levels might . . . be due . . . at least partly to culturally determined circumstances affecting the interactions between fertility and economic and social changes."[34] This dichotomy had the disadvantage that the nations identified as being beyond the threshold had in many cases passed it long ago; and neither the nature of the actual threshold nor the changes sufficient to ensure movement across it were actually detected.[35]

Other attempts to apply or develop threshold analysis have been made. Etienne van de Walle and John Knodel failed to find it a usable tool when analyzing fertility decline in France and Germany.[36] Dudley Kirk proclaimed the value of such an approach in 1971, and in 1975, together with Frank Oechsli, applied it to Latin America, calculating a "Development Index" and relating it to declines in both mortality and fertility.[37] But Oechsli and Kirk's data unmistakably evidence a cultural dichotomy: most of the countries with reduced fertility either are areas of almost purely European settlement in the extreme south or are Caribbean Islands with very mixed cultures and population origins. Island nations have been conspicuous in recent fertility declines, and the United Nations has identified ten and attempted to explain the change in terms of their small size and hence the easy penetration of ideas and health measures.[38] Yet seven of the island nations were settled entirely by immigrant populations while under European control: Réunion, Jamaica, Mauritius, Trinidad and Tobago, Guadeloupe, Martinique and Puerto Rico; one has been entirely Christianized: American Samoa; one is a mixture of an immigrant population and a fully Christianized indigenous one: Fiji; and one has achieved universal Western-style education: Sri Lanka.

In contrast to the approach of the thresholders, there has recently been renewed interest in the innovational explanation. (In the late nineteenth and early twentieth centuries, governments and other institutions almost invariably explained fertility control innovationally, as the spread of pernicious ideas.) Much of this has arisen from the Princeton Office of Population Research European fertility project and its demonstration that fertility declines spread fairly rapidly through linguistic or religious units only to be halted at their borders.[39]

The threshold and innovational approaches share a common problem in explaining the onset of fertility decline. Their data are usually for considerable aggregates of population, and, hence, it is difficult to determine whether the measured drop in fertility is attributable to a single socioeconomic group or not. If it is, then the threshold explanation holds up (provided that the threshold indices are meant to apply to subsections of a society), but the spread of innovation is shown to have an impact only on groups that have already reached some potential state of receptivity as measured by socioeconomic indices and not by attitudinal changes; if it is not, then the threshold indices can be discarded as measures of the sufficient conditions that must be met for demographic change to occur. In any case both approaches have failed as yet to specify the kinds of changes necessary for individuals or couples to alter their fertility behavior and why such alterations take place.

Attempts have of course been made to investigate these changes around the beginning of transition, the most ambitious to date for developing countries being the East-West Population Institute's Value of Children Study.[40] So far the published national reports (on the Philippines and Hawaii) have had a strong social psychological orientation toward beliefs and values—stronger even than the questionnaires upon which they are based. The approach is clearly an aspect of innovational theory and has a good deal in common with explanations that rely heavily on the props; and, although it does not spell it out, the Philippines report could be described as an analysis of the import and diffusion of nonindigenous cultural values. So far, the project has insufficiently investigated the changing material aspects of life and the extent to which changing values could be said to be rationally moving parallel to economic realities.

New Experience Increasingly massive family planning programs in Asia and parts of Africa, Latin America, and Oceania over the last quarter of a century have presented an enormous increase in opportunities to watch and measure fertility transition and to identify the innovators. This should have allowed demographic transition theory to be rewritten with the sureness that arises from large-scale field experiments. This has not happened, and one of the keys to the whole problem may be why it has not happened.

An important reason is undoubtedly described by the well-known precept in other areas of endeavor: applied science has increasingly limited returns, unless based on continuing fundamental research. Too much of the research has taken as its starting point and framework the preexisting conclusions of demographic transition theory. Too many frustrated family planning fieldworkers and administrators have been only too willing to blame the props for the failure to achieve program targets. Most indigenous and all expatriate administrators and advisors are in circumstances in which they economically benefit from controlling their own fertility, and they find it hard to understand why this should not be so for everyone else—irrationality is an easy answer especially when it can be demonstrated that education and demand for the family planning services are highly positively correlated. Probably too much of the research has been program-based instead of concentrating on the mechanisms of change in the society as a whole. Yet this is not the whole explanation. The operational research has permitted the identification of large numbers of innovators—at least in terms of using contraception, if not always in terms of deciding to restrict family size—but research has not clearly established the basic changes that have affected these people. On the face of it this seems hard to believe, and yet it is true for a number of reasons. One (as will be seen below) is that the innovators do not really know themselves; they differ in various ways from their parents and these differences make fertility control rational, but they usually cannot identify the essential differences. Another reason for the failure to identify preconditions is that comparison of the characteristics of family planning acceptors and nonacceptors shows that the former are much more likely to exhibit not merely one "modern" characteristic but a whole interrelated set (more education, nonagricultural employment, higher incomes, and so on), so that there is a chicken-and-egg problem. There has also been a research failure: failure to investigate in detail the way of life and circumstances of individual acceptors parallel to similar studies of the population as a whole.

In relation to the last point it might be noted that there has been over the last half century a considerable advance in economic anthropology, which has been almost entirely ignored by demographers.[41] Fierce debate has raged in economic anthropology between the Formalists and the Substantivists, the former claiming that Western economic analysis can be applied unchanged to all economic life and the latter maintaining that economics serves social ends and that every culture has its own economic theory. The Formalists narrowly define the subject of modern economics as allocation of scarce resources between either unlimited or numerous ends, while the Substantivists contend that rational economic behavior is rational only within a given social context and that these contexts are diverse and often startlingly different from those of the modern West. The Substantivists have also established that, even where money and markets exist, these may embrace only part of a society, and, more importantly, only part of the life of much of the population. The rest of the society, and perhaps the bulk of the life of most of its citizens, falls in the more traditional sector, where it is not rational, and usually not possible, to act out the life of market-economy man. The implications for demographic transition theory are that transition is made possible only by profound changes in the social structures of such societies, and that analyses of the economic rationality of high fertility reach different conclusions in different social structures.

Fundamental Problems of Research Part of the failure to advance demographic transition theory can undoubtedly be blamed on inadequate research. The basic problem has not been inadequate methodology but rather poor application, especially in the application of methods in cultures other than those for which they were developed. The problems will only be summarized here as they have been treated more adequately elsewhere. The general failing, and one that encompasses the others, has been ethnocentricity. Too much research has been done too quickly and on too large a scale with research instruments, and often researchers, brought directly from contemporary Western society. Too often, the representatives of the non-Western society in the research have been completely inculcated with Western research approaches and conclusions in Western universities. As a result, the research approach often predetermines the range of findings and asks questions that provide the appropriate answers almost by an echo effect.[42] What prevents the researcher from worrying about the extent to which the pattern of responses fails to represent the society is the magnitude and flow-chart nature of modern social scientific research: the large sample, the hierarchy of command, the precoded questionnaire, the responses as invisible magnetic recordings on a computer tape, computer editing, the computer print-outs of marginals and cross-tabulations that necessarily balance to the last unit, the written report in a predetermined pattern, and finally the cross-cultural international comparison with other research using similar or even identical instruments.

Four pitfalls of current research have particularly contributed to misunderstanding of the nature of demographic transition.

1. The magnitude and direction of wealth (money, goods, services, guarantees) flows and potential flows are areas of research that are often neglected or misunderstood. Such research is difficult. In premodern societies much of the wealth is still outside the monetized economy. Often money-equivalents are not visualized; services usually have an element of obligation; investments in future security may be discounted in the opposite direction to that to which Western economics is accustomed (discussed further below); the details about wealth have often not been disclosed even to immediate relatives (who exert competing demands and from whom details must often be hidden, more to prevent resentment and to allow equity to prevail, than to deprive people of their just deserts); and there is sometimes also a fear about tax officials and other authorities knowing about earnings. In these circumstances, small-scale, painstakingly thorough research is needed by investigators with a thorough knowledge of the society. Hardly any good research has yet been done. There is a temptation to quote inadequate or incomplete research, with highly misleading results. There would be less danger if the errors were random, but, without question, there is a great understatement of all flows of wealth and potential wealth.

2. The "family" of the fertility survey is often an artifact of the survey. Women are asked about their own reactions and their husbands' reactions, and of course, the women answer in these terms. No one describes the role in decision-making of the husbands' and wives' lineages; no one explains that the husband regards his brother as a nearer relation than his wife in the sense of that close inner circle where one no longer regards expenditure as depriving one personally of wealth; no one explains the intricate system of decision-making and obligations that may far exceed the nuclear family or residential group and in which the nuclear family may not even be a recognizable subunit.

3. The nature of family formation and of related decisions in developing countries is frequently misunderstood. Family size decisions are usually out of the respondents' hands for several reasons: both the physiological side of reproduction and the obeying of cultural practices may seem (sensibly enough) to them to be something they cannot control and hence there is an element of fatalism; family size is often the product of decisions taken for family reasons not primarily aimed at determining fertility; and, where there are decisions to be made, they may not be primarily decisions of the "couple." All these factors must be taken into account when interpreting "Up to God" and "Don't know" responses, which may be closer to the truth than the numerical ones. In these circumstances the value of any "ideal family" type of question is debatable, and the employment of the concept of "norms" misleading.

4. While fertility transition research is essentially a study of change, such investigations have been impeded by too much emphasis on modernization. Change can be understood only if emphasis is given to studying the fundamental nature of the society that is being subjected to new forces. Too many survey questions are focused on the modernizing features, and too many of them have a built-in assumption that everyone is reaching for such change. 90
Demographers have been far too rarely concerned with familiarizing themselves

with the knowledge other social scientists have already accumulated about the society being examined. Perhaps even more serious is the fact that modernization has been accorded such respect (by all development researchers, but specifically by population researchers, in that they regard modernization as being the chief mechanism for reducing fertility and hence eventually containing global population growth) that its components have usually not been analyzed and the all-important distinction has not been made between Westernization, which may proceed at a rate unrelated to economic change, and residual modernization, which must go hand in hand with economic change because it is either a necessary condition or a necessary product.

What we obtain from research that is vitiated by these weaknesses is a reflection of the way a poorer version of our own society might be expected to behave if set down in a Third World context. We fail to appreciate significantly different social and economic structures and the extent to which these yield rewards to the highly fertile.

A Society Experiencing Change

The observations in this section are primarily of Nigerian Yoruba society. The Yoruba are the indigenous inhabitants of Nigeria's Western State (recently subdivided into Ogun, Ondo, and Oyo States) and Lagos State, as well as considerable parts of Kwara State in Nigeria and Southern Benin, or Dahomey. The Western and Lagos States are believed to have contained about 8.5 million people in 1962[43] and contain perhaps 13 million now, of whom over 11 million are Yoruba, out of a total of 13 million Yoruba in Nigeria and Benin. The Yoruba of the Western and Lagos States have been the focus of the largest segment of the Changing African Family Project and of the Nigerian Family Study, and many of the data used here are drawn from that study.[44] The area is well suited to this kind of investigation, because a primitive society (as defined here) existed over most of it until the latter part of the nineteenth century (and aspects of it can still be studied in any rural area); the traditional society is now paramount; and some of the population—largely the urban population and especially the middle classes of the cities (Lagos probably has over 2 million inhabitants and Ibadan 750,000)—are part of transitional society.

The Primitive Society A primitive society is one in which the largest organizational institution is the tribe, the clan, or the village. No overall responsibility is taken by the larger apparatus of State or Church, which means that security within the groupings that exist is not augmented or guaranteed by an outside entity. Indeed, security outside the group is minimal; nearly everyone continues to live among their people of origin; and the size of that group is often the measure of safety.

*Several aspects of such a society are of prime importance for under*standing all pre-demographic-transition societies.

Perhaps the foremost is that the society or economy (for they cannot be separated) of the group is a single system in which the participants have time-honored roles and duties. There is usually communal land (which is essential in nomadic, food-gathering, and most shifting-cultivation systems); residence in propinquity to large numbers of people—mostly relatives—with whom one has lived all one's life; government by these same people; and a simple economy where much cooperation is needed for the larger tasks. The absolute right of individual ownership is unknown. In fact economic relations and social relationships intermingle. Edward Evans-Pritchard wrote of the Sudan, "One cannot treat Nuer economic relations by themselves, for they always form part of direct social relationships of a general kind,"[45] and C. K. Meek of Nigeria, "One of the main distinctions between Native systems of holding land and those of Western societies is that the former are largely dominated by personal relationships, whereas the latter are subject to the impersonal legal conception of 'contract'."[46] Marshall Sahlins summarized the position as, "A material transaction is usually a momentary episode in continuous social relations."[47] Transactions and gifts are not in fact markedly differentiated, especially as the latter are almost invariably also the cause of two-way flows of wealth.

Gifts of goods or services and later reciprocation allow the creation of a security system of mutual obligations (which will be dealt with in this review of the primitive society, even though such systems are of fundamental economic and demographic importance in traditional and transitional societies and survive even into modern society[48]). In all primitive and most traditional societies the maximization of profit or other ends in good times is of small importance compared with the minimization of risks (which often means ensuring survival) in bad times. Describing the Fulbe (or Fulani) of northern Nigeria, C. Edward Hopen reported that they "have an almost pathological concern (and often fear) for the future. Their conversation abounds with such expressions as 'tojaango' (what of tomorrow) and 'gam jaango' (because of tomorrow). . . . The prospect of a secure and relatively care-free old age under the care of their sons will often restrain young women from deserting or divorcing their husbands. Both men and women in many respects show a remarkable disposition to forego present convenience (or pleasure) in the interests of future benefit."[49] Such attitudes are universally reported by field researchers, even among the businessmen of Ghana's capital, Accra.[50]

The fertility implications are obvious. It is in such conditions, where one lives with almost all one's relations and possibly with other families whose ancestors have dwelt near one's own for generations, and where one has no other social environment and no other source of cooperation, and where social organization tends towards gerontocracy, that it is inconceivable that the nuclear family should crystallize out and that such a unit should attempt to gain economic advantage over other units.[51]

It is the survival of the extended family system as economic change occurs that helps to sustain high fertility. This survival is rendered more likely by a system of mechanisms that retain the full rigor of the extended family system even through the primitive and traditional societies. After the observations above, it might seem unlikely that primitive society would need such mechanisms, yet they exist throughout sub-Saharan Africa.[52] The reason is society's awareness that conjugal sexual relations can intensify conjugal emotional relationships, and that parent-child emotions can also become of overriding importance. Therefore, African cultures successfully weaken both types of relationship, because communal residence and occupational cooperation would be endangered if men listened to what their wives said was in their mutual interest rather than what their brothers or fathers said, while matrilineal societies would disintegrate if preference were to be shown for sons and daughters over nephews and nieces. In fact (and this is important in terms of demographic transition), relationships between spouses, even in monogamous marriages, are not very strong in traditional Yoruba society and parents do not exclusively focus their attention on their biological children. Even in 1973 only one-third of Yoruba spouses slept in the same room or ever ate together (admittedly indexes of affection regarded as less significant by Yorubas than by outsiders), and fewer still identified the person to whom they felt closest as their spouse, while children were commonly brought up by a number of kinsmen.[53] This should be seen in the context of traditional Yoruba residence in extended family compounds, which persisted even in Ibadan until only a few years ago.

Networks of relatives are important in the primitive society and remain so in the traditional society. They increase the size of the security system and of the cooperating group in less serious situations; they increase the number of close allies in the political contest in the traditional political system in which success is due to the ability to tap more or better communal resources; they increase the number of relatives who can attend family ceremonies and hence magnify one's social importance and sheer consumption pleasure. In rural Yoruba society it is still taken as one of the immutable facts of existence that family numbers, political strength, and affluence are not only interrelated but are one and the same thing. Furthermore, such a base still forms an excellent springboard to success for young aspirants in the modern sector of the economy.[54] There are only two ways of increasing the size of one's network of relatives and they are interrelated: by reproduction and by the marriage of one's children. Data from the second survey in the Nigerian segment of the Changing African Family Project show that 80 percent of all Yoruba still hold that children are either better than wealth or are wealth, while those who maintain that on balance they consume wealth fall to 6 percent in rural areas; 96 percent agree that increasing the number of relatives by means of marriage is a good thing and 83 percent that they can ask relatives by marriage for help with material things or services to a greater extent than they can ask nonrelatives.

But, if this is the way to wealth and power, why do extra children not press more on resources, especially on the supply of food? The question seems to have no meaning in most primitive societies and in traditional society among the Yoruba, even in densely settled rural areas or among urban populations. Part of the answer is that each new pair of hands helps to feed the extra mouth (to paraphrase the kind of proverb that seems to be found widely in Africa and Asia). Part is the nature of the communal economy, where "a man does not acquire more objects than he can use; were he to do so he could only dispose of them by giving them away."[55] Indeed, in such an economy underuse of resources may be far more common than pressure upon them, a situation generalized in Sahlins' rephrasing of Chayanov's rule: "the intensity of labour varies inversely to the relative working capacity of the producing unit [i.e. the household or family]."[56] Lorimer constructed a model for agrarian societies, which apparently showed that, even if belt-tightening was caused in some families by the birth of extra children, it was only to a small extent while the children were young.[57] Less than one-fifth of Yoruba respondents in the second survey of the Nigerian segment of the Changing African Family Project believed that the birth of an extra child would have even an immediate impoverishing effect.

African children certainly work (except perhaps in the transitional society), beginning at age 5–7 years, as they imitate ever more what their elders of the same sex do. It is often difficult, even among adults, to distinguish work completely from way of life. Nevertheless, the traditional patriarch appreciated that work had to be done, that it was often onerous, and that more could be done and others could perhaps take a larger share of the burden if the family were large. C. Edward Hopen relates that he discussed with a Fulani of northern Nigeria whether the Fulani, who supposedly are filled with joy by fathering large families, would have many children in the happy Moslem Heaven that they describe, only to be told: "No, why will we want children? All the work will be done by the servants of Allah."[58] Pierre de Schlippe, reporting on the Zande of south-west Sudan states that, "The prestige of extensive fields and full granaries was to a great extent achieved by family despotism," including "cruel punishments inflicted on wives and children."[59] This is not now 92 the case among either the Zande or the Yoruba, but in rural areas wives and children obey male instructions to work (see below on the question of schoolchildren). Yoruba children work as they have always done helping to provide nonmarket goods and services, as well as helping with market production. That a man benefits economically in such a society by polygyny is now widely affirmed[60]; it is a small step from this to recognizing that he also gains if he has a large number of children.

Traditional Society In Yoruba society the difference between primitive and traditional society is hardly worth making when analyzing demographic trends; but the establishment of the latter was undoubtedly the necessary precursor for fertility change in the transitional society. However, this has not been the case in all traditional societies, many of which evolved slowly over a long period,[61] and indeed the beginning of fertility transition can almost certainly be found in Europe at a time when it was still very largely premodern. State and Church, long before the advent of the Welfare State, were able to provide some assurance that they would intervene to try to prevent unnecessary deaths at times of community disaster—in Europe, with intermissions, since the time of the Ancient World, and over considerable parts of China over the centuries. This may well have weakened the need for the extended family in that the family was no longer the ultimate guarantor of survival. This was probably particularly the case where the authority of the State impinged most strongly and for the longest periods: for instance, in the Ancient World, in Metropolitan Rome, and, especially, in the City of Rome. It is difficult to examine Augustus's marriage laws without concluding both that the extended family at least was under pressure and that a subsequently reversed fertility decline was under way. Rome, as Gibbon so eloquently related, never really died away in Europe: the Church inherited the marriage laws and the attitudes that framed them, as well as responsibility for those in critical circumstances; the manor guaranteed employment and set conditions on access to land, which not only implied that family nucleation (in the economic sense of responsibilities) was well advanced but also reinforced that nucleation (and possibly held fertility in check by preventing early marriage).[62]

Traditional societies with their greater overall organization either introduced or increased the use of money. This, together with their greater guarantees of security to the traveler, expanded trade. With their national legal systems, they were more likely to move toward freehold tenure of land, although the demographic transition theorist should note how recently communal tenure has been important in non-European parts of the world. In fact, in most of sub-Saharan Africa freehold land still exists on only a very limited scale. All these changes had implications for the family.

Wealth Flows in Primitive and Traditional Societies[63]. As analyzed by an outsider from a modern society, children have demonstrable values of several different types in primitive and traditional societies. They do a great deal of work for or with their parents not only when young but usually during adulthood as well; they accept responsibility for the care of parents in old age; they eventually bolster the family's political power and hence give it economic advantages; they ensure the survival of the lineage or family name and in many societies undertake the necessary religious services for the ancestors.

This list, like much value of children research, obscures two very important points.

The first is that such disaggregation is a product of external observation or, even more significantly, of hindsight. In relatively unchanging societies no one sees these separate bonuses conferred by fertility. The society is made of a seamless cloth: children fit into an unintrospective society where they behave as their parents behaved and where their role is to work when young and to care for the old. This is why they may have great trouble in listing any good things (or bad things) about large families when asked by the researcher. Indeed, the respondents' ability to see clearly the separate aspects of children's value shows that the old system is already crumbling and that children's roles are not as certain as before. These roles, then, become important in what is now the transitional society and help to explain the options and decisions of such a society.

The second point is that the value of children to the lineage and ancestors is not really a prop with a strength of its own. Rather, this aspect of the role of children reflects the fact that the other aspects conducive to high fertility are positive as well. When the other props begin to deteriorate in the transitional society, so does the concern for ancestors (often with the help of imported religions, or new interpretations of existing religions, or the spread of secularism).

Nevertheless it is important for the analyst of a society moving toward transition (and this is true of most developing countries) to identify the nature and magnitude of the intergenerational wealth flows in the society. In pretransitional and essentially rural societies, at least six different economic advantages of children to one or both parents can be distinguished: (1) Situational gain is of particular importance to patriarchal males. The obsession with per capita analysis has obscured this type of gain. In Yoruba society there is nothing approaching an equal division of wealth or consumption within the family: there are inequalities by sex, age, and family status. As the number of children beyond infancy grows, and, indeed, as the number of wives and ultimately the number of children-in-law increases, it is inevitable that the person on top of the pyramid controls more resources and has access to more services (as well as enjoying more obvious power), even if per capita income remains static. (2) Children work in the household and on the farm not only producing goods but providing a range of services that adults regard as wholly or partly children's work and that they are loath to do themselves: carrying fuel, water, messages, and goods; sweeping; looking after younger siblings; caring for the animals; weeding the crops; and so on. (3) Adult children usually assist their parents, especially with labor inputs into farms (which frequently increase as the parents age) and with gifts, to a much greater extent than the older generation readily admits or than is spontaneously reported to survey interviewers by either parents or children. (4) Adult children provide particular assistance in making up the family contributions to community festivities and to such family ceremonies as marriages, funerals, and celebrations connected with births. (5) The care of aged parents, who may insist on having their farms, businesses and households propped up as if they were still running them, can be a major undertaking. (6) Parents can invest in training or education of children so as to increase their ability to make returns (although the motive is usually only partly economic and is much more complex than is baldly stated here).

The key issue here, and, I will argue, the fundamental issue in demographic transition, is the *direction and magnitude of intergenerational wealth flows* or the net balance of the two flows—one from parents to children and the other from children to parents—over the period from when people become parents until they die. In premodern society much of the flow is indirect, because of the existence of extended families, clans, and even villages that share in these flows, and because the child's contribution to the parent may be largely by the augmentation of political strength to allow the tapping of a larger share of the communal wealth. The concept of a net balance is still valid, however, even if difficult to measure. It may even be closer to the truth in the older traditional village to speak of the flow being from the younger to the older in the community as a whole with the parent-child relationships in each family playing only a secondary role.

In all primitive societies and nearly all traditional societies the net flow is from child to parent. This is often partly obscured (especially in recent times) from the researcher by the very mechanisms that help to keep it working and to some degree determine the magnitude of the flow. Parents continually point out to children how much they have done for them and how much the children owe (not specifically as money or goods, but more as duty, which in the end means much the same thing). Such protestations may not have been needed in primitive society; to a large extent they help to provide guarantees in a changing and increasingly uncertain society. Three points should be noted. First, such protests are heard most in societies where the wealth flow is still from child to parent; they are much less a feature of a society where the flow has been firmly established toward the child. Second, the protests are not likely to bear much relation to the size of the family and hence to the size or reality of the outlay. Third, the researcher is likely, on hearing the protests and recording them as responses in his questionnaire, to take them as evidence of the economic disadvantages or even irrationality of high fertility. The protests are likely to be supported by details of actual expenditure, without equal concern for details of the returns, and these the researcher may regard as quantified data. There is evidence from one study of a region adjacent to Nigeria that the work of single, adult sons is so important to fathers that they deliberately use their control of bride wealth and marriage ceremonies to space out and postpone sons' marriages so as to organize an even flow of the labor first of unmarried sons and eventually of grandchildren.[64]

There is then a great divide, a point where the compass hesitatingly swings around 180°, separating the earlier situation in which the net flow of wealth is toward parents and in which hence high fertility is rational and the later situation in which the flow is toward children and in which hence no fertility is rational. Why the divide is where it is, and why the compass swings, will be our major concern when investigating the transitional society.[65]

What this means is that before the divide economic rationality dictates unlimitedly high fertility. On the whole, discussion and even survey work in African primitive and traditional society seem to support this. Fertility is limited for all kinds of noneconomic reasons (some of which, however, like child survival, have economic implications). In Yoruba society, the Nigerian segment of the Changing African Family Project found that easily the most important reason is the spacing of births so as to contain infant and early childhood mortality and, hence, to maximize the number of living children. The second most important reason (at least in the past, because it has now been displaced in importance by delayed marriage) has been the cessation of sexual relations by a woman on the birth of the first grandchild so as to avoid the social and psychological tension arising from competing maternal and grandmaternal obligations. Other reasons have been the cessation of sexual relations in some cases when the husband takes another wife or when he moves elsewhere to work or because the woman feels increasingly old or battered by reproduction. Increasingly, fertility is being held in check by postponed age at marriage, which in the case of females already averages several years past puberty; this postponement arises out of competition with education or job opportunities and holds fertility in check because it is accompanied by continence, less sexual activity than in marriage, contraception, or abortion. When the numbers of children become really large, they raise problems of control, noise, and emotional deprivation even in rural societies. The list of noneconomic reasons is quite formidable and 94 is incontrovertible evidence that economic rationality alone is unlikely to determine fertility in any society.

Similarly, after the economic divide, economic rationality dictates zero fertility. This does not happen, and fertility often falls slowly and even irregularly, again for social and psychological reasons—the extent to which alternative roles are available to women, the degree to which child-centeredness renders children relatively expensive, the climate of opinion, and so on.[66] Fertility does not reach zero for reasons that are entirely psychological and social.

It is then necessary to attempt to measure intergenerational wealth flows, an endeavor that is rendered difficult in pretransitional society by a host of problems: much of the flow is not direct but is derived from the extra political power exerted by a man with many children, especially grown-up sons and daughters married into other families; much of the flow is not money but goods and services; some of the flow forms part of family contributions to meet community obligations and does not reach the parents at all; most people have good reason for diffidence about revealing the total flow of wealth, or at least that received. All of these difficulties except the last diminish as the economy becomes more monetized and society more urbanized, and hence transitional society allows easier measurement. Attempts to measure the near-lifetime return on investment in children as well as the outflow from older children were made in Ghana in 1963, and a more comprehensive attempt to examine intergenerational money flows was made in Nigeria's Western State in 1974–75. Both showed clearly that returns from children are substantial.[67]

It is essential to emphasize that the divide is not mechanistically determined by economic conditions. On the contrary it is almost entirely a social phenomenon (except that parent-child net flows of wealth, with the exception of labor and other services such as care for the very young and very old, are hardly possible in subsistence conditions or in the primitive society), and can be reached only when the economy of the nuclear family has been largely isolated from that of the extended family and when a subsequent change of balance has occurred within the nuclear family. The necessity for economic nucleation arises in several ways: the change of economic balance inside the nuclear family is essentially one of emotion and sentiment, which requires emotional nucleation (and other changes of emotional balance within the family) that is incompatible with the extended family economic system, which also needs a parallel system of emotional obligations to work; the change of economic balance in the nuclear family really means that the parents of the family are wholly in charge of their own family economy.

Even if the divide would probably eventually be reached in any urban-industrial society, attitudes and social organization could long delay its advent. Alternatively, a different set of circumstances could bring it on early, even, in fact, before the creation of the modern economy. This seems to be what happened in Western Europe.[68] The feudal system, built on the inherited ruins of the urbanized civilizations of the ancient world, went far toward making a nuclear family economically viable. Doubtless, economic obligations existed to more distant relatives. But these obligations were supported by moral forces and were susceptible to the weakening or reversal of those forces. This seems to have happened with the rise of Protestantism, which put much store on self-sufficiency of all types and on moderation in expenditure and desires. It allowed a man to tell his relatives that they should be more careful in their expenditures, more frugal in their wants, and more foresighted in planning for times of need. More importantly, it allowed him to do this and cautiously refuse to give any (or much) assistance, while retaining his pride and even preaching his practice. Given that the divide had been reached, fertility could be increasingly controlled, even if, at first, mostly by postponed marriage.

In Africa, substantial support for the thesis that emotional nucleation precedes economic nucleation comes from a study in Ghana where Oppong showed among male undergraduates at two universities a significant correlation between the kind of family and kinship obligations the students believed in and the number of children they wanted and an earlier study by the writer that presented evidence on the extent to which urban elite families were emotionally turning in upon themselves.[69]

The Transitional Society An increasing proportion of the Third World population lives in transitional societies that are laboratories for the study of demographic change and lack of change and for determining the origins of demographic transition. "Transitional" here refers to rapid changes in the way of life, especially changes in the impact of children and in the possibilities available to parents for limiting the number of their children.

Nigeria's second largest city, Ibadan, is such a laboratory.[70] Its population is almost 750,000. Although agricultural links are still strong, only one-sixteenth of males report farming as their main occupation; one-third work in nonmanual occupations and another one-third work as soldiers, policemen, or craftsmen, or in similar jobs requiring a degree of training or imported skills and often with an orientation toward the nontraditional world. One-twelfth of women work in nonmanual occupations; but a similar proportion is employed in skilled occupations and over one-half in marketing, often of a somewhat different order from similar employment in rural areas. Three-quarters of the men and one-half the women 95 have been to school; of the latter, one-quarter have experienced some secondary education and almost one-eighth have completed secondary school. More importantly, in terms of the strains on families frequently depicted by demographers, nearly all their children are now receiving some formal education and the majority are proceeding on to secondary schooling. It is rapidly becoming easier to limit fertility if that is the aim. Sexual abstinence has long been widely known as an approved method of avoiding pregnancy. Modern contraceptives are now available from several clinics, a large number of pharmacies, and other retail outlets; in 1973 one-sixth of all women aged 15–59 years had used modern contraception and one-ninth were currently doing so, while the doubling time for the levels of each category of behavior (i.e. the time taken for the proportions behaving in this way to double) had for many years been only four years.

However, fertility (and "ideal family size") appear to have changed little. Significant differentials exist neither between Ibadan and Yoruba rural areas nor within Ibadan society (except that the small group of very highly educated women exhibit lower fertility at younger ages). Nor were contraceptors less fertile than noncontraceptors within Ibadan.[71] The conventional answer in terms of accepted demographic transition theory would be that attitudinal lags prevented parents from fully assessing the new economic situation, that innovation is not fully accepted and implemented at once because the props do not disintegrate at once, and that insufficiently motivated contraceptors are inefficient. None of these propositions appears to hold good in Ibadan, nor are they likely to elsewhere: the parents' assessment of the economic situation appears to be realistic with no time-lag involved; the innovators (as discussed in the section below) do not seem to be aware of their courage in disregarding the props; the contraceptors are mostly doing precisely what they meant to do with the contraceptives.

High fertility remains rational in nonagricultural urban conditions as long as the flow of wealth is predominantly from the younger to the older generation.[72] This is still overwhelmingly the case in Ibadan. The 1974–75 Survey of the Intra-Family Flow of Money and Assistance in Nigeria's Western State surprised us by showing that the return from investment in children is greater for urban than rural residents and is the greatest of all among the city white-collar and professional class. Yet the reason is not far to seek. The urban population working in the modernized economy have both the means and the understanding of the system to keep their children moving up the educational ladder to the top positions in the modern society—positions with high salaries and fringe benefits, as well as control of the levers of power and hence access to opportunities for more wealth, some, but not all, fraudulently obtained. The parents can provide a background suited to continued study, and they know the headmasters and the people who allocate jobs. Perhaps more unexpectedly, the younger generation do not resent the system because they expect to receive wealth in turn from their own, even more successful, children. In fact, as Adepoju has shown, it is the more successful children who would feel most guilt about not sharing their wealth and who visit their parents most often to share it.[73] Furthermore, as the Nigerian Family Study's biographies of the successful clearly demonstrated, a major joy (perhaps the single most important consumption good for the successful) is meeting all family obligations in a more than generous way —in (as they repeatedly said) seeing distant relatives and even non-relatives recognize the donor's success and generosity.

This picture of the success of the urban middle class is but a segment of a wider picture of a whole modernizing society existing in a situation where wealth flows predominantly from the young to the old and where there are marked differentials in earning powers by rural-urban division and by education. The route from the rural area to the job in the modern sector of the economy is almost solely by extended education. Most parents can no longer manage to travel this way, but their children can. To get children far up the educational ladder and into the high-salary positions three stratagems are necessary: relatives outside the nuclear family must be encouraged to help with school fees or with accommodation and subsistence at centers where the right educational institutions exist; older children must help the younger ones in the same way (the sibling chain of educational assistance); and priority must be given to channeling the most assistance, at least early in the establishment of the sibling chain, to the children with the most chance of success—usually the brightest but occasionally those with unusual application, although the distinction is not often made. The first and second stratagems depend on the retention of the system of mutual obligations; the second and third work best with high fertility. The society, like many others in the Third World, believes that the birth of bright and potentially successful children is a matter of capricious fate to which some kind of probability can be assigned (the lucky dip, or lottery, principle) and that large families are likely to have one or more of such children whose existence far outweighs any disadvantages arising from a larger number of less successful siblings. Poor people have limited investment opportunities in such societies, and economic and political caprice can upset what appears to exist, so educational investment in children is thought to be the best investment in both Nigeria and Ghana, and doubtless in many similar societies. The child who has broken through to a job in the modern economy can assist the parents through flows of wealth (sent regularly and at times of crisis, brought on visits, or spent on visiting parents and siblings) or through influencing authorities and manipulating power; the child can bring honor to the parents by visiting them; and can give them access to the joys of the modern world during their visits or final retire- 96 ment to the child's house. Children in urban areas are usually needed to bring earnings into the household, in circumstances where the total income of a poor household is often the sum of many small parts.[71]

Contraception may in the future be used largely to limit family size, but for the time being there is a substantial and increasing demand for contraceptives in Ibadan for other, more pressing reasons: to substitute for female sexual abstinence after birth (in a world where the message of the enjoyment of sexual relations is increasingly being heard); to permit sexual relations during the increasingly long period before marriage in a situation in which pregnancy might destroy the investment in education or dictate a marriage regarded as less than desirable by the family; or to allow safe extramarital sexual relations in a society in which long periods of abstinence, substantial age gaps between spouses, and late marriage of males have meant that discreet relations of this kind have been to a large measure condoned.

More work needs to be done on individuals and families in dire poverty in both traditional and transitional societies. We have investigated a considerable number of cases in West Africa and one point seems clear: they are most likely to be products of an atypically inadequate family structure—often one that has been greatly eroded by mortality and that was vulnerably small in the first place because of accident or subfertility.

Identification of the Primary Forces of Change The transitional nature of Ibadan society also allows the identification of the extent, nature, and cause of fertility transition. This is best done by identifying the innovators. Two methods were employed in the Changing African Family Project. The first was the isolation of all those women in Ibadan (together with their husbands where the marriage was a first, monogamous one with the husband still present) who had indubitably succeeded in demographic innovation: women already over age 40 years with fewer than six live births achieved by intention and any method of restricting fertility.[75] The second was the examination of all women in the three 1973 Nigerian surveys who, regardless of age at the time, had had fewer than six live births, but desired no more and were at the time employing modern contraception to try to ensure this.

The first point established was that there are still very few demographic innovators. Ibadan contains about 62,500 women over age 40 years, but only 438 or 0.7 percent had intentionally and successfully restricted fertility to less than six births.[76] Women of all ages with fewer than six live births and using modern contraception to avoid further pregnancies numbered less than 2,000 in Ibadan, out of about 153,000 women aged 15-49 (or 1.3 percent) or about 128,000 aged 20-44 (about 1.5 percent).

The size of this demographically innovating group (i.e. under 2,000) can be compared with the number of so-called family planning innovators, for in 1973 the number of Ibadan women practicing modern contraception was over 17,000 or almost nine times as many. In the whole of the Western and Lagos States (which include rural areas but which also contain Lagos with its 2 million people and rapidly changing society as well as many other towns), only 0.5 percent of women are currently demographic innovators according to the first Nigerian survey. The 1.5 percent of demographic innovators in Ibadan can also be compared with the number of socioeconomic innovators: 46 percent of women have had schooling, and 15 percent have experienced at least some secondary education; most have their children of school age in full-time education; one-tenth are employed in the modern sector of the economy; one-third of the husbands work in nonmanual occupations, while no more than one-fourth could be said to be employed in the traditional sector of the economy. Clearly, continuing high fertility is not explained by lack of access to or even use of contraception, or by only limited modernization, or by children still maintaining the occupational roles they filled in traditional rural society.

The problem is, then, to study the demographic innovators in depth and to find out how and when they separated themselves from the rest of the community. The quest should be easy. One might infer from demographic transition theory that the decision to do without the props might well be traumatic, and some demographers have wished that they could talk to the eighteenth-century French couples who first daringly decided to innovate. In fact, at first the most frustrating and then the most illuminating discovery was that the demographic innovators are for the most part unaware that they have done anything unusual. After all, contracepting is no longer unusual, particularly in the educational and social groups to which most belong. The use of such contraception to limit family growth just seemed an obvious thing to do in their economic circumstances.

The fundamental question is then: What were the economic circumstances of this group and how did they differ from others who were supporting children at school? The first hint is given by some of their characteristics: demographic innovators compared with noninnovators are 1.6 times as likely to have been to school and 2.7 times as likely to have been to secondary school; they are 2.0 times as likely to have husbands in nonmanual occupations, 4.5 times as likely to be in such occupations themselves, and 2.5 times as likely to have had fathers in such occupations; they are 6.5 times as likely to have all these characteristics—to have fathers and husbands in nonmanual occupations and to be in such occupations themselves and to have had secondary education. Background and education are more important than current occupational experience or indeed any other contemporary circumstance or experience.

These findings could be said to be consonant with the knocking away of the props. However, the Nigerian segment of the Changing African Family Project contained a battery of questions and propositions of a psychosocial kind, relating to phrases taken from Yoruba proverb or song and of a type that could be made in a semi-philosophic way in everyday conversation. The responses showed clearly that what distinguished the demographic innovators from others was not their lack of superstition or their rationalism but their attitudes toward family and children. They have emotionally nucleated their families; they are less concerned with ancestors and extended family relatives than they are with their children, their children's future, and even the future of their children's children. They are more likely to have been "spoilt" themselves in the sense that their parents gave them more emotion and wealth than they expected back, and this is the way they tend, although usually to a greater extent, to treat their own children.[77]

What causes this emotional nucleation of the family whereby parents spend increasingly on their children, while demanding—and receiving—very little in return? Not the urban-industrial society, at least to the extent that it has developed in Ibadan. The majority of the society, even among the elite, is still one where net wealth flows over a lifetime from child to parent. Nor is that majority system buffeted by the institutional requirements of the modern economy; on the contrary it can adapt not only well but profitably to such a society. It might well be able to continue and improve the adaption for decades, or perhaps generations, except for the factor that has already brought about change among the small minority of demographic innovators.

That factor is undoubtedly the import of a different culture; it is Westernization. Just as Western ethnocentricity has bedeviled Third World research and introduced wholly inappropriate attitudes, assumptions and methods, it has in a perversely negative way upset the whole study of "modernization" (i.e. the social changes that seem to precede, accompany, or follow economic development). Western researchers have all too frequently decided to become "objective" or at least "non-self-centered" by achieving the almost incredible feat of omitting transmitted European cultural traditions from the study of modernization; it is like leaving Hellenization out of an examination of social change in fifth century BC Macedonia or leaving Roman social influences out of a treatise on Britain in the second century AD. This may sound like hyperbole, but it is not. In one of the major texts on social change in the Third World, Alex Inkeles and David Smith fleetingly recognized that the difference in their division of the world into that which was modernized and that which was not was almost entirely a contrast between the West and the rest: "With the exception of Japan . . . all the major nations which we can consider modernized are part of the European tradition."[78] Rather than pursue this theme, they decided not to be "arrogant" and instead broke up the Western tradition into components that could be used for measuring not "Westernization" but "modernization."[79] Throughout William Goode's important study, *World Revolution and Family Patterns*, with its investigation of recent family changes in the Arab, Sub-Saharan African, Indian and Chinese worlds, "revolution," except in the discussion of slower growth over a longer period in the West, is a synonym for "Westernization."[80]

Curiously, it is only the well-trained, over-sensitive Western researcher who does not see and hear the obvious. In West Africa, survey respondents (as well as the conversationalist met in the street, the villager in the compound, and the Lagos newspaper) speak continually of adopting European ways—often, in fact, embarrassing the researcher in rural areas by going on to summarize this as "becoming civilized."

How, then, is the European concept of family relationships and obligations imported? The answer is that the import has been on such a massive scale that the slow erosion of traditional family structures is a measure of cultural durability.

Sailors, traders, and slavers may have disrupted some families, but they preached little and few took their examples as a model. However, in the mid-nineteenth century British colonial administration reached Lagos (less than 160 kilometers from Ibadan) and missionaries arrived at Ibadan itself. According to the Changing African Family Project, by 1973 nearly one-half the population of Ibadan were Christian and only 0.5 percent still described themselves as adhering to traditional African beliefs; two-thirds of those who had achieved small families were Christian. Missionaries and their successors have for over a century preached the Western family as the Christian family: monogamy as God's way instead of polygyny; husbands and wives looking after *their* children.[81] Administrators tended to take the same viewpoint, and nearly all Europeans in the developing colonial society advertised the Western family by example and viewpoint.

The mass infusion of European manners, however, has been relatively recent and it has had two interrelated vehicles: mass education and the mass media. Schooling for a very small minority, mostly male, dates back in Ibadan for over a century, but the movement toward some schooling for most children got under way in Yorubaland only in the 1950s. "The family," as taught by the school, is almost entirely the Western family. Textbooks either come from England or are local products modeled on English prototypes. Readers, used in the first years of schooling, are very

98

much concerned with the family and generally tell of a house with a father who goes out to work, a mother who stays home and looks after the children, and the children themselves, who are good and who can expect help and gifts to rain upon them from their two parents. School teachers, even when their own family lives are not fully Westernized, are unlikely to offer non-Western family precepts to their pupils.[82] Researchers have sometimes tried to relate fertility change to the Westernized context of the syllabus,[83] while activists have introduced a "population awareness" ingredient into existing syllabuses; almost certainly such formal ingredients are trivial compared with the inbuilt assumptions of the system and its teachers. Education systems are not easily changed, and are much more likely to be imported intact. In much of the Third World they are essentially a reflection of the modern West, both in their origins and messages, and rarely mirror life in a largely communal and subsistence village. By the mid-1980s many of the women who flooded as youngsters in the late 1950s into the new primary schools may well be faced with the question of calling a halt to family size rather than continuing to reproduce. Then we will discover what impact their schooling had on their families' social and economic structure and what impact this has for their fertility.

Mass media in Nigeria have only had a marked impact since Independence in 1960. Only the newspapers and magazines require the literacy that comes from schooling, but education is likely to lead to the higher income that facilitates the purchase of a radio or a television set or a cinema ticket and to the interest in the nontraditional world that makes these purchases more probable. All cinema films, most television films that portray family life, much of the magazine content, and a considerable proportion of the newspaper feature content are imported, and the models on which they are based are wholly imported from the West. The same message of nuclear family structure is relayed as is imparted by the schools. But another message is also presented in Nigeria: the great importance of sexual relations. This is luridly presented in newspaper and magazine features, news stories, and question and answer sections. Taking a single important example, the emphasis on sex in the widely read *Lagos Weekend* must boost the market for contraceptives, because until recently the main interpretation has been on the excitement of relations outside marriage. But, with the increase in the proportion of educated (and partly Westernized) wives, it is inevitable that the message will be increasingly interpreted to mean also sexual relations within marriage. Such a change, certainly already well under way among the elite, cannot fail to affect the traditional system of family relationships (as has always been recognized in the society) and by strengthening the conjugal emotional bond will tend to nucleate the family, at first emotionally and ultimately economically.

Transition Theory Restated

In general, in societies of every type and stage of development, fertility behavior is rational, and fertility is high or low as a result of economic benefit to individuals, couples, or families in its being so. Whether high or low fertility is economically rational is determined by social conditions: primarily by the direction of the intergenerational *wealth flow*. This *flow* has been from younger to older generations in all traditional societies; and it is apparently impossible (or, at least, examples are unknown) for a reversal of flow—at the great divide—to occur before the family is largely nucleated both emotionally and economically. A fair degree of emotional nucleation is needed for economic nucleation; and considerable amounts of both are required before parents are free to indulge in ever greater expenditures on their children.

Pre-divide populations do not aim at females conceiving as frequently as possible during the full reproductive span, and post-divide populations do not favor childlessness. The reasons are not basically economic; they are social, psychological, and physiological. It is possible, however, that the marginal economic advantage of each additional child in pre-divide society and disadvantage in post-divide society in some circumstances modifies the impact of the noneconomic determinants. Nevertheless, economic analysis on its own can do nothing to predict the timing of the divide and very little to explain the levels of fertility on either side of it—probably the course of fertility in the twentieth century West owes less to the economics of each additional child born than it does to the extent to which parental emotional and expenditure patterns have become focused on the children and the degree to which their society renders such focusing expensive in terms of alternative uses for money, emotion, and time. Similarly, demographic evidence of fertility change may be valueless in terms of deducing movement toward the divide or estimating the probable timing of the reversal of the intergenerational wealth flow; the fertility change may well represent an adjustment of changing social, psychological, or physiological circumstances.[84]

Extreme external factors may influence this pattern. Pre-divide fertility may be restricted in the Kalahari Desert or on Tikopia because of very finite resources; and post-divide fertility was temporarily very high on the American frontier, where the wealth flow to children was relatively insignificant and where there were few alternative sources of labor and even company. The analysis carried out here has been largely based on Africa where access to land has been fairly unrestricted. The position may be somewhat more like Tikopia in densely settled agrarian areas in Asia. However, the little available evidence suggests that it is not, and that even there farming families do not on the whole see the extra birth as impoverishing and do not tighten their belts as the child 99 grows. The explanation may be partly that we are deceived by a static analysis and see the household or family too little in terms of the coming and going of people over time; partly that the extra child does in due course add sufficiently to production; and partly that in the contemporary world the existence of urban employment takes sufficient strain off the need for providing more land.

For reasons that lie deep in its history, the family was increasingly economically nucleated in Western Europe centuries ago; indeed some social groups may have crossed the divide reversing the intergenerational wealth flow as early as the seventeenth century.[85] This phenomenon had two demographic effects: a direct one, namely that Europe's population growth rate was lower than it would otherwise have been once mortality began to decline; an indirect one, in that European culture accepted the nuclear family as the basic unit of society and included a range of values associated with it among exports to other parts of the world.

An emphasis must be placed here on the export of the European social system as well as its economic system. It is as absurd to deny that this is the central feature of our times as to deny the significance of the Hellenization of southwest Asia, the Romanization of the Mediterranean and western Europe, and the Sinoization of much of southeast and central Asia in other periods. The issue is not whether Western social structure is better or even whether it is more suited to modernization; it is merely that the West has been able to export it because of the overwhelming economic strength it derived from the industrial revolution.

From the demographic viewpoint, the most important social exports have been the concept of the predominance of the nuclear family with its strong conjugal tie and the concept of concentrating concern and expenditure on one's children. The latter does not automatically follow from the former, although it is likely to follow continuing Westernization; but the latter must be preceded by the former. There probably is no close relationship in timing between economic modernization and fertility—and, if true, this may be the most important generalization of our time. If another culture had brought economic development, a culture with a much less nucleated family system, industrialization might well have proceeded far beyond its present level in the Third World without reversing the intergenerational flow of wealth. Conversely, in the present situation, family nucleation and the reversal of the intergenerational wealth flow are likely to penetrate deeply into the Third World in the next half century, almost independently of the success of industrialization, and, almost inevitably, they will guarantee slower global population growth.

Several subsidiary points about the export of the Western economic and social systems should be made. First, this export has made both mortality and fertility declines possible in the Third World. Public health measures were acceptable deep in traditional society, and this has been taken as evidence of the reality of the props, which were so constructed as to encourage the desire for low mortality and high fertility. The props are in fact needless: in pre-divide society economic prosperity increased with the number of surviving children—the noneconomic restraints on fertility were more on the number of pregnancies and on the time-span of reproduction than on numbers of survivors. Second, the whole *system* of extended family obligations and the flow of wealth from younger to older generations may be disrupted by political means (China is the clearest example) with exactly the same effect in reducing fertility (although net wealth flows in a commune are probably relatively low, they are almost certainly from the old to the young). Third, the imminence of the reversal of the wealth flow and of declining fertility is usually hidden because of the increased economic benefits from high fertility in the modernizing economy of pre-divide transitional society. And fourth, the attempts to slow associations over time between mortality decline and various economic development indices on one hand and fertility decline on the other are probably valueless; even where there are direct relationships they usually cannot be proved because of the tendency for so many economic and social changes to move together.

A final note should perhaps be added on the more theoretical aspect of population growth in primitive societies. It can be argued that mortality is determined by environment, way of life, and technology, and varies widely among primitive and traditional societies. Yet, demonstrably, population growth rates over long periods have been very low, thus establishing that fertility levels must have approximated mortality levels. One can go further and maintain that this means that mortality levels determined fertility levels, an argument that not only supports the concepts of props but implies that they were subject to strengthening or relaxing until the right level was reached. A more plausible reading of the African tribal situation, however, is that fertility levels were established independently. Where they were above mortality levels, population grew, and the tribe expanded its area through warfare with its neighbors. When expansion was successfully opposed, mortality rates climbed to meet fertility rates: first, because of increasingly unsuccessful warfare and, subsequently, because of growing pressure on limited resources. Where fertility levels were below mortality levels, the tribe died out.

Research Implications

100

If the society is at every stage rational, and economically rational at that, then it can be studied employing economic tools, as long as it is understood that the researchers must accept the society's own ends. Those ends can be researched only by students of society, and their techniques alone

—and not those of economic inquiry—can attempt to predict the approach to the divide where the wealth flow reverses.

First-class fieldwork *on wealth flows in* pre-divide societies is urgently needed, and that research must start with the identification of all possible types of mobile wealth and the development of methods for detecting flows. A good study of a single village would be worth a great deal; defective work on a nation could be dangerously misleading. Cross-sectional studies have some value, but it will be necessary to build up life-cycle models. Specialized investigations might attempt to discover why children do not seem to press on resources in agrarian areas even when these areas are densely settled.

Sociological and anthropological work is needed to define the extent of the true extended families of obligation and to measure the internal wealth flows. It will also be necessary to measure the strength of each obligation bond—the circumstances (and the likelihood of those circumstances occurring) that will bring it into play and the probable volume of the wealth flow under given conditions. The study of the changing family and the measurement of movement toward the social, emotional, and economic nucleation of the conjugal family are important.

A combined social science assault will probably be needed on the circumstances and conditions of the reversal of the wealth flow—and on the time taken for the flow from the older to the younger generation to grow to such an extent that it exerts a real impact on fertility control decisions.

We also need studies that can easily be done in association with family planning action programs. We must find out the real reasons people want contraceptives and the extent to which contraception has anything to do with restricting fertility. Subtle and sympathetic studies in depth of both demographic innovators and contraceptive innovators are essential for action programs.

Finally, we need to know a lot more about the effect on the family of the lessons learned from the media and in school. Much effort has gone into distinguishing the population content of high school lessons but little study has been done on the family structure almost inadvertently taught in the elementary school.

•

The major implication of this analysis is that fertility decline in the Third World is not dependent on the spread of industrialization or even on the rate of economic development. It will of course be affected by such development in that modernization produces more money for schools, for newspapers, and so on; indeed, the whole question of family nucleation cannot arise in the nonmonetized economy. But fertility decline is more likely to precede industrialization and to help bring it about than to follow it.

Notes

1. See Harvey Leibenstein, "An interpretation of the economic theory of fertility: Promising path or blind alley?", *Journal of Economic Literature* 12 (1974): 457–479, for a survey of primarily economic theory that brings out the lack of concern of that theory with the onset of fertility decline.

2. Although many European countries remained in a state of transition for a long period, such conditions are not likely to recur, partly because of the existence of mass schooling. In contemporary transitional societies, families tend to be clearly in one fertility situation or the other, and hence fertility differentials appear; even whole societies are likely to move rather rapidly through the transition as the social and economic calculus changes.

3. Warren S. Thompson, "Population," *The American Journal of Sociology* 34, no. 6 (May 1929): 959–975; C. P. Blacker, "Stages in population growth," *The Eugenics Review* 39, no. 3 (October 1947): 88–101. In his 1946 publication, *Population and Peace in the Pacific* (Chicago: University of Chicago Press), Thompson largely supported the view put forward by Notestein in 1945.

4. Frank W. Notestein, "Population: The long view," in Theodore W. Schultz (ed.), *Food for the World* (Chicago: University of Chicago Press, 1945), p. 39. The term "demographic transition" is first employed on p. 41 of this article, after reference has been made to "demographic evolution" and "transitional growth."

5. Notestein, "Population: The long view," pp. 40–41.

6. Frank W. Notestein, "Economic problems of population change," *8th International Conference of Agricultural Economists, 1953* (London: Oxford University Press, 1953), pp. 15–18.

7. Some social scientists emphasized isolated parts of the argument: In "Population and family planning programs in newly developing countries" [in Ronald Freedman (ed.), *Population: The Vital Revolution* (Chicago: Aldine, 1965)], J. Mayone Stycos emphasized the possibility of advancement in life. W. F. Ogburn and M. F. Nimkoff [*Technology and the Changing Family* (Cambridge, Mass.: Houghton Mifflin, 1955)] stressed the great departure in the city from rural household economy. And Gösta Carlsson in "The decline of fertility: Innovation or adjustment process" (*Population Studies* 20 (November 1966), wrote of the new life style of the urban industrial society and the export of that style. Others, notably Philip M. Hauser and Otis Dudley Duncan, complained that too many explanations had been given, and that some of the supposed causes were material changes, while others were ones of ideas. [See "Demography as a body of knowledge," in Philip M. Hauser and Otis Dudley Duncan, *The Study of Population: An Inventory and Appraisal* (Chicago: University of Chicago Press, 1959), p. 94.]

8. Kingsley Davis, *Human Society* (New York: Macmillan, 1949), pp. 599–600 and "Institutional patterns favoring high fertility in underdeveloped areas," *Eugenics Quarterly* 2, no. 1 (March 1955): 37.

9. George J. Stolnitz, "The demographic transition: From high to low birth rates and death rates," in Freedman (ed.), *Population: The Vital Revolution* (Garden City, New York: Anchor Books, 1964), pp. 33–34.

10. Eva Mueller, "Economic motives for family limitation: A study conducted in Taiwan," *Population Studies* 27, no. 3 (November 1972): 383.

11. William Rich, *Smaller Families Through Social and Economic Development* (Washington, D.C.: Overseas Development Council, 1973), p. 2. Emphasis added.

12. Stephen Enke, "The economic aspects of slowing population growth,"

101

The Economic Journal 76, no. 1 (March 1966): 54.

13. Michael E. Endres, "Underdeveloped countries and the birth control alternative," in *On Defusing the Population Bomb* (Cambridge, Mass.: Halstead Press, 1975), p. 74.

14. Glenn Thomas Trewartha, *The Less Developed Realm: A Geography of Its Population* (New York: Wiley, 1972), pp. 182–183.

15. Notestein, "Population: The long view," pp. 39–41.

16. United Nations, *Concise Report on the World Population Situation in 1970–75* (New York: United Nations, 1974). Quotations from pp. 17, 2, and 14.

17. Eva Mueller, "The economic value of children in peasant agriculture," in Ronald G. Ridker (ed.), *Population and Development* (Baltimore: Johns Hopkins University Press, 1976).

18. Notestein, "Economic problems of population change," pp. 17–18.

19. Davis, "Institutional patterns favoring high fertility"; Davis and Judith Blake, "Social structure and fertility: An analytic framework," *Economic Development and Cultural Change* 4 (April 1956): 211–235. The relationship of fertility to kinship was stressed a year earlier by Lorimer, but he retained the religious and cultural props. See Frank Lorimer, *Culture and Human Fertility: A Study of the Relation of Cultural Conditions to Fertility in Non-Industrial and Transitional Societies* (Paris: UNESCO, 1954).

20. Thomas K. Burch and Murray Gendell, "Extended family structure and fertility: Some conceptual and methodological issues," in Stephen Polgar (ed.), *Culture and Population: A Collection of Current Studies* (Chapel Hill: Carolina Population Center, 1971), pp. 87–104.

21. William J. Goode, "Industrialization and family change," in Bert F. Hoselitz and Wilbert E. Moore (eds.), *Industrialization and Society* (Mouton: UNESCO, 1963), p. 240. It is possible to argue, at least in the Australian context, that they opted not for high fertility but for early female marriage in frontier conditions where women were scarce, but had an important role to play and that high fertility was the unplanned consequence [L. D. Ruzicka and J. C. Caldwell, *The End of Demographic Transition in Australia* (in preparation)].

22. Colin Clark, *Population Growth and Land Use* (London: Macmillan, 1967), pp. 186–187. It is true, however, that some preindustrial peoples appear to have a family structure nucleated not only in residence, but in closeness of relationships; but nevertheless they shared food and animal skins for clothing on a basis going beyond even distant relatives at the same camp. See Nelson H. Graburn, "Traditional economic institutions and the acculturation of the Canadian Eskimos," in George Dalton (ed.), *Studies in Economic Anthropology* (Washington, D.C.: American Anthropological Association, 1971), pp. 107–111. For the argument that the true extended family is largely a product of agrarian societies, see R. L. Blumberg and R. F. Winch, "Societal complexity and familial complexity: Evidence for the curvilinear hypothesis," *American Journal of Sociology* 77, no. 4 (January 1972): 898–920.

23. David M. Heer and Dean O. Smith, "Mortality level and desired family size," Contributed Papers: *Sydney Conference, International Union for the Scientific Study of Population, 21–25 August 1967* (Canberra: 1967), pp. 26–36.

24. Ansley J. Coale and Edgar M. Hoover, *Population Growth and Economic Development in Low-Income Countries: A Case Study of India's Prospects* (Princeton: Princeton University Press, 1958), pp. 11–12. Their summary is essentially based on Notestein, "Economic problems of population change."

25. See Deborah S. Freedman (with Eva Mueller), "Economic data for fertility analysis," *Occasional Paper* no. 11,

World Fertility Survey (August 1974): 7–8.

26. Richard Nelson, "A theory of the low-level equilibrium trap in underdeveloped economies," *American Economic Review* 46, no. 1 (1956): 894–906; Harvey Leibenstein, *Economic Backwardness and Economic Growth: Studies in the Theory of Economic Development* (New York: Wiley, 1957), pp. 170–173.

27. Julian L. Simon, *The Effects of Income on Fertility* (Chapel Hill: Carolina Population Center, 1974), pp. 163–164 and 130.

28. Ronald Freedman, "The sociology of human fertility: A trend report and bibliography," *Current Sociology* 10/11, no. 3 (1961–2): 40 and 48.

29. W. Parker Mauldin, "Fertility studies: Knowledge, attitude and practice," *Studies in Family Planning* 1, no. 7 (June 1965): 6.

30. Bernard Berelson, "KAP studies on fertility," in *Family Planning and Population Programs,* Berelson et al. (eds.) (Chicago: University of Chicago Press, 1966), p. 658.

31. This view is also at odds with Lorimer's attempt to produce a more sophisticated interpretation of fertility levels, a kind of "plural society" way of looking at the world, when he argued that there is not a simple contrast between the low fertility of developed countries and the high fertility of developing countries but that the latter exhibit a wide range of fertility levels reflecting their social and economic structures and presumably their norms. See Frank Lorimer, *Culture and Human Fertility: A Study of the Relation of Cultural Conditions to Fertility in Non-Industrial and Transitional Societies* (Westport, Conn.: Greenwood Press, Inc., 1954). Carr-Saunders had earlier argued that societies might be able to sustain different levels of fertility and that "the evidence . . . shows that the mechanism whereby numbers may be kept near to the desirable level is everywhere present" [A. M. Carr-Saunders, *The Population Problem: A Study in Human Evolution* (Oxford: Clarendon Press, 1922), p. 230].

32. Simon, *The Effects of Income on Fertility*, p. 105. He buttresses this by deciding that fertility behavior is rational, largely on the basis of the Princeton Office of Population Research demonstration that fertility is nearly everywhere substantially lower than it would be if presumably largely uncontrolled Hutterite fertility behavior were prevalent (see, e.g., p. 11).

33. See Emily L. Jones, "The courtesy bias in South-East Asian surveys," *International Science Journal* 15, no. 1 (1963): 70–76.

34. United Nations, Department of Economic and Social Affairs, *Population Bulletin of the United Nations, No. 7—1963, with Special Reference to Conditions and Trends of Fertility in the World* (New York: 1965), p. 143.

35. Among Western European countries, the first declines in fertility paralleled the beginning of marriage postponement perhaps as early as the seventeenth century [see J. Hajnal, "European marriage patterns in perspective," in D. V. Glass and D. E. C. Eversley (eds.), *Population in History: Essays in Historical Demography* (London: Arnold, 1965), pp. 101–143]; and even the restriction of fertility within marriage began a century ago. Therefore, at the threshold itself, many of the post-threshold societies identified in the study exhibited different index values (a range of socioeconomic and demographic indices was calculated) than their current ones—most, indeed, were then within the range of the contemporary pre-threshold societies. (This assumes that the UN studies mean the threshold to be between Groups 3 and 4. There is some tendency to alternate between the concept of a threshold and that of a continuum.)

102

36. Etienne van de Walle and John Knodel, "Demographic transition and fertility decline: The European case," *Contributed Papers: Sydney Conference, In-*

ternational Union for the Scientific Study of Population, 21–25 August 1967, pp. 47–55.

37. Dudley Kirk, "A new demographic transition?" in National Academy of Sciences, *Rapid Population Growth: Consequences and Policy Implications* (Baltimore: Johns Hopkins Press, 1971), pp. 123–147; Frank W. Oechsli and Dudley Kirk, "Modernization and the demographic transition in Latin America and the Caribbean," *Economic Development and Cultural Change* 23, no. 3 (April 1975): 391–419.

38. United Nations, *Concise Report on the World Population Situation*.

39. A. J. Coale, "The demographic transition reconsidered," *International Population Conference, Liege, 1973* Vol. I (Liege: IUSSP, 1973), pp. 62–63.

40. The only two national survey reports published at the time of writing were Rodolfo A. Bulatao, *The Value of Children: A Cross-National Study, II, Philippines* (Honolulu: East-West Population Institute, 1975), and Fred Arnold and James T. Fawcett, *The Value of Children: A Cross-National Study, III, Hawaii* (Honolulu: East-West Population Institute, 1975). The emphasis on a psychological approach is set out in Fred Arnold et al., *The Value of Children: A Cross-National Study, I, Introduction and Comparative Analysis* (Honolulu: East-West Population Institute, 1975), pp. 5–6. The report on the original workshop is also available but it is more economically oriented than the subsequent project [James T. Fawcett (ed.), *The Satisfactions and Costs of Children: Theories, Concepts, Methods* (Honolulu: East-West Center, 1972)]. There have been separate Value of Children projects, such as the survey carried out as part of the 1973 Nigerian segment of the Changing African Family Project to be described later in this paper.

41. The origins of economic anthropology lie, appropriately for the demographic transition theorist, in premodern European history and economic history, but its genesis as a separate field is to be found in German ethnographic studies of the second half of the nineteenth and the first quarter of the twentieth centuries and French studies of the 1920s. In English a literature also began to develop from the 1920s with the work of Malinowski and Firth, leading to the attempt by Herskovits at the end of the 1930s to compile and synthesize what was known. Controversy and new studies have found a renewed vitality in recent years. For a good review of the field, see Raymond Firth (ed.), *Themes in Economic Anthropology* (London: Tavistock, 1967), with references to the syntheses of Wilhelm Koppers in 1915–16 and Max Schmidt in 1920–21 and the later work by Richard Thurnwald. And on more recent studies, see also, for instance, George Dalton, *Studies in Economic Anthropology* (Washington, D.C.: American Anthropological Association, 1971); Marshall Sahlins, *Stone Age Economics* (Chicago: Aldine-Atherton, 1972); and Scarlett Epstein, "The data of economics in anthropological analysis," in A. L. Epstein (ed.), *The Craft of Social Anthropology* (London: Tavistock, 1967).

42. For example, in the Nigerian Segment of the Changing African Family Project, respondents were asked, "If someone offered you a good job for three years, but you could only take it if you put off having a baby for that time, would you be prepared to try to stop having a baby for three years?" Only one-quarter of both women and men replied "No" and that response was not much higher even in remote villages. Very few Nigerians would be offered a good job (defined by most as meaning one in the modern, white-collar sector) and fewer still with a guaranteed period of employment. In practically no case would a woman have to agree not to have a child (and never in the case of men). Should such an extraordinary offer ever be made, of course many might opt for the good job. The fundamental fact about developing economies is that choices of this kind do not exist and,

therefore, a question of this kind is not appropriate.

43. Chukuka Okonjo, "A preliminary medium estimate of the 1962 mid-year population of Nigeria," in John C. Caldwell and Chukuka Okonjo (eds.), *The Population of Tropical Africa* (London: Longman, 1968), pp. 78–96.

44. The argument will not repeat that of the various research papers from that work but will draw on them: primarily, John C. Caldwell, "Fertility and the household economy in Nigeria," *Journal of Comparative Family Studies*, Special Issue, 1976, and "The economic rationality of high fertility," with supporting data from J. C. Caldwell and Pat Caldwell, "The role of marital sexual abstinence in determining fertility: A study of the Yoruba in Nigeria," *Population Studies* 30, no. 2 (July 1977, forthcoming) and "Demographic and contraceptive innovators: A study of transitional African society," *Journal of Biosocial Science* 8, no. 4 (October 1976); "The achieved small family: Early fertility transition in an African city" (in press); J. C. Caldwell and H. Ware, "The evolution of family planning in an African city: Ibadan, Nigeria," *Population Studies* 31, no. 3 (November 1977) (forthcoming); F. O. Okediji et al., "The Changing African Family Project: A report with special reference to the Nigerian segment"; Oshomha Imoagene, *Social Mobility in Emergent Society: A Study of the New Elite in Western Nigeria*, Changing African Family Monograph, No. 2, Department of Demography, Australian National University and Department of Sociology, University of Ibadan, Canberra, 1976.

45. Edward E. Evans-Pritchard, *The Nuer: A Description of Livelihood and Political Institutions of a Nilotic People* (Oxford: Clarendon, 1940), p. 90.

46. C. K. Meek, *Land Law and Custom in the Colonies* (Oxford: Oxford University Press, 1949), p. 16.

47. Sahlins, *Stone Age Economics*, pp. 185–186; see also Mauss (esp. pp. 37–41).

48. See Larissa Lomnitz, "Reciprocity of favors in the urban middle class of Chile," in Dalton, *Studies in Economic Anthropology*, for a description of the extensive system of reciprocity still existing among the Chilean middle class.

49. C. Edward Hopen, *The Pastoral Fulbe Family in Gwandu* (Oxford: International African Institute, Oxford University Press, 1958), pp. 113–114. For examples from other cultures, see, for instance, Clifton R. Wharton, Jr., "Risk, uncertainty and the subsistence farmer: Technological innovation and resistance to change in the context of survival," and Allen W. Johnson, "Security and risk-taking among poor peasants: A Brazilian case," both in Dalton, *Studies in Economic Anthropology*.

50. Peter C. Garlick, *African Traders and Economic Development in Ghana* (Oxford: Clarendon Press, 1971), pp. 110–118.

51. Although this conclusion seems obvious, misinterpretations on this issue abound. Thus, one economist/demographer, Julian Simon, arrived at the right conclusion by making the unfounded assumption that in high-risk situations one cannot afford to worry about the future and, hence, is irresponsibly fertile. George Peter Murdock's analysis of family types from the Yale cross-cultural survey file in *Social Structure* (New York: Macmillan, 1949) confused the whole position by placing emphasis on such simple characteristics as residence units and groupings during movement, so that his successors began to draw parallels between independent, nucleated families found on the one hand among food gatherers and herders and on the other in industrial societies, and to contrast these with the extended family of settled agriculturists. (See, for example, M. F. Nimkoff and Russell Middleton, "Types of family and types of economy," *The American Journal of Sociology* 66, no. 3 (November 1960): 215–225.) Nothing, as we will see, could be less illuminating. The inward-turning nuclear family where obligations exist

103

largely between spouses and toward their nonadult children is a very recent phenomenon almost everywhere except in the West. In spite of Murdock's followers' attempts to show resemblances between Eskimo and Western families, the former in fact have traditionally shared all the food they caught, and it is hardly possible that a nuclear family could improve its diet at the expense of others (see Graburn, cited in note 22 above).

52. Max Gluckman, *Custom and Conflict in Africa* (Oxford: Basil Blackwell, 1955). See also, on the breaking of the emotional bond between a mother and her first-born in Hausa-Fulani society of northern Nigeria, Jean Trevor, "Family change in Sokoto: A traditional Moslem Fulani/Hausa city," in John C. Caldwell (ed.), *Population Growth and Socioeconomic Change in West Africa* (New York: Columbia University Press, 1975).

53. Data from CAFN 2 (Changing African Family Project: Nigerian Segment, Survey 2). On the traditional upbringing of children by a number of kinsmen, see L. P. Mair, "African marriage and social change," in Arthur Phillips (ed.), *Survey of African Marriage and Family Life*, Part 1 (London: Oxford University Press, 1953), p. 2; and, for survey figures showing fewer than half of children in the Ivory Coast to be with their biological parents, see Remi Clignet, *Many Wives Many Powers: Authority and Power in Polygynous Families* (Evanston: Northwestern University Press, 1970), p. 171.

54. The new elite are more likely to have come from larger rural families than from smaller rural families even when allowance is made for the anticipated differential between the two in the number of children supplied to the succeeding generation (see Imoagene).

55. Evans-Pritchard, *The Nuer*, p. 91.

56. Marshall Sahlins, "The intensity of domestic production in primitive societies: Social inflections of the Chayanov Slope," in Dalton, *Studies in Economic Anthropology*, pp. 30–51. S. P. Reyna, "Pronatalism and child labor: Chadian attitudes to birth control and family size," in *Population Growth and Socioeconomic Change in West Africa*, argues that, even in primitive society, the unit with greater working capacity is able to diversify its activities, thus making use of windfall gains and distant economic opportunities and raising its per capita income.

57. Frank Lorimer, "The economics of family formation under difficult conditions," in *Proceedings of the World Population Conference, Belgrade, 30 August-10 September 1965* (New York: United Nations, 1967), II, pp. 92–95.

58. Hopen, p. 124, fn. 1.

59. Pierre de Schlippe, *Shifting Cultivation in Africa: The Zande System of Agriculture* (London: Routledge and Kegan Paul, 1956), p. 235.

60. Boserup, *Woman's Role in Economic Development* (London: Allen and Unwin, 1970), pp. 27–52.

61. An analysis of the startlingly rapid change that occurred in another southern Nigerian society (the Ibos) with the imposition of colonial government found massive development in trading and other economic adaptations, but nothing worth reporting on the family and reproduction. See Simon Ottenberg, "Ibo receptivity to change," in William R. Bascom and Melville J. Herskovits (eds.), *Continuity and Change in African Cultures* (Chicago: University of Chicago Press, 1959), pp. 130–143.

62. Josiah C. Russell, "Demographic values in the Middle Ages," in George F. Mair (ed.), *Studies in Population: Proceedings of the Annual Meeting of the Population Association of America at Princeton, New Jersey, May 1949* (Princeton: Princeton University Press, 1949), pp. 103–107.

63. Strictly speaking, economists describe these "wealth flows" as "income flows," retaining the word "wealth" for a stock rather than a flow. However, most social scientists assume that "income" ex-

cludes the giving of a helping hand in the house and many other items included in this discussion. Hence, it seemed necessary to use a new term.

64. Stephen P. Reyna, "Making do when the rains stop: Adjustments of domestic structure to climate variations among the Barma," *Ethnology* 14, no. 4 (October 1975): 405–417.

65. This will be discussed by the author at greater length in a book on *The Conditions of Fertility Decline* (in preparation) and in a set of studies which he is editing, *The Persistence of High Fertility: Population Prospects in the Third World, Changing African Family Project Monograph no. 3* (Canberra: Department of Demography, Australian National University, 1977).

66. This pattern is being examined toward the end of fertility transition in Australia in a book by L. D. Ruzicka and John C. Caldwell, *The End of Demographic Transition in Australia* (in preparation). Reported child-centeredness in this population is noted in John C. Caldwell, "Family size norms," in Helen Ware (ed.), *Fertility and Family Formation: Australasian Bibliography and Essays, 1972*, Australian Family Formation Project Monograph No. 1 (Canberra: Department of Demography, Australian National University, 1973), pp. A3–A13.

67. See John C. Caldwell, "The erosion of the family," "Extended family obligations and education," and "Fertility and the household economy in Nigeria," especially the section entitled, "An investigation into the inputs into children and the returns from adult children and from education."

68. It is doubtful if this happened in the traditional stage of any other society, although in Japan families did exhibit "rapid segmentation in each generation," partly because of a kind of primogeniture system, and because fertility levels were probably moderate. [See Ezra F. Vogel, "Kinship structure, migration to the city, and modernization," in R. P. Dore (ed.), *Aspects of Social Change in Modern Japan* (Princeton: Princeton University Press, 1967), pp. 91–92; and Irene B. Taeuber, *The Population of Japan* (Princeton: Princeton University Press, 1958), pp. 52–53.] Extended family help was the rule in India and China [Olga Lang, *Chinese Family and Society* (New Haven: Yale University Press, 1946), p. 169], while in northern Nigeria it could be institutionalized into the *gandu* [Polly Hill, *Rural Hausa: A Village and a setting* (New York: Cambridge University Press, 1972)].

69. Christine Oppong, "Attitudes to family type and family size in West Africa: A study of norms among a Ghanaian student population," *International Journal of the Sociology of the Family* 4, no. 2 (1974); Caldwell, *Population Growth and Family Change in Africa*.

70. Data from the Changing African Family Project, mostly from the first survey in the Nigerian segment.

71. John C. Caldwell and Helen Ware, "The evolution of family planning in an African city: Ibadan, Nigeria." Comparisons were made by age at given parities and changes in parity, and age-specific birth rates were also estimated.

72. This is a different argument from that put forward in Davis, "Institutional patterns favoring high fertility," p. 4, where it is argued that the growth of cities at first reinforces high fertility in rural areas by providing greater outlets for agricultural produce.

73. Aderanti Adepoju, "Migration and socioeconomic links between urban migrants and their home communities in Nigeria," *Africa* 44, no. 4 (October 1974): 385–387.

74. On the economic impact of rural-urban migrant children in Ghana, see John C. Caldwell, *African Rural-Urban Migration: The Movement to Ghana's Towns* (New York: Columbia University Press, 1969); for a discussion of the role of children in the family economy in India, see Mahmood Mamdani, *The Myth of Population Control. Family, Caste, and Class in an Indian Village* (New York: Monthly Review Press, 1972); for a description of

104

the situation in a cloth-weaving town in south India, *see* V. P. Pethe, "Attitudes toward family planning: Case studies," in *Demographic Profiles of an Urban Population* (Bombay: Popular Prakashan, 1964), p. 112.

75. CAFN 3, Changing African Family Project: Nigerian Segment, Survey 3. The latitude allowed with regard to contraception, namely the use of any method, including abstinence, to achieve the small family was necessary because the survey was restricted to older women who had relatively little access to modern contraception.

76. This is a very conservative definition of fertility innovation; however, even some of these women might have achieved this fertility by chance and then have rationalized the position [see R. Lesthaege and H. J. Page, "Relating individual fertility to other variables: Common problems and pitfalls," *Seminar on Marriage, Parenthood, and Fertility in West Africa, Lomé, January 3–9, 1976*. International Sociological Association].

77. There has been previous evidence pointing this way from studies in Ghana. The author, drawing on a 1962–64 research program, emphasized that the family-building practices and attitudes of the new urban elite could be understood only in terms of relationships restructured in terms of a fusion of an existing culture with an imported one (*Population Growth and Family Change in Africa*, especially pp. 52–73 and 183–188). Oppong has shown how presumably rational decisions about desired family size among younger members of this elite reflect the type of family situation they desire and will probably try to construct (see "Attitudes to family type and family size in West Africa," a study of university students).

78. Alex Inkeles and David H. Smith, *Becoming Modern: Individual Change in Six Developing Countries* (London: Heinemann, 1974), pp. 17–18.

79. Their index lists "European influence" only to suggest "See Western bias," and, on following this into the text, we find them preparing a defense because "some of our critics would be prepared to argue that the use of the O. M. [Overall Modernity] scale borders on being a social science form of cultural colonialism" (p. 297). Their scale is of little use for the demographer trying to relate modernization to family and fertility change, for two of its important components are "kinship obligations" and "family size" (both measured negatively) (pp. 25–27 and 34).

80. William J. Goode, *World Revolution and Family Patterns* (Glencoe: Free Press, 1963).

81. There has been some revolt against the identification of Christianity with the West, and the African pentecostal churches, which do not preach Western values, have attracted about one-quarter of Ibadan's Christians (CAFN 1).

82. A comprehensive study of teachers, their family lives and problems, their attitudes, and their fertility will be available when Christine Oppong analyzes the Ghanaian Segment of the Changing African Family Project.

83. Norman H. Loewenthal and Abraham S. David, *Social and Economic Correlates of Family Fertility: An Updated Survey of the Evidence* (North Carolina: Research Triangle Institute, 1972), p. 42.

84. See, for example, past fertility declines reported in William Brass et al., *The Demography of Tropical Africa* (Princeton: Princeton University Press, 1968), pp. 178, 181, 346–347, 512–513.

85. Louis Henry, *Anciennes Familles Genevoises: Etude Démographique XVI–XX Siècle*, INED, Travaux et Documents, Cahier no. 26 (Paris: Presses Universitaires de France, 1956); Sigismund Peller, "Births and deaths among Europe's ruling families since 1500," in D. V. Glass and D. E. C. Eversley (eds.), *Population in History: Essays in Historical Demography.*

105

Population and Economic Change in Developing Countries

Edited by Richard A. Easterlin

The University of Chicago Press

Chicago and London

106

2 Toward a More General Economic Model of Fertility Determination: Endogenous Preferences and Natural Fertility

Richard A. Easterlin, Robert A. Pollak, and Michael L. Wachter

This paper develops a general model of marital fertility, from which, with appropriate empirical restrictions, implications are drawn for research and welfare analysis. The model builds to a considerable extent on prior economic research, but it differs from much of the economic literature on fertility in its emphasis on endogenous preferences and natural fertility. We feel there is need for a formal statement of such a model to serve as an alternative to the "Chicago-Columbia" approach that dominates the current work on economics of fertility (e.g., Schultz 1974). Throughout the paper we shall frequently contrast our framework with this approach. The first section outlines our argument; the second presents a formal statement of the model; the third classifies fertility determination into four special subcases; the fourth discusses some of the general research implications; and finally, an outline of the welfare implications of our model is contrasted with those of the Chicago-Columbia approach.

2.1 Overview

In section 2.2 we will present a general model of the determinants of marital fertility and completed family size. The determinants are seen as working through a family's preferences for consumption, children, and fertility regulation, and through four constraints:

106

Richard A. Easterlin, Robert A. Pollak, and Michael L. Wachter are associated with the University of Pennsylvania. Michael L. Wachter is a research associate of the National Bureau of Economic Research.

The research this paper reports was funded by NICHHD grant HD-05427, NSF grants SOC 74-20292 and SOC 75-14750, and Rockefeller Foundation Grant

1. a budget constraint that reflects the limitations implied by the market prices of goods and services, the wage rates of family members, any nonlabor income, and the time at the disposal of household members;

2. the household's technology, which enables it to convert market goods and the time of family members into the basic commodities that are the arguments of its utility function;

3. a "births function" or "fertility production function" that expresses the number of live births as a function of frequency of intercourse, reproductive span of the household, fertility regulation practices, and the commodities, goods, and practices that govern the probability of conception and the nonsusceptible period of the wife;

4. an "infant" mortality function that expresses infant and child mortality through adulthood as a function of such variables as health and nutrition. Subtracting mortality from fertility gives completed family size.

Maximizing the utility function subject to the budget constraint, the household's technology, the births function, and the infant mortality function yields the optimal solution values for the household's decision variables. We denote the optimal solution values for births by b^0 and for completed family size by N^0.

The model is presented (as in the Chicago-Columbia approach) in a single-period decision-making framework. Parents are viewed as making their basic fertility decisions at the beginning of the marriage and then not altering their behavior over their lifetimes. This requires, however, a distinction between results perceived or anticipated when the decisions are made and the actual outcomes. The distinction reflects the fact that families may not correctly perceive the constraints of the maximization problem. The theoretical model of section 2.2 is developed in terms of perceived magnitudes. Conceptually the model can be altered in a straightforward manner to deal with the actual results. This is an important consideration, since the empirical data are usually for the actual rather than the perceived concepts.

In developing a general model of fertility determination, we concentrate on two considerations that we believe are empirically important but that have been largely ignored in much of the economics literature. First, a family's utility function, whose arguments include a vector of commodities and completed family size, is viewed as endogenous to the society in which it lives. In our model this relationship is incorporated

72029. We are grateful for assistance to Debbie Faigen, Stacy Hinck, Neil Weintraub, and Deborah C. K. Wenger, and for helpful comments and suggestions to Ronald Demos Lee, Harvey Leibenstein, Warren Sanderson, Morton Owen Schapiro, and Anne D. Williams. This paper extends previous work by the authors; see Easterlin (1975, 1978) and Wachter (1972*b*).

through an interdependent preference mechanism, which allows for the transmission of aspirations from one family to another and from one generation to another. Past behavior, whether in a "socialization" or a purely intrafamily framework, determines a family's tastes. Second, a family does not always understand or acknowledge the relationship between its fecundity and its consumption decisions because it lacks accurate information concerning the determinants of births and infant mortality. The composition of the consumption bundle has both a direct effect on utility and an indirect effect that operates through the births production function. When household decisions fail to recognize the fecundity effects, in part or in full, there is a problem of "unperceived jointness."

Interdependent preferences and the births and infant mortality functions, with a given level of unperceived jointness, enrich and complicate the optimal solution function. For example, as we shall see, maximizing the family's utility function subject to the appropriate constraints does not yield demand functions for completed family size (or births) as generally construed in the literature.

Needless to say, practical application of such a model is constrained by the limited amount of available data. On fairly reasonable assumptions, however, various subcases of the general model can be distinguished and estimated. Although they are not necessarily realized in pure form, we think these subcases may often constitute useful approximations to reality. In section 2.3 we develop this classification scheme and discuss its empirical relevance.

The concepts of desired fertility and natural fertility play a central role in our classification scheme. The concepts, although prominent in empirical demographic research, have received little attention from economists. We make these concepts an integral part of our analysis. Desired fertility, b^d, is defined as the number of births a family would choose in a situation termed by demographers a "perfect contraceptive society" (Bumpass and Westoff 1970); that is, one in which the family has access to a contraceptive technology with no economic costs and free of preference drawbacks.

Natural fertility, b^n, is defined as the number of births a family believes it would have if it made no deliberate attempt to influence its fertility. Natural fertility is less than the biological maximum and is consistent with the existence of "social controls" on fertility, such as an intercourse taboo. It constitutes uncontrolled fertility only in the sense that the family itself makes no *deliberate* effort to influence its fertility. Contraceptive devices are not utilized, and unperceived jointness or social taboos or both exclude other methods of deliberately influencing family size. If families did perceive the relationship between their consumption pattern and their fecundity, they would alter the former in order to change their fertility. Conscious and deliberate variations by families in the level of their fertility, however, are not compatible with the concept of a *natural* level of fertility. From the standpoint of the family, b^n is constant and is independent of its family-size preferences.

Natural fertility may be greater than, less than, or equal to desired fertility; that is, a family's desires may range from more to fewer children than it thinks it could produce if its fertility were uncontrolled. If the solution for births is below the family's perceived natural fertility ($b^0 < b^n$), then it practices deliberate fertility control. An optimal solution for births above the desired level ($b^0 > b^d$) implies the existence of "excess" or "unwanted" fertility, as the term is used in the demographic literature.

We utilize our generalized fertility model and the associated concepts of natural and desired fertility to classify societies or populations within societies into several categories. The categorization is useful in that it implies restrictions on the coefficients of the variables that appear in the optimal solution functions. Some groups, especially in less developed countries, may be at or close to their natural fertility levels. These groups can be divided into two subcategories. First there are those that lack the motivation to practice fertility regulation because desired fertility is greater than or equal to the optimal solution. Second, there are those, again largely in less developed countries, where the economic costs or preference drawbacks of fertility regulation outweigh the potential gains. In both these cases, the determinants of fertility are largely independent of the preferences for children. "Demand models," with their emphasis on income and substitution effects, are not relevant. Although income might be a significant determinant of completed family size, its influence would be unintended and would work through improved nutrition and health, which would lead to increased fecundity and decreased infant mortality. Demand models tell a different story, typically suggesting that increases in income lead to an increase in the number of children demanded. For natural fertility societies, demand variables—correctly measured and interpreted—are insignificant.

At the other extreme are groups, largely in developed countries, that can be approximated by the perfect contraceptive society. In this case, births and infant mortality technology functions are not quantitatively important determinants of the level of fertility. The properly specified optimal solution function now contains the preference parameters related to children and may reflect endogenous tastes and household technology, including those aspects concerning child-rearing, as well as the budget constraint.

The general fertility model, which includes endogenous tastes and the births production function, has implications for a number of important demographic questions. We have already indicated its significance for specifying the optimal solution function in different societies for different time periods. It is of particular importance that the parameters of this function vary systematically in quantitative importance as one moves along the continuum from less to more developed economies and/or lower to higher socioeconomic classes within a society. Hence, elasticities of births and completed family size with respect to their arguments will vary systematically both across and within societies. We shall also indicate the model's implications for the analysis of the "demographic transition," long-run fertility swings, secular trends in fertility in both less developed countries and developed countries, and the welfare benefits of various types of fertility-control programs in different societies.

At various points we contrast our analysis with the "Chicago-Columbia" approach, by which we mean the line of inquiry exemplified in two recent special issues of the *Journal of Political Economy*, since published as an NBER volume.[1] That there is a distinctive Chicago-Columbia approach to the economics of fertility hardly requires demonstration. In a review of the volume that brings together the *JPE* work, Allen Kelley observes that "the papers are . . . largely of one voice, showing a common perspective to the analysis of economic problems and to a certain extent a mild intolerance of other approaches to viewing the world of social and economic behavior" (Kelley 1976, p. 517). As examples of spokesmen for the approach, one may cite T. W. Schultz (in his editor's introduction to the *JPE* volume), Michael Keeley (in a reply to a critique by Leibenstein), and T. P. Schultz (in several survey articles).[2] We shall draw particularly on the last two in comparing our framework with the Chicago-Columbia approach, because these articles provide valuable general discussions of that viewpoint.[3]

The Chicago-Columbia approach is most simply characterized by what it emphasizes and deemphasizes. Particular emphasis is placed on cost factors and on the opportunity cost of a wife's time; little or no attention is given to taste factors and to the births production function (the latter relates to what T. P. Schultz calls "supply" factors). T. P. Schultz asserts that "cross-sectional studies of individual countries at all levels of development have confirmed the qualitative predictions of this rudimentary demand theory of fertility" (T. P. Schultz 1976, p. 98).[4]

Our main reservation about this line of work is that its deemphasis of tastes and "supply" factors severely limits its empirical relevance. For developed countries the model is of limited application because it ignores preference variables. This is most strikingly illustrated by the failure of the Chicago-Columbia approach to advance an explanation for the recent fertility swing in the United States.[5] For less developed countries, fitting a "demand" model to data for households whose fertility is largely uncontrolled leads to unwarranted inferences about "demand" elasticities. 108

Furthermore, the subordination of taste considerations lends itself to dubious conclusions about economic welfare and public policy. Minimizing the importance of tastes makes it easier to draw unambiguous inferences about the desirability of policies aimed at reducing "unwanted" fertility, but the lack of attention to tastes make such inferences questionable. At the same time, the approach is unlikely to be helpful to those directing family planning programs, who must make choices between attempting to alter preferences (for example, by allocating resources to advertising the benefits of small families) and simply providing contraceptive information or cheaper services. Hence, we believe that both the analysis of fertility behavior and of the welfare effect of government programs requires a more balanced approach, one in which economic research on preferences and natural fertility takes equal place with the usual concerns of the Chicago-Columbia approach.

2.2 The Formal Model

In this section we develop a formal framework for analyzing marital fertility. We begin by summarizing the household production model, which provides the starting point for our analysis. In the three subsequent subsections we modify the household production model to incorporate a number of additional variables related to the determination of marital fertility and completed family size. In section 2.2.2 we incorporate the basic variables related to fertility into the household production model by adding two new "production" relations, a "births production function" and an *infant mortality* or "deaths function," and then describe two extensions of this model, one incorporating unperceived jointness (section 2.2.3) and the other interdependent preferences (section 2.2.4).

By unperceived jointness we mean a situation in which the family does not correctly recognize the relationship between its fecundity and its consumption or life-style decisions. For example, an increase in nonlabor income might cause an unintended and unanticipated increase in births through the following chain of causation: the increase in nonlabor income causes an increase in consumption of health care services or food, which leads to an improvement in health or nutrition; these in turn cause an increase in fecundity. The essence of unperceived jointness is that the decision to devote additional resources to improved health or nutrition rather than shelter or recreation is made without awareness of its implications for fertility.

By interdependent preferences we mean that the family's tastes are influenced by the consumption and family-size decisions of other families. In the "socialization" version of the interdependent preferences model the family's tastes are influenced by the observed behavior of other families in the society, perhaps those in a suitably restricted socioeconomic group. In the "intrafamily" version, a family's aspirations for both commodity consumption and family size are influenced by the consumption and family-size patterns the husband and the wife experienced in childhood and adolescence.

Our model provides a framework for analyzing a number of important aspects of fertility behavior, but it neglects a number of others. First, we deal exclusively with marital fertility. Second, we do not attempt to explain the determination of age at marriage. Third, our analysis is based on a single-period planning model in which the family makes a once-and-for-all decision about its consumption and fertility at the time of marriage. Those aspects of fertility behavior that are best understood in terms of a sequential decision-making model—for example, the timing and spacing of children—are beyond the scope of the analysis, although in principle it could be extended this way. Fourth, our model treats average fertility outcomes as if they were certain to be realized by the "representative family." That is, we ignore both the discreteness of children and the randomness of the births and deaths functions and focus on the mean experience of a group of identical families. In general, randomness and discreteness have implications for the average fertility of families who are not risk-neutral and whose behavior is therefore sensitive to the variance as well as to the mean outcome. Finally, we ignore the fact that children come in two sexes and that parents may have preferences for the sex composition of their families. Such preferences could be incorporated into a sequential model of fertility that recognized the role of uncertainty. In such a model one would expect sex preferences to influence family size, but such preferences cannot be incorporated into a one-period planning model in any straightforward way.[6]

2.2.1 The Household Production Model

In this section we introduce the standard household production model that serves as the basis for our subsequent discussion of fertility. The model is one in which the household purchases "goods" on the market and combines them with time in a "household production function" to produce "commodities."[7] These commodities, rather than the goods, are the arguments of the household's preference ordering; market goods and time are desired not for their own sake, but only as inputs into the production of "commodities." The n market goods are denoted by $X = (x_1, \ldots x_n)$, and the m commodities by $Z = (z_1, \ldots z_m)$, and the time allocation vector by t; the vector t records how much time each family member devotes to market work and to each household activity. Let R 109 denote the household's preference ordering over commodity vectors, and $U(Z)$ the corresponding utility function.[8]

We represent the household's technology by a production set, T. Thus, the "input-output" vector (Z,X,t) belongs to the set T, $(Z,X,t) \epsilon T$, if and only if the commodity collection Z is producible from the goods collection X and the time-allocation vector t. We could distinguish those uses of time devoted to household production activities from those devoted to market work and include only the former as arguments of the household's technology, but it is harmless to include the entire vector, and we do so for notational convenience. Unless explicitly stated to the contrary, constant returns to scale and/or the absence of joint production are *not* assumed. If the household derives satisfaction or dissatisfaction from time spent at various household or market activities, the times devoted to these activities will appear as components of the vector Z as well as the vector t. Technically, this is a case of joint production, since, for example, time devoted to the activity "cooking" is both an input into the production of a "home cooked meal" and is itself one of the outputs of the activity "cooking"—an output that may yield a utility or disutility quite distinct from that associated with eating the meal itself. Because we have not ruled out joint production, there need not be a one-to-one correspondence between activities and commodities.

We let $\bar{t}_h$ denote the total time available to household member h, and t_{hs} the time which he (or she) allocates to activity s. Thus, the family's time constraint may be written as

$$\sum_{s=1}^{S} t_{hs} = \bar{t}_h \qquad h = 1, \ldots, H$$

where S is the total number of market and nonmarket activities and H the number of household members.

We distinguish between the set of market activities (M) and the set of household production or nonmarket activities (T). Thus, if w_h denotes the market wage rate of household member h, his earnings are given by $\sum_{s \epsilon M} w_h t_{hs}$ and the household's total earnings by $\sum_{h=1}^{H} \sum_{s \epsilon M} w_h t_{hs}$. We let μ denote the household's nonlabor income, and write its budget constraint in the form

$$\sum_{k=1}^{n} p_k x_k \leqq \mu + \sum_{h=1}^{H} \sum_{s \epsilon M} w_h t_{hs}.^{9}$$

"Optimal solution values" for the household's decision variables (Z,X,t) are found by maximizing the utility function $U(Z)$ subject to the constraints

$$(Z,X,t) \in T$$

$$\sum_{s=1}^{S} t_{hs} = \bar{t}_h \qquad h = 1, \ldots, H$$

$$\sum_{k=1}^{n} p_k x_k \leqq \mu + \sum_{h=1}^{H} \sum_{s \in M} w_h t_{hs}.$$

The optimal solution values are functions of the values of the variables the household takes as predetermined: goods prices, P; wage rates, w; nonlabor income, μ; and the household's technology, T. The "optimal solution" is optimal with respect to the household's own preferences, not necessarily with respect to any general social welfare criteria. The optimal solution function shows the relationship between the household's decision variables, (Z,X,t), and the parameters it takes as given, $(P,w,\mu;T)$. The optimal solution function is not a demand function in the conventional sense, nor does it treat commodity consumption as a function of commodity shadow prices. Indeed, commodity consumption and the optimal values of the other decision variables are functions of the predetermined variables: goods prices, wage rates, nonlabor income, and the parameters of the household's technology. Commodity shadow prices (i.e., the partial derivatives of the cost function with respect to commodities) have played an unduly prominent role in household production analysis. The difficulty with treating optimal commodity consumption as a function of commodity shadow prices is that commodity shadow prices reflect not only the constraints which the household faces, but also its preferences. With joint production, commodity shadow prices depend on the household's tastes as well as on goods prices and the household's technology. Our model of fertility builds on the household production model, but we reject the "commodity shadow price" version.[10]

2.2.2 The Simple Fertility Model

In this section we extend the standard household production model to include a number of variables related to fertility: children ever born (b), infant and child deaths (d), completed family size (N), frequency of coitus (a), the reproductive span of the household (Λ), the length of time over which each fertility control technique is practiced (θ) and the "intensity" with which each is practiced (τ), and a vector of "practices," such as lactation (l), which affect either the number of children born or their chances of survival.

To simplify the notation we shall not introduce subscripts to distinguish among fertility regulation techniques, but the framework we develop is well suited for discussing choices among techniques. For exam- 110 ple, if one of the available techniques is a contraceptive pill that is to be taken daily, θ might represent the number of months during which it is taken and τ the ratio of the number of days on which the pill is taken to the number on which it is supposed to be taken.[11] Similarly, we do not use subscripts to distinguish among "practices"; formally, we interpret l as a vector, but we shall use "lactation" (i.e., the number of months of lactation following each birth) as an example of the type of variable we have in mind.

These variables are related to each other and the other variables in the household production model by two biological "production" relationships, a births function, B: $b = B(a,Z,X,l,\theta,\tau,\Lambda)$; and a deaths function, D: $d = D(b,Z,X,l)$; and by the identity defining completed family size: $N = b - d$.

The births function depends not only on frequency of coitus (a) and the household's fertility regulation practices (θ and τ), but also on a number of other variables that are likely to vary systematically from one society to another and from one socioeconomic group to another within a society. To take account of the role of factors such as health and nutrition in determining fecundity, we include the household's consumption of commodities (Z) and its purchase of goods (X) as arguments of the births function. Practices such as lactation that influence fecundity are also included; in the case of lactation, a longer interval of lactation following each birth will, ceteris paribus, imply fewer births, since lactation inhibits ovulation. The family's reproductive span, Λ, depends on age at marriage and age at the onset of permanent sterility. The latter is almost certainly endogenously determined by variables such as health and nutrition, but for simplicity we treat the reproductive span as exogenous.

The child and infant mortality function depends not only on the population at risk (b), but also on health and nutrition, which are reflected in the family's consumption of commodities and its purchases of goods. A variety of "practices" that influence deaths are captured by the vector l, although the components of l that influence deaths need not be the same as those that influence births. The length of the lactation interval, however, will appear in the mortality function because—in many societies, at least—a longer lactation interval is associated with lower infant mortality.

Both the births function and the deaths function represent biological "production" relationships. The existence of these biological relationships is quite distinct from the question whether families in either developed or underdeveloped countries perceive these relationships accurately. In this subsection we proceed on the assumption that families are fully aware of the fertility and mortality implications of their behavior. In the next subsection we drop this assumption of perfect knowledge and introduce the concept of unperceived jointness.

Preferences in the simple fertility model are relatively complicated. The utility function includes not only commodities (Z) and completed family size (N), but also infant mortality (d), frequency of intercourse (a), and the contraceptive variables (θ and τ). If frequency of intercourse (a) were not included in the utility function, then abstinence would be the dominant form of fertility regulation, since it is costless and completely effective. Similarly, if there were no disutility associated with infant and child mortality (d), then infanticide might be the second-choice technique, since it also provides an inexpensive and effective method for limiting completed family size. That these techniques do not play a prominent role in most societies clearly reflects preference drawbacks rather than economic costs. But it is not only these extreme techniques of population control that entail preference consequences or drawbacks; the use of any currently available fertility regulation technique (for a particular length of time and with a particular intensity) is likely to entail preference effects that may play an important role in determining not only their time span and intensity of use, but also the number of births and completed family size. We denote the utility function by $U(Z,N,d,a,l,\theta,\tau)$.[12]

The budget constraint must also be modified to allow for the cost of fertility regulation. We assume that its cost is a function of θ and τ alone and denote it by $\rho(\theta,\tau)$.[13]

The optimal solution to the simple fertility model is the set of values of the decision variables $(Z,X,t,b,N,a,l,\theta,\tau)$ that maximize the utility function $U(Z,N,d,a,l,\theta,\tau)$ subject to the constraints

$$(Z,X,t) \in T$$

$$\sum_{s=1}^{S} t_{hs} = \bar{t}_h \qquad h = 1, \ldots, H$$

$$\sum_{k=1}^{n} p_k x_k + \rho(\theta,\tau) \leqq \mu + \sum_{h=1}^{H} \sum_{s \in M} w_h t_{hs}$$

$$b = B(a,Z,X,l,\theta,\tau,\Lambda)$$

$$d = D(b,Z,X,l)$$

$$N = b - d.$$[14]

The optimal solution values are functions of the variables the household takes as given: goods prices, P; wage rates, w; nonlabor income, μ; the household's technology, T; the births function, B; the deaths function, D; the cost function for fertility regulation, ρ; and the family's reproductive span, Λ.[15]

111

2.2.3 Unperceived Jointness

In this section we modify the simple fertility model by postulating that the household is not aware of all the ways its consumption and expenditure patterns affect fecundity and infant mortality. The resulting model is one in which consumption patterns affect realized fertility and mortality, but the effects are unintended. Consider, for example, a family that is not practicing fertility regulation: if it is unaware of the relationship between nutrition and fecundity, it will allocate its expenditure between food and other goods without taking account of the marginal impact of better nutrition on births. An increase in nonlabor income would lead to greater expenditures on food, and, ceteris paribus, through better nutrition to greater fecundity. But the effect on births would be an unintended consequence of the consumption pattern corresponding to a higher income; the household's allocation of expenditure between food and other goods had nothing to do with its desire for children. The family might regard the unintended increase in fertility as a blessing or a curse; in either case, however, the family could "do better" in terms of its own preferences if it knew the true relationship between nutrition and fecundity. If the family were aware of the true relationship it could allow for it in allocating its expenditure between food and other goods: a family that wanted more children would allocate more to food, while one that wanted fewer children would allocate less. We use the phrase "unperceived jointness" to describe a situation in which the family does not recognize the true relationship between its consumption pattern and its fertility or infant mortality.[16] In this section we formalize the concept of unperceived jointness and examine its implications for marital fertility and completed family size.

Although the definition of unperceived jointness does not formally presuppose a situation in which the family makes no deliberate use of fertility control, the concept is useful primarily in such cases. It is especially useful in the first two of the special cases we described briefly in section 2.1: that is, families who fail to recognize that their consumption and expenditure patterns have any effect on their fecundity and who do not employ deliberate fertility control techniques either because they expect to have fewer children than they desire or because, although they expect to have more children than they want, the economic costs and preference drawbacks of fertility regulation outweigh its advantages.

Unperceived jointness is a powerful concept with a wide range of potential applications to topics other than fertility. For example, health or various narrowly defined health states can be treated as commodities that are affected by many household activities, and it is plausible that the effects of many of these activities on health states are unknown to the household. The assumption that the household correctly perceives the relationship between diet and health is an uncomfortable one, especially in cases where the experts do not agree on the nature of the relationship or have learned of it only recently. Unperceived jointness allows us to recognize that health is related to many aspects of a family's consumption pattern and life-style without assuming that the household is fully aware of these relationships. Although we apply the concept of unperceived jointness only to the births production function and the infant mortality function, it could be applied to the household's knowledge of other aspects of its technology. In the fertility context, we could apply it to the length of the reproductive span, Λ, but for simplicity we shall continue to treat the reproductive span as exogenous.

Unperceived jointness does not imply complete ignorance; families may know a great deal about the effects of their behavior on fertility and infant mortality. Indeed, unperceived jointness is consistent with any assumption about the family's knowledge other than the traditional assumption of perfect knowledge. If we view the family's knowledge of the relationships governing fertility and mortality as a point on a continuum from complete ignorance to perfect knowledge, then unperceived jointness is present everywhere except at the polar case of perfect knowledge.[17]

We denote the perceived births function by $\hat{B}(a,Z,X,l,\theta,\tau,\Lambda)$ and the perceived deaths function by $\hat{D}(b,Z,X,l)$. The simplest specification of the perceived deaths function corresponds to the assumption of complete ignorance and is one in which the mortality rate is a constant, independent of the family's consumption and expenditure pattern (Z,X) and its practices (l): $\hat{D}(b,Z,X,l) = \delta b$. For example, the family might believe that one out of every four (or one out of every four hundred) of its children will die, but it does not believe that its behavior can alter this mortality ratio. The family's perception of the mortality rate might depend on the experience of other families in the society, or on that of other families of similar socioeconomic status.

The simplest specification of the perceived births function is also one of complete ignorance, one in which births are independent of the family's decision variables, at least when the family is not practicing any of the fertility control techniques specified by (θ,τ). This implies a perceived births function of the form $\hat{B}(a,Z,X,l,O,O,\Lambda) = \bar{B}$.[18] The family believes that (if it does not practice fertility regulation) its fertility will be exogenously determined and that $\bar{B}$ children will be born to it. The family's estimate of $\bar{B}$ might reflect its observations of the experience of other families in the society or that of other families of similar socioeconomic status.[19]

Completed family size is by definition the difference between births and deaths. In the polar case of complete ignorance, for a family not practicing fertility regulation, perceived completed family size is given by $(1 - \delta)B$. Actual births, deaths, and family size may depart from these expected levels and are determined by the actual births and deaths functions; hence, the actual values of these variables depend on the family's consumption pattern and on other family decision variables such as those grouped together as "practices" and on frequency of intercourse.

Beyond the simplest case of complete ignorance, we must face the question of how families form expectations and adjust the perceived births and deaths functions in the light of experience and observation. Similar problems, however, arise in any version of the household production model unless we assume that the household has perfect knowledge of its technology. If a family recognizes that its consumption and expenditure patterns affect its fertility, it seems plausible that it would systematically revise the perceived births function to reduce any gap between observed and expected fertility corresponding to any consumption pattern. But such revisions are not possible within the confines of a one-period planning model.[20]

With unperceived jointness there are two analogues of the "optimal solution." The first, the "optimal perceived solution," which we denote by the superscript p, is the vector of decision variables obtained by maximizing the utility function subject to the perceived constraints. The optimal perceived solution corresponds to the values of the births and deaths functions the household expects, not the levels that would be generated by substituting the household's consumption and expenditure patterns into the true births and deaths functions. The second, the "realized solution," which we denote by the superscript r, is the vector of decision variables obtained from the optimal perceived solution by replacing the perceived values for births, deaths, and completed family size by the values of these variables that would be generated by the true births and deaths functions, evaluated at the optimal perceived values of the other variables. In the case of goods purchases and the commodity consumption pattern, the realized solution coincides with the optimal perceived solution.[21] But the realized solution for births and deaths typically differs from the optimal perceived solution when there is unperceived jointness.

Formally, the optimal perceived solution to the model with unperceived jointness is the set of values of the decision variables $\{Z,X,t,b,d,N,a,l,\theta,\tau\}$ that maximize the utility function $U(Z,N,d,a,l,\theta,\tau)$ subject to the constraints

$$(Z,X,t) \in T$$

$$\sum_{s=1}^{S} t_{hs} = \bar{t}_h \qquad h = 1, \ldots, H$$

$$\sum_{k=1}^{n} p_k x_k + \rho(\theta,\tau) \leqq \mu + \sum_{h=1}^{H} \sum_{s \in M} w_h t_{hs}$$

$$b = \hat{B}(a,Z,X,l,\theta,\tau,\Lambda)$$

$$d = \hat{D}(b,Z,X,l)$$

$$N = b - d.$$

We denote the optimal perceived solution values by $\{Z^p,X^p,t^p,b^p,d^p,N^p,a^p,l^p,\theta^p,\tau^p\}$; these values are functions of the variables the household takes as given: goods prices, P; wage rates, w; nonlabor income, μ; the household's technology, T; the perceived births function, $\hat{B}$; the perceived deaths function, $\hat{D}$; and the cost function for fertility regulation, ρ.

The realized solution coincides with the optimal perceived solution for the variables (Z,X,t,a,l,θ,τ), but the realized solution for the demographic variables (b,d,N) is determined by substituting the optimal perceived solution values of the other variables into the true births and deaths functions:

$$b^r = B(a^p,Z^p,X^p,l^p,\theta^p,\tau^p,\Lambda)$$

$$d^r = D(b^r,Z^p,X^p,l^p)$$

$$N^r = b^r - d^r.$$

A fulfilled-expectations equilibrium is a solution in which the realized values of b and d coincide with the optimal perceived values. This does not imply that in a fulfilled-expectations equilibrium the family knows the true births and deaths functions—only that its predictions of b and d are correct. It need not know the effects of changes in X or Z on births or deaths, and it may even believe that b and δ are exogenously given.[22] If births and deaths were truly exogenous, then equilibrium could be reached only through the revision of beliefs about the births and deaths functions. When they are not exogenous, the adjustment toward a fulfilled-expectations equilibrium involves both changes in perceptions and changes in behavior that change the realized levels of births and deaths. In equilibrium, observing the fertility and mortality experience of the family will not cause another family holding similar beliefs to revise its perceptions of these functions.[23]

2.2.4 Taste Formation

In this section we introduce endogenous tastes into our model of marital fertility. Within our one-period planning model, interdependent preferences—that is, preferences that depend on the consumption and family-size decision of other families—are the only admissible specification of endogenous tastes.[24] Such preferences are endogenous to the society, but not to the family itself. The model of interdependent preferences is greatly simplified when it is driven by the *past* rather than the current consumption and family-size decisions of other families; because the lagged specification is at least as plausible as the simultaneous one, we shall rely on it exclusively.[25]

Two versions of the lagged interdependent preferences model are of particular interest. The first is a model of "socialization," whose simplest specification is one in which each family's preferences depend on the average consumption and family size of all families in the previous generation or cohort. This specification can be modified by restricting the relevant group of families to those with a particular social or economic status, or by allowing consumption and family-size patterns in the more distant past to play a role in the formation of tastes. The second version, the "intrafamily" model, is one in which each family's preferences are determined by the consumption and family-size patterns the husband and wife experienced during their childhood and adolescence. The intrafamily version predicts that differences in consumption and family-size patterns within a group of families that are similar with respect to such economic variables as wage rates and nonlabor income as well as such variables as education, social status, and religion will be systematically related to differences in the consumption and family-size patterns experienced by husbands and wives during childhood and adolescence. The socialization version does not imply the existence of any systematic differences within such a group of similar families. The intrafamily specification is a version of interdependent preferences rather than habit formation, because tastes depend on the consumption and family-size decisions of the husband's parents and the wife's parents rather than on their own past consumption decisions. Within the context of lagged interdependent preferences, the socialization and the intrafamily specifications are competing hypotheses about whose past consumption and family-size patterns determine a family's tastes.

The socialization model of interdependent preferences is essentially that presented in Pollak (1976*b*) in a traditional demand analysis context. The intrafamily version has been put forward by Easterlin (1968, 1973) and by Wachter (1972*b*, 1975) as an explanation of the recent fertility and labor force participation rate swings in the United States. The intrafamily version is somewhat more complicated than the socialization model because its specification requires a notation that associates each family with the corresponding "parent families" in the previous generation. Rather than introduce such a notation, we shall discuss only the socialization specification.

We formalize interdependent preferences by postulating that each family's tastes depend on "normal levels" of commodity consumption (Z^*) and family size (N^*), and that these normal levels are related to the past consumption and family-size decisions of other families. Normal levels can sometimes be interpreted as "aspiration levels" or "bliss points," sometimes as "necessary" or "subsistence" levels. The essence is that the normal level of a variable is positively related to the family's preference for the commodity in question or for children, so that, ceteris paribus, one would expect an increase in the normal level of a variable to increase its level in the optimal solution.

We shall not specify an explicit form for the family's utility function, but we assume that its tastes for commodities and children are nonnegatively related to the corresponding normal levels.[26] Since the family's preferences depend on normal levels of consumption and family size, we denote its utility function by $U(Z,N,d,a,l,\theta,\tau;Z^*,N^*)$. The semicolon separating the normal levels of Z^* and N^* from the other variables is intended to indicate that this utility function corresponds to a preference ordering over the variables (Z,N,d,a,l,θ,τ), which depends on the value of the normal variables, not to a preference ordering over the extended set of variables $(Z,N,d,a,l,\theta,\tau,Z^*,N^*)$. A preference ordering over the variables (Z,N,d,a,l,θ,τ) that depends on the values of the normal variables is called a "conditional preference ordering," while a preference ordering over the extended set of variables is an "unconditional preference ordering."[27] The distinction between conditional and unconditional preferences plays a crucial role in the analysis of welfare implications in section 2.5.

From a formal standpoint, normal levels are simply parameters that influence preferences in a nonnegative way toward the variables in question. In some cases (e.g., the linear expenditure system) we can interpret them as "necessary" or "subsistence" levels, while in others (e.g., the additive quadratic utility function) they have plausible interpretations as "bliss points," "target levels," or "aspiration levels." However, there are some situations in which neither interpretation is appropriate.[28]

To complete the socialization version of the interdependent preferences model, we must specify how the normal levels N^* and Z^* are determined by past levels. We shall present only the simplest specification, one relating normal levels to average levels in the previous generation. That is, we let $\bar{Z}$ and $\bar{N}$ denote average levels of Z and N in the previous generation and postulate that Z^* and N^* are given by $Z^* = E^Z(\bar{Z})$ and $N^* = E^N(\bar{N})$. The short-run behavior implied by the interdependent preferences model differs from that implied by the model with constant tastes described in section 2.2.3 in that average past consumption and family size, $\bar{N}$ and $\bar{Z}$, operate through the normal levels N^* and Z^* to determine preferences. The analysis of the effects of changes in prices, wages, nonlabor income, or the household's technology presents no new issues. By hypothesis, an increase in a particular $\bar{z}_i$ increases z_i^*, and one would expect this to cause an increase in the

114

optimal solution value of z_i; similarly, an increase in N will increase N^*, and one would expect a corresponding increase in the optimal solution level of N.

The "optimal solution" to the endogenous tastes model is a set of values of the decision variables $(Z,X,t,b,d,N,a,l,\theta,\tau)$ that maximizes the utility function $U(Z,N,d,a,l,\theta,\tau;Z^*,N^*)$ where

$$Z^* = E^Z(\bar{Z})$$
$$N^* = E^N(\bar{N}),$$

subject to the constraints

$$(Z,X,t) \in T$$
$$\sum_{s=1}^{S} t_{hs} = \bar{t}_h \qquad h = 1, \ldots, H$$
$$\sum_{k=1}^{n} p_k x_k + \rho(\theta,\tau) \leqq \mu + \sum_{h=1}^{H} \sum_{s \in M} w_h t_{hs}$$
$$b = B(a,Z,X,l,\theta,\tau,\Lambda)$$
$$d = D(b,Z,X,l)$$
$$N = b - d.$$

The optimal solution values are functions of the values of the variables the household takes as given: goods prices, P; wage rates, w; nonlabor income, μ; the household's technology, T; the births function, B; the deaths function, D; the cost function for fertility regulation, ρ; the variables that determine the normal values for commodities and family size, $\bar{Z}$ and $\bar{N}$; and the family's reproductive span, Λ.

2.3 Special Cases

The framework we have sketched views fertility as the outcome of maximizing a utility function subject to four constraints: the budget constraint, the household's technology, the births production function, and the infant mortality function. Needless to say, empirical application of such a model is constrained by the limited amount of available data. On certain assumptions, however, subcases of the general model can be identified, some of which are much simpler than the complete model. In section 2.3.1 we develop a classification scheme distinguishing four special cases of fertility determination. We show that under certain assumptions the preferences for children may play no role in explaining fertility; under others, the births production function and infant mortality function may play no essential role, and completed family size is governed largely or wholly by the utility function, budget constraint, and household technology—that is, by the variables traditionally emphasized in economic analyses of fertility. Section 2.3.2 takes up the empirical relevance of the proposed classification scheme. The evidence presented suggests that in the typical less developed country, observed fertility for the bulk of the population may depend on the simple model in which preferences for children play no essential role in determining completed family size, but that in developed countries the situation tends increasingly toward one in which preferences play a central role and the births production function and the infant mortality function play no essential role. In section 2.3.3 we develop some implications of this scheme for research on cross-sectional differentials and time-series trends in fertility.

2.3.1 Special Cases of the General Model

Two concepts, prominent in the demographic literature, are of central importance in the development of our classification scheme—desired fertility, b^d, and natural fertility, b^n.

The definition of desired fertility involves another notion common in the literature, that of the "perfect contraceptive society" (Bumpass and Westoff 1970). In terms of our framework this is a situation characterized by a contraceptive technology with no economic costs and free of preference drawbacks (that is, $p(\theta,\tau) = 0$ and $\frac{\partial U}{\partial \theta} = \frac{\partial U}{\partial \tau} = 0$). The term "perfect contraceptive technology" is sometimes used in the literature interchangeably with "perfect contraceptive society." We prefer the latter, because the former conveys the notion of a situation involving only technological aspects of fertility regulation, whereas clearly subjective preferences are also involved.

Desired fertility, b^d, is defined as the number of births a family would choose in a perfect contraceptive society. Desired fertility is independent of the births production function, but it does not depend solely on preferences: other constraints facing the household, its budget constraint, its technology, and its infant mortality function will all influence desired fertility. Although there is no real-world perfect contraceptive society, we believe there are families in a number of societies that effectively approximate such a situation in that further reductions in the economic costs and preference drawbacks of fertility control would have no effect on their fertility behavior.

Natural fertility, b^n, is defined as the number of births a family believes it would have if it made no deliberate attempt to influence its fertility. It is the value of the births function when its arguments are determined without regard to preferences concerning family size.

The natural fertility case thus assumes that unperceived jointness or social taboos or both essentially fix all the arguments of the births pro-

115

duction function except the fertility control variables relating to contraception and induced abortion, which take on zero values. As in the case of the perfect contraceptive society, we do not argue that the pure case of natural fertility is often observed; instead, we argue that it is a useful empirical approximation.

Natural fertility, as we have defined it, is quite different from a biological maximum level of fertility. Natural fertility will almost certainly fall below the maximum value of the births function because a household's consumption pattern involves deficient health or nutrition or because there are social practices (e.g., with regard to nursing children) that restrict the output of children. In addition, natural fertility is influenced by many facets of the family's behavior. For example, the level of natural fertility may reflect such factors as observance of an intercourse taboo, coital frequency, and the consumption bundle chosen by the family. The central point, however, is that natural fertility is independent of the household's preferences for children; although its preferences for commodities and practices play a major role in determining the values of the arguments of the births production function, the relevant decisions are made without regard for their effect on fertility.

Both behavioral and biological factors shape natural fertility. The issue with regard to behavioral influences is whether the behavior is consciously motivated, at least in part, by considerations of its effect on fertility. If it is not, then such behavioral influences are consistent with natural fertility. The question of the household's motivation is clearly important for predicting the likely response to a policy intervention. If, for example, a family has no motivation to regulate its fertility, there is little reason to suppose that establishing a government family planning program would elicit a response from the population.[29]

We also assume for empirical purposes a constant level of infant mortality that is independent of preferences. This is more troublesome than the comparable assumption applied to the births function, because households are likely to realize that they have some control over infant mortality through their expenditures on children's food and health care. Our assumption is that the degree of social control over these variables is great enough that individual family discretion is not empirically important in altering fertility or completed family size. On this assumption, N^n, the natural level of completed family size, as well as b^n, natural fertility, is independent of family preferences.

The concepts of desired and natural fertility can be used to identify four special cases of fertility determination. Natural fertility may be greater than, less than, or equal to desired fertility; that is, a family's desires may range from more to fewer children than it thinks it could produce if its fertility were uncontrolled. An optimal solution for births below the family's perceived natural fertility ($b^o < b^n$) implies a motivation to practice deliberate fertility regulation. An optimal solution for births above the desired level ($b^o > b^d$) implies the existence of "excess" or "unwanted" fertility, as these terms are used in the demographic literature. Using these concepts of deliberate fertility control and excess fertility, households can be classified into four groups on the basis of the determinants of their fertility:

	Excess or Unwanted Fertility		*Practice of Deliberate Fertility Control*	
Group I	No	$b^o \leqq b^d$	No	$b^o = b^n$
Group II	Yes	$b^o > b^d$	No	$b^o = b^n$
Group III	Yes	$b^o > b^d$	Yes	$b^o < b^n$
Group IV	No	$b^o = b^d$	Yes	$b^o < b^n$

For those in group I, natural fertility is less than or equal to desired fertility. In this "deficit fertility" situation there is no motivation to limit fertility, and hence actual fertility will depend on the determinants of natural fertility.

In contrast, all households in groups II, III, and IV have a motivation to regulate fertility because their natural (or "uncontrolled") fertility would result in a greater number of births than desired ($b^n > b^d$). Whether these families practice fertility control depends on the economic costs and preference drawbacks of control relative to its anticipated benefits.

For those in group II the economic costs and preference drawbacks of fertility control outweigh the benefits, and no deliberate control is practiced. For this group, then, as for group I, actual fertility equals natural fertility. Families in group II differ from those in group I, however, in that natural fertility is greater than desired fertility; hence, families in the two groups will respond differently to changes in the economic costs or preference drawbacks of fertility regulation.

Households in both group I and group II do not deliberately attempt to influence their fertility—group I, because of lack of motivation; group II, because the economic costs or preference drawbacks outweigh the incentive. In both cases, therefore, observed fertility behavior corresponds to the natural fertility level and is independent of preferences for births.

For group III the benefits of fertility regulation outweigh the economic costs and preference drawbacks, and these families practice fertility control. But the economic costs and preference drawbacks of fertility control are such that these families have "excess fertility" in the sense that the number of children called for by the optimal solution exceeds desired fertility. Hence, for families in group III: $b^n > b^o > b^d$. For this group, preferences for commodities and children and all of the constraints—the births production function, the infant mortality function, the budget constraint and the household's technology—enter into the determination of actual fertility. The identification of the factors that distinguish families in group III from those in group II is of substantial interest, since these are the factors that push households across the threshold of fertility regulation and cause them to adopt deliberate fertility control.

For group IV the economic costs and preference drawbacks of fertility control are so low relative to motivation for control that the group regulates its fertility to the point where actual births are equal to desired births. Thus, for group IV we have $b^n > b^o = b^d$. Strictly defined, no individual families are in group IV because no perfect contraceptive society exists. However, we believe that a sizable number of families in developed economies are close enough to this case that it provides a useful empirical approximation.[30] For such families the level of fertility is independent of the births production function.

A simple illustration may clarify our classification scheme. Consider a population of households identical in all respects except for nonlabor income and the preference drawbacks of fertility control. Suppose that there is only a single composite commodity, z, one unit of which is produced from each unit of market goods. Consider the indifference map of economic theory with b measured along the horizontal axis and z along the vertical axis. The curve labeled b^d in figure 2.1 is the "expansion path" or "income-consumption curve" of consumer demand theory,

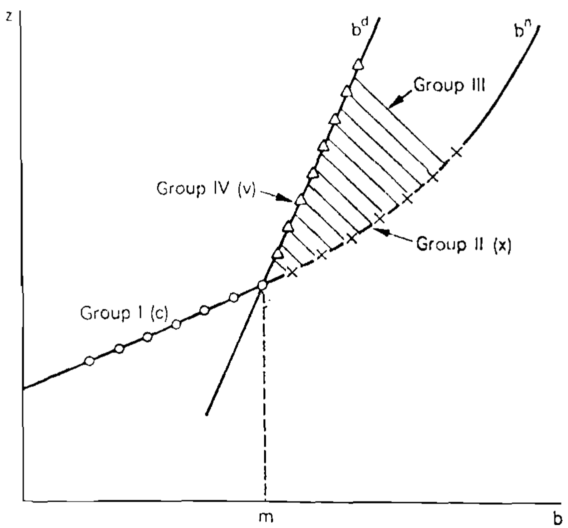


Fig. 2.1 Illustration of four-group classification scheme.

showing the amount of z and b that would be demanded as nonlabor income varied, given tastes and prices. Each point on the curve is obtained from the tangency of an indifference curve and perceived feasible set. One may think of the values of b for various assumed levels of nonlabor income, thus derived, as the "Engel curve" for births—that is, how births would change with the level of nonlabor income. As drawn, the curve shows the number of births increasing with nonlabor income, implying that births are a normal good.

Let us now consider how the ability of households to produce live births might vary with the level of nonlabor income if no deliberate attempt were made to regulate fertility. If nonlabor income were extremely low, then health and nutrition might be so poor that the household would be effectively sterile, that is, $b = 0$. Higher levels of nonlabor income (implying higher input values of health and nutrition in the births production function) would, up to some limit, imply increasing numbers of births. The b^n curve of figure 2.1 traces the path that the potential output of births is assumed to take as nonlabor income grows; that is, it shows how natural fertility might vary with income.

Consider now households whose income is so low as to place them to the left of point m. For these households, desired fertility, b^d, is greater than their reproductive ability, b^n. Hence they would have as many births as they could, and their actual fertility would correspond to natural fertility. These are our group I households; observations for this group would fall along the b^n curve, as shown by the "c" values in the figure.

All households to the right of point m are in an "excess supply" situation; their reproductive potential, b^n, exceeds their desired fertility, b^d. Differences in the actual fertility of these households would arise only from differences in their nonlabor income and the size of the drawbacks they attach to fertility control, because all other factors are assumed to be the same for all households. Households that perceive the drawbacks as so great that they do not practice deliberate fertility control will have observed fertility equal to natural fertility; such households are in our group II, and the observations for this group fall along the b^n curve, as illustrated by the "x" values in figure 2.1. For households who view the preference drawbacks as negligible, observed fertility will equal desired fertility; these households are in our group IV, and the observations for them all along the b^d curve (illustrated by the "v" values in fig. 2.1). Finally, households that practice some deliberate control, but for whom the drawbacks are so great as to result in some excess fertility, will fall in the shaded area between the b^d and b^n curves; these are the group III households.

For some populations the entire b^n curve could lie to the left of the b^d curve, in which case all households would fall in group I, with observed fertility equal to natural fertility. For other populations, the relevant part of the b^n curve might lie wholly to the right of the b^d curve, and if obstacles to fertility control were negligible, all households might fall in group IV, with observed fertility equal to desired fertility. Typically, however, one would expect most societies to include a distribution of households ranging from group I through group IV. When this is so, if one plotted the observations for the population as a whole, one would obtain a scatter of points corresponding to the c, x, and v values as well as some that fall in the shaded area.

If all fertility-determining functions were known, there would be no difficulty in explaining variations among households in observed fertility. When full information is lacking, we suggest using survey response data to divide the population into four groups based on the concepts of natural and desired fertility. For those falling in groups I and II a births production function can be estimated, reflecting the effect of income changes on natural fertility. For those in group IV, it is appropriate to ignore the births production function. For those in group III, we require a model involving preferences for children and fertility control as well as the births production function and infant mortality function.

2.3.2 Some Evidence

Within our general model of fertility determination we have identified four special cases. The empirical evidence currently available, although limited, suggests that it is analytically useful to emphasize these special cases.

The most important evidence relates to the distinction between socially controlled and family controlled fertility (groups I and II versus groups III and IV). For demographers and sociologists, the absence of deliberate family control of fertility is unlikely to raise serious questions, because most noneconomists think of premodern populations as primarily "natural fertility" regimes. Economists, however, are predisposed toward viewing behavior, including reproductive behavior, as a matter of conscious choice. For example, in work on agricultural production behavior in peasant societies, the trend of research has been toward establishing the applicability of rational decision-making models. Thus it has been shown that an unfavorable price movement for a product influences production decisions and causes a contraction in the acreage of the crop planted, in a manner consistent with the predictions of decision-making models (Behrman 1968). By the same token, one might suppose that a decrease in the returns from child labor might lead to curtailment of the output of children.

Reproductive behavior, however, differs from production behavior in an important respect. Babies, since they are a product of sexual intercourse, tend to be produced whether or not they are wanted, whereas rice and wheat do not. Hence, a decision to limit fertility typically requires conscious action, such as abstinence, contraception, or induced abortion. If reproductive behavior is a matter of deliberate choice, then one would expect to find evidence of deliberate practice of fertility control. In fact, the evidence points to the general absence, rather than presence, of deliberate fertility control in less developed countries.

The evidence available is of two types—survey data in which households report on their knowledge and use of fertility control, and census or other data on actual age-specific marital fertility rates.[31] The former come mostly from what are known as "KAP" surveys—surveys of the knowledge of, attitudes toward, and practice of fertility control—which have been conducted in a number of countries since World War II.[32] The other body of data relating to the presence or absence of consciously controlled fertility is quite different; here one draws inferences from the actual fertility behavior of the population, instead of relying on subjective responses. The procedure requires brief exposition, although the essential idea is a simple one.

If no conscious effort were made to limit family size, the age pattern of marital fertility would be governed largely by fecundity and would show a slow decline from ages 20–24 through 35–39, then drop sharply thereafter. If couples were consciously limiting family size, the age pattern of fertility would tend, as age rises, to diverge increasingly in a negative direction from the natural fertility pattern. This is because when a young couple is at the start of the family-building process, there is little incentive to regulate fertility, and hence actual fertility would tend to coincide with natural fertility. However, as a couple ages and family-size grows, approaching or exceeding the desired level, the incentive for deliberate action to restrict family size increases, and correspondingly so does the incentive to adopt deliberate control measures; if such measures are adopted, one would observe the gap between actual fertility and natural fertility increasing over time.

Building on this notion, deriving from Louis Henry's work, that deliberate control involves behavior affecting fertility that is modified as parity increases, Ansley Coale has recently developed a summary index of fertility control, "m," that measures the extent to which an observed age pattern of fertility departs from that believed to characterize a natural fertility regime. An important advantage of the Coale measure (defined in the note to table 2.2) is that it rests on observed behavior, not subjective responses to an interviewer. Moreover, Coale's index would reflect any technique of deliberate control, including abstinence, withdrawal, lactation practices, and induced abortion. In this respect, it avoids two possible problems in the survey data—the possibility that some techniques of deliberate control may have been omitted from the survey, and the possibility of misrepresentation in the responses.[33] A disadvantage is that the Coale measure, unlike survey data, would fail to register a growth in deliberate control if it occurred uniformly at all reproductive ages, for the measure is premised on the assumption that when deliberate control is common, the fertility of older married women is especially low relative to the fertility of younger women. Both a priori reasoning and experience suggest that this is usually true, but the full empirical significance of this qualification remains to be established.[34]

Clearly, one may have doubts about either body of evidence—household surveys of fertility control or inferences from observed fertility behavior—as an adequate indicator of the extent of deliberate fertility control. However, if the results from the two sources are mutually consistent, this would significantly enhance the credibility of each. In fact, as comparison of tables 2.1 and 2.2 shows, this proves true.

Both sets of data show quite limited practice of fertility control in most countries at a premodern or early modern stage of development. In table 2.1 the proportion of the population in such areas reporting that they have ever attempted to control fertility is often about 10% or less. In table 2.2 the index of fertility control, which can range from values about zero (virtual absence of fertility control) to about 2.0, is usually about 0.25 or less.[35] In contrast, in contemporary developed countries, both measures show substantial practice of deliberate control.[36]

The two sets of data also show similar results with regard to rural-urban differences in fertility control. Uniformly, the practice of fertility control is higher in urban than in rural areas.

Finally, for the one case, Taiwan, for which data were readily available for a comparison of the changes over time in the two measures, they show a quite similar trend. In figure 2.2, Coale's index of fertility control is plotted for three dates, 1956, 1965, and 1973. The 1956 value is just about zero, which means that the age pattern of fertility in Taiwan at that date was almost identical with that of a natural fertility regime. Subsequently the index rises sharply to 1965 and again to 1973, implying the rapid adoption and spread of deliberate control. For the last two dates we can compare this pattern with the results of KAP surveys. At each date the survey value is approximately one-half that of "m," and the trend (broken line) lies very close to that shown by the Coale index. Although this is a very simple comparison and the female populations covered by the two measures are not identical, the closeness of the trends indicated by the two measures is encouraging.

Thus we have two bodies of evidence that are mutually confirming—one drawn from personal reports on the knowledge and practice of fertility control and the other based on inferences from observed behavior. It appears that households are, in fact, behaving as they say they are. In most less developed countries, this means that a large proportion of

Table 2.1 **Percentage of Married Women of Reproductive Age Currently Using Contraception, Developed and Developing Countries, Recent Dates**

Country	Date	National	Rural	Urban
		A. Developed Countries		
Australia	1971	66		
Belgium	1966	76	70	77
Czechoslovakia	1970	66	59	69
Denmark	1970	67	64	69
England and Wales	1967	69		
Finland	1971	77	79	76
France	1972	64	59	65
Hungary	1966	64	64	65
Netherlands	1969	59	43	64
Poland	1972	57	51	62
Yugoslavia	1970	59	54	69
USA	1965	64		
		B. Developing Countries		
Africa				
Egypt	1975	21		
Ghana	1976	2		
Kenya	1971	2		
Mauritius	1971	25		
Morocco	1969	1		
Tunisia	1971	12		
Asia				
Bangladesh	1976	5		
India	1969	7–8		
Indonesia	1971	0.5		
Iran	1969	3		
Korea	1972	30		
Malaysia	1969	6		
Nepal	1971	3		
Pakistan	1968–69	6	4	10
Philippines	1972	8		
Taiwan	1971	44		
Thailand	1969–70		13	42
Turkey	1968	35	25	65
Latin America				
Colombia	1974	31	19[a]	35[a]
Costa Rica	1976	34		
Dominican Republic	1976	24		
Ecuador	1974	3		
El Salvador	1976	10		
Guatemala	1974	4		
Haiti	1976	5		
Mexico	1973	13		
Paraguay	1975	10		
Trinidad and Tobago	1971	44		

Source: Nortman (1977), tables 2 and 7.
[a] 1969. Data are for those ever using contraception.

119

Table 2.2 **Coale Index of Fertility Control, *m*, for Females 20–49, Contemporary and Historical Western Populations and Asian Populations by Place of Residence, Specified Dates**

Population	Date	National *m*	Date	Rural *m*	Total Urban or Provincial Urban *m*	Large Cities or Capital *m*
A. Contemporary Western Populations						
Bulgaria	1956	1.67				
Denmark	1963	1.51				
Finland	1960	1.22				
Norway	1960	1.02				
Sweden	1963	1.33				
Australia	1961	1.20				
B. Historical Western Populations						
Bulgaria	1901–5	.02				
Denmark	ca. 1865	.26	ca. 1865	.24	.25	.56
Finland	1871–80	.24				
Norway	1871–75	−.05	1910–11	.31	.86	
Sweden	1751–1800	.23				
6 north French villages			17th–18th cent.	.00		
7 south and central French villages			17th–18th cent.	.02		
14 northwest French villages			17th–18th cent.	.03		
8 Germanic villages			17th–18th cent.	−.00		
1 Swedish village			1745–1820	.13		
Quebec			17th cent.	−.06		
C. Asian Populations						
Japan	1925	.21				
Korea	1961	.03	1960	.01	.36	
Malaysia	1957	.25				
Pakistan	1963–65	−.24				
Philippines	1963–67	.19	1963–67			.69
Sri Lanka	1953	.44				
Taiwan	1956	−.02	1961	.16	.29	.66
Thailand	1960	.11	1968–70	.15	.47	.58
Indonesia			1965–70	.17	.28	
Mysore, India			1952	.26	.16	.56
West Malaysia			1967	.27	.32	.97
China (rural)			1930	.06		
Comilla (Bangladesh)			1963–64	.13		
4 Japanese villages			17th–19th cent.	.18		
Hong Kong			1961			.61
Singapore			1957			.30

households are not deliberately regulating their fertility and thus fall in groups I and II of our classification scheme.

The discussion so far relates to evidence of the division of the population between groups I and II versus groups III and IV. There are no published data that permit the classification of a population into our four groups—a cross-classification based on the practice of fertility control and absence or presence of excess fertility—and hence judgments on empirical importance of the individual groups must be more tentative. However, in the case of Taiwan, for which the availability of unpublished data permit us to derive at least an illustrative distribution of the population among all four groups, the results suggest that all four groups were important in 1965.

The data contain various biases, such as inadequate recall and a tendency after the fact to adjust one's view of desired fertility to realized fertility. Nevertheless, the results shown in table 2.3 may provide a rough idea of orders of magnitude of the four groups at that time. In 1965, the population is divided fairly evenly among the four. For 30% (group I), the number of children was less than or equal to that desired, and consequently, there was no incentive to practice fertility control.[37] Another 26% (group II), although in an excess fertility situation, had not resorted to fertility control, presumably because the costs or preference drawbacks of such control exceeded its benefits. The total of these two groups together amounts to 56%, a majority of the population. The observed fertility behavior of this segment of the population reflects the operation of social controls but not of deliberate family control of fertility; its fertility behavior is independent of preferences for children. The remaining 44% of the population had resorted to deliberate control. This group was almost equally divided between those who had excess fertility (group III), 21%, and those who had not (group IV), 23%.

Source: A, unpublished data kindly provided by Ansley J. Coale; B and C, Knodel (1977, tables 1 and 2), except 1960 data for Korea, which were also provided by Coale.

Note: The index of fertility control, m, is calculated from a comparison of the age-specific marital fertility schedule in the subject population with that presumed to characterize a natural fertility regime according to the following formula:

$$r(a) = M \times n(a) \times e^{m \cdot v(a)},$$

where
a stands for age (from 20–24 through 40–49)
n(a) is an empirically derived natural fertility schedule
r(a) is the marital fertility schedule of the subject populations
M is a scale factor equal to the ratio of r(a) to n(a) at ages 20–24, and
v(a) is an empirically derived function expressing the typical age pattern of voluntary control of fertility.

See Coale and Trussell (1974, p. 187) and Knodel (1977, n. 12).

120

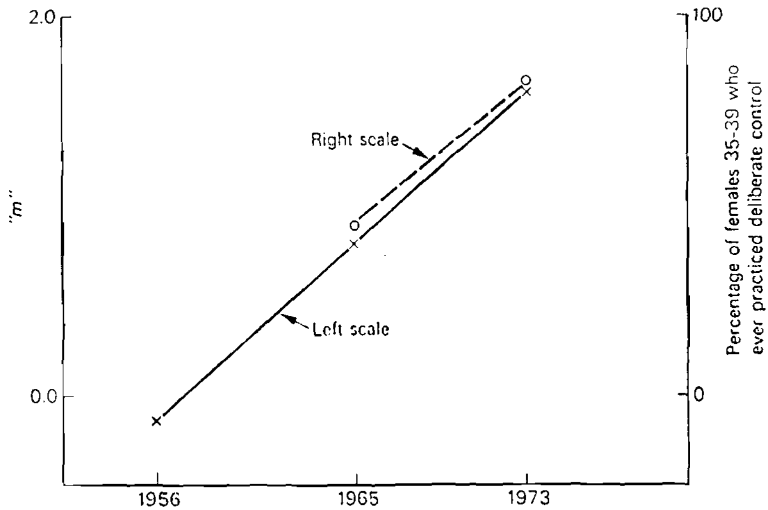


Fig. 2.2 Coale index of fertility control, "m", and survey responses on deliberate control, Taiwan, 1956–73. Index values from Knodel (1977, fig. 5). Survey data from KAP 1 and KAP 4 surveys (cf. table 2.3).

In sum, these data suggest that all four groups in our classification scheme may be empirically important at certain times and places. What stands out most clearly is the importance of social as opposed to deliberate family control of fertility in many less developed countries. Evidence of a pervasive lack of knowledge and use of deliberate fertility control relates especially to rural areas in less developed countries. Since the rural sector typically comprises such a large proportion of a less developed country's population, this means that the behavior over time of the national average of fertility may be largely dominated by the behavior of a natural fertility population. The evidence does not indicate a total absence of deliberate family control of fertility, but it does suggest that such control is usually very limited among premodern and early modern populations.

2.3.3 Research Implications

Our four-group classification scheme, to the extent it has empirical relevance, has important implications for research. First, it implies that for cross-sectional analyses the population should be subdivided based on survey questions regarding deliberate fertility control and excess fertility, and the resulting groups should be analyzed separately. For households in groups I and II, natural fertility models stressing the births production function and ignoring preferences for children are appropriate. For those in groups III and IV, preferences play a crucial role, and we view hypotheses regarding tastes as a high priority area for future research. Our viewpoint is illustrated below in regard to the analysis of fertility differentials and trends.

Fertility Differentials

Our classification scheme suggests that the cross-sectional pattern of fertility differentials by socioeconomic status for a national population is a weighted average of the patterns for the component groups. Pooling the data for all groups is unlikely to lead to correct identification of the underlying relationships. On the other hand, disaggregation of the data into the component groups and separate analysis of each should clarify the basic relationships.

Let us illustrate in terms of a hypothetical example. Suppose that for households in groups I and II, those whose behavior is governed by natural fertility conditions, there would typically be a mild positive relation between socioeconomic status and fertility around a fairly high

Table 2.3 **Percentage Distribution by Practice of Fertility Control and Deficit or Excess Fertility, Wives Aged 35–39 of Unbroken Marriage, Taiwan, 1965**

	Practice of Deliberate Fertility Control					
	Never Practiced			Ever Practiced		
	Total	Desired Family Size		Total	Desired Family Size	
Total		Greater Than or Equal to Actual (group I)	Less Than Actual (group II)		Less Than Actual (group III)	Equal to Actual (group IV)
100	56	30	26	44	21	23

Source: KAP 1 and KAP 4 surveys. We are grateful to Ming-cheng Chang, Ronald Freedman, and Albert Hermalin for making these data available to us and for help in interpreting them. The specific basis for classification is:

1. Excess fertility: the excess for each respondent of living children over the ideal number of living children.
2. Practice of fertility control: based on replies to the question whether the respondent "ever used any birth control."

Because our concern is with marital fertility, the data shown refer to wives, not to all women, and, in order to eliminate the effect on fertility of marital disruption, to wives whose marriage has not been broken. For those who are at an early stage of the reproduction process, one would expect that desired fertility would exceed natural fertility. Hence the data are for women aged 35–39 (the oldest age group available), whose fertility is virtually completed.

121

average level of fertility.[38] Such a pattern might result from the impact of higher income and better health working through the births function. This is illustrated by the groups I–II curve in figure 2.3. Assume further that for households approximating the conditions of a perfect contraceptive society (group IV) there would be a zero correlation between fertility and socioeconomic status around a low average level of fertility —perhaps because systematic variations in taste or cost factors offset a positive income effect. This pattern is suggested by some data on desired family size in the United States. This is shown by the group IV curve in figure 2.3. Finally, let us suppose that for households in group III the pattern of fertility differentials is dominated by differences in the adoption of fertility regulation practices, which are perceived by those in higher socioeconomic status groups to involve fewer preference drawbacks. Then for this group we have a relation between socioeconomic status and fertility given by the group III curve in figure 2.3.[39] The overall pattern of socioeconomic status-fertility differentials would in these circumstances be a weighted average of the patterns for the component groups. By appropriate variations in the underlying assumptions one could produce a great variety of fertility-socioeconomic status patterns.

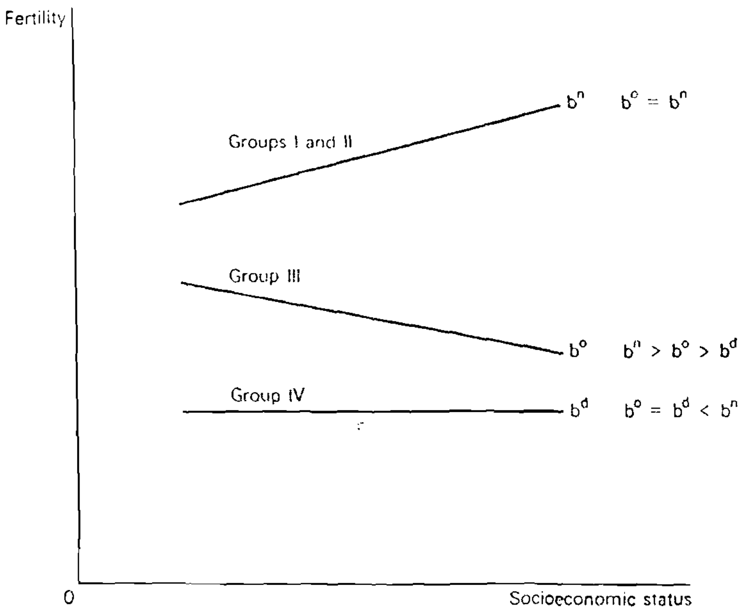


Fig. 2.3 Hypothetical fertility differentials by socioeconomic status.

Fertility Trends

In this area, the most important research questions relate to the demographic transition—the factors behind the shift from high to low fertility during socioeconomic development—and to the long-term outlook for fertility in now-developed countries.

Our classification scheme is compatible with, although it does not require, a view of the demographic transition as a shift from a primarily natural fertility regime (groups I and II) to one eventually largely comprising a "perfect contraceptive society" (group IV), an interpretation consonant with much of the demographic literature. An illustration is provided in figure 2.4, which shows some hypothetical trends during "modernization" (i.e., the transition from a premodern to a modern society) in the levels of natural fertility, desired fertility, and the optimal solution. In the diagram, the process of economic and social modernization is assumed to be correlated with increasing family income and corresponds to a movement to the right along the horizontal axis. The diagram represents only the general nature of the possible relationships during modernization; no implication is intended regarding specific magnitudes.

Natural fertility is assumed to increase during social and economic development, then to level off. This reflects the effect of, for example, increasing income on the health and nutrition of mothers and children, which operates through the births function to increase fertility. Desired family size is assumed to trend downward during the demographic transition, owing perhaps to a change in tastes or to a relative increase in the prices of the inputs required for child-rearing. As drawn, the diagram implies that in premodern societies natural fertility is less than desired fertility (that is, most households are in group I), but the analysis would be essentially the same if most households were in group II. The main point is that initially there is no deliberate practice of fertility regulation.

Consider the trend in the optimal solution implied by our assumptions about natural fertility and desired fertility. At points to the left of m, the optimal solution coincides with natural fertility: parents would have no motivation to practice fertility regulation even if it were free of economic costs and preference drawbacks. At points to the right of m, desired fertility is less than natural fertility, and families would practice fertility regulation if it were available without economic costs or preference drawbacks.

Since fertility control has economic costs and preference drawbacks, we anticipate that initially, as natural fertility edges above desired fertility, the benefits of fertility control would not be great enough to offset 122

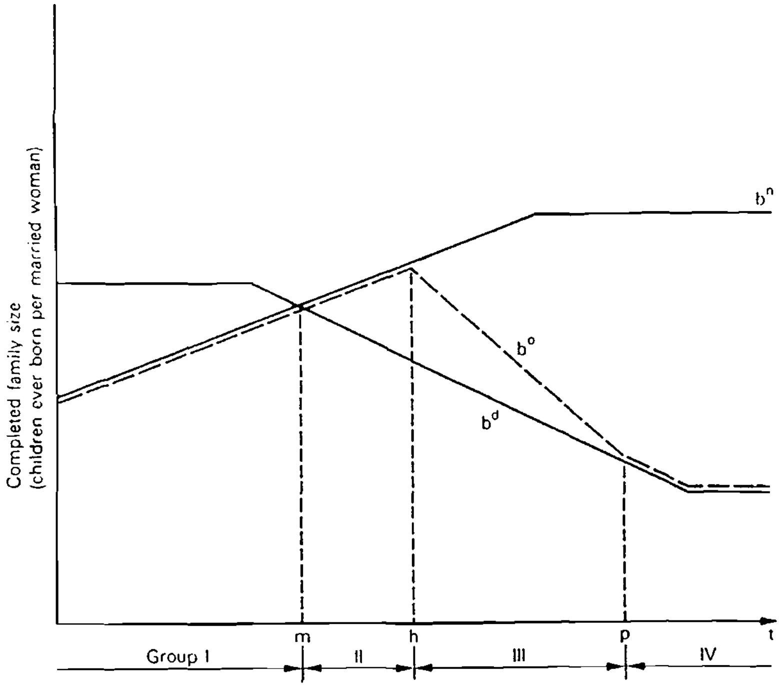


Fig. 2.4 Hypothetical trends in fertility variables associated with economic and social modernization.

its costs, and the optimal solution continues to coincide with natural fertility. As the movement to the right continues, excess fertility and the welfare loss to the family associated with it becomes larger, and a point is reached at which the benefit of fertility control exceeds its economic costs and preference drawbacks. At this point, the "threshold" of fertility regulation, labeled *h* in the diagram, the family adopts fertility control. Beyond *h*, the optimal solution no longer coincides with the natural fertility curve; instead, the optimal solution curve turns downward in the direction of the desired births curve, and, eventually, beyond point *p*, merges with it.

In terms of the previous classification of the population, the situation to the left of point *m* corresponds to a group I situation—no unwanted fertility and no practice of fertility control—as shown in the space beneath the diagram. Between points *m* and *h* there is a group II situation —excess fertility but no practice of fertility control. Between points *h* and *p* there is a group III situation, and to the right of point *p*, a group IV situation. Thus, one might think of the fertility transition as an evolution from a group I situation to a group IV situation. Actually, as we have seen, at any given time the households in a population are distributed among groups I through IV. More realistically, therefore, one might say that in the course of modernization a society gradually shifts from a predominantly group I (and/or group II) situation to a predominantly group IV situation. The main point is that there is a shift in the nature of fertility determination from one where natural fertility factors are largely or wholly dominant and preferences regarding fertility play virtually no role to one in which the influence of natural fertility disappears and conscious choice plays the dominant role.[40]

2.4 Research Implications

Although the four cases identified in section 2.3 depend on special assumptions, a number of other research implications follow from our general model. In this section we discuss the use of preferences as explanatory variables; some issues involving the births and household production functions; and the estimation of elasticities of births and family size with respect to income variables and the wife's wage rate. Our discussion does not depend on the special cases of section 2.3, although it is sometimes explicated more easily by reference to them. To bring out the distinctive features of our approach, we contrast it with the Chicago-Columbia view.

2.4.1 The Role of Preferences as Explanatory Variables

The arguments in section 2.3 regarding the prevalence of unperceived jointness and social taboos suggest that the role of preferences for children in determining observed fertility is smaller in less developed than in more developed countries. Hence, a section devoted to preferences must emphasize issues more relevant to the latter.

In the interplay between preferences and constraints, the Chicago-Columbia approach assumes that systematic variations in fertility are due largely if not entirely to differences in the constraints. Preferences are assumed to be constant across households in cross-sectional studies and over time. Partisans of the Chicago-Columbia approach are generally opposed to investigating taste formation. An example is provided by Michael and Becker (1973, p. 380): "For economists to rest a large part of their theory of choice on differences in tastes is disturbing since they admittedly have no useful theory of the formation of tastes." Going further, Keeley (1975, p. 462) argues that "the household production model lessens the reliance on tastes by incorporating socioeconomic variables in the technology of household production and thus provides a framework where the effects of socioeconomic variables on the shadow prices of home-produced commodities can be systematically analyzed." 123

A related view is that of T. P. Schultz, who asserts that "however conceptualized and quantified, the influence of these 'taste' factors can be properly assessed only *after* the tangible pecuniary returns have also been isolated and taken into account" (Schultz 1976, p. 95, n. 3; emphasis added). But unless one works exclusively with subsamples in which all households have identical tastes, the attempt to isolate "pecuniary returns" must fail. Since one cannot perform controlled experiments, there is no way to segregate the effects of price and income changes from the effects of habits or other types of endogenous taste formation if these phenomena are actually taking place. Hence, estimates of price and income effects obtained from a specification that presupposes fixed tastes are conditional on the maintained hypothesis of fixed tastes. In demand analysis, for example, introduction of habit-formation not only improves R^2 but also changes the estimates of price elasticities, income elasticities, and marginal budget shares.[41]

The notion that economists have an acceptable theory of systematic differences in household technology, but do not have a definitive theory of taste formation, is hardly a justification for excluding or neglecting hypotheses related to tastes.[42] Indeed, economists do not have very satisfactory theories of systematic differences in technologies available to firms in different regions, and the problems posed by household technologies are substantially more difficult because the outputs are often not measured directly. In any case, the testing of alternative models should be left to the empirical arena and not settled by a priori arguments about the proper scope of economics, demography, or other social sciences.

We advocate research on the effect of tastes along with other determinants of variations in fertility. For example, our own work suggests that models incorporating systematic changes in tastes are capable of providing an explanation of recent fertility swings in the United States and some other developed countries.[43] The Chicago-Columbia approach, on the other hand, has been unable to provide a satisfactory explanation of this experience. The Chicago-Columbia approach emphasizes that children are "time intensive in household production." Continuing advances in female education continue to raise the real wage of the wife and consequently the opportunity cost of child-rearing. In this framework, fertility in developed economies should then trend monotonically downward once a regime of low infant mortality is established. Important swings in fertility rates are not anticipated or explained.

In contrast, an endogenous preference mechanism leads to different implications for the long-term outlook in developed countries. Over the long term, it is possible that income growth may be fertility-neutral in the sense that shifts in the budget constraint favoring children are offset by an endogenous preference mechanism functioning as a lagged result of income growth that disfavors children. (This is suggested in fig. 2.4 by the leveling off of the b^d curve, once, say, a low infant mortality regime is established.) These two influences however, may differ in their timing effect and consequently generate longer-term fluctuations in fertility. In this view, recent observations of total fertility rates below the zero population growth level in the United States may be a low point on a long-swings cycle rather than the continuation of a secular decline owing, for example, to increasing female wage rates that increase the cost of child care.[44]

A framework that incorporates research on taste formation along with other social and economic variables will extend the range of empirical problems potentially amenable to treatment. In addition, it may forestall possibly biased results in analyses that omit taste consideration. For if preference variations enter into the determination of fertility variations along with variations in included variables such as wage rates, then failure to take simultaneous account of taste factors will lead to biased estimates of the effects of changes in the included variables.

2.4.2 The Technology of Child-rearing

Central to the Chicago-Columbia approach is the assumption that, as a technological datum, child-rearing is time intensive. Despite its importance, however, there has been little research on the technology of child-rearing. The lack of progress in this area may be due to two factors: the lack of measurable outputs that can be associated with child-rearing and the difficulties of distinguishing between the effects of technology and those of tastes in child-rearing behavior.

We would argue that while children today may be time intensive, especially in their early years, this results from the interplay of tastes, technology, and the budget constraint. There is little evidence to support the notion that the technology requires child-rearing to be time intensive, independent of prices and tastes.

An analysis of the child day-care industry could highlight some of these factors. Has the growth of day-care centers been solely due to a shift in techniques of child care resulting from an increase in the opportunity cost of the wife's time, or has the development of this industry also been due to a change in attitudes toward child-rearing and the role of women? Taking these factors into account, one could develop a number of scenarios in which child-rearing might cease to be a time-intensive activity for the household. For example, if the day-care industry were at an "infant" industry stage (no pun intended), its development, viewed as a dynamic process, could lead to a substantial reduction in the price of child care outside the home. This would lead to a substitution in favor of day-care centers and a decline in household time devoted to child care, but it would be independent of changes in the

124

wife's wage rate. Similarly, a change in preferences in favor of day-care centers as a desirable way to rear children would lead to a decline in household time devoted to child-rearing, although the change might be unrelated to any changes in market prices or wage rates. In these situations it is far from certain that fertility rates would continue to decline, or even remain low, as income and female wages trended upward.

When dealing with the technology of child-rearing, the Chicago-Columbia school tends to treat commonly used socioeconomic variables such as education as proxies for inputs of the household technology. For example, Michael (1973) assumes that differences in education among individuals cause differences in household efficiency but have no effect on tastes. In this framework, the anticipated relationship between education and family size depends upon the relative gains in efficiency in consuming different "commodities" including children. Since the model does not lead to a priori restrictions on the sign of the coefficients, it cannot be separately identified from the preference effects that include education.[45] Is it not possible that education operates through both preferences and technology?

2.4.3 Births Production Function

Another important line of research suggested by our general model is the investigation of the births production function. This is especially important for less developed countries where observed behavior for a substantial share of the population is typically governed by a natural fertility regime. Research on the births production function might be of interest for developed countries, but it would be more difficult because of the smaller proportion of direct observations available.

Suppressing the fertility regulation variables θ and τ, the births production function used in section 2.2 is of the form $b = B(a,Z,X,l,\Lambda)$. Since observations on frequency of intercourse, a, are usually not available, and taking the reproductive period, Λ, as given in an analysis of marital fertility (i.e., focusing on women with the same age at marriage), one would focus on the second and third terms of the relationship, the vectors Z and l. A number of suggestions on relevant empirical variables are available in the literature. Perhaps best known is the line of research stimulated by Rose Frisch (1975), which hypothesizes a positive relation between nutrition and fertility. To date, the most thoroughgoing attempt to test this notion in a context of less developed countries is that of Anderson and McCabe, who find supporting evidence for a biological relationship between nutrition and fertility among younger women in Kinshasa, Zaire (Anderson and McCabe 1975). Another fundamental factor that often receives attention is health conditions. Romaniuc, for example, in a study of data for districts of the Congo, concluded: "the evidence points definitely to the existence of a sequence of events one would logically expect to occur. The birth rate is low because of the high incidence of sterility; the latter is caused by venereal disease, the incidence of which varies with the degree of sexual promiscuity" (Romaniuc 1968, p. 233). In Ceylon, malaria eradication appears to have had a positive effect on births (Barlow 1967). Among other factors (not necessarily independent) that figure in the literature as determinants of natural fertility are lactation practices (Jain and Sun 1972), cultural norms such as intercourse taboos (Leridon 1977), occupational circumstances (e.g., fishing or a nomadic life) (Chen et al. 1974; Romaniuc 1974; Henin 1972), and altitude above sea level (Heer 1967; James 1966).

What is needed in the study of the births production function is an approach analogous to that employed in studying mortality (Auster, Leveson, and Sarachek 1969). Such work would embrace a variety of input variables—nutrition, health, and others of the types just mentioned—that determine fertility as an output. This approach could ascertain the roles of these variables both singly and in conjunction with others as determinants of fertility variations.[46]

As has been noted, in the Chicago-Columbia approach, the prevailing view is that "demand" models of the type used for empirical research on developed countries are, with the addition of a child mortality function, a satisfactory point of departure for empirical research on less developed countries. (Advocates of a demand approach are not confined to those working in the Chicago-Columbia tradition.) When discussed —which is rarely—the need for research relating to the births production function is not emphasized. Thus, T. P. Schultz, in defending the demand model and the disregard of the births production function, argues that:

> If, as seems intuitively reasonable, exogenous differences in a woman's expected fecundity are not usually correlated with exogenous factors affecting her demand for births, proxies for exogenous biological fecundity may be omitted from the demand model of fertility determination and, if this is true, pose no estimation problems [Schultz 1976, p. 93].

We argue that the typical Chicago-Columbia demand model poses a serious misspecification for less developed economies and that Schultz's attempt at salvaging the model does not work.[47] Most troublesome for the demand model advocates is that the "demand" variables, properly interpreted, may simply not be relevant in many less developed countries, except for explaining family size *desires* as distinct from behavior. A prerequisite for a preference model is evidence that the deliberate 125 practice of fertility control is linked to variations in observed fertility. In fact, as discussed above, evidence relating both to age patterns of fertility and to survey responses suggests an absence of deliberate control among much of the population in a typical less developed country.[48] Variables such as income that economists usually interpret as demand variables may, however, be significant in a statistical sense because of their effect on births through the births production function. The variables "fit," but for the wrong reasons. The Schultz argument that the explanatory variables in the demand and supply models are different, so that the demand variable coefficients are uncontaminated, is dubious for the same reason; a woman's fecundity is likely to be positively correlated with income, a factor that also affects her demand for children.

2.4.4 Estimating the Optimal Solution Functions

The empirical fertility literature focuses on estimating the effects on births and completed family size of certain key explanatory variables such as nonlabor income and the wife's wage rate. In the Chicago-Columbia approach, the problem is often treated as one of estimating elasticities (assumed to be constant), and these elasticities are assumed to correspond to the traditional income and substitution effects that would be present in a model in which preferences and the techniques of household production were unchanged.[49]

In our general model, the effects on births and completed family size of changes in nonlabor income and the wife's wage rate operate through at least three distinct lines of causation. Because we are not simply maximizing a utility function subject to a budget constraint, the optimal solution function is not directly comparable to a demand function.

To see this, consider the effect of an increase in the family's nonlabor income. First, the budget constraint shifts out so that the new budget line is parallel to the old one, indicating that the family faces a larger feasible set in the goods space, but that the relative prices of goods are unchanged. This is a possible analogue of the "income effect" of traditional demand theory, although it is not the only possible one. The outward shift in the budget set in the goods space implies a corresponding outward shift in the feasible set in the commodity space. If the household's technology exhibits constant returns to scale, then the new feasible set in the commodity space will be a radial blowup of the old feasible set. Second, if the household's preferences are not homothetic, it may choose to consume commodities in different proportions than before; if this is the case, and if the household's technology exhibits joint production, then the commodity shadow prices at the new equilibrium commodity consumption pattern will differ from those at the old equilibrium, and the change will correspond to a change in the technique of production used by the household. The change in the household's commodity consumption pattern may affect the household's fecundity, even though the household is unaware of the relationship between its consumption pattern and its fertility. Unperceived jointness may operate on the side of infant and child mortality as well as of fecundity, and their net effect will determine completed family size. Third, in the long run, the increase in nonlabor income may lead to an endogenous change in preferences. In the relative income model, for example, an increase in income will in the long run alter aspirations and lead to taste-induced changes in fertility and the participation rate of married women.[50]

In the general model, a change in nonlabor income will affect fertility through all three of the mechanisms described above. Disentangling these separate effects and estimating the underlying structural parameters is a difficult task given the usual limitations on data and our lack of a priori knowledge of technology and tastes. This lack of information, however, does not permit us to assume that induced changes in techniques of production or in tastes are quantitatively unimportant relative to the traditional income and substitution effects. Indeed, we believe the available data suggest that the effects that operate through changing techniques and changing tastes are significant, and that their relative importance varies systematically across societies and across groups within a given society. Our suggestion in section 2.3 that populations be divided into four groups whose fertility behavior should be analyzed separately is our response to this problem. Whether or not one adopts the assumptions necessary to classify populations strictly into these four special cases, the evidence on the practice of deliberate control suggests that for many less developed societies the response of births to changes in nonlabor income and wage rates will operate largely through unperceived effects of consumption patterns on the births and infant mortality functions and that preferences for children will not play a quantitatively important role. For developed economies, on the other hand, response to changes in nonlabor income and wage rates will reflect preferences for children (which may be endogenous) and the household's child-rearing technology as well as the traditional income and substitution effects of the Chicago-Columbia school.

Regressing wage rates and nonlabor income on fertility does not yield sensible estimates of the impact of these variables. The bias would be greater in less developed than in more developed societies, and greater for lower than for higher socioeconomic groups within a society. One way to minimize these biases is to divide the population into groups on the basis of survey responses or income levels and to estimate the parameters separately for each group. In our four special cases, parameters related to preferences for children can be omitted for groups I and II, and those related to the births technology can be omitted for group IV. 126

2.5 Welfare Implications

Our emphasis on unperceived jointness and endogenous tastes requires substantial modifications in the usual formulation of welfare arguments. But even without unperceived jointness and endogenous tastes, our stress on the preference drawbacks as well as the economic costs of fertility regulation as a determinant of fertility control has important implications for evaluation of the welfare effects of policies aimed at reducing "excess" or "unwanted" fertility. These issues are taken up in order below.

2.5.1 Endogenous Preferences

Any type of endogenous tastes considerably complicates welfare analysis. In section 2.2 we discussed a model of interdependent preferences in which a family's tastes depend on the consumption and family-size decisions of others. More specifically, we assumed that each family's preference ordering over its decision variables—vectors of the form $(Z,X,b,d,N,a,l,\theta,\tau)$—depend on the "normal values" of Z and N, which we denote by Z^* and N^*. These normal values might depend on the consumption and family-size patterns it observes in the surrounding society (in the socialization version) or on the levels of Z and N experienced by the husband and wife in childhood (in the intrafamily version). In section 2.2 we described the preference ordering over the decision variables as "conditional" on the values of Z^* and N^* and indicated this by writing the utility function as $U(Z,N,d,a,l,\theta,\tau;Z^*,N^*)$, with the semicolon separating the normal levels Z^* and N^* from the other variables. We did this to distinguish between an "unconditional" preference ordering over the extended set of variables $(Z,N,d,a,l,\theta,\tau,Z^*,N^*)$ and a "conditional" preference ordering over the decision variables (Z,N,d,a,l,θ,τ) that depends on the levels of the normal variables Z^* and N^*.

A conditional preference ordering captures the notion that families with different consumption and family-size experiences may have different tastes and may make different decisions, but it does not permit us to compare situations that correspond to different normal values, Z^* and N^*.[51] Such comparisons must be based on the unconditional preference ordering over the extended set of variables $(Z,X,b,d,a,l,\theta,\tau,Z^*,N^*)$. To see this, suppose that preferences for children are determined by the number of children in the wife's family, independent of the commodity consumption pattern of that family, in the following very simple way: regardless of other considerations, the family attempts to have the same number of children as were present in the family in which the wife grew up, and it is unwilling to trade off children against commodities in attempting to accomplish this. In such a world, the size of the family in which the wife grew up uniquely determines the number of children she will have, but there is no way to use this information to compare the welfare level of women with one sibling with that of women with two siblings. Notice that the woman with one sibling is observed to choose two children rather than three, while the woman with two siblings is observed to choose three children rather than two.[52]

The welfare implications of a model of interdependent preferences must be derived from the unconditional preference ordering, but these preferences are not revealed by the family's choices of the decision variables. Thus the conceptual basis for welfare evaluation in such a model must be quite different from the "revealed preference" approach usually employed by economists. We see two possible bases for welfare evaluation with endogenous preferences. The first is based on direct comparisons of the well-being of different families as reflected by their responses to survey questions which ask them directly about their "happiness" or "well-being."[53] The second approach relies on a different type of interpersonal comparison. Sen (1973, p. 14) discusses this approach:

> If I say "I would prefer to be person *A* rather than person *B* in this situation," I am indulging in an interpersonal comparison. While we do not really have the opportunity (or perhaps the misfortune, as the case may be) of in fact becoming *A* or *B*, we can think quite systematically about such a choice, and indeed we seem to make such comparisons frequently.
>
> Representing (x,i) as being individual i (with his tastes and mental qualities as well) in social state x, a preference relation R defined over all such pairs provides an "ordinal" structure of interpersonal comparisons.

In the case of interdependent preferences, one would ask a family whether it would rather be in the position of family A, which experienced the consumption–family-size pattern α during adolescence, or of family B which experienced the consumption–family-size pattern β during adolescence, but it is possible that the family's choice between the alternatives (A,α) and (B,β) will depend on its own consumption–family-size experiences during adolescence. If individuals are unable to abstract from their own backgrounds and upbringing in making choices of this type, there is little chance of extracting an unconditional preference ordering from responses to such questions. If this is the case, welfare evaluations must rest on direct comparisons of "happiness" or "well-being" reflected either by responses to survey questions or by an appeal to general (and often questionable) assumptions about "human nature."[54]

127

2.5.2 Unperceived Jointness

Even without interdependent preferences, unperceived jointness complicates welfare evaluation. Economists are accustomed to asserting that if a family chooses alternative A when it could have chosen B, then A is at least as good as B according to the family's preferences.[55] The analogous correct version of this assertion is the following: If the family intends to choose A when it believes it could have chosen B, then A is at least as good as B (according to the family's preferences). Unperceived jointness breaks the automatic link between observed consumption and intended choice, since the family that intends to choose A may be observed with A' (e.g., a larger number of children, because of the effect of better nutrition on fertility). Similarly, with unperceived jointness the household's perception of the set of feasible alternatives may be quite different from the true feasible set, and inferences about preferences must be based on the perceived rather than the actual feasible set. These difficulties of inferring preferences from observed choices in the presence of unperceived jointness are not restricted to situations involving nutrition and fertility but apply equally to choices involving diet and health or transportation and safety, or any other context in which unperceived jointness is present.

Welfare inferences—even welfare inferences based on the family's preferences—are difficult to make in the unperceived jointness model because it is difficult to infer the family's preferences from its observed choices. This is clearly true in the short run, when the number of births or deaths realized by the family is different from the numbers it expected on the basis of the perceived births and deaths functions. But it is also true in the long run, when realized and perceived births and deaths coincide. The difficulty with revealed preference-type inferences based on the fulfilled expectations solution is that even in such an equilibrium the family's perception of its feasible set of alternatives is inaccurate.

2.5.3 Unwanted Fertility

We have defined "excess" or "unwanted" fertility as the difference between optimal and desired fertility; that is, $b^o - b^d$. We are concerned with two general causes of unwanted fertility, the economic costs and the preference drawbacks of fertility regulation.[56] A reduction in the economic costs of fertility regulation (e.g., a reduction in the price of condoms, diaphragms, or pills) represents a clear welfare gain to those whose excess fertility is reduced. The introduction of a new fertility-regulation technique (e.g., the pill) also represents a clear welfare gain for those who choose to use it. However, the welfare evaluation of a reduction in unwanted fertility due to an increase in the use of contraceptives because of a change in the family's attitudes toward their use is more complex. Evaluated in terms of the family's new preferences, the change is an improvement, but evaluated in terms of its old preferences it is not. Thus the evaluation of the welfare impact of a government program that operates by changing tastes so as to reduce the preference drawbacks of fertility regulation is necessarily ambiguous.[57]

The view is common in the demographic literature that reduction or elimination of unwanted fertility through public policy would increase welfare. The Chicago-Columbia version of the economics of fertility lends itself to this view because it minimizes the role of preferences in determining contraceptive usage and emphasizes the importance of access to information and efficient use of a contraceptive technique. Thus Becker (1960) attributed the high completed family size of poor families to contraceptive failure owing to inadequate information. Similarly, Michael and Willis (1976) show that in the United States higher levels of formal education are related to lower contraceptive failure rates. If this reflects the greater efficiency of these families in fertility regulation, then a decrease in excess fertility would imply an improvement in the welfare of a family.

We do not assert that government-sponsored programs to control fertility cannot be valuable. But we do insist on distinguishing between benefits that accrue to the families whose excess fertility is reduced and benefits that accrue to others in the society.[58] In evaluating the benefits to the families whose excess fertility is reduced, it is important to understand the mechanisms through which such programs operate. To the extent that such programs operate by changing the preferences of the families whose excess fertility is reduced, there is no clear way to determine whether the families in question have benefited.[59]

A government-sponsored program that reduced unwanted fertility by lowering the economic costs of fertility regulation clearly benefits families whose excess fertility is reduced, provided the costs of the program are paid by others (i.e., by other groups within the society or by outside groups such as the United Nations). If the costs of the program are paid by taxes levied in part on the group whose unwanted fertility is reduced, then the question whether their welfare is increased depends on the balance between the benefits of lower-cost fertility regulation and the costs in the form of higher taxes; there is no presumption that the benefits outweigh the costs.

The strongest case for economic benefits can be made on the grounds of market failure. The argument for the existence of market failure is generally based on the fact that information collection and dissemination is a public or quasi-public good. A governmental unit can internalize both the information costs and the direct costs of establishing a market, whereas individual families cannot. The information–market-failure argument for government intervention presupposes an absence

128

of knowledge on the part of families that particular techniques of fertility regulation are available, a situation that is more likely to exist in less developed countries than in advanced industrial societies.[60]

When the reduction in excess fertility is the result of government proselytizing for the acceptability of contraception in order to reduce its preference drawbacks, we cannot infer that the reduction in excess fertility implies a welfare gain to the family. Even when the fertility control program is associated with a reduction in the economic costs of fertility regulation (e.g., by the free provision of fertility control devices and associated medical care not financed by taxes levied on those who practice fertility regulation), welfare gains cannot be inferred if tastes are changed at the same time.

In evaluating the welfare impact of family planning programs one must distinguish between developed and less developed countries. Preference drawbacks and economic costs underlie excess fertility in both areas, but economic costs are likely to be more significant in less developed countries, whereas preference drawbacks are likely to predominate in developed countries. Family planning programs designed to change preferences regarding the use of fertility regulation may be justified in terms of their benefit to society as a whole, but it is difficult to argue that such programs improve the welfare of the families whose tastes they change and whose excess fertility they thereby reduce. The benefits to families whose excess fertility is reduced by a government program that reduced the economic costs of fertility regulation are likely to be considerably smaller in developed countries than in less developed countries because the costs of access to information are typically much lower in developed countries. Most parents know of the existence, availability, and method of use of "reliable" techniques of fertility regulation (i.e., techniques with low theoretical failure rates). Many, however, continue to report unwanted fertility. Since the economic costs, including information access costs, are low in developed countries, the preference drawbacks must be decisive. Hence the main elements in unwanted fertility in developed countries appear to involve preference and motivation and not the economic costs or lack of information about fertility control.

2.6 Conclusion

Although there have been important advances in the analysis of fertility since the pathbreaking work of Becker (1960) and Leibenstein (1974*a*), the subject has become increasingly fettered by a narrowing view of the determinants of fertility. The framework laid out in this paper is intended to reverse this tendency by emphasizing a number of neglected determinants of fertility that deserve further exploration. Some of the principal views we have advanced are the following:

1. To come to grips with the variety of real-world fertility behavior, models of fertility determination must be expanded to include preferences and the biological production relationships. We propose a framework that includes such considerations in section 2.2, where we emphasize the role of "interdependent preferences" and the "births production function."

2. This model of section 2.2, although rich in analytical potential, is complicated, and practical application is limited by the lack of data. On fairly reasonable assumptions, however, special cases of the model can be distinguished, ranging from one in which fertility is independent of preferences for children to one in which the births production function becomes irrelevant and preferences for children—perhaps endogenously determined—play a crucial role. These special cases are discussed in section 2.3.1.

3. Evidence both from household surveys of fertility control practices and from census and other data relating to actual fertility behavior show that in many less developed countries deliberate efforts by individual families to regulate their fertility are rare. Hence the fertility of the bulk of the population is determined by its "natural fertility." For such countries, time-series and cross-sectional fertility variations may primarily reflect determinants of natural fertility rather than desired family size (see section 2.3.2).

4. Survey data make it possible to subdivide a population into those who practice deliberate fertility regulation and those who do not. We believe the analysis of fertility requires that these groups be treated in different ways. To explain the behavior of those who do not deliberately control their fertility, models stressing "natural fertility" and ignoring preferences for children are appropriate. For those who deliberately control their fertility, models emphasizing preferences for children and the effects of prices and income on desired family size are appropriate. Attempting to analyze the fertility behavior of an entire population without distinguishing between those who deliberately regulate their fertility and those who do not may result in biased estimates of the likely response of fertility to changes in incomes or to wider access to modern techniques of fertility regulation (sections 2.3.3, 2.3.4).

5. There is a need for further research in three relatively unexplored areas: preferences (sections 2.2.4, 2.4.1), the births production function (section 2.4.3), and unperceived jointness. Research on preferences could include both the endogenous formation of preferences for children and the role of the preference drawbacks of fertility regulation as a determinant of observed fertility. The investigation of the births production function should clarify the effect on fertility of practices such as lactation as well as such variables as health and nutrition. Of particular importance in evaluating the births production function is the role of

129

unperceived jointness. In this case the family does not fully incorporate into its behavior the relationship between its fecundity and its consumption decisions. Unperceived jointness is a pervasive problem in economics because individuals are often maximizing without taking account of the full interrelationships among constraints and between constraints and preferences.

6. Because our framework embraces a wider range of fertility determinants than the Chicago-Columbia approach, it is consistent with a greater variety of hypotheses regarding the factors that shape fertility trends and differentials. For example, it is consistent with an explanation of the demographic transition in which, in the early stages, an upsurge in fertility occurs owing to natural fertility factors. It is also consistent with the possibility that there will be substantial long-term fluctuations in fertility in developed countries rather than a monotonic downward trend. The framework also suggests a more cautious approach in evaluating the welfare effect of reducing "unwanted fertility," since its reduction may reflect a change in tastes (e.g., a reduced aversion toward the use of certain fertility regulation techniques) rather than a movement to higher indifference curves on an unchanging indifference map (section 2.5).

Notes

1. See Schultz (1974). No single designation for this approach is fully satisfactory. Here we adopt the term used by one of its advocates, Keeley (1975).

2. See Schultz (1974); Keeley (1975), Leibenstein (1974*b*), and Schultz (1976).

3. T. Paul Schultz's article, although published in 1976, was originally prepared for a 1973 conference. Inevitably, there are differences among members of a "school" on particular points, and injustice may be done to one or another individual in a general discussion. Moreover, there are indications that several of the leading workers may be venturing in directions we advocate. A recent paper by Michael and Willis, for example, departs strikingly from the usual Chicago-Columbia model, and introduces a "natural fertility" concept (Michael and Willis 1976). T. Paul Schultz has encouraged work on natural fertility at the Rand Corporation and has recently given more attention to "biological factors" in a discussion of the relation between infant mortality and fertility. Ben-Porath, whose identification with the Chicago-Columbia approach is in any event uncertain, has explored the issue of intergenerational taste influences (Ben-Porath 1975).

4. T. W. Schultz, on the other hand, is markedly restrained in commenting on the relevance of the Chicago-Columbia approach to less developed countries: "Turning to fertility behavior in low income countries the [Chicago-Columbia] household model as it now stands has not been developed to treat the particular classes of circumstances that constrain the household in these countries. These are countries in which illiteracy abounds, human time is cheap, and the income oppor-

tunities that women have outside the home are mainly not jobs in the labor market. Furthermore, infant mortality is high, life expectancy is low, debilitation during the adult years is substantial for reasons of inadequate nutrition and endemic diseases, and the availability of modern contraceptive techniques, including information about them, is, in general, wanting. These classes of circumstances are not yet at home in the household model" (Schultz 1974, p. 20).

5. This failure is admitted by both Keeley (1975, p. 466) and Schultz (1976, p. 94). Curiously, although the value of a relative income model in explaining this movement is generously acknowledged by these writers, they are not led to reconsider their general stance against research on preferences.

6. For a discussion of the role of sex preferences in determining family size, see Ben-Porath and Welch (1976).

7. The seminal paper in the household production literature is Becker (1965). In Lancaster's model (Lancaster 1966*a*,*b*, 1971) goods possess "characteristics" that are often identified with Becker's "commodities," and the "technology" is linear. Becker often uses fixed coefficient production functions as an expositional device, but linear technology is not an integral part of his model. For a recent sympathetic statement of the household production approach, see Michael and Becker (1973). For a discussion of some of its limitations, see Pollak and Wachter (1975).

8. It is customary to assume that the household's preferences over the commodity space are well behaved in the sense that they can be represented by a continuous utility function that is strictly quasi-concave and nondecreasing in its arguments. If the feasible set in the commodity space is convex, these assumptions guarantee that the utility maximizing collection of commodities is unique.

9. The "cost function," $C(P,w,Z;T)$, is defined as the minimum cost of producing the commodity bundle Z with the technology T at goods prices P and wage rates w. That is,

$$C(P,w,Z;T) = min \sum_{k=1}^{n} p_k x_k + \sum_{h=1}^{H} \sum_{s \epsilon T} w_h t_{hs},$$

subject to $(Z,X,T) \epsilon T$. We can use the cost function to translate the budget constraint from the goods space into the commodity space. The translation of the constraint is the requirement that $C(P,w,Z;T)$ not exceed the family's "full income" (i.e., the household's total earnings if it devoted all of its time to market work):

$$C(P,w,Z;T) \leq \mu + \sum_{h=1}^{H} w_h \bar{t}_h.$$

10. These issues are discussed in Pollak and Wachter (1975), where it is argued that joint production is pervasive in household production situations, especially when the role of time is recognized. For further discussion, see Barnett (1977) and the reply by Pollak and Wachter (1977).

11. Formally, it would be possible to treat the same fertility-regulation technique practiced with different intensities as different techniques. Our formulation is more consistent with ordinary usage and is capable of casting some light on the question why some population groups have higher "failure rates" than others using the same technique.

12. We take the family (more specifically, the parents) rather than the individual to be the basic unit of analysis. The assumption that the family (i.e., the husband and wife collectively) has well-defined preferences begs the issue of aggregating the separate preferences of the husband and wife into a collective preference ordering. Samuelson (1956) provides a classic statement of the problem; Nerlove 130

(1974, p. S 204) describes the resolution of these problems by postulating a "family utility function" as the "Samuelson finesse."

13. If the total cost of fertility can be decomposed into a fixed cost, $\rho_0(\tau)$, and a variable cost, $\rho_1(\tau)\theta$, the cost function takes the form

$$\rho(\theta,\tau) = \rho_0(\tau)k + \rho_1(\tau)\theta,$$

where

$$k = 0 \text{ if } \theta = 0 \text{ and } k = 1 \text{ if } \theta > 0.$$

The cost of fertility regulation might also depend on the fecundity of the family, which may in turn depend on its goods purchases and commodity consumption; abortion is an example of a technique whose cost depends on fecundity.

14. The earnings of children could be incorporated into the model either by expanding it to include an "earnings function" or by interpreting vector t to include the allocation of the time of children. We implicitly adopt the latter course to avoid additional notation.

15. Existence of a solution poses no real problems, but uniqueness is a different matter. We have not ruled out the possibility of multiple solutions. The usual uniqueness argument rests on the assumptions that feasible sets are convex and preferred sets are strictly convex. But some of our variables have no "natural" units of measurement, and there are no market units we could adopt by convention. For example, given any index of the intensity of use of a particular fertility regulation technique, any increasing transformation of that index would serve equally well. But such transformations can alter the convexity properties of feasible sets and preferred sets, so that the usual type of uniqueness argument cannot be made. Of course, the uniqueness of the solution cannot be altered by such transformations, and uniqueness is assured if there exist any units of measurement in which the feasible sets and the preferred sets are both convex and one or the other is strictly convex. Since we cannot establish uniqueness, we cannot guarantee that the optimal solutions are continuous in the variables the family takes as exogenous.

16. The term "unperceived jointness" is motivated by viewing the household as having a single production technology instead of three distinct technologies, one producing births, another infant mortality, and the third the other commodities. We call this single technology the household's "generalized technology." This treatment avoids treating births, deaths, and fertility regulation as distinct from the other commodities by extending the notion of commodities to include all of the arguments of the family's utility function; we refer to these variables as "generalized commodities." The generalized technology exhibits joint production because the same inputs affect the output of more than one generalized commodity: for example, purchased food inputs produce the generalized commodity "nutrition," which is desired for its own sake, but they also influence the output of the generalized commodities "births" and "deaths." The assumption that the family is not fully aware of the relationship between nutrition and births (or deaths) implies that the jointness in the household's generalized technology is at least in part "unperceived."

17. The polar cases here are the extreme points on the continuum from ignorance to knowledge; they do not coincide with the special cases of the classification scheme described in section 2.1. Indeed, that discussion assumed "complete ignorance" in order to define "natural fertility."

18. Since the family's reproductive span, Λ, is not a decision variable for the family, this constant specification is equivalent to $\hat{B}(a,Z,X,l,0,0,\Lambda) = \beta\Lambda$ where the family believes that the ratio β is not affected by its decisions. If the reproduc-

tive span were made a decision variable, then the two specifications would no longer be equivalent, and age at marriage would become a possible mechanism of conscious and deliberate fertility regulation.

19. For example, $\bar{B}$ might be equal to the average fertility of the most recent cohort to have completed its reproductive span, or a weighted average of the experience of such recent cohorts, perhaps restricted to families of similar socioeconomic status. A more complicated specification might make use of the experience of families who had not yet completed their reproductive spans. This would be legitimate even in our one-period planning model, but it would not be legitimate to use the family's own experience or that of other families of its cohort as a basis for prediction.

20. A one-period planning model cannot capture the behavior of a family that did not intend to practice fertility regulation, has more children than it expected, and then begins to practice fertility regulation. Inability to reflect this type of period behavior is a serious drawback of one-period planning models. Two points should be made. First, this defect is relevant only for families that have knowledge of fertility-regulation techniques they would utilize if they knew their true fecundity. Second, the difficulty of incorporating unperceived jointness into the one-period planning model is a point against the one-period planning model, not against unperceived jointness.

21. There is a conceptual difficulty here, again reflecting the confines of the one-period planning model. Presumably, the eventual allocation of expenditure among goods is determined by realized rather than perceived family size: a family that expects two children and has four will buy more "child goods" and fewer "adult goods" than it planned. One can imagine a two-period model in which the consumption pattern in period one determines realized family size, and realized family size determines the consumption pattern in period two. In a multiperiod model, births and mortality in each period would depend on consumption patterns in previous periods, and consumption patterns in each period would depend on actual family size and composition in that period.

22. Similarly, in a Cournot duopoly equilibrium each firm correctly predicts the output of the other firm without perceiving the reaction function that generates that output.

23. Notice that perceived and realized completed family size could be equal even if $b^p \neq b^r$ and $d^p \neq d^r$, if there are offsetting errors, but this is not a fulfilled-expectations equilibrium. We have defined an equilibrium in terms of births and deaths rather than completed family size because a divergence between perceived and realized births will cause a revision of expectations about the births function and a revision of plans.

24. It might be thought that a model of habit formation in which a family's own past consumption levels influenced its taste for goods would be appropriate, but such a specification cannot be developed within the structure of a one-period planning model. For the family's own past consumption experience to play a role, we need a sequential model in which the family makes decisions at more than a single decision point.

25. The simultaneous specification in which each family's preferences depend on everyone else's current decisions is analytically intractable because the preferences of each family in a particular cohort are determined by variables whose values depend on the behavior of all families in that cohort. In the lagged specification the preferences of each family in a particular cohort depend on the behavior of families in earlier cohorts, and hence the model has a recursive rather 131

than a simultaneous structure. With the lagged specification it is only in a "steady state" equilibrium that the full effects of interdependent preferences manifest themselves.

26. The notation of a nonnegative relationship between normal levels and preferences can be formalized as in Pollak (1977, n. 8).

27. This terminology is used in Pollak (1977) to distinguish preferences over goods (X) that depend on prices (P)—"conditional preferences"—from preferences over alternative goods-vector-price vector situations (X,P)—"unconditional preferences." The nomenclature is analogous to that used for conditional probability.

28. The "necessary" level interpretation works for the linear expenditure system, provided that certain parameters assume nonnegative values; but there is no a priori reason to believe that these parameters are nonnegative. For a discussion of both of these systems and references to the literature, see Pollak (1970, 1971).

29. A word is necessary also on the distinction between "socially controlled" and "family controlled" fertility. By "family controlled" or "family regulated" fertility we mean deliberate efforts by individual households to influence their fertility. From this point of view, natural fertility is socially controlled but not family controlled. For example, an intercourse taboo observed as a matter of custom is a social control that affects the level of natural fertility. It does not, however, imply controlled fertility in our sense, since observance of the taboo by individual households is not geared to family-size concerns. For a contrary view, see T. Paul Schultz (1976, p. 92). It should be noted also that while the present concept of natural fertility reflects social controls, such controls are only one of a number of societal conditions that affect natural fertility. War, for example, may reduce natural fertility by separating spouses, but it would not be viewed as a social control on fertility.

30. The focus of our approach is on groups or collections of families rather than on individual families or on society as a whole. This is to avoid both the problems associated with random or stochastic elements in the births and infant mortality functions and the discreteness of children. This allows us to interpret our model as applying to the mean experience of a group of identical families. We do not assume that all families in a particular society belong to the same group—quite to the contrary, important aspects of demographic behavior can be captured only by recognizing the changing balance among the groups we have described.

31. A possible third type of evidence comes from studies in which an attempt is made to formulate and test hypotheses that distinguish between "behavioral" and "biological" determinants of fertility. As explained in the preceding section, a finding in favor of behavioral influences does not necessarily imply controlled fertility in our sense, since the actual issue relates to whether the behavior is motivated by its possible fertility effect. On the other hand, a finding in favor of biological influences can be viewed as support for uncontrolled fertility in our sense. Without pretending to do a systematic survey, our impression is that the results of a number of these studies lean toward the importance of biological factors. (Cf., e.g., Anderson and McCabe [1977], Chowdhury et al. [1976], and Taylor et al. [1976].

32. A useful early summary report on some of these surveys is Mauldin (1965); a recent review is given in Nortman (1977).

33. Some surveys aim explicitly for comprehensive coverage of possible methods. For example, a recent survey in Nigeria asked specifically about traditional methods, the practice of abstinence, and possible use of extended lactation as ways of limiting family size (Caldwell and Igun 1972).

34. A useful discussion of some of the shortcomings of the Henry concept is given in David and Sanderson (1976, pp. 143 ff.).

35. The index value can actually take on mildly negative values. This is because the "standard" age pattern for a natural fertility regime is an average of schedules for ten cases, and a given situation might actually involve a relationship between the age-specific fertility of older and younger women that is higher than the standard natural fertility case.

36. See also Knodel (1977). Knodel's paper is especially pertinent to the present discussion, for it concludes, from calculations of m, that "modern family limitation (i.e., parity-specific fertility control) was largely absent prior to a secular fertility decline in both Europe and Asia" (Knodel 1977, abstract).

37. Some additional data may be noted bearing on the prevalence of "excess demand" situations. Survey data for rural Morocco (1966), West Malaysia (1967), and Kenya (1966) indicate that among wives 35–49 the proportion who want more children is substantial, ranging between about one-fourth and one-half. A recent survey in an area of rural Indonesia states that "despite relatively high levels of ideal family size (average 4.5) . . . , women in Mojalena give birth to an average of only 3.9 children; moreover, owing to high rates of mortality, completed family size averages 2.7 children" (Singarimbun and Manning 1976, p. 175). On the other hand, in Potharam (1964) the proportion was only a tenth or less. The Morocco data are from Lapham (1970); Kenya, from Heisel (1968); and West Malaysia, from Palmore (1969). The figures for Kenya include those for whom the "desire for children" was not ascertained or "up to God," a category that in rural Morocco accounted for only about 4–6% of the respondents. Indonesian data are from Singarimbun and Manning (1976); the Potharam data, from Peng (1965). See also Tabbarah (1971).

38. This is a pattern suggested by the data for rural Mysore (United Nations 1961, chap. 10) and more recently by work on Indonesia (Hull and Hull 1977) and Iran (Ajami 1976).

39. As drawn, this curve lies below that for groups I and II, but one can imagine conditions under which it might lie above it. Clearly, for all of the groups, identification of typical patterns is itself a research issue.

40. One might imagine a corresponding trend in fertility differentials by socioeconomic status as the nature of the underlying determinants changed. Suppose, for example, that the demographic transition involved simply a shift from an initial group I–group II situation through group III to a wholly group IV situation. Then the initial pattern of socioeconomic status–fertility differentials for the population as a whole might be given by the positively inclined b_n curve of figure 2.3 above, reflecting the effect of natural fertility factors. When the society was in the group III situation, the negatively inclined curve would prevail, and in group IV the horizontal curve. Thus one might hypothesize a trend in fertility differentials by socioeconomic status from positive through zero to negative and back to zero again. However, this is only one possibility. The point is that the expected pattern of fertility differentials would shift as the underlying determinants of fertility changed.

41. See, for example, Pollak and Wales (1969), Wales (1971), and Howe, Pollak, and Wales (1977).

42. Taste differences, like differences in technology, can be and often are used as a deus ex machina when other explanations fail. But the fact that specifications involving taste differences (or technological differences) can be misused is not a justification for ignoring them.

43. See, for example, Easterlin (1973) and Wachter (1975). For other work on taste formation in a time series context, see Leibenstein (1974*b*), Lee (1976, 1977), and Lindert (1978). For the application of the relative income model to the related question of labor force participation behavior, see Wachter (1972*b*, 1974).

44. See Easterlin (1973), Lee (1975 *a,b*), and Wachter (1972*b*, 1974).

45. Indeed, some empirical work suggests that the ceteris paribus relationship between education and family size is U-shaped. See, for example, Yoram Ben-Porath (1973).

46. Work at the level of intermediate variables, represented by what are known as "renewal models," seeks to account for fertility through factors such as age at sexual union, frequency of intercourse, probability of conception, length of the nonsusceptible period, and duration of reproductive union. So far as the present framework is concerned, this research is of interest primarily for the guidance it may provide into more fundamental causal factors at work. For example, if the nonsusceptible period (NSP) is an important source of fertility variation between two societies, one may be led to inquire into lactation practices, a seemingly important determinant of NSP and, in turn, into the determinants of these practices. However, the proximate components of fertility do not each depend uniquely on *different* causal factors—for example, a number of the intermediate variables might be affected by nutrition. An excellent concise presentation of renewal models is given by Keyfitz (1971). Economists who have followed this lead in recent work include Michael and Willis (1976), David and Sanderson (1976), and Crafts and Ireland (1976). Leridon (1977) has recently completed a valuable comprehensive survey of the field, which makes accessible in English the pioneering work of the French demographers, led by Henry and Bourgeois-Pichat.

47. We have avoided here the terminology of "demand" and "supply" models. As the optimal solution function illustrates, there are no demand and supply functions in the traditional sense.

48. If one disregards Puerto Rico and Chile, which are uncertain representatives of less developed countries' experience even for Latin America, the studies cited by Schultz as empirical support for the relevance of the Chicago-Columbia type of demand approach to less developed areas are: Egypt 1960, Philippines 1968, Thailand 1960, and Taiwan 1964–69. With the exception of Taiwan, the available evidence indicates extremely low levels of deliberate fertility control in these countries at the times studied. Table 2.3 shows very low indexes of fertility control for the Philippines and for Thailand. In Egypt in 1960 the proportion of married women of reproductive age who had practiced family planning was, in rural areas, 1.5%; semiurban, 12.0%; and urban, 17.0% (Mauldin 1965, p. 9). (The rural proportion of the population in 1960 in Egypt was 62.0%.) Even in regard to Taiwan, as shown in table 2.1 above, in 1965 less than half of married females aged 35–39 had practiced deliberate control. These observations suggest that in the studies cited by Schultz a substantial share of the population, and in some cases almost all the population, is in a natural fertility situation.

49. A good concise exposition is provided in Schultz's appendix (Schultz 1976).

50. See Easterlin (1968) and Wachter (1972*b*).

51. A similar point is made in Pollak (1976*b*) in the context of interdependent preferences, and in Pollak (1977) in the context of price-dependent preferences.

52. Our example assumes a lexicographic preference for family size, but this is not crucial. Notice that, because the relevant utility functions are conditional rather than unconditional, we could multiply the utility function of the woman with one sibling by 100 while leaving the conditional utility function of the woman with two siblings unchanged; such transformations have no effect on the behavior

implied by the utility functions, but the admissibility of such transformations shows that the level of utility cannot be used to compare the satisfaction or well-being in such cases.

53. See Easterlin (1975) for a survey of results of this type.

54. A third approach, based on the long-run behavior implied by the endogenous taste model, makes use of the "long-run" utility function. This approach was proposed by von Weiszäcker (1971) and criticized on conceptual and technical grounds by El-Safty (1976*a*,*b*), Hammond (1976), and Pollak (1976*a*).

55. Strictly speaking, the assertion in the text should refer to an individual rather than a family, but we assume that families, like individuals, have well-defined preferences.

56. There is a third source of excess fertility. No fertility regulation technique (excluding abstinence) is technically perfect even under ideal conditions. Associated with each method of fertility control is a minimum failure rate, termed the "theoretical" failure rate (Leridon 1977, p. 122).

57. A fertility control program which changes preferences for children may reduce fertility without reducing excess fertility. This is not an unlikely result.

58. Those whose fertility is unaffected might benefit from a reduction in the fertility of other groups in the society if the tax and transfer structure caused them to pay a portion of the cost of the unwanted children.

59. For a more detailed exposition of this argument, see Wachter (1972*a*).

60. Costs of fixed information and costs of access to fertility control may be sizable in many less developed countries today. Where modern medicine is not readily available, the costs of acquiring modern contraceptive techniques can be prohibitive. For example, parents in a rural village that has neither a doctor nor a clinic could not import modern contraceptive techniques and associated medical care except at a very high initial or fixed cost. For these families, the traditional methods of abstinence and withdrawal may be the only forms of regulation that can be adopted without violating the budget constraint. To the extent that excess fertility prevails, the fact that these methods are often not utilized attests to their significant preference drawbacks. As development occurs, an increasing proportion of households in less developed countries have the motivation to practice fertility control, but the economic costs are too high for modern techniques and the preference drawbacks too high for traditional fertility regulation. At this stage these areas offer at least the potential of large economic benefits if the government were to organize the necessary infrastructure for dispensing contraceptive information and techniques. The government is in effect capturing an externality by establishing a market for modern contraceptive devices.

Comment Harvey Leibenstein

This is an unusually stimulating paper on a very difficult subject. Its main features, as I see them, are as follows: (1) It emphasizes and employs a demographic view of economic development. (2) It contains a taste-shift factor that is unique for models of this type. (3) It develops

Harvey Leibenstein is Andelot Professor of Economics and Population at Harvard University.

an interesting concept in the notion of "unperceived jointness." (4) It separates the demand for children and the demand for controls. (5) In one sense the model is conventional in that it assumes the maximization of utility.

Among the clearly desirable features of the model are its introduction of the taste-shift factor and its use of a demand for controls equation.

It is very difficult to assess the purely demographic view of economic development. Demographic behavior appears to hang in a void unconnected to economic and social trends. An alternative that might have been considered is to connect the theory to reasonably uniform patterns of development of the type studied by Kuznets, Chenery, and others. For example, fairly specific things can be said about shifts of labor toward urban areas, increases in education, and shifts in broad occupational categories, which usually accompany economic development. A model more explicitly connected with persistent patterns of change would add a feeling of realism, but it is impossible to say at this point if it would have greater explanatory power.

An interesting feature of the model is its use of the concept of natural fertility within marriage as an anchoring point for the predevelopment situation. The difficulty with this approach is that it omits the marriage age as a control variable. The view I am espousing is that there are wide variety of social controls of population even in developing countries, and, furthermore, that the social controls are *substitutes* for private controls. The view emphasized is that we must not underestimate the significance of the *substitution* of some controls for others as part of the process of demographic change.

It may help to keep in mind the following list of *population* controls:

a. Nonmarriage of women—spinsterhood
b. Late marriage
c. Celibacy rules for some professions
d. Taboos on widow remarriage
e. Periods of noncohabitation
f. Infanticide
g. Neglect leading to infanticide
h. Long lactation periods
i. Ritual taboos on intercourse
j. Abortions
k. Contraceptive means
l. Outmigration

We should note that the word used is population rather than fertility. All population controls are to some degree substitutes for each other. Some of these controls are social controls, others are individual controls within the power of family members. But the individual controls are substitutes for social controls. Hence, if we consider only fertility within marriage, we may lose some sense of the capacity for substitution between various types of controls. While the use of the demand for fertility control in the Easterlin model strikes me as an excellent idea, in some contexts fertility controls are likely to be substitutes for existing population controls, and hence a sense of the degree of substitution may help us assess the *net* demand for some specific fertility controls.

The concept of unperceived jointness seems extremely useful and is likely to take care of observed anomalies in the analysis of specific situations. There is some danger in a concept of this sort, since it is unlikely to be observable, in that it may be tempting to use it as a rationalization of any deviation between the results of empirical research and the predictions from a specific model.

It is understandable that Easterlin, Pollak, and Wachter should use a utility-maximizing model, since this is the conventional approach among economists. But this seems to me to be a questionable procedure. First, it leaves out frequently observed characteristics of behavior—repetitive behavior and inertia. Second, and most important, it leaves out changes in degree of rationality as an explanatory factor. In criticizing the maximization assumption, a question that frequently arises is whether there is any alternative. In the pages that follow, I shall present the bare bones of a nonmaximizing model and suggest, albeit quickly and necessarily vaguely, how this model might be used to handle some of the concepts of the Easterlin/Pollak/Wachter model or related models. Below, a brief comparison is made between the standard theory and the one I propose, which I shall refer to as general X-efficiency theory. (For a detailed exposition of these ideas, see Leibenstein (1976, chaps. 5–10.)

Postulates and Basic Variables	*Conventional Microtheory*	*General X-Efficiency Theory*
1. Behavioral postulate	1. Maximization or minimization	1. Selective rationality
2. Units	2. Households and firms	2. Individuals
3. Efforts	3. Assumed given	3. Discretionary variable
4. Interpersonal interactions	4. None	4. Some
5. Inert areas	5. None	5. Important variable
6. Agent-principal relationship	6. Identity of interests	6. Differential interests
7. Motivation as an output	7. Assumed given	7. Significant variable

The basic assumption behind my theory is that people work out a compromise between the way they would like to behave, in the absence of constraints, and the way they would like to see themselves behave in terms of their standards of behavior, or superego. Under selective rationality, individuals do not pursue opportunities for gain to the maximum degree given the constraints, nor do they optimize the pursuit of information. In other words, they select the degree of constraint concern their personalities dictate.

The cost of ignoring constraints is a feeling of pressure. This pressure may be in part the result of ignoring consequences and one's desires to behave in accordance with one's internalized standards (superego). Thus, individuals "choose" a compromise position between *pressures* and *a degree of constraint concern* to operate at a psychologically comfortable level. This implies, first, that individuals do not necessarily or usually pursue gains to be obtained from an opportunity to a maximum degree; and, second, *maximizing behavior is a special case in this system*. The specific compromise an individual makes between the competing demands of his id (unconstrained desires), and his superego (standards), on the average, may be viewed as an index of his personality. If he yields too much to his superego, he will feel pressure to behave in terms of less constraint, and if he behaves with too little constraint he will feel the pressure of his conscience. Thus personality and context select, so to speak, the degree of rationality that will control an individual's decision-making (and performing) behavior. The context may contain strong countervailing pressures to increase the degree to which an individual approaches maximizing behavior.

Since motivation is extremely important in determining behavior, we have to take into account interpersonal interactions and especially peer group interactions that determine the system of approval and disapproval, which in its turn influences choices. At the same time, the distinction between principals and agent is extremely important in such contexts, since if effort is a variable there is no reason to presume that the interests of the agent and the principal are identical. Many choices are carried out by agents, but there is no reason to assume that the agent puts forth the same degree of effort that the principal would in similar circumstances.

An important element in our system of analysis is the concept of inert areas. As its name suggests, this is akin to the notion of inertia. Individuals are presumed to choose effort positions (a set of related effort options) in interpreting their jobs or roles in specific contexts. The basic idea is that once an effort position exists for some time period, an individual may not shift to a new position even though a gain may be achieved thereby, because the cost of moving from one effort position to another is larger than the perceived gain. Thus, individuals may find themselves stuck within inert areas even though, apart from the cost of moving, superior effort positions may exist even from the individual's viewpoint.

In what follows, the idea of inert areas will be used to examine some of the basic notions in the Easterlin/Pollak/Wachter paper in order to illustrate how they could fit into a nonmaximizing framework. Given the space constraints, we can only vaguely suggest how it all works out. Now inert areas are made up of two components: a segment that expresses some aspects of selective rationality (e.g., ignoring very careful calculation), and another segment that involves the cost of moving from one position to another.

Natural Fertility

We may visualize natural fertility as being based on routine behavior patterns utilizing a traditional mix of population controls. These routine behavior patterns are presumed to operate within an inert area. They do not change unless pressure is exerted beyond some minimum level. Thus, natural fertility would not be interpreted here to imply some maximal level of fertility, nor would it imply a complete lack of population controls, including nonmechanical means of contraception (e.g., coitus interruptus); rather, it would imply a situation before the introduction of modern contraceptive means. Thus a situation frequently found in developing countries before sustained fertility decline could be fitted into the natural fertility idea. The transition between the natural fertility state and the partially controlled state would then be observed as pressure increases sufficiently to induce some people to adopt additional controls.

Tastes

The concept of the transmission of taste from one generation to another can also be interpreted in terms of the inert area principle. Up to a point, the inherited taste pattern would persist, but as modernization creates pressures for new tastes and consequent consumption patterns that compete significantly with children, we would expect the old tastes to yield to some degree. Furthermore, we would expect the existing tastes at any one time to be the product of inherited tastes as well as peer-group influences, to the degree that peers adopt modern consumption standards. As fertility declines, a conflict is created between the inherited tastes and the peer group influences, and the rate at which there is a shift from one to the other would be determined by the size of the inert areas.

Techniques of Control

Like the above, techniques of control could also be interpreted through the inert area framework. Namely, the set of controls normally used would be surrounded by inert areas; but, as sufficient pressure is generated and new techniques are introduced, the new techniques gradually become part of the option set of the techniques available. Those with the narrowest inert areas are likely to become the initial adopters. (For an innovation adoption model along these lines, see Leibenstein [1976, pp. 234–39].) One could visualize a variety of stages between old techniques and new ones, representing different degrees of knowledge and confidence. We would not expect that the new techniques to become part of the demand for control until they become noticed, generally known, and tried.

Rationality Increase as a Factor in Fertility Determination

The existing theory does not allow for changes in degrees of rationality in determining eventual fertility decline. Clearly, if a maximizing model is used, this forecloses any increase in rationality. But the degree of rationality may depend on the diffusion of responsibility within which the nuclear family finds itself. Thus, if the nuclear family is part of an extended family in which there is considerable diffusion of responsibility for children and for economic well-being of household members, then there will be little pressure toward a high degree of rationality. As we obtain a shift toward the nuclear family as a separate independent unit and responsibility for economic welfare of the household becomes concentrated, then there is likely to be increased pressure for rational behavior. Exogenous influences, such as the gradual spread of secularization through modern education, will also result in an increase in rationality. In particular, as nuclear families become more responsible for their own welfare, the inert areas that surround their critical choice variables become narrower, and hence they respond to pressure with less inertia.

Comment Warren C. Sanderson

Economic theory teaches us that competition among producers usually benefits consumers. Competition among producers of economic models of fertility behavior is no different. Even though the market is dominated

Warren C. Sanderson is associated with the Department of Economics at Stanford University.

by two large producers, the Pennsylvania school and the "Chicago-Columbia" school, the contest to produce a model that more economists would buy has resulted not only in a substantial improvement in the models themselves, but in a number of valuable "spinoff" developments as well.[1] This essay by Easterlin, Pollak, and Wachter is the formal presentation of the 1976 Pennsylvania school model with a full description of all its novel features and a discussion of why it is superior to what the competition has to offer.

The paper focuses on three features of the new model: "endogenous preferences," supply side factors, and consideration of behavior under imperfect information. I shall comment briefly on each of the three.

"Endogenous preferences" is not a new feature of the 1976 Pennsylvania school model. Quite the contrary, it has been standard equipment on Pennsylvania school models since 1966.[2] What motivates its discussion in the paper, then, is not its novelty, but the manner in which the competition has reacted to its introduction. The members of the Chicago-Columbia school not only have refused to incorporate this feature into their own models, they have positively rejected it as being dangerous to the health of economic theory.

Before continuing the discussion of the "endogenous preferences" specification, it is useful to note that the contending models are not nearly so different as their producers might lead us to suppose. First of all, in the context of a single generation, preferences are just as *exogenous* in the 1976 Pennsylvania school model as they are in any of the Chicago-Columbia school models. Current preferences and behavior are not simultaneously determined in the models of either large producer.

There is a difference between the two types of models when dealing with fertility change over the course of several generations. In the Pennsylvania school models since 1966, preferences change generation by generation in a manner determined within the model. Until recently this stood in sharp contrast to the Chicago-Columbia school models, which maintained that preferences were invariant over time. With the publication of Stigler and Becker (1977) even this difference narrowed. Stigler and Becker argued that tastes are truly invariant, but that household production structures vary over time and space. Therefore consumers now seem to have a choice between a model that deals with intergenerational fertility changes within a framework in which preferences vary and the household production structure does not, and a framework in which the household production structure varies and preferences do not.

That preferences vary both across time and across cultures is plausible enough. Certainly, as the authors argue, this view should not be discarded on theoretical grounds. The same can be said, however, with regard to household production structures. Perhaps one day a combination model will be produced. 136

The second aspect of the 1976 Pennsylvania school model featured in the paper is concern with the biological aspects of fertility behavior. In section 2.2 a births function and a deaths function are introduced into a formal economic model of fertility, and in section 2.3 the concepts of natural fertility and the difference between desired and achieved fertility are incorporated into an economic framework. Again, the best context in which to understand these contributions is that of the competition between the two rival schools of economists. Scholars have produced substantial bodies of literature on the biological determinants of fertility and on the biological and behavioral correlates of infant mortality. These go far beyond anything found in this paper. One contribution of the 1976 model, then, is the integration of past demographic findings into an economic context. Demographers may find little new here, but in the competition between the two large producers, the Pennsylvania school has scored a success in aligning itself somewhat more closely with the results of previous demographic research.

In the same vein, section 2.3 can best be read as criticism of the competition for not incorporating the biological aspects of fertility behavior into their formal model. Without this perspective, a substantial portion of that section may seem rather puzzling. For example, a long argument is made to demonstrate that there are indeed some contemporary cultures in which the volitional practice of fertility control is virtually absent; but there are very few social scientists, if there are any at all, who would contest this point. The rationale of the argument becomes clear when it is viewed as a warning to economists not to use other models in those contexts where the biological aspects of fertility are likely to be important.

The third aspect of the 1976 Pennsylvania school model highlighted in the paper is the notion of behavior under imperfect information. Each couple is viewed as choosing a pattern of goods consumption, time allocation, fertility, and infant mortality that, subject to resource and technology constraints, maximizes their utility. The problem with adopting this view naively, as the authors point out, is that people are often ignorant of the consequences on fertility and infant mortality of various seemingly unrelated aspects of their behavior. To make their model more realistic, the authors suggest that couples be treated as if they maximized their utility subject to their resources, household technology, and possibly incorrect beliefs about the determinants of their experience of fertility and infant mortality. These couples are then assumed to maintain all other aspects of their behavior invariant even though the resulting family size is different from the one they anticipated.

This specification has two serious drawbacks. The first, mentioned by the authors in a note, is that consumption and time allocation should *depend on actual family size* as opposed to a *hypothetical* family size that never materializes. The second problem is more technical in nature. Since the family's consumption alternatives may depend on the earnings of children, it may not be possible to hold all other aspects of behavior constant when the actual family size is substantially below the anticipated one. In economic argot, the procedure proposed in the paper to deal with the problem of imperfect information is not guaranteed to result in feasible solutions. New models often have bugs in them, and I am confident that future technological advances will result in a preferable treatment of behavior under imperfect information.

My final comments concern the formal economic model presented in section 2.2 and its relationship to the arguments made in the other sections of the paper. It is important to note here that although the authors present an economic model of fertility behavior, they never use the model in the framework of a comparative statics analysis. This is a bit like creating an intriguing piece of machinery one never intends to use. The art in creating microeconomic behavioral models is in abstracting from all but the most important factors in a given problem so that the analysis of the model results in falsifiable implications. The model in section 2.2 is not constructed on this principle. Instead, it is specified so generally that in its present form it has no unambiguous implications when any of the exogenous variables are altered one at a time.

Since the model is consistent with almost any kind of behavior, it offers no guidance on what is plausible and what is not. For example, in section 2.3 there are three graphs concerning desired fertility—one in which it is drawn as an increasing function of nonlabor income, one in which it is drawn as invariant with respect to socioeconomic status, and one in which it is drawn as either a constant or a decreasing function of social and economic development. The model is certainly consistent with all three graphs. Indeed, the model is consistent with desired fertility being a *decreasing* function of nonlabor income (even if desired fertility is a "normal" good), a *sinosoidal* function of socioeconomic status, and an *inverted U-shaped* function of social and economic development. In other words, the model in section 2.2 has less substantive connection with what is said in the other parts of the paper than one might wish.

In conclusion, then, I reiterate that this paper is the product of a competitive struggle between two rival producers of economic models of fertility behavior. It contains not only explicit criticisms of alternative models, but numerous implicit criticisms. Although some of the arguments may seem either arcane or pointless to the nonspecialist, they are all aimed at perceived weaknesses in the Chicago-Columbia school's product lines.

Will the 1976 Pennsylvania school model come to dominate the market? It is not clear. After all, different people have different tastes/production functions (choose one or both) when it comes to the formulation and use of economic models of fertility behavior.

Notes

1. For example, Easterlin (1974), Sanderson (1974), and Stigler and Becker (1977).
2. Endogenous preferences were first introduced into the Pennsylvania school's models in Easterlin (1966).

References

Ajami, I. 1976. Differential fertility in peasant communities: A study of six Iranian villages. *Population Studies* 30 (November): 453–64.

Anderson, Barbara A., and McCabe, James L. 1977. Nutrition and the fertility of younger women in Kinshasa, Zaire. *Journal of Development Economics* 4 (December): 343–63.

Auster, Richard; Leveson, Irving; and Sarachek, Deborah. 1969. The production of health, an exploratory study. *Journal of Human Resources* 4 (fall): 411–36.

Barlow, Robin. 1967. The economic effects of malaria eradication. *American Economic Review* LVII (May): 130–48.

Barnett, William A. 1977. Pollak and Wachter on the household production function and its implications for the allocation of time. *Journal of Political Economy* 85 (October): 1073–82.

Becker, Gary S. 1960. An economic analysis of fertility. In *Demographic and economic change in developed countries,* ed. Universities-National Bureau Committee for Economic Research, pp. 209–31. Princeton: Princeton University Press.

———. 1965. A theory of the allocation of time. *Economic Journal* 75 (September): 493–517.

Behrman, Jere R. 1968. *Supply response in underdeveloped agriculture: A case study of four major annual crops in Thailand, 1937–1963.* Amsterdam: North Holland Publishing Co.

Ben-Porath, Yoram. 1973. Economic analysis of fertility in Israel: Point and counterpoint. *Journal of Political Economy* 81 (March/April): S202–33.

———. 1975. First generation effects on second generation fertility. *Demography* 12 (August): 397–405.

Ben-Porath, Yoram, and Welch, Finis. 1976. Do sex preferences *really* matter? *Quarterly Journal of Economics* 90 (May): 285–307.

Bumpass, Larry L., and Westoff, Charles F. 1970. The "perfect contraceptive" population. *Science* 169 (18 September): 1177–82.

Caldwell, J. C., and Igun, A. 1972. Anti-natal knowledge and practice in Nigeria. In *Population growth and economic development in Africa*, ed. S. H. Ominde and C. N. Ejiogu, pp. 67–76. London: Heinemann.

Chen, Lincoln C.; Ahmed, Shamsa; Gesche, Melita; and Moseley, W. Henry. 1974. A prospective study of birth interval dynamics in rural Bangladesh. *Population Studies* 28 (July): 277–97.

Chowdhury, A. K. M.; Khan, A. R.; and Chen, L. C. 1976. The effect of child mortality experience on subsequent fertility: Pakistan and Bangladesh. *Population Studies* 30 (July): 249–62.

Coale, Ansley J., and Trussell, T. James. 1974. Model fertility schedules: Variations in the age structure of childbearing in human populations. *Population Index* 40 (April): 185–258.

Crafts, N. F. R., and Ireland, N. J. 1976. Family limitation and the English demographic revolution: A simulation approach. *Journal of Economic History* 36 (September): 598–623.

David, Paul A., and Sanderson, Warren C. 1976. Contraceptive technology and family limiting behavior: Toward a quantitative history of the diffusion of contraceptive practices in America, 1850–1920. Unpublished manuscript.

Easterlin, Richard A. 1966. On the relation of economic factors to recent and projected fertility changes. *Demography* 3(1): 131–53.

———. 1968. *Population, labor force, and long swings in economic growth: The American experience*. New York: National Bureau of Economic Research.

———. 1973. Relative economic status and the American fertility swing. In *Family economic behavior: problems and prospects*, ed. Eleanor B. Sheldon. Philadelphia: J. B. Lippincott.

———. 1974. Does economic growth improve the human lot? In *Nations and households in economic growth: Essays in honor of Moses Abramovitz*, ed. Paul A. David and Melvin W. Reder. New York: Academic Press.

———. 1975. An economic framework for fertility analysis. *Studies in Family Planning* 6 (March): 54–63.

———. 1978. The economics and sociology of fertility: A synthesis. In *Historical studies of changing fertility*, ed. Charles Tilly. Princeton: Princeton University Press.

El-Safty, Ahman E. 1976*a*. Adaptive behavior, demand and preferences. *Journal of Economic Theory* 13 (October): 298–318.

———. 1976*b*. Adaptive behavior and the existence of Weiszäcker's long-run indifference curves. *Journal of Economic Theory* 13 (October): 319–28.

138

Frisch, Rose E. 1975. Demographic implications of the biological determinants of female fecundity. *Social Biology* 22 (spring): 17–22.

Hammond, Peter J. 1976. Endogenous tastes and stable long-run choice. *Journal of Economic Theory* 13 (October): 329–40.

Heer, David M. 1967. Fertility differences in Andean countries: A reply to W. H. James. *Population Studies* 21 (July): 71–73.

Heisel, Donald F. 1968. *Fertility limitation among women in rural Kenya.* Nairobi: University College, Institute for Development Studies. Discussion Paper 62.

Henin, R. A. 1972. The level and trend of fertility in the Sudan. In *Population growth and economic development in Africa*, ed. S. H. Ominde and C. N. Ejiogu. London: Heinemann.

Hossein, Askari, and Cummings, John Thomas. 1977. Estimating agricultural supply response with the Nerlove model: A survey. *International Economic Review* 18 (June): 257–92.

Howe, Howard; Pollak, Robert A.; and Wales, Terence J. 1977. Theory and time series estimation of the quadratic expenditure system. University of Pennsylvania Discussion Paper #388; *Econometrica*, forthcoming.

Hull, T. H., and Hull, V. J. 1977. The relation of economic class and fertility: An analysis of some Indonesian data. *Population Studies* 31 (March): 43–58.

Jain, Anrudh K., and Sun, T. H. 1972. Interrelationship between sociodemographic factors, lactation, and postpartum amenorrhea. *Demography India* 1 (October): 1–15.

James, William H. 1966. The effect of altitude on fertility in Andean countries. *Population Studies* 20 (July): 97–101.

Keeley, Michael. 1975. A comment on "An interpretation of the economic theory of fertility." *Journal of Economic Literature* 13 (June): 461–67.

Kelley, Allen C. 1976. Review of *Economics of the family: Marriage, children and human capital*, edited by Theodore W. Schultz. *Journal of Economic Literature* 14 (June): 516–20.

Keyfitz, Nathan. 1971. How birth control affects births. *Social Biology* 18 (June): 109–21.

Knodel, John. 1977. Family limitation and the fertility transition: Evidence from the age patterns of fertility in Europe and Asia. *Population Studies* 31 (July): 219–49.

Lancaster, Kelvin J. 1966*a*. Change and innovation in the technology of consumption. *American Economic Review* 56 (May): 14–23.

———. 1966*b*. A new approach to consumer theory. *Journal of Political Economy* 74: 132–57.

———. 1971. *Consumer demand: A new approach.* New York: Columbia University Press.

Lapham, Robert J. 1970. Morocco: Family planning attitudes, knowledge and practice in the Sais Plain. *Studies in Family Planning*, no. 58 (October), pp. 11–22.

Lee, Ronald D. 1976. Demographic forecasting and the Easterlin hypothesis. *Population and Development Review* 2 (September/December): 459–68.

———. 1977. *Fluctuations in U.S. fertility, age structure, and income.* Final Contract Report to NICHHD (August).

Leibenstein, Harvey. 1974*a*. *Economic backwardness and economic growth.* New York: John Wiley.

———. 1974*b*. An interpretation of the economic theory of fertility: Promising path or blind alley? *Journal of Economic Literature* 12 (June): 457–79.

———. 1976. *Beyond economic man.* Cambridge: Harvard University Press.

Leridon, Henri. 1977. *Human fertility: The basic components.* Translated by Judith F. Helzner. Chicago: University of Chicago Press.

Lindert, Peter. 1978. Fertility and scarcity in America. Princeton: Princeton University Press.

Mauldin, W. Parker. 1965. Fertility studies: Knowledge, attitude, and practice. *Studies in Family Planning*, no. 7 (June), pp. 1–10.

Michael, Robert T. 1973. Education in nonmarket production. *Journal of Political Economy* 81 (March/April): 306–27.

Michael, Robert T., and Becker, Gary S. 1973. On the new theory of consumer behavior. *Swedish Journal of Economics* 75 (December): 378–96.

Michael, Robert T., and Willis, Robert J. 1976. Contraception and fertility: Household production under uncertainty. In *Household production and consumption*, Conference on Research in Income and Wealth, pp. 27–94. New York: National Bureau of Economic Research.

Nerlove, Marc. 1974. Household and economy: Toward a new theory of population and economic growth. *Journal of Political Economy* 82 (March/April): S200–218.

Nortman, Dorothy. 1977. Changing contraceptive patterns: A global perspective. *Population Bulletin*, vol. 32, no. 3.

Palmore, James A. 1969. Malaysia: The west Malaysian family survey, 1966–67. *Studies in Family Planning* no. 40 (April): pp. 11–20.

Peng, J. Y. 1965. Thailand: Family growth in Pho-tharam District. *Studies in Family Planning* no. 8 (October), pp. 1–7.

Pollak, Robert A. 1970. Habit formation and dynamic demand functions. *Journal of Political Economy* 78 (July/August): 745–63.

———. 1971. Additive utility functions and linear Engel curves. *Review of Economic Studies* 38 (October): 401–14.

139

———. 1976*a*. Habit formation and long-run utility functions. *Journal of Economic Theory* 13 (October): 272–97.

———. 1976*b*. Interdependent preferences. *American Economic Review* 66 (June): 309–20.

———. 1977. Price dependent preferences. *American Economic Review* 65 (March): 64–75.

Pollak, Robert A., and Wachter, M. L. 1975. The relevance of the household production function and its implications for the allocation of time. *Journal of Political Economy* 83 (April): 255–77.

———. 1977. Reply: Pollak and Wachter on the household production function approach. *Journal of Political Economy* 85 (October): 1083–86.

Pollak, Robert A., and Wales, T. J. 1969. Estimation of the linear expenditure system. *Econometrica* 37 (October): 611–28.

Romaniuc, A. 1968. Infertility in tropical Africa. In *The population of tropical Africa*, ed. J. C. Caldwell and C. Okonjo, pp. 214–24. Princeton: Princeton University Press.

———. 1974. Modernization and fertility: The case of the James Bay Indians. *Canadian Review of Sociology and Anthropology* 11 (no. 4): 344–59.

Samuelson, Paul A. 1956. Social indifference curves. *Quarterly Journal of Economics* 70 (February): 1–22.

Sanderson, Warren C. 1974. Does the theory of demand need the maximum principle? In *Nations and households in economic growth: Essays in honor of Moses Abramovitz*, ed. Paul A. David and Melvin W. Reder. New York: Academic Press.

Schultz, T. Paul. 1976. Determinants of fertility: A micro-economic model of choice. In *Economic factors in population growth*, ed. Ansley J. Coale, pp. 89–124. New York: Halsted Press.

Schultz, Theodore W., ed. 1974. *Economics of the family: Marriage, children and human capital*. Chicago and London: NBER.

Sen, Amartya. 1973. *On economic inequality*. Oxford: Oxford University Press.

Singarimbun, Masri, and Manning, Chris. 1976. Breastfeeding, amenorrhea, and abstinence in a Javanese village: A case study of Mojolama. *Studies in Family Planning* 7 (June): 175–79.

Stigler, George J., and Becker, Garry S. 1977 De gustibus non est disputandum. *American Economic Review* 67 (March): 76–90.

Tabbarah, Riad B. 1971. Toward a theory of demographic development. *Economic Development and Cultural Change* 19 (January): 257–77.

Taylor, C. E.; Newman, J. S.; and Kelly, Narindar U. 1976. The child survival hypothesis. *Population Studies* 30 (July): 263–78.

United Nations Department of Economic and Social Affairs. 1961. *The Mysore population study*. Population Studies no. 34. New York: United Nations.

von Weiszäcker, Carl Christian. 1971. Notes on endogenous change of tastes. *Journal of Economic Theory* 3 (December): 345–72.

Wachter, Michael L. 1972*a*. Government policy towards the fertility of the poor. Fels Center of Government Discussion Paper no. 19.

———. 1972*b*. A labor supply model for secondary workers. *Review of Economics and Statistics* 54 (May): 141–51.

———. 1974. A new approach to the equilibrium labor force. *Economica* 41 (February): 35–51.

———. 1975. A time series fertility equation: The potential for a baby boom in the 1980s. *International Economic Review* 16 (October): 609–24.

Wales, Terence J. 1971. A generalized linear expenditure model of the demand for non-durable goods in Canada. *Canadian Journal of Economics* 4 (November): 471–84.

140

141

On the Economics of Intergenerational Control

Donald O. Parsons

Although economists have commonly identified the family or household as a decision-making unit, few economists have examined the implicit and explicit exchange system that orders family life.[1] This neglect no doubt partially arises from the economist's natural reluctance to deal with exchanges that lack observable transaction prices and quantities. Nonetheless intergenerational exchanges may have powerful effects on the long-run growth of the economy in terms both of physical and human wealth accumulation per capita and of population growth.[2] This article explores the choice structure underlying the relations between the generations with particular attention to the interaction of intergenerational wealth flows and parental control of the activities of offspring.

Central to the analysis is the treatment of families as collections of individuals with potentially strong interdependence of utilities yet with distinct preferences and economic constraints.[3] Parental control of offspring is not a given under this formulation but is an economic outcome, subject to shifts in market opportunities and the volume and form of intergenerational wealth flows. This view of the family process opens up a wide range of problems to economic analysis, including age of independence of offspring, extent of child labor, and pension value of offspring, and it recasts family decisions in such areas as schooling, marriage, and fertility in a potentially useful way.[4]

Economic analyses of intrafamily allocation have traditionally assumed, implicitly or explicitly, a parental control model of the household. Perhaps the most rigorous basis for such an assumption is Gary Becker's theorem on the intergenerational distribution of consumption in which

> a change in the distribution of family income between i and j has no effect at all on the consumption or welfare of either, as long as i continues to transfer resources to j. (Becker, 1974: 1076)

144

A corollary of this theorem is the celebrated, if infelicitously labeled "rotten kid theorem," namely that:

> if a head exists, other members also are motivated to maximize family income and consumption, even if their welfare depends on their own consumption alone. (Becker, 1974: 1080)

The underlying principle is that if the altruism and relative income of one family member are sufficiently large, this individual alone will determine the intrafamily distribution of consumption. A small variation in income sources among members will not affect consumption shares since the head will simply alter his gift-giving to reestablish the consumption shares he feels are appropriate.

These propositions, while providing considerable insight into the family allocation process, assume away a fundamental issue in intergenerational relations, the issue of intergenerational control: family headship is an economic outcome. If incomes among family members are approximately equal and/or altruistic feelings are limited, intrafamily transfers will not occur, a family head will not exist, and consumption shares will depend solely on individual income shares. Even in a family with a head, the identity of the head will and indeed must change over time. An individual will be a household head, controlling the consumption of all other members, for only a portion of his lifetime.

In an individual's early years, transfers must of necessity be from parent or grandparent to child. When the child matures, reaching an age at which he or she is capable of productive activity, the direction of intergenerational transfers may shift. The probability of child-to-parent transfers grows as younger members of the family contribute to the consumption of older and possibly retired family members. The identity of the head, should such an individual exist, and the allocation of consumption shares within the family may, therefore, shift over the life cycle as relative income shares change.

In the first section below, I discuss a formal model of family control and intergenerational consumption shares as a function of initial resource endowments and levels of altruism. A simple economy in which income is the only economic good is considered. A more general model of the family allocation process, which permits exchange as well as gift relationships among family members, is then described. A second good, labeled child services, is introduced; it is assumed that the younger generation holds property rights in this service flow. If the younger generation controls income as well as child services, a gift or altruism process will occur. If the older generation controls income, an exchange process within an altruistic framework will occur. A bargaining model of this family exchange is elaborated and implications for household structure are derived.

The relative incomes of family members are not fixed over time; neither are they independent of family decisions. The decision by the family to invest in the schooling of a child, for example, will have consequences for the relative incomes of parents and offspring at a later time. The older generation can maintain control of the physical assets of the family but, if intergenerational contracts are not feasible, cannot control the income generated by the younger generation's human capital. This conflict between total family income and intergenerational control is of more than theoretical interest. The rapid expansion of knowledge and consequently the demand for more highly educated workers in the twentieth century in the United States and other developed countries may have induced sufficiently large shifts in schooling investments so as to alter intergenerational control and consumption shares.[5] In the second section below, I describe a formal model of the older generation's investment and asset decisions when such decisions can result in the future loss of control of the family's wealth and in adverse shifts in the parents' consumption share.

"Gifts," family control, and intergenerational consumption shares

Becker's theorems on intergenerational consumption shares assume the existence of a family head who provides gifts to all other family members and who, through adjustments of the level of transfers or gift giving, determines intrafamily consumption shares. More generally, of course, such an individual need not exist and family consumption shares may be determined solely by income shares. Indeed the position of family head is a function of family income shares. A simple model of the determination of headship and consumption shares within the family will illustrate this process.[6]

Relative income, consumption shares, and control in the family

Imagine an economy in which three generations are alive at any one time. Denote the three generations as young, adult, and aged. Individuals in the young generation are assumed to lack the physical and emotional strength required for family control; the competition for control of family consumption allocations therefore is between the adults and the aged. The analysis will focus on these two decision agents, although the education of the young plays a crucial, if indirect role in the model.

The distribution of consumption between the generations depends on the initial endowments of income between the generations and on the altruistic impulses each generation has toward the other. One of three economic relationships—adult headship, economic independence, or aged headship—binds the generations, and the regime that prevails depends on the relative wealth of the generations and on the degree of altruism that binds the family. Small variations in wealth endowments do not affect consumption in the adult headship regime or in the aged headship regime. Large variations in wealth endowments may, however, alter consumption shares and even the identity of the family head. Even small variations in wealth shares will affect consumption shares in the intermediate state in which no "head" exists.[7]

The simple insight that intergenerational consumption shares are dependent on wealth shares and on the extent of intrafamily altruism has a variety of potential applications. Neher (1971), for example, develops a model of the pension motive for fertility that assumes all family members share equally in family income. The current model offers the possibility of generalizing such a model. Assuming intergenerational contracts are not feasible, the pension motive for children depends on the asset endowments of the family and on the level of altruism the children feel toward the parents and consequently on the willingness of the children to transfer resources to the aged. This model may also provide insight into the presence or absence of alternative institutional arrangements for the provision of consumption for the aged. Goode (1963: 354) presents an interesting perspective on the interaction of family altruism and government programs when he notes that among the Japanese: "Filial piety is expressed by caring for the parents, but it is also a source of satisfaction to both the old and the young. As a result, Japan has relatively little development of old-age pensions and social security systems." Finally the model may have macroeconomic insights; the neutrality of government debt, for example, depends critically on the independence of intergenerational consumption shares from intergenerational wealth endowments (Barro, 1974).

The intrafamily allocation of goods and services: a bargaining model

The issue of bargaining within the family does not arise in the wealth model since, at any one time, the single commodity (income) can be rationally transferred only in one direction and contracts over time are assumed to be infeasible. Goods and services other than income, of course, are acquired and exchanged within the family, and this raises the possibility of exchange and bargaining within the family. Important in the family economy are child services, where "child" is used to denote not age but an intergenerational relationship. Parents appear to wish to control the behavior of their offspring both as children and as adults (when they themselves are aged). In this section the consequences of the addition to the model of child services as an economic commodity that may be exchanged for income in the family economy are explored.

The model differs little from the previous one if the adult generation controls both wealth and services. Each is allocated to the aged according to the altruistic impulses of the adult generation in this case. Small variations in wealth endowments between the generations do not affect intergenerational consumption shares of utility.

If the aged control the preponderance of physical wealth, however, the whole intrafamily allocation structure changes. The adult generation and the aged exchange commodities in such circumstances; a gifts model is no longer appropriate. Analysis of the determination of utility inequality between the generations requires specification of the bilateral exchange process. A Nash bargaining process is assumed. In this exchange model, variations in income shares induce variations in intergenerational utility shares; no headship in Becker's sense exists. Control by the aged is not absolute. More importantly, the most obvious manifestation of intergenerational control, the size of the flow of "child" services to the aged, is shown to depend positively on the relative control of wealth by the aged.[8]

Empirical verification of the more general child services model is difficult since child services are hard to define, much less measure. Nonetheless a variety of empirical phenomena appear broadly consistent with the model. One would expect, for example, that household structures with two or more adult generations would be more prevalent in situations in which older generations disproportionately control economic wealth and therefore receive large amounts of child services, the provision of which is facilitated by joint tenancy.[9] Organizational costs exist in extended families in which productive activities must be assigned to family members and family output distributed among members.[10] As Goode (1963: 128) notes, "The careful calculation by each adult member in each conjugal unit as to who is contributing or consuming the most often makes relations within the extended family strained." At a deeper level there must be a large psychic cost to large families as couples intrude on one another's lives (Demos, 1972: 563). If the flow of services of the offspring to the parents is sufficiently large, however, the bearing of these organizational costs may be worthwhile; more modest service exchanges are unlikely to justify this form of household structure.

This argument is consistent with the secular trend in household structure that has accompanied the shift to a human capital economy and its likely shift in control of wealth to the younger generation. Although recent historical research indicates that large patriarchal or matriarchal family groupings were less frequent in the past than one might suppose from the folk wisdom, the decline in joint tenancy has been a strong and continuing process over the past few centuries as human capital has grown in importance. The trend has reached its natural limit in modern, high-human-capital economies: virtually all married couples in the industrial countries maintain their own households. Carter and Glick (1970: 149), for example, report that the portion of all married couples in the United States not living in their own household declined from 6.8 percent in 1940 to 1.9 percent in 1965.

Historically, the evidence is fairly persuasive that the extended family is more likely among the wealthy and/or in rural areas (where land is a large part of family wealth). Evidence on the positive relationship between wealth and the frequency of extended families, generally defined as two or more nuclear units, can be found in as widely dispersed times and areas as preindustrial England, mid-nineteenth century Lancashire, mid-fifteenth century Tuscany, mid-eighteenth century Netherlands, and nineteenth and twentieth century United States.[11] Many of these same studies report a consistently higher prevalence of extended families in rural areas than urban.

Similar findings have been reported in the sociological literature across a wide variety of cultures. Goody (1972) cites evidence on such relationships in India and Africa as does Goode (1963) across a wide range of cultures. The parental control hypothesis also resolves in a straightforward way the "paradox" raised by Goode (pp. 12–13), namely that industrialization "requires" the nuclear family; yet the wealthy, those one would expect to be most attuned to the industrial society, are more likely than others to live in extended families.

146

Since the shift from agriculture to industry led to a marked reduction in the ratio of physical to human capital (and therefore a reduction in parental control), it would be expected to induce a shift toward nuclear families in general. Among families, however, the wealthy are likely to be able to afford the purchase of child obedience required for extended families. The physical wealth they intend to transfer to children is more likely to be large compared to the child's labor earnings than is true for the poor.

A less obvious prediction of the model is that extended families should be less common in more "primitive" economies as well, in particular those based on labor-intensive activities such as hunting and gathering. This prediction is born out by the research of Nimkoff and Middleton (1960). Using the data collected on human cultures by George Murdock in his World Ethnographic Sample, Nimkoff and Middleton found among other things that, in the 380 cultures where agriculture either dominated or co-dominated the economy, 228 or 60 percent reported extended families as the dominant family structure. In the 86 cultures where hunting either dominated or co-dominated the economy (excluding co-domination with agriculture), only 27 or 31 percent reported the extended family was the norm (p. 217, Table 2).

The same parental control hypothesis applies to other family processes and decisions as well. The choice of mate and age at marriage are, for example, amenable to a similar sort of analysis. Goode argues that early marriage is itself a parental control device:

> A system that permits marriage just before puberty or in the earlier years of adolescence fits well with a system of high control by elders over mate choice. If the young girl or boy can be kept from members of the opposite sex so that no emotional attachments can be formed, and marriages are arranged early, the bride and groom have neither independent emotional nor independent socioeconomic resources with which to oppose such decisions. Thus it was to the interest of elders who wished to maintain the traditional family system to arrange the marriage at as early an age as possible. (1963: 105)

Marriage is, however, not necessarily a one-point-in-time decision: divorce, desertion, or simple neglect can quickly change the content and the form of the marriage. Parental choice of marriage partner should therefore be correlated with extent of family since the latter is itself a manifestation of parental control. Although objective data on who controls marriage choice are obviously not as easy to come by as data on family form, it is interesting that a number of observers, including Engels and Goode, have noted that in the West the children of wealthy parents are generally the least free in their choice of mate.[12]

The degree of parental control of marriage and the extent of family, of course, are not independent indicators of parental control. The parents will demand greater voice in the choice of mate if the family is extended since they must then live with the choice. Similarly if the marriage customs are patrilocal, the parents should be more interested in the son's mate than the daughter's.

146

Schooling and the intrafamily distribution of consumption

> The industrial world is not based on family ownership of land; it has become the first civilization not based upon landholding. . . . The transformation in modern times is that the elders give the young man and woman an education. This is their inheritance, their dower or dowery. Elders are left with only education, maintenance, and love as levers for controlling their young. (Goode, 1963: 375)

The distribution of asset ownership or property rights between the generations may affect the distribution of consumption between the generations and even the identity of the family head, if one exists. Over time property rights in assets are not exogenously determined (as for example in Samuelson's overlapping generations model of the family economy) but are themselves subject to choices made within the family. Intergenerational transfers must flow initially from older to younger generation. Assets of all kinds are initially "owned" by the older generation.

This observation raises the issue of intergenerational contracting: whether a contract can be written between parents and child that will insure a future payback of parental transfers earlier in life. With an appropriate set of assets and property rights in those assets, it may be possible for parents to control children throughout the parents' lifetime and, with appropriate inheritance rules, beyond. Contracting for physical asset exchanges over time, for example, does occur between the generations. In the area of pensions and retirement, the older generation may yield up assets in order to secure (remaining) lifetime care from the younger generation; in other words it may in effect buy an annuity from the younger generation.[13] Indeed old age care in return for predeath disposal of an estate or implicit or explicit bequest rights is a continuing phenomenon in the Western world. Homans (1941) in his excellent account of English village life in the thirteenth century found contracts of this sort to be widespread. One of the best examples was entered in the court rolls of one of the manors of Ramsey Abbey in the year 1294:

> Elyas of Brickendon surrendered in full court one messuage and a half-yardland at rent in Cranfield, with a wood appurtaining thereto and with all its other appurtenances, and three acres of forland in the same, to the use of John, his son. . . . The aforesaid John will find for the same Elyas and Christiana his wife honorable sustenance in food and drink so long as they live, and they will dwell with the aforesaid John in lodging on the chief messuage. And if it chance (which God forbid) that quarrels and discords arise in time to come between the parties, so that they cannot dwell together peaceably in one house, the aforesaid John will find for the same Elyas and Christiana, or to whichever one of them lives the longer, six quarters of hard corn at Michaelmas, namely three quarters of wheat, a quarter and a half of barley, a quarter and a half of beans and peas, and a quarter of oats. (Homans, 1941: 149)

147

The keen eye of the father for human contingencies—default risk in the modern jargon—as he prepares to transfer his farming rights to his son in return for care for himself and wife does not speak well for his faith in the social compact of the affections of his son.[14]

Nor are such exchanges purely historical phenomena. An analysis of inheritance behavior by Sussman, Cates, and Smith (1970: 149–156) suggests that a major reason for inequality of bequests among children is the practice of implicit maintenance payments, in which one child provides indefinite support to the parent in return for a promised lump sum inheritance. Discord among siblings over such a contract may occur if the parent dies unusually early. Outside the family many nursing homes for the aged are organized on a lump sum basis.

Intergenerational contracts may be costly to enforce, however, and in some cases enforcement may not be feasible. Shoup (1966) finds that few older persons transfer physical wealth during life, even though taxes on gifts are substantially less than on bequests.[15] He attributes this to a desire for parental control, with parents implicitly assuming that an appropriate service contract is not feasible. Contracting difficulties are believed to be particularly prevalent in the financing of human capital. The education of one's offspring has one undesirable property as a pure investment activity by the adult generation: legally enforceable repayment is limited. Schooling activities are optimally carried out when the individual is young, which makes formal contracting between parent and child difficult. Contracts with the very young would be invalid, whatever their nature. Even among adults, social restrictions are imposed on work or income contracts of long duration. At a more practical level, educated young people tend to be quite mobile and therefore offer poor collateral.

The share of wealth invested in human capital, with its corresponding limitation on intergenerational contracting, has expanded rapidly in this century. In the United States in 1900, 50.5 percent of the population aged 5 to 19 years old were enrolled in school. In 1970, 90.6 percent were. This rise was no doubt induced by the explosion of knowledge in this century and the increasing relative importance of human capital in production processes (Adams, 1980). As the quotation from Goode at the start of this section suggests, the shift from a physical-capital-and-manual-labor-based economy to a human-capital-based economy may have profoundly affected family control and the distribution of consumption within the family.

School attendance of offspring is, of course, a choice variable for the parents. With intergenerational contracts not feasible, a decision by the adult generation to invest family assets in the schooling of the young carries with it the loss of direct control over family assets and perhaps of consumption when aged. The alternative of holding the asset in physical form permits the old to retain effective property rights in the asset, but only at a price if the rate of return on human investments is relatively high. The older generation, in making the decision whether to invest in their offspring's education, face 148 a trade-off between high total family income and a less attractive distribution of subsequent consumption within the family.

The joint determination of offspring schooling and subsequent total family income and intergenerational consumption shares is explored in Parsons (1983) under the assumption that intergenerational contracts are infeasible. In the wealth model developed there, a sufficiently high rate of return to schooling will lead to a schooling choice that is the joint family income maximum, with the total family wealth effect outweighing the adverse consumption share effect in the parents' view. The higher the level of family altruism, the lower the rate of return necessary to secure the joint family income maximizing level of schooling. The schooling choice will also be affected by the degree of altruism that binds the family since greater family altruism increases the likelihood that children will transfer wealth to the parents in the future and makes the parents less unhappy with any adverse consumption shift that does occur.[16] An empirical paper by Goldin and Parsons (1983) suggests, however, that at least among one group of families, working-class families in the United States in 1890, the extent of parental altruism was relatively small. The established level of altruism suggests that the parental view of the returns to schooling may be a great deal less than the offspring's view (or that traditionally measured by human capital economists) because of the adverse consumption share effect induced by the intergenerational reallocation of property rights.[17]

The possibility of control oscillating between the adult and aged generations over time may arise in the model outlined here. The discussion has been limited to the conflict for control between the adult generation and the aged since the young generation is assumed to be insufficiently mature physically and emotionally to bid for control of the family. The aged may transfer resources directly to the young for education (thereby eliminating the possibility of the current adult generation ever controlling family assets) if the aged care only about maximizing aggregate family income in the next period. The distribution of family consumption will not directly concern them since they themselves will be dead. In the next generation, however, the family assets would be controlled by the adult population in whom the aged had previously invested and they will be directly interested in consumption shares in the next period. As a result they may choose to invest disproportionately in controllable physical assets until they themselves are aged, at which time they may invest in schooling for their grandchildren.

Conclusion

The recognition that intergenerational transfers involve a complex mixture of gifts and exchanges, induced by both altruism and self-interest, suggests that a variety of economic and family models may be fruitfully recast. Macroeconomic models based on unidirectional intergenerational concepts—for example, Barro's neutrality of government debt argument—may be usefully reconsidered in a broader framework. The simplest sorts of dynamic intergenerational models of education and fertility (Razin and Ben-Zion, 1975; Becker and Tomes, 1979) also assume a unidimensional flow of resources from the old to the young. Pension models of fertility (Neher, 1971) assume rather the opposite. Obviously integration of these models is needed. Goldin and Parsons (1983) have undertaken an empirical study of schooling and child labor among working-class families in the United States in 1890 that attempts to incorporate both parental altruism and self-interest into the analysis. Much more remains to be done.

The models discussed in this article also point the way to a more complete analysis of the system of family, government, and market. The major expansion of knowledge in the twentieth century in the United States and elsewhere no doubt induced the massive expansion in schooling as parents altruistically invested in the human capital of their offspring. It is likely, however, that this shift in family asset position made children a less attractive pension instrument, which in turn plausibly reduced the demand for children and at the same time increased the demand for government old age Social Security. The long-run "crisis" in Social Security financing—the growing proportion of aged to young workers combined with generous transfers to the aged—may not be happenstance but rather internal to the system, the systematic evolution of altruism and self-interest in a human-capital-based economy.

Notes

T. W. Schultz, Yoram Weiss, and Victor Fuchs provided valuable comments on earlier drafts of this paper.

1 Research on intrafamily allocation systems has increased substantially in the last decade. Becker (1974) and Ben-Porath (1980) provide some general considerations of family processes. Ishikawa (1975), Becker and Tomes (1976, 1979), Adams (1980a), Tomes (1981), and Sheshinski and Weiss (1982) have contributed importantly to our understanding of the intergenerational allocation process. Among contributions in other disciplines in the same spirit as the present paper are Goode (1963) and a number of contributions in Laslett (1972).

2 The interaction between wealth allocation and family fertility choice is considered in Razin and Ben-Zion (1975), Caldwell (1976, 1980), and Willis (1981). Barro (1974) pointed out the crucial importance of the family allocation process in macroeconomic behavior.

3 Samuelson (1956) has noted that groups, even the family, will not behave as a consistent consumption unit except under severe assumptions. Becker (1974), however, has made the important point that these assumptions are not as severe as one would first think in the case of the family. For the analysis below, however, the independent choice agents within the family become a critical feature. See also Nerlove (1974). For applications of game theory to family decision-making, see the interesting work by Manser and Brown (1980) and McElroy and Horney (1978).

4 This paper focuses purely on the interaction between parents, considered as a unit, and children, an interaction suggested for study by T. W. Schultz (1973). For the interaction between parents see Becker (1973), Parsons (1980), Manser and Brown (1980), and McElroy and Horney (1978).

5 Adams (1980b) estimates the magnitude of the relative growth of human capital in the United States in the twentieth century.

6 The model is described verbally. A formal derivation can be found in Parsons (1983).

7 Each generation is assumed to have a log linear utility function with own consump-

tion and the other's utility as arguments; formally $U = C_2^\alpha V^\gamma$ and $V = C_1^\alpha U^\gamma$ where U is the aged's utility index, V that of the adult generation. Denoting the aged's share of total family income as k_2, the aged will be the family head if $k_2 > 1/(1 + \gamma)$, the adult generation will be the head if $k_2 < \gamma/(1 + \gamma)$, and the two generations will be independent (with consumption shares equal to income shares) if $\gamma/(1 + \gamma) \leqslant k_2 \leqslant 1/(1 + \gamma)$. The relationships described in the text are outcomes of the formal model (Parsons, 1983).

8 Each generation is assumed to have a log linear utility function with offspring services (S) as an additional argument; formally $U = C_2^\alpha S_2^\beta V^\gamma$ and $V = C_1^\alpha S_1^\beta U^\gamma$. A Nash bargaining process is assumed with the threat point determined by the pure altruistic allocation. The Nash bargaining solution is the only outcome that meets a set of four plausible axioms (criteria), including Pareto optimality, symmetry, invariance with respect to utility functions, and independence of irrelevant alternatives (Luce and Raiffa, 1957: 124–143). Independence between the generations will not occur in this economy, although an interval of exchange (with no family head) will exist within which resource ownership variation will alter intergenerational consumption. The variation in consumption will not be dollar for dollar with variation in income but will depend on the relative importance of consumption goods and child services in the utility structure, that is, on the magnitude of α and β. Again, a formal statement of the model can be found in Parsons (1983).

9 Goode (1963) extends this hypothesis to include parental control of jobs as well as physical and human assets, an extension he feels applies usefully to the Japanese in particular. Hammel (1972) argues that tax incentives may explain one well-documented instance in Europe of truly large extended families, the Balkan Zadruga. Apparently major taxes were imposed per household rather than per adult.

10 Alchian and Demsetz (1972) provide a readily accessible account of the problems of shirking in the firm and the consequence of shirking on firm size.

11 See respectively Laslett (1972), Anderson (1972), Klapisch (1972), van der Woude (1972), and Pryor (1972).

12 Engels (1972: 136) moreover makes the interesting argument that laws and wealth are substitutes in the control by parents of their children's marriage choice. He argues that where parents transfer sufficient wealth to the child, laws on marriage age and parental consent are unnecessary. Only when parents have little property control does Engels expect parents to use laws.

13 Kotlikoff and Spivak (1981) present a formal analysis of the family as an insurance organization.

14 See Demos (1970, particularly pp. 164–170) for similar observations from the Colonial American experience.

15 Adams (1979) makes the point that certain assets with unrealized capital gains may be transferred more inexpensively at death than during life. This observation does not negate Shoup's point as long as other assets without this tax advantage are also transferred by bequest.

16 The single-good economy is considered in a two-period framework, with the controlling generation in the first generation choosing between two types of investments: physical assets, which yield no return or interest, and human capital investments in the young, which yield a positive return $\delta - 1 > 0$. The joint wealth maximizing strategy for the family is to devote all resources to human capital. The older generation will, however, find it optimal to consider adverse second period consumption effects and will choose the joint wealth maximizing strategy if and only if $\delta \geqslant (1/\gamma)^{1-\gamma}$. Otherwise schooling choice will be sensitive to the level of altruism; the share of assets devoted to schooling will be $\gamma/(1 + \gamma)$.

17 Ishikawa (1975) considers this issue at length.

151

References

Adams, James D. 1979. "Equilization of true gift and estate tax rates," *Journal of Public Economics* 9: 59–71.

———. 1980a. "Personal wealth transfers," *The Quarterly Journal of Economics* (August): 159–179.

———. 1980b. "Relative capital formation in the United States," *Journal of Political Economy* 88 (June): 561–577.

Alchian, A. A., and H. Demsetz. 1972. "Production, information costs, and economic organization," *American Economic Review* 62 (December): 777–795.

Anderson, Michael. 1972. "Household structure and the industrial revolution; mid-nineteenth-century Preston in comparative perspective," in *Household and Family in Past Time*, ed. Peter Laslett. London: Cambridge University Press, pp. 215–236.

Barro, Robert J. 1974. "Are government bonds net wealth?," *Journal of Political Economy* 82 (December): 1095–1117.

Becker, Gary S. 1973. "A theory of marriage, Part I," *Journal of Political Economy* 81 (July/August): 813–846.

———. 1974. "A theory of social interactions," *Journal of Political Economy* 82 (November/December): 1063–1094.

———, and Nigel Tomes. 1976. "Child endowments and the quantity and quality of children," *Journal of Political Economy* 84 (August): S143–S162.

———, and Nigel Tomes. 1979. "An equilibrium theory of the distribution of income and intergenerational mobility," *Journal of Political Economy* 87: 1153–1189.

Ben-Porath, Yoram. 1980. "The F-connection: Families, friends, and firms and the organization of exchange," *Population and Development Review* 6, no. 1 (March): 1–30.

Caldwell, John C. 1976. "Toward a restatement of demographic transition theory," *Population and Development Review* 2, no. 3–4 (September/December): 321–326.

———. 1980. "Mass education as a determinant of the timing of fertility decline," *Population and Development Review* 6, no. 2 (June): 225–255.

Carter, Hugh, and Paul C. Glick. 1980. *Marriage and Divorce*. Cambridge, Mass.: Harvard University Press.

Demos, John. 1970. *A Little Commonwealth, Family Life in Plymouth Colony*. New York: Oxford University Press.

———. 1972. "Demography and psychology in the historical study of family-life: A personal report," in *Household and Family in Past Time*, ed. Peter Laslett. London: Cambridge University Press, pp. 561–570.

Engels, Friedrich. 1972. *The Origin of the Family, Private Property and the State*. New York: International Publishers.

Goldin, Claudia, and Donald O. Parsons. 1983. "Industrialization, child labor, and family economic well-being," mimeo.

Goode, J. William. 1963. *World Revolution and Family Patterns*. New York: The Free Press.

Goody, Jack. 1972. "The evolution of the family," in *Household and Family in Past Time*, ed. Peter Laslett. London: Cambridge University Press, pp. 103–124.

Hammel, E. A. 1972. "The Zadruga as process," in *Household and Family in Past Time*, ed. Peter Laslett. London: Cambridge University Press, pp. 335–374.

Homans, George C. 1941. *English Villagers of the Thirteenth Century*. Cambridge, Mass.: Harvard University Press.

Ishikawa, Tsuneo. 1975. "Family structures and family values in the theory of income distribution," *Journal of Political Economy* 83 (October): 987–1008.

152

Klapisch, Christiane. 1972. "Household and family in Tuscany in 1427," in *Household and Family in Past Time*, ed. Peter Laslett. London: Cambridge University Press, pp. 267–282.

Kotlikoff, Laurence J., and Avia Spivak. 1981. "The family as an incomplete annuities market," *Journal of Political Economy* 89 (April): 372–391.

Laslett, Peter (ed.). 1972. *Household and Family in Past Time*. London: Cambridge University Press.

———. 1972. "Mean household size in England since the sixteenth century," in *Household and Family in Past Time*, ed. Peter Laslett. London: Cambridge University Press, pp. 125–158.

Luce, R. Duncan, and Howard Raiffa. 1957. *Games and Decisions*. New York: Wiley.

Manser, Marilyn, and Murray Brown. 1980. "Marriage and household decision-making: A bargaining analysis," *International Economic Review* 21 (February): 31–44.

McElroy, Marjorie B., and Mary J. Horney. 1978. "Nash-bargained household decisions: Toward a generalization of the theory of demand," mimeo.

Neher, Philip A. 1971. "Peasants, procreation, and pensions," *American Economic Review* 61 (June): 380–389.

Nerlove, Marc. 1974. "Household and economy: Toward a new theory of population and economic growth," *Journal of Political Economy* 82, Part II (April): S200–S218.

Nimkoff, M. F., and Russell Middleton. 1960. "Types of family and types of economy," *The American Journal of Sociology* 66 (November): 215–225.

Parsons, D. O. 1980. "The marriage market and female economic well-being," *Journal of Mathematical Sociology* 7: 113–138.

———. 1983. "The economics of intergenerational control," Ohio State University Economics Department Working Paper 83-13.

Pryor, Edward T. 1972. "Rhode Island family structure: 1875 and 1960," in *Household and Family in Past Time*, ed. Peter Laslett. London: Cambridge University Press, pp. 571–589.

Razin, Assaf, and Uri Ben-Zion. 1975. "An intergenerational model of population growth," *American Economic Review* 65 (December): 923–933.

Samuelson, Paul A. 1956. "Social indifference curves," *Quarterly Journal of Economics* 70 (February): 1–22.

———. 1958. "Exact consumption-loan model of interest with or without the contrivance of money," *Journal of Political Economy* 66 (November/December): 467–482.

Schultz, T. W. 1973. "The value of children: An economic perspective," *Journal of Political Economy* 81, Part II (April): S2–S13.

Sheshinski, Eytan, and Yoram Weiss. 1982. "Inequality within and between families," *Journal of Political Economy* 90 (February): 105–127.

Shoup, Carl S. 1966. *Federal Estate and Gift Taxes*. Washington, D.C.: The Brookings Institution.

Sussman, Marvin B., Judith N. Cates, and David T. Smith. 1970. *The Family and Inheritance*. New York: Russell Sage Foundation.

Tomes, Nigel. 1981. "The family, inheritance, and the intergenerational transmission of inequality," *Journal of Political Economy* 89 (October): 928–958.

van der Woude, A. M. 1972. "Variations in the size and structure of the household in the United Provinces of the Netherlands in the seventeenth and eighteenth

153

154

centuries," in *Household and Family in Past Time*, ed. Peter Laslett. London: Cambridge University Press, pp. 299–318.

Willis, Robert J. 1981. "The direction of intergenerational transfers and demographic transition: The Caldwell hypothesis re-examined," in *Income Distribution and the Family*, ed. Yoram Ben-Porath, Supplement to *Population and Development Review* 8 (1982).

154

155

The Relevance of the 'Easterlin Hypothesis' and the 'New Home Economics' to Fertility Movements in Great Britain*

JOHN ERMISCH†

The hypothesis that a family's economic status relative to its aspirations (relative economic status) is an important determinant of its fertility behaviour has been developed and applied to the explanation of swings in American fertility by R. A. Easterlin.[1] Its apparent success in explaining the movements in American fertility has spawned a more rigorous econometric test of the hypothesis in the American context which has produced additional support for the hypothesis.[2] In addition, the hypothesis has provided the basis for some demographic forecasting methods, one of which is being experimented with by the U.S. Census Bureau,[3] as well as mathematical models exploring the dynamic consequences of links between fertility and age structure.[4] While the theory has not been refuted in the American context, a recent application by Butz and Ward[5] of a model derived from the 'new home economics' (pioneered by Becker and Mincer) strongly suggests that relative economic status is not the dominant factor in explaining fertility movements in the U.S.[6] Rather, both current men's earnings and women's wages operate independently in explaining movements in fertility, and in particular the decline in fertility is attributed to rising women's wages. In this paper, we explore the relevance of both the Easterlin hypothesis and the hypotheses derived from the 'new home economics' to the 1955-75 fertility swing in Great Britain. Although Easterlin had already applied his hypothesis to the explanation of the recent fertility swing in England and Wales[7] he used only one measure of relative economic status, 'relative cohort size'

* I would like to thank David Eversley, Elizabeth Overton, Thomas Weiss and the editors for their helpful comments on an earlier draft of this paper, and Peter Davies for his computational assistance. This work was supported by a grant from the Anglo-German Foundation for the Study of Industrial Society, and is part of a larger study of the causes and implications of the recent decline in fertility in the U.K. and the Federal Republic of Germany. The views expressed in this paper are solely those of the author.

† J. Ermisch is a Research Fellow at the Policy Studies Institute, London.

[1] R. A. Easterlin, *Population, Labor Force and Long Swings in Economic Growth* (New York: National Bureau of Economic Research, 1968); 'Towards a Socio-economic Theory of Fertility' in S. J. Behrman, L. Corsa and R. Freedman (eds), *Fertility and Family Planning–A World View* (Ann Arbor: University of Michigan Press, 1969); 'Relative Economic Status and the American Fertility Swing' in E. B. Sheldon (ed.), *Family Economic Behaviour: Problems and Prospects* (Toronto and Philadelphia: Lippincott, 1973).

[2] M. L. Wachter, 'A Time Series Fertility Equation: The Potential for a Baby Boom in the 1980's', *International Economic Review*, 16 (October 1975), pp. 609–624.

[3] For a summary of three different forecasting methods see R. D. Lee, 'Demographic Forecasting and the Easterlin Hypothesis', *Population and Development Review*, 2 (September/December 1976), pp. 459–468.

[4] For example, see R. D. Lee, 'The Formal Dynamics of Controlled Population and the Echo, the Boom and the Bust', *Demography*, 11 (November 1974), pp. 563–585; and P. A. Samuelson, 'An Economist's Non-linear Model of Self-Generated Fertility Waves', *Population Studies*, 30 (July 1976), pp. 243–247.

[5] W. P. Butz and M. P. Ward, *The Emergence of Countercyclical U.S. Fertility* (Santa Monica, Cal.: Rand Monograph R-1605-NIH, June 1977).

[6] The pioneering works of the 'new home economics' are: G. S. Becker, 'An Economic Analysis of Fertility' in Universities–National Bureau Committee for Economic Research, *Demographic and Economic Change in Developed Countries* (Princeton: Princeton University Press, 1960); J. Mincer, 'Market Prices, Opportunity Costs and Income Effects' in C. Christ *et al.* (eds), *Measurement in Economics: Studies in Mathematical Economics in Memory of Yehuda Grunfeld* (Palo Alto, Cal.: Stanford University Press, 1963); and G. S. Becker, 'A Theory of the Allocation of Time', *Economic Journal* 75 (September 1965), pp. 493–517. An explicit criticism of Easterlin's hypothesis appears in A. Sweezy, 'The Economic Explanation of Fertility Changes in the United States', *Population Studies*, 25 (July 1971), pp. 255–267; and Easterlin's reply is in 'Relative Economic Status and the American Fertility Swing,' *op. cit.* in footnote 1, p. 197. 155

[7] R. A. Easterlin and G. A. Condran, 'A Note on the Recent Fertility Swing in Australia, Canada, England and Wales and the United States', in H. Richards (ed.), *Population, Factor Movements and Economic Development: Studies Presented to Brinley Thomas* (Cardiff: University of Wales Press, 1976).

(defined below). We shall first investigate the relation between British fertility movements and movements in other measures of relative economic status available for Great Britain, and then estimate the parameters of and test a model incorporating the basic hypotheses of the 'new home economics'.

We find that we must reject the Easterlin hypothesis on the basis of the measures of relative economic status suggested by the work of Easterlin and Wachter. A variant of the Easterlin hypothesis suggested by Oppenheimer does receive some support from the available evidence, and the evidence provides strong support for the model of fertility behaviour derived from the 'new home economics', which emphasizes the distinction between the effects of changes in men's and women's real wages on fertility decisions. Fertility decline is attributed to rising women's wages and employment opportunities through their direct effect on the opportunity cost of time and children among working wives, and through their effect on the labour force participation rate of married women of childbearing age. The test of this model and the estimates of its parameters are not definitive however, because of poor supporting data and problems of statistical estimation. We nevertheless conclude that both this model and the Oppenheimer variant of the Easterlin hypothesis, as well as other elements of a more comprehensive economic theory of fertility, point to a continuation of low fertility with the possibility of a secular decline, with fertility approaching some lower asymptote. Before proceeding to test the Easterlin hypothesis and the 'new home economics' model we briefly discuss each, as well as other elements of a more comprehensive theory of fertility.

Easterlin regards fertility behaviour as the result of household choices in which the household's resources are weighed against preferences/aspirations which in turn are formed through the experience of young adults (i.e. those of marrying and childbearing ages) in their parents' households and during their previous years in the labour force. To quote from Easterlin[8]:

The basic idea is that if young men—the potential breadwinners of households—find it easy to make enough money to establish homes in the style desired by them and their actual or prospective brides then marriage and childbearing will be encouraged. On the other hand, if it is hard to earn enough to support the desired style of life, then the resulting economic stress will lead to deferment of marriage and, for those already married, to the use of contraceptive techniques to avoid childbearing, and perhaps also to the entry of wives into the labor market.

He postulates that the young generation's conception of the 'desired style of life' is primarily formed by their experiences as adolescents and young adults in the households of their parents. There are, of course, other influences on the formation of their material aspirations but Easterlin suggests that adolescent experience is dominant. A comparison of this desired standard of life with the current earnings of young men is important for their reproductive behaviour: failure to earn enough to attain the desired standard triggers the Malthusian defence mechanisms: postponement of marriage and contraception within marriage, and additionally entry by the wife into the labour force. The 'Easterlin hypothesis' is, therefore, the conjunction of two more *basic hypotheses*: that aspirations are primarily determined by adolescent experience and that fertility is a function of a young man's earnings relative to his aspirations. Since the living standard during a young man's adolescence in his parents' household is mainly determined by his father's income, Easterlin has employed the ratio of a young man's income to that of his father as a measure of a young man's earnings relative to aspirations, or 'relative economic status'. Changes in relative economic status thus depend upon changes in the real wage and employment experience of the son relative to that of the father.

While accepting Easterlin's basic hypotheses Oppenheimer[9] has rejected his suggested measure

[8] 'Relative Economic Status and the American Fertility Swing', *op. cit.* in footnote 1, p. 181.

[9] Valerie K. Oppenheimer, 'The Eastern Hypothesis: Another Aspect of the Echo to Consider', *Population and Development Review*, 2 (September/December 1976), pp. 433–458.

of the desired standard of living on two counts. First, for any given level of family income, the actual standard of living in adolescence will depend upon the number of siblings in his parents' household, especially the number in the expensive adolescent years. Thus, relative economic status will depend on the size of the family in which a young adult grew up, as well as on his father's income. Secondly, and most important, if the mother is in gainful employment, her contribution to family income will also contribute to her family's actual standard of living and to the desired standard of living of her children. These two influences on the children's desired standard of living are in fact likely to be linked; families with a number of adolescent children may face a financial squeeze (if the husband's earnings are not high enough) because of the high costs of adolescent children, and a practical response would be for the mother to enter the labour force. This phenomenon has been labelled the 'second life-cycle squeeze' by Oppenheimer,[10] and it is clear that a succession of large birth cohorts is likely to intensify the second life-cycle squeeze some 12 to 17 years later. In sum, Oppenheimer correctly argues that a better measure of relative economic status in Easterlin's basic hypotheses would be the ratio of the earnings of young men to total family income (including the mother's contribution) of their parents during their adolescence, net of 'essential child-care expenditures'.[11]

The major alternative explanation of fertility movements comes from the model of household decision-making which forms the core of the 'new home economics', especially the model by Willis[12] and its application by Butz and Ward.[13] We describe this model in greater detail below; briefly it considers that fertility responses to changes in men's earnings and women's wages are different in families with and without employed wives, and that the probability that a married woman of childbearing age will enter paid employment is a function both of her husband's earnings and of her own earning capacity. In particular, the opportunity costs of time of children rise for working wives as women's wages increase, and, for any given level of men's incomes, this induces them to have fewer children and space them more closely. The opportunity cost of time of women not in employment is unaffected by increases in women's wages, but their expanding earning capacities induce a greater proportion of women of childbearing age to enter employment, thus increasing the proportion of families whose fertility is negatively affected by rising women's wages and enhancing their negative impact upon aggregate fertility. Potential earnings of women have increased as the educational level of women has risen, in addition to the general rise in real wages. Also changes in the industrial and occupational structure have combined with slow growth in the male labour force, the wage differential between men and women and changes in social attitudes toward women working outside the home, to expand employment opportunities for women, especially for part-time employment.[14] Given the social division of labour within the family, this increase in the

[10] Valerie K. Oppenheimer, 'The Life Cycle Squeeze: The Interaction of Men's Occupational and Family Life Cycles', *Demography*, 11 (May 1974), pp. 227-245.

[11] 'Essential' expenditures are computed using the U.S. Bureau of Labor Statistics equivalence scales in Valerie K. Oppenheimer, 'The Life Cycle Squeeze . . . ', *loc. cit.* in footnote 10, and estimates of the 'cost of children' developed by Epenshade are used in Valerie K. Oppenheimer, 'The Easterlin Hypothesis: Another Aspect of the Echo to Consider,' *loc. cit.* in footnote 9.

[12] R. J. Willis, 'A New Approach to the Economic Theory of Fertility', *Journal of Political Economy*, 81 (March/April 1973), pp. S14-S69.

[13] *op. cit.* in footnote 5.

[14] 83 per cent of the increase in the number of economically active married women during 1961-71 in Great Britain consisted of part-time workers, and at least three-quarters of the increase in the number of part-time women workers is attributable to an increased use of part-timers within industries; the expansion of the service industries is responsible for the remainder of the increase. The shifting industrial structure was responsible for a substantial part of the general increase in the employment of women between 1961 and 1971, but the substitution of women for men and changes in the occupational mix within industries made up a large part of the increase in female employment in the clerical occupations. See J. F. Ermisch, *The Interaction of Economic Change, Demographic Change and Labour Force Participation in the Evolution of the British Labour Force*, Centre for Studies in Social Policy Working Paper (London, 1978). 157

earning capacities of women has had an increasingly greater negative impact upon aggregate fertility.[15]

Another potentially important influence on fertility is the cost of contraception, including subjective (psychic) costs, search costs and monetary outlays. Effective, easy-to-use contraceptive methods became more easily available throughout the post-war period and knowledge of these methods became more widespread. The introduction of new methods such as the IUD and the pill reduced the psychic costs of contraception thus encouraging its use; diffusion of knowledge reduced search costs, as did the spread of family planning clinics and other sources of contraceptives; and their distribution through family planning clinics and the National Health Service reduced money cost. But while the cost of contraception has fallen throughout the post-war period the timing of the fall has varied between 1955 and 1975.

The only other influence on fertility movements which is considered below is the cost and availability of housing. This selection of factors other than relative economic status which influence fertility is not exhaustive, but I believe it to include the *major* competing influences.

Later in the paper those factors which can readily be measured over time are brought together in a multivariate statistical model, but since Easterlin implies an especially large and stable elasticity of fertility with respect to relative economic status, and regards it as the dominant influence on fertility movements, we shall initially test this hypothesis.

In Easterlin and Condran's brief analysis of the recent fertility swing in England and Wales[16] 'relative cohort size', defined as the ratio of men aged 35-64 to those aged 15-34, is used as a proxy for relative economic status. Relative cohort size is considered an important determinant of the relative affluence of young persons since it is plausible that a substantial increase in the size of a cohort in the labour force relative to earlier cohorts, will affect its competitive position in the labour market adversely. As a result, the income of members of the cohort will be relatively lower at each age than that of members of earlier cohorts, most significantly, their fathers'. The opposite is, of course, true for relatively smaller cohorts. Because different age groups within the labour force are not perfect substitutes for one another large (small) cohorts may have lower (higher) earnings than normal. Cohort size may also affect the whole process of human capital formation (including parental time inputs and formal education) and thereby eventually affect income.[17] Aggregate level studies of U.S. experience since the Second World War support this suggestion of a negative influence of cohort size on relative earnings.[18] Since the size of the cohort of young men is included in the denominator of the measure, this suggests that fertility and relative cohort size should be directly related. Figure 1 shows a fairly high positive correlation between relative cohort size and total fertility rate, although the decline in the former leads that of the latter by three years. Such a lag in response is not implausible, hence this evidence by itself

[15] Some may argue that the true opportunity cost of a child is not related to the value of the wife's time, but rather the cost of child care: this could be zero if there is an elderly relative about, or free day nurseries. This would be acceptable if people view such child care as a good substitute for mother's care, but I would contend that most people view it as an inferior substitute. Even if such child care were viewed as a good substitute, a higher wage for a working mother will, through a 'pure income effect,' raise the desired level of child-care quality and therefore produce a demand for better quality and more costly child-care services (e.g. a nanny is substituted for group care), and consequently the opportunity cost of children increases with women's wages among working mothers. Furthermore, the cost of child-care services outside the home (relative to other goods) will rise with real wages since it is a labour-intensive activity characterized by slow labour productivity growth.

[16] *op. cit.* in footnote 7.

[17] See, for example, S. P. Dresch, 'Demography, Technology and Higher Education: Toward a Formal Model of Educational Adaptation', *Journal of Political Economy*, 83 (June 1975), pp. 535-570, and R. B. Freeman, 'The Effect of the Increased Relative Supply of College Graduates on Skill Differences and Employment Opportunities', Harvard Institute of Economic Research Discussion Paper 485 (1976) for an analysis of the effect of cohort size on the returns from college education in the U.S. and the feedback effect on enrolment.

[18] R. D. Lee, 'Fertility, Age Structure and Income in the United States, 1947 to 1974', Paper presented at the Third World Congress of the Econometric Society (Toronto, 1975), and H. H. Winsborough, 'Age, Period, Cohort and Education Effects on Earnings by Race', in K. Land and S. Spilerman (eds), *Social Indicator Models* (New York: Russell Sage Foundation, 1975).

158

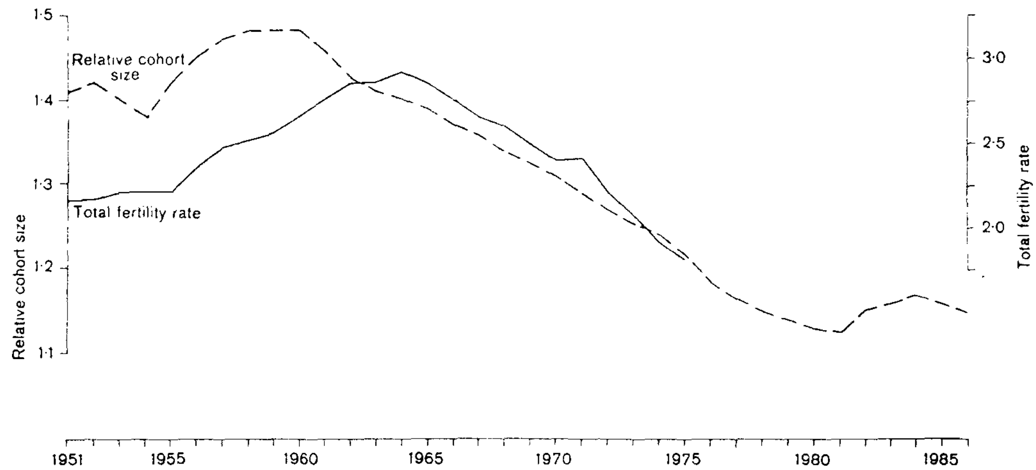


Figure 1. Relative cohort size and fertility (England and Wales). Source: S. E. Overton, *Changes in Fertility in England and Wales*; Centre for Studies in Social Policy Working Paper (London, 1977).

would provide considerable support for the Easterlin hypothesis. Nonetheless, in the context of the evidence presented below it is somewhat less convincing, and Easterlin and Condran admit that the co-variation of fertility and age structure may be purely coincidental.[19]

The strength of the relation between fertility and relative cohort size depends upon the stability and size of two parameters: one relating cohort size to relative income and the other relating relative income to fertility; to the best of my knowledge no direct evidence exists in Great Britain regarding the former parameter, and differences in the institutional structure of the labour market between the U.S. and Britain make it inadvisable to apply U.S. figures to Britain. Furthermore, since the effect of cohort size on the incomes of its members also depends on the level of aggregate demand this parameter may well be unstable. As regards the second parameter, we now consider some direct evidence relating to its size and stability.

Since we initially test Easterlin's hypothesis rather than the variant suggested by Oppenheimer, we shall consider measures of relative economic status similar to those used by Easterlin.[20] The first of these is a 'relative wage'. Since hours worked by the husband and the extent of the wife's gainful employment are both a function of the husband's real wage the latter is a better measure of the resources influencing family decisions than either actual family income or husband's earnings, both of which are endogenous variables resulting from household choice.[21] We therefore use the time series of 'expected men's real wages'. This series is obtained by deflating the money wage by the retail price index and adjusting it by multiplying it by the probability of employment (equal to the complement of the male unemployment rate).[22] The time series of 'expected male real wages' (W_t) is then used to construct a series of the ratio of 'son's wages' to a hypothetical 'father's wage'. When discussing the fertility of married women aged 20-24 (who produce about one-third of all births) their husbands tend to be between 22 and 26 years old and will, therefore have been working for about seven years; more than 60 per cent are in a manual occupation and they will have left the parental home for about five years.[23] The 'son's wage' in year t is, therefore,

[19] *op. cit.* in footnote 7, p. 148.

[20] Relative Economic Status and the American Fertility Swing,' *op. cit.* in footnote 1.

[21] Furthermore, husband's earnings would also include overtime, and overtime earnings are more of a transitory phenomenon reflecting fluctuations in the economy. Fertility decisions are more likely to be based upon 'full' or permanent income and therefore would be little influenced by any changes in husband's earnings which are transitory. Utilization of the husband's real wage as the income variable would implicitly assume a standard workweek.

[22] See Appendix for sources.

[23] See *Population Trends* for the average marriage ages of men and women and the 1971 *Census Economic Activity Tables*, Vol. IV, Table 29 for the social class distribution by age.

159

Table 1. *Measures of relative economic status**

	Wachter definition		Easterlin definition
Year	Relative wage before tax†	Relative wage after direct tax	'Son's' real wage relative to 'father's'
1955			1.07
1956	1.13		1.08
1957	1.12		1.09
1958	1.12		1.11
1959	1.13		1.11
1960	1.16		1.12
1961	1.17		1.13
1962	1.16		1.14
1963	1.15		1.15
1964	1.17		1.16
1965	1.18		1.17
1966	1.18		1.19
1967	1.16		1.20
1968	1.15		1.19
1969	1.13		1.19
1970	1.17		1.19
1971	1.20	1.14	1.19
1972	1.22	1.22	1.20
1973	1.24	1.23	1.20
1974		1.21	1.21
1975			1.23

*See text for definitions and Appendix for sources.
†Three-year-centred moving average

defined as the average of W_t for the seven previous years, and the 'father's wage' is defined as the average of W_t for the ten years preceding t—5. The ratio is the 'relative wage' and is shown in Table 1 for the period 1955-75. As is shown in Figure 2, this relative wage did not fall at any time between 1955 and 1975 whereas the fertility rate of married women aged 20-24 rose until 1964 and then declined, contrary to what would be expected from the Easterlin hypothesis. We have not been able to find age-related wage data as Easterlin did for the whole period 1955 to 1975.[24] Nevertheless, after 1970 relative wages moved in the opposite direction from fertility, and during this period median hourly earnings of young men relative to older men were improving for manual and non-manual workers alike while there was a shift towards non-manual occupations in which the increase in the relative wages of young men was greatest (see Table 2). Although average *weekly* earnings of young men fell slightly relative to those of older men in the manual occupations there was a large increase in the relative average weekly earnings of young men in the non-manual occupations. This is, of course, not the same as a son/father relative wage, but the figures do not suggest a substantial deterioration of the relative wages of young adults after 1970.

As an alternative we consider a measure of relative economic status employed by Wachter in his time series fertility equation, as well as other research on labour force participation.[25] Wachter notes that the lag structure implied by the Easterlin hypothesis would appear to be both flat and long; he, therefore, assumes a simple ten-year moving average for determining 'desired living standard'. We use the time series of 'expected men's real wages' defined earlier as our measure of family resources for the reasons given, and apply the ten-year moving average to that series to

[24] 'Relative Economic Status and the American Fertility Swing,' *op. cit.* in footnote 1.

[25] M. L. Wachter, *loc. cit.* in footnote 2; M. L. Wachter, 'A Labor Supply Model for Secondary Workers', *Review of Economics and Statistics*, 84 (May 1972), pp. 141-151, and 'A New Approach to the Equilibrium Labor Force', *Economica*, 41 (February 1974), pp. 35-51.

160

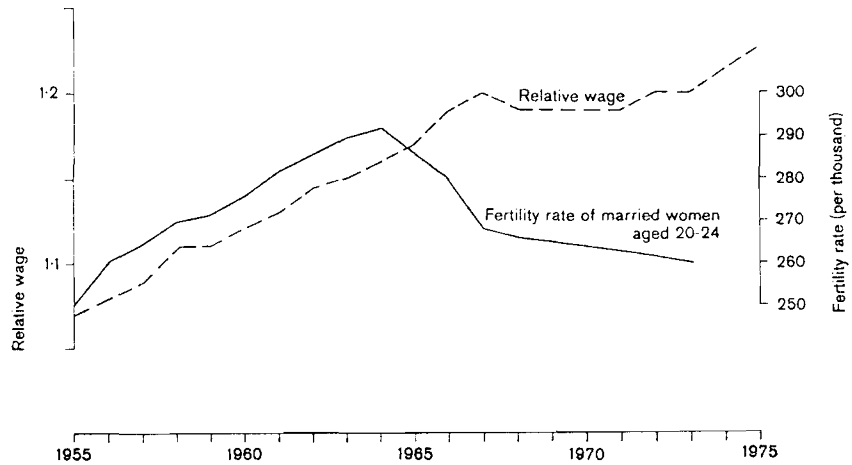


Figure 2. Relative expected wage (Easterlin variant) and fertility. Source: Table 1 and S. E. Overton, *op. cit.* in Figure 1.

Table 2. *Relative earnings by age group: relative to males 40–49**

Age group	Median hourly earnings 1970	1974	1976	Average weekly earnings 1970	1974	1976
Manual						
21–24	0.895	0.917	0.912	0.886	0.900	0.876
25–29	0.965	0.989	0.978	0.971	0.989	0.968
30–39	1.009	1.014	1.013	1.021	1.027	1.016
Non-manual						
21–24	0.624	0.634	0.641	0.565	0.578	0.605
25–29	0.812	0.830	0.827	0.738	0.773	0.776
30–39	0.978	0.982	0.974	0.928	0.937	0.942

Sources: 1970 New Earnings Survey, Tables 81 and 83; 1974 New Earnings Survey, Tables 124 and 126; 1976 New Earnings Survey, Tables 124 and 126.
*Full-time males, paid for a full week

generate a series for 'desired standard of living'. The ratio of expected men's real wages in year *t* to the computed desired standard of living in that year is defined as the relative wage. The movement of this index is shown in Table 1 and Figure 3. As Figure 3 indicates it rises with the total fertility rate, although there are substantial fluctuations between 1961 and 1964 not matched by fluctuations in fertility, and relative wages continued to increase up to 1965–66, when the peak in fertility was already over. There is, therefore, some general agreement with the Easterlin hypothesis up to 1969, but from 1970 onwards the relative wage shoots up while total fertility rate falls more rapidly than before, thus contradicting the hypothesis.

Are there any features of our statistics which can explain this conflict with the Easterlin hypothesis? Throughout, we have used wages before tax, but even if we had used the series for income net of direct tax derived from the direct tax rate for a two-adult, non-pensioner household in the Family Expenditure Survey, we continue to find a substantial increase in the relative wage after 1971 (see Table 1 and Figure 3).[26] We have already seen (Table 2) that the available evidence

161

[26] Since the Family Expenditure Survey figures only take us back to 1961 we can only define the relative, after-direct-tax wage from 1971 onwards. Implicit direct tax rates for other relevant types of households imply similar movements in the relative, after-direct-tax wage.

162

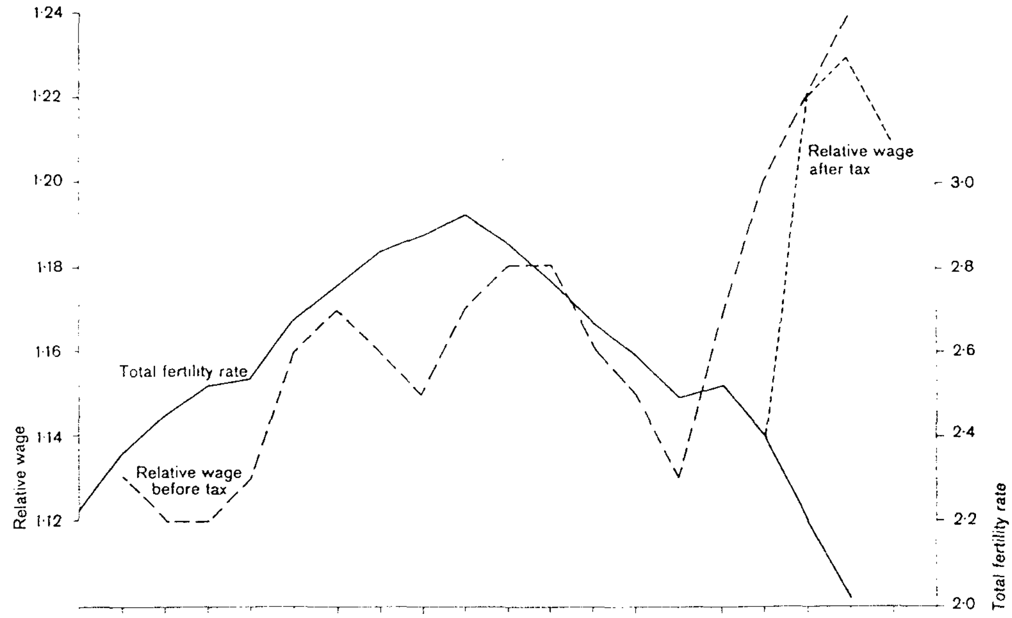


Figure 3. Relative wage (Wachter variant) and fertility. Source: Table 1 and S. E. Overton, *op. cit.* in Figure 1.

does not suggest a deterioration in wages for young men relative to older men after 1970. While wages of manual workers, i.e. those employed in our analysis, have increased relatively to those of non-manual workers since 1971, the majority of young men are in manual occupations.[27] It appears, therefore, that there was a considerable increase in the relative wages of young men in Great Britain after 1969, which negates the Easterlin hypothesis. This result is consistent with the support for the Easterlin hypothesis provided by the U.S. experience since 1970[28] in that real hourly earnings in manufacturing rose by almost 19 per cent between 1970 and 1975 in the U.K. and by only 3 per cent in the U.S. during this same period.[29]

Easterlin also measured relative economic status in terms of the 'son's' unemployment experience relative to that of his 'father'.[30] When we construct measures of 'relative unemployment experience' similar to Easterlin's we find a steady deterioration in the 'son's' unemployment situation relative to 'father's' over the period 1955-75 (see Appendix Figure I), but these measures seem inappropriate in the British context, especially for the period 1955-65. After the Second World War, there was a strong government commitment to full employment. In consequence, unemployment rates stood at about one-fifth the level of the inter-war years, and unemployment rates of young men during 1955-65 were substantially lower than those their fathers had experienced. This strongly suggests an improvement in perceived relative economic status, rather than a deterioration. The movement over time of another measure of relative unemployment experience analogous to the Wachter measure of the relative wage used above also does not tally with the Easterlin hypothesis (see Appendix Figure II). Nonetheless, not much confidence should be

[27] The New Earnings Survey indicates that the ratio of the mean gross hourly earnings of men aged 21 and over (working full time) in manual occupations to the mean gross hourly earnings of men aged 21 and over in non-manual occupations (working full time) increased from 0.63 in 1970 to 0.7 in 1975 (and fell back to 0.68 in 1976); *New Earnings Survey 1976* Part A, Table 16.

[28] R. A. Easterlin, 'The Conflict between Aspirations and Resources', *Population and Development Review*, 2 (September/December 1976), pp. 417-426.

[29] *National Institute Economic Review* (November 1972 and November 1976). Similar comparative results hold after adjusting wages for changes in unemployment rates to obtain' expected real hourly earnings in manufacturing'.

[30] 'Relative Economic Status and the American Fertility Swing,' *op. cit.* in footnote 1.

162

attached to these results, because all these relative unemployment experience measures 'make big things of small differences' in unemployment rates from year to year. Until very recently post-war unemployment experience does not appear to have varied enough to affect longer-term expectations about employment and relative economic status.

The evidence summarized in Figures 2 and 3 leads us to reject the Easterlin hypothesis as formulated and applied by him. We now consider the evidence relating to Oppenheimer's variant discussed earlier. Considering first the effect of the number of siblings, especially those close in age, upon the desired standard of living of young men we find that the adjustment to measures of relative economic status to allow for the impact of fertility echo effects on the standard of living of households in the 'second life cycle squeeze' suggested by Oppenheimer[31] does not remove the contradiction between the evidence quoted earlier and the Easterlin hypothesis. The adjustment for the negative effect on the desired standard of living of the presence of a relatively larger number of adolescents in the 'father's household' following the post-war baby boom at the time when the young adults making fertility decisions between 1970 and 1975 were adolescent enhances the increase in relative economic status observed in Great Britain after 1970. We must, however, also consider the mother's response to the 'second life-cycle squeeze' and the effect of her gainful employment on her family's standard of living and the desired standard of living of her children.

A substantial number of married women aged 35-54 entered the labour force during the early 1960s,[32] possibly in response to a financial squeeze, as members of the large birth cohorts of 1946-49 reached adolescence. The earnings of these working mothers helped to set the desired standard of consumption for their then adolescent children, and the contribution of the mother to family income may have been large enough to offset the effect of the presence of the relatively large number of adolescents in the household. Therefore, since 1970 the increasing contribution of mothers to family incomes during the present generation's adolescence tended to increase the standard of living desired by young couples. The wages of young men may, however, still have risen fast enough between 1970 and 1975 to have led to a rise in relative economic status as defined by Oppenheimer. These conflicting influences on relative economic status make it impossible to be certain about the direction of movement of relative economic status (as measured by the Oppenheimer variant) between 1970 and 1975, but it is clear that her variant does offer a chance of rescue for Easterlin's basic hypotheses during the period when the evidence presented earlier suggested the discrepancy was strongest.

Consider Oppenheimer's variant over the entire period 1955-75. The young adults primarily responsible for the fertility upswing (1955-64) had few siblings for they were survivors of the small birth cohorts of the 1930s, but their mothers were also less likely to be in the labour force than those of members of later cohorts. The net impact of these two factors upon the movement in the desired standard of living is uncertain, but it was probably small relative to that of father's earnings. Movements in the 'relative wage' measures of Figures 2 and 3 are therefore likely to be a good index of movements in relative economic status during 1955-64. In addition to the increase in the 'relative wage' between 1955 and 1964 indicated in Figures 2 and 3 other evidence also suggests that relative economic status was increasing at that time. There were large increases in the proportions of young adults possessing certain qualifications between 1955 and 1964 and as there was an excess demand for qualified persons, their incomes increased.[33] As shown in Table 3, the proportion of qualified persons reaching age 25 in 1952-56 was 25 per cent higher than among those reaching the same age in 1947-51 and the proportion in the subsequent cohort rose by a

[31] 'The Easterlin Hypothesis: Another Aspect of the Echo to Consider', *loc. cit.* in footnote 9.

[32] Between 1961 and 1966 over one million married women entered the labour force of England and Wales as a result of changes in labour force participation rates; two-thirds of these were between the ages of 35 and 54. Cf. J. F. Ermisch, *op. cit.* in footnote 14.

[33] The Advisory Council on Scientific Policy reported in 1961 that excess demand for qualified manpower was likely to exist until 1965, and the evidence seems to bear this out. Cf. A. J. Buxton, 'Some Evidence on the Productivity of Qualified Manpower in Britain', *Bulletin of Economic Research*, 29 (1977), pp. 61-68.

163

Table 3. *Proportion qualified by cohort**

(1)	(2) Per cent qualified of those who attained the age of 25 during the period		(3) Increase over preceding cohort	
Period	Males and females	Males only	Males and females	Males only
1967–71	12.1	13.8	1.0	1.1
1962–66	11.1	12.7	1.1	1.2
1957–61	10.0	11.5	1.2	1.1
1952–56	8.8	10.4	1.8	1.9
1947–51	7.0	8.5	1.1	1.3
1942–47	6.1	7.2	0.2	0.3
1937–41	5.9	6.9	0.7	0.9
1932–36	5.2	6.0	0.8	1.1
1927–31	4.4	4.9		

Source: CSO, *Qualified Manpower in Great Britain,* Studies in Official Statistics No. 29, (HMSO, 1976), Table 4.
* Calculated from the age distribution of the 1971 stock of qualified manpower on the assumption that qualifications are attained by the age of 25 and no difference in mortality between qualified and non-qualified.

further 1.2 percentage points, or 13 per cent. The increase in the labour force due to participation rate changes, and the relative shortage of male workers aged 20–44 also suggest a tight labour market and good employment opportunities for young men.[34] The increase in the marriage rates of women aged 16–24 also provides indirect evidence of the relative affluence of young adults, and this increase contributed 18 per cent of the increase in legitimate births between 1955 and 1964.[35] The average hourly earnings of men under 21 years of age in manual occupations, relative to those aged 21 and over gradually rose by nine per cent between 1955 and 1964.[36] Unemployment among young men in these years was also much lower than that experienced by their fathers, and war-time rationing and shortages probably reduced their material aspirations. All of this strongly suggests a favourable labour market situation and increasing relative economic status for young adults between 1955 and 1964.

It appears that relative wages stagnated (Figure 2) or declined (Figure 3) between 1966 and 1969. Moreover, the labour force participation rate of married women aged 35–54, and thus the average contribution of the mother to the family income during the children's adolescence, increased throughout the 1950s, the period of adolescence of young adults who made fertility decisions during the second half of the 1960s. In consequence, the desired standard of living tended to increase independently of the 'father's' real wage, and relative economic status appears to have fallen between 1966 and 1969. Whether it continued to decline after 1969 depends, as we noted earlier, upon whether the size of the increase of the mother's contribution to family income during the period of adolescence of those young adults who made fertility decisions between 1970 and 1975 was large enough to offset the increase in the relative wage since 1970 (Figures 2 and 3) as well as the negative impact upon the desired standard of living (positive impact on relative economic status) of the relatively large number of siblings of these young adults. There was, in fact, a considerable acceleration in the growth of labour force participation rates of married women aged 35–54 between 1961 and 1966[37] which suggests an accelerated increase in the mother's

[34] The number of working males aged 20–44 in England and Wales decreased by 175,000 due to demographic factors between 1951 and 1956; cf. J. F. Ermisch, *op. cit.* in footnote 14.

[35] S. Elizabeth Overton, *Changes in Fertility in England and Wales*, Centre for Studies in Social Policy Working Paper (London 1977).

[36] *British Labour Statistics: Historical Abstract*, Tables 47–49.

[37] More married women aged 35–54 entered the labour force of England and Wales due to labour force participation rate changes between 1961 and 1966 (670,000) than during the entire decade 1951–61 (600,000). In 1951 the labour force participation rate of women aged 35–54 was 25.7 per cent; by 1961 it rose to 36.7 per cent; and by 1966 it was 49.5 per cent. See J. F. Ermisch, *op. cit.* in footnote 14.

164

average contribution to family income, but we cannot quantify the impact of this on the desired standard of living and the relative economic status of her children. It should also be noted that while real wages of young adults increased between 1970 and 1975 unemployment among them was worsening relative to older workers (see Table 4).

There was also a major acceleration in labour force activity rates of married women aged 35-54 during the first half of the 1950s,[38] and the effect on the mother's average contribution to family income during the adolescence of young adults making fertility decisions in the mid-60s may have been sufficient to offset the increase in relative wages between 1963 and 1966. Again, we are unfortunately unable to quantify these opposing influences on relative economic status, and thus cannot test how well the theory predicts the fertility turning point. We are, therefore, unable to refute or confirm Oppenheimer's variant of the Easterlin hypothesis; while it does receive some support from the evidence of changes in the labour force participation rates of older married women, it is still open to question whether relative economic status, as defined by Oppenheimer, did turn down after 1964 and decline after 1969 as Easterlin's basic hypotheses would require. These results suggest that though relative economic status may not be the dominant influence upon fertility suggested by Easterlin, it may yet be important among a number of factors influencing fertility.

It is clear from the formulation of the 'new home economics' household decision-making model by Willis[39] that the opportunity cost of time for a married woman not in gainful employment is independent of the market rate for female labour and is a positive function of her husband's income. For married women in employment, however, the opportunity cost of the wife's time is equal to the wage she earns, and the probability that a married woman of childbearing

Table 4. *Relative unemployment experience of young men aged 20-29*

	Unemployed aged 20-29		
July	Number (000)	Rate*	Relative to U.K. Rate**
1964	37	0.011	0.70
1965	33	0.010	0.73
1966	34	0.010	0.73
1967	83	0.022	0.85
1968	90	0.025	0.85
1969	89	0.024	0.82
1970	102	0.027	0.82
1971	156	0.041	0.91
1972	168	0.043	0.90
1973	106	0.027	0.82
1974	121	0.031	0.91

Sources: *British Labour Statistics: Historical Abstract*, Table 166; *Yearbook 1970*, Tables 147 and 128; *Yearbook 1974*, Tables 132 and 114. Mid-year population estimates, OPCS.

* Number of unemployed divided by population; this estimate in increasingly biased downwards over time because of the decline in labour force activity rates, especially among men aged 20-24

** Wholly unemployed rate as computed by the Department of Employment

[38] The young adults making fertility decisions between 1963 and 1966 tended to have been born between 1937 and 1943 and to have been adolescents in the 1950s. More than three quarters of the 1951-61 increase in labour force participation rates among married women aged 35-54 appears to have occurred between 1951 and 1956. Cf. J. F. Ermisch, *op. cit.* in footnote 14.

[39] R. F. Willis, *loc. cit.* in footnote 12.

165

age will go out to work is primarily a positive function of the wage she could earn in the market and a negative function of her husband's income.[40] Holding the effects of the husband's income on a non-working wife's opportunity cost of time and the probability of the wife's labour force participation constant, it is suggested that an increase in the husband's income will increase fertility while an increase in the wife's opportunity cost of time increases the cost of children and decreases fertility, since the care of young children is relatively time-intensive. It follows that, other things being equal, the positive effect of the husband's earnings upon fertility of non-working wives will be smaller than that of working wives (because for the former the 'pure income effect' is partially offset by the positive effect of husband's income upon the wife's opportunity cost of time), and that only among working wives is the wife's opportunity cost of time (and children) affected by changes in women's wages. The aggregate fertility response to changes in both men's and (especially) women's wages will, therefore, depend upon the proportions of married women of childbearing age in the working and non-working groups, and the former is itself a function of men's and women's wage rates and employment opportunities.

From Figure 4 we see that women's real wages have risen throughout the post-war period, and men's wages have moved roughly in step with them until after 1969 when the ratio of women's to men's wage rates rose steeply. This rise in women's wages increased the opportunity cost of time and children, and thus reduced fertility among married women in employment, and, in addition, tended to increase the proportion of married women of childbearing age in the labour force since the available evidence suggests that the elasticity of labour force participation by married women of childbearing age with respect to women's wages exceeds in absolute value the elasticity with respect to men's earnings.[41] In consequence, not only did this rise in women's wages have a negative impact upon aggregate fertility, but the size of this impact was increasing because of the growing proportion of married women of childbearing age whose opportunity cost of time was

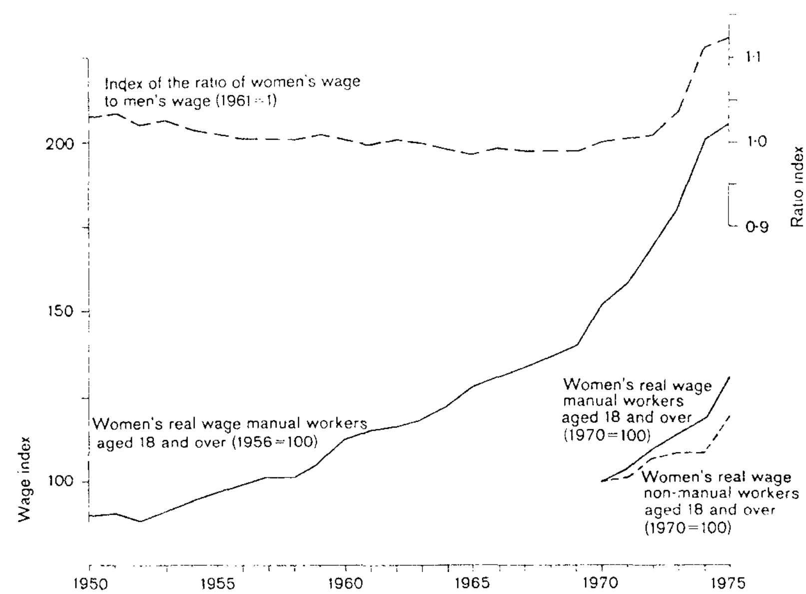


Figure 4. Women's real wages. Source: see Appendix.

[40] Cross-section elasticity estimates for Britain in 1971 indicate a labour supply elasticity for young married women with respect to women's wages larger in absolute value than the elasticity with respect to husband's income, suggesting that an increase in labour force participation by young married women would be observed as real wages increase even if the ratio of the men's to the women's wage remained constant; cf. C. Greenhalgh, 'A Labour Supply Function For Married Women in Great Britain,' *Economica*, 44 (August 1977), pp. 249–266.

[41] *ibid.*

raised by an increase in the wage for women. This increase in the size of the negative effect upon aggregate fertility of increases in women's real wages may explain the fertility turning point.

In an empirical application of the above model to U.S. fertility Butz and Ward[42] consider the response of aggregate fertility to changes in men's earnings and women's wages as a weighted average of the response of working and non-working married women of childbearing age, using the proportion working and not working as weights. They treat the proportion of married women of childbearing age who work outside the home as an endogenous variable. This model follows the U.S. fertility series quite closely during the post-war period, and in particular shows that the fertility downturn was caused by rising women's wages.

We are unable to duplicate the test of this model exactly because we lack information on earnings and wages by age, and, more importantly on women's employment by age and marital status. Nevertheless, we did apply this model to British post-war fertility movements, by using a time series of the ratio of female employees to the female population aged 15-59 (defined as the 'aggregate female employment ratio') as a proxy for the proportion of married women of child-bearing age in paid employment. We also used 'expected men's real wages' (defined above) as a measure of men's earnings and women's real wages as defined in Figure 4. Information from the Censuses of Population and the EEC Labour Force Survey shows that the proportion of married women of childbearing age in gainful employment actually increased more than our 'aggregate female employment ratio'. Our attempts to adjust the latter variable by assuming a linear trend in the proportion of the female labour force who are married and aged 20–44 between each pair of census dates was not successful, because of the collinearity with the other variables. We were, therefore, forced to retain our 'aggregate female employment ratio' as an approximation.

In both the linear and log-linear specifications, the parameter estimates are not in conflict with any of the basic hypotheses of the model (see Appendix), and the elasticity of the total fertility rate with respect to women's real wages rises in absolute value over the post-war period, especially after 1970. The estimates are, however, not statistically precise because of the correlation among the residuals in both specifications,[43] and to some extent because of multicollinearity among the independent variables (correlation coefficients in excess of 0.997) in the log-linear specification. The estimates of the elasticity of the total fertility rate with respect to women's real wages and with respect to 'expected men's real wages' sum to a positive number, so that an equal percentage increase in men's and women's real wages tends to increase fertility after 1975, but in the log-linear specification the effect of such an increase in real wages on fertility is not statistically significantly different from zero (see Appendix). Using the first differences of the variables in the linear specification yields parameter estimates with a negative sum of the elasticities, but this sum does not differ significantly from zero. Between October 1975 and October 1976 expected men's real wages fell by 6.3 per cent; women's real wages rose by one per cent and the total fertility rate fell only slightly (by about 3 per cent). The parameters estimated from the linear specification would predict a fall in the total fertility rate of 38 per cent while use of the parameters estimated from the first differences of the variables in the linear specification would predict a fall in the total fertility rate of 9 per cent. The serial correlation among the residuals

[42] *op. cit.* in footnote 5.

[43] The value of the Durbin-Watson statistic is sufficiently low to reject the hypothesis of no correlation among the original residuals. Serial correlation among the residuals implies that the estimates of the *standard errors of the parameter estimates* shown in the Appendix table are biased and that the parameter estimates themselves are inefficient. If the serial correlation is of the first order (and positive), then these standard error estimates are biased downwards, so the parameter estimates are less precise than the computed standard errors and *t*-ratios would suggest (cf. A. S. Goldberger, *Econometric Theory* (New York: John Wiley and Sons, 1964), pp. 236–246). The estimated first-order autocorrelation coefficient of the residuals is 0.45 and 0.69 in the linear and log-linear specifications respectively. We did not utilize the two-step, generalized least squares technique for estimating the model's parameters using this coefficient, nor did we employ the iterative Cochrane-Orcutt procedure. The low value of the Durbin-Watson statistic also obtained in the specification using the first differences of the variables (although now in the 'inconclusive range') suggests that this specification did not improve the precision of the parameter estimates.

168

suggests that the standard errors of our parameter estimates are larger than those reported in the Appendix, and this, together with the fertility experience since 1975, is an indication that we should put little faith in the relative values of the point estimates of elasticities of fertility with respect to men's and women's wages obtained from these specifications of the model. The errors in our estimates are likely to be quite large, and the sum of these elasticities may in fact be near to or less than zero. Butz and Ward[44] found that the sum of these elasticities was negative in the U.S., and the conjunction of their parameter estimates with the 1971 British proportion of women of childbearing age in gainful employment also yields a negative sum. We also find that our estimated model follows fertility (see Figure 5) less well than Butz and Ward's estimated model. These differences are likely, however, to be due to the imprecision or bias in our parameter estimates.[45] We conclude that the increase in the size of the aggregate elasticity of fertility with respect to women's real wages, which was itself a consequence of rising women's real wages and employment opportunities, made a substantial contribution to the downturn in fertility after 1964, and, in association with the acceleration in the growth of women's real wages, led to the steep decline in British fertility since 1970. Even with our imprecise parameter estimates, this model of fertility change (summarized in Figure 5) does a better job of 'explaining' British fertility movements than the movement in the Easterlin and Wachter measures of relative economic status (Figures 2 and 3 above).

The analysis by Butz and Ward[46] suggests that the relative economic status variables constructed by Easterlin and Wachter may play a role in explaining fertility movements, but not a dominant one. Their results 'strongly indicate that both current income and women's wages operate independently' in explaining fertility movements. Inclusion of the Wachter relative wage

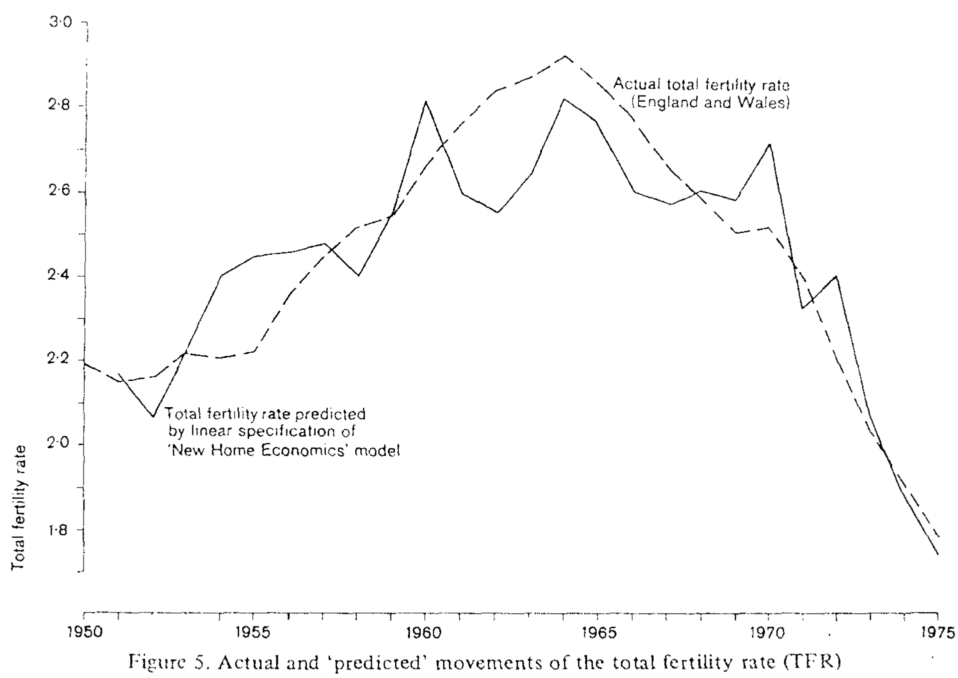


Figure 5. Actual and 'predicted' movements of the total fertility rate (TFR)

[44] *op. cit.* in footnote 5.

[45] A bias in our parameter estimates could arise from bias in our estimate of the proportion of married women of childbearing age in gainful employment. The serial correlation among the residuals might also be taken as an indication that an independent variable moving slowly through time (possibly the 'desired standard of living') has been omitted, thus also causing bias in the parameter estimates. Other specifications of the model are being experimented with.

[46] *op. cit.* in footnote 5.

variable in the statistical analysis of British fertility movements produces either insignificant or perverse results.[47] Nevertheless, in view of the deficiencies of the measures of relative economic status pointed out by Oppenheimer and discussed above, Oppenheimer's variant of Easterlin's basic hypotheses, cannot be completely rejected. As we show below, both Oppenheimer's variant of Easterlin's basic hypotheses and the household decision-making model of the 'new home economics' yield the same forecast for the secular movement of fertility provided we do not take our estimates of the relative magnitudes of the parameters in our specifications of the 'new home economics' model too seriously.

The steady increase in the relative cost of housing throughout 1955–75 was also operating to depress fertility, but since the rate of increase was relatively steady throughout 1972, this cannot explain the turning point in fertility. As housing costs moved more or less in step with real wages it is difficult to separate statistically their effect from that of income. Housing availability may be more important in explaining fertility than housing cost. The war and post-war restrictions on building led to a considerable housing shortage during the early 1950s. Efforts were made to alleviate this shortage during the 1950s and the 1960s,[48] and these measures tend to have a positive influence on fertility which diminishes over time (to the extent that housing shortages restrained family-building). Although there still appears to be a shortage of housing, especially when regional variations are taken into account,[49] this does not appear large enough to be restraining fertility noticeably, so that its alleviation is unlikely to raise fertility.

Finally we briefly consider the influence of the cost of contraception on fertility. There can be little doubt that there has been a movement away from non-appliance methods of birth control toward the pill and various appliance methods. This is a movement toward more effective methods with lower subjective user costs and the switch to these methods was made easier by increases in knowledge about and availability of these methods. Appliance methods were available throughout the period 1955 to 1975, but their costs probably diminished over time. The pill, on the other hand, was introduced during the early 1960s, and as it became increasingly acceptable and available, it further reduced the cost of contraception. Notwithstanding this, it is unlikely that the reduction in the cost of contraception during the 1960s was primarily responsible for the downturn in fertility. The lower cost of fertility control associated with the pill may have facilitated the fertility decline after 1964 but it is unlikely to have been sufficient by itself to produce the downturn in fertility.[50] Nor can changes in the cost of fertility control explain the upswing of fertility between 1955 and 1964.

We have unfortunately been unable to assess the relative contribution of these various influences on British fertility movements. Our analysis, nevertheless, has relatively unambiguous implications for the future movement of fertility.

With birth control costs already near their minimum the influence of changes in such costs on future fertility will be almost nil. We have already noted that the current housing shortage does not appear to be restraining fertility so that its alleviation will not significantly affect future fertility. The impact of future housing costs on fertility will primarily depend upon relative economic status. In fact, future movements in relative economic status, current income and women's real wages are likely to be the dominant influences upon future fertility.

The proportion of couples reaching their mid-20s between 1975 and 1980 whose mothers were in gainful employment during their adolescent years will be greater than among earlier cohorts and they will have had a relatively small number of siblings; those in one-earner households

[47] When entered in addition to the variables in the Appendix Table, the coefficient of the relative wage was insignificantly different from zero; when entered as a substitute for current income it or other variables had coefficients with a sign which significantly contradicted the basic hypotheses of the model.

[48] See Department of the Environment, *Housing Policy, Technical Volume*, Part I (London: HMSO, 1977).

[49] See C. Whitehead, 'Where Have All the Dwellings Gone?', *CES Review*, 1 (1977), pp. 45–53.

[50] The decline in fertility in a number of European countries where the pill is not extensively used supports this.

169

will, therefore, achieve a lower consumption standard than that set in their two-earner, parents' household during adolescence. The young women of this generation will thus be encouraged to take employment and limit the size of their families in order to avoid a decline in relative economic status. Since labour force participation by the wife increases her earning capacity and therefore the opportunity cost of a child[51] a higher proportion of married women employed during the early years of marriage is likely to lead to continued higher participation rates among married women and to lower completed family sizes. In turn, these families will set a standard of consumption for the succeeding generation (reaching their mid-20s at the turn of the century). Unless the earnings of young men in the succeeding generation increase substantially, their wives, in turn, will be encouraged to work and have small families. In sum, the expanding employment opportunities for women during the 1960s and 70s have made it more likely that a prospective fall in relative economic status will result in an increase in married women's employment and to a lower family size; as a result the compensatory responses to family financial squeezes at both the beginning and middle of the life cycle tend to perpetuate themselves.[52]

Young adults reaching their mid-20s, the age at which fertility is at its peak, between 1977 and 1987 are the survivors of the large birth cohorts of 1955-64. Larger families will have tended to reduce family consumption standards during their adolescence, but increased labour force participation by their mothers may have offset this. The disproportionate rise in the labour activity rates of older married women has in fact continued,[53] so the net effect of the desired standard of living of the generation reaching their mid-20s over the next ten years is ambiguous, but is not likely to decline very much. In view of the bleak employment and earnings prospects for young men over the next ten years, caused partly by the size of the birth cohort, but also by slow world economic growth, it appears unlikely that relative economic status will improve.

Both legislation and general economic growth are likely to lead to an increase in women's real wages and, therefore, the opportunity cost of time and children among working women. We have already discussed the impact of a relatively larger growth of women's wages than men's upon fertility after 1970. In addition, more young married women will be induced to enter employment, so that the effect of rising women's wages upon fertility will increase. Our analysis of the 'new home economics' model and the fertility experience since 1975 (as well as Butz and Ward's analysis) suggest that the sum of the elasticities of fertility with respect to men's and women's real wages is near zero or possibly negative. A further rise in labour force participation among young married women is, therefore, likely to make this sum negative soon if, indeed, it is not so already. Economic growth will, therefore, lead to a gradual fall in fertility, and if women's real wages were to rise more rapidly than men's wages, the decline in fertility will be larger. It cannot, however, continue forever, so that economic growth may entail the asymptotic approach to some minimum level of fertility, with random oscillation around this path. Both the Oppenheimer variant and this model therefore point to a continuation of low fertility in conditions of slow to moderate economic growth. While women's employment opportunities are likely to stagnate or expand only slowly because of the excess supply of persons with low qualification and the glut of young men[54] this is likely to reduce only the *rise* in the proportion of young wives who work and not the proportion itself; since women's employment opportunities and real wages are unlikely to fall continuously. The sum of the fertility elasticities is, thus, not likely to become positive. Furthermore, the prospective decline in relative economic status will tend to keep fertility low; if young women

[51] See J. Mincer and S. Polachek, 'Family Investments in Human Capital: Earnings of Women', *Journal of Political Economy*, 82 (March/April 1974), pp. S76–S110.

[52] Labour force participation rates of married women of all ages have continued to increase substantially through 1975, and each succeeding cohort has had a higher labour force participation rate throughout its lifetime than earlier cohorts; cf. *Department of Employment Gazette* (January 1974), p. 13, and J. F. Ermisch, *op. cit.* in footnote 14.

[53] *ibid.*

[54] *ibid.* and J. F. Ermisch, *Demographic Change and the Prospects for Qualified Labour in Britain*, Centre for Studies in Social Policy Working Paper (London, 1978).

170

feel that they must work in order to achieve the standard of living they aspire to, they certainly will not add to their costs by having children when they are unable to find work. A significant upturn in fertility during the next ten years therefore seems unlikely. Beyond that period much will depend upon the rate of economic growth,[55] but as to-day's two-earner, small families are setting the desired standard of living of the succeeding generation, that generation, too, is very unlikely to increase its fertility significantly. We can, therefore, expect a dampened fluctuation of fertility rates around a low level, probably below replacement in Britain because of the low level prevailing now.

The Oppenheimer variant of Easterlin's basic hypotheses and the 'new home economics' only have conflicting implications under conditions of very rapid economic growth or when there is an absolute decline in real wages. In the former case, relative economic status may improve despite the effect of their mother's employment on the desired standard of living of the young couples, and fertility would tend to increase in consequence. In the 'new home economics' formulation (with a negative sum of fertility elasticities) a period of rapid economic growth would be a very expensive time in which to have children, and thus fertility will fall. The converse, would be true in a situation of economic decline, although a sustained decline in women's real wages and employment opportunities could produce a switch in the sign of the sum of fertility elasticities because a lower proportion of married women of childbearing age would be in paid employment. Nevertheless, in the economic circumstances which are likely during the next ten years both the Oppenheimer variant and the 'new home economics' imply a continuation of low fertility. Moreover, if both models operate in conjunction, their opposite effects on fertility in circumstances of rapid economic growth or economic decline also suggest the maintenance of low fertility levels.

The forecasting models we have discussed can be constructed by regarding 'relative cohort size' as an index of relative economic status. Our analysis leads us to question the use of 'relative cohort size' because it does not allow for the effect of the wife's labour force response on the desired standard of living of the next generation, and because our analysis suggests that in any case relative economic status is not the dominant influence on fertility. It is, nevertheless, apparent from Figure 1 that even if these objections were disregarded, the model would still lead to a forecast of continued low fertility for at least the next ten years. In contrast, the Office of Population, Censuses and Surveys projections show a fertility turning point in 1979 reaching the replacement level by 1987.[56]

The basic hypotheses of Easterlin and the 'new home economics' are useful in explaining past and future influences on British reproductive behaviour. But more is needed for an adequate explanation of fertility movement. In particular, we cannot ignore that the decline in fertility since 1964 has occurred throughout Europe. Perhaps the effect on fertility of the rising earning capacities of women (and their consequent greater labour force participation) throughout the industrialized world, and/or the effect upon relative economic status and fertility of the end of a 'long swing' in world economic growth has become apparent since 1964. Since further insight may be obtained by looking at other countries, a similar analysis of the relevance of the Easterlin hypothesis and the 'new home economics' to fertility movements in the Federal Republic of Germany is in progress. We hope it will increase our understanding of the recent fertility swing and the future path of fertility.

EMPIRICAL APPLICATION OF THE 'NEW HOME ECONOMICS' MODEL

In the linear specification the equation showing the aggregate response of fertility to changes in men's and women's wages takes the form:

$$F = \beta_0 + (\beta_1 - \beta_2).K.W_M + \beta_2.W_M + \beta_3.K.W_F + u$$

[55] In addition to the effect of economic growth on the growth of real wages, economic growth may also spur the development of high quality facilities and arrangements for the care of children.

[56] Cf. *OPCS Monitor, Population Projections: mid-1976 based* (31 May 1977).

171

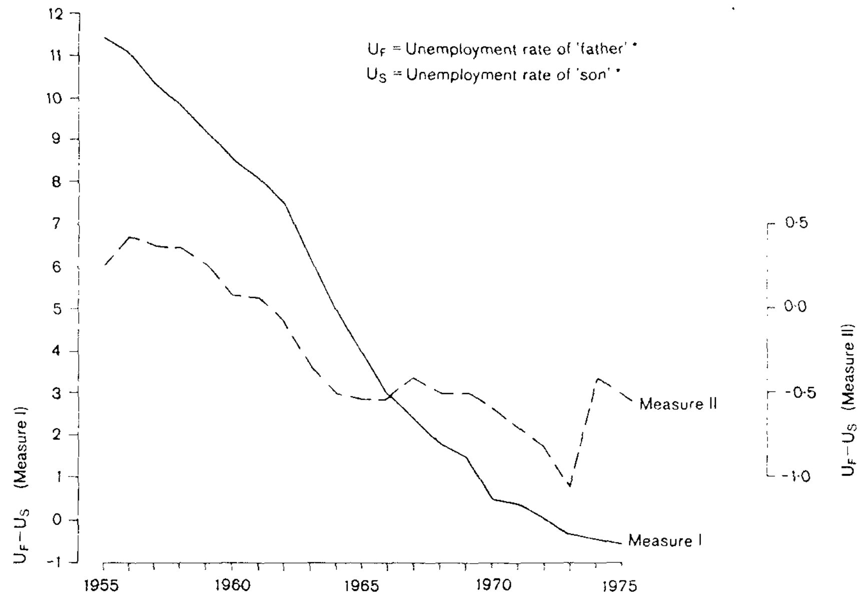


Figure I. Relative unemployment experience (Easterlin variants)

*In Measure I $U_{Ft} = \frac{1}{20} \sum_{t-25}^{t-6} U_i$ (excluding the war years 1939–40 where appropriate), where U_i = men's unemployment rate in year i. In Measure II $U_{Ft} = \frac{1}{10} \sum_{t-15}^{t-6} U_i$. In both measures $U_{St} = \frac{1}{8} \sum_{t-8}^{t-1} U_i$.

Source: see Appendix.

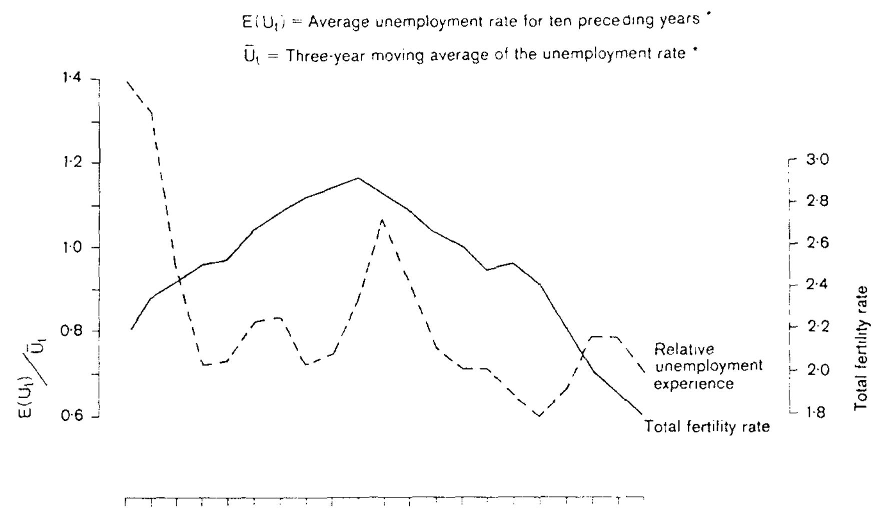


Figure II. Relative unemployment experience (Wachter variant).

*$E(U_t) = \frac{1}{10} \sum_{t-10}^{t-1} U_i$, where U_i = men's unemployment rate in year i.

$\bar{U}_t = \frac{1}{3} \sum_{t-1}^{t+1} U_i$.

Source: see Appendix.

where W_M stands for expected men's real wages, W_F women's real wages and F the total fertility rate while β_1 and β_2 are the fertility responses to changes in expected men's wages in families with and without a working wife respectively. The fertility response to changes in women's wages (which only occurs in families with an employed wife) is given by β_3. The 'aggregate female employment ratio' is denoted as K (although theoretically K should be the proportion of married women of childbearing age in employment), and u is a normally distributed random residual with zero mean. Since K is endogenous we use two-stage least squares to estimate the parameters (i.e. the β_i), using W_M, W_F and W_M and W_F lagged one year as instrumental variables. The model implies that $\beta_1 > 0$, $\beta_2 > 0$, and it has been found that the negative effect on fertility of the higher opportunity cost of time associated with higher women's real wages dominates the income effect, so $\beta_3 < 0$. The log-linear specification merely substitutes the logarithms of W_M and W_F for W_M and W_F in the equation above, and the β_i are then interpreted as elasticities. The first difference specification substitutes the first differences of $K \cdot W_M$, $K \cdot W_F$ and W_M for these variables.

APPENDIX

Men's money wage:	Average hourly earnings of full-time male manual workers aged 21 and over, all industries, U.K: *British Labour Statistics Historical Abstract* (BLSHA), Table 85 and *British Labour Statistics Yearbook 1974* (BLSY74), Table 50.
Retail price index:	BLSHA, Tables 90, 91 and 93; and BLSY74, Table 56 (linked to 1956 base).
Men's unemployment rate:	BLSHA, Tables 160, 161, and 166; BLSY74, Table 114; and *Department of Employment Gazette*, January 1976 (percentage of insured workers only for years before 1943).
Implicit direct tax rate:	From Family Expenditure Survey as reported in R. Harris, 'A Review of the Effects of Taxes and Benefits on Household Incomes 1961–75', *Economic Trends*, January 1977.
Women's money wage:	Average hourly earnings of full-time female manual workers aged 18 and over, all industries, U.K: BLSHA, Table 85 and BLSY74, Table 50.
Women's manual and non-manual money wage:	Mean gross hourly earnings of full-time women workers aged 18 and over, manual and non-manual, *New Earnings Survey 1976*, Part A, Table 16.
Housing Cost Index:	Implicit deflator of housing expenditure from Consumers' Expenditure Tables of *National Income and Expenditure*, various years.
Aggregate female employment ratio:	Ratio of female employees in employment, U.K. (BLSHA, Table 134: BLSY 73, Table 54; BLSY 75, Table 52) to female population aged 15-59, U.K. (*Annual Abstract of Statistics*)

173

APPENDIX TABLE

Two-stage least squares estimates of the parameters of the total fertility rate equation* (1951–75)

Parameter	Linear specification	Log-linear specification	Linear, first difference specification
$\beta_1\ \beta_2$ (coefficient of $K.Wm$ or $K.\ln Wm$)	−0.213*** (0.0588)	4.257*** (2.066)	−0.0503(n.s.) (0.0324)
β_2 (coefficient of Wm or $\ln Wm$)	0.184*** (0.0363)	2.363(n.s.) (2.435)	0.043** (0.0209)
β_3 (coefficient of $K.Wf$ or $K.\ln Wf$)	−0.042*** (0.011)	−5.764*** (0.834)	−0.023** (0.0091)
β_1	−0.029(n.s.) (0.024)	6.623*** (1.023)	−0.0073(n.s.) (0.0155)
R^2	0.814	0.726	0.419
Durbin–Watson statistic	1.102	0.629	1.264
Degrees of freedom	21	21	20

Total fertility rate elasticity estimates

Elasticity of fertility rate with respect to:	Linear specification	Log-linear specification	Linear, first difference specification
Women's real wages (Wf):			
1955–59	−0.86**	−2.81**	−0.474**
1960–64	−0.90**	−2.97**	−0.501**
1965–69	−1.14**	−3.13**	−0.634**
1970–75	−1.99**	−3.29**	−1.106**
1975	−2.90**	−3.44**	−1.608**
Men's real wages (Wm):			
1955–59	3.26**	4.44**	0.754**
1960–64	3.01**	4.56**	0.692**
1965–69	3.35**	4.68**	0.773**
1970–75	4.74**	4.79**	1.084**
1975	5.56**	4.90**	1.268**
Sum of elasticities (1975)	2.66**	1.46(n.s.)	−0.340(n.s.)
Standard error of sum	0.76	(1.394)	(0.566)

* Numbers in parentheses are asymptotic standard errors. The reported R^2 is based upon the use of the values of K calculated from the instruments as regressors. (n.s.) not significantly different from zero at conventional levels of significance.
** Significant at the 0.05 level under $H_0 : \beta = 0$.
*** Significant at the 0.005 level under $H_0 : \beta = 0$.

174

1

Population and Economic Change in Developing countries

Edited by Richard A. Easterlin

The University of Chicago Press, National Bureau of Economic Research, 1980

5 Causes and Consequences of Mortality Declines in Less Developed Countries during the Twentieth Century

Samuel H. Preston

Only a few countries of Africa, Asia, and Latin America can supply suitable data for estimating mortality levels in 1900. Many more can supply such data for 1940 or 1950. Without exception, the estimated levels of mortality prevailing in those years are higher than current levels. For those countries that can provide data at both earlier points, most improvement as indexed by life expectancy at birth has been achieved since 1940. It appears from fragmentary records that life expectancy at birth during 1935–39 was about 30 years in Africa and Asia and 40 years in Latin America. The respective levels in 1965–70 were on the order of 43, 50, and 60 (World Health Organization 1974*b*; United Nations, Population Division 1973).

The magnitude and the demographic character of this improvement have been documented in a number of excellent reviews, and for this reason they need not detain us here (United Nations 1963, 1973, 1974; Stolnitz 1974; Arriaga 1970; World Health Organization 1974*b*). These works suggest that the mortality improvements, when measured by the absolute decline in age-specific death rates, have tended to be largest at ages under 5 (especially infancy) and above 40. The proportionate declines, on the other hand, have been largest in the older childhood ages. Life expectancy gains for females have been larger than those for males. It is likely that gains have been more rapid in urban than in rural areas. In these matters, mortality experience in less developed countries

Samuel H. Preston is associated with the University of Washington.

The author is grateful to Avery Guest, James McCann, Peter Newman, and Masanori Hashimoto for comments and suggestions and to William Grady and Thomas Revis for research assistance. This research was aided by National Institutes of Health Center Grant 1POL HD 09397–01.

175

(LDCs) has roughly recapitulated that in more developed countries (MDCs). Life expectancy differences between MDCs and LDCs have narrowed, although the lagging pace of improvement in Africa has produced greater dispersion within LDCs themselves. The decadal rate of mortality decline in many LDCs surpasses that ever observed in populations of the now-developed world.

This paper has two purposes: to identify the factors responsible for these mortality improvements in LDCs and provide estimates of their relative importance; and to begin tracing the effect of these improvements on demographic and economic processes. Less developed countries are defined regionally to comprise Africa, Latin America, and Asia except Japan. Data on mainland China, North Korea, and what was formerly North Vietnam are not available, and for all practical purposes these countries are also excluded from the set under review. Conclusions reached about the importance of various factors in the mortality decline do not appear to conflict with the impressions of informed observers of these matters in China (Wegman, Lin, and Purcell 1973).

5.1 Causes of Declining Mortality

There is much more consensus on the fact of mortality decline in LDCs than on its causes. Considerable dispute remains about whether the decline has been principally a by-product of social and economic development as reflected in private standards of nutrition, housing, clothing, transportation, water supply, medical care, and so on or whether it was primarily produced by social policy measures with an unprecedented scope or efficacy. A third possibility is that technical changes reduced the relative costs of good health. This possibility is usually subsumed within the social policy position because it is clear to most observers that the major technical changes that have occurred—immunization against a host of infectious diseases, vector eradication, chemotherapy—had to be embodied in social programs in order to affect the mortality of the masses in LDCs. Demographers have almost unanimously favored the social policy–technical change interpretation of mortality decline (Davis 1956; Coale and Hoover 1958; United Nations, Population Division 1974; Stolnitz 1974). As evidence, they have principally cited the unprecedented rate of mortality reduction in many LDCs and certain dramatic examples of obviously effective government intervention, most notably in Sri Lanka and Mauritius. Many specialists in international health (Fredericksen 1961, 1966*a*,*b*; Marshall, Brown, and Goodrich 1971), medical historians focusing on primarily Western populations (McKeown and Record 1962; McKeown 1965), and some economists (Sharpston 1976) have opposed this interpretation, usually claiming that social interventions have been largely ineffective or insufficiently widespread.

Kuznets (1975) and Coale and Hoover (1958) have argued that, in one sense, the distinction between economic development and public health interventions creates a false dichotomy. Development itself strengthens the nation-state, improves communications among nations and hence facilitates the transfer of medical technology, and routinizes scientific advance. While this position is unassailable, it leaves unanswered the question whether the mortality decline was a product of changes in private consumption or of public programs and technical change, regardless of whether the latter were in turn produced by economic development in its broadest sense. Even if public programs and technical changes were merely intervening variables in the relation between mortality and development, the importance of their role remains to be identified.

5.1.1 *Effect of Private Income Levels on Mortality*

That mortality rates are sensitive to private living standards, independent of the national level of economic development, scarcely needs documentation. Studies of mortality differentials among individuals by social or economic class in countries as disparate as India and the United States consistently reveal lower mortality rates among the upper classes (Kitagawa and Hauser 1973; Vaidyanathan 1972). The role of private living standards in creating the pattern of international mortality differentials is more difficult to assess. Richer countries not only have richer people but, in general, have larger and more effective social programs.

Some indication of the importance of private living standards for international mortality differences may be gained by examining the importance of income distribution as a factor in those differences. The international relation between national income per capita and life expectancy is decidedly nonlinear, with life expectancy showing strongly diminishing returns to increases in income (Preston 1975*a*; Vallin 1968). It is reasonable to expect that mortality also responds nonlinearly to individual income levels, in which case the distribution of income within a nation should influence its aggregate level of mortality. In particular, suppose that the relation between individual income and life expectancy is log linear:[1]

$$e^{o}_{0i} = a + b \ln Y_i,$$

where e^{o}_{0i} = life expectancy at birth in income group i

Y_i = level of income received by group i

a, b = constants.

If the national level of life expectancy is simply the aggregate of these individual-level relations, with no contribution from the *national* level of income except insofar as it reflects individual incomes, then the life expectancy for the population, $e^o{}_{0p}$, will be equal to

$$e^o{}_{0p} = f[\sum_i (a + b \ln Y_i)]$$
$$= a + b \ln Y + b \times f \times \sum_i \ln (S_i/f),$$

where Y = mean level of income in the population
S_i = share of total national income earned by group i
f = share of total population represented by group i, assumed to be constant among the groups.

Life expectancy will be a function of mean national income, Y, and of the distribution of income as represented by the term $\sum_i \ln(S_i/f)$. This term, which is related to the entropy measure of income distribution, ranges from 0 if income is perfectly evenly distributed to $(-\infty)$ if one group has no income. Strictly speaking, the weights that permit subgroup life expectancies to aggregate into population life expectancy are provided by births rather than by population size, but the two will be very highly correlated.

To examine the importance of private incomes for national life expectancy, the value of this income distribution measure was computed for fifty-two populations on which income shares were estimated in 5% population segments. The values and sources can be found in appendix table 5.A.2, along with values of $e^o{}_0$ and Y. The importance of private incomes can be inferred from the consistency of coefficients on Y and on the income distribution measure. If national life expectancy is simply a function of private income levels, the coefficients on $\sum_i \ln(S_i/.05)$ should be 0.05 of the coefficient on Y. If national income contributes independently of private income, the ratio should be less than 0.05. The equation as estimated by ordinary least squares on the fifty-two observations is

$$e^o{}_0 = 19.105 + \underset{(.859)}{6.984} \ln \bar{Y} + \underset{(.237)}{.375} \sum_i \ln(S_i/.05)$$

$$R^2 = .651$$
$$\bar{R}^2 = .644.$$

The coefficient of the income distribution term is in fact 0.0536 of that of national income, suggesting that relations between mortality and income at the national level are indeed dominated by relations between mortality and income at the individual level. This result should be treated with great caution because of inaccuracy and incomparability in the measure of income distribution and because the log-linear functional form probably simplifies a more complex relationship. Furthermore, the standard error of the income distributional coefficient is large enough to prevent rejection of the hypothesis that the true coefficient is zero. Nevertheless, one direct implication of the result is that the mortality risks facing a family earning \$10,000 per year or \$100 per year are not strongly influenced by the prevailing level of average income in the nation in which they reside.

The suggestion that private incomes are very influential in determining national levels of life expectancy at a moment in time does not imply, of course, that changes in private incomes have been the dominant factor in mortality changes during this century. The actual changes in income may have been too small, in conjunction with the sensitivity of mortality to income, to account for the observed mortality changes. Before trying to establish the role played by changes in private living standards in LDC mortality declines, it is useful to make an assessment of the causes of death responsible for those declines.

5.1.2 Causes of Death Responsible for Mortality Declines

Interpretation of mortality declines in LDCs would depend on whether the cause of death responsible for the majority of declines were, for example, smallpox, diarrheal disease, or malaria, since it is clear that death rates from these causes are fundamentally responsive to different influences. Unfortunately, the causes of death responsible for mortality change in LDCs have never been documented on a broad scale. A large part of the reason is that most LDCs still cannot supply national-level data on cause patterns, and data for those that can undoubtedly reflect inaccurate diagnoses and incomplete coverage. Problems are magnified when attention is turned to the patterns of earlier years. Nevertheless, it is possible to piece together a picture that provides some useful clues about the order of magnitude of the causes responsible.

First, it is clear that, in high-mortality populations, infectious and parasitic diseases bear almost exclusive responsibility for shortening life below the modern Western standards of 69 years for a male and 75 for a female. Life tables by cause of death have been constructed for 165 populations at varying levels of mortality (Preston, Keyfitz, and Schoen 1972). When the aggregate of infectious and parasitic diseases were hypothetically eliminated from those life tables and life expectancy was recalculated, the common result was to produce a life expectancy between 65 and 70 for males and between 70 and 75 for females, regardless of a population's initial mortality level (Preston, Keyfitz, and Schoen 1973). In 1920, Chilean males would hypothetically have gone from a life expectancy of 28.47 to one of 65.68 and females from 29.85 to 69.76. In Taiwan, males would have enjoyed a life expectancy of

72.27 years instead of 26.68, and females 76.00 instead of 29.18 (Preston, Keyfitz, and Schoen 1972, pp. 150–51, 702–3).[2]

Despite the appeal of life table measures, they are an unnecessarily awkward vehicle for discussing causes of death because the causes are nonadditive in their effect on life table parameters. This problem is averted by the use of age-standardized death rates. Models have been constructed to represent the typical cause-of-death structure for populations at various levels of mortality as indexed by the age-standardized crude death rate from all causes combined (Preston and Nelson 1974). Of the 165 populations supplying data for these models, only 41 were from Africa, Asia, or Latin America, and of these 5 were for Japan and 3 for the Jewish population of Israel. Nevertheless, the results suggest that cause-of-death structures, controlling mortality level, vary less between MDCs and LDCs than they do among regional groups within MDCs. Lower cardiovascular mortality and higher mortality from diarrheal diseases and maternal causes in LDCs represent their only significant divergence from MDC patterns (Preston and Nelson 1974, p. 37). John Gordon, one of the leading epidemiologists whose work focuses on developing countries, states that "infectious disease in the tropics and in some other preindustrial areas is too often viewed elsewhere as a collection of odd processes peculiar to those regions. Such diseases as schistosomiasis, filariasis, paragonimiasis, and all the others do exist. The plight of children, however, [who account for the bulk of annual deaths in LDCs] is the result of the everyday infections of the intestinal and respiratory tracts and with the communicable diseases specific to early life everywhere" (Gordon 1969, p. 218). Even among adults, the exotic tropical diseases are typically much more important sources of morbidity than of mortality.

Thus, there is some justification for allowing LDC cause-of-death patterns to be represented by relationships calculated on the basis of a data set that includes both MDCs and LDCs. The typical cause structures pertaining to populations with age-standardized crude death rates of 0.035 and 0.020 are presented in table 5.1.[3] These are roughly the levels that probably best characterize the average mortality situation in LDCs in 1900 and 1970, since they correspond to life expectancies at birth of 27.5 and 50.[4]

What is perhaps surprising about the table is that the specific "name" infectious and parasitic diseases (the first two listed) account for only an estimated 26.1% of mortality change. More important than all of these diseases combined—tuberculosis, typhoid, typhus, cholera, measles, diphtheria, whooping cough, malaria—is the category of respiratory diseases, which comprises a wide assortment of respiratory difficulties that are concentrated largely in infancy and old age. To be sure, some of the deaths in this category are improperly assigned complications of

Table 5.1 **"Normal" Cause-of-Death Patterns at Standardized Crude Death Rates of 35/1,000 and 20/1,000 and Cause-Patterns of Change**

Cause of Death	Model Value (Mean, Male and Female) of Age-Standardized Crude Death Rate from Cause at Age-Standardized Crude Death Rate from all Causes Combined of		Percentage of Decline Attributable to Cause
	35/1,000	20/1,000	
Respiratory tuberculosis	3.85	1.42	9.5
Other infectious and parasitic diseases	4.00	1.51	16.6
Influenza/pneumonia/bronchitis	7.85	2.87	33.2
Diarrheal disease	2.32	1.34	6.5
Maternal causes	.27	.14	0.9
Certain diseases of infancy	1.37	.88	3.3
Violence	.89	.78	0.7
All other and unknown	15.45	11.06	29.3
Total			100.0

Source: Preston and Nelson (1974).

specific infectious diseases. After undergoing a careful review of the initial medical certification in 1962–64, death certificates assigned to influenza/pneumonia/bronchitis in ten cities in Latin America, San Francisco, and Bristol, England, suffered a net loss of 1.0% of the original deaths assigned at ages 15–74, while the total of specific infectious and parasitic diseases increased by 5% (Puffer and Griffith 1967, pp. 230, 235). In a similar study of deaths before age 5 in thirteen Latin American and two North American areas between 1968 and 1972, respiratory diseases forfeited 22% of their originally assigned deaths, while the specific infectious diseases gained 23% (Puffer and Serrano 1973, pp. 332, 342). Corresponding adjustment of figures in table 5.1 would equalize the contribution of the specific infectious diseases and of the respiratory diseases to mortality decline at about 28–30%. Nevertheless, the "name" infectious diseases remain relatively submerged compared with popular accounts of their role, a point also stressed by McDermott (1966). Part of the reason for overemphasis on the role of the "name" diseases in mortality decline is probably their preeminent importance in the relatively small-scale English decline from 1851 to 1901, as has been elegantly documented in a widely cited paper by McKeown and Record (1962).

Some indication of whether the patterns depicted in table 5.1 provide a suitable representation for LDCs can be gained by examining the few LDC records that are available. In so doing, it is useful to provide more 178 detail on specific members of the infectious and parasitic set. Mortality changes during long periods will be considered in order to avoid sampling periods during which specific public health interventions may have badly distorted the cause pattern of change.

Table 5.2 presents crude death rates from certain diseases of infectious origin for five populations of LDCs in the early twentieth century and for a more recent year. By and large, they support the previous estimate of the relative importance of respiratory tuberculosis and diarrheal diseases. Influenza/pneumonia/bronchitis is somewhat less important a source of decline than depicted in table 5.1, but the reason is probably that by the latter date each of the four populations supplying information on this cause achieved a mortality level far superior to the 20/1,000 age-standardized rate assumed in table 5.1. This cause is generally a more important source of mortality decline in movements between high and intermediate levels than between intermediate and low ones (Preston and Nelson 1974, pp. 31–33). As a very shaky generalization based on these undoubtedly unrepresentative populations, it may not be too far off the mark to assign 2% of the twentieth-century mortality decline in LDCs as a whole to smallpox, 2% to whooping cough, and 1% each to typhoid, typhus, measles, cholera, and plague (the latter estimate accounting for the disease's heavy concentration in India), and 0.5% to diphtheria. The epidemic nature of typhoid, cholera, and plague add even more uncertainty to these figures.

But the major uncertainty relates to the role of malaria. If estimates for British Guiana and India are to be believed, malaria by itself has accounted for 18–35% of large-scale mortality declines, equaling or exceeding the contribution we have estimated for all infectious and parasitic diseases combined. It should be noted that both estimates are based upon assignment to malaria of an arbitrary portion of deaths originally ascribed to "fever." A less arbitrary approach was pursued by Newman (1965, 1970), who had access to regional mortality data before and after an eradication campaign in Sri Lanka, as well as to regional data on malarial endemicity. Newman estimates by indirect techniques that the malaria eradication campaign in 1946 reduced Sri Lanka's crude death rate (CDR) by 4.2/1,000 between 1936–45 and 1946–60 (1970, p. 157). The relative contribution of such a decline to the total mortality reduction depends, of course, on the size of the latter. Had it been experienced by one of the populations in table 5.2, where the average drop in CDR was 26/1,000, it would account for about 16% of the mortality reduction. As a component of the smaller decline during the shorter period considered by Newman in Sri Lanka, it represented 42%. Simple inspection of time-series data on crude death rates suggests that virtually complete malaria eradication reduced the crude death rate by

Table 5.2 **Crude Death Rates (per 1,000) by Cause in Certain LDCs in the Early Part of the Century and Recently**

Cause of Death	Chile, 1920[a]	Chile, 1971[b]	% of Decline	Taiwan, 1920[c]	Taiwan, 1966[d]	% of Decline
Typhoid	.665	.007	2.9	.045	0	.2
Typhus	.439	0	2.0	n.a.	—	—
Malaria	.026	0	0	2.123	0	7.7
Smallpox	.008	0	0	.065	0	.3
Measles	.697	.061	2.8	.358	.039	1.2
Whooping cough	.789	.006	3.5	.046	.004	.2
Diphtheria	.062	.006	.2	.013	.004	0
Influenza/pneumonia/ bronchitis	4.527	1.355	14.1	8.231	.613	27.8
Respiratory tuberculosis	2.404	.209	9.8	1.813	.333	5.4
Diarrhea, dysentery, cholera nostras	2.120	.397	7.7	1.691	.215	5.4
Cholera	n.a.	—	—	.452	0	1.6
Total	30.85	8.389		32.686	5.246	

[a]Chile, Oficina Central de Estadistica, *Annuario Estadistico*, vol. 1 (1920).

[b]World Health Organization, *World Health Statistics Annual*, vol. 1 (1972).

[c]Taiwan, Jinki Dotai Tokei, *Sotoku Kanbo Chosaka* (1920).

[d]United Nations, *Demographic Yearbook* (1967), table 24.

[e]Pani (1917, pp. 192–99).

[f]Mandle (1970, p. 303).

[g]United Nations, *Demographic Yearbook* (1973), table 33.

[h]Compiled from material in Davis (1951, pp. 33–53). The 8.7 malaria estimate assumes that one-third of fever deaths are due to malaria. All estimates are highly suspect because of very deficient coding. Data on cholera, plague, and smallpox are probably most accurate because they were most consistently "notifiable" causes of death at the provincial level. Coale and Hoover (1958, p. 67), using a completely different technique, estimate the pre-spraying program level of malaria death rates in India to be about 6/1,000, including deaths indirectly attributable to the disease.

[i]Crude death rate from United Nations, Population Division, *Selected World Demographic Indicators by Countries, 1950–2000* (1975), p. 125. Average, 1960 and 1965. Distribution of deaths by cause from United Nations, *Demographic Yearbook* (1967), table 24. Data are for medically certified deaths in Poona and Bombay corporations and deaths in public hospitals in Rajasthan. The World Health Organization provides some confirmation of the virtual eradication of malaria from India by noting a hundredfold decline in reported cases between 1952 and 1972. World Health Organization, *Fifth Report on the World Health Situation*, Official Records no. 225 (1975), p. 142.

[j]Bunle (1954).

[k]Deaths under age 2 only.

[l]Malaria and undefined fevers.

[m]Pneumonia and bronchitis.

[n]Tuberculosis, all forms.

179

Table 5.2 (continued)

Cause of Death	Mexico City, 1904–12[e]	Mexico, 1922–25[j]	Mexico, 1972[h]	% of Difference, Mexico City–Mexico	% of Decline, Mexico
Typhoid	.068	.341	.065	0	1.7
Typhus	1.363	.040	.001	4.1	.2
Malaria	.076	1.471	.001	.2	9.0
Smallpox	.733	.826	0	2.2	5.0
Measles	.290	.427	.219	.2	1.3
Whooping cough	.284	.938	.080	.6	5.2
Diphtheria	.169	.062	.002	.5	.4
Influenza/pneumonia/ bronchitis	7.838	3.213	1.664	18.6	9.4
Respiratory tuberculosis	2.485	.653	.152	7.0	3.1
Diarrhea, dysentery, cholera nostras	9.785	1.861[k]	1.337	25.4	—
Cholera	n.a.	n.a.	—	—	—
Total	42.314	25.489	9.081		

Cause of Death	British Guiana, 1911–20[f]	Guyana, 1967[g]	% of Decline	India, 1898–1907[h]	India, 1963[i]	% of Difference
Typhoid	n.a.	—	—	n.a.	—	—
Typhus	n.a.	—	—	n.a.	—	—
Malaria	4.185[l]	0	18.1	6.0–8.7	0	24.3–35.2
Smallpox	n.a.	—	—	.27	.08	.8
Measles	n.a.	—	—	n.a.	—	—
Whooping cough	n.a.	—	—	n.a.	—	—
Diphtheria	n.a.	—	—	n.a.	—	—
Influenza/pneumonia/ bronchitis	3.680[m]	.055	15.7	n.a.	—	—
Respiratory tuberculosis	1.432[n]	.024	6.1	n.a.	—	—
Diarrhea, dysentery, cholera nostras	2.780	.026	11.9	1.96	.14	7.4
Cholera	n.a.	—	—	1.66	1.18	1.9
Plague	n.a.	—	—	1.82	0	7.3
Total	30.049	6.987		43.5	18.8	

about 3/1,000 in Guatemala (Meegama 1967, pp. 231–33), by 5–9 points in Mauritius (Titmuss and Abel-Smith 1968, pp. 49–50), and by 2–3/1,000 in Venezuela (Pampana 1954, p. 504).

The importance of malaria reduction as a source of declining mortality in a country obviously depends upon initial endemicity and the success of antimalarial campaigns. Southern Africa, southern Latin America, and northern Asia were never seriously afflicted with the disease; malaria in tropical Africa is highly endemic but, with few exceptions, it has not been successfully attacked (World Health Organization 1975*a*). A valuable compilation by Faust (1941, p. 12) suggests that recorded crude death rates from malaria in Mexico, Central America, and the West Indies during the 1930s, before any inroads had been made against the disease in rural areas, was 1.66/1,000. But he considers this "a figure probably far too low." The recorded CDR from malaria in cities of Burma in 1939 was 2.14 (Simmons et al. 1944, p. 11). The preprogram malarial CDR in Venezuela was 1.73/1,000 (Pampana 1954, p. 504). But recorded levels are often greatly in error. Newman calculates for Ceylon that each death assigned to malaria represented approximately four deaths that were directly or indirectly caused by it. But such an inflation factor is simply not tenable in Guatemala or Venezuela, where contemporaneous declines in crude death rates and malarial crude death rates suggest that a 2:1 ratio is the most that could have been sustained. Part of the inflation factor reflects malaria's role as "the great debilitator," but another part may be spurious. Spraying with residual insecticides reduces not only malaria but other vector-borne diseases such as yellow fever, typhus, and especially diarrheal disease. As Newman points out, an indirect approach that bases estimates on relations between regional changes in aggregate mortality and changes in spleen rates may well overascribe mortality decline to malaria reductions (but not to insecticide campaigns themselves). Finally, there is an often-quoted estimate, the basis of which is unknown, that worldwide malarial deaths have declined from 2.5 million per year to less than 1 million (*Lancet* 1970, p. 599), figures suggesting for LDCs a decline in crude death rates on the order of 1/1,000. The 2.5 million estimate evidently first appears in Pampana and Russell (1955) and presumably applies to 1955. Russell put the figure at at least 3 million for 1943 (1943, p. 601).

There is obviously much work to be done on this issue. In the present state of semi-ignorance, it seems judicious to adopt a range of CDR declines of 2–5/1,000, within which the twentieth-century LDC mortality decline attributable to antimalarial programs seems to have a better than even chance of falling. Sri Lanka, the best-documented case, falls within this range, and Sri Lanka was apparently intermediate in terms of initial endemicity, although the program there probably enjoyed unusual success and malaria reduction may not be solely responsible for the mortality decline produced by insecticide spraying. The range is also consistent with the apparently widespread malarial CDR declines of 1.5–2.0/1,000 and with inflation factors of 1.4–2.5.

These assorted scraps of information are pieced together in table 5.3, where diseases that seem in the main to be responsible for mortality declines in LDCs between 1900 and 1970 are classified into three groups according to their dominant mode of transmission. An estimate—obvi- 180

Table 5.3 Diseases Responsible for LDC Mortality Declines and Methods That Have Been Used against Them

Dominant Mode of Transmission	Diseases	Approximate Percentage of Mortality Decline in LDCs, 1900–1970, Accounted for by Disease	Principal Methods of Prevention Deployed[a]	Principal Methods of Treatment Deployed[a]
Airborne	Influenza/Pneumonia/Bronchitis	30		Antibiotics
	Respiratory tuberculosis	10	Immunization; identification and isolation	Chemotherapy
	Smallpox	2	Immunization	Chemotherapy
	Measles	1	Immunization	Antibiotics
	Diphtheria / Whooping cough	2	Immunization	Antibiotics
		45		
Water-, food-, and fecesborne	Diarrhea, enteritis, gastroenteritis	7	Purification and increased supply of water; sewage disposal; personal sanitation	Rehydration
	Typhoid	1	Purification and increased supply of water; sewage disposal; personal sanitation; partially effective vaccine	Rehydration, antibiotics
	Cholera	1	Purification and increased supply of water; sewage disposal; personal sanitation; partially effective vaccine; quarantine	Rehydration
		9		
Insectborne	Malaria	13–33	Insecticides, drainage, larvicides	Quinine drugs
	Typhus	1	Insecticides, partially effective vaccines	Antibiotics
	Plague	1	Insecticides, rat control, quarantine	
		15–35		

[a]Major sources: Paul (1964); Morley (1973); Hinman (1966).

ously highly tentative but nevertheless the first that appears to have been offered—of the relative importance of the various diseases in the decline is also supplied. The category totals are somewhat more robust than the figures for individual diseases because diagnostic confusions, disease interactions, and program externalities are more likely to occur within groups than between groups. Since progress was very slow between 1900 and 1935, the listing may serve as an adequate representation of declines between 1935 and 1970 as well. Finally, some indication of the public health and medical instruments that have been deployed against the various diseases is also provided. The modes of transmission that are listed are not mutually exclusive or exhaustive, but the classification is not seriously distortive.

5.1.3 Influences Operating on the Various Causes of Death

Mortality from every disease listed in table 5.3 would be expected to decline when personal living standards rise. Of the many linkages, probably the most important are those between nutritional status and influenza/pneumonia/bronchitis, diarrheal disease, and respiratory tuberculosis. The mechanisms of effect are not well known, but it appears that protein malnutrition impairs the production of circulating antibodies in response to bacterial and viral antigens and that undernutrition can produce atrophy of the organs responsible for the immune response (World Health Organization 1972, pp. 24–25). There is no question of the importance of poor nutrition as a factor underlying high mortality rates in LDCs. The PAHO study of child mortality in thirteen Latin American projects found that immaturity or malnutrition was an associated or underlying cause of 57% of the deaths before age 5 (World Health Organization 1974*a*, p. 279). Immaturity is in turn a frequent product of maternal malnourishment (Mata et al. 1972). The problem is apparently equally severe in Africa, although the data are much more fragmentary (Bailey 1975). Diet supplementation programs in Peru (Baertl et al. 1970) and Guatemala (Ascoli et al. 1967; Scrimshaw 1970) significantly reduced child mortality in test populations, but, oddly, without having substantial effects on indexes of child physical development. Despite the improvement, mortality and morbidity in the Guatemalan villages remained "shockingly high," which was attributed to irregular participation in the feeding program and to the continued heavy burden of infection to which the children were subject (Scrimshaw 1970, pp. 1689–90).

But nutritional status is not exclusively determined by diet, nor is diet determined only by the availability of calories or protein. There is now extensive evidence that infectious diseases themselves are an extremely important source of malnourishment, independent of the child's nutritional state at the time of attack (Mata et al. 1972; Scrim- 181 shaw 1970; World Health Organization 1972, p. 27). Infection increases metabolic demands and often reduces the absorption of nutrients and increases their excretion. Nutritional intake can also be reduced by nausea or through customs denying food to the sick. Infections among pregnant women can reduce birthweights and, among new mothers, milk secretion (Mata et al. 1972; Bailey 1975). Gordon (1969, p. 218) suggests that diet fails by far to explain all the prevalent malnourishment in LDCs. The frequency of inappropriate nutritional practices despite adequate food supplies does not require emphasis (Bailey 1975; Food and Agriculture Organization 1975). The importance of nutritional practices is indicated by the reported reversal of expected social class differences in infant mortality in Chile, a condition attributed to earlier weaning among children of upper-class women (Plank and Milanesi 1973). The point is simply that, even if nutritional status were the only influence on mortality from a disease, mortality declines from that cause do not necessarily imply that an improvement in food supplies has occurred. Nutritional practices and exogenous declines in the incidence of other infections must also be considered candidates for explanation. The "subsistence level" of food production is obviously fictitious if it is presumed to represent a fixed requirement that is independent of the state of prophylactic or nutritional arts.

Despite the undoubted influence of nutritional intake and other components of general living standards on mortality, it is clear from table 5.3 that many other influences have also been at work. Obvious as it may be, it is easy to forget that death from an infectious disease involves an encounter between a pathogenic organism and a vulnerable human host. The rate of death can be altered by changing the rate or terms of the encounter without any prior change in the host. It is not possible and is probably not necessary to document individually the preventive and curative measures that have been utilized for this purpose. With the exception of influenza/pneumonia/bronchitis, it seems likely that preventive measures have been more effective than curative ones.

That preventive measures have been widely deployed in each of the three categories can be demonstrated relatively easily. At the end of 1964, 1.935 billion persons lived in areas that were originally malarious. Of these, 41% were living in areas from which malaria had been eradicated; 16% were living in areas where incidence was very low and was being controlled by case detection and treatment; 24% were living in areas protected by extensive mosquito control measures; and 19% were living in areas without specific antimalarial measures, most of these in tropical Africa (World Health Organization 1975*a*). The cost of programs producing virtually complete eradication has been estimated at 10–30 cents per capita per year, with programs probably extending over a two- to three-year period and a continuing annual cost of 5 cents per capita required for surveillance thereafter (Pampana and Russell 1955, table 1, provides the most complete cost compilation).

India has vaccinated 170 million persons against tuberculosis and in 1968 alone vaccinated 83 million against smallpox (World Health Organization 1975*b*, p. 142). It has finally succeeded in eliminating smallpox completely, as has every other country. Colombia dispensed 5.9 million vaccinations in 1972, one-third of the population size; Egypt vaccinated 15 million persons against cholera in the same year and 25 million against smallpox in 1970. All primary-school entrants in the Philippines are vaccinated against tuberculosis. Barbados has compulsory immunization against diphtheria, polio, smallpox, and tetanus upon school entrance (World Health Organization 1975). An expert committee assembled by WHO estimated that 80% of the 70 million children reaching age one in LDCs each year could be immunized against measles, polio, tuberculosis, pertussis, tetanus, diphtheria, and smallpox at a cost, exclusive of personnel, of $37.5–$60 million, or $0.67–$1.07 apiece (World Health Organization, 1975*c*, p. 2).

Conditions of water supply and sewage disposal have also been markedly improved, despite continued abysmal conditions in many areas. Many of the improvements were normal and integral parts of economic expansion and hence cannot be specifically interpreted as public health interventions of an unprecedented sort. WHO surveys of government officials in 1962 and 1970 indicated that the proportion of LDC urban populations served by house connections to public water supply increased from 0.33 to 0.50 during this eight-year period for the seventy-five countries that replied in both years (World Health Organization 1973, p. 726). No rural population comparisons could be made, but in 1970 only 12% of rural LDC populations had "reasonable access" to community water supply (public fountain or standpipe within 200 meters of a house). The figure was only 6% in India (World Health Organization 1973, pp. 727, 729). New urban house connections were estimated to cost an average of $35 per capita, and providing rural residents with easy access to safe water to cost $12 per capita. In 1970, 69% of LDC urban populations had sewage disposal facilities (27% were connected to public sewerage and 42% had private household systems). New connections to urban public sewerage cost an average of $29 per head. Only 8% of the rural population was judged to have adequate sewage disposal, although the average cost of providing such facilities was estimated at only $4 per capita (World Health Organization 1973, pp. 732–33, 738–43). No trends in sewage facilities could be established, but improvement is probably fairly rapid in urban areas and slow in rural areas. Clearly, the initial cost of such programs is considerably higher than that for programs of vector control and immunization. Water supply improvements were among the very first changes to modify mor-

182

tality patterns in European countries in the mid-nineteenth century. But they are lagging relative to other improvements in LDCs. This may explain why diarrheal disease remains relatively more prominent as a contributor to total mortality in LDCs than it was in European countries at the same general mortality level (Preston and Nelson 1974).

With the exception of water and sewerage improvements and smallpox vaccination, the techniques of preventive and curative health care that have been widely deployed in LDCs are twentieth-century products. Virtually all were facilitated by ultimate acceptance of the revolutionary germ theory of disease at the turn of the century. Even smallpox eradication has benefited from technical improvements such as freeze-dried vaccine and the forked needle (Foege et al. 1975). The next section attempts to identify the relative importance of these technical improvements, as typically embodied in government health programs enjoying some measure of external assistance or support, in LDC mortality improvements during the last three decades.

5.1.4 Structural Changes in Relations between Mortality and Other Development Indexes

That mortality reductions have not merely been residual by-products of socioeconomic development is best illustrated by showing that major structural changes have occurred in the relationship between mortality and other indexes of development. Important technical changes and exogenous increases in government health commitment or foreign health assistance should result in a shift in the average level of life expectancy that corresponds to a particular level of other development indicators. Preston (1975*a*) suggested that such a change had occurred in the relation between mortality and national income, and this section will attempt to supplement that observation by introducing new variables and a larger sample of countries. Data have been gathered on national levels of life expectancy, per capita income (in 1970 U.S. dollars), daily calorie consumption, and literacy for thirty-six nations in or about 1940, including seventeen LDCs and several others that today would be classified as LDCs if 1940 conditions had persisted. The LDC estimates of life expectancy are based largely upon indirect demographic techniques such as intercensal survival analysis rather than upon vital statistics. Data on these same variables have been generated for 120 nations in or around 1970. The data and sources are presented in appendix tables 5.A.1 and 5.A.2.

A preliminary indication that structural changes have occurred is presented in table 5.4. Countries are cross-classified by level of per capita national income (in 1970 U.S. dollars) and by daily calorie consumption per head. Complete information was available for only twenty-nine countries in 1940. Nevertheless, it is clear that, within every

Table 5.4 **Mean Life Expectancy at Birth of Countries in Various Ranges of National Income and Caloric Consumption, 1940 and 1970**

Daily Calories Per Capita	National Income per Capita in 1970 U.S. Dollars <150	150–299	300–699	700+	
<2,100	42.7 (17)	51.5 (8)	53.3 (5)	69.5 (1)	47.5 (31)
	38.3 (5)	36.0 (1)	(0)	(0)	37.9 (6)
2,100–2,399	42.6 (16)	49.9 (14)	56.2 (7)	71.4 (2)	49.1 (39)
	40.0 (1)	43.9 (2)	46.1 (1)	(0)	43.4 (4)
2,400–2,899	45.4 (1)	57.9 (8)	61.3 (10)	68.0 (7)	61.4 (26)
	(0)	44.1 (2)	50.4 (4)	59.6 (2)	51.1 (8)
2,900+	(0)	(0)	(0)	71.6 (24)	71.6 (24)
	(0)	(0)	58.7 (2)	65.2 (9)	64.0 (11)
	42.7 (34)	52.4 (30)	57.8 (22)	70.8 (34)	55.9 (120)
	38.6 (6)	42.4 (5)	52.2 (7)	64.1 (11)	52.2 (29)

Source: Appendix tables 5.A.1 and 5.A.2.
Note: 1970 countries appear in the top rows, 1940 countries in the bottom rows. The number of countries is shown in parentheses.

one of the nine cells where both 1940 and 1970 populations appear, average life expectancy was higher at the later date. The average intra-cell gain is 8.7 years of life expectancy.

A somewhat more precise indication of the magnitude of structural changes can be obtained by regressing life expectancy on income, calories, and literacy separately for 1940 and 1970 observations. Because of nonlinearities expected on obvious inductive and deductive grounds, natural logarithms of calorie consumption and income are used as regressors. Daily calorie consumption is measured from 1,500, approximately the average level required to meet minimum daily metabolic demands. Literacy, a personal dichotomous variable, cannot act nonlinearly at the individual level, since it takes on only two values. Barring spillover effects whose existence in income relations was called into question in section A, the proportion literate should be linearly related to life expectancy at the aggregate level. No claim is made that the resulting equations are perfectly specified, but simply that the socioeconomic variables included are the only ones available in the 1930s. It seems unlikely that relations between terms omitted and terms present 183 would have changed in such as way as to influence the outcomes described below.

The equations as estimated by ordinary least squares are the following:[5]

$$1970:\ e^{o}_{0} = \underset{(7.4090)}{17.1464} + \underset{(.6524)}{4.2488} \times \ln \overline{Y} + \underset{(.0212)}{.2086} \times LIT + \underset{(1.3492)}{.3170} \times \ln CAL$$

$$N = 120,\ R^2 = .860 \qquad \bar{R}^2 = .858$$

$$1940:\ e^{o}_{0} = \underset{(18.5102)}{-13.1035} + \underset{(2.3860)}{5.4352} \times \ln \overline{Y} + \underset{(.0626)}{.1654} \times LIT + \underset{(3.7176)}{2.9470} \times \ln CAL$$

$$N = 36,\ R^2 = .856 \qquad \bar{R}^2 = .845,$$

where e^{o}_{0} = life expectancy at birth, average male and female
$\overline{Y}$ = national income per capita, 1970 U.S. dollars
LIT = percentage literate of the adult population
CAL = excess of daily calorie consumption per capita over 1,500.

Coefficients of all three variables in both equations are properly signed. The explanatory power of the regression equations is virtually identical for the two years. Income and literacy terms are highly significant in both periods and retain approximately the same magnitude. This stability was unexpected because of the high degree of colinearity among regressors. The coefficients indicate that a 10 percentage point increase in literacy is associated at both points with a gain in life expectancy of approximately 2 years, and that a 10% gain in national income by itself increases life expectancy by approximately one-half year. Coefficients of the calorie term decrease over time but are insignificant in both periods. It is very unlikely that the availability of calories for daily consumption has no influence on mortality. The calorie variable is probably subject to greater measurement error than the other two, and the influence of calorie availability is probably being reflected through them. The constant term increases by about 30 years between 1940 and 1970, although by itself this change is not readily interpreted, since the zero-points on variables are well below the range of observed experience. The hazards of extrapolation are shown by the negative (though insignificant) intercept for 1940.[6]

The substantive significance of the structural shift, as reflected primarily in the intercept, is probably best illustrated in the following way. Each of the 120 countries in 1970, including 94 LDCs, can supply estimates for each of the three regressors. It is therefore possible to estimate what life expectancy would be for every country at its current developmental level if no structural change had occurred in the relation between mortality and socioeconomic development. This estimate is simply obtained by substituting values of the three regressors for 1970 into the 1940 regression equation. Differences between actual life expectancy and that predicted if 1940 relations had continued to prevail indicate the amount of change in life expectancy attributable to the structural shift. A weighted average of such differences will indicate the importance of the shift for LDCs as a whole.[7] Results of this exercise are presented in table 5.5.

Estimates presented in this table indicate that life expectancy for LDCs as a unit (exclusive of China, North Korea, and North Vietnam) would have been 8.66 years lower in 1970–75 if life expectancy had continued to be related to other development indexes as it was in 1940. Excluding South Vietnam, where special factors were obviously distorting life expectancy, the figure is 8.84. This is an estimate of the amount of increase in life expectancy that is attributable to factors exogenous to national levels of income, literacy, and calorie consumption.[8] What fraction of the total gain in life expectancy during the period this 8.84-year structural shift represents is difficult to assess. WHO estimates that life expectancy in LDCs was 32 years in 1935–39 (30 in Africa and Asia and 40 in Latin America) and 49.6 years in 1965–70 (World Health Organization 1974, p. 23). The earlier figure is based on very little information, but if we accept it, the implication is that the structural change accounts for about half (50.2%) of the total gain in life expectancy during these nearly equivalent 30-year periods. This estimate is lower than the 79.5% (9.7/12.2) figures estimated for MDCs and LDCs combined by Preston (1975, p. 238) not so much because the estimated Δe^o_0 attributed to structural shifts differ (the difference is in fact only 0.86 years) but because the estimated gains in life expectancy differ (12.2 years for the world as a whole between 1938 and 1963 by Preston versus 17.6 years for LDCs between 1935–39 and 1965–70 by WHO). Part of the discrepancy in the estimates probably results from differences in the universe covered. Because MDCs had achieved by 1940 levels of developmental indicators high enough that relatively little gain in life expectancy was to be expected from advances in living standards, it is likely that exogenous factors represented a larger fraction of the gains that occurred there than they did in LDCs. Another part may reflect differences in the periods covered, since the present estimate pertains to a somewhat later period. Suggestions that the pace of mor-

Table 5.5 **Life Expectancy in 1970–75 and Life Expectancy Predicted if 1940 Relations between Life Expectancy and Levels of Literacy, Income, and Calorie Consumption Had Continued to Prevail**

Africa	Predicted e^0_0*	Actual e^0_0	Difference	*Latin America*	Predicted e^0_0*	Actual e^0_0	Difference	*Asia*	Predicted e^0_0*	Actual e^0_0	Difference
Algeria	41.42	53.20	11.78	Argentina	61.84	68.20	6.36	Afghanistan	30.16	40.30	10.14
Angola	37.32	38.50	1.18	Bolivia	39.49	46.80	7.31	Bangladesh	31.49	35.80	4.31
Botswana	34.13	43.50	9.37	Brazil	50.85	61.40	10.55	Burma	41.23	50.00	8.77
Burundi	40.39	39.00	−1.39	Chile	56.62	62.60	5.98	Cyprus	56.21	71.40	15.19
Central African Rep.	35.16	41.00	5.84	Colombia	51.36	60.90	9.54	India	40.11	49.50	9.39
Chad	30.29	38.50	8.21	Costa Rica	56.07	68.20	12.13	Indonesia	39.54	47.50	7.96
Congo	42.47	43.50	1.03	Dominican Rep.	48.38	57.80	9.42	Iran	43.23	51.00	7.77
Dahomey	33.60	41.00	7.40	Ecuador	47.14	59.60	12.46	Iraq	42.57	52.70	10.13
Egypt	41.95	52.40	10.45	El Salvador	45.09	57.80	12.71	Israel	63.10	71.00	7.90
Ethiopia	30.39	38.00	7.61	Guatemala	43.86	52.90	9.04	Jordan	43.59	53.20	9.61
Gabon	41.65	41.00	−.65	Guyana	51.21	67.90	16.69	Khmer Rep.	47.19	45.40	−1.79
Gambia	33.47	40.00	6.53	Haiti	29.48	50.00	20.52	Korea, Rep. of	52.20	60.60	8.40
Ghana	43.10	43.50	.40	Honduras	44.10	53.50	9.40	Laos	32.95	40.40	7.45
Guinea	30.43	41.00	10.57	Jamaica	54.93	69.50	14.57	Lebanon	55.11	63.20	8.09
Ivory Coast	41.97	43.50	1.53	Mexico	55.24	63.20	7.96	Malaysia	50.43	59.40	8.97
Kenya	37.37	50.00	12.63	Nicaragua	48.51	52.90	4.39	Nepal	31.63	43.60	11.97
Liberia	35.68	43.50	7.82	Panama	55.47	66.50	11.03	Pakistan	37.05	49.80	12.75
Libyan Arab Rep.	52.23	52.90	.67	Paraguay	47.84	61.90	14.06	Philippines	47.95	58.40	10.45
Madagascar	39.55	43.50	3.95	Peru	46.72	55.70	8.98	Saudi Arabia	42.56	45.30	2.74
Malawi	33.05	41.00	7.95	Puerto Rico	59.60	72.10	12.50	Singapore	55.22	69.50	14.28
Mali	28.99	38.00	9.01	Trinidad and Tobago	54.17	69.50	15.23	Sri Lanka	46.86	67.80	20.94
Malta	58.07	70.80	12.73	Uruguay	59.54	69.80	10.26	Syria	44.14	54.00	9.86
Maritius	49.47	65.50	16.03	Venezuela	55.14	64.70	9.56	Taiwan	54.32	69.40	15.08
Mauritania	33.50	38.50	5.00					Thailand	48.09	58.00	9.91
								Turkey	49.03	56.90	7.87

Table 5.5 (continued)

Africa	Predicted e^0_0*	Actual e^0_0	Difference	*Latin America*	Predicted e^0_0*	Actual e^0_0	Difference	*Asia*	Predicted e^0_0*	Actual e^0_0	Difference
Morocco	39.92	52.90	12.98	Mean difference, Latin America			= 10.90	Vietnam, Rep. of	49.08	40.50	−8.58
Mozambique	38.98	43.50	4.52	1970 population-weighted mean difference, Latin America			= 9.54	Yemen	31.53	44.80	13.27
Niger	30.20	38.50	8.30					Yemen, P.D.R.	31.79	44.80	13.01
Nigeria	37.52	41.00	3.48								
Rhodesia	38.39	51.50	13.11								
Rwanda	29.66	41.00	11.34								
Senegal	36.25	40.00	3.75								
Sierra Leone	35.25	43.50	8.25								
Somalia	28.37	41.00	12.63								
South Africa	50.59	51.50	.91								
Sudan	34.70	48.60	13.90								
Togo	33.93	41.00	7.07								
Tunisia	42.32	54.10	11.78								
Uganda	37.62	50.00	12.38								
Cameroon	45.40	41.00	−4.40								
Tanzania	32.00	44.50	12.50								
Upper Volta	28.92	38.00	9.08								
Zaire	37.16	43.50	6.34								
Zambia	45.02	44.50	−.52								
Mean difference, Africa			= 7.05	Mean, all LDCs			= 8.61	Mean difference, Asia			= 9.14
1970 population-weighted mean difference, Africa			= 7.22	1970 Population-weighted mean, all LDCs			= 8.66	1970 population-weighted mean difference, Africa			= 8.90

*Based on substitution of 1970 values of literacy, income, and calorie consumption into 1940 regression relating e^0_0 to these variables.

tality decline in LDCs has slowed in the past decade (Hansluwka 1975; World Bank 1975) imply that the shift in the mortality/development relation may have essentially ended by the early 1960s, while gains in living standards continue to exert an influence on mortality. In any case, the estimated amount of the structural shift is consistent between the two estimates at about 9 years of life expectancy at birth.

The structural shift has evidently been least pronounced for African countries and most pronounced for Latin America. Africa has unquestionably experienced the least penetration by modern public health measures of any region. The problem is not simply poverty but also a widely dispersed population that increases program costs (World Health Organization 1975*b*, p. 17). Several of the African countries have lower life expectancies in 1970–75 than could have been expected based on 1940 relations. The apparent advantage enjoyed by Latin American countries may be due to their special relations with the United States. The United States has been by far the largest bilateral donor in international health aid, and the bulk of aid appears to go to Latin American countries, either directly or through the Pan American Health Organization (World Bank 1975, pp. 68–69). It is worth noting that, of the ninety-four LDCs, Sri Lanka and Mauritius are two of the four whose estimated structural changes are largest. It is unfortunate that so much attention has focused on these unusual cases.

Attributing to all countries the relations prevailing in countries for which data are available is always risky. The preceding analysis of change can be complemented by one that focuses exclusively on the cases that can be documented. Each of the thirty-six countries providing data in 1940 can also supply data in 1970. According to the previous formulation, we should expect Δe^o_0 to be linearly related to $\Delta \ln Y$, $\Delta \ln CAL$, and ΔLIT, with a relatively large positive intercept reflecting the structural shift. In the first specification of the model, we add three terms believed to reflect factors responsible for a portion of the structural shift. The first (MAL) is an estimate of the degree of malarial endemicity in 1940, which is a proxy for the effect of antimalarial programs on Δe^o_0. Each of the thirty-six countries with endemic malaria has had a major antimalarial campaign. It is hoped that the coefficient of this term will provide a clearer indication of the effect of antimalarial activities on gains in life expectancy than was previously available. The second (AID) is an estimate of the average annual per capita nonmilitary aid in United States dollars received from bilateral and multilateral donors between 1954 and 1972. The third (WAT) is an estimate of per capita aid received for water and sewerage projects between 1965 and 1970 (U.S. dollars). The latter two variables are assumed to be proxies for the amount of total per capita health aid received between 1940 and 1970. Their values are generally highest for the Latin American countries. Values of these variables are presented in appendix table 5.A.3.

The estimated equation with all six terms present is the following:[9]

$$\begin{aligned}
\Delta e^o_0 = {} & \underset{(2.8468)}{6.5212} + \underset{(2.4111)}{3.4500} \times \Delta \ln \overline{Y} + \underset{(.0927)}{.0354} \times \Delta LIT \\
& + \underset{(4.9362)}{.5605} \times \Delta \ln CAL \\
& + \underset{(.9411)}{3.1328} \times MAL + \underset{(.2376)}{.1460} \times AID + \underset{(.3668)}{.1955} \times WAT
\end{aligned}$$

$$R^2 = .595$$

$$\overline{R}^2 = .506.$$

Each of the coefficients has the expected positive sign. The coefficient of income remains similar in absolute value to that estimated in the cross-sectional regressions, but that of literacy declines by a factor of five and calories remain an insubstantial factor. Receipt of external aid contributes positively but insignificantly to mortality improvement. The most interesting result refers to the constant term. For a country essentially free of malaria in 1940 ($MAL = 0$), it is estimated that life expectancy would have increased by 6.52 years in the absence of socioeconomic development and external aid during the three decades. For a country in which malaria was highly endemic ($MAL = 3$), the corresponding gain is 15.92 years, of which 9.40 is attributable to factors associated with malarial endemicity. The average life expectancy for the seventeen LDCs was 39.29 in 1940 and 59.42 in 1970, giving an average gain of 20.13 years. The average malaria endemicity score for these seventeen was 2.59. Of the total gain in e^o_0, 8.11 years (3.1328×2.59), or 40%, is attributable to factors associated with malarial endemicity, and the constant term of 6.5 years, or 23%, represents other exogenous factors. The sum of 72% is considerably higher than that implied by the previous procedure. The external aid terms contribute an additional 1.13 years, or 5.6%.[10]

Whether or not antimalarial programs themselves produced the gain of 8.11 years attributed to malarial endemicity remains in serious doubt. The malaria score is correlated with life expectancy in 1940 at —.873. It is thus acting as a proxy for the initial level of mortality from a host of potentially eliminable infectious and parasitic diseases. When the initial level of life expectancy is entered as an independent variable, the 187
magnitude of the structural change remains roughly the same but the portion attributable to the malarial term declines to zero:

$$\begin{aligned}\Delta e^o_0 = {} & \underset{(8.5108)}{31.4722} + \underset{(2.0533)}{3.6048} \times \Delta \ln Y + \underset{(.0790)}{.0430} \times \Delta LIT \\ & - \underset{(4.2327)}{.9865} \times \Delta CAL \\ & - \underset{(1.3059)}{.0211} \times MAL + \underset{(.2036)}{.0750} \times AID + \underset{(.3139)}{.2939} \times WAT \\ & - \underset{(.1328)}{.4063} \times e^o_0(1940)\end{aligned}$$

$$R^2 = .720$$

$$R^2 = .643.$$

Other coefficients are not affected in such a way as to substantially alter interpretations, but the coefficient of malarial endemicity becomes effectively zero. The amount of structural change for the seventeen LDCs, the gain that is not accounted for by changes in Y, LIT, or CAL, is 16.46 years of life, or 81.8% of the average gain during the period.[11] The estimated structural change is close to that estimated directly above, but malaria's role in it is now negligible. When malarial endemicity is operationalized as a series of dummy variables, none of the dummy coefficients is significant, and the relation between mortality change and endemicity is nonmonotonic. Other functional forms and variable operationalizations should be investigated; at the moment all we can conclude is that the longitudinal analysis provides no better fix on malaria's role than the largely inconclusive cause-of-death analysis.

Two estimates have been advanced of the fraction of LDC gains in life expectancy between 1940 and 1970 that are attributable to structural change. The first estimate of one-half was based on a regression-decomposition technique that assumed all nations had relations between mortality and development indexes in 1940 identical to those prevailing in nations that could supply data for that year. The second estimate of approximately 80% was based solely upon examination of trends in the latter group. There is an important technical reason to favor the former estimate, namely, that measurement error is likely to be a more important source of distortion in longitudinal than in cross-sectional data. Measurement error biases coefficients toward zero. If random measurement error were all that was reflected in our measured changes in income, literacy, and calorie consumption, then the entire change in life expectancy would be absorbed in the constant term and attributed to structural change, regardless of the actual importance of these factors. Development levels are undoubtedly better measured than development rates, giving greater stability to the analysis based on a comparison of cross sections. There are, however, indications that a fraction of the gain larger than one-half would be attributed to structural shifts if analysis had focused more narrowly on the period between 1940 and 1960.

The estimate of one-half is roughly consistent with the preceding cause-of-death analysis. Influenza/pneumonia/bronchitis has accounted for perhaps a third of the mortality decline. No effective preventive measures have been deployed against these diseases, the effectiveness of immunization being minimal, and there are suggestions that antibiotics, sulfa drugs, and curative services are not widely enough available in LDCs to have substantially altered the disease picture (Sharpston 1976; Bryant 1969, pp. 314–23). Diarrheal diseases probably account for another 9% or so of the decline, and the principal method of control has been improvements in water supply and sewerage that, because of their expense, are closely associated with economic development.[12] It is likely that social and economic development—especially as reflected in water systems, nutrition, housing, and personal sanitary knowledge—have operated largely through these diseases. In the case of other diseases it appears that programs of a narrowly public health nature that have embodied inexpensive new techniques, especially vector control and immunization, have been the decisive forces in mortality reductions.

5.1.5 The Role of MDCs in LDC Gains

Many have argued that MDCs have played a decisive role in the mortality declines experienced by LDCs, although the case has not been well documented. Certain of the influences are clear enough. Sulfa drugs, antibiotics, and most vaccines and insecticides, including DDT, have been developed in laboratories within MDCs. MDCs contributed 5,764 technical assistance workers in health services to LDCs in 1968 (Organization for Economic Cooperation and Development, n.d., pp. 276–77). Governmental health agencies were often created under colonial auspices. The role of external financing has also been stressed, but the accounts have focused on the dramatic examples in relatively small countries where international campaigns have often been undertaken largely for their demonstration value.

Health assistance in the developing world began with the work of medical missionaries, who were established in the Philippines in 1577 and in China by 1835 (Maramag 1965; Bowers 1973).[13] The early efforts of colonial governments were designed primarily to protect the colonials from epidemic diseases (Beck 1970). Correspondingly, cooperative international health efforts principally attempted to protect Europe and North America from imported cholera, plague, and yellow fever (Howard-Jones 1974). An evolving social conscience in the interwar years led to greater concern with the health of the native population itself and to the establishment of local medical colleges (Beck 1970). The most effective international efforts of the period were undoubtedly those of the Rockefeller Foundation, which led a successful campaign to eradicate yellow fever from Latin American cities in 1916–23, repelled the invasion of *Anopheles gambiae* into Brazil in 1938, financed medical schools around the world, and was "probably the largest single factor in improving the public health education of the world up to the creation of the WHO" (Goodman 1971, pp. 381, 266, 377–82). Its antimalarial activities began in 1915 and included demonstrations of the superior cost effectiveness of vector control compared with treatment and the feasibility of complete eradication. The antimalarial activities were considered by Russell, one of the world's leading malariologists, to be of fundamental importance in ultimate control of the disease. "In instance after instance the foundation provided the catalyst, or the inexpensive mainspring, or the seed money that resulted in control of the disease" (Russell 1968, p. 644).

The total amount of money appropriated by the Rockefeller Foundation from 1914 to 1954 for antimalarial activities, exclusive of salaries and overhead was only $5 million (Russell 1968, p. 644). This is a vivid illustration that contributions to mortality change are inaccurately reflected on financial ledgers. International aid for health purposes is a small part of total health expenditures in LDCs and probably always has been. But its cost effectiveness has certainly far surpassed the average for internally financed appropriations, which are too often focused on expensive curative services in urban areas.[14]

Only crude indications are available of the relative magnitudes of internal and external sources of health expenditures in LDCs. In 1970, government health expenditures were estimated for LDCs containing 1.89 billion people. Total government expenditure on health in these areas came to $7.67 billion, or about $4 per capita.[15] Private health expenditures in LDCs are probably slightly larger than public expenditures, judging from comparisons that can be made in seven countries.[16] Addition of private and public expenditure in countries not represented would bring the total annual expenditure perhaps to the range of $20–30 billion.

In contrast, the largest single source of international assistance for health, the World Health Organization, dispersed only $115 million in 1972, a figure that includes family planning activities and some dispersals to MDCs (World Bank 1975, pp. 68–69). The annual budget of the World Health Organization in 1970 was less than that of Massachusetts General Hospital! (Goodman 1971, p. 223). Of its regular budget, the United States contributed 31% and the USSR 13%, with no other country making a contribution larger than 7% (Goodman 1971, p. 220). The second largest source of international assistance in 1972, USAID, contributed $42 million. All together, the ten largest multilateral or bilateral sources of health aid contributed $300.7 million in international assistance for health programs in 1972 (World Bank 1975, p. 68), probably between 1% and 2% of total health expenditures in LDCs. To this should be added a portion of the $79 million in loans and credits made by the World Bank for water supply and sewerage construction in that year (World Bank 1975, p. 48).

It seems very likely, then, that total external health aid received by LDCs is less than 3% of their total health expenditures. The figure may have been somewhat higher earlier in the postwar period. The cumulative United States contribution to antimalarial activities through national research and international assistance has been estimated to be about one-half billion dollars (Russell 1968), but the annual contribution has declined drastically (Weller 1974; World Health Organization 1975*a*). But even the cumulative total is a paltry figure compared with annual expenditures in LDCs themselves. MDC contributions to mortality declines in LDCs have not been primarily financial; according to the estimates of the preceding section, the financial contributions are associated with an increase in e^{o}_{0} of about one year in the seventeen LDCs between 1940 and 1970. Instead, they seem to have consisted of the development of low-cost health measures exploitable on a massive scale, demonstration of their effectiveness in relatively small areas, training and provision of personnel, and occasionally the initiation of large-scale programs whose major cost was often absorbed by the recipient country.[17] When action appeared to be remarkably cost effective and timely, such as a campaign to eradicate smallpox from West and Central Africa, the entire burden of effort was occasionally absorbed by an MDC (Foege et al. 1975).

5.2 Consequences of Mortality Reductions

In this section we can do no more than begin to sketch in the major influences of these mortality declines on populations of the less developed world, since these declines affect virtually every aspect of individual and collective life in a manner that undoubtedly varies with a host of initial conditions present in the population. It is probably wise to begin with the most concrete and least variable effects, the demographic ones.

Other things remaining the same, mortality declines increase the rate of population growth. The initial effect obviously is to increase the crude rate of natural increase by the absolute amount of the decline in the crude death rate. To a close approximation, the long-run effect of a permanent decline is to increase the rate of natural increase by the average (unweighted) decline in age-specific death rates between age zero and the mean age of childbearing (Preston 1974). This effect is almost fully realized within two generations. In neither case is the growth response strongly conditioned by the prevailing level of fertility. Not a shred of doubt remains that the vast majority of the acceleration in world population growth during the twentieth century is attributable to mortality decline rather than to a rise in fertility.

It is important to recognize that changes in rates of population growth typically have very different effects on demographic, economic, and social processes depending on their source. Coale and Hoover (1958) in their classic study were careful to point out that they were studying the economic implications of variation in fertility, but the study has often been misinterpreted as suggesting the deleterious effects of rapid population growth per se. Application of a modified Coale-Hoover model by Barlow (1967, 1968) demonstrated much more ambiguous economic effects when the source of growth acceleration was mortality decline. Loose discussions of relations between population and economic growth are usually aimed implicitly at the fertility component, even though it is mortality variation that has been the root of trends in population growth.

The fundamental reason why effects differ is a difference in the ages of persons affected. Changes in fertility initially affect only the number of zero-year-olds, and permanent changes permanently affect the age distribution of the population. Mortality changes typically affect all ages, and age distributional changes are relatively minor (Coale 1956; Stolnitz 1956). Such as they are, the short-run age distributional changes induced by mortality decline typically increase the proportion of the population at ages below 5 or 10 and above 40 and decrease the proportion at other ages. The pertinent index is the age-specific death rate, μ_x. When this declines by more than the population-weighted average, the proportion of the population in the immediately succeeding ages will rise. Since mortality declines have tended to be largest in absolute (but not proportional) terms at the extremes of life, the dependency burden initially rises. The long-run effect of a permanent decline in mortality is typically to increase the proportion of the population at ages below 20 and above 75. The pertinent age-specific index here is the cumulative change (unweighted) in age-specific death rates since age zero relative to an appropriately defined average (Preston 1974). The long-run effect on the dependency burden is also positive. With the gross reproduction rate fixed at 2.5, a rise in female life expectancy from 30 to 50 to 70 years in stable populations characterized by "West" mortality patterns increases the ratio of those outside of labor force age (15–64) to those within from 0.635 to 0.764 to 0.847 (Coale and Demeny 1966, pp. 82, 98, 114). All of this increase is sustained in the ages below 15. These changes are not trivial, but they are rather small relative to those induced by movements of fertility within its observed range.

It follows from this discussion and the formal analysis that supports it that if all ages experience an identical decline in death rates (usually termed a "neutral" decline), the age composition of the population will be unaffected in both the short run and the long run. The probability of survival from age x *to* age $x + n$ is equal to $\exp\{-\int_x^{x+n} \mu(t)\,dt\}$, where $\mu(t)$ is the death rate at exact age t. A decline in mortality by amount k at all ages will raise all n-year survival probabilities by the factor, $\exp\{kn\}$. Since the population at each age grows by the same factor (including infants via the greater survivorship of prospective parents), the proportionate age distribution is unaffected. Barring behavioral changes, a decline that is neutrally distributed among population subgroups, however defined (e.g., occupational or educational groups), will not affect population composition. The point is worth emphasis: it is *differential* mortality change that affects population composition. A change that is equally shared affects only size and growth. To the degree that typical mortality changes have been differentially distributed, the first-order changes induced in population composition have been economically unfavorable. Not only have the very young and the very old profited disproportionately, but so have women and unskilled or semiskilled workers.[18] Unlike programs of human resource development, which usually aim directly at an upgrading of population composition in ways that relate to production, programs of mortality decline have typically increased population size and reduced, at least initially, the desirability of its configuration.

Although mortality and fertility variation have very different effects on population composition, the mechanism by which they influence population size is the same in the long run: changes in the annual number of births. The principal long-run effect of mortality decline on population size arises not from the greater survivorship of persons who would have been born in any event, but from the larger number of births that are produced. To see this clearly, suppose that a neutral mortality change occurs to an initially stable population such that all ages experience a permanent reduction in death rates of .02. If age-specific fertility rates remain unchanged, the rate of population growth will increase by .02 and the rate of increase in the annual number of births will also rise by .02. Consider the number of 20-year-olds in the population 60 years after the mortality decline. The original number born into this cohort will be larger as a result of the mortality decline by $[e^{.02(40)} - 1] = 123\%$, whereas their improved survivorship after birth will have increased their numbers by $[e^{.02(20)} - 1] = 49\%$, a growth factor less than half as large. More than half the members of the cohort would not have been born had mortality not declined. This fraction continues to grow over time, but the improved survivorship factor does not. Stated more vividly, any LDC child "saved" from death today adds only one to the population size for a time. But the progeny of that child will ultimately be infinite in numbers if current rates of mortality and fertility are maintained by all generations. The prevailing practice in health economics of ignoring the offspring of the population "saved" seriously misrepresents the effect of health programs on populations (see reviews in Weisbrod 1975 and Klarman 1967).

5.2.1 Economic and Behavioral Responses

In discussing aggregate economic and behavioral responses to mortality decline, it is useful to recognize that all of the responses must make themselves felt through one of four indexes. This follows directly from a formal identity:

$$CDR = CBR + CRNM - R_p + R_{pc},$$

where CDR, CBR, $CRNM$ = crude rates of death, birth, and net migration

R_p, R_{pc} = proportionate rate of growth of total production and of production per capita.

When the crude death rate declines, one of the terms on the right side must change to keep the identity in balance. The first three terms on the right side—birthrates, migration rates, and the economic growth rate—primarily reflect behavioral adjustments to mortality change. The fourth—the growth rate of output per capita—is basically a default option, inevitably activated if none of the other three terms change. If none of the four terms on the right can change, or change for very long, the decline in CDR cannot be sustained. This is the basic Malthusian model, in which the "passion between the sexes" placed a floor on the crude birthrate, a subsistence level of production bounded R_{pc} from below, migration was defined as impossible, and slow technical change and rapidly diminishing returns to labor constrained R_p from above.

As an identity, any term in it could be isolated on one side of the equation and the others forced to "respond" to its changes. The justification for isolating the CDR is provided in the first part of the paper: a substantial fraction of changes in CDR have been induced by factors independent of any term on the right side, and it is reasonable to view them as being forced to respond to it. To the extent that declines in CDR have been produced by increases in rates of growth of production per head, the equation as presented is misleading.[19]

In most of the remainder of this section we will consider the various ways populations appear to have responded to mortality declines, taking each of the possibilities in turn. The review attempts to be positive and historical rather than normative.

Declines in Crude Birthrates

There are a multitude of ways that changes in mortality can induce changes in fertility. Three of the effects are quasi-biological. Declines in mortality change the age structure in such a way as to reduce the proportion of the population in the childbearing years and to reduce crude birthrates if age-specific fertility rates remain constant. Using the earlier example, gains in female life expectancy from 30 to 50 to 70 years, with age-specific fertility held constant at a level that produces a gross reproduction rate of 2.5, changes the crude birthrate from 38.78/1,000 to 37.12 to 35.99. The decline in the crude death rate over this range is from 33.34 to 5.78, so that the decline in the birthrate compensates for 10.1% of the decline in crude death rates through this age-structural route (Coale and Demeny 1966).

A second biological mechanism operates through breast-feeding. Breast-feeding inhibits ovulation, particularly in poorly nourished populations. Survival of the previous birth, by extending lactation, tends to delay the arrival of the next birth. Estimates of the average amount of net delay range as high as 12–13 months in Senegal, Bangladesh, and certain preindustrial European populations, although an estimate of about 7 months is probably more representative of poor agrarian populations (see the review in Preston 1975*b*). Since average interbirth intervals in such populations average about 30–35 months, the compensating variation in fertility to a change of infant mortality rates can be as high as 35% but is more likely to be in the area of 20%. That is, if all a woman's children die in infancy, her average interbirth intervals will be shorter by 20% and she will have approximately 20% more births over her reproductive life than if all had survived. In urban Latin American settings, where breast-feeding is usually short if it occurs at all, the compensating variation from this source is negligible (Rutstein and Medica 1975).

A third quasi-biological influence operates in the opposite direction but is probably fairly weak. Mortality declines make it more likely that marriages will survive through the end of the partners' reproductive periods. In areas such as India where sanctions against widow remarriage are strong, reductions in the incidence of widowhood probably exert an upward pressure on fertility. Arriaga (1967) has argued that this mechanism is responsible for substantial postwar increases in child-woman ratios in certain Latin American populations. However, his argument neglects the age-structural changes that are directly produced by mortality declines. The maximum effect of mortality changes on a woman's completed fertility can be estimated by assuming that no remarriage is possible and that childbearing occurs at a constant rate throughout her reproductive life up to the death of her husband or for 25 years, whichever comes first. Then her completed fertility is simply proportional to the expected number of years lived by the husband in the first 25 years of marriage. Assuming that males are age 20 at marriage, this expectation goes from 20.97 years in a population with a male life expectancy at birth of 30.08 to 24.49 when life expectancy is 68.56 190

(Coale and Demeny 1966, pp. 7, 23). Fertility would increase by 16.8%, or by perhaps 6 points, while the death rate would decline by perhaps 25 points. Complete prohibition of widow remarriage could thus boost the growth acceleration induced by mortality decline by some 25%, making it roughly equivalent to but opposite in sign from the lactational effect. However, it does not appear that taboos on widow remarriage are sufficiently widespread outside of India to have anywhere near this effect.

In addition to the three quasi-biological links, there are many possible behavioral ones. Since these are the subject of another conference paper (chap. 3), they will not be reviewed here (see also O'Hara 1975). Suffice it to say that the magnitude of one relationship has now been investigated rather carefully in a variety of populations in Asia and Latin America, as well as in preindustrial Europe. It has repeatedly been shown that, among women of a particular parity, those who had experienced one additional child death subsequently bore, on average, far fewer than one additional birth. Furthermore, some of the additional childbearing that did occur could be traced indirectly to the biological link identified above. The studies have attempted to control other characteristics of women believed to influence fertility (see studies by Chowdhury, Khan, and Chen; Heer and Wu; Rutstein and Medica; and Knodel in Committee for . . . Demography 1975). The largest "replacement" effect was 0.28 identified in Taiwan by Heer and Wu. It may be that the dead children were replaced in advance, but the simple demography of death makes this an inefficient reproductive strategy. The large majority of children who die before adulthood do so in the first two or three years of life, and their death can be observed and reacted to by parents during their own reproductive period.

The apparent failure of parents in many LDCs to behave as though they were pursuing a single reproductive target framed in terms of surviving children should not be surprising. It is clear that many of the social norms and sanctions that regulate reproductive behavior in LDCs refer to age at entry into union, frequency of intercourse, postnatal abstinence, number of partners, lactation, remarriage, and so on, rather than specifically to the number of children born or surviving (Polgar 1972). It is not necessary to reject the view that reproducers are goal-directed, but only to recognize that social norms and expectations have established other goals than "the" number of children. These norms and expectations are not responsive to an individual's experience with child death, although they may be responsive to aggregate mortality rates. Such conditioning is in fact the basis of functionalist theories of fertility. Davis (1955) has repeatedly argued that high-mortality populations must adopt a set of institutions and customs producing high fertility or else face extinction. The major adaptive institution he points to is the extended family, which encourages early entry into union by arranging marriages and stimulates fertility by removing many of the child-rearing costs from the parents. The expected positive effect of family extension on fertility has not been observed in most empirical studies at the household level (see the review in Burch and Gendell 1970). More to the point, it is not clear how the group-selection processes that are supposed to have created such institutions in the first place would operate to change them when mortality conditions relax.

There is one aggregate-level linkage between death rates and birthrates that deserves mention because of its apparent importance in preindustrial Europe. In a spatially limited system where land is the basis of wealth and accession to land the prerequisite for marriage, the rate of marriage and hence childbearing will depend upon the rate at which land becomes available, hence on mortality. Exogenous declines in mortality will, more or less automatically, reduce fertility by slowing the turnover of land and delaying marriage. Such a system was apparently an important feature of demographic-economic relations in Western Europe from 1600 to 1800 and may account for the generally late age at marriage in these populations (Habakkuk 1971; Wrigley 1969; see Eversley 1957 for a vivid numerical description of the expected responses to an epidemic). More important, economic historians have suggested that this mechanism, the "European marriage pattern," was a fundamental basis of the industrial revolution since it facilitated capital accumulation by severing the link between the level of mortality and the rate of growth of income per capita (Wrigley 1969; Habakkuk 1971). How important this mechanism is in contemporary less developed countries is unclear. Increased rates of population growth have been accompanied by declining proportions married throughout much of Asia since 1960 or so (Smith 1976). However, many other modernizing influences have also been at work. In Asia, at least, the prevalence of an extended family system in rural areas reduces the dependence of marriage on individual acquisition of land and presumably attenuates the link between population pressure and individuals' marital behavior. In urban areas, however, a "suitable" job may come to play an analogous, though no doubt less decisive, role.

That the sum of responses of fertility to mortality declines in LDCs has been quantitatively weak is shown clearly in the history of population growth rates. Durand (1967, p. 7) puts the annual rate of population growth for LDCs at 0.3% for 1850–1900 and 2.1% for 1950–65. He further suggests that, with few exceptions, birthrates have not declined. The United Nations Population Division "estimates and conjectures" that CDRs in LDCs were 38/1,000 in 1850–1900 and 17/1,000 in 1960–70, whereas CBRs were 40/1,000 and 41/1,000 in the two periods (United Nations 1971, p. 7). These figures suggest no compensating variation in fertility whatever, although mortality declines may have reduced the increase in fertility that would otherwise have occurred. It is true that several of the postulated relations should operate with a lag, and widespread (but small) declines in fertility since 1970 in LDCs might be partially attributed to prior mortality decline. But the evidence is that changes in birthrates have not been a major mechanism of adjustment to date.

Declines in Crude Rates of Net Migration

An increase in out-migration for the world as a whole is clearly impossible, and it has been scarcely more of an option for LDCs as a bloc. Perceived cultural, economic, and political difficulties attendant upon migration from LDCs to MDCs have resulted in MDC immigration quotas that are fixed at a point where they represent a tiny fraction of annual natural increase in LDCs. International migration among LDCs faces similar obstacles (see the discussion in Myrdal 1968, pp. 1459–62). In this matter the present LDC situation is again markedly at variance with that of European populations in the eighteenth and nineteenth centuries, when a substantial amount of population increase was drained off via overseas or transcontinental migration. Friedlander (1969) stresses the importance of this "safety valve" for delaying fertility reductions in England and Sweden. My calculations indicate that 15–25% of persons born in Sweden in the middle decades of the nineteenth century died outside its borders.

When subnational territories are considered, the migration response probably becomes more consequential, because the export of population growth from one area to another faces fewer legal, cultural, and institutional impediments. In Indonesia, the Philippines, and Ceylon, government programs have attempted to redistribute population from dense, rapidly growing areas to sparsely populated ones, although relative to natural increase the movements have been small (Myrdal 1968, pp. 2139–49). How important mortality declines in rural areas have been for rural-urban migration in LDCs has simply not been identified, to my knowledge. The region with apparently the sharpest mortality decline since 1900, Latin America, has also had the most rapid rural to urban migration (United Nations 1971). But both regional peculiarities may have been caused by its more rapid economic growth. Unlike the European situation, net rural-urban migration is not required to maintain constant proportions in the two sectors. Rural birthrates are higher, but so in general are rural death rates. Rates of natural increase are not widely disparate (Davis 1973). It has often been suggested that, in Asia at least, a large fraction of rural natural increase (which accounts for the large bulk of the annual volume) simply cannot be exported to urban areas because the cities cannot create enough new jobs. Even if these predictions prove pessimistic, it remains that, for national aggregates, changes in net migration rates have not been and will not be an important response to mortality decline.

Increases in Growth Rates of Total Production

The effect of population growth on economic growth is obviously a topic whose scope and complexity are too vast to be adequately reviewed here. While the effects of changes in death rates on output are difficult to partial out (see Denison 1962 for one attempt), economic history and economic theory are quite consistent with the view that R_p has been the principal respondent to changes in CDR.

A decline in CDR generally increases the growth rate of the labor force in roughly equal measure. As noted above, the labor force growth rate is typically incremented in the short run by a slightly smaller amount than the population growth rate. But even after accounting for this tendency, if we accept standard estimates of the elasticity of production with respect to changes in the quantity of labor input on the order of 0.5 to 0.8, more than half of the "response" can be accounted for without any change in other inputs or in technology.

There are reasons to believe that other inputs have typically changed in a reinforcing fashion. In some instances (South America, Southern Africa, and much of Asia) the growing population has had access to unutilized land not markedly inferior to that already under production (Myrdal 1968). According to D. Gale Johnson (1974, p. 89), between 1935–39 and 1960 approximately 75% of the increased grain output in developing countries resulted from expansion of the planted area. In some cases the mortality reduction campaigns themselves have liberated large areas of previously inhospitable territory. Taylor and Hall (1967) cite examples of such effects from antimalarial programs in Nepal, Sri Lanka, Sardinia, and Mexico (see also Sorkin 1975). Schultz (1964, pp. 63–70) shows that the response is also present when mortality changes in the opposite direction. An estimated 8% fall in the Indian labor force resulting from the 1918–19 influenza epidemic was accompanied by a reduction of 3.8% in acreage sown.

Expansion of land use obviously cannot continue indefinitely, and to the (apparently minor) degree that fixed factors are important in production, the particular effect of accelerated population growth on per capita production will be negative. For most purposes it is more important to know the effects of mortality decline on the capital supply. With one possibly important exception, the effects should not be radically different from those of increased fertility. The age distributional effects are similar in nature, though muted in the case of mortality change. When mortality declines, the proportion of the population in a stage of dissaving increases—permanently. Households face an increased depen- 192 dency burden and governments a greater press of immediate consumption demands. Business optimism regarding future demand for their products rises with the growth rate of potential consumers. Business profits and internally financed investments may increase as the labor supply curve shifts outward, unless the shift also reduces consumers' purchasing power. These effects have been reviewed elsewhere in conjunction with fertility effects, and it is not profitable to reconsider them here.

The one possibly important difference is that members of a lower-mortality population can look forward to longer lives in which to reap the benefits of personal investment. There is no such effect when fertility is the source of growth acceleration. The present value of investments with a long gestation period, such as extended schooling, retirement equities, or children, necessarily rises when mortality rates fall. Mushkin (1964), Schultz (1975), and others have suggested that such effects may represent an important economic benefit of reduced mortality.

It seems indisputable that such effects operate, and are probably reinforced by the increased proportion of the population who are in the investment stage. Nevertheless, it appears that the effects are relatively small and can be easily overwhelmed by minor variations in discount rates. To illustrate, define the present value of an investment in the standard manner prior to a change in mortality:

$$P.V. = \int_0^\infty p(a) \left[\frac{B(a) - C(a)}{e^{ra}}\right] da,$$

where $p(a)$ = the probability that the investor will survive a years from the time of investment according to the mortality schedule in effect prior to the mortality decline

$B(a), C(a)$ = benefits and costs of the investment realized or incurred in year a

r = subjective rate of discount, continuously applied.

Now superimpose a neutral mortality decline of 0.01 per year, equivalent to a reduction in the CDR of 10/1,000, that is, a large decline. This is the average reduction in age-specific death rates between ages 15 and 50 when male life expectancy at birth increases from 30.1 to 51.8 years (Coale and Demeny 1966, pp. 7, 16). The new present value is

$$P.V.' = \int_0^\infty p(a)\, e^{.01a} \left[\frac{B(a) - C(a)}{e^{ra}}\right] da$$

$$= \int_0^\infty p(a) \left[\frac{B(a) - C(a)}{e^{(r-.01)a}}\right] da.$$

That is, such a major change in mortality conditions has the same effect on present value as reducing the discount rate by only 0.01 and retaining the initial mortality conditions. In view of the wide variation that seems to prevail in discount rates, as partially reflected in the common analytic practice of applying several that differ by 0.05 or even more, it is readily seen that mortality prospects are a relatively minor influence on present value.

As a more concrete illustration, we will compute the internal rate of return on investment in schooling and show how it is affected by empirically observed variation in mortality. We have chosen Mexico for the illustration because Carnoy (1967) provides all of the necessary information except life tables.[20] We will compute the internal rate of return for a 15-year-old male who has completed grade 8 from his subsequent completion of grades 9, 10, and 11. Three mortality schedules are used: the male life table of Mexico, 1921, having a life expectancy at birth of 33.66 years (Arriaga 1968); the male life table of Mexico, 1966, having a life expectancy at birth of 59.49 years (Keyfitz and Flieger 1971); and immortality. Results are shown in table 5.6.

It is clear from this table that the increment to the private or social rate of return that results from replacing an e^{o}_{0} of 33.7 by immortality is only 1.5 points. This is approximately the difference it makes to use continuous rather than once-a-year compoundings. The reason for the weak effect is simply that mortality rates in young and middle adulthood are not high enough even in very high mortality populations to substantially alter expected payoffs. The large variability in rates experienced in childhood do not figure in the calculation, and the large variability at older ages is heavily discounted.

The effect of mortality variation would be somewhat stronger if individuals were predominantly risk-aversive and made decisions on the basis of the entire distribution of expected outcomes rather than simply on the basis of the mean. Higher mortality adds greater variability to the distribution of expected outcomes as well as reducing the mean. It increases the chance that zero or negative returns will accrue to investment, and this increase is faster than is the reduction in mean. Nevertheless, for the bulk of investors—those in early and middle adulthood

Table 5.6 **Internal Rate of Return from Completing Grades 9, 10, and 11 in Mexico, 1963, under Varying Assumptions about Mortality**

Assumption	Private Rate of Return	Social Rate of Return
Mexican life table of 1921	12.8	9.9
Mexican life table of 1966	13.9	10.9
Immortality	14.3	11.4

Sources: Carnoy (1967); Keyfitz and Flieger (1971); Arriaga (1968).

—it does not appear that even very radical mortality change could exert much influence on perceived investment profitability.

On balance, it does not appear that the effects of population growth acceleration on capital formation when mortality is the source should be markedly different *in nature* from the effects when fertility is the source. The Coale-Hoover model postulates a less-than-proportionate increase in capital supply when the population grows faster via higher fertility, and this feature is retained when the effects of mortality change are simulated (Barlow 1967, 1968). Largely for this reason, Barlow concludes that the antimalarial campaign in Ceylon ultimately had negative effects on per capita income growth, even after the improved health of the labor force is accounted for. Confidence in this conclusion depends on one's confidence in the savings-investment assumptions of the Coale-Hoover model.

It is also possible that mortality reductions foster economies of scale in production and an intensification of individual work effort. Again, the subject is very complex, and in most respects the analysis of mortality variation need not differ from that of fertility variation. A postulated difference is that mortality control programs have an important demonstration effect (Mushkin 1964; Fein 1964; Malenbaum 1970; and others). That is, they demonstrate that individuals can control their own destiny through the rational application of science and technology. They attest to the power of man in contrast to that of the supernatural and hence spark work effort and a stronger motivation toward self-fulfillment. Malenbaum (1970) argues that this is the basis of observed associations between mortality rates and labor productivity, but there are surely other mechanisms that offer more plausible interpretations. The contention has not been put to a rigorous test. It seems inconsistent with observations that poor illiterate farmers in LDCs have always responded quickly to new and profitable opportunities and technologies (Johnson 1974). A sense of personal control over the environment is one of the strongest components of "modern" attitudes (Sack 1974, p. 90). But the importance of mortality declines in the development of this attitude and the influence of the attitude on output remain to be demonstrated, important as the issue may be.

Changes in Growth Rates of Production per Capita

With constant technology and no changes in nonlabor factors of production, it is reasonable to expect that a CDR decline of 10/1,000 would raise the growth rate of total production only by 0.6–0.7% or so. If no birthrate or migration adjustments were forthcoming, the rate of growth of per capita income would decline by 0.3–0.4%. This is not a large amount relative to prevailing growth rates, and it is not surprising that the mortality declines have left no unmistakable imprint on per capita national economic growth rates.

Nevertheless, there are circumstances where the mortality decline has apparently had a decisive negative effect on economic well-being. Perhaps the most vivid account is offered by William Allan (1965, especially chap. 21). Allan outlines a cycle of land degeneration in East Africa that was, he argues, initiated by a mortality reduction. Population pressure at the kin-group level led to land subdivision and fragmentation, since landowners were expected to share their holdings with needy kin. Subdivision led in turn to a shortening of the fallow period and to soil depletion and erosion. Declining yields led to accelerated shortening of the fallow and to ultimate soil exhaustion. "Perhaps the greatest 'sin' of the suzerain powers was the saving of life, the lives of millions of men who under the old conditions would have died in early childhood, or in later life, of famine, disease, and violence" (Allan 1965, p. 338). Ultimately the response was out-migration, but not before a stage of economic misery was encountered. The initial stages of intensified land use are those described by Boserup (1965), but the outcome is very different. Instead of self-sustaining growth supported by a newly developed work ethic, the result was simply impoverishment. Obviously, some soils can support extreme intensification and multiple cropping and others cannot. Population pressure by itself is clearly not sufficient for sustained technological change, nor does it appear to be necessary once the possibilities of trade are opened up.

It is clear that general statements cannot be made about how populations have reacted or will react to exogenous mortality declines. The reaction will depend on a wide variety of initial conditions. In agrarian populations it appears that the most important conditioning factors are type of soil, land tenancy and kinship-marriage systems, density, possibilities for out-migration, savings and investment relationships, and the saliency of surviving-children goals. A great deal of work in recent economic and demographic history remains to be done before the quantitative details of the outline sketched in these sections can be confidently filled in.

194

Appendix

Table 5.A.1 Estimates of National Indexes about 1940

Country	Life Expectancy at Birth	Year of Estimate	Source	Daily Calories per Capita Available at Retail Level	Year of Estimate	Source: (6) Unless Noted
Australia	66.09	(40)	(2)	3,128	(40)	
Belgium	59.08	(38–40)	(3)	2,885	(40)	
Canada	64.62	(40–42)	(1)	3,109	(40)	
Czechoslovakia	56.79	(37)	(1)	2,761	(40)	
Chile	38.10	(40)	(4)	2,481	(40)	
Colombia	36.04	(38)	(4)	1,860	(38)	(7)
Denmark	66.31	(40)	(2)	3,249	(40)	
Egypt	38.60	(36–38)	(1)	2,199	(40)	
Finland	57.39[a]	(40)	(1)	2,950	(40)	
Greece	54.37	(40)	(1)	2,523	(40)	
Guatemala	30.40	(40)	(4)	—	—	
Honduras	37.50	(40)	(4)	2,079	(40)	
Hungary	56.57	(41)	(1)	2,815	(40)	
India	32.27	(41)	(5)	2,021	(40)	
Ireland	60.02	(40)	(1)	3,184	(40)	
Japan	49.12	(39–41)	(3)	2,268	(40)	
Korea	48.90	(38)	(1)	1,904	(40)	
Luxembourg	60.07	(38–42)	(3)	2,820	(34–38)	(8)
Mexico	38.80	(40)	(4)	1,909	(40)	
Netherlands	65.43	(40)	(2)	2,958	(40)	
New Zealand	67.00	(36)	(2)	3,281	(40)	
Nicaragua	34.50	(40)	(4)	—	—	
Panama	42.40	(40)	(4)	—	—	
Peru	36.50	(40)	(4)	2,090	(40)	
Philippines	46.26	(38)	(1)	2,021	(40)	
Portugal	51.06	(40)	(2)	2,461	(40)	
Puerto Rico	46.09	(39–41)	(1)	2,219	(40)	
Spain	50.18	(40)	(1)	2,788	(40)	
Sweden	66.64	(40)	(1)	3,052	(40)	
Switzerland	64.88	(40)	(2)	3,049	(40)	
Taiwan	47.80	(41)	(1)	2,153	(40)	
Thailand	40.02	(37–38)	(1)	2,173	(40)	
Turkey	33.91	(35–45)	(15)	2,619	(40)	
United Kingdom	61.64	(40)	(2)	3,005	(40)	
United States	63.74	(39–41)	(1)	3,249	(40)	
Venezuela	39.91	(41)	(1)	—	—	

National Income per Capita in 1970 U.S. Dollars	Year of Estimate	Source: (9) Unless Noted	Percentage Illiterate	Age Range	Year of Estimate	Source: (10) Unless Noted	Population in Thousands (1940)(13)
1,128	(38–40)		2.5	—	(40)	(14)	7,039
715	(38–39)		4.5	—	(40)	(14)	8,301
1,041	(38–40)		3.16	10+	(40)		11,682
438	(38)		3.6	—	(37)	(14)	14,429
371	(38–40)		28.2	10+	(40)		5,063
190	(38–40)		44.10	10+	(38)		8,702
971	(38–40)		1.5	—	(40)	(14)	3,832
167	(39)		84.9	10+	(37)		16,008
419	(39)		8.8	—	(40)	(14)	3,698
187	(38)		34.4	—	(40)	(14)	7,319
78	(40)		65.4	7+	(40)		2,201
109	(41–42)		66.35	10+	(45)		1,146
318	(38–40)		6.0	10+	(41)		9,344
67	(38–39)		86.5	10+	(41)	(10,11)[b]	316,004
665	(40)		1.5	—	(40)	(14)	2,993
260	(39)		—	—	—		71,400
—	—		68.6	10+	(30)		21,817
795	(39)		4.4	—	(40)	(14)	292
138	(40)		51.5	10+	(40)		19,815
889	(38)		1.5	—	(40)	(14)	8,879
1,055	(38–40)		1.5	—	(40)	(14)	1,573
105	(40)		63.0	7+	(40)	(12)	825
374	(40)		35.25	10+	(40)		620
89	(40)	(16)	56.35	10+	(40)		7,033
113	(38)		37.75	10+	(48)		15,814
—	—		48.7	10+	(40)		7,696
413	(38–40)		31.5	10+	(40)		1,880
361	(40)	(16)	23.2	10+	(40)		25,757
1,091	(40)		0.1	10+	(30)		6,356
1,246	(40)		1.5	—	(40)	(14)	4,234
—	—		78.7	5+	(40)		6,163
128	(39)	(11)	46.25	10+	(47)		14,755
212	(39)		79.1	10+	(35)		17,620
1,334	(38–40)		1.5	—	(40)	(14)	41,862
1,549	(40)		4.2	10+	(40)	(13)	132,594
291	(40)		56.50	10+	(41)		3,803

195

Table 5.A.1 (continued)

Sources:
1. United Nations, *Statistical Yearbook* (1967).
2. Preston, Keyfitz, and Schoen (1972).
3. Keyfitz and Fleiger (1968).
4. Arriaga (1968).
5. Estimated by Davis (1951, pp. 62–63).
6. Estimates prepared by the Food and Agriculture Organization of the United Nations, cited in "Food, Income, and Mortality," *Population Index* 13, no. 2 (April 1947): 96–103.
7. United Nations, *Statistical Yearbook* (1951).
8. United Nations, Food and Agriculture Organization, *Production Yearbook* (1958).
9. Estimates prepared by the Technical Group, U.S. Bureau of the Budget; cited in "Food, Income, and Mortality," *Population Index* 13, no. 2 (April 1947): 96–103. All figures have been converted to 1970 U.S. dollars by application of the consumer price index from U.S. Bureau of the Census, *Statistical Abstract of the United States*, various issues.
10. United Nations, *Statistical Yearbook* (1949–50).
11. United Nations, *Statistical Yearbook* (1955).
12. United Nations Educational, Scientific and Cultural Organization, *Basic facts and figures: Illiteracy, libraries, museums, books, newspapers, newsprint, film and radios* (Paris, 1952).
13. United Nations, *Demographic Yearbook* (1960).
14. Banks (1971).
15. Calculated by author from estimates of e_5^0 presented in Shorter (1968). The Coale-Demeny "South" model mortality pattern was assumed to apply.
16. United Nations, *National Income Statistics of Various Countries, 1938–1948* (Lake Success, N.Y., 1950).

[a]Average 1936–40 and 1941–50.
[b]Interpolated from data for 1931 and 1951.

Table 5.A.2 National Indexes about 1970

	Life Expectancy at Birth 1970–75 (1)	Percentage Illiterate of the Adult Population (2)	Year of Esti-mate (2)	1970 National Income per Capita (1970 U.S. $) (3)	1970 Daily Calories Available for Consumption, per Capita (4)	1970 Population (in 1,000s) (1)	Index of Income Inequality (5)	Year of Esti-mate	Coverage
Africa									
Algeria	53.2	52.5	1971	295	1,710	14,330	—		
Angola	38.5	87.5	1973*	280	1,910	5,670	—		
Botswana	43.5	87.0	1971	132	2,040	617	—		
Burundi	39.0	35.0	1974*	68	2,330	3,350	—		
Central African Rep.	41.0	82.0	1975*	122	2,170	3,350	—		
	38.5	90.0	1975*	70	2,060	1,612	—		
go	43.5	65.0	1970*	281	2,160	3,640	—		
ahomey	41.0	80.0	1975*	81	2,250	1,191	—		
Egypt	52.4	62.0	1975*	202	2,360	2,686	—		
Ethiopia	38.0	93.0	1975*	72	2,150	33,329	−6.5660	1965	Nat. Household
Gabon	41.0	88.0	1974*	468	2,210	24,855	—		
Gambia	40.0	90.0	1971*	99	2,370	500	−17.2442	1960	Nat. Population
Ghana	43.5	56.5	1971	236	2,200	463	—		
Guinea	41.0	92.5	1971*	79	2,040	8,628	—		
Ivory Coast	43.5	80.0	1973*	325	2,490	3,291	—		
Kenya	50.0	75.0	1975*	130	2,350	4,310	−9.8088	1970	Nat. Income Recip.
Liberia	43.5	88.0	1970	181	2,040	11,247	−13.1644	1969	Nat. Income Recip.
Libyan Arab Rep.	52.9	68.0	1974*	1,450	2,540	1,523	—		
Madagascar	43.5	60.0	1975*	123	2,350	1,938	—		
Malawi	41.0	75.0	1976*	68	2,150	6,932	−9.8964	1960	Nat. Population
Mali	38.0	90.0	1972	50	2,170	4,360	−7.0502	1969	Nat. Household
						5,047	—		

Table 5.A.2 (continued)

	Life Expectancy at Birth 1970–75 (1)	Percentage Illiterate of the Adult Population (2)	Year of Esti-mate (2)	1970 National Income per Capita (1970 U.S. $) (3)	1970 Daily Calories Available for Consumption, per Capita (4)	1970 Population (in 1,000s) (1)	Index of Income Inequality (5)	Year of Esti-mate	Coverage
Mauritius	65.5	20.0	1974*	223	2,370	824	—		
Mauritania	38.5	95.0	1972*	147	2,060	1,162	—		
Morocco	52.9	78.6	1971	225	2,400	15,126	—		
Mozambique	43.5	80.0	1972	228	2,190	8,234	—		
Niger	38.5	94.0	1973*	70	2,180	4,016	—		
Nigeria	41.0	74.0	1973	135	2,290	55,073	—		
Rhodesia	51.5	95.0	1972	257	2,550	5,308	−14.8362	1968	Nat. Income Recip.
Rwanda	41.0	90.0	1973*	57	2,160	3,679	—		
Senegal	40.0	95.0	1971*	201	2,300	3,925	−12.1394	1960	Nat. Population
Sierra Leone	43.5	90.0	1974*	150	2,240	2,644	—		
Somalia	41.0	95.0	1974*	85	1,770	2,789	—		
South Africa	51.5	56.0	1974*	680	2,730	21,500	−14.3543	1965	Nat. Population
Sudan	48.6	80.0	1973	109	2,130	15,695			
Togo	41.0	90.0	1976*	125	2,160	1,960			
Tunisia	54.1	60.0	1972	257	2,060	5,137	−9.7467	1961	Nat. Population
Uganda	50.0	70.0	1976*	127	2,230	9,806			
United Rep. of the Cameroon	41.0	35.0	1976*	183	2,230	5,836	—		
United Rep. of Tanzania	44.5	71.0	1967	94	1,700	13,273	−8.1535	1967	Nat. Household
Upper Volta	38.0	90.0	1972	62	1,940	5,384	—		
Zaire	43.5	65.0	1971	118	2,040	21,638	—		
Zambia	44.5	52.7	1969	345	2,040	4,295	—		

Table 5.A.2 (continued)

	Life Expectancy at Birth 1970–75 (1)	Percentage Illiterate of the Adult Population (2)	Year of Esti-mate (2)	1970 National Income per Capita (1970 U.S. $) (3)	1970 Daily Calories Available for Consumption, per Capita (4)	1970 Population (in 1,000s) (1)	Index of Income Inequality (5)	Year of Esti-mate	Coverage
Asia									
Afghanistan	40.3	92.5	1973	83	1,950	16,978	—		
Bangladesh	35.8	90.0	1973	111	1,860	67,692	−3.5534	1967	Nat. Household
Burma	50.0	30.0	1974*	73	2,230	27,748	—		
Cyprus	71.4	18.0	1973	688	2,460	633	−3.3164	1966	Urban Household
India	49.5	40.0	1971	93	2,060	543,132	−7.6103	1968	Nat. Household
Indonesia	47.5	40.0	1971	98	1,920	119,467	—		
Iran	51.0	65.5	1974	352	2,080	28,359	−8.2776	1968	Urban Household
Iraq	52.7	70.0	1974*	311	2,250	9,356	—		
Israel	71.0	12.8	1974*	1,654	2,970	2,958	−3.1183	1970	Urban Household
Japan	73.3	2.0	1975*	1,636	2,310	104,331	−3.3217	1963	Nat. Household
Jordan	53.2	62.5	1972*	260	2,470	2,280	—		
Khmer Rep.	45.4	15.0	1973	123	2,410	7,060	—		
Korea, Rep. of	60.6	8.5	1970	252	2,420	30,721	−2.2963	1971	Nat. Household
Laos	40.4	75.0	1970*	71	2,080	2,962	—		
Lebanon	63.2	14.0	1975*	521	2,380	2,469	—		
Malaysia	59.4	24.0	1970	295	2,400	10,466	−10.0406	1970	Nat. Household
Nepal	43.6	86.0	1971	80	2,050	11,232	—		
Pakistan	49.8	83.0	1973	164	2,280	60,449	—		
Philippines	58.4	16.5	1970	225	1,920	37,604	−8.6939	1971	Nat. Household
Saudi Arabia	45.3	75.0	1973*	495	1,920	7,740	—		
Singapore	69.5	24.5	1974	918	2,080	2,075	—		

Table 5.A.2 (continued)

	Life Expectancy at Birth 1970–75 (1)	Percentage Illiterate of the Adult Population (2)	Year of Estimate (2)	1970 National Income per Capita (1970 U.S. $) (3)	1970 Daily Calories Available for Consumption, per Capita (4)	1970 Population (in 1,000s) (1)	Index of Income Inequality (5)	Year of Estimate	Coverage
Asia (continued)									
Sri Lanka	67.8	22.0	1971	160	2,240	12,514	−4.4313	1970	Nat. Household
Syria	54.0	60.0	1970	258	2,530	6,247	—		
Taiwan	69.4(7)	5.0(6)	1965	295(7)	2,662(7)	14,334	−3.6354	1964	Nat. Household
Thailand	58.0	18.0	1970	167	2,330	35,745	−8.3579	1962	Nat. Household
Turkey	56.9	44.0	1970	348	2,770	35,232	−11.6389	1968	Nat. Household
Vietnam, Rep. of	40.5	23.0	1971	232	2,340	17,952	−3.8293	1964	Rural Household
Yemen	44.8	82.5	1975*	77	1,970	5,767	—		
Yemen, P.D.R.	44.8	90.0	1970*	96	2,020	1,436	—		
Latin America									
Argentina	68.2	8.0	1973*	1,065	3,150	23,748	−6.2544	1961	Nat. Household
Bolivia	46.8	58.5	1973	191	1,840	4,780	—		
Brazil	61.4	33.0	1970	376	2,600	95,204	−17.7304	1970	Nat. Household
Chile	62.6	12.0	1970	618	2,460	9,369	−8.4001	1968	Nat. Household
Colombia	60.9	21.5	1973	358	2,250	22,075	−10.6656	1970	Nat. Economic Active Population
Costa Rica	68.2	10.0	1973	522	2,470	1,737	−6.4897	1971	Nat. Household
Dominican Rep.	57.8	32.0	1970	334	2,060	4,343	—		
Ecuador	59.6	30.0	1970	255	2,040	6,031	−17.3614	1970	Nat. Economic Active Population
El Salvador	57.8	40.0	1971	283	1,890	3,516	−8.1625	1969	Nat. Population
Guatemala	52.9	62.0	1974*	343	2,120	5,298	−2.8074	1966	Rural Household

Table 5.A.2 (continued)

	Life Expectancy at Birth 1970–75 (1)	Percentage Illiterate of the Adult Population (2)	Year of Estimate (2)	1970 National Income per Capita (1970 U.S. $) (3)	1970 Daily Calories Available for Consumption, per Capita (4)	1970 Population (in 1,000s) (1)	Index of Income Inequality (5)	Year of Estimate	Coverage
Latin America (continued)									
Guyana	67.9	14.0	1974*	319	2,080	709	—		
Haiti	50.0	90.0	1974*	100	1,720	4,325	—		
Honduras	53.5	53.0	1974*	259	2,180	2,553	−14.9221	1968	Nat. Household
Jamaica	69.5	18.0	1970	600	2,300	1,882	—		
Mexico	63.2	24.0	1970	655	2,560	50,313	—		
Nicaragua	52.9	47.0	1971	423	2,380	1,970	—		
Panama	66.5	21.5	1973	646	2,520	1,458	−11.2564	1969	Nat. Economic Active Population
Paraguay	61.9	38.0	1973	230	2,800	2,301	—		
Peru	55.7	44.3	1970	293	2,310	13,248	−14.7827	1971	Nat. Economic Active Population
Trinidad and Tobago	69.5	30.4	1970	732	2,360	955	—		
Uruguay	69.8	9.0	1975*	799	2,860	2,955	−6.8128	1967	Nat. Household
Venezuela	64.7	35.2	1971	954	2,460	10,559	−10.5676	1962	Nat. Household
North America									
Canada	72.4	1.0	1975*	3,369	3,190	21,406	−4.0518	1965	Nat. Household
United States	71.3	1.0	1969	4,289	3,270	204,879	−6.6384	1966	Nat. Household
Puerto Rico	72.1	27.9	1970	1,744	2,450	2,743	—		

Table 5.A.2 (continued)

	Life Expectancy at Birth 1970–75 (1)	Percentage Illiterate of the Adult Population (2)	Year of Esti-mate (2)	1970 National Income per Capita (1970 U.S. $) (3)	1970 Daily Calories Available for Consumption, per Capita (4)	1970 Population (in 1,000s) (1)	Index of Income Inequality (5)	Year of Esti-mate	Coverage
Europe									
Austria	71.2	1.0	1974*	1,730	3,340	7,447			
Belgium	72.9	2.0	1975*	2,421	3,390	9,638			
Bulgaria	71.8	5.0	1975*	2,726	3,300	8,490	−1.5225	1962	Nat. Workers
Czechoslovakia	69.3	0.0	1974*	3,013	3,190	14,339	−1.1290	1964	Nat. Workers
Denmark	73.9	1.0	1974*	2,898	3,230	4,929	−4.9775	1966	Nat. Income Recip.
Finland	70.4	0.0	1975*	1,998	3,020	4,606	−9.8462	1962	Nat. Income Recip.
France	72.6	3.0	1975*	2,550	3,210	50,670	−11.4365	1962	Nat. Household
Germany, W.	70.6	1.0	1970*	2,752	3,230	60,700	−7.5125	1964	Nat. Income Recip.
Greece	71.8	15.6	1971	1,051	2,900	8,793	—		
Hungary	69.5	2.0	1975*	2,244	3,180	10,338	−2.1548	1969	Nat. Population
Ireland	71.8	1.0	1974*	1,244	3,420	2,954	—		
Italy	72.0	7.0	1975*	1,591	3,170	53,565	—		
Luxembourg	70.8	2.0	1975*	2,613	3,390	339	—		
Malta	70.8	12.0	1974*	721	2,680	326	—		
Netherlands	73.8	2.0	1973*	2,232	3,290	13,032	—		
Norway	74.5	0.0	1974*	2,458	2,920	3,877	—		
Poland	70.1	2.2	1970	5,766	3,270	32,473	−2.2046	1964	Nat. Workers
Portugal	68.0	45.0	1970*	684	2,890	8,628	—		
Spain	72.1	19.9	1970	884	2,620	33,779	−5.1072	1965	Nat. Household
Sweden	73.3	0.1	1975*	3,724	2,800	8,043	−6.1269	1963	Nat. Income Recip.
Switzerland	72.4	0.0	1973*	2,963	3,250	6,267	—		
United Kingdom	72.3	10.0	1975*	1,990	3,140	55,480	−4.0915	1968	Nat. Household

Table 5.A.2 (continued)

	Life Expectancy at Birth 1970–75 (1)	Percentage Illiterate of the Adult Population (2)	Year of Esti-mate (2)	1970 National Income per Capita (1970 U.S. $) (3)	1970 Daily Calories Available for Consumption, per Capita (4)	1970 Population (in 1,000s) (1)	Index of Income Inequality (5)	Year of Esti-mate	Coverage
Oceania									
Australia	72.4	1.5	1975*	2,633	3,050	12,552	−3.6895	1968	Nat. Household
New Zealand	72.0	2.0	1975*	2,008	3,330	2,820	—		

Sources: (1) United Nations, *Selected World Demographic Indicators by Countries 1950–2000* (Population Division, Department of Economic and Social Affairs of the United Nations, 1975).
(2) Unstarred: United Nations Educational, Scientific, and Cultural Organization, *Statistical Yearbook, 1973* (Paris, 1974), table 1.4; starred: United States, State Department, *Background Notes*, individual country volumes, various years 1970–76.
(3) United Nations, *Statistical Yearbook, 1974*, tables 181, 188.
(4) United Nations, Food and Agriculture Organization, *The State of Food and Agriculture 1974: World Review* (Rome, 1975).
(5) Jain, Shail, "Size Distribution of Income" (International Bank for Reconstruction and Development, Bank Staff Working Paper no. 190, November, 1974).
(6) Kenneth Clark et al., *Area Handbook for the Republic of China* (Washington, D.C.: Department of the Army, 1969), p. viii.
(7) Taiwan, Council for International Economic Cooperation and Development, *Taiwan Statistical Data Book* (1972). All figures refer to 1970.

Table 5.A.3 National Indexes Used in Analysis of Mortality Change, 1940–70

	Malaria Ende-micity in 1943[a] (1)	Annual Average International Economic Aid Received 1954–72[b c] (2)	External Assistance Received for Community Water Supply and Sewage Disposal Projects 1966–70[b] (3)
Australia	1	–0	0
Belgium	0	–0	0
Canada	0	–0	0
Chile	1	13.29	.14
Colombia	3	5.13	2.30
Czechoslovakia	0	–0	0
Denmark	0	–0	0
Egypt	2	7.61	.18
Finland	0	–0	0
Greece	2	19.55(4)	0
Guatemala	3	3.35[d]	3.45
Honduras	3	4.05[d]	.90
Hungary	0	–0	0
India	3	1.71	0
Ireland	0	–0	0
Japan	1	–0	0
Korea (South)	1	10.15	.52
Luxembourg	0	–0	0
Mexico	3	1.50	.34
Netherlands	0	–0	0
New Zealand	0	–0	0
Nicaragua	3	5.78[d]	3.40
Panama	3	10.79	19.48
Peru	2	4.13	3.55
Philippines	3	2.02	.58
Portugal	1	–0	0
Puerto Rico	2	–0	0
Spain	1	9.22(4)	0
Sweden	0	–0	0
Switzerland	0	–0	0
Taiwan	3	7.09	0
Thailand	3	1.63	.09
Turkey	3	–0	0
United Kingdom	0	–0	0
United States	0	–0	0
Venezuela	3	2.82	3.71

Sources: (1) Shattuk (1951, p. 4; 1943 map prepared by U.S. Army Medical Intelligence Branch); Faust (1941); Boyd (1949).
(2) United Nations, *Statistical Yearbook* (1958–74).
(3) World Health Organization, *World Health Statistics Report* (1973), vol. 26, no. 11.
(4) Organization for Economic Cooperation and Development, *Development Cooperation: Efforts and Policies of the Members of the Development Assistance Committee, 1973 Review* (Paris, 1973).

Notes

1. The zero-order correlation between e^o_0 and the natural log of national income per capita for the 120 countries in appendix table 5.A.2 is 0.859; the correlation between e^o_0 and national income itself is only 0.693.

2. Causes eliminated include influenza/pneumonia/bronchitis, diarrheal disease, and maternal mortality, and also a proportion of "other and unknown causes" equal by age to the proportion of known causes assigned to infectious diseases at that age.

3. The age distribution used for direct standardization is that of a female "West" stable population with $e^o_0 = 65$ and $r = .01$ (Coale and Demeny 1966).

4. WHO in conjunction with the United Nations Population Division has estimated that life expectancy for LDCs as a whole was 49.6 in 1965–70 (World Health Organization 1974*b*, p. 23). Life expectancy in 1900 is the author's guess based upon life tables calculated by Arriaga for Latin America and on life tables for India, Taiwan, and Japan around the turn of the century. Corresponding ASCDRs were computed by the author based on relationships between the two mortality measures in the set of 165 populations employed by Preston and Nelson (1974).

5. Pairwise deletion was employed for missing data in the 1940 regressions. That is, results are based upon correlation matrices computed exclusively on the basis of data that were available. *N* is taken as 28, the number of cases for which observations were complete. One observation was missing on literacy, three on national income, and four on calorie consumption. For no country was more than one piece of information missing. The data on LDCs were considered too valuable to sacrifice all information because one piece was missing. Standard errors are shown in parentheses.

6. Several alternative specifications of these equations were employed, but results were not appreciably altered. These included polynomial representations of income and calories and measurement of *CAL* from different base points. The R^2's for equations including first- and second-degree terms for *Y* and *CAL* were lower than those for the specification presented in the text despite the addition of two variables. The use of other base points for *CAL* left R^2 unaffected to four decimals. Population-weighted regressions were rejected on the grounds that they gave too much weight to India, where measurement of variables was believed to be unusually poor, especially in 1940.

7. There are of course many ways to attribute differences to changes in values of variables and to changes in coefficients, none of them clearly preferable. In this instance we are constrained by the unavailability of data for most LDCs in 1940. Substitution of 1970 values of the regressors into the 1970 equation will not, of

---

[a]0 = virtually none; 1 = low; 2 = moderate; 3 = high.
[b]Current U.S. dollars per capita.
[c]Includes net official flow of external resources to individual countries from developed market economies and multilateral agencies and bilateral commitments of capital by centrally planned economies. "Military expenditures and contributions are excluded as far as possible."
[d]Information not available for 1966 through 1968.
[e]Information for 1970–72 from source 4.

200

course, usually yield the correct 1970 life expectancy for a country. But the predictions must be very nearly correct in the aggregate, since the regression plane must pass through the mean of the variables and since 94 of the 120 observations are LDCs.

8. This figure is close to the 9.7 estimate derived by Preston (1975*a*, p. 238), who considered only income and used cruder, regional income distributions of 1963 to evaluate changes between 1938 and 1963.

9. Once again, pairwise deletion is employed for cases of missing data.

10. The mean values of *AID* and *WAT* for the 17 LDCs were 4.77 and 2.21, respectively.

11. $16.46 = 31.47 - .0211(2.59) - .4063(39.39) + .0750(4.77) + .2939(2.21)$. Of the 16.46 year gain, 1.01 years is attributed to external aid.

12. A large majority of countries in the WHO survey listed lack of financing as the principal barrier to expanded water supply and sewerage systems (World Health Organization 1973).

13. For an account of the impact of the Catholic diocese of Oklahoma on mortality in a Maya village, see Early 1970. The experience was probably repeated hundreds of times.

14. For ample documentation, see World Bank (1975) and Bryant (1969).

15. Compiled from United Nations, *Statistical Yearbook*, 1974, table 197. Figures refer to thirty-nine LDCs.

16. WHO has compiled estimates of private consumption expenditures on health care in certain LDCs (World Health Organization 1970). The percentage of total private consumption that is spent on health can be compared with the proportion of GNP represented by government health expenditure, a procedure that reduces incomparabilities resulting from differences in the years of estimate. The percentages are the following (private consumption appearing first): Sierre Leone, 2.9 and 0.9; Jordan, 0.6 and 2.8; Thailand, 3.6 and 1.2; the Philippines, 1.7 and 0.5; Malaysia, 2.4 and 2.5; Panama, 4.0 and 2.2; Jamaica, 1.1 and 2.7. Government health expenditures are from World Bank (1975, annex 3).

17. WHO, for example, does not absorb any materials costs but views its role exclusively as providing technical, advisory, and educational assistance (Goodman 1971, pp. 203–4). Dispute over the provisions of assistance for material led to the "resignation" from WHO of Soviet-bloc countries between 1949 and 1955.

18. The class distribution of gains in mortality is not well documented. There was a sharp contraction of age-standardized mortality ratios for various classes in England during the twentieth century (Antonovsky 1967, p. 63). Even constant ratios would entail reductions in absolute differences, which is the pertinent index when population composition is considered.

19. This is a less serious problem than it might at first appear. The coefficients of $\ln Y$ presented earlier range from 3.6 to 7.0 (the latter observed when no other term except income distribution is present in the cross-sectional analysis). A 1-point increase in the rate of per capita economic growth would thus be associated with at most a gain in life expectancy of 0.07 years. A gain in e^o_0 of 2 years is roughly associated with a drop in CDR of 0.0015. Thus, a 0.01 gain in the rate of economic growth would be expected to reduce the CDR by at most 0.07 $(.0015/2) = 0.000053$, or by 0.53% of the change in R_{pc}.

20. Earnings differentials were obtained from a 1963 survey of 4,000 urban wage earners. Adjusted earnings differentials presented in Carnoy's (1967) table 3 are employed. Retirement was assumed to occur at age 65. The continuously compounded rate of return was calculated using the formula presented in the text, with

P.V. set equal to zero. For calculation of social costs, the average annual public expenditure per student by grade is added to direct personal outlays and income forgone.

Comment J. D. Durand

I will attempt to fit some of Preston's findings with regard to the determinants of mortality, and related results in Ronald Lee's paper (chap. 9), into a sketch of salient features of the evolution of determinants of the overall levels and trends of mortality during recent centuries and decades.

Results of recent work in this field at the University of Pennsylvania suggest that the typical form of the trend of expectation of life in countries making the transition from premodern to modern regimes of mortality may be represented by an essentially logistic curve, which can be divided into fairly distinct phases as follows:

0—pretransitional phase, in which expectation of life fluctuates around a nearly constant long-term level;

1—initial phase of transition, in which expectation of life increases irregularly at a relatively slow long-term average rate;

2—"takeoff" phase, in which expectation of life rises at a steady, rapid rate;

3—final phase, in which expectation of life rises slowly and appears to be approaching a ceiling at a high level.

I will attempt to draw a tentative general sketch of principal causal factors that have contributed to the increases in expectation of life (e) during each of these phases of the transition, in terms of the following simplified formula: $e = f(y, k, s, n, \ldots)$ where y stands for income per head, k for knowledge of the causes of disease and death and methods of prevention and treatment (including what the layman knows as well as what the physician knows), s for social action in the broad field of health protection, and n for natural factors.

This is not a comprehensive formulation of the determinants of mortality. Preston shows that the distribution of income and nutrition are influential; among other factors that may have played significant roles in the gains of life expectancy during modern times are the advance of popular education and the decline of fertility; no doubt urbanization has influenced the trends in various countries, and other factors could be 28

J. D. Durand is professor of economics and sociology and a research associate of the Population Studies Center, University of Pennsylvania.

mentioned. However, probably the greater part of the increases in expectation of life in most parts of the world during the last two centuries can be attributed to the growth of *y*, the advances of *k* and *s*, and some favorable changes in *n*. I do not presume that the effects of these factors have been simply additive. It is not to belittle the value of Preston's regression models to postulate, for example, that the effects of given changes in each and all of the factors have varied with the levels of *e*, as is shown vividly by the smallness of recent gains of life expectancy in countries where the highest levels have been achieved.

Pretransitional Regimes

Lee's study of preindustrial England provides a most valuable example of conditions and factors of mortality in a pretransitional state. In this case, we may disregard factors *k* and *s*, assuming that they were constant in effect, at least up to the eighteenth century. From the thirteenth to the eighteenth century, there were important variations of mortality in England, both long-term and short-term, and Lee finds that these were due mainly to noneconomic factors; that is, presumably natural factors in the main. The identification of these natural factors remains an unsolved puzzle. Lee says, "they may have been climatic, or the by-product of independent epidemiological changes, or the result of voyages of exploration."[1] As regards climate, Le Roy Ladurie's work is rather discouraging to hopes of finding in its variations a satisfactory explanation for the long swings in mortality, but the question of its influence on the hazards of disease as well as on agriculture has by no means yet been disposed of. There were also important long- and short-term ups and downs of *y*, represented in Lee's analysis by indicators of wages and prices, and he finds that their influence on mortality was not negligible, although it was less potent than the influence of *n*. One of the most interesting features is the low ceiling over *e*. Apparently even the wealth of kings and dukes would not purchase more than about 25 to 35 years of life expectancy, depending on the conditions of *n*. The ceiling is much higher today, but it is still firm. Unlimited growth of national income per head seems unlikely under present conditions to bring expectation of life for the two sexes much above 75 years.

England's economic situation during the centuries shortly before the industrial revolution was relatively favorable compared with that of most other countries, as Lee points out. Both *e* and *y* were probably considerably lower in most of the rest of the world, and it is likely that the influence of changes of *y* over time may have been stronger elsewhere than it was in England. Lee suggests this with reference to Goubert's observations on mortality in Beauvais during the seventeenth and eighteenth centuries. However, I would hypothesize that *n* was a major factor in both temporal variations and international differences in mortality under the pretransitional regimes throughout the world.

Preston (1975) gives a chart of the changing relation between e and y in international cross sections about 1900, 1930, and 1960. If data were available to draw such a chart with reference to conditions around 1750 and earlier dates, I presume that the correlation between e and y would be seen to have been weaker, the curve representing the relation between the two variables would exhibit a less steep positive slope, and it would shift erratically up and down from one date to another under the influence of changing natural factors. A major feature of the transition to modern regimes of mortality has been progressive neutralization of the influences of n as a result of the growth of y and advances of k and s.

First Phase of the Transition

Although it is not easy to define the date of beginning of the mortality transition in any country, the secular trend of slowly rising expectation of life identified with the first phase of the transition was clearly general in Western Europe during the nineteenth century, and indications of accelerating rates of population growth suggest that it was also widespread in Eastern Europe, North Africa, and Latin America. The trend of mortality in the United States before the closing decades of the nineteenth century remains an unresolved question.

In the countries in the vanguard of industrial development during the nineteenth century, the growth of income per head was undoubtedly a major factor contributing to the decline of mortality in this first phase, but the effect of increasing y was reinforced by advances in k and s. Under the heading of k, in addition to the important innovation of smallpox vaccination, I surmise that increasing understanding of the importance of hygiene and proper feeding of children, linked with the advance of popular education in the industrializing countries, played an influential part. Under s were such social actions as protection of water supplies, urban sewerage, swamp drainage, quarantine practices, restriction of child labor, and regulation of conditions of women's employment. Such health-protective social actions were not confined to the wealthiest countries; Sanchez-Albornoz (1974) traces their development in Latin American cities during the nineteenth century.

With regard to the historical antecedents of Preston's (1975) chart of changing relations between e and y during the twentieth century, I 202
postulate, although I cannot provide statistical proof, that the developments related to the first phase of the mortality transition during the nineteenth century had the following effects: making the correlation between e and y stronger than it had been under the pretransitional

regimes; gradually shifting upward the curve of e values corresponding to given levels of y; and making the slope of the curve steeper—that is, widening the differences in e between richer and poorer nations.

Second Phase of the Transition

The decided quickening of the rate of gain in life expectancy that marks the beginning of the second phase of the transition took place in the 1890s or about a decade earlier or later in the more developed countries of Europe and America. It seems clear that this turn of the trend was primarily a result of the first revolution in death-control technology produced by the validation and wide acceptance of the germ theory of disease. The effect of this was not limited to the new techniques of immunization and therapy for particular diseases that began to be invented late in the nineteenth century. Meanwhile, increasing income and health-protective social actions continued to contribute to gains in e, and it seems a reasonable hypothesis that the tightening control of fertility may also have contributed to the quickening reduction in child mortality.

The less developed countries in Latin America, Asia, and Africa were slow to get much benefit from the advances in k at this stage. They were handicapped in applying the new knowledge by low income, low levels of popular education, small resources at the disposal of the governments, and perhaps colonial administrations' lack of interest in taking very costly actions to protect the health of the indigenous people. So the beginnings of the second phase of the transition were delayed in most of these countries until after World War I, and in many until the 1940s or 1950s. As a result, the differences in life expectancy between more and less developed countries widened during the early decades of the present century, and the slope of the curve of e in relation to y grew steeper as it shifted upward more rapidly in the higher than in the lower brackets of per capita income.

It might be tempting to infer that countries had to reach some threshold of income and development in other respects to be eligible for rapid progress in the reduction of mortality under the conditions of this period. But some observations imply that if this were true, the level of the threshold was not high enough to explain fully why so many less developed countries were so long retarded in entering the second phase of the mortality transition:

1. The case of the eastern and southern European countries: Although they were considerably less developed than the northwestern European countries, they were quick to join their richer neighbors in the sharp acceleration of gains in e around the turn of the century or shortly afterward. In spite of handicaps in income, education, and other

aspects of development, the countries in eastern and southern. Europe generally managed to keep pace with those of northwestern Europe in rates of gain in life expectancy until about the 1940s, when they began to overtake the lead of the latter.

2. The case of Cuba: In a new study of the trend of mortality in Cuba since the late nineteenth century, Diaz Briquets (1977) shows that a spectacular reduction of mortality was achieved there during the few years of United States military occupation following the Spanish-American War, by a campaign of sanitary reforms and mosquito control instigated and aided by the army of occupation. He estimates that the crude annual death rate in the city of Havana dropped from a prewar average of 32 per 1,000 in 1891–95 to 20 in 1903–7, and a large decrease was achieved in the rest of the country also, in the face of general poverty and illiteracy.

3. The case of Taiwan under Japanese rule during early years of this century is another precocious example of the effective transfer of *k* and *s* from a more developed to a less developed country (Barclay 1954). This is even more remarkable than the Cuban case, because when the Japanese arrived Taiwan was a good deal less developed than Cuba, and the Japanese themselves had not yet reached a very high level of either *e* or *y*.

4. A decided upturn of the trend of *e* took place during the 1920s in a number of less developed countries (Cuba, Japan, Taiwan, and others). Diaz Briquets (1977) observes that this seems to have been especially characteristic of countries where export industries were dominant, and he suggests that an economic boom in such countries in the 1920s owing to expanding demand and rising prices for their exports might account for their having entered the second phase of the mortality transition earlier than other less developed countries did. The interest of their more developed trading partners in making these countries healthy places to do business with and in may also have been a factor.

The relevance of the trend of income to the life expectancy gains in less developed countries is illustrated in reverse by the example of Cuba in the 1930s and early 1940s, when the misfortunes of the international market for sugar cast Cuba into economic doldrums. Diaz finds that the decline of mortality in Cuba was checked and probably temporarily reversed during this period, and that worsening nutrition and diminishing public and private expenditures on health services were important factors.

A second revolution in the technology of disease control began about 1935 and progressed rapidly during the 1940s and 1950s, with major advances of *k* especially in the fields of immunization, chemotherapy, and chemical control of disease vectors. This time, the less developed 203 countries were the principal beneficiaries. Although measures of trends in e since 1940 are lacking for many of these countries, especially among those at the lowest levels of development, it is apparent that substantial gains since that time have been practically universal in the less developed regions of the world. Preston's findings suggest that 50% to 80% of the gains between 1940 and 1970 in less developed countries may be attributable to k and s factors, but advances on these fronts have not gone so far as to make economic factors irrelevant. Gains in e since the 1940s have been less spectacular in the least developed countries, particularly in Africa, than in those that were somewhat more developed, and Preston finds a positive association between rates of increase in e and y among less developed countries.

A tendency toward slackening rates of gain in e in less developed countries is apparent in the 1960s and 1970s. While this might be due partly to slowing economic growth, Preston links it with a diminishing rate of "structural change," that is, slowdown of the advance of k and s. He observes that only a few innovations of major importance to health technology have been made during the last decade. The implication is that upward shifting of e in relation to y may be drawing to an end and that e gains in less developed countries henceforth may depend mainly on their ability to move up on the scales of y and related social developments. This has been suggested in a number of recent studies. However, there may still be a good deal of scope for raising e in less developed countries where it remains relatively low, through the pursuit of s actions to take fuller advantage of existing k at their present levels of y. To cite once again the example of Cuba: the series of Cuban life tables compiled by Diaz Briquets (1977) shows expectation of life at birth for the two sexes increasing from 58.8 years in 1953 to 72.1 in 1971 (70.6 for males, 73.9 for females). The 1971 figure compares favorably with that of the United States, especially for males. Diaz presumes that most of the gain in Cuba since 1953 has taken place since the establishment of the socialist government, and he attributes the high rate of gain since that time mainly to more equal distribution of income and government policies aimed at equalizing access to health services for all categories of the population.

Although the less developed countries were the main beneficiaries of the new advances in k since the 1930s, the more developed countries also benefited to an important extent. Examination of the trends in a number of more developed countries shows that the 1940s were a bumper decade for gains in life expectancy. Preston's analysis indicates that a major share of the gains in more developed as well as less developed countries since 1930 has been due to the upward shifting of the curve of e in relation to y and other indicators of development, which may be attributed to the advances of k and s.

Third Phase of the Transition

In more developed countries, the rates of gain in *e* have slowed conspicuously since the 1950s, and in many of them hardly any gains have been registered during the past ten years. The same tendency is noticeable in recent statistics from some less developed countries that have attained levels of *e* comparable to those of more developed countries. This feature is commonly interpreted as meaning that the expectation of life in countries where it is now highest is approaching a ceiling that cannot be surpassed by increasing income or by other means unless a new revolution in medical technology is achieved—a revolution that would make possible important gains in control over the so-called degenerative diseases. If this interpretation is correct and if the less developed countries where *e* is still well below such a ceiling manage to continue progress in reducing their mortality rates through economic development and fuller application of the present medical knowledge, the time may come when levels of mortality will be nearly equalized among countries around the world. The advances of *k* and *s* would then have neutralized, to a large extent, the influence of *y* as well as that of *n*.

An interesting aspect of the recent trends in countries where *e* is high is that they seem to be leveling off at considerably different values of *e*. If they are coming up against a ceiling, the level of the ceiling seems not to be the same in all countries. In the United States, for example, *e* seems to be stagnating some four or five years below the level achieved in the Scandinavian countries. On the surface, these differences do not seem very consistently related to per capita income. I mentioned earlier that Cuba's estimated expectation of life in 1971 compared favorably with that of the United States. The position of Puerto Rico and Hong Kong is similar, although they are far below the United States in per capita income. Perhaps a part of the explanation for such anomalies might be found in factors associated with advanced economic development that are unfavorable to health and longevity. I think it would be interesting to make a systematic study of factors related to the different levels of *e*, and of mortality rates for sex and age groups, causes of death, and so forth, at which the trends seem recently to have been stalling in many countries. The distribution of income, governmental action in fields of health care, environmental pollution, diets, and behavioral patterns relevant to health are among the factors that might usefully be examined.

Note

1. McNeill (1976) argues that major factors in the long-range trends of mortality and population growth in Europe between the fourteenth and eighteenth centuries were new diseases resulting from increased contacts with the Orient during the Mongol conquests and a subsequent gradual adaptation to these diseases, as well as variations of climate.

204

Comment Victor R. Fuchs

The paper by Preston has two principal purposes: to explain the increase in life expectancy in the less developed countries over the past several decades and to consider the effect of this increase on population size, output, and output per capita. The bulk of the paper is concerned with the first question, and I shall limit my comment to that. Furthermore, I shall examine only one aspect of Preston's multifaceted discussion—the attempt to partition the gain in life expectancy into the portion attributable to increased per capita income and the portion due to a structural shift in the relationship between life expectancy and per capita income.

Using the data Preston provides, I have run regressions of life expectancy (*LE*) on the natural logarithm of per capita income (*LnY*) for four separate groups of countries: LDCs in 1940 (*L40*), MDCs in 1940 (*M40*), LDCs in 1970 (*L70*), and MDCs in 1970 (*M70*).[1] The results are presented in table C5.1. I have also plotted the predicted (from the regressions) relationship between life expectancy and per capita income for each group in figure C5.1. The curves are plotted over the range of per capita income actually observed for each group.

Table C5.1 **Results of Regressing Life Expectancy on Logarithm of per Capita Income across Less Developed and More Developed Countries in 1940 and 1970**

$LE = a + b \ln Y$	L40	M40	L70	M70
b	4.99	6.65	7.82	.13
σ_b	(1.95)	(1.21)	(1.10)	(.85)
a	13.3	17.1	12.5	70.8
σ_a	(10.0)	(8.0)	(6.7)	(6.6)
R^2	.27	.68	.77	−.08
N	16	15	16	15

Inspection of figure C5.1 suggests that the structural relation between life expectancy and per capita income in the LDCs in 1940 was different from that in the MDCs in the same year. Given the level of income, life expectancy seems to have been appreciably higher in the MDCs. It is therefore inappropriate to pool the two groups of countries, as Preston does, without allowing for differences in structure. It should also be

Victor R. Fuchs is professor of economics at Stanford University and is a research associate of the National Bureau of Economic Research.

The author acknowledges the financial support of the Robert Wood Johnson Foundation, the research assistance of Sean Becketti, and the helpful comments of Robert J. Willis.

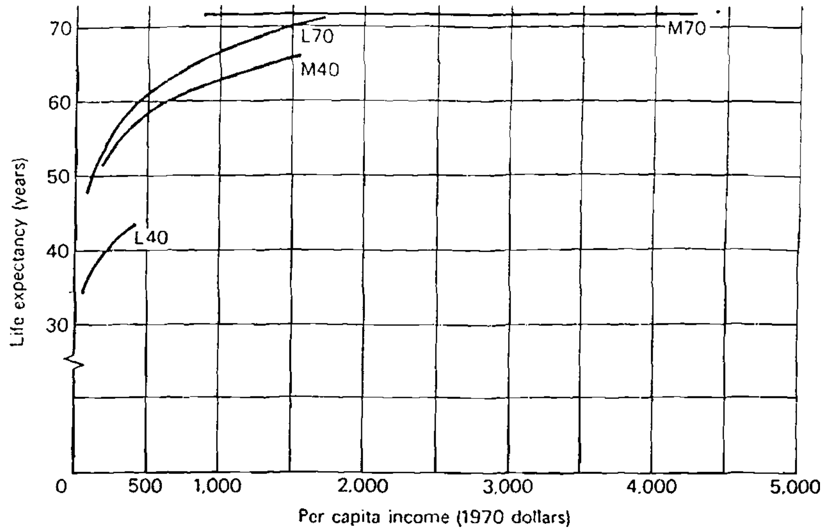


Fig. C5.1 The relationship between life expectancy and per capita income across less developed and more developed countries in 1940 and 1970.

noted that in the MDCs in 1970, the relationship between life expectancy and per capita income has disappeared, a phenomenon I have called attention to before (Fuchs 1965, 1974). There can be no question, therefore, of attempting to assess shifts in the function between the MDCs in 1940 and 1970 or between the LDCs and MDCs in 1970.

I have run regressions that pool LDCs and MDCs in 1940 and LDCs in 1940 and 1970 with dummy variables inserted to allow for differences in intercepts and slopes. These results are reported in table C5.2.

The first regressions (part A) constrain the slopes of the pooled groups to be equal but allow the intercepts to vary. We see that the shift coefficients are large and highly significant in both cases. This means that, compared with the LDCs in 1940, both the MDCs in 1940 and the LDCs in 1970 had substantially higher life expectancy for any given level of per capita income. The second set of regressions (part B) constrains the intercepts to be equal and allows the slopes to vary within each pair of groups. We now find that the slopes do differ significantly. The third set of regressions (part C) allows both the intercepts and the slopes to vary, and with this specification none of the interactions are statistically significant.

If one does not demand statistical significance, it is possible to answer Preston's question fairly unambiguously along the following lines. The mean life expectancy of the LDCs rose from 38.8 years in 1940 to 59.6

205

Table C5.2 Pooled Regression Results with Interactions

	L40 + M40	L40 + L70
A. $LE = a + b \ln Y + c$ Intercept		
b	5.83	6.83
σ_b	(1.14)	(1.04)
c	−13.4	−14.6
σ_c	(2.10)	(1.70)
a	22.4	18.4
σ_a	(7.54)	(6.34)
$\bar{R}^2$	.91	.89
N	31	32
B. $LE = a + b \ln Y + d$ Slope		
b	7.00	7.77
σ_b	(1.00)	(.95)
d	−2.29	−2.69
σ_d	(.35)	(.30)
a	14.7	12.8
σ_a	(6.50)	(5.74)
$\bar{R}^2$	.91	.90
N	31	32
C. $LE = a + b \ln Y + c$ Intercept $+ d$ Slope		
b	6.65	7.82
σ_b	(1.63)	(1.27)
c	−3.79	.78
σ_c	(13.60)	(11.80)
d	−1.66	−2.83
σ_d	(2.31)	(2.15)
a	17.1	12.5
σ_a	(10.7)	(7.73)
$\bar{R}^2$	.91	.90
N	31	32

years in 1970, a rate of increase of 1.4% per annum (see table C5.3).[2] Over that same period the mean per capita income (in 1970 dollars) rose from \$194 to \$560. We can estimate what the change in life expectancy would have been as a result of income change alone by moving along either the 1940 predicted relation or the 1970 predicted relation. The former tells us that life expectancy would have changed from 39.6 years to 44.8 years, an increase of 0.4% per annum. If we calculate the change along the 1970 curve, we get a predicted increase of 0.5% per annum, from 53.7 years to 62.0 years.

Alternatively, we can look at the implied rates of change attributable to structural shift by comparing predicted life expectancies at the same per capita income in the two years. At \$194 the implied change is 1.0% per annum; at \$560 it is 1.1% per annum. Thus, either approach indicates that about one-third of the observed change in life expectancy in

Table C5.3 **Life Expectancy in Less Developed Countries in 1940 and 1970: Actual and Predicted Changes**

	Mean Life Expectancy (Years)
Actual 1940 (A40)	38.78
Actual 1970 (A70)	59.59
Predicted 1940 (P40:40) (from L40 regression)	39.56
Predicted 1970 (P70:40) (from L40 regression)	44.85
Predicted 1940 (P40:70) (from L70 regression)	53.68
Predicted 1970 (P70:70) (from L70 regression)	61.97
	Rates of Change (Percentage per Annum)
A40 to A70	1.4
P40:40 to P70:40	0.4
P40:70 to P70:70	0.5
P40:40 to P40:70	1.0
P70:40 to P70:70	1.1

the LDCs between 1940 and 1970 can be attributed to the growth of per capita income, and about two-thirds to a shift in the life expectancy–income relationship. Preston presents one estimate of 50% due to structural change and another of 80% due to that source. The results presented here are quite consistent with those estimates.

Notes

1. The LDCs are all in Asia, Africa, and Latin America, and all had life expectancies below 50 years in 1940. The MDCs are all in Europe, North America, and Australia, and all had life expectancies above 50 years in 1940.
2. The change is expressed in percentage per annum in order to minimize the problem of interaction between shifts in the function and movements along the function.

206

Comment Richard W. Parks

Preston presents an interesting regression test of the relative importance of private and national income in determining life expectancies. In light of the Kuznets and Fishlow discussions of interaction between the income distribution and the age distribution of the population, it may be useful to point out a possible bias in the Preston regression in section 5.1.1.

The income distribution as commonly measured does not correct for the age distribution of the population. Thus, in a hypothetical world with *no* differences among individuals in their lifetime income streams, we will observe considerable income inequality as conventionally measured if individuals follow the usual life-cycle pattern of earning and saving followed by retirement. For the determination of the effect of income on life expectancy, it appears that a permanent income rather than a measured income concept makes more sense, but given the data available to him, Preston relies on the distribution of measured income.

Preston's distribution measure, which I shall call $D = \Sigma\ ln(S_i/.05)$, takes values ranging from $-\infty$ to 0 on a scale representing increasing equality. Thus we can represent the partial relationship between life expectancy e^o_o, and D as shown in figure C5.2.

We expect a positive association. Suppose we start at point A with given (unequal) distribution and low life expectancy. If incomes were to become more equal (in a life-cycle sense), we would expect to find a new point at B showing greater equality and higher life expectancy. However, even with the pattern of life-cycle income corresponding with point B, there is likely to be an effect on the measured income distribu-

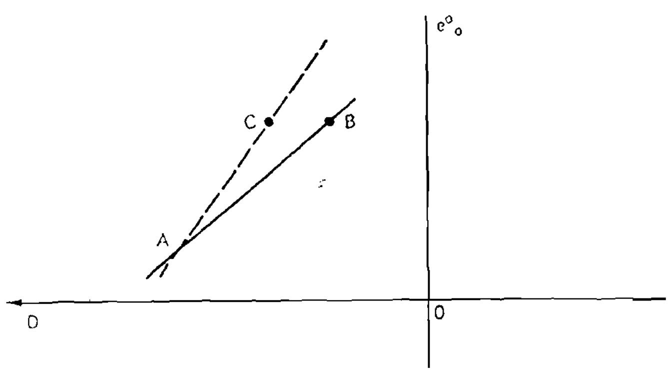


Fig. C5.2

Richard W. Parks is associated with the Department of Economics, University of Washington.

tion arising from the altered life expectancy. An increase in the share of population in nonproductive years, for example, older age groups, will have the effect of increasing the observed inequality. The observation based on measured income will be at point *C*, giving an upward bias to the slope coefficient. Since the crucial test of the relative importance of national and private incomes in the determination of life expectancy depends on the size of the slope coefficient, the upward bias would tend to suggest the absence of an effect for national income even when it was in fact important.

References

Allan, William. 1965. *The African husbandman.* London: Oliver and Boyd.

Antonovsky, Aaron. 1967. Social class, life expectancy, and overall mortality. *Milbank Memorial Fund Quarterly* 45: 31–73.

Arriaga, Eduardo. 1967. The effect of a decline in mortality on the gross reproduction rate. *Milbank Memorial Fund Quarterly* 45: 333–52.

———. 1968. *New life tables for Latin American populations.* Berkeley: Institute of International Studies, University of California.

———. 1970. *Mortality decline and its demographic effects in Latin America.* Berkeley: Institute of International Studies, University of California.

Ascoli, Werner; Guzman, Miguel A.; Scrimshaw, Nevin S.; and Gordon, John E. 1967. Nutrition and infection field study in Guatemalan villages, 1959–1964. IV. Deaths of infants and preschool children. *Archives of Environmental Health* 15 (October): 439–49.

Baertl, Juan M.; Morales, Enrique; Verastegui, Gustavo; and Graham, George. 1970. Diet supplementation for entire communities: Growth and mortality of children. *American Journal of Clinical Nutrition* 23(6): 707–15.

Bailey, K. V. 1975. Malnutrition in the African region. *Chronicle of the World Health Organization* 29: 354–64.

Banks, Arthur. 1971. *Cross-polity time-series data.* Cambridge: MIT Press.

Barclay, George W. 1954. *Colonial development and population in Taiwan.* Princeton: Princeton University Press.

Barlow, Robin. 1967. The economic effects of malaria eradication. *American Economic Review* 57(2): 130–48.

———. 1968. *The economic effects of malaria eradication.* Economic Research Series no. 15. Ann Arbor: University of Michigan School of Public Health.

207

Beck, Ann. 1970. *A history of the British medical administration of East Africa, 1900–1950*. Cambridge: Harvard University Press.

Boserup, Ester. 1965. *The conditions of agricultural growth: The economics of agrarian change under population pressure*. Chicago: Aldine.

Bowers, John. 1973. The history of public health in China to 1937. In *Public Health in the People's Republic of China*, ed. Myron E. Wegman, Tsung-Yi Lin, and Elizabeth F. Purcell, pp. 26–45. New York: Josiah Macy, Jr., Foundation.

Boyd, M. F. 1949. *Malariology*. Vols. 1 and 2. Philadelphia: W. B. Saunders.

Bryant, John. 1969. *Health and the developing world*. Ithaca: Cornell University Press.

Bunle, Henri. 1954. *Le mouvement naturel de la population dans le monde de 1906 à 1936*. Paris: Institut d'études demographiques.

Burch, Thomas K., and Gendell, Murray. 1970. Extended family structure and fertility: Some methodological and conceptual issues. *Journal of Marriage and the Family* 32(2): 227–36.

Carnoy, Martin. 1967. Rates of return to schooling in Latin America. *Journal of Human Resources* 2(3): 359–74.

Coale, Ansley J. 1956. The effects of changes in mortality and fertility on age composition. *Milbank Memorial Fund Quarterly* 34: 79–114.

Coale, Ansley J., and Demeny, Paul. 1966. *Regional model life tables and stable populations*. Princeton: Princeton University Press.

Coale, Ansley J., and Hoover, Edgar M. 1958. *Population growth and economic development in low-income countries*. Princeton: Princeton University Press.

Committee for International Coordination of National Research in Demography. 1975. *Seminar on infant mortality in relation to level of fertility*. Paris.

Davis, Kingsley. 1951. *The population of India and Pakistan*. Princeton: Princeton University Press.

———. 1955. Institutional patterns favoring high fertility in underdeveloped areas. *Eugenics Quarterly* 2: 33–39.

———. 1956. The amazing decline of mortality in underdeveloped areas. *American Economic Review* 46 (May): 305–18.

———. 1973. Cities and mortality. *Proceedings*, International Population Conference, Liège, Belgium, 3: 259–81.

Denison. Edward F. 1962. *The sources of economic growth in the United States and the alternatives before us*. New York: Committee for Economic Development.

Diaz Briquets, Sergio. 1977. Mortality in Cuba: Trends and determinants, 1880 to 1971. Ph.D. diss., University of Pennsylvania.

Durand, John R. 1967. A long-range view of world population growth. *Annals of the American Academy of Arts and Sciences* 369: 1–8.

Early, John. 1970. The structure and change of mortality in a Maya community. *Milbank Memorial Fund Quarterly* 48: 179–201.

Eversley, D. E. C. 1957. A survey of population in an area of Worcestershire from 1660 to 1850 on the basis of parish registers. *Population Studies* 10: 253–79.

Faust, Ernest Carroll. 1941. The distribution of malaria in North America, Mexico, Central America, and the West Indies. In *A symposium on human malaria*, ed. F. R. Moulton, pp. 8–18. Washington, D.C.: American Association for the Advancement of Science.

Fein, Rashi. 1964. Health programs and economic development. In *The economics of health and medical care. Proceedings of the conference on the economics of health and medical care, May 10–12, 1962*, ed. H. E. Klarman. Ann Arbor: Bureau of Public Health Economics and Department of Economics, University of Michigan.

Foege, William H.; Millar, J. D.; and Henderson, D. A. 1975. Smallpox eradication in West and Central Africa. *Bulletin of the World Health Organization* 52: 209–22.

Food and Agriculture Organization of the United Nations. 1975. *Population, food supply, and agricultural development*. Rome: FAO.

Fredericksen, Harald. 1961. Determinants and consequences of mortality trends in Ceylon. *Public Health Reports* 76 (August): 659–63.

———. 1966*a*. Determinants and consequences of mortality and fertility trends. *Public Health Reports* 81 (August): 715–27.

———. 1966*b*. Dynamic equilibrium of economic and demographic transition. *Economic Development and Cultural Change* 14 (April): 316–22.

Friedlander, Dov. 1969. Demographic responses and population change. *Demography* 6(4): 359–82.

Fuchs, V. R. 1965. Some economic aspects of mortality in the United States. National Bureau of Economic Research Study Paper.

———. 1974. Some economic aspects of mortality in developed countries. In *The economics of health and medical care*, ed. Mark Perlman, pp. 174–93. Proceedings of a conference held by the International Economic Association at Tokyo. London: Macmillan.

Goodman, Neville M. 1971. *International health organizations and their work*. Edinburgh: Churchill Livingstone.

Gordon, John. 1969. Social implications of nutrition and disease. *Archives of Environmental Health* 18(2): 216–34.

Habakkuk, H. J. 1971. *Population growth and economic development*. New York: Humanities Press.

208

Hansluwka, Harald. 1975. Health, population, and socio-economic development. In *Population growth and economic development in the Third World*, ed. Léon Tabah, 1: 191–250. Dolhain, Belgium: Ordina.

Hinman, E. Harold. 1966. *World eradication of infectious diseases.* Springfield, Ill.: Charles C. Thomas.

Howard-Jones, N. 1974. The scientific background of the International Sanitary Conferences, 1851–1938. *Chronicle of the World Health Organization* 28(10): 455–70.

Johnson, D. Gale. 1974. Population, food, and economic adjustment. *American Statistician* 28(3): 89–93.

Keyfitz, Nathan, and Fleiger, Wilhelm. 1968. *World population: An analysis of vital data.* Chicago: University of Chicago Press.

———. 1971. *Population: Facts and methods of demography.* San Francisco: W. H. Freeman and Company.

Kitagawa, E. M., and Hauser, P. M. 1973. *Differential mortality in the United States: A study in socioeconomic epidemiology.* Cambridge: Harvard University Press.

Klarman, Herbert E. 1967. Present status of cost-benefit analysis in the health field. *American Journal of Public Health* 57(11): 1948–53.

Kuznets, Simon. 1975. Population trends and modern economic growth: Notes towards an historical perspective. In *The population debate: Dimensions and perspectives*, 1: 425–33. Papers of the World Population Conference, Bucharest: New York: United Nations.

Lancet. 1970. Strategy of malaria eradication: A turning point? *Lancet* editorial. 1970 (21 March): 598–600.

McDermott, W. 1966. Modern medicine and the demographic-disease pattern of overly traditional societies: A technological misfit. *Journal of Medical Education* 41(9): 138–62.

McKeown, Thomas. 1965. Medicine and world population. In *Public health and population change*, ed. Mindel C. Sheps and Jeanne Clare Ridley. Pittsburgh: University of Pittsburgh Press.

McKeown, Thomas, and Record, R. G. 1962. Reasons for the decline of mortality in England and Wales during the 19th century. *Population Studies* 16: 94–122.

McNeill, William H. 1976. *Plagues and peoples.* Garden City, N.Y.: Anchor Press/Doubleday.

Malenbaum, Wilfred. 1970. Health and productivity in poor areas. In *Empirical studies in health economics*, ed. Herbert E. Klarman, pp. 31–57. Baltimore: Johns Hopkins University Press.

Mandle, Jay R. 1970. The decline in mortality in British Guiana, 1911–1960. *Demography* 7, no. 3 (August): 303.

Maramag, Ileana. 1965. The cost of public health programs in the Philippines. *Journal of the American Medical Women's Association* 20 (September): 848–57.

Marshall, Carter L.; Brown, Roy E.; and Goodrich, Charles H. 1971. Improved nutrition vs. public health services as major determinants of world population growth. *Clinical Pediatrics* 10 (July): 363–68.

Mata, Leonardo, et al. 1972. Influence of recurrent infections on nutrition and growth of children in Guatemala. *American Journal of Clinical Nutrition* 25 (November): 1267–75.

Meegama, S. A. 1967. Malaria eradication and its effect on mortality levels. *Population Studies* 26(3): 207–38.

Morley, David. 1973. *Paediatric priorities in the developing world.* London: Butterworth's.

Mushkin, Selma J. 1962. Health as an investment. *Journal of Political Economy* 70 (part 2, supplement): 129–57.

———. 1964. Health programming in developing nations. *International Development Review* 6(1): 7–12.

Myrdal, Gunnar. 1968. *Asian drama: An inquiry into the poverty of nations.* Vols. 1–3. New York: Random House.

Newman, P. 1965. *Malaria eradication and population growth: With special reference to Ceylon and British Guiana.* Research Series no. 10. Ann Arbor, Michigan: Bureau of Public Health Economics, School of Public Health, University of Michigan.

———. 1970. Malaria control and population growth. *Journal of Development Studies* 6: 133–58.

O'Hara, Donald J. 1975. Microeconomic aspects of the demographic transition. *Journal of Political Economy* 83(6): 1203–15.

Organization for Economic Cooperation and Development. N.d. *Resources for the developing world: The flow of financial resources to less developed countries, 1962–1968.* Paris: OECD.

Pampana, E. J. 1954. Effect of malaria control on birth and death rates. *Proceedings of the World Population Conference, 1954*, Rome, pp. 497–508. New York: United Nations.

Pampana, E. J., and Russell, Paul F. 1955. Malaria: A world problem. *Chronicle of the World Health Organization* 9(2–3): 33–100.

Pani, Alberto. 1917. *Hygiene in Mexico.* New York: G. P. Putnam's Sons.

Paul, Hugh H. 1964. *The control of diseases.* 2d ed. Baltimore: Williams and Wilkins.

Plank, S. J., and Milanesi, M. L. 1973. Infant feeding and infant mortality in rural Chile. *Bulletin of the World Health Organization* 48: 203–10.

Polgar, Steven. 1972. Population history and population policies from an anthropological perspective. *Current Anthropology* 13(2): 203–11.

Preston, S. H. 1974. Effect of mortality change on stable population parameters. *Demography* 11(1): 119–30.

209

———. 1975*a*. The changing relation between mortality and level of economic development. *Population Studies* 29(2): 231–48.

———. 1975*b*. Introduction. In *Seminar on infant mortality in relation to the level of fertility*, pp. 10–22. Paris: Committee for International Coordination of National Research in Demography.

Preston, S. H.; Keyfitz, N.; and Schoen, R. 1972. *Causes of death: Life tables for national populations*. New York: Seminar Press.

———. 1973. Cause-of-death life tables: Application of a new technique to worldwide data. *Transactions of the Society of Actuaries* 25 (December): 83–109.

Preston, S. H., and Nelson, V. E. 1974. Structure and change in causes of death: An international summary. *Population Studies* 28(1): 19–51.

Puffer, R. R., and Griffith, G. W. 1967. *Patterns of urban mortality*. Scientific Publication no. 151. Washington, D.C.: Pan American Health Organization.

Puffer, R. R., and Serrano, C. V. 1973. *Patterns of mortality in childhood*. Scientific Publication no. 262. Washington, D.C.: Pan American Health Organization.

Russell, Paul F. 1943. Malaria and its influence on world health. *Bulletin of the New York Academy of Medicine, September*, pp. 599–630.

———. 1968. The United States and malaria: Debits and credits. *Bulletin of the New York Academy of Medicine* 44(6): 623–53.

Rutstein, Shea, and Medica, Vilma. 1975. The effect of infant and child mortality in Latin America. In *Seminar on infant mortality in relation to the level of fertility*, pp. 225–46. Paris: Committee for International Coordination of National Research in Demography.

Sack, Richard. 1974. The impact of education on individual modernity in Tunisia. In *Education and individual modernity in developing countries*, ed. Alex Inkeles and Donald B. Holsinger, pp. 89–116. Leiden: E. J. Brill.

Sanchez-Albornoz, Nicholas. 1974. *The population of Latin America: A history*. Berkeley: University of California Press.

Schultz, Theodore. 1964. *Transforming traditional agriculture*. New Haven: Yale University Press.

Schultz, T. Paul. 1976. Interrelationships between mortality and fertility. In *Population and development: The search for selective interventions*, ed. Ronald G. Ridker, pp. 239–89. Baltimore: Johns Hopkins University Press.

Scrimshaw, Nevin S. 1970. Synergism of malnutrition and infection. *Journal of the American Medical Association* 212(10): 1685–92.

Sharpston, Michael J. 1976. Health and the human environment. *Finance and Development* 13(1): 24–28.

Shattuk, George. 1951. *Diseases of the tropics.* New York: Appleton-Century-Crofts.

Shorter, Frederick. 1968. Information on fertility, mortality, and population growth in Turkey. *Population Index* 34(1): 5.

Simmons, James Stevens; Whayne, Tom F.; Anderson, Gaylord West; Horack, Harold MacLachian; and collaborators. 1944. *Global epidemiology: A geography of diseases and sanitation.* Vol. 1. Philadelphia: J. B. Lippincott.

Smith, Peter C. 1976. Asian nuptiality in transition. Paper presented at the Seventh Summer Seminar in Population, East-West Population Institute, Honolulu.

Sorkin, Alan L. 1975. *Health economics: An introduction.* Lexington, Mass.: D. C. Heath, Lexington Books.

Stolnitz, George J. 1956. Mortality declines and age distribution. *Milbank Memorial Fund Quarterly* 34: 178–215.

———. 1974. International mortality trends: Some main facts and implications. Background Paper for United Nations World Population Conference E/CONF.60/CBP/17, Bucharest.

Taylor, Carl E., and Hall, Marie-Francoise. 1967. Health population and economic development. *Science* 157 (August): 651–57.

Titmuss, Richard M., and Abel-Smith, Brian. 1968. *Social policies and population growth in Mauritius.* London: Frank Cass.

United Nations, Population Branch. 1963. Population bulletin of the United Nations no. 6 (with special reference to the situation and recent trends of mortality in the world). New York.

United Nations, Department of Economic and Social Affairs. 1971. *The world population situation in 1970.* Population Study no. 49. New York.

———. 1973. *The determinants and consequences of population trends.* Vol. 1. New York: United Nations.

United Nations, Population Division. 1974. Recent population trends and future prospects. World Population Conference Paper E/CONF. 60/3.

Vaidyanathan, K. E. 1972. Some indices of differential mortality in India. In *Studies on mortality in India*, ed. K. E. Vaidyanathan. Gandhigram: Institute of Rural Health and Family Planning.

Vallin, J. 1968. La mortalité dans les pays du Tiers Monde: Evolution et perspectives. *Population* 23: 845–68.

Wegman, Myron E.; Lin, Tsung-yi; and Purcell, Elizabeth F., eds. 1973. *Public health in the People's Republic of China.* New York: Josiah Macy, Jr., Foundation.

Weisbrod, Burton A. 1975. Research in health economics: A survey. *International Journal of Health Services* 5(4): 643–61.

210

Weller, T. H. 1974. World health in a changing world. *Journal of Tropical Medicine* 77(4), suppl. 54: 54–61.
World Bank. 1975. *Health: Sector policy paper.* Washington, D. C.
World Health Organization. 1970. Special subject: Health expenditure. In *World health statistics annual, 1966–1967* 3: 236–45.
———. 1972. *Human development and public health.* Technical Report Series no. 485. Geneva: WHO.
———. 1973. Community water supply and sewage disposal in developing countries. *World Health Statistics Report* 26(11): 720–83.
———. 1974*a*. Childhood mortality in the Americas. *Chronicle of the World Health Organization* 28: 276–82.
———. 1974*b*. Health trends and prospects, 1950–2000. *World Health Statistics Report* 27(10): 672–706.
———. 1975*a*. The malaria situation in 1974. *Chronicle of the World Health Organization* 29: 474–81.
———. 1975*b*. *Fifth report on the world health situation.* Official Records of the World Health Organization, no. 225. Geneva: WHO.
———. 1975*c*. *First WHO seminar on expansion of the use of immunization in developing countries.* Offset Publication no. 16. Geneva: WHO.
———. 1976. Smallpox eradication in 1975. *Chronicle of the World Health Organization* 30: 152–57.
Wrigley, E. A. 1969. *Population and history.* New York: McGraw-Hill.

211

212

UNA TEORÍA DEL DESARROLLO ECONÓMICO *

Gustav Ranis y *John C. H. Fei*

Este artículo pretende hacer una contribución a la teoría del crecimiento mediante un riguroso análisis del proceso de transición a través del cual una economía subdesarrollada espera pasar de la fase de estancamiento a la de crecimiento autosostenido. Puesto que la totalidad de las economías que llevan la etiqueta del "subdesarrollo" rechazan la fácil generalización, nos ocuparemos aquí principalmente del excedente de fuerza de trabajo, de la escasa variedad de recursos de la mayor parte de la población agrícola en condiciones de desocupación disfrazada y tasas elevadas de crecimiento de la población. Confiamos realizar nuestra tarea inspirándonos liberalmente en el acervo de ideas aceptadas y, en seguida, procedemos a entrelazarlas en un modelo general explicativo del crecimiento económico.

Nuestro análisis comienza en el punto de partida de una economía casi estancada o de iniciación del llamado proceso de despegue.[1] Rostow lo define como un periodo de dos o tres décadas, durante las cuales la economía se transforma de manera tal que el crecimiento económico se lleva a cabo, subsecuentemente, en forma más o menos automática; se caracteriza por una disminución de la proporción de la población rural, por la duplicación de las tasas de ahorro y por el florecimiento inicial y continuo de la industria, estimulado por la disponibilidad de mano de obra excedente [11, pp. 25-32]. Esta noción intuitiva, bien conocida, se ha escogido como nuestro punto de partida. Para nuestro instrumental analítico básico, sin embargo, utilizamos ampliamente el trabajo de Arthur Lewis.

En sus célebres artículos, Lewis [3] [4] presenta un modelo de dos sectores e investiga la expansión del sector capitalista o industrial, tal como es alimentado por la oferta de fuerza de trabajo barata que proviene del sector de subsistencia o agrícola.[2] El desarrollo consiste de una redistribución de los trabajadores agrícolas excedentes, cuya contribución a la producción pudo haber sido cero o insignificante, para la industria, donde se convierten en miembros productivos de la fuerza de trabajo con un salario

* Los autores son, respectivamente, profesor asistente de la Universidad de Yale y profesor asociado en el Antioch College. Este artículo se comenzó cuando ambos se hallaban asociados al Instituto de Economía del Desarrollo, Karachi, Pakistán. Se agradecen cumplidamente los comentarios de Bela Balassa y John M. Montias de la Universidad de Yale. Véase The American Economic Review, vol LI, núm. 4, septiembre de 1961, pp. 533-565. Versión al castellano de Héctor Rodríguez Licea.

1 No se trata de subestimar la importancia del periodo de acondicionamiento previo (véase [1] y [9] cuando se movilizan fuerzas institucionales potencialmente expansionistas y capacitan al sistema para responder en forma positiva a estímulos que se presentan al azar.

2 Deseamos subrayar la ausencia de cualquier relación unívoca entre el sector de subsistencia y la agricultura o entre el sector capitalista y la industria en la mayoría de las economías menos desarrolladas. La existencia de importantes islas de producción comercializada en el sector primario y de apreciables enclaves de subsistencia en las industrias de servicios y en pequeña escala, no impide, sin embargo, que Lewis utilice esta terminología abreviada.

EL TRIMESTRE ECONÓMICO Vol XXIX n.º 114
México Abril-junio 1962. Fondo de Cultura Económica

212

213

igual (o atado) al salario institucional de la agricultura. Este proceso continúa hasta que la curva de la oferta de mano de obra industrial comienza a declinar.

Sin embargo, Lewis no ha presentado un análisis satisfactorio del sector agrícola o de subsistencia. Parece claro que este sector también debe crecer para que el mecanismo que describe no se vea conducido a detención prematura. El desenvolvimiento de la noción de un necesario equilibrio en el crecimiento nos lleva entonces a una definición lógica del fin del periodo de despegue.

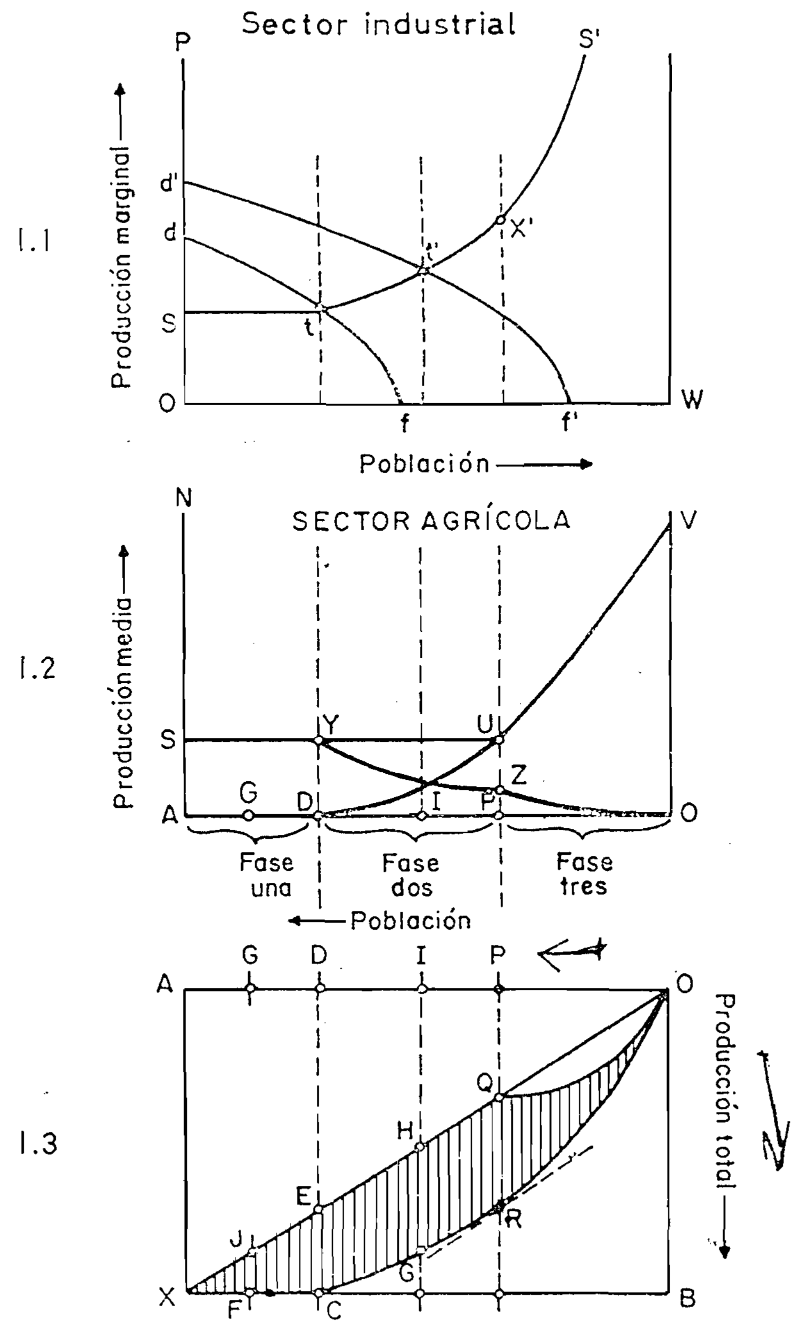
Finalmente, la economía debe resolver el problema malthusiano con objeto de asegurar el buen éxito del proceso de desarrollo en forma equilibrada. Consideraciones de esta naturaleza han dado lugar a la llamada teoría del "esfuerzo crítico mínimo" [2], que se ocupa de la magnitud del esfuerzo necesario para salir más que temporalmente del estancamiento. Mostraremos en el curso de nuestro análisis que el concepto del esfuerzo crítico mínimo no presupone ninguna magnitud absoluta de esfuerzo, sino que contiene una dimensión automática de tiempo que permite la variación del esfuerzo con la duración del proceso de despegue.

Así pues, la contribución de este artículo es construir una teoría del crecimiento económico de la cual las ideas anteriores, rigurosamente formuladas, constituyen sus partes componentes. En la Sección I presentamos los supuestos básicos estructurales de nuestro modelo, con énfasis en el análisis del papel del sector agrícola. La Sección II generaliza el análisis "estático" anterior admitiendo la posibilidad de un cambio de productividad en el sector agrícola. En la Sección III introducimos cambios en la productividad industrial y en la noción del "criterio de crecimiento equilibrado", por medio del cual se define formalmente la terminación del proceso de despegue. La Sección IV continúa con una precisa formulación matemática de nuestra teoría, la que nos permite hacer ciertas predicciones condicionales cuantitativas como una primera prueba de su pertinencia empírica. Finalmente, en la Sección V, integramos el crecimiento de la población así como otros fenómenos complejos en nuestro modelo, e investigamos la noción del esfuerzo crítico mínimo en relación con la duración del proceso de despegue.

I. Los supuestos básicos

Nuestro modelo formal explicativo se presenta con la ayuda de la gráfica 1. La gráfica 1.1 representa el sector industrial y las 1.2 y 1.3 el sector agrícola. La primera es la conocida gráfica de Lewis que mide la fuerza de trabajo industrial en el eje horizontal OW y su productividad marginal física (PMF) en el eje vertical OP. La curva de la demanda de trabajo (*v.gr.* la curva *dtf* de la *PMF*), junto con la curva de la oferta de trabajo

213



GRÁFICA 1

214

215

($Stt'S'$), determina el empleo de la fuerza de trabajo industrial (St). Puesto que la curva de la productividad marginal física depende del volumen del acervo de capital, en cooperación con la fuerza de trabajo, un incremento en el acervo de capital conduce a un desplazamiento de la curva *PMF* hacia la derecha, *v. gr.*, de *dtf* a *d't'f'*. La curva de la oferta de trabajo "ilimitado" de Lewis se define por la porción horizontal de la curva de la oferta, en el ejemplo *St*. Cuando la curva de la oferta vuelve hacia arriba, termina su carácter ilimitado. Nuestro primer problema es investigar las condiciones de este punto de inflexión. Esto nos lleva a concentrar la atención en el sector agrícola.

En la gráfica 1.3 la fuerza de trabajo agrícola se representa en el eje horizontal *OA* (leyendo de derecha a izquierda) y la producción agrícola aparece en el eje vertical *OB* (de la *O* hacia abajo). La curva *ORCX* describe la productividad física total del trabajo (*PFT*) en el sector agrícola. Se supone que esta curva tiene una porción cóncava *ORC* que muestra una productividad marginal del trabajo agrícola que disminuye gradualmente y una porción horizontal *XC* donde desaparece el producto marginal. La porción de cualquier fuerza de trabajo en exceso de *OD* puede considerarse redundante en el sentido de que su retiro de la agricultura no afecta la producción de la misma.

En el punto inicial, supondremos que toda la fuerza de trabajo *OA* está ocupada en la agricultura y produce un volumen agrícola total de *AX*. Supongamos que la producción agrícola *AX* se consume en su totalidad por la fuerza de trabajo agrícola *OA*. Entonces el salario real es igual a *AX/OA* o la pendiente de *OX*. La persistencia de este nivel de salarios está apoyada por fuerzas institucionales independientes del mercado, puesto que bajo supuestos de competencia el salario real descendería a cero, en igualdad con *PMF*. Lo llamaremos salario institucional.

El punto *R* de la curva de producción total es el punto en el que la *PMF* es igual al salario institucional, *v.gr.*, la línea punteada tangencial en *R* es paralela a *OX*. Entonces podemos definir *AP* como la fuerza de trabajo agrícola en condiciones de desocupación disfrazada, puesto que más allá de *P*, *PMF* es menor que el salario institucional.[3]

Obsérvese que las gráficas 1.1, 1.2 y 1.3 se hallan "alineadas". Cualquier punto sobre el eje horizontal de las gráficas 1.1 a 1.3 representa una forma particular en que la población total o la fuerza de trabajo *OA* se distribuye entre los dos sectores; por ejemplo, en el punto *P* (gráficas 1.2 y 1.3) la fuerza de trabajo agrícola es *OP* y la fuerza de trabajo (ya asignada) industrial es *AP*. Si en el punto de partida, la población total *OA* se ocupa en el sector agrícola, el proceso de distribución durante el despegue puede repre-

215

[3] La redundancia es un fenómeno tecnológico, v.gr. determinada por la función producción. Por otra parte, el desempleo disfrazado depende de la función producción, del salario institucional y el monto de la población agrícola. En otras palabras, es un concepto económico.

sentarse por una serie de puntos, A, G, D, I, P, etc., sobre OA, que se mueven gradualmente hacia O.[4]

Los importantes conceptos de desocupación disfrazada, fuerza de trabajo excedente y salario institucional pueden representarse con mayor claridad con ayuda de la gráfica 1.2 en donde la producción agrícola por trabajador se mide en el eje vertical *AN*. Los puntos *ADUV* representan la curva de la productividad marginal física (*PMF*) del trabajo. La distancia vertical *AS* es igual al salario institucional (representado también como *PU*, igual a la *PMF* del trabajo agrícola en *U*, alineado con *P* y *R* en la gráfica 1.3). Pueden ahora distinguirse tres fases en el proceso de redistribución: *1)* Fase 1 es la escala en que $PMF = O$, por ejemplo, el sector en que la curva de la productividad total en la gráfica 1.3 es horizontal. Esta fase señala la fuerza de trabajo excedente, *AD*. *2)* La fase 2 es la escala en la cual una *PMF* positiva es menor que el salario institucional. Las fases 1 y 2 juntas indican la existencia de desocupación disfrazada de la fuerza de trabajo, *AP*. *3)* La fase 3 es la escala en donde la *PMF* es mayor que la tasa institucional de salario que se supone prevalece en el punto de partida.

Suponemos que el salario institucional *AS* prevalece durante las fases 1 y 2 y que una tasa de salarios igual a la *PMF* se encuentra en la fase 3. Sólo cuando la desocupación disfrazada ha sido absorbida, *v.gr.*, en la fase 3, la contribución marginal del trabajo a la producción se hace tan grande o más que el salario real institucional. Como resultado, el terrateniente puede entonces pujar con ventaja por la mano de obra; puede decirse que el sector agrícola se ha comercializado al abandonarse el salario institucional y las fuerzas competitivas del mercado generan las condiciones de equilibrio comúnmente aceptadas. Bajo esos supuestos, el salario agrícola real en términos de bienes agrícolas se define por la curva *SUV* en la gráfica 1.2, consistente de una porción horizontal *SU* y una porción ascendente, *UV*. Ésta puede ser conocida como la curva de oferta precio del trabajo agrícola. Indica, para cada nivel de salario real, el volumen de fuerza de trabajo que puede liberarse del sector agrícola.

La transición a la fase 3 constituye un punto culminante de importancia en el proceso de desarrollo. Al completarse la transferencia de los trabajadores en condiciones de ocupación disfrazada, ocurrirá un cambio, forzado por las circunstancias, en la conducta del patrón, *v.gr.* la aparición de un sector agrícola plenamente comercializado. Esta culminación puede definirse como el fin del proceso de despegue. No conocemos otra forma de restablecer un criterio no arbitrario para una economía que alcanza el umbral del llamado crecimiento autosostenido.[5]

[4] El supuesto actual de una población estática se eliminará después.

[5] Ya sea que el crecimiento pueda realmente ser o no "autosostenido", según la frase de Rostow, básicamente no es un problema dócil a los instrumentos del análisis económico tradicional. El papel de las tasas de ahorro y de los niveles de ingreso *per capita* en la generación de su movimiento permanece indefinido. Todo lo que afirmamos aquí es que, después del punto de inflexión, el salario real

Volviendo ahora a la gráfica 1.3 vemos que, al liberarse a los trabajadores agrícolas, comienza a aparecer un excedente de bienes agrícolas. Esa porción de la producción agrícola total que excede a las necesidades de consumo de la fuerza de trabajo agrícola con un salario institucional, se define como el excedente agrícola total (*EAT*). La cantidad de *EAT* puede considerarse como una función del volumen de mano de obra redistribuida en cada etapa. Por ejemplo, si se retira a los trabajadores agrícolas en la medida de *AG* en la fase 1 y se les redistribuye, se requiere *JG* para alimentar a los agricultores restantes y se obtiene un *EAT* de dimensión *JF*. El *EAT* a cada punto de distribución en las fases 1 y 2 se representa por la distancia vertical entre la línea recta *OX* y la curva de la productividad física total *ORCX*. (Para la fase 3, debido a la elevación de la tasa de salarios, el *EAT* es en cierta medida menor que esta distancia vertical, e iguala la distancia vertical entre la curva *OQ* y la curva de la productividad total.)

El *EAT* puede considerarse como recursos agrícolas librados al mercado a través de la redistribución de los trabajadores agrícolas. Dichos recursos pueden traspasarse mediante las inversiones de la clase terrateniente y/o la política impositiva gubernamental y puede utilizarse en apoyo de las nuevas aportaciones industriales.[6] El excedente agrícola promedio, o *EAP*, puede definirse ahora como el excedente agrícola total disponible por cada uno de los trabajadores industriales asignados.

La curva del *EAP* se representa con la curva *SYZO* en la gráfica 1.2. En la fase 1, al aumentar el *EAT* en forma lineal con la distribución de la fuerza de trabajo excedente de *A* a *D*, podemos representar a cada trabajador asignado como si llevase consigo su propia cesta de subsistencias. La curva del *EAP* de la fase 1 coincide así con la curva *SY* de salarios institucionales. Sin embargo, en la fase 2, puesto que era positiva la *PMF* de los trabajadores agrícolas ahora repartidos, no habrá una producción agrícola suficiente para alimentar a todos los recién llegados a la industria, al nivel institucional de salarios. Así pues, en tanto que todavía aumenta el *EAT*, el *EAP* comienza a declinar.[7] Además, puede verse fácilmente

en la agricultura está determinado por las fuerzas competitivas impersonales del mercado, una transformación cualitativa que constituye una condición necesaria (si no suficiente) para que el crecimiento se haga automático y rutinario. Es este el punto que Lewis parece tener en mente [4, p. 26] cuando habla de "dos diferentes etapas de desarrollo económico con dos conjuntos diferentes de resultados" y cuando describe la segunda etapa como una situación en que "todos los factores de la producción son escasos [y los] . . .salarios dejan de ser constantes al proseguir la acumulación".

6 No obtante que el modelo podría comprenderlos fácilmente, omitimos los costos de transferencia de recursos así como la posibilidad de que pueda ser imposible inducir a los que se quedaron en la agricultura para liberar todo el excedente.

7 Puede tomarse la siguiente analogía con el análisis de la empresa individual para mostrar con claridad la relación entre los conceptos marginal, total y promedio. Podemos considerar la curva del producto agrícola total (*ORCX*) y la curva del consumo agrícola total (*OX*) en la gráfica 1.3 como análogas a la curva del ingreso total y a la curva del costo total, respectivamente. La diferencia entre esas curvas es la curva del beneficio total que es equivalente a nuestra curva del *EAT*. La curva del beneficio total alcanza un máximo cuando el costo marginal iguala al ingreso marginal. Esto ocurre en el punto *U* de la gráfica 1.2 —debido a que *SU* es la curva del costo marginal y *ADUV* es la curva

durante la fase 3 el *EAP* declina aún más rápidamente (y *EAT* también desciende) al hacerse operante el salario agrícola ahora comercializado.

Podemos ahora considerar la derivación del punto de inflexión de Lewis en el sector agrícola. Lewis mismo [4, pp. 19-26] explica el punto de inflexión de un modo más bien libre, el cual ocurre cuando uno de los siguientes sucesos pone fin a la curva horizontal de la oferta de mano de obra: *a)* el deterioro en la relación de intercambio en el sector industrial, y *b)* la terminación del excedente de fuerza de trabajo en el sector agrícola. Pero en nuestro modelo cualquier explicación semejante debe tomar en cuenta la determinación básica de la curva de la oferta de la fuerza total de trabajo por las condiciones postuladas para el sector no industrial.

El "deterioro en la relación de intercambio" del sector industrial ocurre como resultado de una escasez relativa de bienes agrícolas que buscan intercambiarse por bienes industriales en el mercado. En nuestro modelo, se recordará, este excedente se mide por el excedente agrícola total (*EAT*) y sobre una base por trabajador industrial, por excedente agrícola promedio (*EAP*). Entonces aparece una tendencia a la elevación de la curva de la oferta industrial al entrar a la fase 2 porque en esos momentos comienza a sentirse una escasez de bienes agrícolas medidos en *EAP* —que provoca un deterioro en la relación de intercambio del sector industrial y una elevación en el salario industrial real medido en términos de bienes industriales—. Así, vemos que la desaparición del excedente de fuerza de trabajo en el sector agrícola es una causa del punto de inflexión de Lewis.

El "agotamiento del excedente de fuerza de trabajo" debe interpretarse en primer lugar como un fenómeno de mercado más bien que como una escasez física de mano de obra; se indica por un incremento del salario real en el origen de la oferta. Si suponemos que el salario real del trabajador industrial es igual al salario real agrícola,[8] entonces hay una tendencia a la elevación de la curva de la oferta de trabajo (*Stt'S'* en la gráfica 1.1) cuando se llega a la fase 3. Con la desaparición de la desocupación disfrazada de la fuerza de trabajo y la comercialización del sector agrícola, el salario real agrícola comienza a aumentar (véase la gráfica 1.2). Esto conduce a un incremento del nivel del salario real industrial si es que el

del ingreso marginal—. La curva del *EAP* en la gráfica 1.2 es equivalente a una "curva de utilidad media".

8 "Gobernado por" puede ser una descripción más realista. Lewis [3, p. 150] señala que la urbanización, los costos de transferencia, etc. pueden requerir un salario real industrial a un margen o "colina" (que considera aproximadamente del 30 %) constante por arriba del salario institucional de la agricultura; en tanto que, por simplicidad de exposición, nuestro modelo mantiene inicialmente una estricta igualdad entre las dos tasas de salario, más tarde este supuesto se elimina (Sección V). En su segundo artículo [4], Lewis también se refiere a ciertos "factores exógenos", incluyendo la formación de sindicatos y probablemente otros cambios en el medio institucional. Dicha "colina" que crece dinámicamente puede también acomodarse en el modelo, pero no se ha considerado en esta primera aproximación.

218

219

patrón industrial quiere competir con buen éxito con el terrateniente por la utilización de la oferta de trabajo ahora "limitada".

Juntando los dos factores (*a* y *b*), podemos decir que al redistribuir la fuerza de trabajo de la agricultura al sector industrial, la curva de la oferta industrial sube (*v.gr.*, aparece el punto de inflexión de Lewis), en el primer caso (en *t*), debido a una escasez de bienes agrícolas atribuible a la desaparición de la fuerza de trabajo agrícola excedente; y que esta tendencia ascendente del salario real industrial se acentúa más tarde (en X′) por el movimiento hacia arriba del salario real agrícola que se explica por la desaparición completa del desempleo disfrazado de la fuerza de trabajo y la comercialización del sector agrícola.

Para facilitar nuestro análisis posterior, vamos a tratar el límite entre las fases 1 y 2 (*v.gr.*, punto Y de la gráfica 1.2) como el "punto de escasez" que significa el comienzo de escasez de productos agrícolas como lo indica el hecho de que *EAP* desciende por abajo del salario mínimo; vamos también a referirnos al límite entre las fases 2 y 3 como el "punto de comercialización" que significa el comienzo de la igualdad entre la productividad marginal y el salario real en la agricultura. El punto de inflexión de Lewis coincide así con el punto de escasez y el movimiento ascendente del salario real industrial se acentúa en el punto de comercialización.[9]

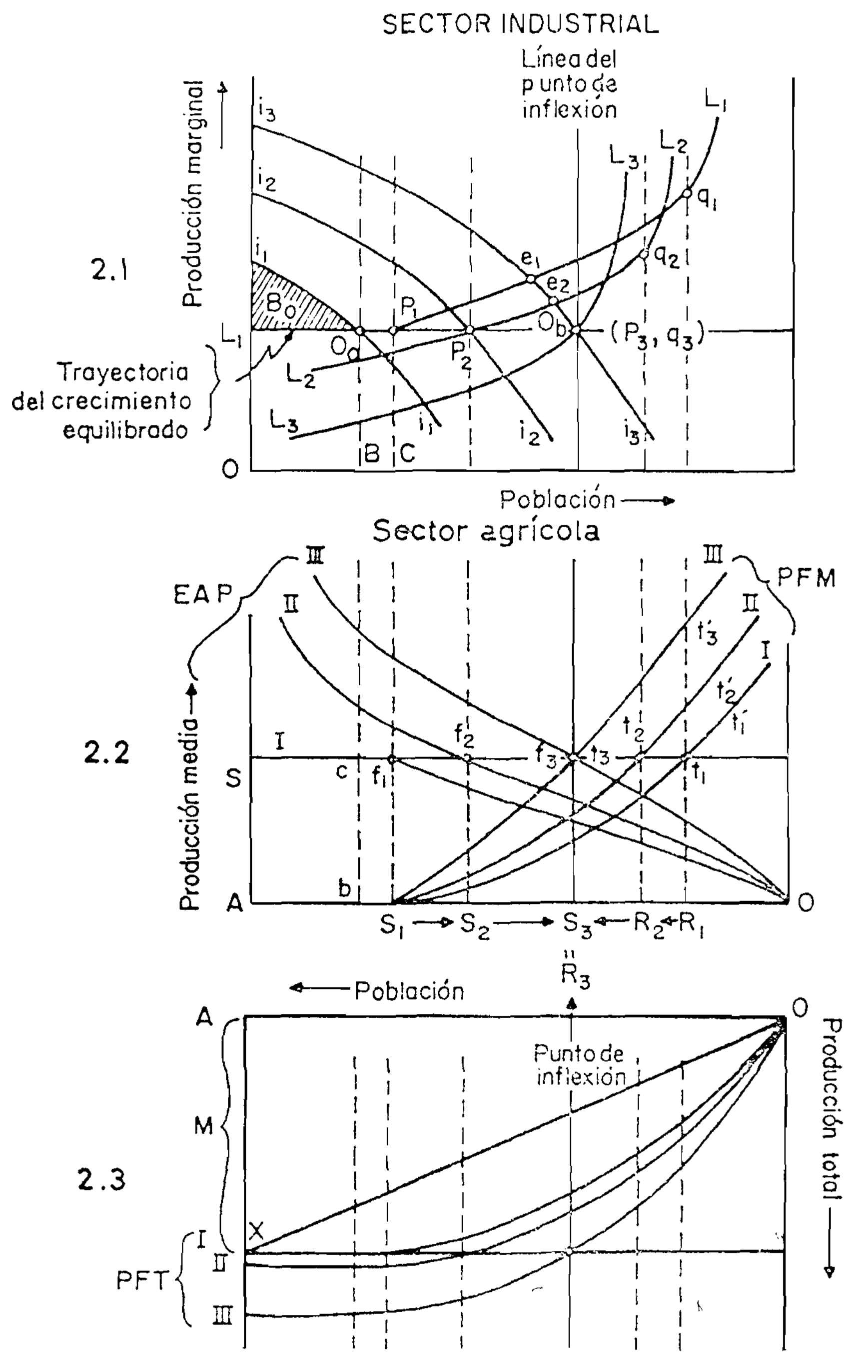
Existen dos factores que pueden contribuir a posponer el punto de inflexión de Lewis: 1) incrementos de la productividad agrícola, y 2) crecimiento de la población. Ambos factores operan en forma muy diferente: el primero se considera en general como una bendición, ya que eleva la producción agrícola excedente; el segundo, se considera casi invariablemente como una maldición, puesto que aumenta la oferta de mano de obra excedente. En primer lugar examinaremos la importancia de un incremento de la productividad agrícola. El análisis sobre el crecimiento de la población se hará más tarde.

II. Cambios en la productividad agrícola

El aumento de la productividad del trabajo en el sector agrícola puede describirse por un desplazamiento "hacia arriba" de la curva de la productividad física total (*PFT*) de la gráfica 1.3. Esos incrementos de la productividad se representan en la gráfica 2.3 por una secuencia de las curvas *PFT* marcadas I,II,III. . . , etc., entre las cuales la curva I es la curva inicial *PFT* (como en la gráfica 1.3) y II,III. . . representan las curvas *PFT*

[9] Desde un punto de vista estrictamente lógico, la curva de la oferta de trabajo industrial debe derivarse de la totalidad de condiciones que surgen de nuestro análisis del sector agrícola. Las condiciones importantes incluyen: *1)* la curva del salario real agrícola, *2)* la curva del *EAP*, y *3)* un cuadro de las preferencias de los consumidores especificando las preferencias por bienes agrícolas vs. bienes industriales. *Las limitaciones de espacio nos impiden llevar a cabo una derivación rigurosa del salario* real industrial en cada punto a través del mecanismo de la relación de intercambio.

219



SECTOR INDUSTRIAL
Línea del punto de inflexión
2.1
Producción marginal
Trayectoria del crecimiento equilibrado
Población
Sector agrícola
EAP
PFM
2.2
Producción media
$S_1 \to S_2 \to S_3 \leftarrow R_2 \leftarrow R_1$
R_3
Población
Punto de inflexión
2.3
PFT
Producción total

GRÁFICA 2

221

después de porciones sucesivas de inversión agrícola. (Por el momento no suponemos cambios en la productividad industrial.)

Supongamos que al aumentar la productividad agrícola el salario institucional permanece inalterado; *v.gr.* en la gráfica 2.2, *SA* es igual a la pendiente de *OX* en las gráficas 1.3 y 2.3 que están determinadas por la curva inicial *PFT*.[10] Podemos ahora seguir en la gráfica 2.2 la secuencia de las curvas I,II,III (todas contienen la porción plana AS_1) de la productividad marginal física del trabajo y la secuencia de las curvas I,II,III del excedente agrícola promedio que corresponden a las curvas I,II,III de la productividad física total en la gráfica 2.3. De acuerdo con el método ya indicado, podemos determinar las tres fases para cada nivel de productividad, *v.gr.*, la secuencia de puntos de escasez, $S_1, S_2, S_3 \ldots$ y la secuencia de los puntos de comercialización, R_1, R_2, $R_3, \ldots$ La referencia a esos puntos facilitará nuestro análisis de los efectos de un incremento en la productividad agrícola en la curva del precio de la oferta del trabajo agrícola y en la curva del *EAP*.

Según la gráfica 2.2, por cada cantidad de trabajo empleado en el sector agrícola, un incremento de la productividad agrícola también desplaza hacia arriba la curva de la productividad física marginal.[11] Como consecuencia, la curva del precio de la oferta del trabajo agrícola se transformará de $St_1t'_1$ a $St_2t'_2$ a $St_3t'_3 \ldots$ etc. con un acortamiento de su porción horizontal (*v.gr.*, la fase 3 llega primero) como la secuencia de los puntos de comercialización R_1, R_2, $R_3 \ldots$ se desplaza gradualmente de derecha a izquierda. Por otra parte, la secuencia de los puntos de escasez S_1, S_2, $S_3 \ldots$ etcétera se mueve gradualmente de izquierda a derecha. Esto se debe al hecho de que, por cada cantidad de trabajo asignado al sector industrial, el *EAP* aumenta con el incremento de la productividad física total; la cantidad de alimentos consumida por el trabajo agrícola permanece inalterada, dejando más *EAT* (y en consecuencia *EAP*) a los trabajadores industriales. Así, el efecto de nuestro incremento de la productividad agrícola es un desplazamiento hacia arriba de la curva *EAP* (a las posiciones *II,III*. . .).

Tarde o temprano, el punto de escasez y el punto de comercialización coinciden, la distancia S_1R_1, S_2R_2, $S_3R_3 \ldots$ se desvanece y se elimina la fase 2. En la gráfica 2.2 ese punto de coincidencia se representa por $R_3=S_3$. Éste será el punto de inflexión. Allí existe un nivel de productividad agrícola que, si se alcanza, producirá este punto de inflexión. (En la gráfica 2.3

10 Por supuesto, es posible permitir el aumento del salario agrícola determinado institucionalmente; pero como la economía se hace más capitalista parece muy dudoso que las fuerzas que no pertenecen al mercado en la agricultura sean reforzadas y en esa forma prevenir el cierre de la laguna del salario artificial de la productividad marginal. Una segunda salvedad, quizás más poderosa, surge del hecho de que el nivel del salario institucional de la agricultura puede ser lo bastante cercano al de subsistencia calórica de manera que su elevación puede constituir una forma muy productiva de inversión. Sin embargo, no consideramos esta posibilidad en el contexto del presente modelo. Con respecto a la posición relativa del nivel del salario industrial, véase nota 8.

11 Esto es un supuesto razonable si el desplazamiento en *PFT* es proporcional.

221

este nivel de productividad agrícola se representa por la curva III de la *PFT*.)

Investiguemos ahora el efecto de un incremento de la productividad agrícola en la curva L_1L_1 de la oferta industrial representada en la gráfica 2.1. Por una parte, el desplazamiento hacia arriba de la curva *EAP* moverá hacia abajo la curva de la oferta industrial *enfrente* del punto de inflexión. Esto se debe al hecho de que un incremento del *EAP* abatirá la relación de intercambio del sector agrícola y, con el mismo salario institucional (en términos de bienes agrícolas) pagado a los trabajadores industriales, el salario industrial (en términos de bienes industriales) debe declinar. Por la otra, el desplazamiento hacia arriba de la curva *PMF* que es acompañado por un mayor salario real en el sector agrícola *después* del punto de inflexión, eleva la curva de la oferta industrial después de ese punto. Así vemos, por ejemplo, que la curva L_2L_2 cruza la curva L_1L_1 por abajo, indicando que en última instancia el "efecto de la relación de intercambio" (debida a un incremento del *EAP*) ha sido vencido por el "efecto del salario real" (debido a un incremento de la *PFM*). Para los propósitos de este trabajo, sin embargo, no nos preocupa mucho la fase 3 que se encuentra más allá del punto de inflexión.

Examinemos ahora más de cerca las posiciones relativas de las curvas de la oferta industrial antes de llegar a la fase 3. Permítase que la porción horizontal L_1L_1 de la curva inicial L_1L_1 de la oferta industrial se extienda al punto de inflexión P_3, y llámese trayectoria del crecimiento equilibrado a esta línea horizontal segmento L_1P_3 (por razones que se explicarán ampliamente en la sección siguiente). Podremos entonces pretender que todas las curvas de la oferta industrial entre L_1L_1, es decir, la inicial, y L_3L_3 o sea la que corresponde al punto de inflexión, crucen la trayectoria del crecimiento equilibrado en los puntos de escasez respectivos. Esto se debe al hecho de que en el punto de escasez de cada caso (*v.gr.*, punto f_2 en la gráfica 2.2 en el caso de la curva de la oferta industrial L_2L_2 en la gráfica 2.1) la tasa de salarios de subsistencia y el *EAP* toman el mismo valor que el que prevalecía en la fase 1 antes de que se hubiese registrado cualquier incremento de la productividad agrícola. En consecuencia, debe prevalecer el mismo salario real, en términos de bienes industriales, en el punto de escasez que existía previamente. En suma, antes del punto de inflexión, la curva de la oferta de trabajo industrial se encuentra arriba (abajo) de la trayectoria del crecimiento equilibrado cuando la curva *EAP* yace debajo (arriba) de la línea horizontal *Sa*, causando un deterioro (mejoría) de la relación de intercambio del sector industrial.

La importancia económica de la igualdad entre nuestro punto de inflexión y el punto (final) de escasez es que, antes del punto de inflexión, la economía se mueve a lo largo de su trayectoria de crecimiento equilibrado en tanto que explota (o aprovecha de la mejor manera) su fuerza 222 de trabajo agrícola subocupada por medio de los incrementos de la productividad agrícola. La importancia económica de la igualdad entre nuestro punto de inflexión y el punto de comercialización es que, después del punto de inflexión, la curva de la oferta de trabajo industrial se eleva finalmente cuando entramos a un mundo donde el sector agrícola deja de estar dominado por fuerzas institucionales que no pertenecen al mercado y toma las características de un sistema capitalista comercializado.

III. Cambios en la productividad industrial y crecimiento equilibrado

Además de la inversión en el sector agrícola, el otro aspecto de importancia del crecimiento que debe considerarse es el proceso simultáneo de inversión en el sector industrial. Conocemos, por añadidura, que tales actividades en los dos sectores no constituyen actividades independientes. Esto es así porque desde el punto de vista de la producción, los dos sectores deben proporcionar los mercados para los productos de cada uno; y, desde el punto de vista del insumo, el sector industrial debe proporcionar las oportunidades de empleo para absorber a los trabajadores liberados por el sector agrícola. La consideración de esta interdependencia básica durante el proceso de despegue no es otra cosa sino la consideración del problema del "crecimiento equilibrado", un concepto clave en la literatura del desarrollo.[12] El propósito de esta sección es formular rigurosamente el problema del crecimiento equilibrado e investigar su importancia en el contexto de nuestro modelo.

Con referencia a la gráfica 2.1 vemos que durante el proceso de despegue la curva de la demanda de fuerza de trabajo, i_1i_1, i_2i_2... se desplaza gradualmente hacia arriba a la derecha al acumularse el capital real en el sector industrial. En forma simultánea la inversión que ocurre en el sector agrícola mueve la curva de la oferta de fuerza de trabajo L_1L_1, L_2L_2... hacia abajo en la misma dirección. El problema central del crecimiento equilibrado comprende la sincronización a través del tiempo de los cambios de las dos secuencias de curvas. En cualquier momento del proceso de despegue, la cuestión decisiva es decidir cómo debe distribuirse el fondo total de inversión entre los dos sectores, para asegurar que sean "armoniosos" desde el punto de vista del criterio del insumo y de la producción.

El criterio de producción, es decir, la existencia de mercados mutuos, especifica que la asignación de fondos de inversión debe hacerse en tal forma que sostenga continuamente los incentivos de inversión en ambos sectores de la economía. En el contexto de nuestro modelo, esto significa 223

[12] Véase especialmente R. Nurkse [5] y [6, p. 192]: "Sin la reorganización [agrícola] el excedente de fuerza de trabajo en la agricultura se mantiene en gran parte en estado de potencia. Por otro lado, la reorganización puede ser impracticable sin una política activa que absorbe la mano de obra excedente."

224

que la relación de intercambio entre los dos sectores no debe deteriorarse de modo importante en contra de ninguno de los sectores. El criterio del insumo, por otro lado, especifica que la distribución del fondo de inversión debe capacitar al sector industrial a demandar, con el salario real industrial constante de acuerdo con el criterio de producción, el número preciso de trabajadores ahora liberados, como un resultado de la actividad de inversión en el sector agrícola. Ahora procederemos a mostrar que existe una trayectoria de crecimiento equilibrado que satisface esas condiciones, como parte integral de nuestro modelo.

La curva de la demanda inicial de mano de obra industrial en el punto de partida se indica por i_1i_1 y la curva de la oferta inicial por L_1L_1 en la gráfica 2.1, con las unidades OB de trabajo ya empleado en el sector industrial. (En tanto que es factible suponer que existe ya algún establecimiento industrial durante el periodo de preacondicionamiento y se hereda en el comienzo del proceso de despegue, también es factible suponer que la fuerza inicial de trabajo industrial OB es muy pequeña.) En este nivel de ocupación el sector industrial obtiene una utilidad representada por el área sombreada B_o (gráfica 2.1) que puede tomarse como representante del fondo de inversión de la economía en esta etapa.[13] Este fondo de inversión se va a asignar en parte al sector agrícola, elevando así la productividad agrícola y desplazando la curva de la oferta industrial a la derecha, y en parte al sector industrial, aumentando así el acervo de capital industrial y moviendo hacia la derecha la curva de la demanda industrial.

Si se pretende satisfacer el criterio del crecimiento equilibrado, la nueva curva de la demanda industrial, *v.gr.*, i_2i_2, y la nueva curva de la oferta industrial, *v.gr.* L_2L_2, deben interceptarse en un punto, *v.gr.*, P_2, que se encuentra en la trayectoria del crecimiento equilibrado (L_1P_3). De otra manera se viola la condición de estabilidad de la relación de intercambio. En P_2, donde se satisface el criterio de crecimiento equilibrado, el sector industrial habrá absorbido O_0P_2 obreros adicionales, que es el mismo número de trabajadores liberado por el sector agrícola (es decir, cf_2 en la gráfica 2.2 es igual a O_0P_2 en la gráfica 2.1).

Así, a medida que tiene lugar el proceso de inversión en ambos sectores, la trayectoria del crecimiento equilibrado describe la trayectoria del crecimiento real si el criterio del crecimiento equilibrado es satisfecho. Por supuesto, es probable que de vez en cuando el curso del crecimiento real se desvíe del curso de crecimiento equilibrado en una dirección o en otra. Sin embargo, tal desviación pondrá en juego fuerzas equilibradoras contrarrestantes que tienden a restablecerlo en la trayectoria del crecimiento

224

13 Si es que puede pasarse por alto el consumo capitalista. Debe notarse que el sector agrícola (gráfica 2.2) no contribuye al fondo de inversión puesto que la producción agrícola total (área OaSA) es justamente adecuada para satisfacer las necesidades de consumo de los trabajadores agrícolas (área Obca) y las necesidades de consumo de los trabajadores industriales (área ASeb).

225

equilibrado. De hecho, es probable que el curso del crecimiento real oscile alrededor de la trayectoria del crecimiento equilibrado.

Por ejemplo, si el curso del crecimiento real se encuentra arriba del curso de crecimiento equilibrado, digamos en e_2 en la gráfica 2.1 (como sería el caso si la inversión en el sector agrícola hubiera movido la curva de la oferta industrial a L_2L_2 y la inversión en el sector industrial hubiera desplazado la curva de la demanda industrial a i_3i_3), tenemos entonces un caso de sobreinversión en el sector industrial. La escasez de alimentos producirá un deterioro en la relación de intercambio del sector industrial y provocará un incremento del salario real industrial. Esta situación tenderá a desalentar la inversión en el sector industrial, y a estimular la misma en el agrícola, impulsando así el curso del crecimiento real a volver hacia la trayectoria del crecimiento equilibrado. Puede suponerse que la política gubernamental seguirá la misma dirección si el sistema de precios resulta inadecuado. En esta forma, la economía que se mueve sobre el curso de crecimiento real que coincide u oscila alrededor de la trayectoria del crecimiento equilibrado, se dirige hacia el punto de inflexión, P_3, definido previamente.[14]

IV. Importancia empírica del modelo básico

Con objeto de formular con mayor rigor nuestro modelo y hacerlo susceptible de verificación estadística, deben ahora aceptarse ciertos supuestos restrictivos no necesarios en nuestro análisis cualitativo previo. El primer supuesto, de que la productividad física marginal del trabajo cambia a una tasa constante, al variar la ocupación en el sector agrícola, se ocupa de la forma de la curva inicial de la productividad física total. Esto significa que la curva inicial *PFM* (I en la gráfica 2.2) se compone de dos segmentos en línea recta: un segmento horizontal, AS_1, que coincide con el eje horizontal, y un segmento $S_1t'_1$, para el orden de la productividad física marginal positiva. Los dos segmentos se unen en el punto S_1 que marca la fuerza de trabajo agrícola excedente (AS_1 en la gráfica 2.2). Bajo esos supuestos, puede mostrarse (véase el Apéndice) que la curva *PFT* toma la forma siguiente:

$$(1) \qquad y = \begin{cases} M\left[\left(-\dfrac{x}{TL}\right) + \left(\dfrac{x}{TL}\right)\right] & \text{para } x < TL \\ M & \text{para } x > TL \end{cases}$$

[14] La porción "ilimitada" de la curva de la oferta de trabajo de Lewis puede así interpretarse como una curva de oferta *ex post* definida como el centro de todos los puntos en nuestro curso de crecimiento equilibrado en condiciones de incrementos continuos de la productividad agrícola. Sin embargo, ni Lewis ni nosotros debemos descartar la posibilidad de que el curso de crecimiento real pueda de hecho ascender suavemente, más bien que permanecer horizontal. Ese curso de crecimiento implicaría el ascenso gradual de los niveles de salario real industrial durante el periodo de despegue. (Véase también la nota 8.)

225

donde las variables x y y, y los parámetros M, T y L tienen la siguiente intepretación económica y diagramática (gráfica 2.3): *i)* $y =$ producción agrícola total (medida hacia abajo desde el punto O); *ii)* $x =$ fuerza de trabajo empleada en el sector agrícola (medida hacia la izquierda del punto O); *iii)* $M =$ producción agrícola máxima (la distancia AI); *iv)* $L =$ monto de la población en el punto de partida (la distancia OA); *v)* $T =$ la fracción de L que no es excedente, *v.gr.* TL es la fuerza de trabajo no excedente (la distancia OS_1 en la gráfica 2.2) y $(1-T)L$ es la fuerza de trabajo excedente (la distancia S_1A). El parámetro T o coeficiente del no excedente puede tomar cualquier valor no negativo. Si T es menor que 1, $(1-T)L$ es la fuerza de trabajo excedente en el punto de partida. Si T es mayor que 1, $(T-1)L$ es la adición a la fuerza de trabajo agrícola L que podría tolerarse antes de que cualquier porción de la fuerza de trabajo agrícola se haga excedente, *v.gr.*, de productividad física marginal igual a cero. El caso de T menor que 1 se representa en la gráfica 2.3.[15]

Nuestro segundo supuesto restrictivo es que un incremento de la productividad agrícola desplaza proporcionalmente "hacia arriba" toda la curva de la *PFT*. En otras palabras, la nueva curva *PFT* se obtiene multiplicando la curva inicial *PFT* por una constante k que se llamará coeficiente de productividad. Al aumentar sucesivamente el valor de éste, se genera una secuencia de curvas *PFT* (II,III, etc.) según se muestra en la gráfica 2.3.[16]

De las curvas de la *PFT* podemos fácilmente derivar expresiones para el salario institucional, las curvas de la productividad física marginal (*PFM*), y las curvas promedio del excedente agrícola (*EAP*):

$$W = M/L \tag{2}$$

(salario agrícola representado por la distancia AS en la gráfica 2.2 o la pendiente de OX en la gráfica 2.3)

$$y' = \frac{2kM}{(TL)^2} \quad (-x + TL) \tag{3}$$

(curva de la productividad física marginal para la fuerza de trabajo agrícola no excedente en la gráfica 2.2; $O < T \leq 1$)

$$EAP = \frac{ky - xW}{L - x} \tag{4}$$

(curva del excedente agrícola promedio)

15 Hay quienes creen, *v.gr.* Harry Oshima [8, p. 259] que la *PFM* de la mano de obra agrícola en un área subdesarrollada nunca desciende realmente a cero. Esta posición está representada por el segundo caso, esto es, $T > 1$, porque probablemente nadie negará que, con una cantidad fija de tierra, habrá un cierto monto de población agrícola lo bastante grande para reducir la *PFM* igual a cero. En tanto que ambos casos se tratan sistemáticamente en el apéndice, por razones de facilidad en la exposición sólo presentamos el caso de $O < T \leq 1$ en el texto. Sin embargo, las condiciones para ambos casos se incorporan en este trabajo.

16 Nótese que bajo esos supuestos todas las curvas *PFM* contienen el mismo segmento horizontal AS_1.

226

Esas variables son funciones de x (esto es, la fuerza de trabajo agrícola), con M, T y L como parámetros y k el coeficiente de productividad exógena.

El objetivo principal de nuestro modelo es derivar una expresión para *PFTA*, el punto de inflexión de la fuerza de trabajo agrícola, representado por la distancia OS_3 en la gráfica 2.2. *PFTA* es una fracción, V_t de la población total L, esto es, $PFTA = V_t L$. Resolviendo el valor de k en el punto de inflexión, puede derivarse de (1) a (4) la siguiente expresión V_t:[17]

$$V_t = 1 + T - \sqrt{1 + T^2} \tag{5}$$

Este porcentaje de la población agrícola en el punto de inflexión (V_t) depende sólo de T, el coeficiente del no excedente. Desde el punto de vista económico, esto significa que nuestro modelo es independiente de la dimensión (esto es, la escala) de la economía (descrita por el monto absoluto de población, L, o la cantidad absoluta de producción agrícola inicial, M).

Para someter nuestro modelo a su primera prueba de pertinencia empírica, examinemos (cuadro 1) los valores de V_t en un orden de valores de T (de .7 a 3) que representa un espectro razonable que cubre la mayoría de los países. Una pequeña T, o un pequeño coeficiente de no excedente, significa que un país se halla inicialmente mal dotado de recursos naturales, esto es, que existe una baja relación trabajo-tierra. Aunque son escasas las estimaciones precisas, la mayor parte de los observadores interesados están de acuerdo en que la fuerza de trabajo excedente puede alcanzar el 30 % en las regiones densamente pobladas de Asia, *v.gr.*, Pakistán, India, Ceilán. El coeficiente de no excedente de $T = .7$ representa así el país con la dotación inicial de recursos más desfavorable. En el otro extremo del espectro están ciertos países occidentales, posiblemente Dinamarca,

Cuadro 1

T	.7	.8	.9	1	1.1	1.2	1.3	1.4	2	3
V_t	.48	.52	.55	.58	.61	.64	.66	.68	.75	.80

que han completado ya su proceso de despegue. Por supuesto, se tienen aún menores conocimientos estadísticos del coeficiente del no excedente para cualquier país en el punto relevante de su historia; hemos escogido más o menos arbitrariamente un valor superior de $T = 3$, aunque de ninguna manera nos comprometemos con dicha cifra.[18]

227

[17] Según se muestra en el apéndice.

[18] Nótese que V_t se acerca a 1 al aproximarse T al infinito en forma tal que el valor de V_t no es muy sensitivo al cambio en T cuando ésta se hace más grande. En consecuencia, no debemos preocuparnos demasiado con el límite superior del orden de valores postulado en el cuadro 1. Incidentalmente, una gran T no debe confundirse con la posibilidad de que la producción primaria, digamos en Austria, pueda siempre organizarse sobre la base de la plantación y, por lo tanto, sin que forme parte jamás del sector "agrícola", según lo hemos definido (nota 2). Como se dijo antes, nuestro modelo no es pertinente cuando la economía entera se encuentra comercializada desde el principio.

En este orden razonable de valores para T, los valores correspondientes para V_t se extienden aproximadamente del 50 al 80 %. Esto significa que al final del proceso de despegue nuestro modelo "predice" que del 20 al 50 % de la fuerza de trabajo total debe haberse distribuido en el sector industrial. Nociones aceptadas con respecto a esas magnitudes sugieren que nuestros resultados son también razonables.

En este cuadro podemos ver también que el valor de V_t aumenta con el incremento del valor de T, una relación en general válida que puede establecerse fácilmente obteniendo la primera derivada de (5). La interpretación económica de esta relación es clara: mientras mayor sea el coeficiente del no excedente, más favorable será (relativamente) la dotación inicial de recursos; y mientras más favorable sea esta dotación, más probable será que la economía esté orientada hacia la agricultura (medida por un valor relativamente grande de V_t) en el punto de inflexión. Por el contrario, mientras más pequeño sea el coeficiente del no excedente, más desfavorable será la dotación inicial de recursos y más probable será que la economía se oriente hacia la industria (medida por un valor relativamente pequeño de V_t) en el momento de completarse el proceso de despegue.[10] En el primer caso (orientación agrícola), asociado con algunas economías más avanzadas, nuestra teoría "predice" entonces un punto de inflexión hacia arriba de la fuerza de trabajo agrícola del 65 % (cuando T es mayor que 1.2). En el último caso (orientación industrial), asociado con los países subdesarrollados contemporáneos de Asia, nuestra teoría "predice" un punto de inflexión hacia abajo de la fuerza de trabajo agrícola en un 55 % (cuando T es menor que .9). Es evidente que, si el proceso de despegue se ha de completar con buen éxito, los países con recursos pobres que son los que nos interesan principalmente, tendrán que redistribuir un porcentaje mayor de su fuerza de trabajo total a la industria que lo que asignaron algunos de los países occidentales mejor dotados. Y esta tarea ya difícil se complica aún más por el hecho de que estos países están por lo regular sometidos a una severa presión demográfica en esta etapa. Procedemos ahora a integrar en nuestro modelo esta importante faceta del problema del desarrollo.

V. El crecimiento de la población y el esfuerzo mínimo

Supongamos que, en el curso del despegue, la economía experimenta un incremento demográfico del 100 %. Indicaremos con L_t la población total en el punto de inflexión. Entonces

$$L_t = (1 + s)L \tag{6}$$

228

[10] Puesto que, en nuestro sistema, sólo el sector comercializado se halla en posición de obtener utilidades y de ahorrar, la conclusión es congruente con la predicción de Lewis [4, p. 27] de que "los márgenes de utilidades serán menores en los países que alcanzan con mayor rapidez su segunda etapa (punto de inflexión), y serán más elevados en los países donde la segunda etapa se retarda más tiempo".

donde L es el monto de la población en el punto de partida. Con tal incremento de población la función del excedente agrícola promedio (EAP) se convierte:

$$\text{EAP} = \frac{ky - xW}{L(1+s) - x} \tag{7}$$

Cuando esta ecuación se emplea en lugar de (4), podemos derivar la siguiente expresión:

$$V_t = 1 + \frac{T}{1+s} - \sqrt{1 + \left(\frac{T}{1+s}\right)^2} \tag{8}$$

donde V_t es el punto de inflexión de la fuerza de trabajo agrícola ($PFTA$) expresada como una fracción de L_t, esto es, $PFTA = V_t L_t$. (En otras palabras, cuando hay un incremento de población, utilizamos la población total L_t en el punto de inflexión, más bien que la del punto de partida como la base para computar la fracción $PFTA$.)

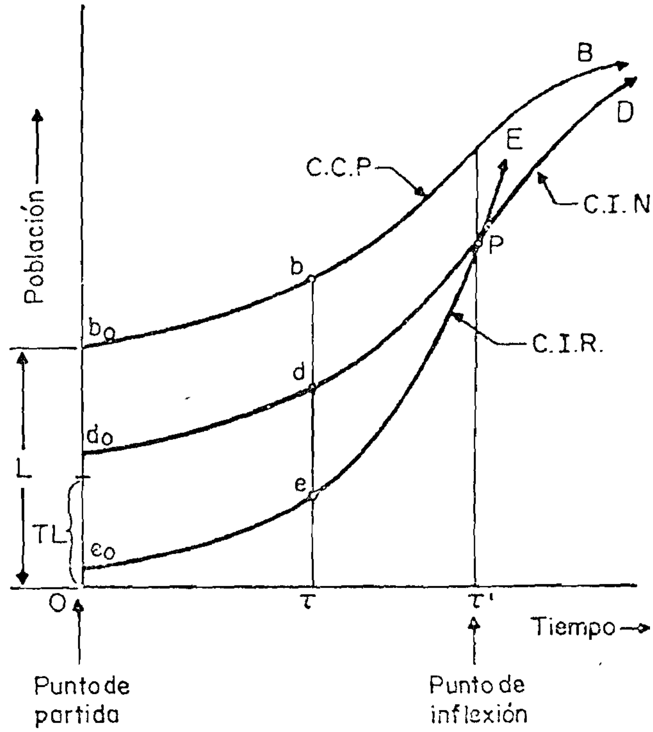
Comparando (5) y (8), vemos que nuestro análisis en la última sección, suponiendo que no hay crecimiento de población, se reduce ahora a un caso especial. Además, por lo que se refiere al efecto en V_t, el incremento de población es equivalente a un descenso en el valor del coeficiente del no excedente T. Esto destaca el hecho de que ambos fenómenos constituyen un empeoramiento de la base de recursos de la economía. Se infiere que, para un valor dado de T, mientras mayor sea el incremento de población (esto es, una s mayor) menor será el valor de V_t y en consecuencia mayor será la orientación industrial de la economía en el punto de inflexión.

La importancia de la expresión (8) puede investigarse ahora, más minuciosamente. En la gráfica 3, el tiempo se mide en el eje horizontal y la población en el vertical. La población L inicial o del punto de partida se representa por la distancia Ob_0 en el año O y el crecimiento de población a través del tiempo se describe por la curva b_0bB, que llamaremos curva del crecimiento de población (CCP). El crecimiento de población se tratará como un fenómeno conocido exógeno a nuestro modelo.

Al crecer la población, el sector industrial obviamente tendrá que absorber más fuerza de trabajo en el momento en que se alcance el punto de inflexión. De hecho, el sector industrial tendrá que absorber no sólo más fuerza de trabajo en términos absolutos sino también un porcentaje mayor del monto mayor de población total. Podemos entonces hacernos la siguiente pregunta hipotética: ¿Cuál será la dimensión absoluta de la fuerza de trabajo industrial, L_{ti}, y de la fuerza de trabajo agrícola, L_{ta}, en el punto de inflexión, si el proceso de despegue se completa en τ años? El punto de inflexión de la población total L_t, se representa por la distancia $b\tau$. Puesto que, para una τ dada, la curva del crecimiento de la población nos da los valores para L y L_t, podemos determinar inmediatamente el factor múltiple $1 + s(\tau)$ en (6). [Nótese que s se expresa ahora como una función $s(\tau)$ de τ.] Cuando el valor de $1 + s(\tau)$ se sustituye en (8), obtenemos:

$$(9) \qquad V_t(\tau) = 1 + \frac{T}{1 + s(\tau)} - \sqrt{1 + \left(\frac{T}{1 + s(\tau)}\right)^2}$$

como la fracción de la población total en el sector agrícola en el punto de inflexión. Ahora se expresa como una función de τ el periodo de tiempo



GRÁFICA 3

específico para completar el proceso de despegue, tratando T, el coeficiente del no excedente, como parámetro. De esta ecuación podemos determinar fácilmente la dimensión absoluta del punto de inflexión de la fuerza de trabajo industrial L_{ti} y del punto de inflexión de la mano de obra agrícola L_{ta} como función de τ.

$$(10) \qquad \begin{aligned} (a)\ & L_{ti} = [1 - V_t(\tau)] \quad [1 + s(\tau)]\, L \\ (b)\ & L_{ta} = V_t(\tau)\, [1 + s(\tau)]\, L \end{aligned}$$

donde $V_t(\tau)$ se define en (9).

La curva correspondiente a (10a), esto es, d_0dD, aparece en la gráfica 3. A esta curva la llamaremos de la industrialización necesaria (*CIN*).

230

La distancia vertical entre la *CIN* y la *CCP* se representa por (10*b*). La *CIN* indica el monto absoluto de población que debe absorber el sector industrial si el punto de inflexión se presenta en el momento indicado sobre el eje horizontal. Como podemos ver directamente en la ecuación (9), el valor de $V_t(\tau)$ se acerca a cero al aumentar τ. Esto significa que *CIN* se inclina hacia la *CCP* al alargarse el tiempo necesario para el despegue. La importancia económica de este fenómeno es que mientras más tardado sea alcanzar el punto de inflexión, esto es, más tiempo tendrá de establecerse el enemigo malthusiano, y más pesada será la carga del sector industrial en términos de la absorción de los trabajadores agrícolas necesarios. La *CIN* indica las necesidades de absorción total para cada τ o duración del proceso de despegue.

Este importante concepto de la curva de la industrialización necesaria puede interpretarse en términos de la tesis del esfuerzo mínimo crítico. Esto quiere decir que, por cada valor de τ, debe llevarse a cabo una cierta actividad de inversión mínima tanto en el sector industrial como en el agrícola durante cada año del proceso de despegue, del año 0 al año τ. Porque como hemos visto, la inversión en el sector industrial debe ser adecuada para proporcionar oportunidades de empleo a la mayor fuerza de trabajo industrial, y la inversión en el sector agrícola debe ser suficiente para incrementar la productividad agrícola lo bastante para alimentar la población creciente a la vista de una posible reducción de la mano de obra agrícola. Así, ya sea que el proceso de despegue pueda de hecho completarse en τ años, depende de que el esfuerzo necesario se realice durante los años que vienen.

Para aclarar más este punto, postulemos en yuxtaposición con la curva de la industrialización necesaria (*CIN*) arriba descrita, una curva de la industrialización real (*CIR*), que muestra la cantidad de trabajo que realmente absorbe la industria en cada punto a través del tiempo. La ecuación de esta curva se escribe:

$$E_i = \phi(t) \tag{11}$$

en donde t es igual a tiempo y E_i al monto real de la fuerza de trabajo industrial en el tiempo t. Esta curva se denota por c_0eE en la gráfica 3. En el momento τ, por ejemplo, de la fuerza de trabajo total o de la población $b\tau$ la cantidad que absorbe *realmente* el sector industrial es igual a $c\tau$. Al mismo tiempo, como hemos visto, la cantidad de trabajo que *necesita* haber sido asignada a este sector es $d\tau$ si el punto de inflexión se presenta en este momento. En consecuencia, en este caso, es imposible alcanzar el punto de inflexión en el tiempo τ. De aquí se desprende que el proceso de despegue puede completarse con buen éxito sólo si la *CIR* y la *CIN* se cruzan, *v.gr.*, en el punto P, después de τ años.

La posición de *CIR* depende entonces del esfuerzo nacional, medido

231

en términos de los gastos de inversión en ambos sectores, que se realizan en el curso del proceso de despegue. Con un mayor esfuerzo nacional la pendiente de la *CIR* es mayor e intercepta *CIN* en una fecha anterior, esto es, una τ más pequeña. En forma recíproca, con un esfuerzo nacional más pequeño, *CIR* se eleva más lentamente y cruza *CIN* en una fecha posterior; alternativamente, no llega nunca a cruzarla.

Para investigar este problema, supongamos, por simplicidad, que la industria absorbe realmente la mano de obra a una tasa anual constante i. *CIR* en (11) toma entonces la siguiente forma concreta:

$$E_t = L(1 - V)e^{it} \qquad (13)$$

donde i, la tasa de crecimiento de la fuerza de trabajo industrial, puede tomarse como un índice resumido del esfuerzo nacional.[20]

Con objeto de poder estimar la duración del proceso de despegue con la ayuda de nuestro modelo, supongamos que la población crece a la tasa constante r. La *CCP* se representa entonces por

$$P = Le^{rt} \qquad (14)$$

y, por esta *CCP* particular, la *CIN* en (10*a*) se convierte en

$$L_{tt} = [1 - V_{t(\tau)}]\, e^{rt}L \qquad (15)$$

donde, utilizando (9), $V_t(\tau) = 1 + Te^{-rt} - \sqrt{1 + (Te^{-rt})^2}$.

Sabemos por nuestro examen anterior que, si el proceso de despegue se completa en τ años, la *CIN* y la *CIR* deben cruzarse en $t = \tau$. Así, el valor de τ debe satisfacer la siguiente ecuación [obtenida mediante la igualación de (13) y (15)]:

$$L(1 - V)e^{i\tau} = L\,1 - V_t(\tau)e^{r\tau} \qquad (16)$$

[20] Y $I - V$ es la fracción de la población inicial ocupada en la industria. La importancia de i como índice de esfuerzo nacional, por supuesto, de ninguna manera es una cuestión simple. Una i mayor significa una tasa anual más acelerada de absorción de mano de obra por el sector industrial; pero esto, debe recordarse, requiere tanto una tasa de inversión mayor en el sector agrícola para alimentar a la población creciente (a la vista de una posible disminución absoluta de la fuerza de trabajo agrícola), como una tasa mayor de inversiones en el sector industrial, para absorber a los nuevos trabajadores agrícolas liberados —con asignaciones entre los dos sectores, obedeciendo a nuestro criterio de crecimiento equilibrado—. El esfuerzo nacional en i es así una función del monto absoluto del fondo de inversión que puede estar disponible cada año durante el curso del proceso de despegue y una función de la eficiencia de su uso en los dos sectores. Por ejemplo, para el sector industrial, suponemos que sólo tiene lugar la ampliación de capital, entonces i también indica la tasa anual necesaria de inversión. Con respecto a las necesidades de inversión del sector agrícola, la tasa de incremento de la producción agrícola debe ser cuando menos igual a la de la población total, y la tasa anual necesaria de incremento de la productividad agrícola puede ser determinada únicamente. Se admite que la medida real de sacrificio se encuentra en la tasa de acumulación de utilidades. Pero su encadenamiento con la tasa de industrialización y la tasa de cambio de la productividad agrícola que se encuentra detrás de i requiere un conocimiento preciso de la eficacia relativa de la inversión y del efecto del cambio tecnológico en los dos sectores. Un análisis de este aspecto del problema del crecimiento equilibrado dinámico está siendo investigado por los autores, pero nos llevaría más allá de los confines del trabajo presente.

232

lo cual significa la intersección de las dos curvas. La ecuación (16) nos permite resolver i explícitamente en términos de τ:

$$(17) \qquad i = r + \frac{In(1/1 - V)}{\tau} + \frac{In[1 - V_t(\tau)]}{\tau}$$

Podemos, por lo tanto, determinar el esfuerzo mínimo anual, tal como lo resume i, por cualquier valor dado de τ. En forma inversa, si conocemos i, podemos determinar τ, la duración del proceso de despegue.

Antes de someter este resultado a otra prueba condicional, queda por generalizar nuestro modelo, lo cual es un importante paso más cercano a la realidad. Además de las necesidades de consumo de los trabajadores agrícolas e industriales con el salario institucional, puede haber otras pretenciones (o mercados) para la producción agrícola. Específicamente, el sector industrial puede requerir materias primas y el trabajador industrial puede demandar un salario con prima sobre el nivel de salarios institucionales de la agricultura. Podemos clasificar esas demandas como proporcionales a la fuerza de trabajo industrial, $L(1 + s) - x$, con d como el factor de proporcionalidad.[21] Otros mercados omitidos antes, incluyen las necesidades de consumo de los terratenientes y la demanda de exportaciones de productos agrícolas.[22] Podemos clasificar estas demandas (por falta de una mejor hipótesis) como proporcionales al crecimiento de la producción agrícola total, con $1 - \theta$ como factor de proporcionalidad. Cuando restamos esos renglones adicionales de la producción agrícola total, obtenemos

$$(18) \qquad EAP = ky - xW - dW[L(1 + s) - x] - (1 - \theta)\, ky$$

en lugar de (7). Con la incorporación de esas complicaciones a nuestro modelo, y con una CCP dada, puede derivarse la siguiente expresión (véase Apéndice) para V_t, el punto de inflexión de la mano de obra agrícola como una fracción del punto de inflexión de la población total:

$$(19) \qquad V_t(\tau) = \frac{1}{(\theta + 2d)\ (1 + s(\tau))} \Big[(\theta + d)T + [1 + s(\tau)]\ (1 + d) - \\ - \sqrt{[dT - (1 + s(\tau))\ (1 + d)]^2 + (\theta + 2d)\ T^2\, \theta}\Big]$$

V_t es así una función de los parámetros T, θ, d y s, y así vemos que nuestra formulación previa en (8) se convierte en un caso especial de (19) si dejamos $d = s = 1 - \theta = O$. Por lo demás, puede mostrarse que al aumentar T o al decrecer s, el valor de V_t crece, lo cual es la conclusión idéntica que se obtuvo para el caso simplificado.

21 Por conveniencia, en el cómputo, d puede medirse en términos de unidades de salario institucional, W. Por ejemplo, supóngase que este apoyo "adicional" de trabajadores industriales está representado por materias primas, más premios de salarios en la cantidad de $ 2 por trabajador y el salario institucional es de $ 4 por obrero; entonces d es igual a .5.

22 Los autores investigan, actualmente, todas las implicaciones de economía abierta del modelo.

233

El análisis de la duración del proceso de despegue puede entonces resumirse una vez más por (17) arriba citada, pero con $V_i(\tau)$ que ahora se define por (19) en lugar de (15). Empleando (17), podemos ahora obtener varios valores de i con valores diferentes de τ, r, T, V, θ y d. Los resultados se presentan en el cuadro 2.[23] Ellos nos permiten determinar el esfuerzo mínimo anual para cualquier valor dado de τ. En forma recíproca, si conocemos el esfuerzo promedio anual i que puede obtenerse, entonces derivamos τ, la duración del proceso de despegue.[24]

Cuadro 2. Esfuerzo mínimo anual promedio (i) ($\theta = .9$, $d = 1.3$, $V = 8$)
(Por ciento)

r \ τ (años)	($T = .7$) 5	20	35	50	r \ τ (años)	($T = .9$) 5	20	35	50
1.0	8.15	3.05	2.29	1.99	1.0	6.00	2.58	2.07	1.85
2.0	9.52	4.34	3.51	3.16	2.0	7.47	3.96	3.36	3.08
2.5	10.19	4.96	4.10	3.71	2.5	8.20	4.62	3.97	3.66
3.0	10.86	5.56	4.67	4.26	3.0	8.92	5.26	4.57	4.21
3.5	11.53	6.16	5.22	4.79	3.5	9.63	5.89	5.14	4.76

La aplicación del lado izquierdo del cuadro 2 es quizá más importante para las regiones de Asia que tienen un fuerte excedente de fuerza de trabajo, *v.gr.* India y Pakistán. Con un crecimiento anual de población estimado en cerca del 2.5 %, el esfuerzo mínimo anual en términos del crecimiento anual del sector industrial debe ser mayor del 10 % si el despegue se va a lograr en un periodo de cinco años. Si el país, en forma realista, establece una meta de 20 o 50 años para completar el despegue, el sector industrial debe crecer sólo al 4.96 o 3.71 %, respectivamente. Además, si los programas de control de la población ahora en realización tienen éxito en reducir a 1 % la tasa de crecimiento de la población, la carga equivalente en la economía, en términos de esfuerzo mínimo debe reducirse más todavía al 3.05 y 1.99 % respectivamente. Para el caso de un país latino-

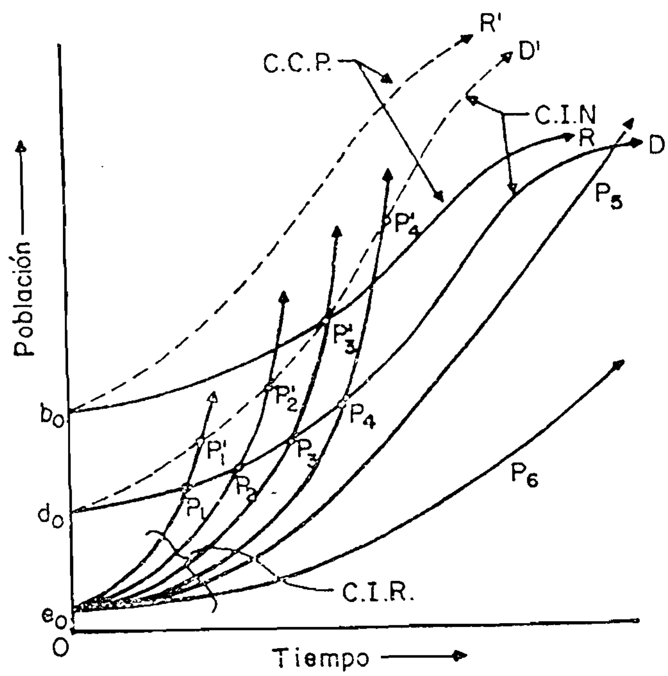
[23] Los parámetros V, θ y d se han estimado partiendo de importantes datos empíricos, principalmente para el Japón hacia fines del siglo XIX. La estimación para V(= .8) se basa en (7), θ(= .9) en (10), y d (1.3) en gran parte sobre los resultados de un estudio reciente no publicado de insumo-producto para 1953-54 por el Indian Statistical Institute, Calcuta. Estimaciones independientes para T son más difíciles de lograr. Hemos utilizado dos de las "estimaciones aproximadas" que se hacen con más frecuencia y que cubren el orden de lo plausible (.7 y .9). Una r que se comporta bien puede variar de 1 a 3.5 y τ de 5 a 50.

[24] Los autores se hallan ocupados actualmente en la comprobación de la validez empírica del modelo y así de su valor de predicción, examinando la medida en que proporciona un marco consistente de explicación en el caso de los países cuyo despegue se ha completado ya. Los índices teóricos de ejecución que muestra el modelo, pueden, por ejemplo, compararse con los índices de ejecución real de economías dadas, durante periodos determinados. Como la elaboración de este esfuerzo nos llevaría más allá de los límites del presente artículo, nos conformamos con atestiguar los primeros estimulantes resultados con respecto al caso del Japón.

234

americano o africano donde podemos ser un poco más optimistas con relación a la dotación inicial de recursos, el lado derecho del cuadro 2 puede ser más pertinente.

Con la ayuda de la gráfica 4, los resultados de esta sección pueden resumirse en forma de una serie de "teoremas" estáticos comparativos. Es



Gráfica 4

evidente que, mientras mayor sea el esfuerzo anual, más corto será el proceso de despegue. Esto confirma simplemente la conocida ventaja de las economías capaces y deseosas de someterse a la austeridad temporal para lograr un mayor fondo de inversión. Para la misma población industrial inicial, Oe_o en la gráfica 4, y la misma curva de crecimiento de la población, b_oR, y por ende, la misma curva de industrialización necesaria, d_oD, puede mostrarse una serie de curvas de industrialización real, e_oP_1, e_oP_2... Claramente, mientras mayor sea el esfuerzo real anual, mayor será la pendiente de la *CIR* y más pronto (esto es, a una τ menor) se presentará el punto de inflexión marcado por la intersección de *CIN* y *CIR* (en P_1, P_2...).

Mientras mayor sea la tasa de crecimiento de la población, más larga será la duración del proceso de despegue o mayores probabilidades de que el despegue se haga imposible. Con la ayuda de la gráfica 4 podemos fácil-

235

mente trazar el efecto de un incremento en r, cambiando *CCP* de b_oR a, digamos, b_oR'. En consecuencia, *CIN* también se desplaza hacia arriba (de d_oD a d_oD') y el punto de inflexión se retrasa, pues su intersección con la importante *CIR* ahora ocurre en P'_1, P'_2, ... Por otra parte, en el caso de *CIR* e_oP_5, que interceptaba la vieja *CIN* en P_5, el punto de inflexión puede ser ahora imposible de lograrse. El mismo esfuerzo anual es ahora insuficiente para satisfacer la presión de un crecimiento acelerado.

Si el esfuerzo mínimo que puede lograrse de la economía no es bastante grande, el buen éxito del despegue puede resultar por completo imposible. Esta observación sólo confirma la noción de que algunas economías pueden no tener capacidad para llegar al punto de inflexión, no importa el tiempo que estén dispuestas a esperar, debido a que su dotación de recursos o sus motivaciones son inadecuadas. Esta situación se representa en la gráfica 4 por *CIR* e_oP_6 la cual, como se observa, no intercepta *CIN* en ningún punto independientemente del periodo de tiempo permitido. Es sólo quizá en este sentido que podemos hablar de un único esfuerzo mínimo crítico como la tasa mínima anual de crecimiento de la fuerza de trabajo industrial que es tangente a la importante *CIN*. Si la i desciende por debajo de este mínimo crítico, el valor de τ será infinitamente grande. En tal país, puesto que no se presenta el punto de inflexión y el proceso de despegue no se completa con éxito, podemos decir, sin violar el sentido común, que el proceso no ha comenzado realmente. La economía sólo experimenta una salida temporal del estancamiento y está, de hecho, todavía en su etapa de preacondicionamiento.

Como podemos ver en la ecuación (17), el despegue puede ocurrir sólo si $i > r$; si $i = r$ o $i < r$, no importa lo grande que sea τ, el despegue es imposible. Si r aumenta, por lo regular debido a un descenso de la mortalidad, la economía debe reducirla otra vez, mediante una disminución de la fertilidad a través de un programa de planeación de la familia o bien, debe aumentar su esfuerzo de desarrollo nacional, i, apretándose aún más el cinturón. Debe así señalarse que el concepto de un esfuerzo mínimo crítico no puede tener vida independiente sino que debe definirse en términos de una tasa dada de crecimiento de la población así de una fecha dada como meta para completar el proceso de despegue. Se requiere un "gran empujón" no para salir en forma definitiva del estancamiento, sino para proporcionar un esfuerzo sostenido durante el tiempo para resistir las presiones malthusianas próximas y las aspiraciones de crecimiento de una sociedad dada. Usando la ya familiar analogía, no basta que un avión alcance una velocidad inicial que le permita escapar de la atracción de la gravedad de la tierra; debe tener capacidad para transportar suficiente combustible que le permita pasar las montañas de los alrededores y llegar a su destino a una velocidad dictada por la ambición del piloto.

En este trabajo hemos intentado construir un modelo explicativo de

236

la transición de las economías menos desarrolladas, del estancamiento al crecimiento autosostenido. En el curso de este ensayo se han formulado rigurosamente un número de nociones familiares en la literatura sobre el desarrollo y se han asimilado en lo que consideramos un patrón significativo. La reformulación de los supuestos que apoyan la curva de oferta ilimitada de fuerza de trabajo de Lewis nos permitió definir el proceso de despegue (*take-off*) en forma no arbitraria y, con la ayuda de un concepto del crecimiento equilibrado a corto plazo y de una tesis retocada del esfuerzo mínimo a largo plazo, elaborar las condiciones para completarlo con éxito.

REFERENCIAS

1. B. F. Hoselitz, "Non-Economic Factors en Economic Development", *Am. Econ. Rev., Proc.*, mayo de 1957, pp. 47, 28-41
2. Harvey Leibenstein, *Economic Backwardness and Economic Growth*. Nueva York,, 1959.
3. Arthur Lewis, *Development with Unlimited Supplies of Labour*, The Manchester School, mayo de 1954, 22, 139-92. Existe versión en castellano. Véase EL TRIMESTRE ECONÓMICO, vol. XXVII, núm. 108, p. 629.
4. ——, *Unlimited Labour: Further Notes*, Manchester School, enero de 1958, *26*, 1-32.
5. Ragnar Nurkse, *Problemas de formación de capital en los países insuficientemente desarrollados*, 2ª ed., Fondo de Cultura Económica, México, 1960.
6. ——, "Reflections on India's Development Plan", *Quar. Econ.*, mayo de 1957, *71*, 188-204.
7. Kazushi Ohkawa, *The Growth Rate of the Japanese Economy Since 1878*, Tokio, 1957.
8. Harry Oshima, "Underemployment in Backward Economies — An Empirical Comment", *Jour. Pol. Econ.*, junio de 1958, *66*, 259-64.
9. Gustav Ranis, "Economic Development: A Suggested Approach", *Kiklos*, 1959, *12*, 428-48.
10. ——, "Financing Japanese Economic Development", *Econ. Hist. Rev.*, 1959, *11*, 440-54.
11. W. W. Rostow, "The Take-Off into Self-Sustaining Growth", *Econ. Jour.*, marzo de 1956, *66*, 25-48.

APÉNDICE

I. PRODUCTO TOTAL Y FUNCIONES DEL PRODUCTO MARGINAL

En las gráficas A.1 y A.2, diremos que el punto O es el origen y mediremos la población agrícola, x, sobre el eje horizontal a la izquierda de O. La función *PFT* (productividad física total), $f(x)$, y la función *PFM* (productividad física marginal), $f'(x)$, se miden sobre el eje vertical; hacia abajo para $f(x)$ y hacia arriba para $f'(x)$.

237

La población agrícola incial será igual a L (ubicada en el punto S en ambas gráficas A.1 y A.2) y el producto total de $x = L$ es M (ubicado en el punto S'). La fuerza de trabajo no excedente en cada caso es igual a TL (esto es, localizada en el punto P). La definición de la fuerza de trabajo no excedente es $f'(x) = O$ porque $x \geq TL$.

Al derivar la función *PFT*, deben distinguirse los dos casos, es decir, $0 < T \leq 1$ (gráfica A.1) y $T > 1$ (gráfica A.2). El primer caso significa que una parte de L, o sea $(1 - T)L$, ya es excedente. El segundo caso significa que la oferta existente de tierra *podría* haber tolerado un incremento ulterior [en la cantidad de $(T - 1)L$] de población más allá de la población inicial, L, antes de que porción alguna de la población se haga excedente. Suponiendo que $f''(x) = 0$ (esto es, la función *PFM* es una línea recta), la función *PFT*, $f(x)$, debe satisfacer las condiciones siguientes para los dos casos que acabamos de establecer:

(1) (*a*) $f''(x) = 0$ (la curva de la *PFM* se forma de líneas rectas)
(*b*) $f'(x) = 0$ (la curva de la *PFM* es una línea horizontal más allá del punto P) para $x \geqq TL$
(*c*) $f(0) = 0$ (la curva *PFT* comienza en el origen)
(*d*) $\begin{cases} f(x) = M \text{ para el caso } 0 < T \leqq 1 \text{ (gráfica A.1) porque } x \geqq TL \\ f(L) = M \text{ para el caso } T > 1 \text{ (gráfica A.2)} \end{cases}$

Es fácil comprobar que la función *PFT*, $f(x)$, tomará las siguientes formas si todas las condiciones en (1) se satisfacen:

$$(2) \quad (a)\ y = \begin{cases} M[-(x/TL)^2 + 2(x/TL)] & \text{porque } x \leq TL \\ & \text{para el caso } T \leq 1 \text{ (gráfica A.1)} \\ M & \text{porque } x > TL \end{cases}$$

$$(b)\ y = [M/(2T-1)]\,[-(x/L)^2 + 2T(x/L)] \quad \text{porque } x \leq TL\text{[25]} \quad \text{para el caso } T > 1 \text{ (gráfica A.2)}$$

Un incremento de la productividad agrícola se define como un desplazamiento proporcional hacia arriba de toda la curva *PFT*. Esto puede expresarse como sigue:

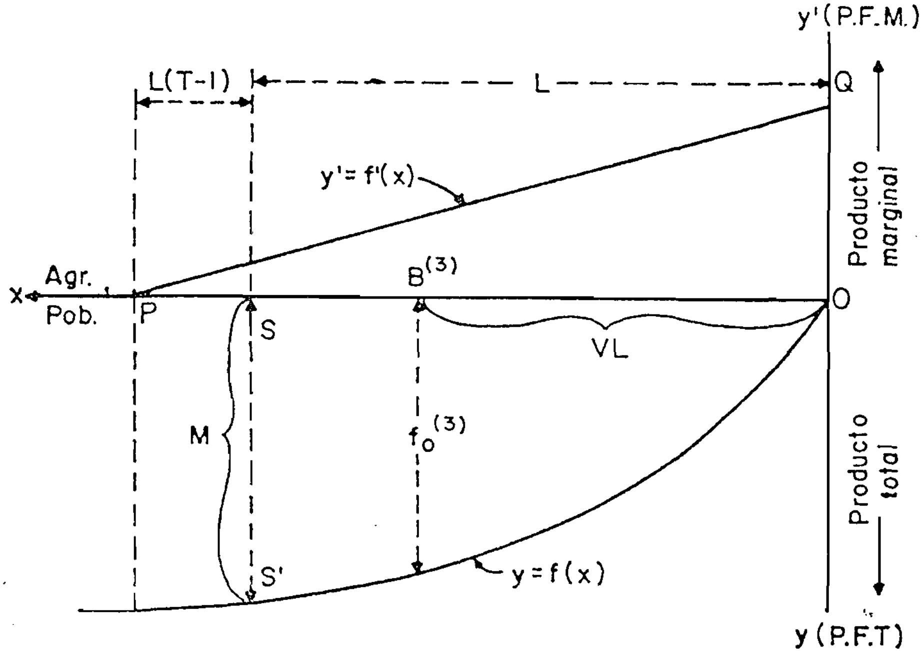
$$(3) \quad (a)\ y = f^{(1)}(k,x) = \begin{cases} kM[-(x/TL)^2 + 2(x/TL)] & \text{porque } x \leq TL \\ & \text{para el caso } T \leq 1 \text{ (gráfica A.1)} \\ kM & \text{porque } x > TL \end{cases}$$

$$(b)\ y = f^{(3)}(k,x) = [kM/(2T-1)]\,[-(x/L)^2 + 2T(x/L)] \quad \text{para el caso } T > 1 \text{ (gráfica A.2)}$$

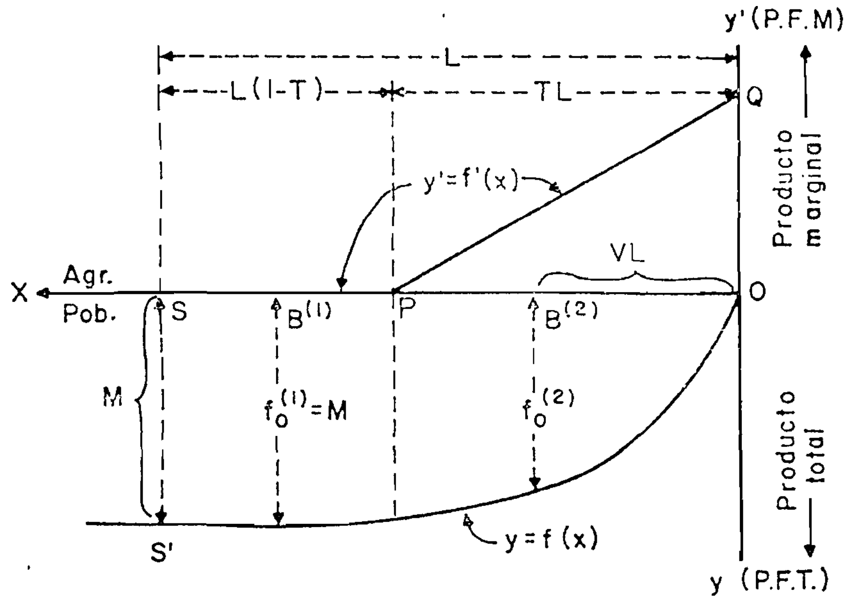
En otras palabras, después de haber tenido lugar un incremento de la productividad agrícola, la nueva función *PFT* es una k múltiple de las funciones en (2). La constante $k \geq 1$ se conocerá como el "coeficiente

[25] La *PFT* es constante para $x \geqq TL$. Sin embargo, puesto que la población agrícola declinará durante el proceso de despegue, no nos ocuparemos con esta porción de la función *PFT* en nuestro análisis.

238


y'(P.F.M.)
L(T-1)
L
Q
y'= f'(x)
Producto marginal
Agr.
x
Pob.
P
B(3)
O
S
VL
M
f0(3)
Producto total
S'
y = f(x)
y (P.F.T)

GRÁFICA A-2



y' (P.F.M)
L
L(1-T)
TL
Q
y'=f'(x)
VL
Producto
marginal
X
Agr.
Pob.
S
B(1)
P
B(2)
O
M
f(1)o=M
f(2)o
Producto
total
S'
y=f(x)
y (P.F.T.)

GRÁFICA A-1

de productividad" por que mide el grado de incremento de la productividad agrícola. (La selección de la notación $f^{(3)}(k,x)$ en (3b) es para facilitar la exposición posterior.)

De (2) y (3), pueden derivarse las funciones *PFM*:

$$(4)\quad (a)\ y' = \begin{cases} [2kM/(TL)^2]\,[-x+TL] & \text{porque } x \leq TL \\ & \text{para el caso } T \leq 1 \text{ (gráfica A.1)} \\ 0 & \text{porque } x > TL \end{cases}$$

$$(b)\ y' = [2kM/(2T-1)L^2]\,[-x+TL] \qquad \text{para el caso } T > 1 \text{ (gráfica A.2)}$$

II. Producción total y el salario institucional en el punto de partida

En el punto de partida, una parte de la población inicial, L, puede haber sido ya asignada al sector industrial. Diremos que la población agrícola en el punto de partida es igual a VL donde $0 \geq V \leq 1$ es la fracción de L en agricultura en ese momento. El punto de partida es indicado por los puntos $B^{(i)}(i = 1, 2, 3)$ en las gráficas A.1 y A.2. Se escogen esas anotaciones para distinguir tres casos posibles:

Caso uno: $V \geq T$ porque $T \leq 1$ (representada por el punto $B^{(1)}$ en la gráfica A.1).

Caso dos: $V < T$ porque $T \leq 1$ (representada por el punto $B^{(2)}$ en la gráfica A.1).

Caso tres: $V < T$ porque $T > 1$ (representada por el punto $B^{(3)}$ de la gráfica A.2).

Esos casos llevarán el índice $i = 1, 2, 3$ en este apéndice. (Para el caso uno, la $PFM = 0$; para los casos dos y tres la *PFM* es positiva.)

La producción agrícola total en el punto de partida se denota por $f_0^{(i)}(1 = 1, 2, 3)$. Los valores de $f_0^{(i)}$ pueden computarse de (3):

$$(5)\quad \begin{aligned} &(a)\ f_0^{(1)} = f^{(1)}(k=1, x=VL) = M \\ &\qquad\qquad (\text{porque } i = 1 \text{ en la gráfica A.1}) \\ &(b)\ f_0^{(2)} = f^{(1)}(k=1, x=VL) = [MV/T^2]\,[-V+2T] \\ &\qquad\qquad (\text{porque } i = 2 \text{ en la gráfica A.1}) \\ &(c)\ f_0^{(3)} = f^{(3)}(k=1, x=VL) = [MV(2T-1)]\,[-V+2T] \\ &\qquad\qquad (\text{porque } i = 3 \text{ en la gráfica A.2}) \end{aligned}$$

Denotaremos la tasa del salario institucional por $W^{(i)}(i = 1, 2, 3)$. El valor de $W^{(i)}$ está determinado por el requisito de que en el punto de partida la producción agrícola total $f_0^{(i)}$ debe ser suficiente para cubrir exactamente:

1) consumo de la población agrícola (VL) a la tasa de salarios $W^{(i)}$;
2) el consumo de la fuerza de trabajo industrial $[(1 - V)L]$ a la tasa de salario $W^{(i)}$;

240

3) el consumo del terrateniente y otros usos proporcionales a la producción agrícola total que se supone es una fracción, $1-\theta$, de $f_0^{(i)}$;
4) la demanda de productos agrícolas como materias primas y otros requisitos supuestamente proporcionales a la fuerza de trabajo industrial $[(1-V)L]$.[26]

Esto puede escribirse como sigue:

$$f_0^{(i)} = W^{(i)}VL + W^{(i)}(1-V)L + (1-\theta)f_0^{(i)} + dW^{(i)}(1-V)L \tag{6}$$

(El parámetro d es el "coeficiente de insumo", *v. gr.*, la cantidad de productos agrícolas usados como materias primas industriales por obrero empleado en el sector industrial, y se mide en términos de unidad de salario.)

Del (6) el salario institucional puede determinarse como:

$$W^{(i)} = (\theta/RL)f_0^{(i)} \text{ puesto que } i = 1, 2, 3 \tag{7}$$

de donde

$$R = 1 + d - dV \tag{8}$$

Puede obtenerse de (5) y (7) una expresión explícita de $W^{(i)}$, definida en términos de los parámetros M, L, θ, V, T, d (introducidos hasta ahora en nuestro sistema):

$$\begin{aligned} &a)\ W^{(1)} = \theta M/RL \\ &b)\ W^{(2)}\ \theta MV(-V+2T)/RLT^2 \\ &c)\ W^{(3)} = \theta MV(-V+2T)/RL(2T-1) \end{aligned} \tag{9}$$

III. Fuerza de trabajo agrícola equilibrada (*FTAE*) y fuerza de trabajo agrícola comercializada (*FTAC*)

Denominemos al monto de la fuerza de trabajo agrícola en el punto (comercialización) de escasez (véase la Sección I), fuerza de trabajo agrícola equilibrada (comercializada), esto es, *FTAE* (*FTAC*). Queremos determinar *FTAE* y *FTAC* como una función de k. Supóngase que hay un 100 % de incremento de la población total en el sector agrícola después del punto de partida. La población total aumenta de L a $(1+s)L$. Si x es el monto de la población agrícola, la población industrial es $L(1+s)-x$. Para esta distribución de la población total entre los dos sectores, la demanda de productos agrícolas contiene los siguientes componentes [cuando se dan los mismos supuestos de comportamiento (1)-(4), identificados en la última sección]:

[26] Obsérvese que en el texto introducimos inicialmente una versión simplificada de nuestro modelo en el cual 3) y 4) no se toman en consideración, esto es, se supone que es igual a 0 y θ igual a 1. En esas circunstancias, todos los otros resultados en este apéndice, que refleja el modelo completo, podrían simplificarse adecuadamente. En forma similar, por supuesto, inicialmente abstraemos del crecimiento de la población en el texto, esto es, suponemos $s = 0$. El lector interesado puede verificar los resultados presentados en el texto, haciendo que $s = d = (1-\theta) = 0$.

(10)
a) consumo de los terratenientes $A^{(i)}(x) = (1-\theta)f^{(i)}(k,x)$
b) consumo de los trabajadores agrícolas $B^{(i)}(x) = W^{(i)}x$
c) materias primas industriales $C^{(i)}(x) = dW^{(i)}[L(1+s)-x]$
d) consumo de la mano de obra industrial $D^{(i)}(x) = W^{(i)}[L(1+s)-x]$

donde $W^{(i)}$ se define en (7) o (9) y $f^{(i)}(k,x)$ se define en (3). [Obsérvese que, para $i = 1, 2, f^{(i)}(k,x)$ ambos se definen por (2*a*).] Puesto que la oferta de la producción agrícola es igual a la demanda, entonces

$$f^{(i)}(k,x) = A^{(i)}(x) + B^{(i)}(x) + C^{(i)}(x) + D^{(i)}(x) \tag{11}$$

La solución para x en (11) es la fuerza de trabajo agrícola equilibrada (*FTAE*). Ésta es sólo una forma alternativa de decir que *FTAE* está determinada por el requisito de que el excedente agrícola promedio (*EAP*) debe igualar el salario institucional $W^{(i)}$ que es la expresión que hemos usado en el texto. *EAT* (excedente agrícola total) y *EAP* (excedente agrícola promedio) se definen como:

$$EAT = f^{(i)}(k,x) - A^{(i)}(x) - B^{(i)}(x) - C^{(i)}(x)$$
$$EAP = EAT/[L(1+s)-x]$$

La ecuación (11) puede también obtenerse igualando *EAP* con $W^{(i)}$.

Para resolver x en (11), transfórmese U^* en *FTAE* expresada como una fracción de la población total, esto es, $FTAE = U^* L\,(1+s)$. Sustituyendo:

$$x = U^* L(1+s) \tag{12}$$

en (11) arriba citado, el valor de U^* puede determinarse. En otras palabras, es la fracción U^*, más bien que la cantidad absoluta de *FTAE*, que se determinará. Sustituyendo (7) y (12) en (11), tenemos:

$$f^{(i)}(k,x)/f_0^{(i)}(1+s)\ R^*/R \tag{13}$$

donde

$$R^* = 1 + d - dU^* \tag{14}$$

y donde el lado izquierdo puede computarse con (3), (5) y (12). La ecuación (13) se convierte entonces:

(15)
a) $(kR/T^2)\,[-(U^*(1+s)^2 + 2U^*T(1+s)] - (1+s)\,(1+d) + {}$
$+ dU^*(1+s) = 0$ puesto que $i = 1$
b) $(ZR/T^2)\,[-(U^*(1+s))^2 + 2U^*T(1+s)] - (1+s)\,(1+d) + {}$
$+ dU^*(1+s) = 0$ puesto que $i = 2,3$

de donde para (15*b*),

$$Z = kT^2/V(-V+2T) \tag{16}$$

Esas ecuaciones definen U^* como una función de k, el coeficiente de productividad. Indicando esta ecuación en su forma explícita, por

$U^* = U^*(k)$, y observando que Z toma el lugar de k en (15b), tenemos, después de una simplificación,

$$
(17)\quad
\begin{aligned}
&a)\ U^* = U^*(k) = \frac{T}{2kR(1+s)}\,[2kR + dT \\
&\qquad - \sqrt{(2kR + dT)^2 - 4kR(1+s)(1+d)}], \text{ puesto que } i = 1 \\
&b)\ U^* = U^*(Z), \text{ puesto que } i = 2,3
\end{aligned}
$$

Para computar la fuerza de trabajo agrícola comercializada (*FTAC*), primero se iguala (*PFM*) en (4) al salario institucional $W^{(i)}$ en (9):

$$
(18)\quad
\begin{aligned}
&a)\ W^{(1)} = [2kM/(TL)^2]\,(-x + TL), &&\text{puesto que } i = 1 \\
&b)\ W^{(2)} = [2kM/(TL)^2]\,(-x + TL), &&\text{puesto que } i = 2 \\
&c)\ W^{(3)} = [2kM/(2T-1)\,L^2]\,(-x + TL) &&\text{puesto que } i = 3
\end{aligned}
$$

La solución de x en (18) nos da la *FTAC*. Indicando *FTAC* por $V^*(1+s)L$ (esto es, V^* es la fracción de la población total que es *FTAC*), la expresión:

$$
(19)\qquad x = V^*(1+s)L
$$

puede sustituirse en (18) con objeto de resolver V^* como una función $V^*(k)$ de k. Después de sustituir (19) y (9) en (18), derivamos:

$$
(20)\quad
\begin{aligned}
&a)\ V^* = V^*(k) = [T/(1+s)]\,(1 - T\theta/2kR) &&\text{puesto que } i = 1 \\
&b)\ V^* = V^*(Z) \quad \text{puesto que } i = 2,3,
\end{aligned}
$$

donde Z en (20b) se define como en (16).

IV. Coeficiente ($k^{(i)}$) de productividad del punto de inflexión y punto de inflexión de la fuerza de trabajo agrícola (*PFTA*)

El coeficiente de productividad del punto de inflexión $k^{(i)}$ es ese nivel de coeficiente de productividad que iguala $U^*(k)$, (17), y $V^*(k)$, (20). Si resolvemos k estableciendo $U^*(k) = V^*(k)$, tenemos:

$$
(21)\quad
\begin{aligned}
&a)\ k^{(i)} = (1/2R)\,[(1+s)(1+d) - dT + \\
&\qquad + \sqrt{(dT - (1+s)(1+d))^2 + (\theta + 2d)T^2\theta}] &&\text{puesto que } i = 1 \\
&b)\ k^{(i)} = [V(-V + 2T)/T^2]\,k^{(1)} &&\text{puesto que } i = 2,3
\end{aligned}
$$

expresando $k^{(i)}$ como una función de los parámetros, V, s, d, T, θ. El punto de inflexión de la fuerza de trabajo agrícola, V_t, (*PFTA*) es la *FTAE* (= *FTAC*) cuando el coeficiente de productividad, k, asume el valor del punto de inflexión (21). Así, sustituyendo (21) en (20), tenemos:

$$
V_t = A + B - \sqrt{(A+B)^2 - 2AB\,\frac{2d+\theta}{d+\theta}},
$$

donde

243

$$A = \frac{1 + d}{\theta + 2d}$$

$$(22) \qquad B = \frac{Q(d + \theta)}{\theta + 2d}$$

$$Q = \frac{T}{1 + s}$$

El valor de V_t, que es el mismo para los tres casos ($i = 1, 2, 3$), se considera como una función de los parámetros, d, T, s, θ. También debe notarse que no participan en (22) los parámetros L (la población inicial) y M (la producción agrícola inicial total). La importancia económica de este hecho es que la dimensión absoluta, esto es, la escala de la economía, medida en términos de L y/o M, no es pertinente para los argumentos de este trabajo.

En (22) vemos que V_t es no-negativa. Por lo demás, puede mostrarse que $V_t \leq 1$ si se satisface la siguiente condición:

$$(23) \qquad T \leq \frac{(1 - \theta/2)\,(1 + s)}{(1 - \theta)}$$

Esto, según hemos señalado en el curso del examen empírico de nuestro modelo en el texto, permite postular todos los valores razonables de T, para producir $0 \leq V_t \leq 1$. Suponiendo que se satisface (23), puede demostrarse fácilmente que

$$\frac{\partial V_t}{T\partial} \geq 0 \text{ y } \frac{\partial V_t}{\partial s} \leq 0.$$

244

HOUSEHOLD AND POPULATION EFFECTS ON AGGREGATE CONSUMPTION

245

Frank T. Denton and Byron G. Spencer*

I. Introduction

CASUAL observation suffices to suggest important connections between demographic variables and consumption-saving behaviour. Patterns of spending vary with the age and family responsibilities of individuals over their lifetimes, and one might expect this to show up at the aggregate level when the age composition of the population changes. Then too, there are possible economies of scale in household consumption — "two can live as cheaply as one," the saying goes — and one might expect these to manifest themselves at the aggregate level also, in response to changes in average household size.[1] Our purpose in this paper is to develop a consumption function in which demographic influences are explicit and to use it to test for household and population effects.[2]

An elegant and appealing theory of the influence of age on consumption and saving is provided by the "life-cycle" hypothesis proposed originally by Modigliani and Brumberg (1954). Our own approach starts within a different framework at the microlevel and leads to a different formulation of the consumption function at the macrolevel. At the same time, it allows for a lifetime pattern of consumption and saving by individuals that may reflect their changing family responsibilities and income levels as they grow older. Population and household variables appear explicitly in our formulation, so that our purpose of allowing for direct links between demographic variables and aggregate consumption and saving is well served.[3]

II. The Basic Model

We begin by defining a "dependency matrix" $A = [a_{ij}]$.[4] Each column of A represents one individual, viewed as a *consuming unit;* each row represents one individual, viewed as a *spending unit.* The matrix has n rows and n columns, corresponding to the n individuals in the population. If $a_{ij} = 1$, the j^{th} individual is fully dependent on the i^{th} individual for his consumption requirements. If $a_{ij} = 0$, there is no dependency relationship. In order to allow for shared responsibility, we permit also the elements of A to be positive fractions: for example, if $a_{ij} = .5$ and $a_{kj} = .5$, individuals i and k are jointly and equally responsible for individual j. If each individual were responsible for himself and only himself, A would be an identity matrix. But children may be dependent on their fathers, wives on their husbands, old people on younger relatives, and so on. Thus we expect $A \neq I$.

Received for publication June 17, 1974. Revision accepted for publication February 26, 1975.

* This paper is one of a series reporting results obtained from a continuing project involving the study of economic-demographic relationships. We gratefully acknowledge the support of this project at various stages by McMaster University, the Canada Council, and the Ford and Rockefeller Foundations. Some of the material was presented at Professor Richard Stone's Growth Project Seminar at Cambridge University, and we are grateful to the participants for their comments. We also thank Angus Deaton, Leslie Robb, and an anonymous referee for helpful comments on earlier drafts of the paper. We note with appreciation the valuable research assistance provided by Christine Feaver.

1 A useful summary and discussion of much of the relevant literature is provided in United Nations (1973, ch. XIII, p. 452), where it is noted that "a number of findings . . . suggest that demographic factors may exert an influence on savings . . . . The problem lies mainly in assessing the impact of population in quantitative terms and its relative importance as compared to other determinants . . . . The findings suggest that population and demographic factors may play a significant, though not crucial role, in private savings."

2 One of our aims was to extend the economic-demographic model presented in Denton and Spencer (1973a) by incorporating a consumption function sensitive to demographic change. A first extension along these lines is presented in Denton and Spencer (1973b).

3 One of the assumptions appealed to in establishing consistent aggregation properties for the original "life-cycle" consumption function is that of constancy through time of the age distribution of the population. (See Ando and Modigliani, 1963, pp. 58–59.) Such an assumption is clearly unacceptable for our purposes, since our aim is to study the effects of changes in age distribution. However, for an analysis within the "life-cycle" framework which allows for changes in age distribution, see Heien (1972).

4 Conceptualization may be facilitated if we begin with the use of some matrix notation even though it is not used in the subsequent algebraic development.

THE Review of Economics and Statistics Vol LVIII feb. 1976 número 1 Published for Harvard University by North-Holland Publishing Company.

245

We next define an n-dimensional vector $b = [b_1\ b_2 \ldots b_n]$ in which each element represents one individual and indicates whether he is the head of a household, in the sense of being responsible for its consumption requirements. If $b_i = 1$, the i^{th} member of the population is a household head; if $b_i = 0$, he is not. As before, we allow for shared responsibility by permitting also the elements of b to be positive fractions.

The consumption that takes place within a given household is related in part to the individuals who make up the household and in part to the household itself. To some extent the distinction can be made in terms of particular classes of goods: food and clothing, for example, are most closely associated with individuals while the family car and the refrigerator in the kitchen are most closely associated with the household. Of course, this categorization is imperfect. A larger family may require a larger refrigerator; clothing can be passed down from one child to another within the household; and so on. But this causes no difficulties in principle, since we can allow some types of consumption to be related both to the household and to its individual members. All that is required is that we be able to express total consumption within a household as the sum of a constant term and a component that varies with the number and personal characteristics of its members.

Assume that at any given time there are certain basic levels of consumption associated with individuals and households. These are determined in part by nutritional requirements, requirements for basic clothing and shelter, and so on, and in part by social customs. Whatever the reasons, individuals, *qua* spending units, attempt to meet these basic levels of consumption before making any further spending or saving decisions. Basic consumption levels we think of as *committed consumption.* Consumption beyond basic levels we think of as *discretionary consumption.*

Let π_1 be the total cost of committed consumption associated with a household and assume it to be the same for all households. Let π_{2j} be the total cost of committed consumption associated with the j^{th} individual. We assume that each individual, *qua* spending unit, behaves as follows: (1) he first meets any commitments that he may have as a provider for himself, his dependents, and his household; (2) he then allocates whatever income remains between consumption and saving in accordance with a fixed proportional savings rule, assumed to be the same for all individuals. Thus, his total expenditure on consumer goods is given by

$$C_i = Z_i + \beta(Y_i - Z_i) = (1 - \beta)Z_i + \beta Y_i \quad (1)$$

where Y_i is income and Z_i is committed expenditure, defined by

$$Z_i = b_i\pi_1 + \sum\nolimits_{j=1}^{n} a_{ij}\pi_{2j}. \quad (2)$$

Summing over all n individuals, aggregate consumption is given by

$$C = (1 - \beta)Z + \beta Y \quad (3)$$

where $C = \Sigma C_i$, $Y = \Sigma Y_i$, and $Z = \Sigma Z_i$. Assuming that π_2 varies only with age,[5] Z can also be defined as

$$Z = \pi_1 H + \Sigma\pi_{2k}N_k \quad (4)$$

where H is the total number of households and N_k is the population in the k^{th} age group. (Note that $\Sigma b_i = H$.) As a third possibility, Z may be expressed as

$$Z = \Sigma(\pi_1 r_k + \pi_{2k})N_k \quad (5)$$

where r_k is the so-called "headship rate" for the k^{th} age group, i.e., the proportion of people in the group who are heads of households. The introduction of headship rates focuses attention more directly on the consequences of changes in the age composition of the population. It also permits explicit consideration of the effects of changes in headship rates such as those associated with changes in age at marriage.[6]

It is convenient to break Z into two parts, as follows. Let θ be the value of π_2 for an adult. We may then write

$$Z = \theta\Sigma(\delta_1 r_k + \delta_{2k})N_k = \theta(\delta_1 H + \Sigma\delta_{2k}N_k) = \theta z \quad (6)$$

[5] For convenience, we think of π_2 as varying only with age. However, there is no conceptual problem in allowing the parameter to vary also with sex or any other population characteristic.

[6] The "headship rate" is a convenient concept in demographic analysis. For our purposes, it matters little that the criteria employed in census taking for designating the head of a household may differ from the "responsibility" criterion defined above. The incorporation of r into equation (5) introduces a meaningful demographic parameter without creating any complications as far as the theory is concerned.

246

where $\delta_1 = \pi_1/\theta$ and $\delta_{2k} = \pi_{2k}/\theta$. The variable $z = \delta_1 H + \Sigma\delta_{2k}N_k$ may be interpreted as the number of weighted consuming units: an adult is given a weight of 1 and a child a weight less than 1; a household is given some appropriate weight δ_1.

Now let us introduce the notion of changing tastes. We expect a society's views on what are basic needs to change as income levels change. Let us suppose that individuals make comparisons with other individuals and households. They observe the levels of consumption per consuming unit that obtain elsewhere in society and gradually revise their ideas about basic needs in accordance with what they perceive to be the "permanent" standard of living of others. How rapidly the revision takes place will depend on how heavily an individual weighs the present against the past, on perception lags, and so on. A convenient way of incorporating the revision of standards is to assume that basic needs are adjusted in proportion to a weighted average of consumption per consuming unit in previous periods, with a geometrically declining lag structure. Writing $c_t = C_t/z_t$ for consumption per spending unit in period t, we define the society's permanent standard of living, as perceived by its members, to be

$$c^*_t = (1-\lambda)\sum_{j=0}^{j=\infty} \lambda^j c_{t-j-1}; \qquad 0 < \lambda < 1 \tag{7}$$

and rewrite equation (1) in dynamic form as

$$C_{it} = (1-\beta)Z_{it}(c^*_t/c^*_0) + \beta Y_{it} \tag{8}$$

where c^*_0 is the permanent standard of living in the initial period to which δ_1 and the δ_2 parameters relate.[7] Aggregating, we obtain

$$C_t = \alpha(1-\beta)z_t c^*_t + \beta Y_t \tag{9}$$

where $\alpha = \theta/c^*_0$.[8]

III. Utility Maximization and the Proportional Savings Rule

The assumption of a constant proportional savings rule is consistent with optimizing behaviour of the following sort. Assume that an individual has a utility function of the form $U = U(C',C'',W,dW)$ where C' is the rate of committed consumption, C'' is the rate of discretionary consumption, W is (non-human) wealth, and dW is the rate of accumulation of wealth. The individual is assumed to derive utility from his own consumption, the consumption of his dependents (possibly), and the existence of a stock of financial or real assets which give him protection against an uncertain future and provide him with social status. In addition, he judges his rate of achievement by the rate at which he adds to his wealth. That is to say, he derives utility from the act of saving as well as from his stock of previous savings.

Assume further that U is separable to the extent that $U = U_1(C',W) + U_2(C'',dW)$. At any point in time, C' and W are given. The individual derives utility from them but his optimizing decision reduces to the maximization of U_2 subject to an income constraint. As a final assumption, let U_2 be homogeneous of degree k. Noting that $C'' = \beta(Y-C')$ and $dW = (1-\beta)(Y-C')$, U_2 can then be written as $U_2 = (Y-C')^k\ U_2(\beta,1-\beta)$. Since Y and C' are given at any point in time, maximizing U_2 is equivalent to choosing a value of β which maximizes $U_2(\beta,1-\beta)$. The maximizing value of β will be the same for all values of Y and C'. We expect β to lie in the open interval 0–1.

The foregoing may be extended to allow for interpersonal and intertemporal differences in utility functions. Let the utility function for individual i at time t be written as $U_{it} = U_{1it}(C'_{it},W_{it}) + U_{2it}(C''_{it},dW_{it})$ and specify that $U_{2it} = F_{it}(U_2(C''_{it},dW_{it}))$, where F_{it} represents any monotonic transformation. U_{2it}, and hence U_{it}, will then be maximized for each individual at each point of time by the same value of β, as required by the model described in the previous section. Thus, utility functions may vary among individuals in the population and over time, within the class defined by monotonic transformations of the original U_2 function.

IV. Some Properties of the Model

The model has a number of properties worth noting. First, it allows for variations in saving behaviour over the life cycles of individuals as

[7] The consumption of each individual is assumed negligible in relation to total consumption in the economy, so that c^* may be calculated without excluding individual i.

[8] Note that the parameter θ relates to the initial period and does not change through time: changes in the cost of committed consumption are reflected entirely in changes in c^*.

247

their commitments and incomes vary. Second, it responds to changes in the age distribution of the population. Third, it allows for scale effects in consumption: other things equal, an increase in average size of household will cause less than a proportionate increase in total consumption. Fourth, the aggregate savings ratio varies in periods of disequilibrium but approaches constancy as the population-income-consumption system approaches steady-state growth. Fifth, the steady-state aggregate savings ratio depends on the rates of growth of population and income.

The first three properties are clear from the discussion of section II. The other two may be demonstrated as follows. We begin by rewriting equation (9) as

$$c_t = \alpha(1-\beta)c^*_t + \beta y_t \tag{10}$$

where $y_t = Y_t/z_t$. Substituting (7) into (10) to eliminate c^*_t, subtracting λc_{t-1} from both sides, making repeated substitutions, and then dividing through by y_t, we obtain

$$c_t/y_t = C_t/Y_t = \beta + \beta\alpha(1-\beta)(1-\lambda) \sum_{j=1}^{j=\infty} [\alpha(1-\beta)(1-\lambda) + \lambda]^{j-1}(y_{t-j}/y_t). \tag{11}$$

(We assume $0 < \lambda < 1$, $0 < \beta < 1$, $\theta > 0$, and $c^*_0 > 0$. The condition that $\alpha(1-\beta)(1-\lambda) + \lambda$ be less than 1 in absolute value is required for stability. This implies $\alpha < 1/(1-\beta)$.) In steady-state growth, with Y and z growing at constant (though possibly different) rates, y_{t-j}/y_t is independent of t and the right side of the equation is constant. Thus the overall consumption and savings ratios approach constancy as the system approaches steady-state growth.

Now let y (and hence c) have steady-state growth rate g per period. Noting that $c^*_t = c_t(1-\lambda)(1-\lambda+g)^{-1}$, equation (10) provides a basis for writing

$$c_t/y_t = C_t/Y_t = \beta(1-\lambda+g)/ \{(1-\lambda)[1-\alpha(1-\beta)] + g\}. \tag{12}$$

Thus the consumption and savings ratios depend on the rates of growth of population and income which underly g. For α, β, and λ in the open interval 0–1, which is what we expect, a positive value of g will reduce C_t/Y_t and hence increase the savings ratio. This implies that the steady-state savings ratio will vary directly with the rate of growth of income and inversely with the rate of growth of population.

V. Estimation of the Aggregate Consumption Function from Time-Series Data

There are several matters to be considered in connection with the estimation of the aggregate function from time-series data. First, there is the question of an operational definition of z. To make the problem manageable, let us assume that the δ_2 parameters increase linearly from birth, reach a maximum of 1 at age 18, and remain constant for all subsequent ages. Letting δ_{20} be the value of δ_2 for a child under the age of 1, we may then write

$$z_t = \delta_1 H_t + \delta_{20} X_{1t} + X_{2t}, \tag{13}$$

where $X_{1t} = \sum_{k=0}^{k=17} (1-k/18)N_{kt}$ and $X_{2t} = \sum_{k=0}^{k=17} (k/18)N_{kt} + \sum_{k>17} N_{kt}$. Thus we have reduced to 2 the number of parameters involved in the specification of z.

The second question has to do with the form in which the consumption function is to be estimated. Here we proceed as follows. We rewrite equation (9) to incorporate a random disturbance u_t:

$$C_t = \alpha(1-\beta)z_t c^*_t + \beta Y_t + u_t. \tag{14}$$

We then subtract $\lambda(z_t/z_{t-1})C_{t-1}$ from both sides of (14), do some substitution and rearrangement, and obtain

$$C_t = \xi_1(z_t/z_{t-1})C_{t-1} + \xi_2 Y_t + \xi_3(z_t/z_{t-1})Y_{t-1} + v_t \tag{15}$$

where $\xi_1 = \alpha(1-\beta)(1-\lambda) + \lambda$, $\xi_2 = \beta$, $\xi_3 = -\lambda\beta$, and $v_t = u_t - \lambda(z_t/z_{t-1})u_{t-1}$. This eliminates c^*, which is convenient, since c^* depends on the unknown parameter λ and on all previous values of C and z, and hence is not directly observable. However, a new problem arises: even if u_t were serially uncorrelated, v_t would not be, and this, in conjunction with the presence of C_{t-1} in the equation, introduces some well known complications.

A third question relates to the obvious simultaneity in the determination of the contemporaneous income and consumption variables. We must regard income as being endogenous in

248

some larger but unspecified system of equations and this implies the familiar problem of correlation between v_t and Y_t. It also implies correlation between v_t and Y_{t-1}, inasmuch as the serial correlation in v_t implies that v_t and C_{t-1} are not independently distributed.

The strategy that we have adopted for estimating the consumption function from time-series data is as follows: (1) We choose equation (15) as our basic equation. (2) To deal with the simultaneity and serial correlation problems, we use a two-stage procedure which involves replacing Y_t, Y_{t-1}, and C_{t-1} with Y^{**}_t, Y^{**}_{t-1}, and C^{**}_{t-1}, the latter being computed by regressing the Y and C variables on a set of instrumental variables regarded as exogenous in the context of a "complete" (but unspecified) model of the economy. Equation (15), thus modified, is then fitted by ordinary least squares. (3) We derive estimates of α, β, and λ directly from the estimates of ξ_1, ξ_2, and ξ_3 in equation (15), while using a scanning procedure for δ_1 and δ_{20}, the other two parameters of interest. That is to say, we refit equation (15) with alternative values specified for δ_1 and δ_{20}, within an a priori admissible region, and examine the resulting sums of squares of residuals and other characteristics of the equations.

VI. Some Estimates Based on Canadian Time-Series Data

We now present some estimates of the aggregate consumption function derived from Canadian annual time-series data for the period 1928–1971, excluding the years 1940–1946. Consumption is defined as personal expenditure on consumer goods and services, expressed in 1961 dollars. Income is defined as personal disposable income, adjusted to eliminate the effects of short-term fluctuations in farm income,[9] and deflated by the national accounts implicit price index for personal expenditure. Population series are based on census data and official intercensal estimates. The household series was estimated from census and other data.[10]

In the first stage of the estimation procedure, values for Y^{**}_t, Y^{**}_{t-1}, and C^{**}_{t-1} were generated by regressing Y and C on a set of 18 "exogenous" variables. These represent current or one-period lagged variables of the sort that are typically exogenous in Canadian macroeconometric models.[11] In the second stage, equation (15) was fitted with $(z_t/z_{t-1})C_{t-1}$, Y_t, and $(z_t/z_{t-1})Y_{t-1}$ replaced by $(z_t/z_{t-1})C^{**}_{t-1}$, Y^{**}_t and $(z_t/z_{t-1})Y^{**}_{t-1}$.[12] The results are presented in table 1. Results are shown for 12 combinations of the δ_1 and δ_{20} parameters: $\delta_1 = 0,1,2,3$, in combination with $\delta_{20} = .1,.5,1.0$.[13]

The table reports values for ξ_1, ξ_2, and ξ_3, which are the coefficients computed directly by fitting equation (15), together with their asso-

[9] Canadian farm income, as defined for national accounting purposes, is subject to wide fluctuations which tend to be related to fluctuations of farm inventories — especially inventories of grain — and to have little connection with aggregate consumption behaviour. Although we have fitted the consumption function using both adjusted and unadjusted income variables, we have found the former to perform better. Net income received by farm operators from farm production is separated into two parts by passing a 5th degree trend polynomial through the series and treating the deviations as the short-term component.

[10] The household series is based on census totals from the censuses of 1921, 1931, 1941, 1951, 1956, 1961, 1966, and 1971. Intercensal annual estimates were calculated by interpolation on the basis of related series. Prior to 1961, the population 20 years of age and over was used to interpolate between census years. For the periods 1961–1966 and 1966–1971, the interpolation was based on the series of published household estimates provided by the annual Statistics Canada Survey of Household Facilities and Equipment.

[11] The set includes the following: government current expenditure on goods and services in constant dollars, both current and lagged; government capital expenditure in constant dollars, both current and lagged; exports of goods and services in constant dollars, both current and lagged; net income received by farm operators from farm production in constant dollars (i.e., deflated by the implicit price index for personal expenditure), both current and lagged; number of households; population 0–4 years of age, 5–9, 10–14, 15–19, and 20 and over; linear and quadratic time-trend variables; a dummy variable representing postwar transition effects (value 1 in 1946, 0.75 in 1947, 0.50 in 1948, 0.25 in 1949, and 0 in all other years); and a dummy variable representing once-and-for-all postwar shift effects (value 1 in all years after World War II and value 0 in all years prior to World War II).

[12] A dummy variable designed to represent possible postwar transition effects in the period 1947–1949 was also tried. However, this variable proved to be of no significance and was discarded.

[13] The equation was estimated also for a number of other combinations of δ_1 and δ_{20} within these ranges. However, since the sums of squares of residuals, estimated parameter values, and other relevant statistics change smoothly, no significant additional information is provided by the other results.

249

TABLE 1. — ESTIMATED CONSUMPTION FUNCTIONS BASED ON CANADIAN TIME-SERIES DATA, 1928–1971 (EXCLUDING 1940–1946), FOR SELECTED VALUES OF δ_1 AND δ_{20}

$$C_t = \xi_1(z_t/z_{t-1})C_{t-1} + \xi_2 Y_t + \xi_3(z_t/z_{t-1})Y_{t-1}$$

Specified Values of δ_1, δ_{20}		Coefficients Estimated Directly			Derived Parameters				
δ_1	δ_{20}	ξ_1	ξ_2	ξ_3	α	β	λ	S^*	ρ
0	.1	.8305	.5316	—.3649	.98	.53	.69	265	.10
		(8.0)	(5.1)	(2.3)					
0	.5	.8311	.5332	—.3665	.98	.53	.69	262	.08
		(8.1)	(5.3)	(2.4)					
0	1.0	.8333	.5424	—.3776	.99	.54	.70	259	.07
		(8.2)	(5.6)	(2.5)					
1	.1	.8389	.5423	—.3843	.98	.54	.71	266	.11
		(8.1)	(5.1)	(2.4)					
1	.5	.8387	.5405	—.3819	.98	.54	.71	264	.09
		(8.2)	(5.3)	(2.5)					
1	1.0	.8395	.5441	—.3859	.98	.54	.71	262	.08
		(8.2)	(5.5)	(2.5)					
2	.1	.8438	.5493	—.3965	.97	.55	.72	267	.11
		(8.1)	(5.2)	(2.5)					
2	.5	.8434	.5463	—.3927	.98	.55	.72	265	.10
		(8.2)	(5.3)	(2.5)					
2	1.0	.8436	.5471	—.3932	.98	.55	.72	263	.09
		(8.2)	(5.4)	(2.6)					
3	.1	.8470	.5542	—.4047	.97	.55	.73	276	.12
		(8.2)	(5.3)	(2.6)					
3	.5	.8465	.5508	—.4006	.97	.55	.73	266	.11
		(8.2)	(5.3)	(2.6)					
3	1.0	.8465	.5500	—.3993	.98	.55	.73	264	.09
		(8.3)	(5.4)	(2.6)					

ciated "t-ratios."[14] It also reports the values of α, β (which is equal to ξ_2), and λ derived from these coefficients, as well as the standard error of estimate, corrected for degrees of freedom, S^*, and the first-order coefficient of serial correlation in the residuals, ρ.[15]

A striking feature of the table is the degree to which the results are insensitive to the values specified for δ_1 and δ_{20}. The parameter values change very little and the S^* surface is relatively flat. The function fits well in every case. However, this is scarcely surprising since close fits are a common characteristic of consumption functions. Of more interest is the fact that all parameters correspond with a priori expectations: β and λ are positive and lie in the interval 0–1; α is positive and less than $1/(1-\beta)$. (Thus the stability requirement noted following equation (11) is satisfied.) The values of ρ suggest that there is little serial correlation in the residuals.[16]

It will be noted that S^* falls as δ_1 falls and as δ_{20} rises. The t-ratios also improve somewhat as S^* falls. The lowest value of S^* occurs when δ_1 is 0 and δ_{20} is 1. This suggests an absence of both household and age-distribution effects on aggregate consumption, though the evidence must be regarded as rather weak in view of the small variations in S^*.

14 The inclusion of "t-ratios" (ratios of coefficients to standard errors) should not, of course, be taken to imply that the two-stage coefficient estimators have Student's t-distribution, except insofar as the latter might be viewed as an approximation to the unknown distributions of these estimators.

15 Coefficients of determination ($\bar{R}^2$) were also computed, as additional measures of overall goodness of fit. In every case $\bar{R}^2$ was greater than 0.999. However, it should be noted that the equation and estimation procedure that we have used results in residuals which do not sum to 0. In such a case, the usual interpretation of $\bar{R}^2$ is not strictly valid and some caution is warranted. See Aigner (1971, pp. 85–90) for a discussion related to this point in the context of a regression equation with no intercept term.

16 If u_t in equation (14) were serially uncorrelated, we would expect v_t in equation (15) to be negatively serially correlated. On the other hand, if u_t were positively serially correlated, which would not be an unusual situation, the tendency toward negative serial correlation in v_t would be offset, in greater or lesser degree. The low positive values of ρ in table 1 might be viewed as consistent with the latter possibility.

250

VII. Some Estimates Based on International Cross-Section Data

A basic difficulty in the use of time-series observations on demographic variables is that these variables tend to change slowly and to be highly intercorrelated, making difficult the task of identifying their separate influences. As an alternative approach, which is less likely to be affected by this problem, we have estimated the same consumption function using international cross-section data. Specifically, we have fitted equation (15) to data for 21 Organization for Economic Cooperation and Development (OECD) member countries. The results are reported in tables 2 and 3.

TABLE 2. — ESTIMATED CONSUMPTION FUNCTIONS BASED ON CROSS-COUNTRY DATA, USING GROSS NATIONAL PRODUCT AS INCOME MEASURE, FOR SELECTED VALUES OF δ_1 AND δ_{20}

$$C_t = \xi_1(z_t/z_{t-1})C_{t-1} + \xi_2 Y_t + \xi_3(z_t/z_{t-1})Y_{t-1}$$

Specified Values of δ_1, δ_{20}		Coefficients Estimated Directly			Derived Parameters			
δ_1	δ_{20}	ξ_1	ξ_2	ξ_3	α	β	λ	S^*
0	.1	.8591 (17.5)	.2383 (2.8)	—.1442 (1.3)	.84	.24	.61	328
0	.5	.8701 (19.1)	.2854 (3.5)	—.2004 (1.9)	.79	.29	.70	305
0	1.0	.8767 (21.0)	.3291 (4.1)	—.2502 (2.4)	.72	.33	.76	282
2.5	.1	.8574 (17.3)	.2325 (2.8)	—.1371 (1.2)	.85	.23	.59	331
2.5	.5	.8644 (18.2)	.2607 (3.1)	—.1709 (1.6)	.82	.26	.66	317
2.5	1.0	.8706 (19.4)	.2913 (3.6)	—.2068 (2.0)	.78	.29	.71	301
5.0	.1	.8568 (17.3)	.2302 (2.7)	—.1344 (1.2)	.85	.23	.58	332
5.0	.5	.8618 (17.9)	.2503 (3.0)	—.1585 (1.5)	.83	.25	.63	322
5.0	1.0	.8669 (18.7)	.2732 (3.3)	—.1856 (1.7)	.80	.27	.68	310

TABLE 3. — ESTIMATED CONSUMPTION FUNCTIONS BASED ON CROSS-COUNTRY DATA, USING NATIONAL INCOME AS INCOME MEASURE, FOR SELECTED VALUES OF δ_1 AND δ_{20}

$$C_t = \xi_1(z_t/z_{t-1})C_{t-1} + \xi_2 Y_t + \xi_3(z_t/z_{t-1})Y_{t-1}$$

Specified Values of δ_1, δ_{20}		Coefficients Estimated Directly			Derived Parameters			
δ_1	δ_{20}	ξ_1	ξ_2	ξ_3	α	β	λ	S^*
0	.1	.9619 (15.5)	.4124 (6.4)	—.3810 (4.5)	.85	.41	.92	356
0	.5	.9639 (17.3)	.4475 (7.5)	—.4193 (5.3)	.78	.45	.94	323
0	1.0	.9606 (19.3)	.4819 (8.7)	—.4525 (6.1)	.68	.48	.94	293
2.5	.1	.9610 (15.3)	.4081 (6.3)	—.3758 (4.4)	.86	.41	.92	361
2.5	.5	.9621 (16.4)	.4290 (6.9)	—.3984 (4.8)	.82	.43	.93	340
2.5	1.0	.9614 (17.6)	.4519 (7.6)	—.4218 (5.4)	.77	.45	.93	319
5.0	.1	.9608 (15.2)	.4064 (6.3)	—.3738 (4.3)	.86	.41	.92	362
5.0	.5	.9615 (16.0)	.4213 (6.7)	—.3899 (4.6)	.84	.42	.93	348
5.0	1.0	.9614 (16.8)	.4383 (7.2)	—.4075 (5.0)	.80	.44	.93	331

251

Data were assembled for each country for a recent census year for which household figures were available, and for the preceding year as well, as required for the fitting of equation (15). Income and consumption data from OECD (1970) were expressed in U.S. 1963 dollar equivalents. Household data and detailed population age-group data from United Nations (1971) were used to generate estimates of z for census years based, as before, on equation (13).[17]

Equation (15) was fitted by ordinary least squares,[18] using two alternative income measures. Because of lack of data for personal disposable income in many countries, gross national product and net national income were used. Table 2 reports results based on GNP for all 21 countries. Table 3 reports results based on national income for 20 of the 21 countries.[19]

Results for combinations of $\delta_1 = 0, 2.5, 5.0$ with $\delta_{20} = .1, .5, 1.0$ are shown in the tables. As before, the lowest value of S^* was obtained for $\delta_1 = 0$ and $\delta_{20} = 1.0$. That this result should be obtained with two such different types of data — Canadian time-series and international cross-section data — adds considerably to the weight of evidence suggesting an absence of household and population age effects on total consumption.

The time-series and cross-section results also agree in other important respects. The α, β, λ parameters of tables 2 and 3 are all within the expected ranges: as before, β and λ are in the 0–1 interval and α is less than $1/(1-\beta)$. The fits are good[20] and the t-ratios are generally high, especially when δ_1 and δ_{20} are assigned values which minimize S^*. It will be noted that the estimates of β, the marginal propensity to consume, are higher when the income measure is national income than when it is GNP. Furthermore, using personal disposable income, in the Canadian time-series case, yields even higher estimates of β, as one might expect.

VIII. Concluding Remarks

The most striking result of this study has been the accumulation of evidence to suggest that aggregate consumption is not affected directly by variations in average household size or in the age distribution of the population. The same result is obtained using two quite different sets of data — Canadian time-series and international cross-section data.[21] Of course, this does not imply an absence of direct household and age effects on particular *categories* of consumption, but only on the aggregate.

[17] Values for X_1, X_2, and H were estimated for the year preceding the census year by assuming that the percentage change in X_1 between the two years was equal to the percentage change in the population under 15, and that the percentage changes in X_2 and H were equal to the percentage change in the population 15 and over. Because of inconsistency from year to year in the published United Nations age-group data, percentage changes for population under 15 and 15 and over were based on data contained in OECD (1972). The latter data provide less age detail but incorporate adjustments to intercensal years and hence are more appropriate for measuring changes through time.

[18] The arguments of section V supporting the use of a two-stage estimation procedure apply also in the case of the cross-section version of equation (15). However, the choice of a single set of instrumental variables would be much more difficult in the case of 21 different countries than in the case of a single country. Accordingly, we have not attempted to go beyond the use of ordinary least squares in fitting the equation to cross-section data.

[19] The countries and the census years for which household data were assembled are as follows: Austria (1961), Belgium (1961), Canada (1966), Denmark (1965), Finland (1960), France (1968), Germany (1961), Greece (1961), Luxembourg (1966), Iceland (1960), Ireland (1966), Italy (1961), Japan (1965), Netherlands (1960), Norway (1960), Portugal (1960), Spain (1960), Sweden (1965), Switzerland (1960), United Kingdom (1966), United States (1960). Net national income figures were not published for Sweden in the OECD publication and this country was omitted in obtaining the results in table 3.

[20] Coefficients of determination were computed and were found to be greater than 0.999 in every case. However, the problem of interpreting $\bar{R}^2$ when the residuals do not sum to 0 remains. See footnote 15.

[21] It is worth noting that the population and household data that we have used reflect substantial variation in demographic circumstances. The Canadian data span the period from the late 1920's to the early 1970's. The crude birth rate fell during the 1920's and through most of the 1930's, rose rapidly in the 1940's, remained at a high level throughout the 1950's, and then fell sharply during the 1960's and early 1970's. Taking the period 1928–1971, the birth rate per 1000 population ranged from a high of 28.9 to a low of 16.8. Immigration was also highly variable over the period. Reflecting these changes, the average size of households ranged between 3.6 and 4.8 persons, and the percentage of population under 15 years of age ranged between 27.6 and 34.0. In the case of the international cross-section data, the range is 2.8 to 4.2 for household size and 20.9 to 34.8 for the percentage of population under 15. 252

It should be stressed also that our results do not imply that demographic factors have no influence on consumption and savings ratios. Other things constant (including income), a larger population implies more committed consumption and a lower savings ratio. Furthermore, we have been concerned only with the consumption function itself. But demographic factors affect other aspects of the economy, in particular the supply of labour and hence the levels of output and income. For a population of given size, a lower proportion of working adults implies a lower level of aggregate income. With a given level of committed consumption, based on the population size, this in turn implies a lower savings ratio, even though the consumption function itself contains no age-sensitive parameters. Thus our results are not inconsistent with the hypothesis that the savings ratio is influenced by birth rates or other demographic factors, or with empirical evidence in support of this hypothesis.[22]

In the context of our basic theoretical model, developed in section II, the apparent absence of household and age effects implies, in equation (2), that π_1 is 0 and that π_{2j} is the same for all j. Even so, individual household consumption will vary with the number of persons in the household. Over the life cycle of a typical individual household, its size will increase as children are born and decrease subsequently as the children grow up and leave home. Also, the income of the household head may vary with age. Thus, the savings ratio may be influenced at the microlevel by life-cycle factors, even though our theoretical framework differs from that of the "life-cycle hypothesis" developed originally by Modigliani and Brumberg (1954).

The Modigliani-Brumberg (M-B) hypothesis assumes a utility function in which current and future consumption levels enter as arguments. Ando and Modigliani (1963) tested the M-B hypothesis with aggregate data but did not allow for the effects of changes in the age distribution of the population or interest rates. Heien (1972) did allow for such effects, assuming a particular form for the M-B utility function. The basic difference between the M-B type of hypothesis and ours lies in the assumptions regarding an individual's perception of the future and of the significance of his personal wealth. The M-B assumption is that he anticipates his future circumstances and makes a rational allocation between current and future consumption. The assumption underlying our approach, as presented in section III, is that he bases his savings decision on his present consumption commitments, his view that personal wealth is both a protection against an uncertain future and a measure of social status, and his perception of the rate at which he adds to his wealth as a measure of his rate of achievement. This leads to the specification of a utility function in which utility is derived from the act of saving as well as from the stock of savings, and subsequently to a justification of the proportional savings rule embodied in our theory. Given the fact that the future is indeed uncertain, this assumption seems quite plausible. Whether or not it represents a better approximation to human decision-making than the "rational planning" assumption implicit in the M-B hypothesis is a matter of conjecture.

[22] For some specific evidence suggesting that the savings ratio is influenced by the proportion of dependents in the population, see the results of cross-country regression analysis reported by Leff (1969). See also the comments by Adams (1971) and Gupta (1971), and the reply to these comments by Leff (1971). Leff's analysis, and the discussion centering on it, are concerned primarily with the differences in savings ratios between developed and underdeveloped countries. Our own empirical results are based on data for developed countries only, it may be noted.

REFERENCES

Adams, N. A., "Dependency Rates and Savings Rates: Comment," *American Economic Review*, 61 (June 1971), 472–475.

Aigner, D. J., *Basic Econometrics* (Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1971).

Ando, A., and F. Modigliani, "The 'Life Cycle' Hypothesis of Saving: Aggregate Implications and Tests," *American Economic Review*, 53 (Mar. 1963), 55–84.

Denton, F. T., and B. G. Spencer, "A Simulation Analysis of the Effects of Population Change on a Neoclassical Economy," *Journal of Political Economy*, 81 (Mar./Apr. 1973a), 356–375.

———, *Analysing the Economic Effects of Changes in Fertility: A Simulation Approach*, McMaster University Department of Economics, Working Paper No. 73-06 (June 1973b).

Gupta, K. L., "Dependency Rates and Savings Rates:

253

Comment," *American Economic Review,* 61 (June 1971), 469–471.

Heien, D. M., "Demographic Effects and the Multiperiod Consumption Function," *Journal of Political Economy,* 80 (Jan./Feb. 1972), 125–138.

Leff, N. H., "Dependency Rates and Savings Rates," *American Economic Review,* 59 (Dec. 1969), 886–896.

———, "Dependency Rates and Savings Rates: Reply," *American Economic Review,* 61 (June 1971), 476–480.

Modigliani, F., and R. E. Brumberg, "Utility Analysis and the Consumption Function," in K. K. Kurihara, ed., *Post-Keynesian Economics* (New Brunswick, New Jersey: Rutgers University Press, 1954), 277–319.

Organization for Economic Cooperation and Development, *Labour Force Statistics 1959–1970,* Paris (1972).

———, *National Accounts of OECD Countries 1950–1968,* Paris (1970).

United Nations, *Demographic Yearbook 1970,* New York (1971).

———, *The Determinants and Consequences of Population Trends,* 1 (1973) (Department of Economic and Social Affairs, Population Studies no. 50).

254

Population Studies Vol XXVII, London 1973 -255-

Population Growth, the Dependency Rate, and the Pace of Economic Development

ALLEN C. KELLEY*

I. THE DEPENDENCY RATE AND ECONOMIC GROWTH

The analysis of population's impact on the economy has frequently been developed in the context of the dependency-rate argument. The dependency rate, typically measured as the proportion of the total population outside the labour force, is a summary statistic which is intended to capture the influence of a population's age structure on the process of economic growth. Unfortunately, there has been substantial confusion surrounding the economic interpretation of the dependency rate. This confusion derives from the fact that the dependency rate has been used as a proxy for *several* age-specific influences of population on economic growth. Additionally, for any one of these influences of population, the impact on the economy will be determined by the particular economic model within which the dependency rate is being analysed. In other words, depending on which age-specific economic aspect of the dependency rate is being examined, and depending on which economic model forms the basis of the analysis, it is possible that an increase in the dependency rate may be associated with either an increase or a decrease in the economy-wide growth of output per head. As a result, the widely used dependency-rate statistic may not be a particularly useful predictor of economic – demographic – growth-rate interrelationships unless the analyst makes explicit his underlying economic framework, and unless the particular economic influences for which the dependency rate is taken as a proxy are delineated.

The Dependency Rate and the Economy-Wide Saving Rate: An Overview

This paper focuses on only one – and possibly the least understood – of the dependency rate economic influences, the saving rate. It is frequently hypothesized that there is an *inverse* relationship between the rate of aggregate saving and the dependency rate. We will argue that this hypothesis is based on a somewhat restrictive set of assumptions relating to life cycle saving, and on a highly aggregative view concerning the economic behaviour of the dependent population. Moreover, we shall show that under plausible assumptions, the dependency rate may be *positively* associated with the aggregate rate of saving.

This argument is based on examining separately the two components of the dependent population – those who are in their pre-labour-force years, and those who are in their post-labour-force years. As the dependency rate changes, so does its *composition* between the proportion of the dependent population distributed between these two components. And, contrary to the typically employed assumption of equal dissaving rates for the various members of the dependent population, the dissaving rate may *not* be the same for the pre- and post-labour-force cohorts. Indeed, at certain stages of economic development, the negative saving rate of the elderly could well be greater than the rate for those who have not yet entered the labour force. Thus, decreases in the dependency rate – resulting, say, from a reduction in the fertility rate – may be associated with *decreases* in the economy-wide saving rate, given the possibility of a greater dissaving rate of the older as compared with the younger cohort of the dependent population, and coupled with a compositional change in the age structure of the dependent population as the aggregate dependency rate changes.

* A portion of this paper was presented at a National Bureau of Economic Research Seminar for the Commission on Population Growth and the American Future. I am grateful for the comments on an earlier version of this paper by Professors Richard Billsborrow, Glen G. Cain, Richard A. Easterlin, Charlie Hirsham, Nathaniel H. Leff, Peter H. Lindert, Soren T. Nielsen, Nathan Rosenberg, Julian L. Simon and Jeffrey G. Williamson. A grant from the Population Council facilitated the computational work of this study. 255

As an empirical illustration of these analytical arguments, two alternative cases are considered in Section III. First a numerical example is provided which examines the dependency-rate – saving-rate connections in a stable population framework. Secondly, the case of a specific transitional population is presented: the American dependency rate over the historical period 1870–1970, and population projections over the period 1970–2020. It is shown that the commonly hypothesized inverse relationship between the dependency rate and the aggregate saving rate may not hold throughout all of American experience.

Before we take up these specific relationships in detail, however, it would be useful to place the analysis in perspective by first considering the two competing economic frameworks within which the dependency rate is commonly analysed, and the four economic influences for which the dependency rate is taken as a proxy.

Competing Economic Models and the Dependency Rate

The dependency rate is frequently analysed within the alternative contexts of the Keynesian or the neo-classical models. The neo-classical framework, the most popular in the economic growth and development literature, focuses on the expansion of potential output growth. The model is supply-oriented and stresses long-run growth performance. Short-run problems of unemployment, or 'deficiency' of aggregate demand – highlighted by the Keynesian framework – are suppressed in the neo-classical model by an assumption of full employment. Indeed, given the full-employment assumption, the neo-classical model simultaneously explains potential *and* actual output expansion.

In justification of the full-employment assumption, it is sometimes asserted that problems of unemployment or underemployment are of a short-run, adjustment nature. The Keynesian framework may therefore be more appropriate for examining deviations from the long-run trajectory of economic performance. An important exception to this position was the concern of the stagnationists who, in the 1930's, argued that there may be tendencies toward a long-run deficiency of aggregate demand and toward secular unemployment.[1]

A debate on the relative merits of the competing models is outside the scope of the present paper. However, it is important to recognize that the interpretation of the economic impact of the dependency rate depends critically on the specific model employed. Indeed, the supply-oriented neo-classical model and the demand-oriented Keynesian model may frequently yield opposite predictions with respect to the impact of the dependency rate on actual output growth per head. For example, a decrease in the saving rate – due, say, to alterations in the dependency rate – will *stimulate* actual economic expansion if there is Keynesian unemployment; it will *dampen* output growth in the fully employed neo-classical economy. While we have elected to focus attention below on the neo-classical framework, the analysis could easily be expanded to account for Keynesian problems of demand deficiency. Moreover, the choice of models is *not* critical to our main argument. Our argument is not primarily concerned with the relationship between the dependency rate and economic growth, but rather with the relation between the dependency rate and one factor influencing growth, namely, the aggregate rate of saving. Thus, the interpretation of the saving rate – growth connection is not critical to the analysis. However, since the growth question does constitute a prime focus of the dependency-rate literature, it would seem appropriate at this point to give some attention to the problems involved in using the dependency-rate statistic in alternative models of economic growth.

Economic Behaviour and the Dependency Rate

Four age-specific influences of population on the economy, for which the dependency rate has been used as a proxy, are often considered in the literature. The first, and most commonly cited economic – demographic connection, relates to the role of the individual as an economic 'producer', an active (or

[1] Alvin H. Hansen, 'Economic progress and declining population growth', *American Economic Review*, 29 (1939), pp. 1–15.

256

potential) member of the labour force. An increase in the proportion of the population in the labour force (a reduction in the proportion of dependants) *ceteris paribus* leads to a greater growth in income per head. In this instance, the 'burden of dependency' is therefore taken as the proportion of the population below and above the customary, or socially determined, working age. This measure changes over time and is clearly arbitrary. Measurement difficulties are particularly acute in determining female participation rates, rural employment rates, and part-time labour force employment.[2] For example, between 1860 and 1940, the declining dependency ratio in the United States was offset by an increase in leisure.[3] The dramatic rise in female employment during this period also diminishes the usefulness of the dependency measure as an index of economy-wide manpower proportions.

A second connection between age and economic behaviour relates to aggregate consumption. In particular, the larger the proportion of the population who are children, the smaller the aggregate consumption demands. This effect, at the micro-economic level, is often measured in an adult-equivalent framework; that is, the child's consumption is considered as a percentage of an adult's.[4]

A third interaction between age and consumption relates not to aggregate consumption, but rather to its composition. Younger populations place a heavy demand on social services, particularly education; older populations place relatively strong demands on medical facilities. Educational and medical facilities are sometimes termed as relatively 'unproductive' investment goods; thus, higher dependency rates may be associated with lower investment in 'productive' capital – machines, producer durables, and so forth. This argument must be modified, of course, if the rate of return on 'human', as well as 'physical' capital, is incorporated into the analysis.

The fourth dependency-rate – economic connection – and the one highlighted in this paper – is associated with saving behaviour. This possibly represents the most fundamental connection between the dependency rate, as a demographic statistic, and economic behaviour. That is, saving represents the *combined* influence, or the *net* impact, of the first two life-cycle hypotheses discussed above – the aggregate production and the aggregate consumption behaviour of an individual over his lifetime. It is often alleged that the rate of aggregate saving and the dependency ratio are inversely related. The underlying theoretical basis of this argument rests on the twofold role of an individual over his life cycle: he consumes (draws upon society's scarce resources); he produces (contributes to the expansion of society's output). The difference between these activities yields a 'surplus', denoted as savings, which may be either positive or negative in value. In early and late stages of the life cycle, savings are negative since the consumption requirements of the individual exceed the value of his production (income); during the working years, savings are typically positive. (This set of arguments will be referred to below as the 'naive' model of the saving-rate – dependency-rate connection.)

II. QUALIFICATIONS TO THE 'NAIVE' SAVING-RATE – DEPENDENCY-RATE ARGUMENT

There are several modifications to the 'naive' dependency-rate argument as it pertains to saving which appear critical to an analysis of this demographic statistic as a proxy for economic behaviour. In considering these qualifications, it is analytically important to examine separately the saving behaviour of the two components of the dependency ratio: the pre-working population (children), and the post-working population (the retired). This bisection of the ratio, together with the empirical and theoretical impacts of each component of the dependent population on saving, represents a critical, yet seldom examined, element in dependency-rate analysis.

[2] Allen C. Kelley, 'Demographic change and economic growth: Australia, 1861–1911', *Explorations in Entrepreneurial History*, 5 (1968), pp. 228–235.

[3] Elizabeth W. Gilboy and Edgar M. Hoover, 'Population and immigration', in Seymour E. Harris (Ed.), *American Economic History* (New York, 1961).

[4] Glen G. Cain, 'Issues in the economics of a population policy for the United States', *American Economic Review*, 61 (1971), pp. 408–417.

257

 258

The Impact of Children on Saving: The Theory

The 'naive' dependency-rate analysis of the impact of children on saving is based on the notion that children yield relatively little income, yet consume significantly; they therefore exert a *negative* impact on the saving rate. The fallacy of this framework is that the relevant resource costs of children to society are not fully specified. Considering first only their impact on household saving, children do not exert their influence in isolation, but manifest it through the total family budget. The decisions regarding the size of the resource costs of a child are *not* made by the individual child, but rather by his parents; the *family* expenditure budget is thus the appropriate unit of analysis.

In contrast to frequent practice and 'common sense', the relevant dependency measure of child raising in this instance is *not* the sum of the various expenditures on children – or an adult-equivalent proxy – but rather the *net* increment to the household expenditure budget due to the presence of or an increase in the number of children.[5] There will be a substitution in the household budget between expenditures on children and on other goods and services. Depending on the nature of this allocation, the saving-rate 'cost' of raising an additional child can be positive or zero. For example, Henderson hat found that family size has a small effect on total expenditure. The impact on the *structure* of the budget is much larger.[6] Moreover, if the presence of an additional child induces more work by parents, or if the child contributes directly to family income, the net impact on the household saving rate may even be positive.[7] Over time, of course, the measured costs of child raising will depend not only on the prices of other goods and services relative to the costs of raising children, but also on the relative 'tastes' for children *vis-à-vis* alternative uses of the household's resources.[8]

As an important qualification to the analysis of the impact of children on the family budget, it should be noted that the argument becomes more complex if the saving-rate impact of children on other sectors of the economy is considered. Simon, after exploring the matter in some detail, concludes: 'The effect on *saving* of the social spending for children and other children's services is most unclear . . . there is no basis on which to estimate either the elasticity of spending on schools, or even harder, the extent to which the incremental expenditures on schools substitute for other social investment rather than causing new tax levies'.[9] Plausible hypotheses could be introduced on the impact of children on government and business saving which would either reinforce or overturn the findings presented for the household sector. The issue remains an empirical one. Below, we will examine empirically the *combined* impact of children on household, government, and business saving rates. For the one case considered, children exert a negative impact on saving, although this negative influence greatly diminishes as countries become more economically advanced.

The Impact of Children on Saving over Time and Space

The impact of children on family saving will probably vary with the level of economic development. In very poor countries or regions where the family budget approaches subsistence levels, substitution

[5] Ansley J. Coale and Edgar M. Hoover, *Population Growth and Economic Development in Low-Income Countries* (Princeton, 1958) pp. 261–270; K. Gupta, 'Dependency rates and savings rates: Comment', *American Economic Review*, **61** (1971), p. 471.

[6] A. M. Henderson, 'The cost of a family', *The Review of Economic Studies*, (1949–50).

[7] R. A. Easterlin, 'The effects of population growth on the economic development of developing countries', in R. D. Lambert (Ed.), *The Annals of the American Academy of Political Science* (Philadelphia, 1967); Nathaniel Leff, 'Dependency rates and savings rates', *American Economic Review*, **59** (1969), p. 887; and Nassau A. Adams, 'Dependency rates and savings rates: Comment', *American Economic Review*, **61** (1971), p. 472.

[8] Gary S. Becker, 'An economic analysis of fertility', in National Bureau of Economic Research: *Demographic and Economic Change in Developed Countries* (Princeton, 1960), pp. 209–240; Robert J. Willis and Warren Sanderson, *Economic Models and Fertility Behavior: Some Examples and Implications* (New York, 1970), p. 26; R. A. Easterlin, 'Towards a socioeconomic theory of fertility: A survey of recent research on economic factors in American fertility', in S. J. Behrman *et al.* (Eds.), *Fertility and Family Planning, A World View*, (Ann Arbor, 1969), pp. 127–156; and R. A. Easterlin, 'On the relations of economic factors to recent and projected fertility changes', *Demography*, 3 (1966), pp. 131–153.

[9] Julian L. Simon, 'The Influence of Population Growth on per Worker Income in Developed Economies' (University of Illinois, 1971), p. 10.

258

induced by an additional child may be largely against saving; that is, there will be increased family consumption with little incremental income, given the under-utilization and low productivity of labour. Kuznets speculates that the negative saving impact of larger families is likely to be much larger in less developed countries, and that in developed countries, children are more likely to be at the expense of consumption and leisure that of saving. He provides limited evidence to support this position.[10] In contrast, Gupta postulates that when income levels are very low, '. . . there is no margin left for saving . . . it simply means that people are only sharing poverty'. He continues, '. . . demographic factors, like the dependency ratio, become operative and significant only when the per capita income of the *working population* reaches a level where it can provide more than a minimum standard of living, thus generating potential savings'.[11]

For countries with significant economic growth, and population well above subsistence levels, yet still primarily rural, the impact of an additional child on family saving is unclear. This situation may characterize the United States during the nineteenth and early twentieth centuries. The incremental resource costs of an additional child may be low, given the substitution against other forms of consumption and the scale economies of child raising; additionally, the contribution of children to family income may be positive and even exceed the incremental consumption expenditures.

In summary, the untested conjectures outlined above are sufficient to illustrate that on theoretical grounds the simple postulated impact of children as a 'burden' on society's economic resources available for capital formation is ambiguous in direction, and in its quantitative magnitude. Moreover, the size of the impact (whether positive or negative) may change systematically as development takes place. Unfortunately, the empirical studies of the relevant relationships are very limited. We shall review some of these studies below and provide additional evidence, drawn from American historical experience, to attempt to provide a preliminary quantification of the resource costs to society of children.

The Impact of the Retired on Saving

The dependency-rate argument typically focuses on the resource costs to society of a young population due to high birth rates. But the argument is two-edged. A low birth rate will yield an older population the resource costs of which, in precisely the same theoretical framework, also represent a social dependence.[12] Indeed, these resource costs could dramatically increase in the future if serious attention were to be directed to the acute social problem of poverty among the elderly, a component of the population the high consumption standards of which are increasingly being frustrated by rising medical costs, taxes, and inflation.[13]

III. SPECULATIONS ON THE 'EFFECTIVE' DEPENDENCY RATE

The dependency ratio, as typically measured, abstracts from a differential age-specific rate of saving of the dependent population. However, saving rates are likely to vary between the pre-working and the post-working ages of the population; equally important, as development takes place, the simple dependency ratio fails to capture the 'effective' impact on savings for which it is taken as a proxy. The 'effective' dependency rate should be constructed so as to *weight* the two components of the dependent population – the children and the aged – according to their relative impact on saving.

The importance of this modification to the dependency-rate statistic can be seen by examining a recent study by Nathaniel H. Leff, who has estimated the differential impact of the 0–14 and the 65 + age groups on the savings rates for both the underdeveloped (LDC) and the Western developed

[10] Simon Kuznets, 'Population change and aggregate output', in National Bureau of Economic Research: *Demographic and Economic Change in Developed Countries* (Princeton, 1960), pp. 331–32.

[11] Gupta, *loc. cit.*, in footnote 5, p. 471.

[12] W. Eizenga, *Demographic Factors and Saving* (Amsterdam, 1961), pp. 6–21.

[13] Leff, *loc. cit.*, in footnote 7, p. 889.

259

countries (DC).[14] While Leff's model is only one of many possible formulations of the relationship, it is generally illustrative of the general principles involved. The tentative nature of Leff's findings has been pointed out by several scholars.[15] Recalculations by them have changed somewhat the impact of D_1 and D_2 levels, although the *differential impacts* of D_1 and D_2 still hold. It is this differential impact of the two dependency rate effects which is emphasized in the present study. Leff employs the model

$$S/Y = \alpha(Y/N)^{\beta}(g)^{\gamma}(D_1)^{\delta}(D_2)^{\epsilon}, \tag{1}$$

where Y/N = income per head, g = average annual growth rate of income per head, D_1 = percentage of population in the 0–14 age group, and D_2 = percentage of population in the 65+ age group. As countries move from low average incomes of around \$250 to those in the \$1500 range, Leff finds that the negative impact on savings of the 0–14 dependence component declines from −1·2297 to −0·4324; there is also a slight rise in the savings impact of the older age group from −0·4455 to −0·4916. (These specific figures represent the statistically significant estimated parameters δ and ϵ.) While his calculations apply largely to aggregate savings parameters, the theoretical interpretation still rests with the impact on individual saving.[16] The main exception to the micro-economic interpretation lies, of course, in the provision of population-sensitive social and educational services. The two sets of weights which will be used in the numerical examples below are summarized in Table 1. It is apparent that account must be taken of the 'stage' of economic development in selecting weights to be applied to the two components of the dependent population in determining the 'effective' dependency-rate impact on saving.

TABLE 1. *Weights used to assess the dissavings impact of two cohorts of the dependent population by level of income per head*

Dependent population cohort	LDC weights ($Y/N \simeq$ \$250)	Developed country weights ($Y/N \simeq$ \$1500)
0–14 (D_1)	−1·23	−0·43
65+ (D_2)	−0·45	−0·49

SOURCE: See text.

The way in which these weights enter into the calculation of the 'effective' dependency-rate effect on saving is readily illustrated by considering the determinants of changes in Leff's estimated economy-wide saving rate. If we abstract, for the moment, from changes in Y/N and g (and assume they, with α, are a constant K), then differentiating (1) with respect to time (assuming all variables are time-dimensioned), we find that

$$\frac{d\,(S/Y)}{dt} = K[\delta(dD_1/dt)/D_1 + \epsilon(dD_2/dt)/D_2]. \tag{2}$$

We assume that δ and ϵ are negative; that is, both components of the dependent population exert a 'burden' on savings. Moreover, we assume that as the dependency rate increases, D_1 increases and D_2 decreases; that is, $dD/dt > 0$ implies $dD_1/dt > 0$, $dD_2/dt < 0$.

Trends in the saving rate depend on the relative rate of change of the two dependency-rate components – each moving in an opposite direction – and on their relative weights, δ and ϵ. The rate of

[14] *Ibid.*, pp. 886–895.

[15] Richard E. Billsborrow, 'Dependency rates and aggregate saving rates: Corrections and further analyses' (Mimeographed, 1973). Also Gupta, *loc. cit* in footnote 5.

[16] W. Eizenga, *op. cit.*, p. 74; A. C. Kelley, 'Demographic changes and American economic development', in *Research Reports of the Commission on Population Growth and the American Future: Economic Aspects of Population Change* (Washington, D.C., 1972).

260

saving can therefore move in either direction; additionally, and possibly more relevantly, changes in the saving rate may be non-linearly related to changes in the dependency rate. Indeed, in countries with very high mortality rates, where D_2 could initially be at very low levels, changes in D_2 might dominate the total effect, even in the case, where $\delta > \epsilon$. For the low-income countries, however, the most likely case is one in which $\delta(dD_1/dt)/D_1$ would dominate $\epsilon(dD_2/dt)/D_2$. In the high-income countries, in contrast, the opposite may well prevail. Moreover, there may be a tendency for the relationship between the aggregate dependency rate, D, and the direction and magnitude of its impact on saving, to vary *systematically* as economic development takes place.

The most important qualification to this argument lies, of course, with the relationship between the dependency rate and Y/N and g. Given the above argument, which demonstrates a potential indeterminacy between the direction of association between the saving and dependency rates, it is by no means clear what this relationship will be. Indeed, not only is the *direction* of the association between D and S/Y ambiguous, but, as argued above, the impact of changes in S/Y on g will vary depending on whether one employs a Keynesian or a neo-classical framework. As a result, for present purposes we shall explicitly recognize that our analysis is of a partial-equilibrium nature and confine our observations to the *ceteris paribus* impact of D on S/Y.

A Numerical Illustration Employing a Stable Population Framework

A convenient way to illustrate the effect of changes in the dependency rate is to examine the saving-rate implications of the age distributions for several alternative stable populations. Examples are provided in Table 2. The stable populations are taken from Coale and Demeny,[17] and represent alternative intrinsic rates of population increase and three levels of mortality. (The West tables for males are utilized.) While the number of possible numerical illustrations employing alternative stable populations is unlimited, the ones shown here are sufficiently representative to illustrate most of the analytical points developed in this study.

The trends reflected in Columns 1 and 3, which provide alternative representations of the dependency ratio for several rates of population growth, are well documented in the literature and require no comment. Column 2, which presents a measure (D_2/D) of the *composition* of the dependent population at various levels of dependency, represents a more interesting analytical feature underlying the 'effective' dependency-rate calculations below. Higher dependency rates are associated with a lower proportion of the post-labour-force age group; moreover, the level of D_2 varies inversely with the level of mortality.

The key results, presented in Columns 4 and 5, represent changes in S/Y at alternative levels and compositions of the dependency rate. Index numbers of S/Y are presented to facilitate comparison. Using the LDC weights, it is seen that, with one exception, the widely discussed inverse relationship between the dependency rate and the saving rate holds. In the high-mortality case, the rapid percentage changes in the D_2 cohort outweigh the fact that the absolute value of δ is almost three times that of ε. For practical purposes, however, this high level of mortality is not typical, and the more general finding of an inverse dependency-rate – saving-rate relationship in low-income countries prevails.

As one moves to the case of the developed countries (Column 5), the relationship alters dramatically. There is a *positive* association between the 'effective' dependency rate and S/Y. Moreover, the magnitudes of the changes in S/Y are equal to or greater than those observed in the LDC case. These results clearly indicate the need to apply extreme caution in generalizing across stages of development, and across countries, on the relationship between the dependency ratio and the economy-wide saving rate.

The changes in the 'effective' dependency rate may appear at first glance to be implausibly large. Historically, variations in the average saving rate in excess of 25% (say an index change of 100 to 125) are seldom observed. In defence, it should be noted that the comparisons in Table 2 relate to cases where the rate of population growth varies between 3% and 0. In terms of calendar time, a change of this magnitude may reflect decades or centuries. Additionally, the specific cases where the changes in 261

[17] Ansley J. Coale and Paul Demeny, *Regional Model Life Tables and Stable Populations* (Princeton, 1966).

the saving rate are highest in Table 2 are typically 'implausible' situations; for example, the 'developed country' weights employing a high-mortality assumption. Finally, recall that the experiments in Table 2 reflect partial-equilibrium relationships. *Actual* changes in the saving rate depend on many factors, some moving in the opposite direction to the trends found in Table 2.

Even with these several qualifications, however, it is notable that the changes in the saving rate are, in some instances, large. Demographic factors, and, in particular, changes in the age distribution, may constitute an important determinant of the economy-wide saving rate, and through this influence, may exercise a quantitatively significant impact on the rate of economic growth.

TABLE 2. *Illustrations of the 'effective' saving-rate impact of the dependent population for two alternative cohort-specific dissaving effects, and for three stable populations.*

r*	Dependency rate (%) (0–14, 65+) Total population (1)	Dependent population in 65+ age group (%) (2)	Dependency-rate index ($r(0)=100$) (3)	'Effective' dependency rates for alternative weights† index ($r(0)=100$) LDC (4)	DC (5)
		Level 1 (high mortality)			
0·00	40·13	5·23	100·0	100·0	100·0
0·01	44·82	2·90	111·7	105·1	119·4
0·02	50·62	1·56	126·1	111·3	143·7
0·03	56·24	0·82	140·1	123·7	178·5
		Level 12 (medium mortality)			
0·00	34·01	20·00	100·0	100·0	100·0
0·01	38·08	13·50	112·0	91·4	115·5
0·02	43·00	7·51	126·4	89·2	129·0
0·03	48·50	4·08	142·6	91·5	153·1
		Level 24 (low mortality)			
0·00	35·94	44·38	100·0	100·0	100·0
0·01	37·50	29·07	104·3	83·5	106·6
0·02	40·83	17·39	113·5	75·6	118·7
0·03	45·40	9·78	126·3	73·6	137·5

SOURCE: Coale and Demeny, 1966, p.122.

*r = intrinsic rate of population growth.

† The average saving rates in Columns 4 and 5 were computed as follows. For each of the two sets of weights in Table 1 (LDC, DC), there is a corresponding set of mean values for the variables attached to the estimated α, β, and γ parameters in Leff's equations. (The mean values of these variables were provided for me by N. Leff.) The equations for S/Y were next solved for the alternative dependency-rate proportions, 0–14 and 65+, corresponding to the indicated stable population growth rates. This yielded values of S/Y corresponding to alternative dependency rates. The S/Y's were then computed as an index with $S/Y_{r(0)} = 100$.

The American Dependency Rate and the 'Effective' Impact on Saving

While the stable population framework illustrates well the general properties of the saving-rate–dependency-rate connections considered in this paper, an equally relevant application of the model relates to specific populations. In this section we shall examine the course of the American dependency rate, and inquire whether variations in this statistic have acted as a contributing or an inhibiting factor to rates of saving. Recall that high-birth-rate countries, the developing countries of to-day being a case in point, are characterized by high dependency ratios; these, in turn, are frequently considered as one of the constraints on economic growth.[18] Since early American development was also characterized by 262

[18] Charles P. Kindleberger, *Economic Development* (New York, 1965), Chaps. 6, 15; Benjamin Higgins, *Economic Development* (New York, 1959), pp. 17–19.

high birth rates and a very young population, the question arises whether, historically, the decline in the dependency rate has acted as a contributing factor to American economic progress.

The solid line in Fig. 1 displays an index of American dependency rates over the last century, together with rates projected up to the year 2020. (For simplicity, the dependency rate in Fig. 1 represents a single index, ages 0–14, 65+. While the measure (0–19, 65+) was also examined, it was found that the conclusions below are not sensitive to the particular measure employed.) The projections are for a family size of 2·11 children. Larger family sizes would, of course, result in higher dependency rates.

With the exception of the post-war baby boom, there is a decline, over most of the period, in the dependency rate and in the population's youth. The underlying calculations reveal that this trend results both from a decline in the pre-working ages and a significant rise in the retirement cohort. The overall dependency ratio falls between six and nine percentage points while the reduction in the pre-working cohort ranges between eleven and 15 percentage points. The decline in the overall dependency rate is thus associated with a marked 'compositional effect' in which the weight is shifted from the young to the old. To the extent that the magnitude of the dependency effect differs between the young and the elderly – a point developed at length in the preceding section – the analysis of the dependency rate must consider both the overall trend and the compositional effect associated with it.

The broken and dotted lines in Fig. 1, which employ the LDC and the 'developed country' weights, respectively, display changes in the average saving rate due to alterations in the dependency

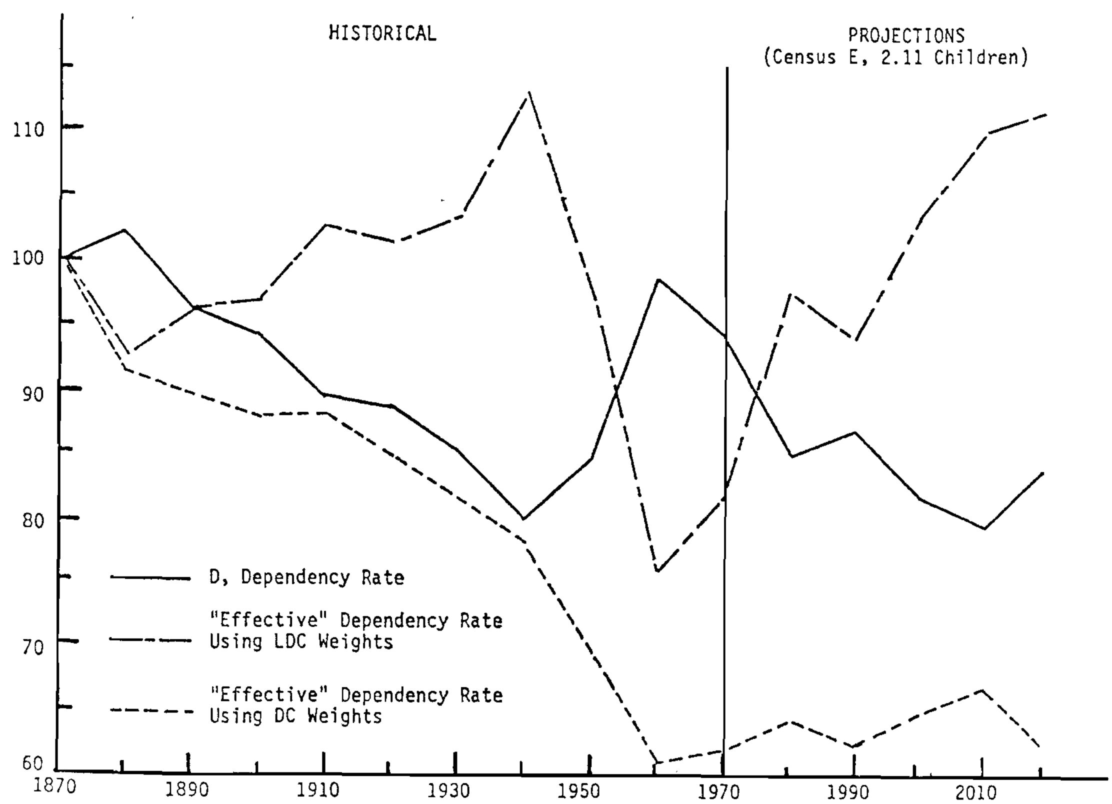


FIGURE 1. *Index of dependency rates under alternative assumptions regarding 'effective' impact on saving: U.S. Historical Dependency Rates*, 1870–1970; *Series E* (*Census*) *Projections*, 1970–2020.

263

ratio. They represent two alternative hypotheses capturing the 'effective' dependency impact on th economy-wide rate of saving. The critical importance of examining the age-compositional chang taking place within the aggregate dependency ratio, as development takes place, is apparent. Th alternative hypotheses on the relative negative impact of age-specific saving rates reveal marked di ferences in the resulting series.

The LDC weights (the broken line) present a story consistent with the widely-hypothesized inver relationship between the dependency rate and the rate of saving. The solid and broken lines general move in opposite directions. From 1870—1940, saving rates increased as the 'burden of dependenc diminished. Moreover, should the economy in the next decades move toward zero population growt the lower rate of dependency would result in a significant rise in the saving rate to a level equalled on once before in the period under consideration. Employing a neo-classical interpretation of Americ economic growth, one could conclude that the long-term reduction in the birth rate, and its continu decline in the future, represents a notable contribution to the pace of American economic progress.

Or does it? Turning to the 'effective' dependency-rate series employing the 'developed countr weights (the dotted line), a different scenario is revealed. Over the period 1870–1940, the reduction the dependency ratio was associated with an approximately equal *decline* in the economy-wide savi rate; economic growth was thus *inhibited* by this age-distributional aspect of long-term population growt As to the future, while one might expect a rise in the saving rate due to a reduction in the dependen ratio, the quantitative impact of this effect may be relatively slight; from 1970–2020, the 'effectiv dependency rate is almost flat.

A choice between the competing hypotheses for the United States is difficult. Even though nin teenth-century America was in the process of early industrialization and growth, its income per he was substantially above the level of the LDC's of to-day. On the other hand, the 'developed countr weights would represent an excessively high average income. Our best guess would be a set of weigh which result in a trend somewhere between the two 'effective' dependency-rate series presented i Fig. 1. For the future, the 'developed country' weights appear most appropriate. If this set of conje tures were accepted, then one could conclude that variations in the dependency ratio have exerted on a slight impact on the economy-wide rate of saving in the U.S. Historically, say over the period 187 1940, an average of the broken and dotted lines would result in a nearly flat 'effective' saving-rate seri moreover, after 1970, a similar finding is revealed. It is thus quite possible that even though the U has historically experienced a quantitatively large secular decline in the rate of population growth, t positive or negative impact on savings, if we focus on the dependency-rate effect, has been slight. Wh the dependency-rate model, as commonly formulated, may provide a useful proxy for saving-r trends in the less developed countries of to-day, its role in interpreting trends in the United States a in other economically advanced economies may be limited indeed.

264

VOLUME 29 NUMBER 1 MARCH 1975

POPULATION STUDIES

A Journal of Demography

Editors: D. V. GLASS and E. GREBENIK

PRICE: £2.50 (U.S. $7.50)

Published by:

THE POPULATION INVESTIGATION COMMITTEE
London School of Economics: Houghton Street, Aldwych, London WC2A 2AE

265

The Impact of Children on Household Saving: Age Effects Versus Family Size

THOMAS J. ESPENSHADE

In a recent paper in this journal,[1] Allen Kelley argued that the relation between the dependency burden and the aggregate savings rate is potentially ambiguous, not only in magnitude but also in direction, and that given alternative sets of plausible assumptions, the impact of children on household savings behaviour could be positive, negative, or non-existent. If Kelley is correct, then the true nature of the dependency rate-savings rate relation cannot be determined on the basis of theoretical considerations, but must be resolved empirically. Kelley contends that further tests of this relation are needed at the household level, and adduces the speculation that whatever the resulting findings, they might be specific to the particular stage of economic development in question.

In another place, Kelley presents the results of one micro-investigation.[2] It is based upon an 1889 sample, conducted in ten states of the United States, of the incomes, expenditures, and savings of 1,956 households whose heads were employed in either the iron, coal, or steel industry. In this study, Kelley allows for both direct and indirect effects of children on household savings rates. Holding the level of income constant, he finds family size to be negatively related to the household savings rate. A second equation, however, shows that the presence of children exerts a positive influence on family income and thus that, in this indirect way, children tend to increase household savings rates. Combining the direct and indirect impact, Kelley finds that the addition of the first child to the family has no appreciable effect on savings rates, that the addition of the second causes a notable increase, and that the marginal effect on savings rates of subsequent children is negative. These results, using as a control the mean income of the total renter sample, are presented in the upper portion of Table 1.

TABLE 1. *Impact of family size on household savings rates*

Source	Household savings rates (per cent)						
	Number of Children						
	0	1	2	3	4	5	6+
Kelley's calculations*							
Direct effect	7·91	7·91	7·91	3·75	0·21	−4·13	−5·55
Indirect effect	0·00	0·00	3·82	3·19	2·90	5·78	8·20
Total	7·91	7·91	11·73	6·94	3·11	1·65	2·65
Author's recomputations using all least-squares estimates†							
Direct effect	8·20	8·17	6·09	4·23	0·49	−3·85	−5·26
Indirect effect	0·00	1·55	3·83	3·21	2·92	5·81	8·24
Total	8·20	9·72	9·92	7·44	3·41	1·96	2·98

* Table 2, p. 26, in Allen C. Kelley, 'Demographic Changes and American Economic Development: Past, Present, and Future', *loc. cit.* in footnote 2.

† *Ibid.*, Table 1, p. 25.

[1] Allen C. Kelley, 'Population Growth, the Dependency Rate, and the Pace of Economic Development', *Population Studies*, November 1973, 27, 3, pp. 405–414.

[2] Allen C. Kelley, 'Demographic Changes and American Economic Development: Past, Present, and Future', in Commission on Population Growth and the American Future, *Research Reports, Volume II, Economic Aspects of Population Change* (Washington, D.C., 1972), pp. 9–44.

265

Upon closer examination, Kelley's results appear to turn on somewhat arbitrary methods of computation. To derive the magnitudes in Table 1, Kelley assigned a value of zero to those coefficients which were not significantly different from zero in the original equations.[3] If, instead, he had retained the coefficients as estimated, his conclusions would have been *dramatically* altered (compare the lower half of Table 1). In this case, it is the first child which exerts a positive impact on savings rates, whereas the addition of a second child has hardly any effect. Despite these differences, the general impression which emerges from Table 1 is that based on Kelley's 1889 sample, additional children exert a fairly sizeable and non-linear influence on household savings rates; the addition of the first or second child causes savings rates to increase, whereas subsequent birth orders result in declines.

EVIDENCE FROM THE 1960–1961 CONSUMER EXPENDITURE SURVEY

The 1960–1961 Survey of Consumer Expenditures conducted by the U.S. Bureau of Labor Statistics constitutes another rich data source for analysing the impact of children on household savings behaviour. In contrast to the sample Kelley uses, the 1960–1961 CES is a national sample of families at all levels of income and has the additional advantage of including the ages of children as well as their number. The following equation was estimated from 3,888 urban consumer units which were either of the husband-wife-with-own-children type where all children were under age 18 or of the husband-wife-with-no-others-present type where the age of the head did not exceed 45 years. Using household annual net savings (S) as the dependent variable,[4] we obtained.

$$\begin{aligned} S = &-5{\cdot}3143 + 0{\cdot}2574\ Y + 0{\cdot}0720\ F - 2{\cdot}1952\ A + 0{\cdot}1131\ Y^2 \\ &(1{\cdot}6078)\ (0{\cdot}1180)\quad (0{\cdot}8458)\quad (0{\cdot}8440)\quad (0{\cdot}0031) \\ &+ 0{\cdot}0141\ F^2 + 0{\cdot}3304\ A^2 \qquad R^2 = 0{\cdot}539 \\ &(0{\cdot}0814)\quad (0{\cdot}1506) \end{aligned} \tag{1}$$

where Y equals annual income after tax, F equals family size inclusive of the parents, and A is a proxy for the age of the children.[5] The standard errors of the estimated coefficients are shown in parentheses.

Inspection of Equation (1) reveals that income exerts a *positive and statistically significant* impact on household saving. In contrast to Kelley's finding, however, the family size variable has no significant effect on savings in either its linear or in its squared form. The corresponding t-ratios leave little doubt about this conclusion. The new and interesting discovery which emerges from an analysis of the 1960–1961 CES data is that although family size has no effect on household saving, the age of the children does. Both the coefficients on A and on A^2 are statistically significant at the five per cent level. Moreover, the direction and magnitude of the impact on household saving depends upon the average ages of the children. Calculating the partial derivative of savings (S) with respect to age of the children (A) in Equation (1), we have

$$\frac{\partial S}{\partial A} = -2{\cdot}1952 + 0{\cdot}6608\ A. \tag{2}$$

[3] The coefficients in question are those of C_1, C_2, and A_3 in Equation 1 and that of C_1 in Equation 2. See Allen C. Kelley, 'Demographic Changes and American Economic Development: Past, Present, and Future', *loc. cit.* in footnote 2, Table 1, p. 25. It is not unusual to drop from regressions those variables which are not statistically significant, but it is customary practice then to re-estimate the equations with the insignificant variables omitted. The fact that the insignificant variables in this case are components of dummy variables greatly enhances the difficulties of re-estimation.

[4] Annual net savings are defined as the net change in household assets minus liabilities during the year.

[5] Savings were measured in hundreds of dollars and income in thousands of dollars. The variable A was such that $A = 0$ if there were no children, $A = 1$ if the oldest child was under six, $A = 2$ if the oldest child was six to eleven and the youngest was under six, $A = 3$ if all children were six to eleven or if the oldest was twelve to seventeen and the youngest was under six, $A = 4$ if the oldest child was twelve to seventeen and the youngest was six to eleven, and $A = 5$ if all children were twelve to seventeen.

266

Thus, for values of $A < 3{\cdot}32$, an increase in the mean age of children causes a reduction in household savings, whereas increases in this age for values of $A > 3{\cdot}32$ augment household saving. In terms of the scheme we are using here, a value of $A = 3{\cdot}32$ corresponds to an average age of children of about 9·5 years. *In other words, children have the greatest positive impact on family savings when they are nearing the completion of high school.* One possible interpretation of this interesting finding is that parents at this time are beginning to anticipate the costs of a college education and attempting to save accordingly. This conjecture would appear to be corroborated by Turchi's finding that only when children reach high school age does their presence exert a positive influence on the mother's market work. Turchi interpreted this effect 'perhaps as a response to the approaching onset of college.'[6]

Possibly, an alternative interpretation of these results is that they reflect an increase in the proportion of parents who are saving for their children's education, and not necessarily an increase in the amount saved by individual families. This pattern of behaviour was in fact detected by a University of Michigan team who found that of those parents who had children in college between 1955 and 1960 and who had specifically set aside money for their children's education, eleven per cent had begun saving twenty years or more before the children went to college, 24 per cent had begun saving 15 to 19 years in advance, 26 per cent between ten and 14 years, 17 per cent between five and nine years, and 22 per cent had begun saving less than five years in advance.[7] The Michigan study, unfortunately, does not examine the amount saved by individual families and how this varies with the age of the children.

Kelley has emphasized that the impact of children on household saving behaviour could plausibly depend upon the level of economic development. This hypothesis would appear to be supported by a comparison of some of the available evidence. Using data from 1889, Kelley found that family size had a pronounced effect on family savings rates; this effect was positive for the first or second child but became negative with the addition of third and subsequent children. The evidence from the 1960–1961 Consumer Expenditure Survey suggests, however, that by the time the United States had achieved a substantially higher level of living, family size had ceased to have much influence. Eizenga's study[8] fits neatly into this pattern; using data from an intermediate period, he obtains an intermediate result. His analysis of the 1950 U.S. Consumer Expenditure Survey data shows that the impact of children on family saving is relatively slight, especially for children beyond the first.[9] The tentative hypothesis which is suggested by this comparative evidence, therefore, is that the importance of family size *per se* as a determinant of household saving behaviour diminishes as a society reaches higher stages of economic and social development, although other aspects of the presence of children (for example, their ages) may affect savings rates.

The purpose of these comments has been threefold: (1) to point out that Kelley's assessment of the effect of family size on household savings rates is sensitive to his manipulation of the original equations and is altered when more conventional procedures are followed, (2) to argue that in modern American society the savings behaviour of families may be tied more to the ages of the children than to their number, and (3) to provide some comparative evidence which tends to substantiate Kelley's conjecture that the way and extent to which children influence household savings patterns depends upon the underlying level of economic development.

[6] Boone A. Turchi, 'The Cost of Child-Rearing: Elements of a Socioeconomic Model of Fertility', Paper presented at the annual meeting of the Population Association of America, Toronto, Ontario, April 13–15, 1972, p. 20.

[7] John B. Lansing, Thomas Lorimer, and Chikashi Moriguchi, *How People Pay For College* (Survey Research Center, Institute for Social Research, The University of Michigan, September 1960), p. 43.

[8] W. Eizenga, *Demographic Factors and Savings* (Amsterdam: North-Holland Publishing Company, 1961).

[9] Interpreting a family of size two as a childless couple, etc. and using a figure of \$4,000 to approximate the mean income after tax for the sample, we have the following savings rates implied by Eizenga's data: for childless couples, 4·24 per cent; for one-child families, 0·04 per cent; for two-child families, 0·42 per cent; for three-child families, −0·36 per cent; and for couples with four or more children −1·58 per cent.

266

Projected Consumption Patterns for a Stationary Population

DONALD EILENSTINE AND JAMES P. CUNNINGHAM

I. INTRODUCTION

Recently economists have demonstrated a renewed interest in the population problems of a mature society. This revival of interest has been sparked by the general recognition of the relationship between population growth and environmental problems and has led to general acceptance of the proposition that '... a zero rate of population growth is the only equilibrium rate that can be sustained'.[1] Consequently the literature produced during the discipline's last period of similar concern, a period running from the late 1920's through the 1940's, needs re-examination.[2] At that time economists were primarily occupied with the implications of a declining rate of population growth and most anticipated the arrival of a stationary population within the foreseeable future. For most of the economists of this earlier period the onset of a stationary or declining population was fraught with dangers for mature capitalism.

The literature of this period, including the classic Presidential Address by Alvin Hansen before the American Economic Association, contains a number of suggestions about the probable impact of zero population growth on the percentage of income saved and on patterns of consumption. Whether the issues are approached from the macro-economic or the micro-economic perspectives, consumer behaviour occupies the primary place. If a stationary population is to display significantly different configurations it must mean that consumer behaviour, in the aggregate, is different in the two societies. Lionel Robbins and G. C. Billing wrote early articles that suggested, in the case of Billing quite strongly, that a stationary population with a larger proportion of older people would exhibit a significantly different pattern of consumption from a growing population.[3] Hansen was even more explicit on this point. He stated that '... the rate of population growth must necessarily play an important role in determining the character of the output; in other words, the composition of the flow of final goods'.[4] Hansen went on to argue that '... a shift from a rapidly growing to a stationary or declining population may so alter the composition of the final flow of consumption goods that the ratio of capital to output as a whole will tend to decline'.[5] A later article by A. R. Sweezy contained similar arguments.[6]

In methodological terms the earlier discussion of the impact of a stationary population upon the economy seems somewhat far removed from what would appear to be a reasonable approach to the same questions to-day. Though the impact of zero population growth upon the economy would appear to demand systematic statistical investigation the earlier literature is strangely non-quantitative. In place of systematic statistical information one finds guesses and speculation. Some of the guesses were repeated sufficiently often so that they virtually became enshrined in the

[1] Glen G. Cain, 'Issues in the economics of a population policy for the United States', *American Economic Review*, **61** (June 1971), p. 409.

[2] Vincent Tarascio, 'Keynes on the sources of economic growth', *Journal of Economic History*, **31** (June 1971), pp. 429–444.

[3] Lionel Robbins, 'Notes on some probable consequences of the advent of a stationary population in Great Britain', *Economica*, **9** (April 1929), pp. 71–82, and G. C. Billing, 'Some economic effects of a stationary population', *Economic Record*, **11** (December 1935), pp. 167–175.

[4] Alvin H. Hansen, 'Economic progress and declining population growth', reprinted in *Readings in Macroeconomics* (H. G. Mueller, Ed.) (New York, 1966), p. 270.

[5] *Ibid.*

[6] A. R. Sweezy, 'Population growth and investment opportunity', *Quarterly Journal of Economics*, **55** (November 1940), pp. 64–79.

Population Studies Vol. XXVI, nº 2 July 1972. London

literature of the period.[7] One searches in vain for precise statistical estimation of the values of the parameters involved in the impact of a stationary population upon the economy.[8] The objective of this paper is to attempt to derive statistically the pattern of consumption associated with a stationary population and compare it with that of a society characterized by a growing population. While this line of inquiry will not resolve all issues associated with the impact of the cessation of population growth upon the economy it should shed some light on the extent of economic adjustments required, were population growth to cease.

II. DEMOGRAPHIC CALCULATIONS

The first task was to construct a stationary population equivalent to the real growing population of the United States in 1960. What was needed was the composition of such a population in terms of spending units. The age structure of the stationary population is readily available,[9] but a review of the literature failed to discover any projections of composition by families or households.

Table 1 presents a plausible structure for a non-growing population, indicative of the shifts to be anticipated in age, number, and size of households. It was necessary to make minor adjustments in the data on the growing population to conform with consumption data.

TABLE 1. *Distribution of spending units classified by age of head and size. Percentages – growing and stationary population*

Age of head of spending unit	Size of spending unit 1	2	3	4	5	6 and over
Under 25	0·51 (0·44)	1·63 (1·88)	1·60 (1·07)	0·87 (0·60)	0·31 (0·21)	0·17 (0·11)
25–34	0·89 (0·92)	2·79 (7·36)	3·64 (2·46)	4·94 (3·33)	3·37 (2·30)	2·86 (1·93)
35–44	1·09 (1·01)	3·06 (6·63)	3·65 (2·47)	5·47 (3·70)	4·14 2·30)	4·66 (1·93)
45–54	1·69 (1·66)	6·20 (9·27)	4·67 (3·17)	3·66 (2·48)	2·05 (1·39)	2·25 (1·52)
55–64	2·49 (2·63)	7·99 (6·86)	3·11 (3·80)	1·35 (1·66)	0·65 (0·80)	0·75 (0·93)
65–74	2·80 (2·91)	6·92 (7·72)	1·66 (1·14)	0·54 (0·36)	0·25 (0·17)	0·28 (0·19)
75 and over	1·64 (3·00)	2·58 (5·32)	0·54 (0·42)	0·16 (0·11)	0·07 (0·05)	0·07 (0·05)

NOTE: Values for the stationary population are in parentheses.

[7] An example of this is Keynes's estimate that one-half of capital formation for the period 1860–1913 could be attributed to the impact of population growth. John Maynard Keynes, 'Some economic consequences of a declining population', *Eugenics Review*, **29** (April 1937), pp. 13–17.

[8] One exception to this general indictment is the article by Leon Goldenburg, ' Savings in a state with a stationary population', *Quarterly Journal of Economics*, **61** (November 1964), pp 40–65, which does attempt to derive statistically some estimates of the impact of declining rates of population growth upon the volume of savings However, this artilce was purely historical and was subjected to serious methodological criticism by H. W. Arndt, 'Savings in a state with a stationary population: comment', *Quarterly Journal of Economics*, **62** (August 1948), pp. 623–628.

[9] *Vital Statistics of the United States: 1960*, Vol. II, *Mortality*, Part A, Washington, 1963, Tables 2–7 (Life Tables).

The procedure and assumptions employed in these calculations are outlined below. The age structure for the growing population is taken from the 1960 Census[10] and for the stationary population from the life tables.

Adjusting the Growing Population

(1) The census report gives husband–wife families, comprising some 88% of the total, by size and age of head in ten-year categories.[11] This conforms exactly with the needs of the consumption data. However, other kinds of families are shown by larger age groups. In each such instance the proportion of husband–wife families in the appropriate ten-year group was used to distribute the other families within the larger grouping.

(2) For households of unrelated individuals, data on primary individuals, i.e. heads of households, and secondary individuals by age class are given in the census.[12] These were assigned from a Poisson distribution with mean $= \dfrac{\text{Secondary individuals}}{\text{Primary individuals}} = 0{\cdot}3048$. This implies that the age of the head does not affect the probability that there are other individuals living with him. There seemed to be no information on which to base a statement of age-specific probabilities needed to compute a separate distribution for each age class.

Computing the Stationary Population

(1) The first step, embodying perhaps the most extreme assumption, was to compute the total number of dependants, excluding spouses, for each age class of heads of spending units in the stationary population. It was assumed that this number would have the same ratio to the total number of people aged less than 20 years in the stationary and in the growing population.

(2) The dependants were then distributed as follows:

(*a*) For each age class in the growing population the number of dependants in families of each size was expressed as a proportion of the total dependants of heads of spending units in the age class.

(*b*) The totals of dependants in the stationary population were distributed in the same proportions, and the corresponding number of family units in each cell was obtained by division.

(3) It was assumed that marriage rates would remain unchanged, i.e. that the proportion of persons in a given age class forming husband–wife families would be the same in the stationary as in the growing population. Corresponding assumptions were made for other categories. This proportion was then multiplied by the total number of persons in each stationary age class to obtain the totals for numbers of families.

(4) For each age class, the sum of the cells obtained in (2) above was substracted from the totals obtained in (3) to yield the number of two-person families.

(5) For unrelated households, the numbers of primary and secondary individuals were adjusted to the stationary age structure and distributed as for the growing population, using a Poisson distribution with mean = 0·2706.

[10] *United States Census of Population: 1960*, Vol. 1, *Characteristics of the Population*, Part I, *Summary*, Washington, 1964, Tables 155. Hereinafter this source is referred to as *U.S. Census*.
[11] *U.S. Census*, Table 187.
[12] *U.S. Census*, Table 181.

Aggregate Characteristics of the Populations

The stationary population constructed by the authors was within 1% of the total number of individuals in the actual growing population.[13] Given the approximate methods of calculation, it is felt that this aggregate error is acceptable.

A striking feature of the stationary population, to be discussed later, is the change in the total number of spending units: 60,714,118 against 53,125,202 in the growing population, an increase of 14·3%. This is a reflection of an older population and the assumed invariance of propensities to marry, or otherwise form households.

III. COMPARATIVE CONSUMPTION PATTERNS

The previous section of this paper outlined the demographic impact, in terms of spending units, of a stationary population. In order to derive consumption patterns for the two populations it is necessary to have consumption data for spending units corresponding to the age and size classifications of the demographic data. By re-arranging data contained in the *Survey of Consumer Expenditures, 1960–61*, it is possible to secure sample data organized by size of the spending unit and by the age of its head.[14] From the *Survey* data have been so obtained for the following categories:

- Money Income After Taxes
- Expenditures for Current Consumption
- Food, Total
  - Food Prepared at Home
  - Food Away From Home
- Alcoholic Beverages
- Tobacco
- Housing, Total
  - Shelter
    - Rented Dwelling
    - Owned Dwelling
    - Other Shelter
  - Fuel, Light, Refrigeration, Water
  - Household Operations
  - House Furnishings and Equipment
- Clothing, Clothing Materials, Services
- Personal Care
- Medical Care
- Recreation
- Reading
- Education
- Transportation
  - Automobile
  - Other Travel and Transportation
- Other Expenditures

[13] This includes persons living in group quarters as well as those in households. The consumption *Survey* did not include these people, so they are not included in Table 1. The source for this category is *U.S. Census*, Table 182.

[14] All data on consumer behaviour utilized in this paper are taken from *Survey of Consumer Expenditures, 1960–61, Consumer Expenditures and Income, Cross-Classification of Family Characteristics, Total United States, Urban and Rural 1960–61*, Supplement 2 to BLS Report 237–93 (USDA Report CES 30), Washington, June 1966, Tables 12a–12g. Hereinafter this report is referred to as *Survey*.

Aggregate Consumption Changes

By multiplying the population weights of Table 1 by the appropriate weights for the various items of consumer expenditure and aggregating the results it is possible to obtain the weighted value of the different categories in the overall pattern of consumer expenditures for both the stationary and the growing population. By comparing the results of these computations it is possible to see the shifts in consumption patterns created by imposing on the 1960 consumption patterns an age and size distribution of spending units, such as would have prevailed had the United States at that time had a stationary population.

For the purposes of this paper 'money income after taxes' is considered as equivalent to the economist's concept of 'disposable income', 'and expenditure for current consumption' is considered as equivalent to 'consumption' in national income accounting analysis. The ratio of these two categories is therefore assumed to be the equivalent of the average propensity to consume. The values for the average propensity to consume for the various age–size classifications are given in Table 2.

TABLE 2. *Expenditures for current consumption/money income after taxes. Percentages*

Age of head of spending unit	Size of spending unit 1	2	3	4	5	6 and over
Under 25	97·43	100·38	105·39	102·21	99·37	102·59
25–34	88·94	93·69	94·36	94·84	96·41	99·87
35–44	92·06	87·79	94·09	90·20	93·90	94·28
45–54	87·56	85·56	86·46	90·83	90·12	93·34
55–64	84·79	83·50	88·31	88·54	88·62	89·24
65–74	92·63	86·34	82·35	97·23	82·99	85·23
75 and over	95·89	94·44	70·74	92·89	84·63	103·93

Multiplying Table 1 by Table 2 yields the following results. The average propensity to consume for the growing population of 1960 is 0·908. The average propensity to consume for the assumed stationary population is 0·905. The absolute difference is therefore only 0·3% and when it is remembered that the consumption percentages are derived from sample data it is obvious that the impact of the stationary population on aggregate consumption is not significant. The macro-economic analysis therefore reveals no significant impact on the economy, through the medium of the percentage of disposable income utilized for current consumption expenditures, by the arrival of a stationary population.

Sectoral Shifts in Consumption

It would be possible for the macro-economic variable to show no significant variation and still have significant sectoral shifts revealed by a more disaggregated analysis. Computations identical to those described above for aggregate consumer expenditure were carried out for the 23 categories of consumer expenditure listed earlier. The results of these computations are shown in Table 3. Of the 23 categories analysed six showed no shift at all, eight a shift of 0·1, two a shift of 0·2, two a shift of 0·3, four a shift of 0·4, one a shift of 0·6, and one a shift of 0·7.

Population Studies

The largest shift in absolute terms is 0·7 in 'total housing' and the second largest shift in absolute terms comes in 'rented dwelling' (0·6). In relative terms the largest shift is in expenditure on education which declines 0·1 on an initial value of 0·9, a shift of approximately 11%. The second largest relative shift is in 'rented dwelling' which increased by slightly more than 10% on a base value of 5·8. The next largest category of shifts occurred in medical care, which increased in relative terms by about 4%, and in clothing, clothing materials, and services, which decreased by about the same amount.

Just as the macro-economic perspective revealed no dramatic changes in the patterns of consumer behaviour, following the onset of a stationary population so the more disaggregated computations displayed in Table 3 reveal what may be considered minimal sectoral changes in the pattern of consumer behaviour. The arrival of a stationary population would apparently not mean dramatic shifts in the economy and it certainly would not mean that there would be a dramatic shift away from goods requiring relatively large capital inputs.

TABLE 3. *Consumer expenditures by category. Percentages*

Food, Total			24·5 (24·2)
Food Prepared at Home		19·6 (19·2)	
Food Away From Home		4·9 (5·0)	
Tobacco			1·8 (1·8)
Alcoholic Beverages			1·6 (1·6)
Housing, Total			29·6 (30·3)
Shelter		13·5 (13·9)	
Rented Dwelling	5·8 (6·4)		
Owned Dwelling	6·9 (6·8)		
Other Shelter	0·7 (0·7)		
Fuel, Light, Refrigeration, Water		5·1 (5·2)	
Household Operations		5·8 (6·0)	
House Furnishings and Equipment		5·2 (5·2)	
Clothing, Clothing Materials, Services			9·8 (9·4)
Personal Care			2·9 (2·8)
Medical Care			7·0 (7·3)
Recreation			3·8 (3·7)
Reading			0·9 (0·9)
Education			0·9 (0·8)
Transportation			15·0 (14·9)
Automobile		13·4 (13·2)	
Other Travel and Transportation		1·6 (1·7)	
Other Expenditures			2·2 (2·2)
Total			100·0 (99·9)

NOTE: Values for the stationary population in parentheses.

IV. RATIONALIZING THE RESULTS

The results presented in the previous section are at variance with most of the literature previously presented by economists on this question. It is perhaps desirable to provide an explanation for these unexpected results. Fortunately, simple linear regression analysis provides at least a partial answer.

The regression results are presented in Tables 4 and 5. The first column, $\sum \overline{X}$, may be taken as a measure of the importance of the particular category of consumer expenditure in total expenditure. The second column, $\sum b$, measures the responsiveness of the particular category of consumer expenditure to the age of the head of the spending unit and the size of the spending unit respectively. The third column, $\sum b/\sum \overline{X}$, measures the responsiveness of the category of consumer expenditure

TABLE 4. *Regression results. Percentage consumer expenditures by category/age*

Category	$\Sigma\bar{X}$	Σb	$\Sigma b/\Sigma\bar{X}$
Other Expenditures	0·125	0·016	0·128
Fuel, Light, Refrigeration, Water	0·297	0·035	0·118
Other Shelter	0·036	0·004	0·111
Medical Care	0·427	0·046	0·108
Owned Dwelling	0·349	0·021	0·060
Food Prepared at Home	1·193	0·066	0·055
Reading	0·053	0·002	0·038
Food, Total	1·515	0·051	0·034
Personal Care	0·174	0·003	0·017
Household Operations	0·329	−0·003	−0·009
Other Travel and Transportation	0·098	−0·001	−0·010
Clothing, Clothing Materials, Services	0·617	−0·007	−0·011
Housing, Total	1·678	−0·021	−0·013
Education	0·061	−0·001	−0·016
Shelter	0·771	−0·033	−0·043
Tobacco	0·110	−0·005	−0·045
Automobile	0·811	−0·039	−0·048
Transportation	0·902	−0·054	−0·060
Food Away From Home	0·319	−0·021	−0·066
House Furnishings and Equipment	0·283	−0·020	−0·071
Alcoholic Beverages	0·094	−0·008	−0·085
Recreation	0·228	−0·023	−0·101
Rented Dwelling	0·389	−0·059	−0·152

TABLE 5. *Regression results. Percentage consumer expenditures by category/size*

Category	$\Sigma\bar{X}$	Σb	$\Sigma b/\Sigma\bar{X}$
Education	0·073	0·009	0·123
Food Prepared at Home	1·396	0·141	0·101
Clothing, Clothing Materials, Services	0·720	0·061	0·085
Food, Total	1·768	0·087	0·049
Tobacco	0·130	0·004	0·031
Automobile	0·948	0·026	0·027
Recreation	0·269	0·007	0·026
Personal Care	0·203	0·004	0·020
Fuel, Light, Refrigeration, Water	0·358	0·005	0·014
Transportation	1·065	0·004	0·004
Medical Care	0·499	−0·003	−0·006
Owned Dwelling	0·404	−0·004	−0·010
House Furnishings and Equipment	0·331	−0·006	−0·018
Other Expenditures	0·145	−0·003	−0·021
Other Shelter	0·043	−0·001	−0·023
Reading	0·063	−0·005	−0·080
Household Operations	0·385	−0·031	−0·081
Housing, Total	1·956	−0·165	−0·084
Shelter	0·901	−0·122	−0·135
Food Away From Home	0·379	−0·058	−0·153
Alcoholic Beverages	0·111	−0·019	−0·171
Other Travel and Transportation	0·116	−0·020	−0·172
Rented Dwelling	0·455	−0·120	−0·264

273

 274

to the age–size variables in relation to the importance of the item in the total package of consumer expenditure.

The important conclusion to be drawn from these two tables is that the age and size variables affect relative consumer spending in contradictory ways. For twelve of the 23 categories of consumer expenditure the impact of ageing associated with the stationary population is opposite to the impact of the reduction in the size of the spending units. These twelve categories are: 'food, total'; 'food prepared at home'; 'food away from home'; 'alcoholic beverages'; 'housing, total'; 'shelter'; 'rented dwelling'; 'fuel, light, refrigeration, water'; 'household operations'; 'house furnishings and equipment'; 'personal care'; and, 'other travel and transportation'. The twelve categories of consumer expenditure displaying offsetting tendencies with regard to the age and size variables account for 52·7% of total consumer expenditure of the growing population and 53·3% of consumer expenditure of the stationary population.

There are only five categories of consumer expenditure in which both the age and size shifts associated with the movement from a growing to a stationary population would tend to increase the relative importance of this item in the overall pattern of consumer behaviour. These categories are: 'other shelter'; 'owned dwelling'; 'medical care'; 'reading'; and, 'other expenditure'. These five categories together account for only 17·7% of total consumer expenditure in the growing population and 17·9% in the stationary population. There are six categories of consumer expenditure for which the regression analysis indicates that both the size and age impacts of zero population growth would be negative. These six categories are: 'clothing, clothing materials, services'; 'tobacco'; 'education'; 'recreation'; 'transportation'; and, automobile'. These six categories account for 29·5% of total consumer expenditure in the growing population and for 28·7% of total consumer expenditure in the stationary population.

Looking back at the earlier literature one can find suggestions, as cited earlier, about the impact of ageing of the population on the propensity to consume and on sectoral patterns of consumption. One can also find discussions, though later than the literature cited earlier, of the impact of family size on the propensity to consume.[15] What is lacking in the earlier discussion is a clear recognition of the possibility that the age and size variables could work to offset each other's impacts on consumption patterns. It is precisely this possibility which the evidence presented above seems to support.

V. CONCLUSIONS AND UNRESOLVED ISSUES

The basic conclusion of this paper is evident. The consumption patterns of a stationary population are sufficiently like those associated with a growing population, so that there is no real reason to fear economic disorder from this source with the cessation of population growth. In terms of the population discussion of the period between the late 1920's and the end of the 1940's this is sufficient to allay the major concerns expressed by Hansen, Billing, Robbins, Sweezy, and others. However, there remain unresolved issues associated with the economic impact of a stationary population.

The process of calculating the spending unit distribution of the stationary population, as mentioned earlier, yielded a side-result. For a population with a given total a stationary population will contain a larger absolute number of spending units. This result did not affect any of the calculations in this paper, since all the other calculations were in percentage, not absolute, terms. However, these extra households pose an interesting problem for the economist concerned with the probable economic effects of zero population growth. Would the increase in the work force implied in the increase in households find productive employment and thereby contribute to an increase in the level of national income? This would perhaps be a micro-economic answer. Or would these extra people in the work force constitute a drag on the labour-market leading to a depression of real wages 274

[15] C. L. Barber, 'Population growth and the rate of investment', *American Economic Review*, 63 (March 1953), pp. 133–139.

per head with a less-than-full employment equilibrium? This would be the answer implied in a fairly orthodox Keynesian macro-model. The results of this study do not yield an answer to this question.

Also it should be noted that the structure of the assumed stationary population is an approximation, derived in part from information on the counterpart growing population. It might turn out that this assumed population would not stand rigorous demographic testing for stability. A related shortcoming of the present study is its static nature. Besides assuming the presence of an equilibrium stationary population, no account was taken of oscillations in the growing population. Furthermore, it is well known that the achievement of an equilibrium stationary population will require a long period of time. During the time when the society is moving toward a stationary state major demographic variables will tend to oscillate around the final equilibrium position.[16] These oscillations could generate potentially serious macro-economic consequences for the performance of the economy. While this paper indicates there is nothing inherent in the consumption patterns of a stationary population that would preclude the successful operation of the economy in the presence of a population that had ceased to grow, it does not preclude a series of difficult adjustments in the time path that would lead to the ultimate realization of zero population growth.

Given the apparent inevitability, perhaps even logical necessity of zero population growth it is time that economists began to take an empirical look, static and dynamic, macro and micro, at the probable consequences of a stationary population. This paper represents one step in this analysis.

[16] Tomas Frejka, 'Reflections on the demographic conditions needed to establish a U.S. stationary population growth', *Population Studies*, **22** (November 1968), pp. 379–397.

276

How a Trend Towards a Stationary Population Affects Consumer Demand*

THOMAS J. ESPENSHADE

During the great depression of the 1930s economists in both the United States and Europe tried to analyse the economic consequences of declining rates of population growth. Not only were birth rates in many industrial countries at the lowest levels ever, but they coincided with high rates of unemployment. Of the many economists who held that demographic trends were partly responsible for the adverse economic conditions, a prominent example was John Maynard Keynes. According to his so-called stagnation thesis, population growth stimulates investment demand in two ways: more people need more goods and services and, hence, more investment in factories and machinery; and with population growing, businessmen are more likely to regard their investment misallocations as less serious than when the growth is slow or nil.[1] A minority of writers were more optimistic about the economic consequences of slower rates of population growth. For example, Thompson argued that with a lower ratio of consumers to producers the population would enjoy a higher standard of living and the education of children should improve.[2]

With the baby boom following World War II this argument was temporarily set aside. But in recent years, birth rates in developed countries have resumed their downward course. In fact, in the early 1970s, total fertility rates were near, at, or even below the replacement level in 20 industrial nations.[3] Whether as a response to the new demographic circumstances or simply by coincidence, economists have renewed their inquiry into the economic consequences of a stationary population. The recent reports of the Commission on Population Growth and the American Future may be the most comprehensive examination of this issue. What difference would it make, the Commission asked, for income per head, aggregate demand, the supply of savings, the rate of technological progress, and the income distribution if there were two children in the average family instead of three? It concluded, in part:

In the long run, no substantial benefits will result from further growth of the Nation's population, rather that the gradual stabilization of our population through voluntary means would contribute significantly to the Nation's ability to solve its problems. We have looked for, and have not found, any convincing economic argument for continued population growth. The health of our country does not depend on it, nor does the vitality of business nor the welfare of the average person.[4]

This contradicts the assessment of the 1930s, when a declining population was viewed with apprehension, and many European nations, including France, England, Italy, and Germany, implemented policies in attempts to stimulate their birth rates.[5]

* This paper has benefited from discussions with P. Hanumantha Rayappa and from his generous research assistance. Dan Garnick, Campbell Gibson, Paul Glick, Morris Goldman, Arthur Norton, Norman Ryder, Joseph Spengler, and James Spitler provided crucial data or information at various stages of the research.

[1] John Maynard Keynes, 'Some Economic Consequences of a Declining Population', *Eugenics Review*, **29**, (1937), pp. 13-17.

[2] Warren S. Thompson, 'The Economic Consequences of Slow Population Growth in the United States', in *Ohio Conference of Statisticians on Business Research, Proceedings, 1939* (Columbus: Ohio State University, 1940), pp. 3-11.

[3] Charles F. Westoff, 'The Populations of the Developed Countries', *Scientific American*, **231**, (1974), pp. 109-120.

[4] Commission on Population Growth and the American Future, *Population and the American Future* (Washington, D.C.: U.S. Government Printing Office, 1972), p. 4.

[5] David V. Glass, *The Struggle for Population* (Oxford: Oxford University Press, 1936); David V. Glass, *Population Policies and Movements in Europe* (Oxford: Oxford University Press, 1940); Joseph J. Spengler, *France Faces Depopulation* (Durham, N.C.: Duke University Press, 1938).

276

Population Studies, **32**, 1. Printed in Great Britain March. 1978

The Commission's report was not definitive and research on the economic consequences of zero population growth or a decline has continued.[6] This paper follows in this context.

THE PROBLEM DEFINED

Shown in Table 1 are data at ten-year intervals on the American expenditures for total personal consumption and for smaller categories. Students of introductory economics will recognize this sum as one component in the familiar equation,

$$(\mathrm{GNP}) = C + I + G.$$

That is, the gross national product (GNP) – or the total money value of all final goods and services produced by the economy in a given year – is equal to the sum of consumption (C), investment (I), and government (G) expenditures. According to Table 1, expenditures on all types of goods and services have increased. This is hardly surprising, if only because the population grew from about 123 million in 1930 to almost 205 million in 1970. What is important here is how the distribution changed. Durables became an increasingly important component of total consumption while non-durables, particularly food and drink, declined, especially after 1940. Following an abrupt decline during the 1930s, the proportion of total consumption defined as services steadily increased up to 1970. How is this composition of personal expenditure likely to change in the future? In particular, what difference would it make if the population moved toward zero growth?

Several economists have speculated how consumption patterns would be affected.[7] The reasoning is that each age group demands special goods and services: infants use toys and nursery equipment; older children and teenagers consume educational services; the elderly require increased medical attention. Moreover, with the prolonged duration of relatively high American birth rates after World War II, a transition to a stationary population would bring about significant changes in the age structure. For example, assuming a slight improvement in mortality, the attainment of a stationary state would cause the percentage of the population under age 18 to fall to 24 from its 1970 value of 34, and of that aged 65 and over to rise from almost 10 in 1970 to 16.[8] With these changes in age distribution the median age would rise from 27·9 years in 1970 to 37·3 years. If the age distribution were actually transformed so drastically presumably one economic consequence would be a shift in consumption away from items used by the young towards those that the retired and the elderly consume.

Just how the composition of consumer demand can be expected to change with a gradual trend toward a stationary population is analysed in the remainder of this paper. First, the historical determinants of changing patterns of consumer demand are examined. Then these determinants are projected to the year 2020 on the basis of two alternative assumptions: an

[6] Donald Eilenstine and James P. Cunningham, 'Projected Consumption Patterns for a Stationary Population,' *Population Studies*, **26**, (1972), pp. 223–231; William J. Serow, 'Economic and Demographic Implications of a Stationary Population: A Case Study of the United States', Unpublished doctoral dissertion (Duke University, 1972); William J. Serow, 'The Implications of Zero Growth for Agricultural Commodity Demand', *American Journal of Agricultural Economics*, **54**, (1972), 955–963; Joseph J. Spengler, *Declining Population Growth Revisited*, Monograph 14 (Chapel Hill: Carolina Population Center, University of North Carolina, 1971); Joseph J. Spengler (ed.), *Zero Population Growth: Implications* (Chapel Hill: Carolina Population Center, University of North Carolina, 1975).

[7] Ansley J. Coale, 'Population Change and Demand, Prices, and the Level of Employment', in National Bureau of Economic Research, *Demographic and Economic Change in Developed Countries* (Princeton, N.J.: Princeton University Press, 1960), p. 369; Allen C. Kelley, 'Demographic Changes and American Economic Development: Past, Present, and Future', in Commission on Population Growth and the American Future, *Economic Aspects of Population Change*, Research Reports (Washington, D.C.: U.S. Government Printing Office, 1972, vol. 2, p. 32 n; Lionel Robbins, 'Notes on Some Probable Consequences of the Advent of a Stationary Population in Great Britain', *Economica*, **9**, (1929), pp. 71–82.

[8] U.S. Bureau of the Census, 'Illustrative Population Projections for the United States: The Demographic Effects of Alternate Paths to Zero Growth', *Current Population Reports*, Series P-25, No. 480 (Washington, D.C.: U.S. Government Printing Office, 1972); p. 16; U.S. Bureau of the Census, 'Population Profile of the United States: 1974', *Current Population Reports*, Series P-20, No. 279 (Washington, D.C.: U.S. Government Printing Office, 1975), p. 12.

278

TABLE 1. *Expenditures for personal consumption, United States, 1930–70 (billions of 1958 dollars)*

Item	1930		1940		1950		1960		1970	
	Absolute	Per cent	Absolute	Per cent	Absolute	per cent	Absolute	per cent	Absolute	Percent
Total personal consumption expenditure	130·4	100·0	155·7	100·0	230·5	100·0	316·2	100·0	477·5	100·0
Durable goods, total	12·9	9·9	16·7	10·7	34·7	15·1	44·9	14·2	83·8	17·5
Automobiles and parts	4·7	3·6	6·7	4·3	15·9	6·9	20·0	6·3	34·4	7·2
Furniture and household equipment	6·6	5·1	8·2	5·3	15·1	6·6	18·7	5·9	36·9	7·7
Other	1·6	1·2	1·8	1·2	3·7	1·6	6·2	2·0	12·6	2·6
Non-durable goods, total	65·9	50·5	84·6	54·3	114·0	49·5	149·7	47·3	206·5	43·2
Food and drink	35·9	27·5	48·7	31·3	63·2	27·4	80·9	25·6	101·2	21·2
Clothing and shoes	16·4	12·6	17·8	11·4	21·8	9·5	26·6	8·4	39·1	8·2
Gasoline and oil	2·7	2·1	3·9	2·5	6·5	2·8	11·8	3·7	18·7	3·9
Other	11·0	8·4	14·1	9·1	22·5	9·8	30·4	9·6	47·5	9·9
Services, total	51·5	39·5	54·4	34·9	81·8	35·5	121·6	38·5	187·2	39·2
Housing	13·9	10·7	15·4	9·9	26·8	11·6	44·9	14·2	72·6	15·2
Household operations	7·6	5·8	8·4	5·4	11·7	5·1	18·8	5·9	29·8	6·2
Transport	4·2	3·2	4·5	2·9	8·5	3·7	10·1	3·2	12·6	2·6
Other*	25·8	19·8	26·1	16·8	34·8	15·1	47·9	15·1	72·3	15·1

Source: U.S. Bureau of Economic Analysis, *Survey of Current Business*, August 1965 and July 1973.

* 'Other services' includes, among other things, shoe cleaning and repair, cleaning, dyeing, pressing, alterations, storage and repair of garments, laundering in establishments, barbershops, beauty parlours, physicians, dentists, other medical care, and admission to selected spectator amusements.

278

average of approximately three births per woman and one of slightly more than two. Finally, the consumption patterns corresponding to these two projections are compared.

While the attention of economic demographers traditionally has been focused on the pattern of expenditures for consumption, one could extend this analysis to include investment and government expenditures as well. Purchases of goods and services by government (federal, state, and local) are a rising proportion of GNP and include such areas as education, housing, transport, and medical care that would be expected to be sensitive to demographic change. On the other hand, the power of the federal government to tax and to spend in order to counteract periods of economic recession and inflation might mean that the demographic influences on these expenditures are not always the most important ones. In any case, the complexity of these additional issues is sufficient to warrant separate study and will not be dealt with here.

HISTORICAL DETERMINANTS OF DEMAND PATTERNS

Consumer expenditures in the past provide a basis for projecting them into the future. For this purpose, I have separately analysed each of the eleven minor categories of consumption in Table 1 (automobiles and parts, furniture and household equipment, and so on) for the 38 years between 1929 and 1970, excluding the war years 1942–45. All figures are in constant 1958 dollars and have been converted to consumption per head by dividing by the estimated population of each year.[9]

According to the theory of consumer choice, income and relative prices help to explain trends in consumption per head. By constraining the total budget, income determines the overall level of expenditures but also affects the relative proportions used to purchase necessities and luxuries, and thus the distribution among various consumer categories. Relative prices also influence the allocation among competing goods and services. Moreover, since consumption patterns sometimes adjust slowly to changes in income, current purchases may also be affected by previous buying habits.[10] Demographic factors to be assessed in association with the economic variables include not only the age distribution but also family size because of possible economies of scale in household consumption. While such other factors as marital and family status, age and sex composition of the family and its members' labour force status may influence consumer behaviour, these additional hypotheses are more easily tested using cross-sectional rather than time-series data.

The concepts used in the present study are operationalized as follows: Income is measured by total personal consumption expenditures per head.[11] To get relative price variables the ratio of expenditures on a particular item (say food and drink) in current to constant dollars is divided by the ratio of total consumption expenditures in current to constant dollars. Previous buying habits are defined as expenditures per head on particular items during the preceding year. Alternative measures of age structure include the proportions of the population under age 5, under age 10, under age 15, and 65 years and over.[12] Data on average household size are used to measure changes in family size.[13]

[9] Annual data on consumption expenditures from 1929 to 1964 in both current and constant dollars are from U.S. Department of Commerce, 'The National Income and Product Accounts of the United States: Revised Estimates, 1929–64, *Survey of Current Business*, **45**, 8 (1965), p. 47. Data for 1965–70 are from successive July issues of the *Survey of Current Business* between 1969 and 1973. Annual population estimates come from the sources cited in footnote 12.

[10] T. M. Brown, 'Habit Persistence and Lags in Consumer Behavior', *Econometrica*, **20**, (1952), pp. 355–371.

[11] According to Houthakker and Taylor, little empirical basis exists for choosing as a measure of income between total personal consumption expenditures per head and disposable personal income per head. H. S. Houthakker and L. D. Taylor, *Consumer Demand in the United States: Analysis and Projections*, 2nd ed. (Cambridge, Mass.: Harvard University Press, 1970).

[12] Annual age distribution data are from U.S. Bureau of the Census, *Current Population Reports*, Series P-25, Nos. 141, 311, and 519.

[13] For the period 1947–1970, data on average household size have been put together from various numbers (in parentheses in what follows) of U.S. Bureau of the Census, *Current Population Reports*, Series P-20: 1947 (173); 1948 (21); 1949 (26); 1950, 1955, 1960–70 (266); 1951 (38); 1952 (44); 1953 (53); 1954 (67); and 1956–59 (176). Data for 1929–1946 were kindly supplied by Paul Glick, U.S. Bureau of the Census.

279

To assess the importance of these economic and demographic factors least-squares regressions were computed for each of the eleven consumption categories using income, relative prices, previous year's consumption, age distribution and household size as explanatory variables. For each dependent variable alternative sets of explanatory variables were tried to see which regression performed best.[14]

The final equation for each consumption category is shown in Table 2. The income term appears in all eleven regressions, and last year's consumption in all except one, automobiles and parts. The relative price variable, on the other hand, is given in just four.[15] Age distribution and household size are included for each of the durable and non-durable categories; only with the services categories (housing, household operations, transport and other services) are the two demographic variables largely absent. No simple generalization can be offered concerning either the magnitude or the direction of the demographic effects. Changes in average household size are associated with both economies and diseconomies of scale (indicated, respectively, by a negative or positive coefficient). Similarly, an ageing of the population causes the per head consumption of some items to increase (for example, food and drink) and of others to decline (automobiles and parts, gasoline and oil). In general, Table 2 confirms that demographic variables exert a statistically significant influence on spending for consumption. Whether or not a trend toward a stationary population will affect the composition of total personal consumption expenditures depends also on how important quantitatively these demographic forces happen to be.

ASSUMPTIONS UNDERLYING THE PROJECTIONS

Two alternative projections of consumption expenditures to the year 2020 have been prepared, one related to a steadily growing population and the other to a gradual reduction of fertility to the level of zero growth. These two postulates are the Series B and E projections developed by the U.S. Bureau of the Census[16] and used also by the Commission on Population Growth. Roughly, Series B corresponds to an average of three children per family and Series E to one of two children.

Projections of personal consumption require an extrapolation of each of the explanatory variables in Table 2 in addition to projections of total population. Total population size and the number in each five-year age category are given by the U.S. Bureau of the Census.[17] In the absence of better information relative prices are assumed to remain constant at their 1973 levels.[18] Of several projections of gross national product to the year 2000, prepared by the U.S.

[14] The criteria involved in the final selection process included (1) a high value for R^2, (2) coefficients with the proper signs (in practice, this criterion was important only for the relative price variables, whose coefficients were expected to be negative), (3) statistically significant coefficients (generally, this was taken to mean a coefficient which exceeded twice its standard error, although this stipulation was relaxed in some cases, especially for prices), (4) the magnitude of the Durbin–Watson statistic (the meaning of this term is discussed by J. Johnston, *Econometric Methods*, McGraw Hill, New York, 1963), p. 192, and (5) regressions that subsequently produced reasonable projections for the expenditure category in question (imposing this criterion altered the equation that otherwise would have been chosen in only three instances – automobiles and parts, food and drink, and clothing and shoes – in which cases it necessitated eliminating the square of the total personal consumption per head term from the final regression. If the squared term had been kept, for example, in the food and drink equation, the non-linear nature of the relation up to 1970 would have meant that purchases projected beyond 1970 on the basis of the same relation would have declined, starting in about 1990, despite continued growth in population and in total consumption. This situation, while not impossible, is unlikely to occur.

[15] This finding parallels an earlier observation by Houthakker and Taylor, *op. cit.* in footnote 11, p. 165.

[16] U.S. Bureau of the Census, 'Projections of the Population of the United States, by Age and Sex (Interim Revisions): 1970 to 2020', *Current Population Reports*, Series P-25, No. 448 (Washington, D.C.: U.S. Government Printing Office, 1970).

[17] *Ibid.*

[18] While a different assumption might affect the quantitative results, it should not alter them qualitatively, since the same set of relative prices underlies both Series B and Series E projections of consumption behaviour. For 1973, the relative price data for the eleven consumption categories, beginning with automobiles and parts and ending with other services, are 0·7761, 0·7792, 0·8476, 1·0640, 1·0396, 0·9067, 0·9360, 0·9733, 0·9624, 1·1543, and 1·2800, respectively. These data are from the July 1974 issue of the *Survey of Current Business*.

280

TABLE 2. *Regressions of expenditures for personal consumption per head on selected economic and demographic variables, United States, 1929–70*

Dependent Variable*	Independent variables†									R^2
c_t	Constant	x_t	p_t	c_{t-1}	$d_{1,t}$	$d_{2,t}$	$d_{3,t}$	$d_{4,t}$	h_t	
Automobiles and parts	0·2237	0·0695	−0·1741				0·3238		−0·0484	0·9575
	(1·7786)‡	(4·2452)	(−3·2245)				(2·9047)		(−2·3595)	
Furniture and household equipment	−0·0740	0·0873		0·2656			−0·1983		0·0188	0·9945
	(−3·6744)	(10·6897)		(3·5187)			(−6·3702)		(4·4905)	
Other durables	−0·0257	0·0240	−0·0049	0·4416	−0·0746				0·0050	0·9979
	(−2·6961)	(9·2109)	(−1·4924)	(7·9453)	(−4·2240)				(3·2613)	
Food and drink	0·4344	0·0223		0·5790		−0·2164			−0·0722	0·9861
	(4·4401)	(2·5227)		(7·0019)		(−2·2956)			(−4·0171)	
Clothing and shoes	0·0020	0·0563		0·3039			−0·2167		0·0211	0·9727
	(0·0839)	(7·7712)		(3·6372)			(−5·1207)		(3·6572)	
Gasoline and oil	0·0339	0·0090		0·8785				−0·1830	−0·0072	0·9988
	(3·1633)	(5·3529)		(25·4601)				(−2·8532)	(−3·8276)	
Other non-durables	0·0680	0·0678		0·2933		−0·0459			−0·0168	0·9981
	(4·5269)	(8·8538)		(3·6229)		(−1·8968)			(−5·4705)	
Housing	0·0024	0·0149	−0·0089	0·9520						0·9996
	(0·4502)	(3·2875)	(−3·6800)	(42·4426)						
Household operations	−0·0068	0·0183		0·7732						0·9957
	(−5·3138)	(5·3483)		(14·1216)						
Transport	0·0660	0·0097	−0·0279	0·4673					−0·0078	0·9937
	(7·3051)	(5·7956)	(−6·2139)	(8·6629)					(−4·3363)	
Other services	−0·2380	0·1305		0·2620					0·0618	0·9963
	(−7·6641)	(9·6189)		(3·2285)					(7·7999)	

* Personal consumption expenditure per head in year t on the item in question, in thousands of 1958 dollars.

† The independent variables are defined as follows:

'Constant' is the constant term in the regression;

x_t is total personal consumption expenditure per head in year t, in thousands of 1958 dollars;

p_t is the relative price in year t of the good in question, calculated by dividing the ratio of current to constant (1958) dollar expenditure for that good by the corresponding ratio for total personal consumption expenditure (see Houthakker and Taylor, *op. cit.* in footnote 11, pp. 55–59;

c_{t-1} is the previous year's value of the dependent variable, c_t;

$d_{1,t}$ is the proportion of the population under age 5 in year t;

$d_{2,t}$ is the proportion of the population under age 10 in year t;

$d_{3,t}$ is the proportion of the population under age 15 in year t;

$d_{4,t}$ is the proportion of the population aged 65 years and over in year t;

h_t is average household size in year t.

‡ Figures in parentheses are the t-ratios associated with the regression coefficients.

281

Bureau of Economic Analysis, the two selected for use here, corresponding to Series B and E, incorporate a decline of one per cent per year in the annual number of hours worked per person.[19] Labour force estimates made by Johnston[20] form the basis of GNP projections to 2020.[21] According to Kutscher the percentage of GNP used for total personal consumption expenditure has remained relatively constant since World War II, ranging between 60 and 66 per cent.[22] Thus, total personal consumption expenditure per head was projected by multiplying the GNP figures by 0·65 and then dividing the product by that year's estimated population.

The starting point for projecting average household size is the high series of the number of households in the United States to 1990.[23] Since most of those who are likely to be household heads in 1990 were alive in 1970, a single projection series for the total number of households can adequately characterize both Series B and Series E population trends which differ in fertility. To estimate the average size of the household, projected total population was multiplied by the average proportion of the total population living in households in 1960, 1965, and 1970 (0·969) and then dividing this product by the projected number of households. For calculating average household size beyond 1990s the linear trends in the proportion of successive age groups functioning as heads of households between 1970 and 1990 were projected to 1995, 2000, and 2020 and then applied to both Series B and E projections of age brackets to produce the number of households under the alternative population postulates. The population in households was determined as before and division by the number of households established their average size.

The projected values of each of the variables are shown in Table 3. In both projections the total population grows continuously though the rate of increase is, of course, smaller with Series E. Despite the trend toward a stationary population underlying Series E growth will not have ceased by 2020 because (1) the transition to a fertility of about two children per couple is assumed to be gradual, (2) once fertility at this level is achieved, growth will continue for a time as the age distribution conforms to that of a stationary population, and (3) it is postulated that 400,000 immigrants will continue to arrive each year. With the mortality trends assumed in the projections the attainment of complete stationarity would require that 20·3 per cent of the population are under age 15 and 16·0 per cent 65 years old and over. Differences in age structure develop between the two projected populations with the proportion under age 15 larger with high fertility and smaller with low. Both projections of average household size continue the long-term decline from 4·12 persons in 1929 to 3·14 in 1970. While both projections reflect an assumed increase in age-specific headship rates the decline in household size is steeper for Series E because of the smaller number of children per family.

The variable with the largest relative increase is total consumption per head which, as a consequence mainly of an assumed continued rise in productivity, grows from an initial value of $2331 in 1970. The Series E population shows somewhat higher consumption standards for two

[19] These projections were made as part of the study on Population Growth and the American Future and are in constant 1958 dollars and compatible with the assumed population trends. U.S. Bureau of Economic Analysis, 'Selected National Data on Economic Growth', in Commission on Population Growth and the American Future, Economic Aspects of Population Change. Research Reports (Washington D.C.: U.S. Government Printing Office, 1972), Vol. 2.

[20] Denis F. Johnston, 'Illustrative Projections of the Labor Force of the United States to 2040', in Commission on Population Growth and the American Future, *Economic Aspects of Population Change*, Research Reports (Washington, D.C.: U.S. Government Printing Office, 1972), Vol. 2.

[21] The exact procedures are described in U.S. Water Resources Council *1972 OBERS Projections: Regional Economic Activity in the U.S.* (Washington, D.C.: U.S. Government Printing Office, 1974), Vol. 1 and conform to those used previously by the U.S. Bureau of Economic Analysis (*op. cit.* in footnote 19).

[22] Ronald E. Kutscher, 'The U.S. Economy in 1985 – Projections of GNP, Output, Income, and Employment', *Monthly Labor Review*, **96**, 12 (1973), pp. 27–42.

[23] U.S. Bureau of the Census, 'Demographic Projections for the United States', *Current Population Reports*, Series P-25, No. 476 (Washington, D.C.: U.S. Government Printing Office, 1972).

[redacted] 283

TABLE 3. *Projected values of the independent variables in the Series B and the Series E projections, United States, 1975–2020*

Item	1975	1980	1985	1990	1995	2000	2020
Series B*							
Total expenditures for personal consumption per head†	2·6838	2·9408	3·1315	3·3704	3·7245	4·1598	7·3601
Proportion of the population under age 5	0·0977	0·1102	0·1134	0·1078	0·1034	0·1052	0·1072
Proportion of the population under age 10	0·1797	0·2011	0·2154	0·2132	0·2040	0·2015	0·2082
Proportion of the population under age 15	0·2742	0·2776	0·2998	0·3082	0·3027	0·2954	0·3030
Proportion of the population aged 65 and over	0·0981	0·0992	0·0991	0·0994	0·0965	0·0899	0·0913
Average household size	3·02	2·97	2·97	3·00	2·96	2·92	2·74
Total population (millions)	219·101	236·797	256·980	277·286	297·884	320·780	440·253
Series E*							
Total expenditures for personal consumption per head†	2·7846	3·1415	3·4558	3·7779	4·1735	4·6472	8·1994
Proportion of the population under age 5	0·0812	0·0849	0·0859	0·0821	0·0767	0·0738	0·0704
Proportion of the population under age 10	0·1630	0·1629	0·1673	0·1647	0·1562	0·1484	0·1411
Proportion of the population under age 15	0·2594	0·2414	0·2421	0·2432	0·2363	0·2258	0·2114
Proportion of the population aged 65 and over	0·1001	0·1042	0·1075	0·1113	0·1117	0·1083	0·1344
Average household size	2·98	2·86	2·76	2·71	2·62	2·55	2·27
Total population (millions)	214·735	225·510	236·918	247·726	257·345	266·281	299·177

* The Series B and Series E population projections are from U.S. Bureau of the Census, *op. cit.* in footnote 16. They correspond, respectively, to ultimate completed cohort fertility rates of 3·1 and 2·11.

† In thousands of 1958 dollars.

283

reasons: the rates of female labour-force participation are assumed to be higher in the lower-fertility population because of the reduced responsibilities for young children at home; and, secondly, the Series E population contains a lower proportion at non-productive ages, below 15 and above 65. However, the differences between the two series in consumption per head are not great and never exceed 13 per cent.

ALTERNATIVE PROJECTIONS OF CONSUMPTION PATTERNS

Projections of personal consumption expenditures were produced by combining the data in Table 3 and the 1973 values of relative prices with the historical regressions in Table 2. For each expenditure item the predicted expenditure per head was multiplied by that item's 1973 relative price.[24] For each year of the projection period the projected values of the individual expenditure items were added to see whether their sum equalled the value of total consumption per head tabulated in Table 3. If not, the projection was repeated with an adjusted value for total consumption per head.[25] Finally, the projected average figures were multiplied by total population size.

TABLE 4. *Series B projections of personal consumption expenditure, United States, 1975–2020 (billions of 1958 dollars)*

Item	1975	1980	1985	1990	1995	2000	2020
Total expenditure for personal consumption	588·0	696·4	804·7	934·6	1109·5	1334·4	3240·3
Durable goods, total	74·2	90·4	107·0	126·8	153·0	186·9	488·8
	(12·62)*	(12·99)	(13·30)	(13·57)	(13·79)	(14·00)	(15·08)
Automobiles and parts	32·7	39·0	46·4	53·6	62·5	73·7	173·2
	(5·56)	(5·61)	(5·76)	(5·74)	5·63)	(5·52)	(5·35)
Furniture and household equipment	30·4	37·7	44·1	52·9	65·4	81·8	227·2
	(5·16)	(5·41)	5·48)	(5·67)	(5·89)	(6·13)	(7·01)
Other	11·2	13·7	16·5	20·3	25·2	31·3	88·4
	(1·90)	(1·97)	(2·05)	(2·17)	(2·27)	(2·35)	(2·73)
Non-durable goods, total	244·4	280·1	314·1	354·7	414·1	488·9	1061·4
	(41·56)	(40·22)	(39·03)	(37·96)	(37·32)	(36·64)	(32·76)
Food and drink	126·9	139·7	152·3	166·6	187·8	211·8	374·8
	(21·58)	(20·06)	(18·93)	(17·83)	(16·92)	(15·87)	(11·57)
Clothing and shoes	44·7	52·5	59·1	68·5	81·3	98·1	237·4
	(7·61)	(7·54)	(7·35)	(7·33)	(7·33)	(7·35)	(7·33)
Gasoline and oil	24·6	30·7	36·6	43·0	54·1	69·9	184·4
	(4·18)	(4·41)	(4·54)	(4·60)	(4·88)	(5·24)	(5·69)
Other	48·2	57·1	66·1	76·7	90·9	109·1	264·9
	(8·19)	(8·21)	(8·21)	(8·20)	(8·19)	(8·17)	(8·17)
Services, total	269·4	325·9	383·6	453·0	542·4	658·6	1690·1
	(45·82)	(46·80)	(47·67)	(48·47)	(48·89)	(49·36)	(52·16)
Housing	125·5	152·8	180·3	213·0	256·2	312·4	810·9
	(21·34)	(21·95)	(22·41)	(22·79)	(23·09)	(23·41)	(25·03)
Household operations	33·1	40·2	47·3	55·8	67·0	81·5	210·1
	(5·63)	(5·77)	(5·88)	(5·97)	(6·04)	(6·11)	(6·48)
Transport	15·5	18·2	20·7	23·5	27·4	32·2	72·2
	(2·64)	(2·61)	(2·58)	(2·52)	(2·47)	(2·42)	(2·23)
Other	95·3	114·7	135·3	160·7	191·8	232·5	596·9
	(16·21)	(16·47)	(16·81)	(17·19)	(17·29)	(17·43)	(18·42)

* Figures in parentheses indicate percentage of total personal consumption expenditure.

[24] Multiplying by relative price corrects for any changes in relative prices between 1958 and the projection date. On this point, see Houthakker and Taylor, *op. cit.* in footnote 11, pp. 53–54.

[25] Houthakker and Taylor, *op. cit.* in footnote 11, pp. 52 53.

The results with the Series B and the Series E projections are shown in Tables 4 and 5. In general, the trends of 1930–1970 are continued. The differences in the dollar volume of consumption would be expected, given the differences in population, but the two projections show a decidedly similar distribution of total consumption expenditures with no great differences throughout the projection period. Under the assumptions made in constructing the projections, a trend toward a stationary population would have a negligible effect on the pattern of consumption. This is essentially the same conclusion as that of Eilenstine and Cunningham,[26] who based their analysis on projections of cross-sectional relations. The projections in Resek and Siegel imply a similar result.[27]

Two points help explain these results. First, as seen in Table 3, with both projections total expenditure per head increases substantially between 1970 and 2020, by 216 per cent with Series B and 252 per cent with Series E. On the other hand, differences in this variable between the two series, with a maximum of only 12·1 per cent (in 1995), are negligible in comparison. Secondly, the direct effects of demographic change are generally offset by indirect ones. Population change affects consumption spending directly through changes in age structure and in

TABLE 5. *Series E projections of personal consumption expenditure, United States, 1975–2020 (billions of 1958 dollars)*

Item	1975	1980	1985	1990	1995	2000	2020
Total expenditure for personal consumption	597·9	708·4	818·7	935·9	1074·0	1237·5	2453·1
Durable goods, total	76·0	92·5	109·4	127·6	148·9	174·0	371·2
	(12·70)*	(13·05)	(13·36)	(13·64)	(13·86)	(14·06)	(15·13)
Automobiles and parts	32·4	37·8	44·1	50·5	57·6	65·5	127·2
	(5·42)	(5·33)	(5·39)	(5·40)	(5·36)	(5·29)	(5·18)
Furniture and household equipment	31·7	40·0	47·6	56·1	66·3	78·8	176·3
	(5·30)	(5·65)	(5·82)	(5·99)	(6·17)	(6·37)	(7·19)
Other	11·8	14·7	17·6	21·0	25·0	29·7	67·7
	(1·98)	(2·07)	(2·15)	(2·24)	(2·33)	(2·40)	(2·76)
Non-durable goods, total	248·8	288·1	325·8	363·7	410·7	464·9	817·6
	(41·61)	(40·66)	(39·79)	(38·86)	(38·24)	(37·57)	(33·33)
Food and drink	128·8	144·1	158·8	172·7	189·7	207·0	302·9
	(21·53)	(20·34)	(19·40)	(18·45)	(17·66)	(16·72)	(12·35)
Clothing and shoes	45·9	54·5	62·0	70·3	80·2	92·5	183·0
	(7·68)	(7·69)	(7·57)	(7·51)	(7·47)	(7·47)	(7·46)
Gasoline and oil	25·0	31·2	37·3	43·2	51·8	63·2	129·0
	(4·19)	(4·40)	(4·56)	(4·62)	(4·83)	(5·10)	(5·26)
Other	49·1	58·3	67·7	77·4	88·9	102·3	202·7
	(8·20)	(8·23)	(8·26)	(8·27)	(8·28)	(8·27)	(8·26)
Services, total	273·2	327·9	383·6	444·6	514·4	598·6	1264·3
	(45·69)	(46·29)	(46·85)	(47·50)	(47·90)	(48·37)	(51·54)
Housing	127·8	155·1	182·8	212·6	247·2	288·6	613·0
	(21·37)	(21·89)	(22·33)	(22·72)	(23·02)	(23·32)	(24·99)
Household operations	33·7	40·7	47·9	55·6	64·5	75·2	158·7
	(5·63)	(5·75)	(5·85)	(5·94)	(6·01)	(6·08)	(6·47)
Transport	15·7	18·4	21·1	23·8	26·9	30·5	55·7
	(2·62)	(2·60)	(2·58)	(2·54)	(2·51)	(2·46)	(2·27)
Other	96·1	113·7	131·8	152·6	175·7	204·4	437·0
	(16·07)	(16·05)	(16·09)	(16·30)	(16·36)	(16·52)	(17·81)

* Figures in parentheses indicate percentage of total personal consumption expenditure.

[26] Eilenstine and Cunningham, *loc. cit.* in footnote 6.

[27] Robert W. Resek and Frederick Siegel, 'Population Growth Rates and Consumption Demand', *Eastern Economic Journal*, 1 (1974), pp. 282–290.

285

household size, indirectly through its influence on differentials between the two series in total expenditure per head.

The direct effects of demographic change are shown in Table 6, based on Series E values for age distribution and average household size and Series B values for total expenditure per head and total population. By comparing Table 6 with Table 4 for the year 2020, for example, it may be seen that a transition to an older population with fewer persons per household increases food consumption by 14 per cent and decreases expenditures on gasoline and oil by nine per cent and on other services by five per cent. On the other hand, when the effects of an increase in total consumption per head are also taken into account (compare Tables 5 and 6 for the year 2020), proportionate expenditures on food and drink decline and those on gasoline and oil and other services increase. As can be seen by comparing all three tables, these offsetting propensities are to be found for other time periods and for most consumption categories. In general, the changes in consumption spending occasioned by the narrowly demographic consequences of a trend toward a stationary population are counteracted by the increase in total consumption per head accompanying the reduction in fertility. In other words, as the rate of population growth approaches zero, the items that consumers would generally purchase proportionately more of

TABLE 6. *Projections of personal consumption expenditure illustrating the direct effects of demographic change, United States, 1975–2020* (billions of 1958 dollars)*

Item	1975	1980	1985	1990	1995	2000	2020
Total expenditure for personal consumption	588·0	696·4	804·7	934·6	1109·5	1334·4	3240·3
Durable goods, total	73·9	89·2	104·9	124·2	150·2	183·8	485·2
	(12·56)†	(12·81)	(13·03)	(13·29)	(13·54)	(13·77)	(14·97)
Automobiles and parts	32·0	37·3	43·8	51·1	60·2	71·2	169·1
	(5·44)	(5·36)	(5·45)	(5·46)	(5·42)	(5·34)	(5·22)
Furniture and household equipment	30·5	38·1	44·7	53·3	65·5	81·9	228·6
	(5·19)	(5·47)	(5·56)	(5·71)	(5·91)	(6·14)	(7·05)
Other	11·4	13·8	16·3	19·8	24·5	30·7	87·5
	(1·93)	(1·99)	(2·03)	(2·12)	(2·21)	(2·30)	(2·70)
Non-durable goods, total	247·9	289·7	331·0	376·7	439·5	518·0	1103·9
	(42·16)	(41·61)	(41·13)	(40·31)	(39·62)	(38·82)	(34·07)
Food and drink	130·3	148·9	168·0	187·5	212·8	241·4	426·9
	(22·15)	(21·38)	(20·88)	(20·07)	(19·18)	(18·09)	(13·17)
Clothing and shoes	45·2	53·6	60·9	70·1	82·7	99·5	241·2
	(7·68)	(7·70)	(7·57)	(7·51)	(7·46)	(7·46)	(7·44)
Gasoline and oil	24·2	29·8	35·5	41·5	51·9	66·6	167·4
	(4·12)	(4·28)	(4·41)	(4·44)	(4·68)	(4·99)	(5·17)
Other	48·3	57·4	66·6	77·5	92·1	110·5	268·4
	(8·21)	(8·24)	(8·28)	(8·30)	(8·30)	(8·28)	(8·28)
Services, total	266·3	317·4	368·9	433·7	519·7	632·5	1651·2
	(45·28)	(45·58)	(45·84)	(46·40)	(46·85)	(47·40)	(50·96)
Housing	124·3	149·9	175·5	207·2	249·7	305·0	801·2
	(21·15)	(21·53)	(21·81)	(22·17)	(22·51)	(22·86)	(24·73)
Household operations	32·8	39·4	46·1	54·3	65·3	79·6	207·6
	(5·58)	(5·66)	(5·73)	(5·81)	(5·89)	(5·97)	(6·41)
Transport	15·6	18·4	21·3	24·5	28·6	33·7	75·0
	(2·65)	(2·64)	(2·65)	(2·62)	(2·58)	(2·53)	(2·31)
Other	93·5	109·7	125·9	147·7	176·1	214·3	567·5
	(15·91)	(15·75)	(15·65)	(15·80)	(15·87)	(16·06)	(17·51)

These projections use the Series E values in Table 3 for age distribution and household size variables and the Series B values in all other instances.

* Figures in parentheses indicate percentage of total personal consumption expenditure.

are precisely the ones that the simultaneous increase in total consumption per head would induce them to consume proportionately less of, and vice versa. Only for automobiles and parts and for furniture and household equipment would the direct and indirect effects of demographic change consistently reinforce each other, but even in these cases the changes would be slight.

CONCLUSION

According to a frequently repeated maxim a gradual transition to a stationary population will bring with it a change in the composition of consumer expenditures: with a zero rate of population growth, there would be a larger proportion of adults and elderly and a smaller proportion of children, and the type and quantity of the goods and services that people consume vary with age. The analysis in this paper does not support this expectation. While in the past such demographic factors as age distribution and household size have, indeed, been statistically significant, with alternative projections of expenditures to the year 2020, assuming that population will grow at rates equivalent to a three-child or a two-child family, the two series did not differ significantly in the allocation of consumer purchases.

This outcome results both from the large growth in total consumption per head characteristic of both projections, and from the tendency of the direct and indirect effects of population change to offset one another. A trend toward zero population growth means an increase in the average age in the population and a decrease in average household size. But it also means a relative rise in levels of living, and the goods and services which an older society would acquire relatively more of are largely those of which a wealthier society would purchase relatively less, and vice versa.

Continuing the projections to a stationary population, which would not have been achieved by the year 2020, is unlikely to influence the results. Even by 2020, the two series differ significantly in age structure and in household size. Moreover, the ratio of total consumption per head in the Series E projection to that in the Series B projection is unlikely to change appreciably thereafter.

A more important limitation might be that the analysis is not sufficiently detailed. What are referred to here as minor categories of consumption actually comprise more than 80 smaller expenditure categories, and an analysis based on these more detailed classifications might show demographic influences hidden in the aggregations that were used.

The extrapolations shown in Tables 4, 5, and 6 are not, of course, predictions for the future. With the poor record demographers have built up in their efforts to anticipate demographic changes, especially those in birth rates, it is unlikely that population forecasts can be accurate, and the consumption patterns that accompany them even less so. Rather, the projections are simply the logical outcomes of specified assumptions. It is assumed, for example, that the relations given in Table 2 will hold in the future. This might not be true. Changing consumer tastes, such as a reduced reliance on the private automobile as a primary means of transport, could cause certain consumption items to fall in relative importance and others to rise. Important new products could similarly affect the balance.

In short, it does not appear that the economy would have to undergo any major adjustments in the composition of output of consumer goods and services as the result of a slower rate of population growth. While after a time the volume of output would be smaller in the lower-fertility population, its structure would seemingly remain essentially unaffected.

287

288

Population and Development Review - Income Distribution and the Family Y. Ben Porath (Ed.) Supp To Vol. 8, 1982

The Direction of Intergenerational Transfers and Demographic Transition: The Caldwell Hypothesis Reexamined

Robert J. Willis

> Although demographic transition theory has been characterized as a historical generalization in search of an explanatory principle, writings on the transition have in fact offered a plethora of explanations for fertility decline. [Bulatao, 1979: 1]

Among the plethora of possible explanations for the demographic transition, singular importance has recently been given in a series of papers by Caldwell to a change in the direction of intrafamily intergenerational transfers and an associated shift in parental motivation to bear and rear children from economic to psychic advantages, which accompanies socioeconomic development and modernization. As a brief summary of his argument, Caldwell writes

> The fundamental thesis is that fertility behavior in both pretransitional and post-transitional societies is economically rational within the context of socially determined economic goals and within bounds largely set by biological and psychological factors. Two types of society can be distinguished: one of stable high fertility, where there would be no net economic gain accruing to the family (or to those dominant within it) from lower fertility levels, and the other in which economic rationality alone would dictate zero reproduction. The former is characterized by "net wealth flows" from younger to older generations, and the latter by flows in the opposite direction. These flows are defined to embrace all economic benefits both present and anticipated over a lifetime. [1978:553]

The fertility transition itself is seen as a period of destabilization caused by a shift in the direction of the net flow of wealth (or net intergenerational transfers, to use the terminology that I adopt in this paper) whose underlying causes incorporate a number of economic and social changes associated with economic development, including such factors as a shift from familial to nonfamilial modes of production and the introduction of mass education, both of which raise the costs of children and undermine the moral basis of intergenera- 288 tional relationships within the traditional family (see, e.g., Caldwell, 1976, 1978, 1980).

In two recent papers (1980, 1981), I have begun a long-term project whose goal is to construct an economic theory of the demographic transition. The initial work on this project has a number of parallels with Caldwell's theory and with an independent statement of a similar view by Ryder (1976); my work also emphasizes the connection between fertility and the intergenerational distribution of income and consumption within the family and attempts to trace the mechanisms by which socioeconomic development may lead to fertility transition.

This paper represents both a summary and extension of my recent work and related theoretical literature with emphasis on its application to demographic transition theory. At the theoretical level, I have been attempting to integrate microeconomic theories of fertility and family behavior associated with, among others, Becker (1960) and Willis (1973), and the altruistic model of family behavior developed by Becker (1974, 1981) and Becker and Tomes (1976), into the dynamic general equilibrium framework of the overlapping general model invented by Samuelson (1958). The broad purpose of this endeavor is to develop a theory that takes explicit account of the interaction between the family and other social and economic institutions such as firms, markets, and government during the course of economic growth and development, and that deals with both the causes and consequences of changes in family fertility behavior, investment in human and physical capital, and intergenerational transfers at both the microlevel and macrolevel of analysis.

The new theoretical contribution in this paper is to bring together and demonstrate the connection between two seemingly unrelated lines of research that both originate with Samuelson's overlapping generations model. One line, represented by the important papers of Arthur and McNicoll (1978) and Lee (1980), applies the tools of formal demography to the Samuelson model to derive expressions that relate the effects of variations in the rate of population growth to steady-state per capita economic welfare. The second line began with an excellent paper by Gale (1973), which was extended significantly in a recent PhD thesis by Kim (1981). The focus of their work concerns the role of money and financial institutions in permitting an economy made up of overlapping generations to provide decentralized competitive markets for borrowing and lending that allow the achievement of a so-called Golden Rule state, in which the rate of interest is equal to the rate of population growth and per capita welfare is maximized. In both lines of research, the pattern of life-cycle production and consumption and the direction of intergenerational transfers can be shown to play a crucial role.

The main substantive conclusions of the present paper follow from the assumption that modern economic development is driven by sustained technological change that is translated into growth in real income via investments in human and physical capital. At a given rate of population growth, a rise in the productivity of human and physical capital due to improved technology increases the mean age at which it is optimal for an individual to produce his lifetime wealth relative to the age at which he consumes his lifetime wealth, in steady state. This, in turn, tends to shift the steady-state direction of net intergenerational transfers from a situation in which the young transfer to the old under primitive technological conditions to a situation in which the old transfer to the young with more advanced technology. In the former case a rise in population growth increases per capita wealth and welfare, and in the latter case it reduces them.

The potential power of this framework to explain demographic transition is perhaps best illustrated by the following application of its comparative steady-state implications. Assume a microeconomic model of fertility such that parents have a positive preference for number of children and that they are perfectly altruistic in the sense that they guide and finance the wealth-maximizing pattern of investments in their children's human and physical capital so that the rate of return to capital is always kept equal to the rate of population growth. In this case, the level of fertility will always be chosen such that the net direction of intergenerational transfers is from parents to children although children may make transfers back to their parents over part of their lifetime. Further assume that parental preferences are such that the fraction of parents' lifetime wealth that is devoted to children is constant across societies that differ in technology and hence in levels of income and wealth. Under these assumptions, in steady state the more technologically advanced the society: (1) the lower the rate of population growth; (2) the higher the amount of physical capital per person; (3) the higher the amount of human capital per person; (4) the later in his life cycle an individual's net productivity will exceed his consumption; and (5) the more likely it is that parents will make positive net bequests to their children at death rather than rely on their children for old age support.

In the next section of this paper the theoretical framework is developed and the main analytical results are presented. In the following section a number of the implications of the framework are outlined informally. The paper concludes with an alternative version of the Caldwell hypothesis that reverses the line of causation that he emphasizes. This hypothesis, which I call the "parental altruism" hypothesis, emphasizes the crucial role that the family plays in retarding or promoting economic development.

Overlapping generations and intergenerational relations

Any viable society must solve a set of economic problems associated with the maintenance of its productive potential over time and the distribution of its output of goods and services to its members. Since Adam Smith, economists have stressed the benefits of the institutional specialization and division of labor and the gains from trade that may be achieved by a decentralized system of prices and markets. Beginning with Smith, economists have also freely con-

290

ceded that there are many reasons that decentralized competitive markets may fail to exist or may function improperly. Only recently, however, have economists realized that the simple demographic fact that a population consists of individuals belonging to different generations may imply that there are barriers to trade that cannot be overcome by the operation of any Smithian Invisible Hand based on the pursuit of individual self-interest.

This barrier, which has come to be called the overlapping generation friction (Wallace, 1980), was first illustrated by Paul Samuelson in his seminal consumption loans paper (1958). To circumvent this barrier to intergenerational trade, Samuelson suggested that society must resort to nonmarket institutions or "social contrivances." He mentioned three possibilities: (1) a system of private intergenerational transfers within households; (2) a public tax and transfer system such as social security; or (3) the creation of real or artificial assets such as paper money, gold, land, or government bonds that can serve as a store of value.

In this section, I will first use the simplest version of Samuelson's overlapping generations model to introduce some issues, concepts, and terminology that will serve as the basis of a survey of recent applications of this model to macroeconomic aspects of the relationship between the direction of intergenerational transfers and population growth. In the last part of the section, I present some new theoretical results that will help clarify some issues concerning the microeconomic role of family structure, fertility behavior, and economic development.

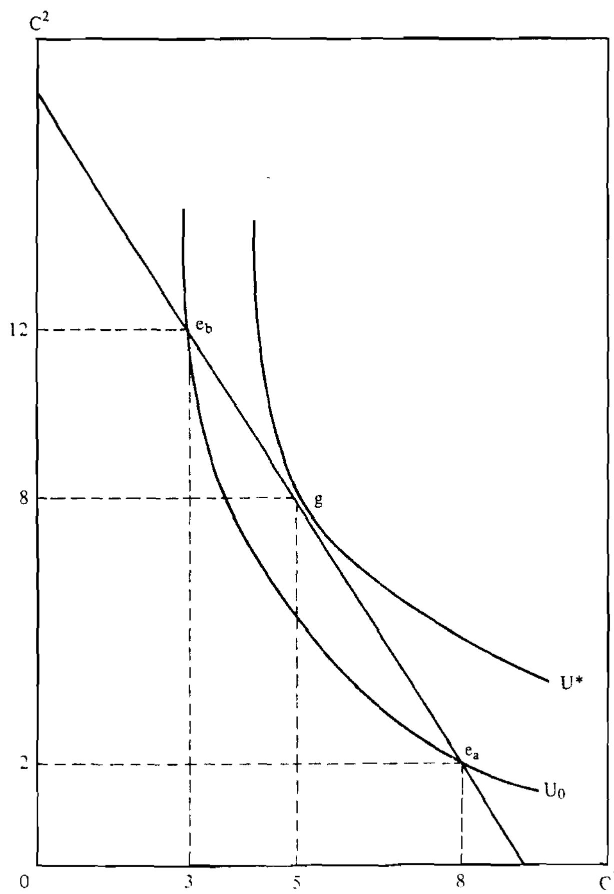
The two-period model

Consider a simple abstract economy in which identical individuals live for two periods of time so that in any given period the population is made up of the "young" and the "old" in proportions determined by the rate of population growth. For example, if each father has one son the rate of population growth is zero and there are equal numbers of each age; if each father has two sons the population doubles each generation and there are twice as many young as old each period. More generally, if the fertility rate is B sons per father, the ratio of young to old is B and the rate of population growth is $n = B - 1$ per generation.

Each individual is endowed with an income of Y^1 units of perishable consumption goods when he is young and Y^2 units when he is old. This endowment may be interpreted to reflect the labor productivities of the young and old. For example, the endowment $(Y^1, Y^2) = (8,2)$ corresponds to the endowment point e_a in Figure 1 where the horizontal axis, marked C^1, measures the consumption of a typical young person and the vertical axis, marked C^2, measures the consumption of a typical old person. If the young were less productive than the old, the endowment might be $(Y^1, Y^2) = (3,12)$ as indicated by point e_b in Figure 1. In the absence of any transactions, each individual in any generation will simply consume his endowment in each period of his life, achieving the level of lifetime utility indicated by the indifference curve passing through his endowment point. The endowment points e_a and e_b in Figure 1 are chosen so 291 that they each provide the same level of lifetime utility indicated by the indifference curve U_0.

Now suppose that the population is doubling every generation (i.e., the growth rate is almost 3 percent per year and the length of a generation is 25 years) so that there are two young persons for every old person in the population in any given period. Also assume that the endowment point is e_a in Figure 1. Clearly, it would be feasible for each person in every generation to achieve a life-cycle consumption path anywhere along the straight line passing through points e_a, g, and e_b in Figure 1. Because this line determines all feasible consumption paths, I shall call it the "social budget constraint." For example, in each period assume that a social security tax of 3 units of consumption is levied on each young person and that the proceeds are transferred to the old. Since there are twice as many young as old, each old person receives a social security transfer of 6 units. The post-transfer age distribution of consumption is $C^1 = 8 - 3 = 5$ and $C^2 = 2 + (2 \cdot 3) = 8$ which corresponds to point g in Figure 1. Given that population growth, endowments, and the tax and transfer program

FIGURE 1 Consumption of the young, C^1, versus consumption of the old, C^2



292

remain the same in the next period, point g also corresponds to the consumption path that each young person expects to follow during his life.

In fact, of all the points along this social budget constraint, the highest attainable level of lifetime utility for members of each generation is attained at point g where the indifference curve U^* is tangent to the constraint. Economists call U^* the Golden Rule level of utility. In this example, the direction of intergenerational transfers required to achieve the Golden Rule was from the younger to the older generation. Obviously, if the endowment point is at point e_b instead of e_a, the direction of transfers required to reach point g is reversed, that is, each old person is taxed 4 units and each young person receives a transfer of 2 units.

An important feature of the social budget constraint should be noted: the rate of interest is equal to the rate of population growth along the constraint. Samuelson calls this the biological rate of interest. The proof is simple. By definition, the rate of interest is the rate at which future consumption can be increased by sacrificing current consumption. Thus, if the rate of interest is r, a young person who gives up one unit of consumption may increase his old age consumption by $1+r$ units. Since the social budget constraint indicates the trade-off between current and future consumption, it follows that its slope is equal to $1+r$ in absolute value. But the budget constraint was constructed by asking at what rate the consumption of the old can be increased by taxing the young. The answer to this question is that for each unit that the young are taxed, the old receive $B = 1+n$ units. Hence, $r = n$.

The social budget constraint in Figure 1 may now be interpreted as the wealth constraint of a representative individual born in any period t. That is, each point on the constraint satisfies the equation

$$Y_t^1 + \frac{Y_{t+1}^2}{1+r} = C_t^1 + \frac{C_{t+1}^2}{1+r}$$

which, in words, means that the present discounted values of the individual's life-cycle income and consumption streams are equal when these streams are discounted at the biological interest rate $r = n$.[1]

Suppose that there exists a competitive market in which individuals may borrow or lend at the biological rate of interest. A typical young person who faces the wealth constraint in Figure 1 maximizes his lifetime utility, increasing it from U_0 to U^* by choosing a life-cycle consumption path at point g. Consider the three ways that this might be achieved. (1) If his endowment point were at e_a, the young person would lend 3 units of his endowment at the biological interest rate of 100 percent. In old age, he would then use the principal and interest to increase his consumption by 6 units above his endowment. (2) Conversely, if his endowment point were at e_b, he would borrow 2 units when young and decrease his old age consumption by 4 units to repay the principal and interest on this loan. (3) Finally, if his endowment point were at g, he would be neither a borrower nor a lender. 293

In an important paper, Gale (1973) pointed out that Samuelson only considered situations corresponding to the first of these cases in which the young wish to be lenders when the rate of interest is equal to the rate of population growth. Fittingly, he called this the Samuelson case. He called the second situation in which the young wish to be borrowers the classical case and the final situation the balanced case. As I shall argue below, the Gale classification turns out to be very important for a range of questions concerning the role of the family and other economic institutions in economic development and demographic transition. It also has a close but hitherto unnoticed relationship to some superficially unrelated results obtained by economic demographers such as Arthur and McNicoll (1978), Lee (1980), and Willis (1980, 1981) concerning the effects of changes in the rate of population growth on economic welfare. Hence, it is appropriate to consider the meaning of this classification carefully at this point.

According to ordinary accounting principles, a lender has a positive credit that is offset by the negative credit or liability of a borrower. Later, when the transaction has been completed, the borrower's liability is discharged by the payment of his debt. From this point of view, the unbalanced Samuelson and classical cases are very puzzling.

Consider the Samuelson case first. If all the young people in society wish to lend when $r = n$, they have positive aggregate credit balance. We may then ask who are the borrowers on the other side of the transaction to whom we may assign the offsetting liability? Clearly, they cannot be the current old people because they will be dead in the next period and in no position to repay their debt. The only people who could repay the loan are the young of the next period who are as yet unborn. Thus, it is the unborn who have an aggregate liability that offsets the aggregate credit of this period. Clearly, however, the unborn young of the next period cannot meet with the young of this period to sign a loan contract. This barrier to intergenerational trade is what I referred to earlier as the overlapping generation friction, and it is this barrier that Samuelson's three social contrivances are designed to overcome.

One of these contrivances, a pay-as-you-go social security system, provides a familiar example. In one sense, a social security system is simply a program that taxes workers (i.e., the young) and transfers the proceeds to the retired (i.e., the old) each period. The taxes and transfers offset one another and no aggregate credit or liability is carried into the next period. When this point is made it invariably generates strong protests, as the current social security debate in the United States vividly illustrates. The current old contend that they are not welfare recipients; rather, they argue that they are entitled to the social security benefits that they receive as payment for the social security taxes that they paid in the past. In effect, their argument implies that the current young incurred a liability before they were born that they are morally bound to discharge by paying social security taxes now. If the young live up to the terms of this "social contract," they will simultaneously discharge their liability to the current old and impose a liability on the unborn young of the next period.

294

295

The classical case is just the reverse of the Samuelson case. At the biological interest rate, $r = n$, the young wish to be net borrowers, the aggregate credit balance of the current young is negative, and this may be considered to be offset by an aggregate positive credit of the unborn young of the next period. Similarly, the old of this period discharge their aggregate liability by making transfers to the young.

The overlapping generation model that I have used to illustrate the Samuelson and classical cases is extremely simple in that (1) people are assumed to be identical and to live only two periods, (2) labor is the only factor of production, and (3) consumption goods are not storable. However, many of the important results of this model carry over to more complex models in which generational overlap is greater because it is assumed that people live many periods or in continuous time and are subject to mortality risks at each age; people have different tastes and different life-cycle productivity profiles; and output is produced with inputs of natural resources, physical capital, and human capital, as well as "endowed labor" (i.e., the exogenous productive capacity of a worker). In particular, Gale (1973) proved that an economy with a constant population growth rate n may be either in a balanced or in an unbalanced (i.e., Samuelson or classical) Golden Rule competitive equilibrium with $r = n$ no matter how many periods individuals live. The main difference between this case and the two-period case is that here a balanced equilibrium may involve trade between members of different generations, whereas in the two-period model the balanced equilibrium is autarkic (i.e., individuals simply consume their endowment).

For example, suppose that people live for three periods as "young," "middle-aged," and "old" and that only the middle-aged have income. In this case, a young person wishes to borrow from a middle-aged person to finance his current consumption, and the middle-aged person is willing to lend to the young person in order to finance his old age consumption with the repayment of principal and interest by the person to whom he made the loan in the last period. Of course, the institutional mechanism by which this transaction takes place need not involve a face-to-face meeting between the borrower and the lender.

More commonly, the middle-aged person views himself as a saver when he places part of his income in a bank or some other financial intermediary that promises to pay him an interest rate r and stands ready to allow him to withdraw his savings plus accrued interest whenever he chooses. The bank then uses these deposits to make loans to the young. If we assume that the cost of running the bank is zero and that there is competition among banks, and if we neglect default risk, the bank will be forced to charge an interest rate on loans equal to the interest rate r that it pays depositors. In a competitive balanced equilibrium, the interest rate varies such that the aggregate demand for loans is equal to the supply of loanable funds. In equilibrium, the aggregate liabilities of each bank (i.e., its deposits) are equal to its aggregate credits (i.e., its loans) so that its aggregate credit balance is zero. Hence, the aggregate credit balance of the banking system as a whole is zero. 295

Only in a special case, however, does balance occur when the competitive rate of interest is equal to the rate of population growth. Typically, there is either (1) an excess demand for loans when $r = n$ because the young wish to borrow more than the middle-aged wish to save or (2) an excess supply of loans when $r = n$ because the middle-aged wish to save more than the young wish to borrow. The first case corresponds to Gale's classical case because the aggregate liabilities of borrowers exceed the aggregate credit of lenders (i.e., savers) in each period. The second case corresponds to his Samuelson case because the population as a whole has a negative credit balance.

Ordinary market institutions such as competitive borrowing or lending markets with or without financial intermediaries cannot support an unbalanced Golden Rule equilibrium with $r = n$. At the same time, the absence of such institutions in a given society implies that individuals in that society fail to achieve the highest possible level of per capita welfare—i.e., the Golden Rule level of lifetime utility—because they are unable to achieve the best feasible pattern of life-cycle consumption given their pattern of life-cycle income. This is illustrated for the two-period model in Figure 1 by the fact that individuals in a balanced (i.e., autarkic) equilibrium achieve the utility level U_0 instead of the higher Golden Rule level U^* when their endowment point is at either e_a or e_b, which correspond, respectively, to the Samuelson and classical cases. In the former case, individuals are relatively affluent while young and suffer a relatively impoverished old age; the reverse is true in the latter case.

In the two-period case, the direction of intergenerational transfers required to achieve the Golden Rule is from the younger to the older generation in the Samuelson case and the reverse in the classical case. The definition of the direction of intergenerational transfers required to achieve the Golden Rule is less clear in the general case in which people live more than two periods. For instance, in the three-period borrowing and lending example given above, the middle-aged people, in effect, make transfers both to the young (in the form of loans) and to the old (in the form of repayment of principal and interest on the loans that they received in the previous period).

Note, however, that the Gale classification tells us the net direction of intergenerational transfers required to reach the Golden Rule. If the equilibrium is balanced when $r = n$, the transfers to the young equal the transfers to the old—that is, the aggregate population is neither a net borrower nor a net lender—and we may say that net intergenerational transfers are zero. In the Samuelson case, the aggregate population is a net lender, transfers to the old exceed transfers to the young, and the net direction of transfers is from the younger to the older generation. Conversely, in the classical case, the population is a net borrower and the net direction of intergenerational transfers is from the older to the younger generation.

Population growth and welfare: The general model

Whether an economy will be balanced, Samuelson, or classical depends on three underlying facts: (1) the preferences of individuals for current versus

296

297

future consumption; (2) the shape of the life-cycle income profile of individuals in each generation; and (3) the rate of population growth. If we abstract from the institutional mechanisms required to support the Golden Rule state and take the rate of population growth as fixed and equal to the rate of interest, consumers who live T periods find an optimal life-cycle consumption path, $[C(1), C(2), \ldots, C(T)]$ by maximizing their lifetime utility function $U = U(C(1), C(2), \ldots, C(T))$ subject to their lifetime wealth constraint

$$W = Y(1) + \frac{Y(2)}{1+r} + \cdots + \frac{Y(T)}{(1+r)^{T-1}}$$

where $[Y(1), Y(2), \ldots, Y(T)]$ is the life-cycle income profile. This maximization problem is illustrated in the two-period case in Figure 1.

We are now in a position to exploit a hitherto unnoticed duality relationship between the Gale classification and some results first obtained in an elegant paper by Arthur and McNicoll (1978), which were reviewed and extended in another excellent paper by Lee (1980). Arthur and McNicoll wished to determine whether an increase in the steady-state rate of population growth increases or decreases the level of lifetime utility of a typical member of the population when it is assumed that the rate of interest is equal to the rate of population growth. Following Samuelson (1976), they noted that an increased rate of population growth reduces the amount of output available for consumption in the steady state because more resources must be devoted to investment in order to equip each worker with a given amount of capital equipment. Because this "capital-widening effect" is always negative, a lower rate of population growth always increases individual welfare in a population without age structure (i.e., in which people live one period so that there is no generational overlap). When there is generational overlap, however, an increase in n may lead to a positive "intergenerational transfer effect," which, in principle, may outweigh the negative capital-widening effect; alternatively, the intergenerational transfer effect may be negative so that it reinforces the capital-widening effect.

The intuition underlying the proximate determinants of the sign of the intergenerational transfer effect is most easily seen by invoking the definition of the direction of net intergenerational transfers that follows from the Gale classification. If the economy corresponds to the Samuelson case, the aggregate population is a net lender, and the direction of net intergenerational transfers is from the younger to the older generation. An increase in the rate of population growth increases the biological rate of interest, and, in turn, an increase in the rate of interest benefits lenders. Hence, the sign of the intergenerational transfer effect is positive. Obviously, the classical case works in the opposite direction because an increase in the interest rate reduces the welfare of net borrowers. The reader can confirm these results diagrammatically in Figure 1. A rise in the rate of population growth causes the social budget constraint passing through the endowment point to become steeper. Given that the 297 consumption point is initially at point g, it is clear that the attainable level of welfare is higher after the increase in the rate of population growth when the endowment point is at e_a, but lower if it is at e_b.

Before displaying the duality between the Gale classification and the Arthur and McNicoll results, it is convenient to merge the capital-widening effect with the intergenerational transfer effect, which are presented separately by Arthur and McNicoll. This is done by defining the net income, $Y(x)$, of an individual of age x to be equal to the individual's age-specific "endowed" labor productivity minus his age-specific investment in physical or human capital plus the age-specific returns to his past investment in physical or human capital. The resulting life-cycle net income profile, $[Y(1),Y(2), \ldots ,Y(T)]$, has some negative or zero elements at early ages because the cost of investment in a child's human capital exceeds his labor productivity. Depending on the extent and duration of human capital investment, the net labor productivity profile eventually becomes positive, rises to a peak during adulthood, and begins to fall at old age, reaching zero upon retirement (see Ben-Porath, 1967).

For simplicity, I have dispensed with the intermediation by firms between savers and investors by treating physical capital as if individuals accumulate machines with their net saving at each age, paying the cost of the new machines as they are acquired and receiving the rental income from their accumulated stock of machines at each age. Dissaving during old age is then treated as if the elderly were selling off their stock of machines (presumably to younger people) as they approached the end of their lives.

Since machines can be sold, but human capital is embodied in the individual and will become unproductive upon death or retirement, capital theory implies that people shift the mix of their total investment from human to physical capital as they age. Total net investment declines and becomes negative eventually if individuals do not have a bequest motive. If there are bequests, net investment may be positive through life. The life-cycle net income profile is therefore likely to be lower than the net labor productivity profile during the middle years of life when individuals are net savers and higher than the net labor productivity profile in old age. It is important to note that age-specific net income provides a measure of the net social product of an individual in terms of both his labor productivity and the net productivity of the capital that he owns.

Given this definition of age-specific net income, per capita net income is equal to per capita consumption in each period, where the per capita measures are computed by taking the sum of age-specific net income or consumption in each period, weighted by the number of individuals in each age group in that period. Let us define the following terms: x $(=1, \ldots, T)$ is the age of an individual at his next birthday; $p(x)$ is the probability that he survives to age x conditional on survival to age $x - 7$; $C(x)$ and $Y(x)$ are the age-specific flows of consumption and net income during the xth year of life; $g = 1+n$ is the population growth factor (i.e., the net reproduction rate) where n is the rate of population growth per period; $\rho = 1+r$ is the discount factor where r is the rate of interest; and age is measured in discrete units $x = 1,2, \ldots ,T$, where T is the upper limit to the length of life. In what follows, it is assumed that investment 298 in physical and human capital is always adjusted so that the rate of return to capital is equal to the biological interest rate, that is, $r = n$ or equivalently, $\rho = g$. For expositional ease, I shall also adopt the convention of referring to ρ and g rather than r and n as the rates of return to capital and the rate of population growth.

I define the social budget constraint as

$$N_t(0) \sum_{x=1}^{T} \frac{p(x)C(x)}{g^{x-1}} = N_t(0) \sum_{x=1}^{T} \frac{p(x)Y(x)}{g^{x-1}} \tag{1}$$

where $N^t(0)$ is the number of newborns living in the population in period t and $N_t(x) = p(x)N_t(1)g^{-(x-1)}$ is the number of persons of age x living in the population. The social budget constraint simply verifies that aggregate consumption is equal to aggregate net income. For notational convenience, let C_L and Y_L denote aggregate consumption and net income divided by the number of newborns.

When $\rho = g$, as it does with Golden Rule savings and investment, it follows immediately from equation (1) that the present discounted values of expected lifetime consumption and production are equal. That is,

$$W(\rho) = \sum_{x=1}^{T} \frac{p(x)C(x)}{\rho^{x-1}} = \sum_{x=1}^{T} \frac{p(x)Y(x)}{\rho^{x-1}} \tag{2}$$

where $W(\rho)$ is an individual's expected lifetime wealth at the moment that he is born, evaluated at the discount factor ρ. Note that $W(\rho) = C_L = Y_L$ so that lifetime wealth may be computed from cross-section data under steady-state conditions with $\rho = g$.

Following Arthur and McNicoll (1978) and Lee (1980), I define the average ages of consuming a_c and producing a_y, calculated from the cross section, as

$$a_c = \frac{1}{C_L} \sum_{x=1}^{T} \frac{xp(x)C(x)}{g^{x-1}} \tag{3}$$

$$a_y = \frac{1}{Y_L} \sum_{x=1}^{T} \frac{xp(x)Y(x)}{g^{x-1}} \tag{4}$$

Since $\rho = g$, a_c may also be interpreted as the mean age at which a person consumes his lifetime wealth, $W(\rho)$, and a_y as the mean age at which he produces his lifetime wealth.

Their major results, modified to accord with my definition of net income and expressed in discrete instead of continuous time, are contained in the following two propositions: 299

Proposition 1: An increase in the population growth rate increases, does not affect, or decreases the steady-state level of per capita welfare when $\rho = g$, depending on whether the mean age of consumption (a_c) is greater than, equal to, or less than the mean age of production (a_y).

This result is given quantitative expression in

Proposition 2: The steady-state elasticity of lifetime wealth, $W(\rho)$, with respect to an increase in the rate of population growth is

$$\epsilon = \frac{\partial \ln W(\rho)}{\partial \ln g} = a_c - a_y \tag{5}$$

Proposition 2 is a slightly different expression of Proposition 2′ in Lee (1980: 1,134) in which he shows that $a_c - a_y$ is the uniform percentage amount by which consumption may be increased or decreased when population growth is increased by a given percentage amount relative to any arbitrary life-cycle consumption path satisfying the social budget constraint.

In Propositions 1 and 2, the economy is assumed to be on a Golden Rule (GR) growth path in which savings and investment are adjusted so as to equate the rate of return to capital to the rate of population growth (i.e., $\rho = g$), thereby achieving the maximum level of steady-state per capita welfare corresponding to any given rate of population growth. However, it is clear from these propositions that the level of per capita welfare on a GR path varies with the rate of population growth. Following Samuelson (1975), I define the "Goldenest" of Golden Rule (GGR) paths as that rate of population growth g^* for which the GR level of per capita welfare is maximized. The characterization of GGR population growth within the T-period framework can then be stated as

Proposition 3: If it exists, the GGR rate of population growth g^* that maximizes per capita welfare occurs when $\rho^* = g^*$ and $a_c = a_y$.

That $a_c = a_y$ is a necessary condition for GGR population growth follows immediately from the fact that an individual's lifetime utility is an increasing function of his wealth, $W(\rho)$, and that $W(\rho)$ is maximized when equation (5) is set to zero. The clause "if it exists" prefacing Proposition 3 calls attention to the fact that $\rho^* = g^*$ and $a_c = a_y$ are necessary but not sufficient conditions for an "interior" maximum of steady-state per capita welfare. Intuitively, g^* represents a situation in which the welfare gain from positive intergenerational transfers from the young to the old associated with higher population growth is offset at the margin by the welfare losses due to capital-widening and increased child-rearing costs. Under certain conditions, however, a rate of population growth g^* such that $a_c = a_y$ may be associated with minimum rather than maximum per capita welfare. Under other conditions, it

300

may be the case that $a_c > a_y$ or $a_c < a_y$ for all possible population growth rates, $0 < g < \infty$. In this paper, I shall assume that there does exist a g^* such that $a_c = a_y$ and that this value of g^* corresponds to maximum per capita welfare.[2]

The aggregate credit balance

A model with essentially the same formal structure was studied by Gale and recently extended significantly by Kim (1981), but from a very different point of view. They concentrated almost exclusively on the possible role of institutions such as money and financial intermediaries in supporting Golden Rule equilibria in individualistic competitive markets. However, as I shall show, there are important duality relationships between Gale's classification and Propositions 1–3 above. In addition, the Gale–Kim results suggest a reason why intergenerational transfers within the family play a vital role in the process of economic development, and their framework provides a natural starting point for microeconomic models of family structure, fertility, and the interaction of families and market institutions.

Let $A(x,\rho)$ be the net financial assets of a given person on his xth birthday who may borrow or lend at the market interest rate r where $\rho = 1 + r$. Assume that the individual begins and ends life with no financial assets so that $A(0,\rho) = A(T,\rho) = 0$ and, for simplicity, neglect mortality prior to T by assuming that $p(x) = 1$ for $x < T$. If $A(x,\rho)$ is positive, the person has positive credit and is, in effect, a net lender; if $A(x,\rho)$ is negative, he has negative credit and is a net borrower. Since I have defined net income, $Y(x)$, as net of the costs of current investment and gross of income from the capital that the individual has accumulated, note that $A(x,\rho)$ does not include claims to physical capital, although, of course, the future earning power of an individual's human or physical capital is implicitly used as collateral for any borrowing he may do.

By definition, the life-cycle path of financial asset accumulation and decumulation can be described by

$$\begin{aligned} A(1,\rho) &= Y(1) - C(1) \\ A(2,\rho) &= Y(2) - C(2) + \rho A(1,\rho) \\ &\vdots \\ A(T,\rho) &= Y(T) - C(T) + \rho A((T-1),\rho) = 0 \end{aligned} \tag{6}$$

where $\rho A(x - 1,\rho)$ is the net interest income from positive assets or the net interest payments on negative assets of a person of age x.

Note that the first equations in this series occur during childhood when net income is negative because labor earnings are zero or negligible and the costs of investing in human capital are positive. If children were independent agents they would be borrowing during this period to finance both consumption and human capital investment so that $A(x,\rho)$ would grow progressively more negative and the child's debt would grow. In reality, of course, it is parents who finance the human capital investment and consumption of the child. This 301

302

illustrates the point often made in the human capital literature that the family serves as a nonmarket surrogate capital market for the child.[3]

From this point of view, parents who finance a child's consumption and investment according to the set of equations (6) so as to maximize his lifetime utility function $U(C(1),C(2), \ldots, C(T))$ are acting as perfect agents for him in the sense that they are doing exactly what he would do to maximize his lifetime wealth and utility if he were able to act independently. I shall later consider whether it is reasonable to expect parents to follow such "Golden Rule" behavior; I will not consider the more troublesome question of whether it is reasonable to assume that children have well-defined lifetime preferences. Perhaps it is easiest to regard the parts of the lifetime utility function pertaining to childhood consumption as reflecting the parents' view of what children *should* prefer and the parts pertaining to adult consumption as reflecting what their parents *taught* them to prefer during their upbringing.[4]

At some point during the life cycle, the individual begins to produce more than he consumes—that is, $Y(x)$ becomes greater than $C(x)$—and he begins to pay off the debt that he has incurred in the first phase of his life. There are a number of institutional mechanisms by which he might pay this debt. Most obviously, if he had originally contracted to borrow from another person or from a bank, the transfer of the excess of his current income over current consumption simply represents the other side of a private quid pro quo exchange.

Alternatively, if we assume that his parents financed his initial consumption and investment, he owes no legal debt to anyone. He may nonetheless make net transfers back to his parents, although it is not obviously in his narrow self-interest to do so. He may have children and transfer resources to them to finance their consumption and investment in human capital, although again this raises the question of whether it is in his self-interest to do so. Still another possibility is that he pays social security taxes, which are transferred to his and other people's elderly parents, and school taxes, which subsidize the human capital investment of his and other people's children. Finally, he may simply acquire nominal assets such as paper money, gold, or government bonds to hold as a store of value. As Samuelson (1958) first pointed out, under certain conditions such nominal assets may yield the biological rate of interest $\rho = g$. However, within a given period the individual who gives up real goods and services in exchange for an intrinsically worthless piece of paper is, in truth, making a net transfer to whoever held the piece of paper prior to the transaction. If the "social contrivance of money" is functioning properly, this transfer will be repaid with interest later when the individual trades his money for goods with someone else.

What all of these institutional mechanisms have in common from a social point of view is that transfers from the younger to the older generation represent a means by which an individual pays off some of, all of, or more than the social cost of his upbringing. As I shall show shortly, the answer to the question of whether an individual ever pays off the social cost of his own

302

upbringing is given by the sign of the aggregate credit balance in the society as a whole. If the aggregate credit balance is positive, the younger generation makes net transfers to the older generation; if it is negative, the older generation makes net transfers to the younger generation. From the point of view of a purely self-interested older generation, children are net economic assets in the former case and they impose net economic costs in the latter case.

Later, I shall argue that parental preferences for the number and quality of children imply that the latter case tends to hold empirically. In effect, such parental preferences imply that the older generation is willing to make "altruistic" net transfers to the future generation because of an interest in the number and welfare of its offspring. In addition, I shall argue that parental altruism may be a requirement for a successful transition to a modern high-income, low-fertility society in response to technological change.

It is, however, possible for individuals to pay the social costs of their upbringing if $Y(x)$ rises sufficiently fast relative to $C(x)$ so that $A(x,\rho)$ becomes positive. If $A(x,\rho)$ does become positive, an individual becomes a net lender for a portion of his life, in which case he must become a net borrower (or dissaver) toward the end of his life in order to end up with a zero credit balance. If a sufficient portion of the population has positive $A(x,\rho)$, the aggregate credit balance may be positive.

The value of the aggregate credit balance in a given time period is simply the sum of the credit balances of each of the individuals living in that period as indicated in the set of equations (6). Following our previous convention of dividing aggregate values by the number of newborns in the population, we find that the aggregate credit balance per newborn is

$$A(\rho) = \sum_{x=1}^{T} \frac{A(x,\rho)}{g^{x-1}} = \sum_{x=1}^{T} \frac{Y(x)-C(x)}{g^{x-1}} + \frac{\rho}{g} \sum_{x=1}^{T} \frac{A(x,\rho)}{g^{x-1}} \tag{7}$$

From the social budget constraint in equation (1), the first term in the right-hand expression is zero. Hence

$$A(\rho)(g-\rho) = 0. \tag{8}$$

Equation (8) admits of two types of solution. First, if the Golden Rule condition $\rho = g$ is met, $A(\rho)$ may be negative, zero, or positive and the economy is, therefore, classical, balanced, or of the Samuelson type. The Golden Rule equilibrium maximizes per capita welfare at any given rate of population growth. Second, if the Golden Rule equilibrium is unbalanced and society does not provide institutions permitting unbalance, the interest rate diverges from the population growth rate (i.e., $\rho \neq g$) until the aggregate supply and demand for loans is equated and $A(\rho)$ is zero. Any such non-Golden Rule balanced equilibrium yields a lower level of per capita welfare than the corresponding unbalanced Golden Rule equilibrium. 303

In the Samuelson case, there is an excess supply of loans and in the classical case there is an excess demand for loans when $\rho = g$. Intuitively, one might expect that the balanced equilibrium corresponding to the Samuelson case would have $\rho < g$ and that corresponding to the classical case would have $\rho > g$ when supply equals demand. Recently, however, Kim (1981) proved that this intuition is correct only under certain conditions involving the nature of lifetime preferences and production technology.[5]

The duality between the Gale classification and the results given in Propositions 1–3 and equation (5) is displayed in the following equation, which is proved in the Appendix:

$$A(\rho) = W(\rho)[a_c - a_y]; \; \rho = g \tag{9}$$

The interpretation of this equation is given in the following two propositions:

Proposition 4: The economy is Samuelson, balanced, or classical in the sense of Gale (1973) as the mean age of consuming (a_c) is greater than, equal to, or less than the mean age of producing (a_y) when $r = n$,

and

Proposition 5:

(a) If $r = n$ and $A(\rho)$ is positive, a representative individual receives over the course of his lifetime net intergenerational transfers from future generations whose present discounted value at birth is equal to $A(\rho)$.

(b) If $r = n$ and $A(\rho)$ is negative, a representative individual gives over the course of his lifetime net intergenerational transfers to future generations whose present discounted value at birth is equal to $-A(\rho)$.

(c) If $r = n$ and $A(\rho)$ is zero, no net intergenerational transfers take place.

Finally, it was stated earlier in Proposition 3 that under certain conditions the rate of population growth that maximizes per capita welfare is g^*, where investment in physical and human capital is carried to the point at which $g^* = \rho^*$ and $a_c = a_y$. The obvious counterpart to this proposition follows immediately from equation (9) and Proposition 4: The "Goldenest" of Golden Rule states is a balanced equilibrium with $A(\rho) = 0$. In this GGR state, each individual born into the society exactly repays the social cost of his upbringing and, from a social point of view, parents' expenditures on children while the children are young and the parents are in their middle years are repaid with interest later via intergenerational transfers from the middle-aged to the old, that is, from middle-aged children to their old parents (see Willis (1981) for an alternative derivation of this result). 304

Implications

The results contained in Propositions 1–5 point toward several issues that are relevant to the Caldwell thesis and, more broadly, to the potentially crucial role of the family in either retarding or promoting the process of economic development. In this section, I shall illustrate some of these issues by using the simplifying analytical device of comparing the steady states associated with different degrees of technological advancement and/or differences in parental demand for children. In the conclusion, I briefly consider the possible implications of the theory under the analytically more difficult but more relevant non–steady-state conditions associated with the historical process of economic development and demographic transition.

Technology, the direction of intergenerational transfers, and population growth

Kim (1981) has shown that the more important capital is in production, the more likely it is that $A(\rho)$ will be negative at any given rate of population growth when $\rho = g$. The logic underlying this result becomes transparent within the framework of this section. The higher the productivity of capital due to technology, the more it pays a wealth-maximizing individual to invest in human and physical capital during the early periods of his life and, consequently, the higher the mean age at which he produces his lifetime wealth (i.e., the higher a_y is). On the other hand, at a given rate of interest, there is little reason to believe that an increase in wealth caused by higher capital productivity has a systematic effect on the mean age at which an individual consumes his wealth (i.e., a_c). For example, if an increase in wealth caused a proportional increase in consumption at each age, a_c would remain constant.

Under conditions of primitive technology that are characteristic of most pretransitional societies, the technical productivity of both human and physical capital tends to be relatively low. At a given rate of population growth, this implies that the mean age of producing tends to be relatively low or, equivalently, that individuals begin producing more than they consume at a relatively early stage of life, even if parents invest optimally in their children. If we assume that high mortality rates in pretransitional societies do not reduce a_y by more than a_c, the theory therefore suggests that $A(\rho)$ is more likely to be positive and the net direction of intergenerational transfers is more likely to be from the younger to the older generation at a given rate of population growth.

An alternative and more interesting way to look at the relationship between technology and the mean age of producing is to begin by rewriting equation (9) to obtain the relationship

$$\pi = -\frac{A(\rho)}{W(\rho)} = a_y - a_c$$

305

where π is a measure of the net social cost of children, $-A(\rho)$, relative to per capita lifetime wealth, $W(\rho)$. Assume that parents have a preference for the number of children. In this case they increase their fertility above the GGR level: they sacrifice the per capita welfare of their children in order to increase the number of children, and cause $A(\rho)$ to be negative. π is then the fraction of their lifetime wealth that parents spend on each child. (If π is negative, it is the fraction of their lifetime wealth that they receive per child.)

Suppose that parental preferences for the number of children are such that π is approximately constant across countries with different levels of technology and, hence, different levels of income and wealth. Under this assumption, the more technologically advanced the society: (1) the lower the rate of population growth; (2) the higher the amount of physical capital per person; (3) the higher the amount of human capital per person; (4) the later in his life cycle an individual's net productivity exceeds his consumption; and (5) the more likely it is that parents make positive net bequests to their children at death rather than relying on their children for old age support.

Instead, suppose that Caldwell is correct in his assertion that the direction of intergenerational transfers shifts during the course of economic development and demographic transition from a pretransitional situation in which the younger generation makes net transfers to the older generation, to a post-transitional situation in which the direction of intergenerational transfers is reversed. Then the pattern described in the preceding paragraph is accentuated because the level of technology and π are positively correlated.

Within my framework, a negative value of π is associated with net transfers from the younger to the older generation. Since both parents and children could be made better off by an increase in the rate of population growth when $A(\rho)$ is negative, this suggests the possibility that fertility is constrained by biological factors in such situations, which is consistent with Caldwell's view and the view of many other scholars.

On the other hand, Mueller's (1976) calculations (which she admits are crude) suggest that the undiscounted sum of lifetime net production minus lifetime consumption is negative in peasant societies. Since the rate of population growth is positive in such societies, her calculations imply that $A(\rho)$ must be positive. If so, it is in the narrow self-interest of parents to reduce their fertility. If, in fact, fertility is biologically constrained this implies that either (a) parents have a preference for number of children or (b) they are unable to control their fertility and, hence, are unable to maximize their narrow self-interest by reducing fertility. Again, there is much support in the literature for each of these positions. It would seem to me well worthwhile to attempt to replicate Mueller's study with better data to attempt to measure $A(\rho)$ in both developing and developed societies.

Even if $A(\rho)$ is negative in all societies, Caldwell's basic point may be correct. Under primitive technological conditions, as suggested above, population growth tends to be relatively high, optimal investment in physical and human capital tends to be low, and each individual tends to begin producing more than he consumes at a relatively early age. If the (not necessarily coresidential) extended family is the social institution through which a child's accumulated "debt" is paid off, it follows that parents (or older family members) receive a flow of net transfers from the young over a considerable period of time. If $A(\rho)$ is negative, this simply means that the child does not fully pay for himself.

The transfers that he provides do, however, reduce the net costs of rearing him from his parents' point of view. Of equal importance, the pattern of life-cycle transfers, first from parents to children and later from children to parents, provides a means by which each individual's life-cycle consumption is allowed to diverge from his life-cycle net productivity pattern in a way that increases the lifetime utility of each individual by shifting consumption from the years of peak productivity in middle-age toward the periods of dependency at the beginning and end of life. If risk and the potential of intrafamily transfers to provide insurance are considered, the economic value of children in the traditional family setting is enhanced still further (see, e.g., Cain, 1981).

In contrast, in a highly developed society in which the technological productivity of investment in human and physical capital is very high and the parental taste for number of children is sufficiently strong, it is possible that wealth-maximizing, Golden Rule investment may require that the direction of intergenerational transfers be from the older to the younger generation over most of an individual's life cycle. For example, in a three-period overlapping generation model (Willis, 1981), I have shown that the income from the physical capital holdings of the elderly may exceed their desired old age consumption in the Golden Rule state. In this case, parents not only pay the costs of rearing their children to adulthood, but also make transfers to them during their adult years and/or leave them possible bequests.

While I have not checked this formally, I believe that the value of $\pi = -A(\rho)/W(\rho)$ need not differ in the situations described in the preceding three paragraphs. In the low-wealth, high-fertility society, the value of $A(\rho)$ is made negative by the relatively large fraction of children in the population; in the high-wealth, low-fertility society, the value of $-A(\rho)$ relative to $W(\rho)$ is kept from becoming too large by the relatively small fraction of young people in the population. However, as wealth increases it is plausible that π increases because the wealth elasticity of demand for number of children is positive, making it more likely that the direction of transfers is from the older to the younger generation throughout the life cycle.

Institutional issues

The institutional structure of the family required to carry out the Golden Rule program of investment in human and physical capital and the pattern of intergenerational transfers needed to achieve the Golden Rule pattern of life-cycle consumption in the technologically primitive society are radically different from those in advanced societies. In the former, the economic returns from children are important to parents. Consequently, the family needs to provide mechanisms to permit parents to capture these returns. One possible mech- 307 anism is for the parents to make the child's inheritance of the family land or other parental assets conditional on his performance in making net transfers to his parents while they are living. Another is for the parents to attempt to inculcate in the child a sense of obligation toward his parents and a sense of loyalty to the family during the period of child rearing.

Such mechanisms are necessary because it is always in the child's narrow self-interest to default on his "debt" by failing to make transfers. Thus, the traditional family needs to discover means to prevent such opportunistic behavior if it is to succeed in achieving the Golden Rule consumption path for each of its members. In the case of a technologically advanced society, however, parents may not expect to receive economic returns from their children at any point. In this case, there is no need for parents to devise means to enforce obedience or family loyalty, and a fully nuclear modern family structure in which the elderly live entirely independently of their grown offspring is possible.

As I have pointed out elsewhere (Willis, 1980, 1981), the development of impersonal asset markets (e.g., money, gold, land) presents a threat to the traditional family because it increases the attractiveness to the young of opportunistic behavior. For instance, if the young can save for their old age by accumulating money, they need not rely on the family for old age security. If parents cannot rely on transfers from their older children, they may react by reducing the human capital investment per child and by allocating more of the child's time to productive labor in order to capture the returns from their investments while they retain physical control of the child.

In effect, the existence of alternative nominal (i.e., intrinsically unproductive) assets such as gold or government bonds creates a divergence between the social and private rates of return to investment in children. That is, private and social returns tend to coincide when intergenerational transfers take place within the family because parents see the connection between their reproductive behavior and their investment in the human capital of the young and the returns to these investments, which they capture partly in the form of the psychic benefits from the number and lifetime welfare of their children and partly from the economic returns that they expect to receive in the form of transfers from their children. On the other hand, when intergenerational transfers from the younger to the older generation take place via social security tax and transfer programs or the accumulation of financial assets during middle age and decumulation during old age, there is a divergence between the private and social rates of return to investments in children. In effect, parents pay the costs of rearing their children, but they receive the economic returns from other people's children. Hence, their private economic incentive to invest in the human capital of their own children is reduced.

This potential divergence between the social and private returns to investment in children is most important when the technological productivity of human and physical capital is relatively low and the transfers from the middle-aged to the older population are most important. I have shown that the divergence disappears when altruistic parents wish to make net transfers to their grown children or leave net bequests to them (Willis, 1981). This suggests that the modern institutional development of financial markets, governmental transfer programs, and the like have a potentially destructive influence in technologically backward societies by reducing parental investment incentives, but they facilitate the efficient allocation of resources and spreading of risk in technologically advanced societies in which relatively wealthy parents wish to make net transfers to their grown offspring. Similar remarks may pertain to the incentives of parents of different levels of wealth in a given society.

Another institutional issue is suggested by a proposition conjectured (but not proved) by Kim (1981). He argues that no decentralized private institution such as a bank or other private intermediary can exist to support a Golden Rule equilibrium in the "classical case" in which $A(\rho)$ is negative. After explaining this conjecture, I shall propose that it suggests an important role for the family in providing the level of investment in human and physical capital required for successful economic development.

To see the logic behind Kim's conjecture, return to the simple two-period model illustrated in Figure 1, where $\rho = g = 2$, the Golden Rule consumption point is at point g where $(C^1,C^2) = (5,8)$, and two alternative endowment points $e_a = (Y^1,Y^2) = (8,2)$ and $e_b = (Y^1,Y^2) = (3,12)$ are considered. Applying the formulas for $W(\rho)$, a_c and a_y presented in the previous section in equations (2)–(4), we see that $W(\rho) = 9$, $a_c = 1.44$ and $a_y = 1.00$ at point e_b and $a_y = 1.67$ at point e_b. From equation (9), it follows that point e_a represents the Samuelson case with $A(\rho) = 3.6$ and $\pi = -A(\rho)/W(\rho) = -.4$, indicating that 40 percent of the lifetime wealth of each individual comes from transfers from the younger generation. In contrast, point e_b represents the classical case with $A(\rho) = -6.3$ and $\pi = -.67$, indicating that nearly 70 percent of an individual's lifetime wealth is due to transfers from the older generation.

In the Samuelson case, the intergenerational transfers required to reach the Golden Rule consumption point g are Pareto improving. Thus, beginning at the endowment point e_a in some period t, both the old and the young of that period as well as members of future generations are made better off in terms of their narrow self-interest by a shift to the Golden Rule point g. Clearly, the old in period t are better off by the transfer of 3 units from each of the young. The young, too, are better off even though each gives up 3 units of consumption in period $t+1$ if they can be confident of receiving a transfer in old age of 3 units from each of the young next period. Hence, the lifetime utility of the current young rises from U_0 to U^* as a result of the improved pattern of life-cycle consumption that they receive.

The "social contrivance of money" (or a market in any nominal asset) provides a decentralized market mechanism by which Golden Rule transfers from the younger to the older generation can be achieved. In the simplest terms, it "works" because the young wish to save and are willing to trade goods to the old in return for worthless pieces of paper in the expectation that these pieces of paper will have purchasing power in the future. These expectations will tend to be fulfilled if the young of the next period also wish to save and they, in turn, believe in the future purchasing power of money, and if the monetary authority does not opportunistically attempt to transfer income from the old by printing money and destroying the purchasing power of their monetary assets via inflation.

At root, however, the social contrivance of money "works" (if it does) because, like any other voluntary economic arrangement, it makes all parties to the arrangement better off or at least no worse off than they would be if they choose not to participate in it. That is, voluntary transactions, institutions, and so on must be Pareto improving. Money provides a means by which the coalition of the current old, the current young, and members of all future generations can make themselves better off than they would be without money.[6]

This is not true in the case of a classical equilibrium. If we begin at point e_b in Figure 1, notice that transfers from the old to the young are required to reach the Golden Rule point g. Clearly, the old are made worse off by carrying out the transfer and it is in their self-interest to refuse to do so, no matter how much the transfer increases the welfare of the current young or of persons in future generations. Conversely, an improvement in the welfare of future generations from, say, U_0 at point e_b to U^* at point e_a can be achieved only if the current elderly are harmed.

This is the basis for Kim's conjectured theorem. For example, imagine that a bank has somehow come into existence from which the young may borrow at the biological rate of interest $n = r = 100$ percent. The bank can be self-sustaining because the desired borrowing of each period's young (i.e., 2 units) can be financed with the principal plus interest (i.e., 4 units) that each member of the current old population pays to discharge his own debt. Note, however, that the bank must be owned by someone or some group of people in the population. If the owner is old, clearly he is better off consuming the bank's working capital than he would be if he used the working capital to make loans to the young. Hence, the bank will be liquidated. If the owner is young, he is willing to make loans at a sufficiently high rate of interest; however, in the next period, when he is old, he will wish to liquidate the bank.

More broadly, Kim's argument suggests that no set of private financial intermediaries can exist with a positive aggregate net worth that offsets the negative aggregate credit balance of the individuals in a classical equilibrium with $A(\rho) < 0$, because it always pays the owners of such an intermediary to liquidate it themselves or to sell the intermediary to someone else who wishes to liquidate it. In equilibrium, enough banks go out of business that a non-Golden Rule balanced equilibrium emerges with $A(\rho) = 0$ and $\rho > g$.

In my view, this line of argument suggests that the only institutional mechanism (apart from government) by which a society can sustain a classical equilibrium with $A(\rho) > 0$ is the institution of the family in which parents are willing to sacrifice utility from their own consumption in order to increase the level of utility of their offspring.

Following Becker (1974) and other economists, I shall refer to this behavior as parental altruism, where *altruism* refers to the voluntary in-kind or monetary transfers that cannot be explained as part of a past, current, or future quid pro quo exchange. According to this terminology, parents have altruistic preferences if they are willing to bear children, invest in their human capital, and make other transfers to them such as bequests even if they expect no current or future returns from their children.

Of course, as is illustrated in the microeconomic theories of fertility of Becker (1960), myself (1973), and others, it is always possible to describe the altruistic preferences of parents as an exchange in which they receive an (unobservable) flow of "psychic returns" or "child services" from the number and quality of their children in return for (observable) expenditures of time and money.

Whatever terminology is used, however, the key point is that there is an intergenerational conflict of interest when parents seek to capture economic returns from their children, whereas there is an intergenerational coincidence of interest when altruistic parents capture the returns from their children in the form of psychic benefits, because the child gets to capture some of the economic return from the parents' investment. In effect, parental altruism broadens the possible scope of Pareto-improving intergenerational coalitions within families because parents view their children' wealth and welfare as part of their own.

Conclusion: The parental altruism hypothesis and demographic transition

To this point, I have been comparing the steady-state features of economies that differ in their degree of technological advancement and, possibly, in parents' preferences for the number and quality of children. Obviously, however, the story of the economic and demographic transition from relatively primitive "pretransitional" societies with high fertility, high mortality, and low income to modern "post-transitional" high-income societies with low fertility and mortality cannot be told without explicitly treating the very dramatic departures from the steady state that are inherent in the transition process itself. While a full treatment of this problem is well beyond the scope of this paper, I shall conclude my reexamination of the Caldwell hypothesis by advancing an alternative version that reverses the line of causation that he emphasizes and, instead, emphasizes the crucial role that the family and, especially, parental altruism plays in the process of economic development. This hypothesis, which I shall call the parental altruism hypothesis, is outlined as follows.

The process of successful economic development itself is likely to require that the direction of intergenerational transfers be reversed, assuming that Caldwell is correct in his supposition that the direction is initially from the younger to the older generation. Assume an initial pretransitional equilibrium in which children are net economic assets (i.e., $A(p)$ is negative) and fertility is constrained by biological supply factors because there is excess demand for children when their net price is negative. Now suppose that there is an exogenous technological improvement that increases the productivity of labor and/or capital if it is embodied in new investment (e.g., human capital investment in the skills, health, or knowledge of labor, or physical capital investment in new types of equipment). The degree to which this technological improvement is translated into growth in real income is greater the more positive is the response of real savings and investment to the technological improvement. This response, in turn, depends on the degree of parental altruism because a rise in the rate of investment requires that the older generation be willing to reduce its own consumption in order to finance the investment in human capital (the young) and the technologically superior equipment with which the young will work.

A purely self-interested older generation refuses to reduce its consumption either directly or indirectly via intermediary institutions that operate in impersonal private markets because the fruits of the investment are reaped by the younger generation after the older generation is dead. The younger generation itself can only finance increased investment by abandoning its obligations to the older generation, by sacrificing current consumption for future rewards, or by reducing its own investment in the still younger generation, possibly through fertility reduction or through a reduction in its investment per child.

On the other hand, an older generation of altruistic parents regards the success of their sons and daughters as a substitute for their own success. Technological change leads them to reduce their own consumption in order to increase their investment in their own offspring, who will live in a more productive and prosperous environment than the parents themselves do. The rate of population growth caused by technological change is subject to the conflicting forces of a rise in the wealth of the family line, which leads to increased desired fertility, and a rise in the optimal investment in human and physical capital (per child), which reduces the desired level of fertility. Eventually, however, the comparative steady-state results point clearly toward a reduction in the rate of population growth and an increase in wealth per capita as the transition to the steady state is completed.

Appendix
Proof of equation 9

In the text, equation (9) asserts the following relationship:

$$A(\rho) = W(\rho)[a_c - a_y];\ \rho = g \tag{9}$$

Proof:

From equation (6),

$$A(x,\rho) = Y(x) - C(x) + \rho A(x-1,\rho) \quad x = 1,2, \ldots, T$$

where $A(0,\rho) = A(T,\rho) = 0$. Using recurrence, we rewrite these equations as

$$
\begin{aligned}
A(1,\rho) &= Y(1)-C(1)\\
A(2,\rho) &= Y(2)-C(2)+\rho[Y(1)-C(1)]\\
A(3,\rho) &= Y(3)-C(3)+\rho^2[Y(1)-C(1)]+\rho[Y(2)-C(2)]\\
&\vdots\\
A(T,\rho) &= Y(T)-C(T)+\rho^{T-1}[Y(1)-C(1)]+\rho^{T-2}[Y(2)-C(2)]\\
&\quad + \ldots + \rho[Y(T-2)-C(T-2)]+[Y(T-1)-C(T-1)]
\end{aligned}
\tag{A1}
$$

Now we use the definition of $A(\rho)$ in (7) and the expression for the $A(x,\rho)$ terms in $(A1)$ to write

$$
A(\rho) = T[Y(1)-C(1)]+(T-1)\left[\frac{Y(2)-C(2)}{g}\right] + \ldots + \frac{Y(T)-C(T)}{g^{T-1}} \tag{A2}
$$

Similarly, we use the definitions of a_c and a_y in (3) and (4), noting that $W(\rho) = C_L = Y_L$ and assuming that $p(x) = 1$ for $x \leq T$, to write

$$
\begin{aligned}
-W(\rho)(a_c-a_y) &= [Y(1)-C(1)]+\frac{Y(2)-C(2)}{g}\\
&\quad + \ldots + T\left[\frac{Y(T)-C(T)}{g^{T-1}}\right]
\end{aligned}
\tag{A3}
$$

Now we sum (A2) and (A3) to reach

$$
A(\rho) - W(\rho)(a_c - a_y) = (T+1)\left[\sum_{x=1}^{T} \frac{Y(x)-C(x)}{g^{x-1}}\right] \tag{A4}
$$

Since the term in brackets is zero by the social budget constraint (1), the assertion (9) is proved.

Notes

This is a completely revised version of a paper of the same title that was presented at the IUSSP seminar, "Individuals and Families and Income Distribution," in Honolulu, 6–10 April 1981. I wish to thank Yoram Ben-Porath, Jack Caldwell, Oliver Kim, and Ronald Lee for helpful discussions and to discharge them from any responsibility for errors.

Research for this paper was supported in part by a grant from the National Institute of Child Health and Human Development to the National Opinion Research Center and by a grant from the Ford Foundation to the National Bureau of Economic Research.

1 This proposition is true for any steady-state overlapping generations model no matter how many periods of life. See Lee (1980).

2 Space limitations preclude a full analysis of the assumptions about technology, tastes, and life-cycle productivity that are required to guarantee the existence of such a g^*. Deardorff (1976) showed that Samuelson (1975) had actually found a welfare minimum in a particular two-period model with capital. Deardorff (1976) and Samuelson (1976) provide a discussion of this issue for this model. I have derived sufficiency conditions for a

three-period model with capital (Willis, 1981), but I have not yet fully analyzed the general T-period model of this paper. However, preliminary analysis suggests that, roughly, the following assumptions are required for the existence of g^*: (1) capital is not too important in production; (2) labor and capital are not too substitutable in production; (3) consumption at a given age is not too substitutable for consumption at other ages; and (4) the variance of the age of consumption is larger than the variance of the age of net productivity.

3 One of the earliest and best expositions of the implications of this point is Becker (1967). Also see Willis and Rosen (1979) for evidence that family background variables influence schooling decisions in a manner that is consistent with this view.

4 Students of demography will immediately think of the Easterlin relative income argument in this connection. In many ways, the formal structure of this problem is similar to the interesting dynamic principal–agent problem studied by Thaler and Sheffrin (1981) who employed a "two selves" model to deal with the issue of self-control.

5 Kim also provides examples in which a balanced equilibrium does not exist and others in which there are multiple balanced equilibria. I shall ignore these possible complications and stick with the "intuitive" case in the text.

6 This argument may be verified by considering the case of some future generation that chooses not to save, perhaps because its life-cycle endowment is at point e_b. If the preceding generation knows this, they will not wish to save by purchasing money because they know it will not have purchasing power. Similarly, by backward recursion money will not have purchasing power in any generation. This well-known potential "unraveling" of a monetary equilibrium under perfect foresight occurs because the generation whose endowment point is at e_b would not benefit from the system of intergenerational transfers implicit in a monetary equilibrium. In effect, future generations have veto power over social compacts and will exercise this power if the compact makes them worse off.

References

Arthur, W. Brian, and Geoffrey McNicoll, 1978. "Samuelson, population and intergenerational transfers." *International Economic Review*, February: 241–246.

Becker, Gary S. 1960. "An economic analysis of fertility." In *Demographic and Economic Change in Developed Countries Universities*. National Bureau Conference Series No. 11. Princeton University Press.

———. 1967. *Human Capital and the Personal Distribution of Income: An Analytical Approach*. Woytinsky Lecture No. 1. Ann Arbor: Public Administration, University of Michigan.

———. 1974. "A theory of social interactions." *Journal of Political Economy*, November/December.

———. 1981. *A Treatise on the Family*. Cambridge, Mass.: Harvard University Press.

———, and Nigel Tomes. 1976. "Child endowments and the quality and quantity of children." *Journal of Political Economy* 84 (August): S143–S162.

Ben-Porath, Yoram. 1967. "The production of human capital and the life cycle of earnings." *Journal of Political Economy* 75 (May): 352–365.

Bulatao, Rodolfo A. 1979. "On the nature of the transition in the value of children." Papers of the East-West Population Institute, No. 60-A, March.

Cain, Mead. 1981. "Risk and insurance: Perspectives on fertility and agrarian change in India and Bangladesh." *Population and Development Review* 7, no. 3 (September): 435–474.

Caldwell, John C. 1976. "Toward a restatement of demographic transition theory."

Population and Development Review 2, nos. 3 and 4 (September/December): 321–366.

———. 1978. "A theory of fertility: From high plateau to destabilization." *Population and Development Review* 4, no. 4 (December): 553–578.

———. 1980. "Mass education as a determinant of the timing of fertility decline." *Population and Development Review* 6, no. 2 (June): 225–255.

Deardorff, Alan. V. 1976. "The growth rate of population: Comment." *International Economic Review*, June: 510–515.

Gale, David. 1973. "Pure exchange equilibrium of dynamic economic models." *Journal of Economic Theory* 6: 12–36.

Kim, Oliver. 1981. "The overlapping generations model: Equilibrium, optimality and institutional requirements." Ph.D. Dissertation, State University of New York at Stony Brook, July.

Lee, Ronald. 1980. "Age structure, intergenerational transfers, consumption and economic growth." *Revue Economique*, November.

Mueller, Eva. 1976. "The economic value of children in a peasant society." In *Population and Development: The Search for Selective Interventions*, ed. R. G. Ridker. Baltimore: The Johns Hopkins University Press.

Ryder, Norman B. 1976. "Some sociological suggestions concerning the reduction of fertility in developing countries." Papers of the East-West Population Institute, no. 37.

Samuelson, Paul A. 1958. "An exact consumption loan model of interest with or without the social contrivance of money." *Journal of Political Economy*, December: 467–482.

———. 1975. "The optimum growth rate for population." *International Economic Review*, October: 531–538.

———. 1976. "The optimum growth rate for population: Agreement and evaluations." *International Economic Review* 17, no. 2 (June): 516–525.

Thaler, Richard H., and H. M. Shefrin. 1981. "An economic theory of self control." *Journal of Political Economy* 89, no. 2 (April): 392–406.

Wallace, Neil. 1980. "The overlapping generations model of fiat money." In *Models of Monetary Economies*, eds. J. Karaken and N. Wallace. Federal Reserve Bank of Minneapolis, pp. 49–82.

Willis, Robert J. 1973. "A new approach to the economic theory of fertility behavior." *Journal of Political Economy*, Supplement (March/April): S14–S64.

———. 1980. "The old age security hypothesis and population growth." In *Demographic Behavior: Interdisciplinary Perspectives on Decision Making*, ed. T. Burch. Boulder, Colo.: Westview, pp. 43–68.

———. 1981. "On the social and private benefits of population growth." Mimeo.

———, and Sherwin Rosen. 1979. "Education and self-selection." *Journal of Political Economy*, October: S7–S36.

316

DEMOGRAPHY© Volume 15, Number 4 November 1978

A publication of the Population Association of America.

WHAT WILL 1984 BE LIKE? SOCIOECONOMIC IMPLICATIONS OF RECENT TWISTS IN AGE STRUCTURE*

Richard A. Easterlin
Department of Economics and Population Studies Center, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Abstract—Since 1940, under conditions of restricted immigration and high and sustained growth in aggregate demand, shifts in the relative number of younger versus older adults have had a pervasive impact on American life. Before 1960, younger males were in increasingly short supply and their relative economic position substantially improved; after 1960, the opposite was true. Since the early sixties, as the relative condition of young adults has deteriorated, marriage has been increasingly deferred and fertility reduced. The labor force participation of young women has risen at above average rates, and that of older women has risen at below average rates. Changes in the age structure of the working age population have also contributed to a combination of rising unemployment and accelerating inflation. Cohort divorce rates, suicide among young males, crime rates, and political alienation have worsened. The rise in college enrollment rates has been interrupted, and SAT scores have declined. In contrast, in the period 1940–1960, changes in these various magnitudes were typically of a more favorable sort. The United States is now at the start of a new period of growing scarcity of young adults as a result of the birth rate decline that set in after 1960. This implies that the 1980s will see a turnaround or amelioration in a wide variety of these social, political, and economic conditions, some of which have been taken as symptomatic of a hardening social malaise.

INTRODUCTION

The period around 1960 marked a turning point in many aspects of American experience. I need not rehearse for this audience the precipitous decline in fertility rates that has since occurred. Also well-known is the dramatic shift in the age pattern of increases in female labor force participation rates—the acceleration for younger women and slowdown for older women. But beyond these developments, there have been other major economic, social, and political changes. On the economic side, the unexpected combination of rising unemployment and accelerating inflation has proved a source of embarrassment to economists (and, thereby, a delight to many noneconomist colleagues). On the social scene, there has been an acceleration in divorce, a rise in suicide rates among the young, and an upturn in crime rates. In the political arena, there has been a growth in alienation from the established system. Some of these developments have been taken as indicating a growing social malaise. Long-established cultural attitudes, too, appear to have changed. There has been growing antipathy toward childbearing and, more generally, toward population growth, and a questioning, as never before, of traditional women's roles. Academic life, a central concern to many here, has been shaken by many of these developments

* Presented as the Presidential Address to the Population Association of America at its annual meeting in Atlanta, Georgia, April 1978.

and others closer to home, such as the unexpected break in the rise of college enrollment rates and the steady decline in the SAT scores of students entering college.

I think it is possible that these seemingly disparate developments—and, in fact, many others—are, in part (and I stress, in part), due to a common cause, to a new type of relationship between population and the economy centering on changes in the age structure of the working age population. These age structure effects operate largely via age-specific rates. Because of this, they have been overlooked in demographic research that focuses almost exclusively on compositional effects of age structure changes. And, if I am correct about the importance of shifts in age structure, then the near-term future will be different, perhaps strikingly so, from that foreseen by those who are extrapolating recent changes into the future.

My interest is not in year-to-year fluctuations or in very long-term trends, but in so-called long swings or Kuznets cycles, heretofore of 15 to 25 years' duration, that have marked the historical record of U.S. economic and population change since at least early in the nineteenth century (Kuznets, 1958, 1961; Abramovitz, 1961, 1968; Easterlin, 1968). Hence, the "1984" in my title—I am interested not in where we will be next year or in the year 2000, but a decade or so hence. Thus, 1984 stands as a convenient symbol for the 1980s.

The analysis rests on the view that any projection of the future must be based on a theory consistent with the past. In consequence, I shall spend the bulk of my time on the nature and causes of past experience.

I shall range over several disciplines. In so doing, I follow in the tradition of a great social scientist who introduced me to demography, and whom I miss very much, Dorothy Swaine Thomas. Throughout her career—in her work on social aspects of the business cycle, on Swedish population and industrialization, and on U.S. population redistribution and economic growth—Dorothy Thomas went beyond the usual confines of demography. It seems most appropriate, therefore, that this talk be dedicated to her. I should also like to acknowledge two other mentors who have perhaps unknowingly left their mark on this talk, Simon Kuznets and the chairman of this session, John Durand. Finally, I must admit what you will see shortly when we come to the heart of the analysis, that I have drawn freely on recent work of other scholars.

THE BREAK WITH PAST EXPERIENCE

Let me start with the assertion that the 1940s marked the dawn of a new age in the relation between population and the economy. Chart 1 summarizes the essential features of the pre- and post-World War II contrast. The main point of the chart is that a reversal has occurred in the long swing roles of aggregate demand and labor supply. Before World War II, swings in labor supply and, thereby, population arose chiefly from immigration and occurred usually only in response to corresponding major swings in aggregate demand. Now, sizable swings in labor supply, involving concurrent changes in the proportion of young to old in the working age population, occur as an echo of prior movements in the birth rate. As a result of government management, aggregate demand fluctuates only mildly compared with its fluctuations in the past. Thus, changes in the supply of labor now occur largely independently of aggregate demand. Before World War II, aggregate demand was the active factor in long swings, and labor supply was a passive factor; at present, the opposite is largely the case. Associated with this change in demand/supply relationships has been an approximate doubling of the duration of long swings.

What is the cause of this new era in the relation between longer term economic and demographic swings? Three causes can be given.

Chart 1.—Contrasting Patterns of Economic and Demographic Long Swings Before and After World War II

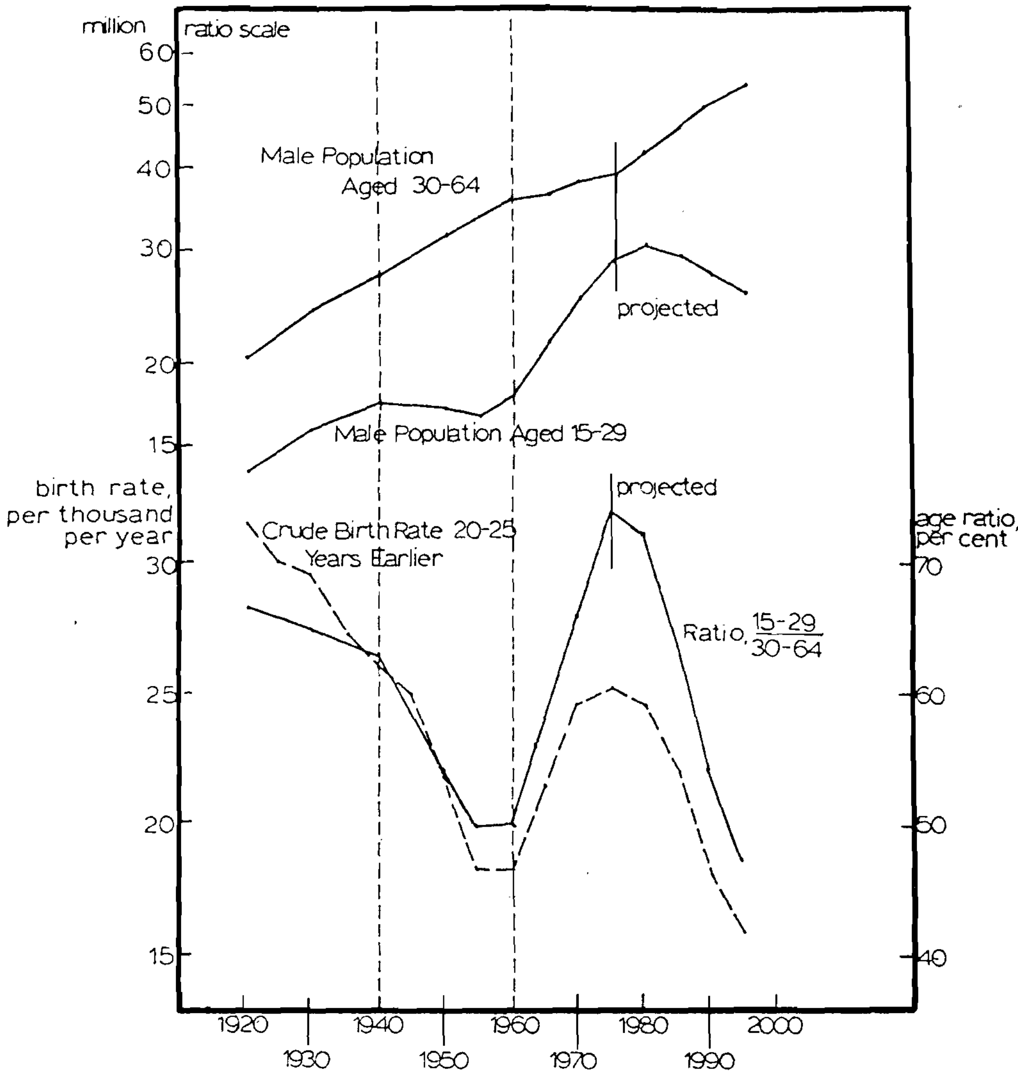
	Before World War II	Since World War II
Aggregate demand	Active role: Private investment booms initiate a major swing in aggregate demand independently of labor supply conditions.	Passive role: Relatively high and sustained growth in aggregate demand maintained by monetary-fiscal policy.
Labor supply	Passive role: Swings in labor supply occur because of immigration movements induced by aggregate demand swing.	Active role: Swings in labor supply and in proportion of young to old in the working age population occur independently of aggregate demand as a lagged effect of birth rate.
Duration of swings	15–25 years	35–40 years?

First, severely restrictive immigration legislation of the 1920s meant a sharp curtailment in the labor reserve that had traditionally supplied the demands of major economic booms. Thus, in the 1940s and 1950s when labor force growth from native sources was at an all-time low, a massive influx of immigrants to satisfy the labor demand of a major economic boom did not occur as in the past. In effect, the role that immigration had served to buffer the impact of economic booms on the native population was eliminated by the restrictive legislation.

Second, the Employment Act of 1946 committed the federal government to maintaining a high and growing level of aggregate demand through monetary and fiscal policies. Also, the substantial rise in the government's share of the GNP after World War II compared to its share before that time helped to stabilize aggregate demand. Between 1942 and 1974, the annual unemployment rate rose above 6 percent in only two different years, and even then by less than one percentage point. Compared with any prior 30-year period, this is an unprecedented stretch of relatively uninterrupted growth in labor demand.

Finally, and most important for the present purpose, the declining birth rate of the 1920s and 1930s and the rising birth rate of the 1940s and 1950s caused, with a 20- to 25-year lag, first a growing scarcity of younger workers and then a growing abundance. This is shown in Figure 1. (The underlying data for Figure 1 and all subsequent figures appear in Appendix A.) Note that between 1940 and 1960 there is a noticeable interruption in the pre-1940 growth of the population aged 15 to 29; after 1960, growth is resumed at a more rapid rate than in the pre-1940 period. This pronounced fluctuation in the growth of the younger age group is not matched in the curve for the older group. As a result, the proportion of younger to older working age population, the solid line at the bottom of the figure, shows a sharp fluctuation in the period after 1940. Through the late 1950s, younger persons are growing in relative scarcity; subsequently, there is a growing abundance of younger persons. The broken line curve in the bottom half of Figure 1 shows that this swing in the proportion of young to old is largely due to a corresponding movement in the birth rate 20 to 25 years earlier.

These shifts in age structure—in the proportion of young to old in the working age population—under conditions of high and sustained growth in aggregate demand and restricted immigration have had ramifications that we are only now



Source: Appendix Table 1.

Figure 1.—Male Population Aged 15 to 29 and 30 to 64, Actual and Projected, 1920–1995, and Crude Birth Rate 20 to 25 Years Earlier

starting to appreciate. Let me turn to the nature of these effects and the causal mechanisms at work.

EFFECT OF AGE STRUCTURE ON THE RELATIVE ECONOMIC POSITION OF YOUNG MALES

Consider first the effect of these age structure shifts on the relative economic position of young males. Not all age structure effects work through relative income, but many of them do. To make the argument as clear as possible, I shall make some very simple assumptions.

The essence of the reasoning is outlined in Chart 2. To start with, disregard females and suppose that the labor supply consists of only two types of labor—

younger and older males—as in the upper panel. Younger males are a relatively inexperienced and low-skilled group that is fairly new in the labor market, with rather tentative job commitments. This group, a "career-entry" group, is engaged in a considerable amount of job search with consequent high job turnover. Older males are an experienced, skilled group that occupies higher level career jobs and has relatively low job turnover. The degree of substitutability between these two groups is low, but not zero. Both groups are fully in the labor force; that is, their labor force participation rates are close to 100 percent.

Now, assume that the growth in the economy's aggregate demand for labor comprises some normal division between younger and older workers. (One need not assume that the demand for each group grows at equal rates, only that for each group the rate of growth in demand remains constant over time.) Suppose, now, that every two decades or so a substantial shift occurs in the relative supply of younger versus older workers, reflecting corresponding changes in the working age population. In one period, younger workers are relatively scarce, in the next they are relatively abundant.

Given the steady growth in demand for both groups, these changes in the supply of labor will create imbalances in the labor market for younger and older workers. What will be the consequences of these imbalances? As shown in the upper panel of Chart 2, a scarcity of younger

Chart 2.—The Easterlin-Wachter Model: Ceteris Paribus Effects of Shifts in Relative Scarcity of Young Adults

Assumption A: Working Age Population Comprises Only Younger and Older Males

Independent variables
- Labor demand: growing at trend rates for both groups
- Labor supply: *large shifts in younger versus older males due to corresponding* shifts in the working age population

Adjustments in dependent variables	*If Scarcity of Young*		*If Abundance of Young*	
	younger males	older males	younger males	older males
Wages	+	−	−	+
Unemployment rates	−	+	+	−
Occupational mobility	+	−	−	+

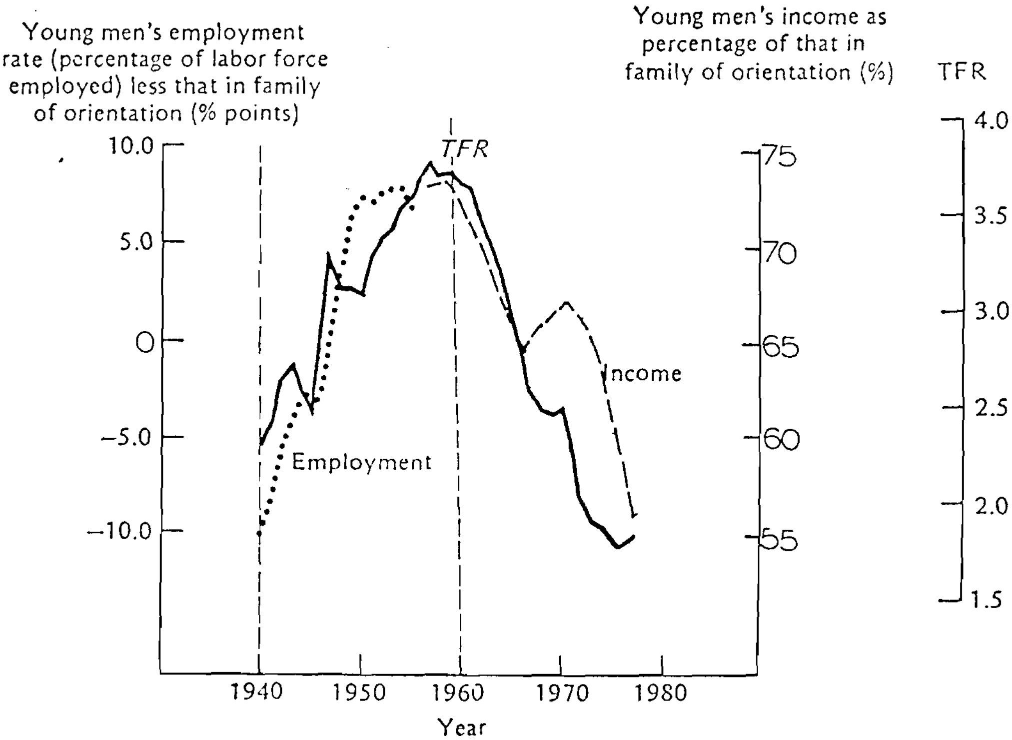
Assumption B: Working Age Population Comprises Younger and Older Males and Females

Independent variables
- Labor demand: Growing at trend rates for all four groups
- Labor supply: large shifts in younger versus older persons due to corresponding shifts in the working age population

Adjustments in dependent variables	*If Scarcity of Young*		*If Abundance of Young*	
	younger males	older males	younger males	older males
	(same as for Assumption A)			
	younger females	older females	younger females	older females
Wages	(+)	(−)	(−)	(+)
Unemployment rates	(−)	(+)	(+)	(−)
Labor force participation rates	−	+	+	−
	younger males and females		younger males and females	
Marriage	+		−	
Fertility	+		−	

workers will affect favorably their relative wages, unemployment rates, and upward mobility. The opposite would be true when there is an abundance of younger workers. There would also tend to be some substitution between older and younger workers, but by assumption such possibilities are limited. There might also be feedback effects from the labor supply side to aggregate demand—affecting both the size and composition of aggregate demand. (For some discussion of this, see Wachter and Wachter, forthcoming.) Thus, in periods when young males are scarce, their relative economic position would improve; when young males are abundant, their relative economic position would deteriorate. (Appendix B evaluates a possible qualification to this argument.)

A variety of evidence indicates that this has, in fact, been the case. The dotted and broken lines of Figure 2 (updated from an earlier paper of mine) give some illustrative measures—for relative employment through the mid-fifties, and relative income thereafter. These show that the relative economic position of young males since 1940 has, by and large, varied directly with their relative scarcity as indicated in Figure 1. [For additional evidence relating to the period of improvement, see Easterlin (1968), pp. 114–118, and Easterlin (1973). For more evidence relating to the period of deterioration, see Wachter (1976a, 1977).]



Source: Appendix Table 2.

Figure 2.—Total Fertility Rate, 1940–1977, Relative Employment Experience of Young Adult Males, 1940–1955, and Relative Income Experience of Young Adult Males, 1957 On

EFFECT OF AGE STRUCTURE ON FERTILITY AND FEMALE LABOR FORCE PARTICIPATION

Next, let us add to the analysis the effects of age structure on fertility and female labor force participation. Here, I shall use what I have elsewhere called the Easterlin-Wachter model (Easterlin, 1978). The working age population is now taken to consist of four groups—younger and older males and younger and older females. Three additional assumptions are made.

First, there is a traditional division between male and female roles—that is, males are the primary "breadwinners" and have labor force participation rates close to 100 percent, whereas females' attachment to the labor market is less permanent (participation rates are substantially below 100) and their primary responsibility is childbearing, child rearing, and care of the home. This assumption implies no judgment about the desirability of these sex roles; I view it simply as a reasonable assumption about reality.

Second, there are three classes of jobs: (a) "career jobs," involving considerable experience and skill, that are typically filled by older males because of their continuing labor force attachment and accumulated experience; (b) "career-entry jobs," that are typically held by younger males; and (c) "noncareer jobs," typically held by women. As a result, there is only limited substitution possible between younger and older males and between women and men, but a high degree of substitution possible between older and younger women.

Third, marriage and childbearing vary directly with the income of younger relative to older men. [In economics, the relative income hypothesis was first applied to savings behavior in Duesenberry (1949).] The reasoning is that the relative income of younger men may be taken as a rough index of the primary breadwinner's ability to support a young household's material aspirations. These aspirations are formed by the material environment that the spouses experienced as they grew up, which depends, in turn, largely on their parents' incomes. Hence, when young males' income is high relative to older males', it means that they may more easily support the aspirations that they and their potential spouses formed in their families of origin. Young people will then feel freer to marry and have children.

Now, let us consider the effect of imbalances in the labor market arising from sizable shifts in the age distribution of the working age population (Chart 2, lower panel). As before, it is assumed that the growth of labor demand for all labor force groups (this time four in number) is at constant, though not identical, rates. One might also suppose there was a normal trend increment in labor force participation rates of younger and older females. Suppose, now, that there is a scarcity of younger workers as a result of a corresponding change in the working age population. With regard to the labor market for younger and older males, the adjustments would be the same as those in the upper panel of Chart 2; that is, younger males would experience relatively favorable changes in wages, unemployment rates, and upward mobility. Now, however, one must consider additionally the implications of the improved relative income of young males for the situation of females. First, it would be easier for young men and women to marry and start families. This would, in turn, dampen and perhaps eliminate the normal increment in labor force participation of young females as the marital status and child dependency distribution of this group shifted toward categories with below average participation rates. The scarcity of young women would tend to have favorable effects on their relative wages and unemployment rates, but, because there is possible a high degree of substitution of older for younger women, these effects would be relatively moderate—this is the reason for the parentheses in the lines for these variables in the bottom panel of Chart 2. The more important effect would be the replacement of younger by older females. This means that the less than normal increment in participation rates for younger females would be compensated by a more than normal increase in the rates for older females. [Wachter's (1972) empirical analysis of the effect of relative versus permanent income on labor force participation rates of younger and older females supports this inference. Other studies noting a possible link between the differential changes in participation rates of younger and older women and the relative scarcity of young persons in the post-World War II period are Bancroft (1958), Oppenheimer (1970), and United Nations (1962).] The entry of older women into the labor market might be further stimulated by the situation of older males, whose relative income situation, as shown in Chart 2, is unfavorable. If we compare in simple price-quantity terms the labor market adjustments of males and females, for males the scarcity of young persons induces an adjustment primarily in terms of relative income, for females, primarily in terms of relative quantities, that is, participation rates.

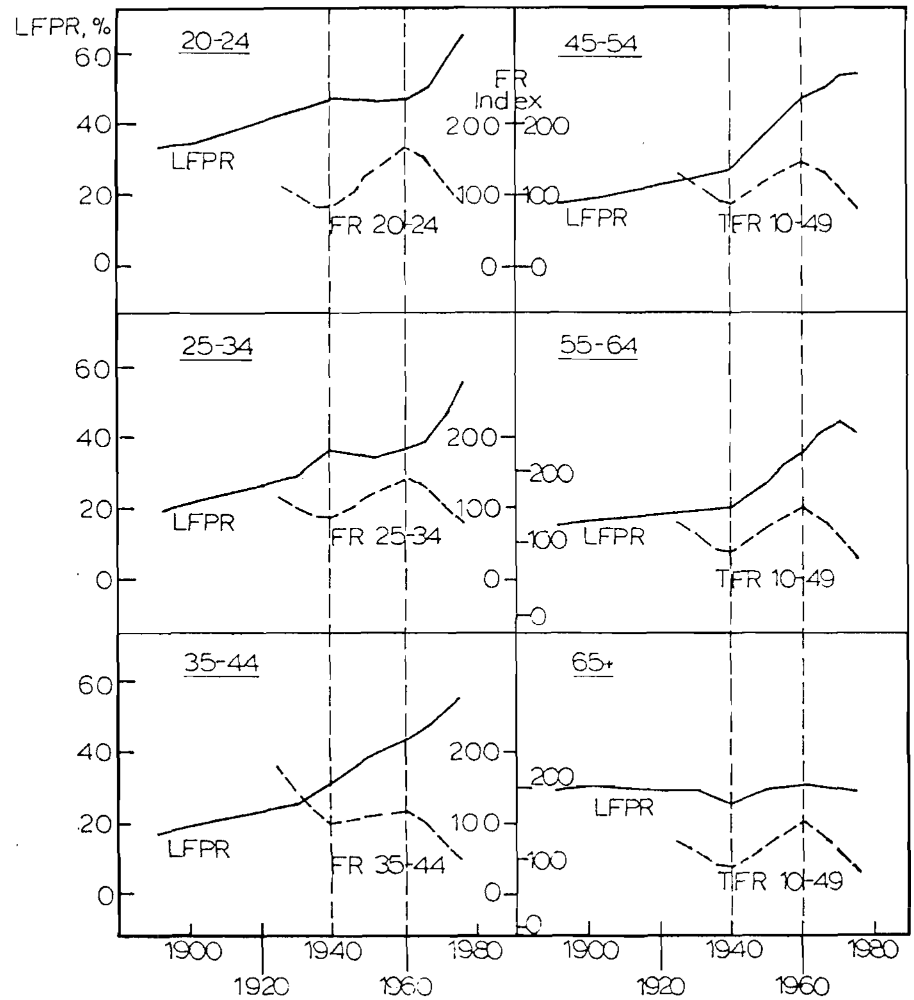
This analysis also shows that a cohort carries its fortunes, good or bad, depending on its size, throughout its life cycle. The cohort implications of Chart 2 are seen by a comparison of the first and last columns. Note how the signs remain the same—as a scarce cohort ages, it carries with it relatively favorable wage and employment conditions. (The argument is most clear-cut if the younger are taken to be those aged 20 to 39, the older those aged 40 to 59, and a turnaround in age structure occurs every 20 years.)

Admittedly, this is a highly simplified view of the makeup of the labor supply and the causal factors at work. One might, for example, distinguish within the group of younger males a "noncareer" group of disadvantaged workers, and within females a "career" group of college graduates. I believe, however, that this view, simple as it is, captures an important part of the forces shaping post-World War II experience. Consider the patterns of fertility and female labor force participation shown in Figure 3. (In the figure, the fertility rate in a given five-year period is plotted against the labor force participation rate at the end of the period; for example, the fertility rate for 1971–1975 is plotted against the participation rate for 1975.) Starting around 1940, there was a marked break with previous experience. (The fertility trends before the 1920s, for which data are not available, would, of course, be downward.) Between 1940 and 1960, the upward trend in participation rates of young women was interrupted, and their fertility rose markedly. Participation rates of older women, which had previously shown only mild increases, shot up dramatically. Since 1960, there has been a reversal in this age pattern of participation rates—rates for younger women have risen sharply, while those for older women have risen only slowly or leveled off. At the same time, a sharp downturn in fertility has occurred.

Referring back to the lower panel of Chart 2, one finds that the analytical model just sketched generates behavior of just this type—an inverse association between the growth in labor force participation rates of younger versus older women, an inverse association between the growth in labor force participation of younger women and their fertility, and a positive association between the growth in labor force participation of older women and the fertility of younger women. But fertility is not causing participation rate change or vice versa. Rather, all of these developments flow from the imbalance in the labor markets of younger versus older workers due to shifts in the relative supplies of the two groups.

I am not claiming that the model explains completely the magnitudes of change observed; it clearly neglects, for example, the determinants of the general upward trend in labor force participation rates of females as a whole. It does capture, however, the essential outlines of the long-term swings since 1940, and, for

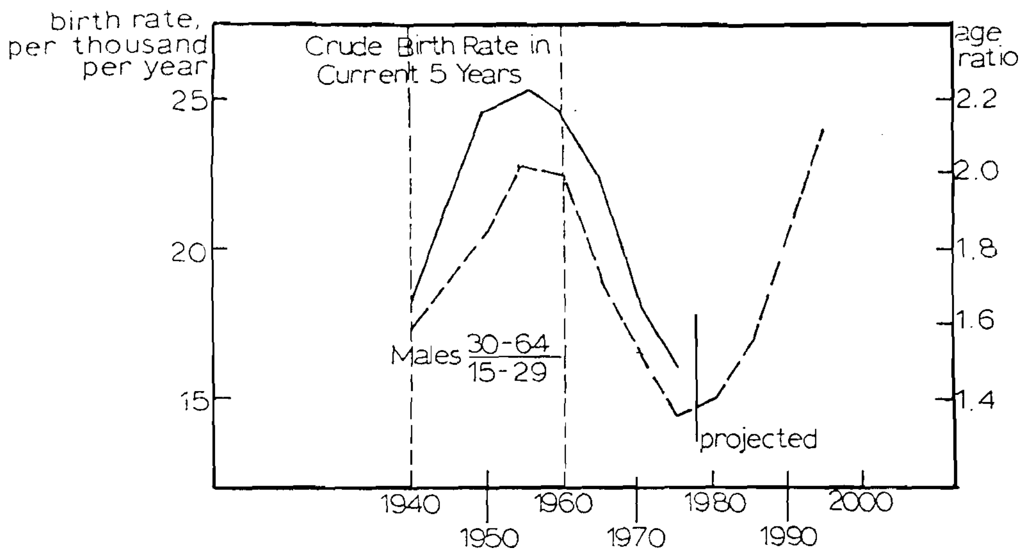
324



a
Comparable data for 1910 not available.
Source: Appendix Tables 3 and 4.

Figure 3.—Labor Force Participation Rates of Females Aged 20+, by Age Group, Decennially, 1890–1950ᵃ, and Quinquennially, 1950–1975, and Index (1940–1945 = 100) of Fertility Rates for Specified Age Groups, Quinquennial Averages, 1921–1925 to 1971–1975

some variables, it may even do well on magnitudes. This is suggested by Figure 2, in which the total fertility rate shows a clear parallel with relative income and its proxy based on the unemployment rate. It is further illustrated in Figure 4, in which the birth rate since the late 1930s is plotted against age structure—in this case, the ratio of older to younger males, the inverse of the solid curve at the bottom of Figure 1. The two curves are not identical, but the similarity is striking—the swing in the birth rate looks very much like that in the relative number of older to younger males. (Figure 4 looks much the same if the total fertility rate is used instead of the



Source: Appendix Table 5.

Figure 4.—Ratio of Males Aged 30 to 64 to Males Aged 15 to 29, Actual and Projected, 1940 on, and Crude Birth Rate, 1935–1940 to 1970–1975

crude birth rate, and Figure 2 looks much the same if the crude birth rate is used instead of the total fertility rate.)

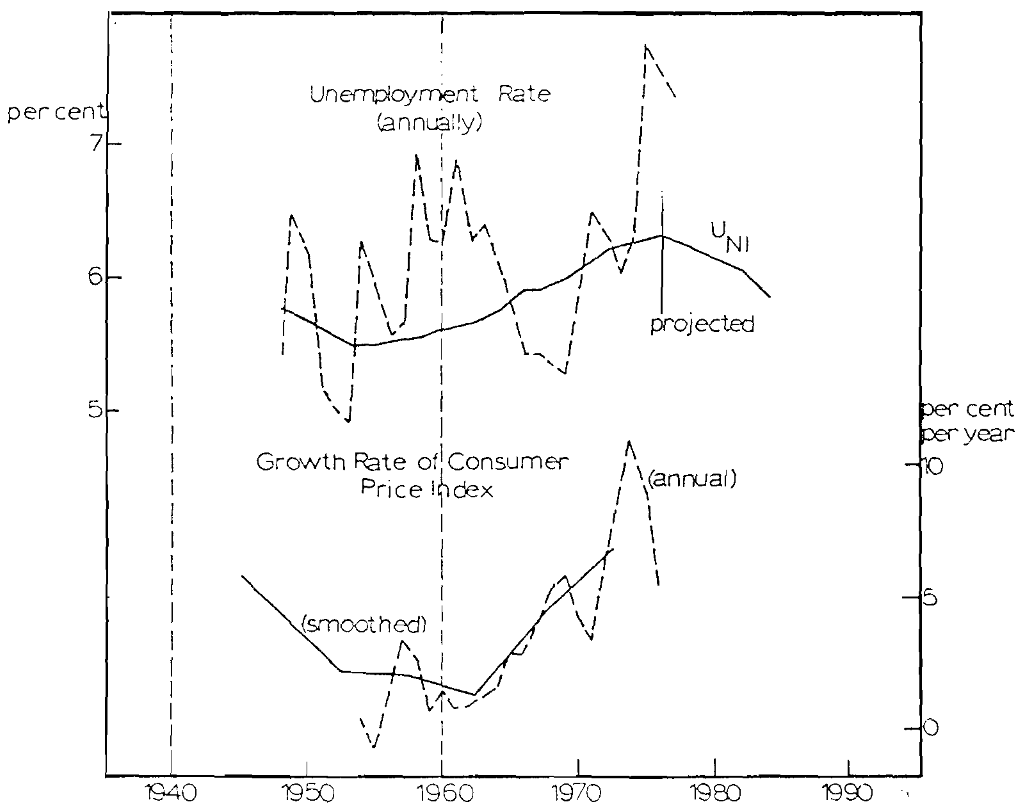
EFFECT OF AGE STRUCTURE ON ECONOMIC STABILITY

We can build on the foregoing analysis to show how shifts in age structure since 1940 have contributed to the macroeconomic problem of rising unemployment combined with accelerating inflation. Here, I shall again borrow in part from research done by my coworker, Michael Wachter (Wachter, 1976a, 1976b; Wachter and Wachter, forthcoming; Easterlin et al., 1978).

First, note that the developments sketched in Chart 2 would tend, in a period of growing abundance of younger workers, to raise the economy-wide average unemployment rate. The economy-wide rate can be viewed as an average of two component rates—a low rate for older workers and a high rate for younger workers, the latter reflecting their newness in the labor market, their job-seeking activity, the tentativeness of their job commitments, and so on. Without any change in the unemployment rate for each older or younger group, a rise in the proportion of younger workers in the labor force (that is, in the proportion of those subject to high unemployment rates) would, other things being constant, raise the economy-wide unemployment rate. This is the standard age composition effect of demographic analysis. But, as shown in the right side of Chart 2, an abundance of younger workers would also affect age-specific unemployment rates, raising those of younger workers relative to those of older workers, and this would further increase the economy-wide unemployment rate. The rates for older workers are typically quite low, and there is little room for a change of much magnitude in the rates for these groups. But there could be a substantial change in the rates for younger persons, males and females. In fact, since the 1950s there has been a marked deterioration in the relative unemployment rates of younger men and women.

Consider, now, the implications of this for economic stabilization policy. Monetary and fiscal policies to correct unemployment conditions are geared to affecting the overall level of aggregate demand, but not the structure of that demand. Specifically, monetary and fiscal policies are *not* shaped to influence differentially the relative demands for two groups, such as younger and older workers, so as to compensate for changes in their relative supply. But, as we have just seen, under given conditions of aggregate demand, an increase in the relative supply of younger workers would tend to raise the economy-wide unemployment rate. If monetary-fiscal policy is initiated to correct a rising unemployment rate caused by an altered mix of the labor supply, the extra demand does not correspond to this altered mix. Given imperfect substitution between older and younger workers, the extra demand generates relatively little improvement in employment conditions and spills over into price increases—hence, the phenomenon of a rising unemployment rate coupled with accelerating inflation.

The relevance of this argument to recent American experience is illustrated by the series plotted in Figure 5. The solid line curve labeled U_{NI} charts the course of the "full employment" rate that would be consistent with the changing effect of the population's age mix on labor supply conditions in our economy, as estimated by Wachter. Note how U_{NI} swings upward,



Source: Appendix Table 6.

Figure 5.—Unemployment Rate U_{NI} and Growth Rate of Consumer Price Index Since World War II

after the late 1950s, due to the rise in the proportion of young adults in the working age population shown in Figure 1. As it swings upward, however, the actual unemployment rate in the 1960s, shown by the broken line, which reflects the impact of aggregate demand and, thus, government monetary-fiscal policy as well as supply, moves below U_{NI}. Reference to the two lower curves shows that it is just about at this time that the average annual rate of price increase rises from a level not much above zero to the increasingly high levels of recent years. This rise in the inflation rate is, in part, the consequence of attempting to correct an imbalance on the labor supply side with inappropriate aggregate demand policies.

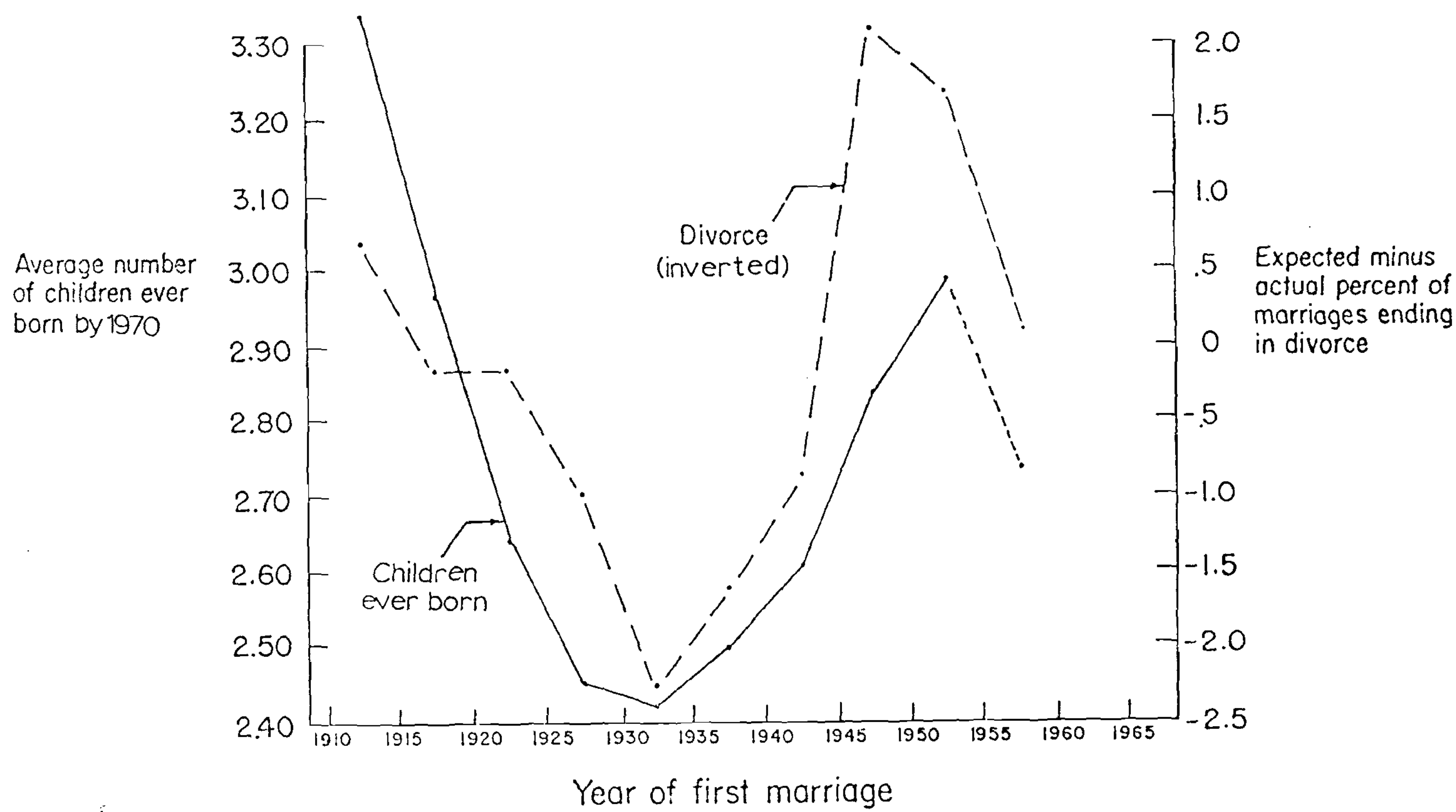
Accelerating inflation as U_{NI} rises is not an inevitable result of aggregate demand management. Over the past decade, the monetary and fiscal authorities seem to have underestimated the effect of the demographic swing on the sustainable, noninflationary unemployment rate. Maintaining old unemployment targets for too long resulted in overexpansive demand policies and an increase in the inflation rate. Moreover, a higher inflation rate persisted for some time after aggregate demand policy had been readjusted downward, because of the lagged response of expectations, and the inertia built into the economy due to the presence of fixed price and wage contracts. If the sustainable unemployment rate through aggregate demand management could be estimated with accuracy, and if the government adopted that rate as the target rate for demand management, then the correlation between rising unemployment and accelerating inflation would be broken, insofar as that correlation is due to the altered labor supply mix.

EFFECT OF AGE STRUCTURE ON SOCIAL VARIABLES

Let me turn briefly to some social and political effects of the recent shifts in age structure. The theoretical underpinning for such effects is straightforward. Relative income or relative status notions have never been the private preserve of economists. Quite the contrary—economists came to the scene (relatively) late and even today are, for the most part, strongly resistant to such notions. In contrast, there is a strong tradition in sociology running from Durkheim's theory of *anomie* through Stouffer's work on relative deprivation down to the present, which sees relative status considerations as pervasive in social phenomena (Durkheim, 1951; Stouffer et al., 1949). Let me cite two examples from recent work by demographers who have strayed somewhat beyond the usual confines in their field.

Figure 6 is taken from a paper by Samuel Preston and John McDonald (1977) that presents historical estimates of divorce by marriage cohort and then—our interest here—speculates on the causal factors at work. The point of Figure 6 is that a trend-adjusted series for cohort divorce rates shows a long swing much like that in fertility, although the relationship is inverse. Thus, the trough in children ever born for the cohort first married in 1930–1934 is accompanied by a peak in the trend-deviation for divorce. The next turning points in the two series are close, but not identical: corresponding to the fertility peak for the marriage cohort of 1950–1954, there is a trough in the trend deviation in divorce for the cohort of 1945–1949. In general, for the marriage cohorts that produced the baby boom, divorce was considerably below the rate expected on the basis of past trends. For the most recent marriage cohorts, as their fertility has turned downward, divorce has risen toward trend rates.

A close association between fertility and another social magnitude, suicide among young males, has been pointed out and studied by Martin O'Connell (1975). For the period 1940–1976 (excluding the war years 1941–1945), the correlation between the total fertility rate of females aged 15 to 49 and the suicide rate of males aged 15 to 24 is −.88, between the fertility rate and the suicide rate of males aged 25



Source: Appendix Table 7.

Figure 6.—Children Ever Born and Deviations From Trend in Divorce for White Marriage Cohorts of 1910-1914 Through 1955-1959

to 34, also −.88. Note that again the relationship is inverse—higher fertility goes with lower suicide.

Such correlations of divorce and suicide with fertility do not necessarily imply causality. I have already argued that the inverse correlation between the growth in labor force participation of younger females and their fertility is due, not to a cause-effect relation between the two, but to a common response to the same cause, relative income. Both the Preston-McDonald and O'Connell studies suggest that the same is true for the correlations they observe. [Simon (1968, 1969, 1975) developed analogous relative-income-type models to explain fertility and suicide.] Both the Preston-McDonald and O'Connell studies, drawing on the literature, advance a relative income hypothesis (though in somewhat different form) and test for its explanatory power. The results are positive—a deterioration in relative income of the young makes for more frequent divorce and a rise in suicide among young males; and improvement has the opposite effect.

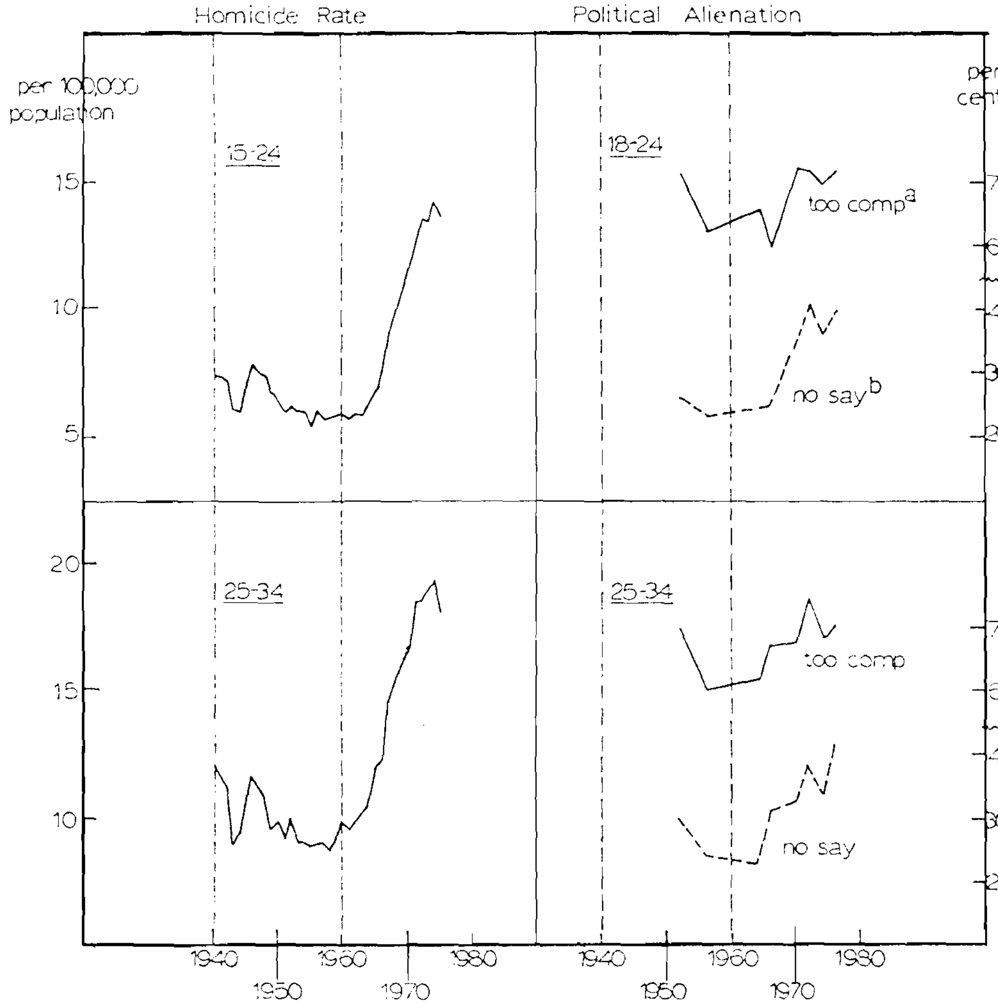
If a relative income mechanism is influencing divorce and suicide, should not its effects be apparent in other social phenomena? The answer, of course, is yes; and there are some hints in the record that they are. Figure 7 presents homicide rates, which may perhaps be taken as representative more generally of crime rates, and two measures of political alienation among the young. Although the series are short and not adjusted for trends, there is a common pattern—a suggestion of slight improvement to around 1960 (a downward movement of the curves) and a noticeable deterioration thereafter (an upward movement). (Evidence that the recent rise in crime rates reflects changes in age-specific rates as well as in the age composition of the population is given in Wellford, 1973.) I am not claiming that age structure and "relative income" are the whole story of the post-1940 change in crime rates and political alienation. In fact, House and Mason (1975) fail to find evidence of significant differences among demographic age groups in the data for political alienation. But these are examples of variables that in sociological theory should reflect relative status influences, and there is a hint in the evidence that this is the case.

EFFECT OF AGE STRUCTURE ON HIGHER EDUCATION

I want to conclude this discussion of the effects of age structure by briefly noting two studies that identify important impacts on the college environment—one on college enrollment rates and the other on SAT scores. In each case, in contrast to the foregoing, age structure has its effect with a substantial lag. As before, my interest is not in effects working through age composition as such (although for college enrollments this effect is an important one), but in influences working on age-specific rates.

Let me start with SAT scores, which, as shown in Figure 8, have been declining since the early 1960s. Robert Zajonc, a social psychologist at the University of Michigan, from whose work this chart is taken, has offered an explanation of this decline based on average birth order (1976). Various studies have found a low but significant association between birth order and intelligence—higher birth order being correlated with slightly lower intelligence. The post-World War II baby boom resulted in an increase in average birth order. This is shown by the lower curve of Figure 8, which is plotted in inverted form. Those born in a given year take their SATs about 17 years later. Hence, according to Zajonc, the decline in SATs that started in the 1960s can be traced back to baby boom cohorts with progressively rising average birth order. As Figure 8 shows, when the curves for SATs and average birth order are plotted with appropriate allowance for the lead-lag relationship, they do exhibit a substantial similarity.

Since about 1970, there has been a puzzling interruption in the uptrend in enroll-



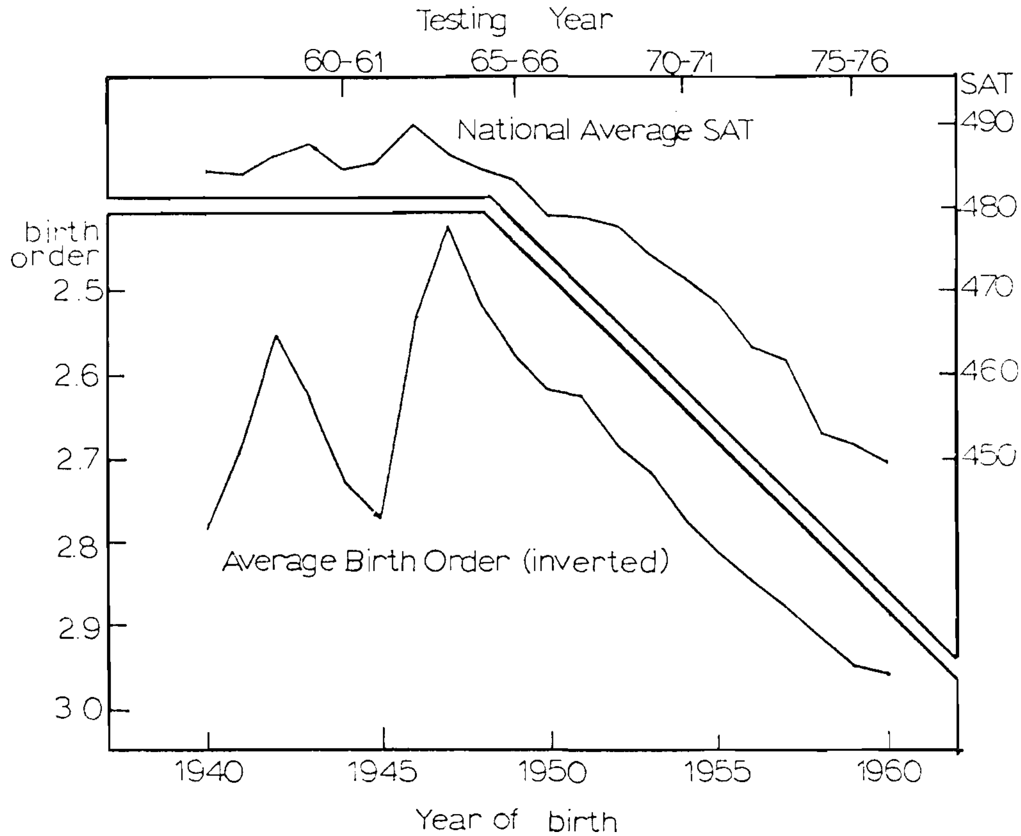
[a]too comp=% agreeing with statement "Sometimes politics and government seem so complicated that a person like me can't really understand what's going on."

[b]no say=% agreeing with statement "People like me don't have any say about what the government does."

Source: Appendix Tables 8 and 9.

Figure 7.—Homicide and Political Alienation Among the Young

ment rates for those of college age—a leveling off among females and a decline among males. Goldberg and Anderson (1974) have advanced an explanation for this based on changes in the proportion of households that have at the same time two or more children of college age. The financial burden of putting a child through college is obvious. But, Goldberg says, consider the financial predicament of those with two children of college age at the same time—not to mention three! Through ingenious work with the Public Use Tapes, Goldberg has constructed several measures of the changing magnitude of this predicament in the post-World



Source: Appendix Table 10.

Figure 8.—Average SAT Score and Average Birth Order for Annual Birth Cohorts of 1940 Through 1960

War II period. The measures show an upsurge in this problem about the time of the interruption of the rise in college enrollment rates. The movement in the proportion of households with two or more children of college age is itself, of course, an echo of prior fertility movements. Thus, the implication is that the postwar baby boom is currently taking its toll on college enrollment rates because of the rise in the proportion of households experiencing the financial strains of having two or more children of college age at the same time.

Another interpretation of the recent decline in enrollment rates is suggested by Richard Freeman's work (1976, especially pp. 64 ff.). The decline may be due to a drop in the returns from higher education, caused in part by the recent increase in the supply of those of college age, an effect of the post-World War II baby boom. The Freeman and Goldberg interpretations may be viewed as complementary, because both involve in different ways lagged effects of the baby boom. The Goldberg argument differs from those previously presented in this paper in that the theoretical argument implies errors in judgment.

I have spoken of the movements in enrollment rates and SATs as reflecting age structure changes, but for each of these variables the causal chain involves several links. In both cases, the immediate determining factor is compositional—the population mix with regard to birth order or proportion with two or more college age children. As has been noted, however, each of the latter depends, in turn, on prior movements in fertility. And, by the reasoning previously stated, the post-World War II swing in fertility has been largely governed by age structure and its relative income effects. By this set of connections, therefore, age structure shifts have, with a lag, affected college enrollment rates and SAT scores.

RESEARCH IMPLICATIONS

Space is running short, but I would like at least to touch in outline form on some of the research possibilities suggested by the present analysis.

1. First, much more work is needed on age structure changes themselves and the underlying mechanisms through which they work.
   a. I have focused here on the proportion of younger to older working age population, with the age of 30 as the dividing line. Experimentation is needed with alternative age ratios. For example, the ratio of those aged 20–29 to 30–39 shows a time series movement since 1940 much like that for the ratio shown in Figure 1. The appropriate age ratio might well differ, of course, from one subject to another, and from one society to another.
   b. Perhaps too much stress is laid here on the importance of relative numbers. There is need to look into the relative "quality" of different age groups, as reflected, say, in measures of educational attainment and work experience (although these may, in part, be dependent on relative numbers).
   c. The appropriate measure of relative income is itself a subject for research. One would not expect the same measure to be applicable to different age groups and all subjects of study.
   d. I have perhaps focused too much on economic indicators of relative deprivation, to the exclusion of psychological ones. As Keyfitz (1972) and others have suggested, the experience of relative deprivation may have its origins earlier in life, for example, in large families and crowded classrooms.
2. One would expect that the mechanisms described here, whatever their specific nature, would have left their imprint on a variety of other phenomena. One clue, suggested by the Preston-McDonald and O'Connell studies, might be the degree of correlation of other economic, social, and political indicators with fertility. In some cases, the effect might appear in the level of a series, as for fertility itself; in others, it might be in movements around the trend, as for female labor force participation and divorce. The possibilities range over the various disciplines.
   a. In the area of demography proper, one would expect that pressures on the young of the sort I have been discussing would leave their mark on illegitimacy measures, and I believe there is some evidence to this effect. Thus, O'Connell (1978) finds that in recent decades the proportion of premarital conceptions legitimated by marriages shows a positive swing conforming to the relative economic position of young males. Estimates of the first-birth order illegitimacy ratio (the number of first-birth order illegitimate births per 1,000 total first-order births) also show a sharp contrast before and after 1960 consistent with this relationship. Beyond the area of fertility, mortality and migration might be affected, too. The results above for suicide and homicide may be symptomatic of a more general pattern in stress-related morbidity and mortality, as Ingrid Waldron and Joseph Eyer (1975) have suggested. In

the area of migration, Hope Eldridge's (1964) work has demonstrated how cohort migration was affected by boom and bust Kuznets cycles of the past. It seems reasonable to suppose that, with the reversal of demand and supply roles noted in Chart 1, recent and current migration patterns might show an effect on cohort size.

b. Turning to economics, the analysis has already suggested effects on a number of other labor market variables, such as job turnover and occupational mobility. One might well expect an impact, too, on the size distribution of income through various wage and work channels. On the demand side, effects might be expected on demands for many specific goods, such as schooling, housing, and other child-related items.

c. In sociology, an impact on family formation, dissolution, and structure is suggested by the previous discussion of marriage and divorce. Also, variations in the economic gap between young and old carry obvious implications for the degree of intergenerational conflict.

3. There is need, too, to extend this line of inquiry to other bodies of data.

a. The possibility of investigating these relationships in evidence for other times and places is obvious, although it must be done with care. It was only under the special conditions that emerged after 1940 in the United States that age structure variations came to play such a pervasive role. One would not expect to find evidence of the same mechanism in earlier periods or in markedly different conditions. The most obvious candidates are other developed countries which have experienced recent major shifts in age structure similar to the United States. Several exploratory studies suggest that parallel movements in fertility and relative income or age structure have occurred in several other countries. Reversals in the age pattern of growth in female labor force participation rates also seem to have occurred elsewhere (see O'Connell, 1975, forthcoming; Easterlin and Condran, 1976; Brunborg and Lettenstrøm, 1976; Sweetser and Peipponen, 1967; Leridon, 1978; and United Nations, 1962). But studies of other times and places must be conducted in a context sensitive to the special historical and institutional circumstances of each country.

b. My attention has focused on time-series data, but clearly the mechanisms described here should be evident in cross-section data as well. The results so far of cross-section studies have been mixed, probably in part because relative income mechanisms are obscured in cross-section data by factors that change little or not at all over time. But cross-section studies may ultimately clarify the specific nature of the mechanisms at work.

4. In concluding this section, I'd like to add a note on implications of the present analysis for graduate training in demography. Whereas demographers typically emphasize the compositional effects of age structure shifts, the present analysis focuses on effects working chiefly through age-specific rates. These effects stem from mechanisms identified in economic and sociological theory, usually in connection with such notions as relative income or relative deprivation. To the extent that demographic analysis moves more in the present direction, greater training in these underlying disciplines will be necessary.

ALTERNATIVE VIEWS

What of competing interpretations of recent experience? The dramatic decline since 1960 in the fertility of young women and the rapid rise in their labor force participation has excited much scholarly discussion. A variety of reasons have been given for these developments, such as exceptional employment opportunities for young women, new possibilities for regulating fertility, and new conceptions of women's roles due to the women's movement.

With regard to employment opportunities for young women, the argument is that these have expanded rapidly since 1960, thereby encouraging substitution of work in the marketplace for childbearing and work at home. The basic question, however, is not whether employment opportunities of young women have been expanding since 1960, but whether the labor market for young women was better in, say, the 15 years after 1960 than in the 15 years before, that is, whether this market was better when young women were increasingly plentiful in the population, as after 1960, than before, when they were increasingly scarce. I believe most labor market students would argue, as I would, that the post-World War II labor market for young women was, like that for young men, exceptionally good, and that young women opted for marriage and childbearing, not because of inadequate employment opportunities, but in the face of very good opportunities. [The recent Butz-Ward analysis (1977) adopts the opposite view.] Developments since 1960 can hardly be attributed to an improvement in labor market conditions for young women. The economy's occupational structure did not change in a way especially favoring women after 1960, as compared with before 1960 (see Appendix C). Changes since the late 1940s in the unemployment rates of females aged 20 to 24 imply a stronger labor market through 1960 than thereafter.

Interpretations of the last decade's shift of young females from home to workplace that stress the women's movement or the development and adoption of new fertility control measures are, at least, more consistent with the facts. However, these developments may themselves be attributable partly to age structure effects operating via a relative income mechanism. With regard to fertility control, after 1960 with the growing abundance of young men and the relative deterioration in their labor market position, families felt greater pressure to adopt available fertility control methods—old or new—to restrict births. The rapid spread of new measures of fertility control was, in part, induced by the pressures on young adults described here.

To turn to the women's movement, the recent acceleration in the growth of young women's labor force participation rates is sometimes cited as indicative of the impact of the movement (although a look at the occupational mix of the growth in female employment might give one pause—see Appendix D). It should be noted, however, that the present model, which adopts a traditional view of male and female roles, predicts an upturn in labor force participation among younger women in the period after 1960. Thus, this upturn is not presumptive evidence of the force of the women's movement.

Rather than the women's movement being the cause of the recent accelerated rise in younger females' participation rates, the research findings of Karen Mason and others suggest that the reverse appears more likely to be the case—namely, that the rise in labor force participation led to attitudinal change favorable to the women's movement (Mason et al., 1976). And, since the accelerated rise in labor force participation is, in turn, ascribable to the mechanisms described here, changing age structure was thus one of the forces recently furthering the women's movement. Elsewhere, in regard to the antinatalist shift in population attitudes, I have similarly argued that attitudinal change is primarily an effect, rather than a cause, of changed behavior (Easterlin, 1973, pp. 206–212).

A more general objection to explanations centering on fertility control or the women's movement is that they are confined to the period of the 1960s onward. A basic tenet of the approach adopted here is that a plausible interpretation of experience since 1960 should be consistent with the contrasting experience of the baby boom period. Neither of these views can offer any explanation of the baby boom. Indeed, so far as fertility control possibilities are concerned, the baby boom occurred despite what must have been a major advance in contraceptive knowledge among young persons as a result of information disseminated by the armed forces in World War II and the Korean War.

Moreover, the behavior that those explanations seek to interpret is much narrower than is true here. They apply primarily to fertility and young women's labor force participation. The relative income interpretation I am suggesting seems to fit not only these developments but, as we have seen, a much wider range of evidence, including such magnitudes as divorce, suicide, homicide, and alienation. These other variables are ones that theory would lead one to expect would fit a relative deprivation-type model, and they do. The fact that the present interpretation is consistent with a wider range of socioeconomic data, and that it fits pre-1960 as well as post-1960 experience, tends to strengthen one's confidence in it.

I am not claiming that age structure and relative income are the be-all and end-all in the explanation of these and other developments of the last few decades. The women's movement may have a lasting effect on females' roles through institutionalizing the attitude changes that have occurred, and fertility control developments have not necessarily been wholly induced. It is quite possible that these developments have exerted an independent effect on behavior. Many additional factors have doubtlessly influenced the movements of the variables surveyed here, some specific to an individual variable and others, such as the Vietnam War, more generally. I am arguing only that the evidence is consistent with the view that age structure and relative income have been much more important and pervasive influences in experience since 1940 than has heretofore been recognized.

THE FUTURE

We come finally to the question posed in the title: what will 1984 be like? I have argued that shifts in the relative number of younger versus older adults under conditions of restricted immigration and high and sustained growth in aggregate demand have had a pervasive impact on American life since 1940. If I am correct, then what happens in the near future depends, in important part, on the age structure outlook.

The age structure outlook is shown in Figure 1, which presents the projection to 1995 of the younger and older working age population. A high degree of confidence can be placed in this projection since it is based almost wholly on persons already born. As one can see, growth in the group aged 15 to 29 is currently slowing, and, after 1980, the absolute number will actually begin a period of decline. The *proportion* of those aged 15 to 29 to those aged 30 to 64 is slated to turn down even earlier, showing a decline from 1975 to 1980. This decline becomes more precipitous in the decade of the 1980s. Thus, the outlook is for a growing relative scarcity of younger adults, even more pronounced than in the 1940s and 1950s.

The implication of this new twist in age structure is that there will be a return to the patterns of the 1940s and 1950s. Specifically, with regard to the developments discussed here, there would be:

1. an improvement in the relative income of young males;
2. an upturn in the crude birth rate;
3. a return to the pattern of larger increases in participation rates for older than for younger women;
4. an abatement of the problem of ris-

ing unemployment and accelerating inflation;

5. an eventual drop below trend levels in cohort divorce rates;
6. a decline in the suicide rates of young males;
7. an improvement in crime rates and political alienation;
8. a resumption of the rise in college enrollment rates; and
9. an upturn in SATs.

More generally, I am suggesting that the 1980s will see a turnaround or amelioration in a wide variety of social, political, and economic developments of the last decade or so, some of which have been taken as symptomatic of a hardening social malaise. And, although I am probably alone in predicting such a sweeping reversal, other scholars, reasoning along lines like those presented above, have arrived at similar conclusions for a number of specific magnitudes. Ronald Lee (1976, 1977) and David Goldberg (1978) have each developed fertility projections that foresee a sizable upturn. (See also Goldberg et al., 1977; Wachter, 1975.) Michael Wachter's (1977) projections of labor force participation rates show a return to the pattern of higher increases for older than for younger women in the 1980s. Goldberg and Anderson (1974) anticipate a resumption of the rise in college enrollment rates as the financial squeeze on parents of college-age students moderates in the early 1980s. Robert Zajonc (1976) predicts an upturn in SATs starting around 1980.

A few straws in the wind are provided by the experience of the last couple of years, which suggests that some changes of the last decade may have been arrested. Fertility rates are no longer falling, and the rise in homicide and suicide rates has ceased, at least temporarily. It is too early to tell whether such movements are indicative of a turnaround, but the timing of such developments is consistent with the present analysis.

I want to emphasize that I am talking about the effect of changing age structure via "age-specific" rates, not via "age composition." Both effects should, of course, be taken into account. Heretofore, demographic analysis has been largely or wholly confined to age composition effects. In the current economy, however, age structure effects via age-specific rates are often much more important than the traditional age composition effects stressed in demography. Experience with regard to the crude birth rate provides a good example. In the post-World War II period, the effect on the crude birth rate of age composition changes has been much less than that due to changes in age-specific rates. Moreover, the age composition effect has been in the opposite direction from the effect of age-specific rate changes, and has dampened only mildly the effect on the crude birth rate of the latter. For example, the negative impact on the crude birth rate of the decline since 1960 in age-specific rates has been offset only slightly by an age composition shift favorable to a higher birth rate.

I realize that some of you think that my suggestion that there may be a substantial rise in fertility over the next decade or so is just plain foolish, if not irresponsible. Although fertility is admittedly a difficult magnitude to predict, how can I ignore or discount so heavily the weight of many factors tending to depress fertility? In much of the demographic community, present thinking about American fertility is, I believe, reasonably well captured in the following quotation (Population Index, 1948):

> Short-run birth rates . . . are unpredictable unless based on a combination of demographic, attitudinal, and business cycle research not yet achieved. . . . Long-run birth rates involve even more difficult questions. . . . In regard to these uncertain factors there is at least a minimum of agreement among demographers. No one anticipates the restoration of levels of fertility that could be regarded as high in a world setting. The range of uncertainty is between rates somewhat below permanent replacement and rates slightly above such replacement.

That statement, for which numerous parallels expressed with equal assurance could be found in current writing (see, for example, Westoff, 1978), appeared in *Population Index* exactly 30 years ago, when the post-World War II baby boom was taking off. The fertility prediction was wrong then because, in my view, it ignored a major demographic influence tending to turn fertility behavior around—the growing scarcity of young adults and the resulting rise in their relative affluence. But, on behalf of the writer, a plea of ignorance could justifiably be entered at that time. I think such a plea cannot now be made and that demographers, after enumerating all of the factors making for continued low fertility, should acknowledge that there is at least one important factor working in the opposite direction, a factor that, to judge from the record, has dominated all others over the past 40 years.

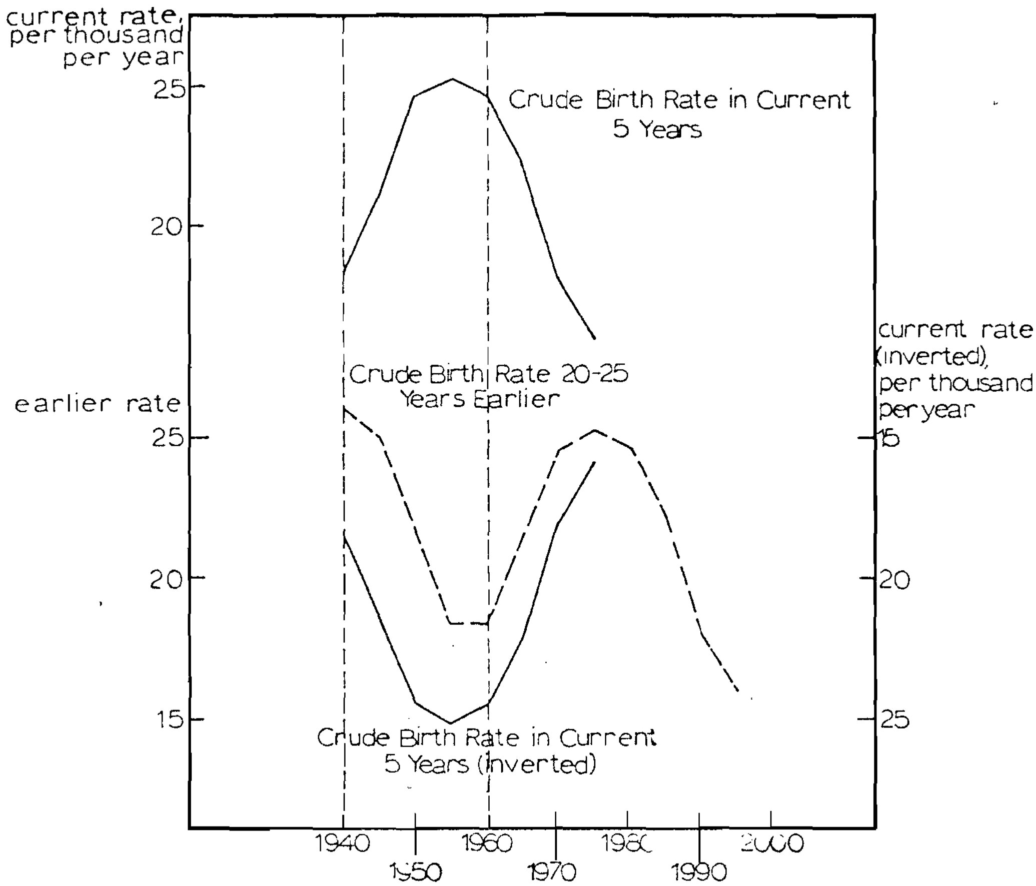
Let me turn from the 1980s to the question of the longer-term outlook. Will the changes in age structure and the attendant effects that we are witnessing continue into the more distant future? Clearly, this hinges on the question of the longer-term outlook for the age structure itself.

The age structure, at any given time, echoes with a lag chiefly the birth rate. This is shown in Figure 1, where movements in the age ratio are seen to follow movements in the birth rate some 20 to 25 years earlier. Thus, the longer-term future of age structure depends on the outlook for the birth rate.

According to the analysis I have presented, however, under current conditions the birth rate is largely being shaped by ongoing developments in age structure. This was the point of Figure 4. We have, then, the possibility of a self-generating mechanism that, in simplest terms, is of the following sort. Current fertility depends on current age structure. Current age structure depends on past fertility. Hence, deleting the mediating role of age structure, current fertility depends on past fertility. (An equivalent formulation, dropping out the mediating role of the birth rate, yields, alternatively, current age structure depending on past age structure.)

The empirical plausibility of this reasoning is illustrated in Figure 9. The current birth rate is the solid line at the top of the figure; the past birth rate, that shaping the current birth rate via its impact on current age structure, is the broken line curve in the middle. Inverting the current birth rate curve, one obtains the solid line curve at the bottom. Note how closely the current swing in the birth rate echoes in inverted form that in the birth rate 20 to 25 years earlier. It seems safe to say that, without the theory just presented, no one would have anticipated this dramatic similarity or even thought of juxtaposing the two series this way. And, if the current birth rate fluctuation so closely mirrors that of the past, may not the future birth rate show a swing mirroring that of the present? And so on into the future? I do not know the answer, and, even if the analysis were correct up to this point, unforeseen factors may subsequently intervene. But the record so far seems consistent with the possibility of a self-generating mechanism that produces repeated swings in the birth rate and age structure and, thus, in socioeconomic conditions of a wide variety. A new Kuznets cycle may have replaced the old—one with a duration of around 40 years rather than 20.

Today, as in the past, the economic fortunes of different generations are shaped in important part by circumstances beyond their control. Before World War II, it was the luck of the draw as to the state of aggregate demand—whether one came of age in a boom or bust period; since World War II, it has largely depended on whether one comes from a large or a small cohort—from a "baby boom" or a "baby bust." But such conditions are not immutable. We have learned from the study of past experience to moderate substantially the variations due to aggregate demand fluctuations—to tame the old Kuznets



Source: Appendix Table 1.

Figure 9.—Crude Birth Rate in Current Five Years Compared With Crude Birth Rate 20 to 25 Years Earlier

cycle. It is reasonable to suppose that, because of recent experience, public policy may eventually be turned, as well, to modifying variations in the fortunes of generations due to their relative numbers, to tempering the new Kuznets cycle.

ACKNOWLEDGMENTS

The research on which this paper reports was funded by NICHHD grant HD-05427, held jointly with Robert A. Pollak and Michael L. Wachter. For comments and assistance I am grateful to Robert Cohen, Frank R. Lichtenberg, Morton Owen Schapiro, and Deborah C. K. Wenger. I have also benefited from the comments of Eileen Crimmins, Robert Higgs, and Michael L. Wachter.

REFERENCES

Abramovitz, M. 1961. The Nature and Significance of Kuznets Cycles. Economic Development and Cultural Change :225–248.

———. 1968. The Passing of the Kuznets Cycle. Economica 35:349–367.

Bancroft, Gertrude. 1958. The American Labor Force. New York: John Wiley and Sons.

Brunborg, Helge, and G. S. Lettenstrøm. 1976. Fertility Trends in Norway since 1965. Paper presented to the Nordic Demographic Symposium, Rungsted, Denmark, June 10–12, 1976.

Butz, William P., and M. P. Ward. 1977. The Emergence of Countercyclical U.S. Fertility. Santa Monica, Calif.: The Rand Corporation.

339

Duesenberry, James S. 1949. Income, Saving, and the Theory of Consumer Behavior. Cambridge, Mass.: Harvard University Press.

Durkheim, Emile. 1951. Suicide, A Study in Sociology. New York: Free Press.

Easterlin, Richard A. 1968. Population, Labor Force, and Long Swings in Economic Growth: The American Experience. New York: National Bureau of Economic Research.

———. 1973. Relative Economic Status and the American Fertility Swing. In Eleanor Sheldon (ed.), Family Economic Behavior: Problems and Prospects. Philadelphia: J. B. Lippincott Company.

———. 1978. Fertility and Female Labor Force Participation in the United States: Recent Changes and Future Prospects. Paper prepared for the International Union for the Scientific Study of Population Conference on Economic and Demographic Change: Issues for the 1980s, Helsinki, August 28, 1978.

———, and G. A. Condran. 1976. A Note on the Recent Fertility Swing in Australia, Canada, England and Wales, and the United States. Pp. 139–151 in Hamish Richards (ed.), Population, Factor Movements, and Economic Development: Studies Presented to Brinley Thomas. Cardiff, Great Britain: University of Wales Press.

———, M. L. Wachter, and S. M. Wachter. 1978. Demographic Influences on Economic Stability: The United States Experience. Population and Development Review 4:1–23.

Eldridge, H. T. 1964. A Cohort Approach to the Analysis of Migration Differentials. Demography 1:212–219.

Employment and Training Report of the President. 1977. Washington, D.C.: U.S. Government Printing Office.

Freeman, Richard B. 1976. The Over-Educated American. New York: Academic Press.

Goldberg, David. 1978. Future of American Fertility: Some Speculations. Paper presented to the annual meeting of the Population Association of America, Atlanta, April 13–15, 1978.

———, and A. Anderson. 1974. Projections of Population and College Enrollment in Michigan, 1970–2000. Paper presented to the Governor's Commission on Higher Education, Lansing, Michigan.

———, J. M. Coble, A. Anderson, and F. Cartford. 1977. Projections of Population and Employment for the Great Lakes States: 1970–2000. Paper prepared for the Upper Great Lakes Regional Commission. Lansing, Mich.: Upper Great Lakes Regional Commission.

Grabill, Wilson H., C. V. Kiser, and P. K. Whelpton. 1958. The Fertility of American Women. New York: John Wiley and Sons.

House, J. S., and W. M. Mason. 1975. Political Alienation in America, 1952–1968. American Sociological Review 40:123–147.

Keyfitz, N. 1972. Population Waves. Pp. 1–37 in Thomas N. E. Greville (ed.), Population Dynamics. New York: Academic Press.

Kuznets, Simon. 1958. Long Swings in the Growth of Population and in Related Economic Variables. Proceedings of the American Philosophical Society 102:25–52.

———. 1961. Capital in the American Economy: Its Formation and Financing. Princeton: Princeton University Press.

Lee, Ronald D. 1976. Demographic Forecasting and the Easterlin Hypothesis. Population and Development Review 2:459–468.

———. 1977. Fluctuations in U.S. Fertility, Age Structure, and Income. Report to the National Institute of Child Health and Human Development. Bethesda, Md.: National Institute of Child Health and Human Development.

Leridon, H. 1978. Fecondité et Structures Demographiques: Une Hypothese sur l'Evolution de la Fecondité depuis 1940. Population 33:441–447.

Mason, K. O., J. L. Czajka, and S. Arber. 1976. Change in Women's Sex-Role Attitudes, 1964–1974. American Sociological Review 41:573–596.

O'Connell, Martin. 1975. The Effect of Changing Age Distributions on Fertility and Suicide in Developed Countries. Unpublished Ph.D. dissertation. Philadelphia: Department of Sociology, University of Pennsylvania.

———. 1978. A Cohort Analysis of Teenage Fertility in the U.S. Since the Depression. Paper presented to the annual meeting of the Population Association of America, held in Atlanta, April 13–15, 1978.

———. Forthcoming. The Effect of Changing Age Distribution on Fertility: An International Comparison. In Julian Simon (ed.), Research in Population Economics, vol. 1. Urbana: University of Illinois Press.

Oppenheimer, Valerie K. 1970. The Female Labor Force in the United States: Demographic and Economic Factors Governing its Growth and Changing Composition. Population Monograph Series No. 5. Berkeley: University of California.

Population Index. 1948. The Population Forecasts of the Scripps Foundation. Population Index 14 (3):188–195.

Preston, Samuel H., and J. McDonald. 1977. The Incidence of Divorce Within Cohorts of American Marriages Contracted Since the Civil War. Paper presented at the annual meeting of the Population Association of America, April 1977, St. Louis, Missouri.

Simon, J. L. 1968. The Effect of Income on the Suicide Rate: A Paradox Resolved. American Journal of Sociology 74:302-303.

———. 1969. The Effect of Income on Fertility. Population Studies 23:327–341.

———. 1975. Response to Barnes's Comment. American Journal of Sociology 80:1460–1462.

Stouffer, Samuel A., et al. 1949. The American Soldier: Adjustment During Wartime Life, vol. I Princeton: Princeton University Press.

Sweetser, F. L., and P. Peipponen. 1967. Postwar Fertility Trends and Their Consequences in Finland and the U.S. Journal of Social History 1:101-118.

United Nations. 1962. Demographic Aspects of Manpower. New York: United Nations.

U.S. Bureau of the Census. 1973. Census of Population, 1970, Detailed Characteristics, Final Report, United States Summary. Washington, D.C.: U.S. Government Printing Office, Table 226, pp. 761-763.

Wachter, Michael L. 1972. A Labor Supply Model for Secondary Workers. Review of Economics and Statistics 54:141–151.

———. 1975. A Time-Series Fertility Equation: The Potential for a Baby Boom in the 1980s. International Economic Review 16:609–624.

———. 1976a. The Changing Cyclical Responsiveness of Wage Inflation. Brookings Papers on Economic Activity 1:115–167.

———. 1976b. The Demographic Impact on Unemployment: Past Experience and Outlook for the Future. Pp. 27–100 in National Commission for Manpower Policy, Demographic Trends and Full Employment. Special Report No. 12. Washington, D.C: U. S. Government Printing Office.

———. 1977. Intermediate Swings in Labor Force Participation. Brookings Papers on Economic Activity 2:545–576.

———, and S. M. Wachter. Forthcoming. The Fiscal Policy Dilemma: Cyclical Swings Dominated by Supply Side Constraints. Pp. 71-99 in Thomas J. Espenshade and W. J. Serow (eds.), The Economic Consequences of Slowing Population Growth. New York: Academic Press.

Waldron, I., and J. Eyer. 1975. Socioeconomic Causes for the Recent Rise in Death Rates for 15-24-Year-Olds. Social Science and Medicine 9:383–396.

Wellford, C. F. 1973. Age Composition and the Increase in Recorded Crime. Criminology (May):61–71.

Westoff, C. F. 1978. Some Speculations on the Future of Marriage and Fertility. Family Planning Perspectives 10:79-83.

Zajonc, R. B. 1976. Family Configuration and Intelligence. Science 192:227–236.

Effects of Cohort Size on Earnings: The Baby Boom Babies' Financial Bust

Finis Welch

University of California, Los Angeles

The arrival of the post–World War II baby boom cohorts in the job market raises many questions of effects associated with a rapidly declining average age of the labor force. This paper first summarizes 1967–75 wage behavior, showing that relative wages between schooling groups have not changed for prime-aged workers, but there is some evidence, for new job-market entrants, that wages of more educated workers have fallen relative to wages of less educated workers. However, changes among schooling groups are small in comparison to those between new entrants and peak earners within schooling group. The evidence is very direct: as work-experience distributions shifted toward increased proportions of young workers, their relative wages fell. After examining a career-phase model in which workers at different phases are imperfect substitutes, estimates of empirical relationships between cohort size and wages are presented. The main result is that income-depressant effects of (own) cohort size decline over the career but do not vanish altogether. Initial effects include reductions in wage rates and in hours and weeks worked, while persistent effects extend only to wages.

If everyone left school and started work on his twentieth birthday, the group of entrants into the U.S. labor market in 1977 would have been 44 percent larger than that in 1962. In fact the 1977 group would be this century's largest, and 15 years later—that is, in 1992—the entering group would be only slightly larger than in 1962. According 344

This project is supported by grants from the Rockefeller and National Science Foundations. I am indebted to Lee Lillard and James P. Smith for many valuable discussions during this work, to Marta Johnson for computational assistance, and to Richard Freeman for comments on an earlier draft. Views expressed are my own and I am responsible for any errors.

[*Journal of Political Economy*, vol. 87, no. 5, pt. 2]

to the assumption of entry at age 20, the number of new entrants has increased each of the last 20–25 years following the 1930s depression-related trough in number of births. But the number of births peaked in 1957 and has continued downward so that the number of new labor-market entrants will fall for at least the next 18 years (for which data on number of births are now available) with no clear indication of an ensuing rise.[1] Instead of labor markets adjusting to increasingly large numbers of entrants, adjustment will be to a declining pool.

It is, of course, true that not everyone begins work at the same age, and the higher the level of school attainment the later the age of job-market entry, so that cohort peaks for college graduates may not have passed. But in the last decade we witnessed a historically unique event—the market's responses to large entering cohorts.

Not only have labor markets been pressed by entry of those just out of school, but the work force has also been increased by the growing participation of married women and the influx of veterans of Vietnam. In the 8 years from March 1967 to March 1975, the (18–64-year-old) civilian labor force grew 21 percent. More important, new entrants during this period had more schooling than earlier cohorts. For example, even though the labor force grew by one-fifth, numbers of participants with 5–8 years of schooling fell 32 percent, as did numbers with 1–3 years of high school. In contrast, during these 8 years the number of high school graduates increased 35 percent, and both for those with 1–3 years of college and for college graduates the number of persons in the civilian labor force jumped an astonishing 64 percent.[2]

The unprecedented change in the age composition of the work force has undoubtedly affected earnings patterns, and this paper attempts to measure these effects. The main conclusion is that pressure of a work force whose average age is rapidly declining reduces wages of new entrants. This supports Richard Easterlin's and Michael Wachter's views of imperfect substitution among workers of different age or work experience.[3] While there is evidence that large cohorts

[1] The peak-sized baby boom cohorts may pull birth totals upward as they reach peak fertility ages, but the early evidence for these cohorts is of very low birthrates. They are so low, in fact, that numbers of births could conceivably continue the downward trend; see Butz and Ward (1979).

[2] These data refer to table A of "Educational Attainment of Workers" (U.S. Department of Labor 1975).

[3] Easterlin's cohort-effect fertility models are well known and in them imperfect substitution plays a central role. More recently, Wachter and Wachter and Easterlin have pursued ideas of imperfect substitution for implications of intermediate cycles both in labor force participation and economic stability. See, for example, Wachter (1977), Easterlin (1978), and Easterlin, Wachter, and Wachter (1978). Michael Wachter (1977) in particular pointed to recent changes in earnings of younger relative to older

receive reduced wages throughout their careers, most of the loss seems to be concentrated in the early years. Evidently the baby boom cohorts are taxed, but their future seems brighter than their experience to date.

The empirical analysis concentrates on earnings within school-completion level. The partition by schooling is used for two reasons. First, I think new entrants and experienced workers are more sharply distinguished at higher levels of education, and if they are, wage effects of expanded cohorts will be larger for them. Second, there is some evidence that during the early 1970s earnings of young college graduates fell relative to earnings of high school graduates. It obviously matters for assessments of the future of education as an investment and as an avenue of socioeconomic mobility whether this change signals a longer-run depression in markets for the products of higher education or is only an aberration as markets digest large entering cohorts.

The next section describes data sources and summarizes changes occurring between 1967 and 1975. Following that, a fairly simple model of career phases is examined for implications of effects of cohort size on earnings. Finally regression results are presented with estimates of these effects.

The Data Base and the Empirical Setting

The data refer to white males described in the 1968–76 March *Current Population Surveys* (*CPS*). Each of these nine surveys includes from 130,000 to 152,000 people. Of these, from 25 to 27,000 are included in the analytical population. They are civilian, white male, age 14 to 65, not now in school (as major activity last week), who either worked 50–52 weeks in the previous year or report the reason for working fewer weeks as something other than being in school or retired. Those self-employed or working without pay are also excluded.

The data are organized on the basis of years of school completed (8–11 years, 12 years, 1–3 years of college, or 4 or more years of college) and estimated years of work experience. Table A1 (see Appendix) reports number of observations by schooling class for each survey year.

One feature of the *CPS* public-use surveys that is especially noteworthy for earnings or wage comparisons is that nontrivial pro- 342

---

males and interpreted the decline as a result of the increased proportion of the population that is "young." While he did not analyze the relationship between wages and factor ratios, his presumption of this relationship served as the justification for examination of relationships between age-specific labor-force participation rates and the age distribution of the population. My approach takes participation as given and concentrates on wage determination.

portions of observations consist of respondents who apparently worked during the year in question, yet who did not report their earnings. In the *CPS* missing observations are filled with an imputation. Unfortunately the imputation procedure used in the first eight surveys differs from that of the ninth so that summary statistics for the 1976 survey (i.e., for 1975 earnings) are not comparable to other years.[4] To further confound analyses, individual records for the first eight surveys contain no flag to identify cases when earnings are imputed. Family records do however identify imputation of total family earnings, which presumably means that earnings for at least one family member are imputed. In contrast, the 1976 survey contains flags for individual imputations but none for families.

I exclude all individuals in the first eight surveys who are in families for which family earnings are imputed. For the final survey, a flag is constructed for families in which any member has imputed earnings, and all persons in these families are excluded. Table A2 (see Appendix) reports proportions by schooling level and survey year of otherwise valid observations excluded on this basis. These proportions range from 7 to 22 percent. They are generally higher for groups with more schooling and are trended, showing higher exclusion rates in more recent years. Unfortunately, exclusions are more frequent for the 1976 survey, suggesting that the constructed family-imputation flag may not correspond perfectly to those for earlier surveys.

Summary statistics reported in this section contain no correction for earnings not reported. There are, however, corrections included in the subsequent regressions to suggest that those not reporting are atypical and, on average, have higher wages than those reporting earnings.

Another feature of the *CPS* surveys is that work experience is not reported. While age, year of birth, and years of schooling are given, we do not know dates when the individuals observed entered the work force. This is troublesome because the purpose here is to examine effects of changing age or experience distributions on earnings patterns. Of course, missing information on dates of job-market entry and accumulated work experience is common to most "census-styled" data, and researchers concerned with income-experience relationships have resorted to various schemes for inferring experience. Here, too, experience is imputed. In fact the assumption of continuous participation following school completion and initial entry is explicit, so the imputation problem is that of estimating ages of job-market entry. This imputation is described in detail by Welch and Gould (1976). Briefly, the procedure estimates a year-of-job-market-

[4] I am grateful to Richard Freeman for pointing this out to me.

entry distribution for each birth–school completion cohort. Account is taken of the fact that more recent birth cohorts finished school at earlier ages. A person observed in one of the *CPS* surveys who worked in the preceding year is assumed to be representative of his cohort subject to the extra information that he began work last year or sometime earlier. Instead of treating an individual as being in a single year on the job, the full work experience is carried so that the number of people in their first year is the sum across individual observations of first-year probabilities. Accordingly, when examining statistics like wages by single experience years, averages are computed using individual probability weights. An example of these year-on-the-job probabilities is given for high school graduates in table A3 (see Appendix). Throughout the summary discussion and subsequent regressions, the data refer to national averages within schooling group for single years of job-market experience. For earnings comparisons, individuals who did not work last year and those who reported an average weekly wage of less than $10 are excluded, yet these people are included in cohort size calculations are those who failed to report earnings. Since earnings data refer to the year prior to the March survey, the dates used refer to earnings rather than survey year. That is, the reference period includes 1967–75.

To summarize changes over this period, I first examine earnings ratios between high school graduates and those with different amounts of schooling. Next, I turn to evidence, within schooling levels, of the general decline in average age of the work force and then compare earnings of recent job-market entrants to earnings of more experienced workers.

Table 1 reports average weekly earnings ratios, 1967–75, for two work-experience intervals, recent entrants (1–5 years) and peak earners (23–27 years). Each ratio is the (geometric) average weekly wage for the indicated schooling group relative to the wage of high school graduates. Table 2 gives corresponding ratios for annual earnings. Although these calculations are based on what by traditional standards are large numbers of observations, the adjacent-year comparisons suggest an uncomfortably high degree of variation relative to overall trend. Regardless of whether this variation refers to ordinary sampling error or to year-to-year variance in labor markets, it should be kept in mind in making trend comparisons.

Relative to high school graduates, new entrants with 8–11 years of schooling pretty much came full cycle over the period. Both weekly and annual earnings declined to 1971 and rose to 1967 levels by 1975. For peak earners, very little happened between 1967 and 1974. Relative earnings fell in 1975 which was by far the worst year of the survey vis-à-vis aggregate unemployment.

TABLE 1

WEEKLY WAGES OF THOSE WITH 8–11 YEARS OF GRADE SCHOOL, 1–3 YEARS OF COLLEGE AND COLLEGE GRADUATES, RELATIVE TO WAGES OF HIGH SCHOOL GRADUATES, 1967–75
(Ratios of Geometric Means)

YEARS OF SCHOOL COMPLETED	1967	1968	1969	1970	1971	1972	1973	1974	1975
A. New entrants (less than 5 years of work experience):									
Grade school:									
8–11 years	.75	.71	.69	.69	.65	.71	.71	.74	.76
College:									
1–3 years	1.09	1.16	1.10	1.10	1.05	1.09	1.07	1.13	1.13
4 or more years	1.49	1.50	1.53	1.54	1.55	1.50	1.48	1.46	1.45
B. Peak earners (23–27 years of work experience):									
Grade school:									
8–11 years	.86	.86	.86	.86	.89	.88	.86	.87	.86
College:									
1–3 years	1.19	1.18	1.19	1.19	1.18	1.17	1.18	1.18	1.12
4 or more years	1.54	1.49	1.51	1.50	1.51	1.52	1.49	1.54	1.47

TABLE 2

ANNUAL EARNINGS OF THOSE WITH 8–11 YEARS OF GRADE SCHOOL, 1–3 YEARS OF COLLEGE AND COLLEGE GRADUATES, RELATIVE TO EARNINGS OF HIGH SCHOOL GRADUATES, 1967–75
(Ratios of Geometric Means)

YEARS OF SCHOOL COMPLETED	1967	1968	1969	1970	1971	1972	1973	1974	1975
A. New entrants (less than 5 years of work experience):									
Grade school:									
8–11 years	.66	.62	.59	.61	.57	.63	.63	.63	.64
College:									
1–3 years	1.12	1.20	1.13	1.16	1.14	1.12	1.14	1.18	1.26
4 or more years	1.55	1.57	1.66	1.72	1.72	1.66	1.57	1.59	1.68
B. Peak earners (23–27 years of work experience):									
Grade school:									
8–11 years	.84	.83	.83	.83	.87	.85	.83	.85	.83
College:									
1–3 years	1.18	1.15	1.21	1.18	1.19	1.17	1.18	1.20	1.17
4 or more years	1.56	1.50	1.57	1.52	1.56	1.55	1.53	1.59	1.56

For those who attended but did not complete college, weekly wages of new entrants followed a similar pattern of relative decay until 1971 and returned at least to initial levels by 1974–75. For peak earners, weekly wages fell in 1975 relative to those of high school graduates but remained roughly constant until then. Annual earnings show no pattern. Only the 1975 jump for new entrants seems important.

Finally, for college graduates, the new-entrant weekly wage pattern is one of rising relative wages until 1971 and falling relative wages thereafter. Relative wages in 1975 are lower than in 1967. This, in a nutshell, is what has been called "the new depression in higher education." But even for new entrants, the annual earnings pattern contrasts with weekly wages. The rising profile to 1971 is consistent, but the subsequent decay is less pronounced. By 1974, relative annual earnings exceeded 1967–68 levels and the 1975 recession seems to have dramatically expanded weeks worked for college relative to high school graduates. For peak earners, there is little systematic change over the period in earnings of college relative to high school graduates.

Noting the decline in earnings of new-entrant college relative to high school graduates between 1969 and 1973 or 1974, Richard Freeman expressed concern that this trend signaled a longer-run depression in the earning power of higher education (see Freeman 1975, 1976, 1977). He conjectured that it occurred despite and not because of the deteriorating general level of economic activity during this period because the preponderance of the evidence suggests countercyclic motion in relative income of college graduates. Had Freeman awaited the 1975 data, or had his analysis extended backward to 1967 (the first annual survey available), his conclusion might have been less pessimistic.

Nonetheless, it is clear that comparisons like those of tables 1 and 2 are confounded by cyclic instability. It is true that college graduates appear to be generally less vulnerable to recessions than are those with less schooling. Most of the literature on this subject explains this phenomenon by reference to one of two closely related observations.

One is that capital goods are less substitutable for college-trained manpower than for high school graduates (see Rosen 1968; Griliches 1969; Welch 1970; Berndt and Christensen 1974). Since in the short run capital is largely fixed, reduced recessionary product demand is accompanied by disproportionate reliance on capital substitutes for reduced output. The other idea is that more educated workers carry larger firm-specific investments in them and, like other forms of capital, their employment is protected during recessions (see Oi 1962; Becker 1964; Rosen 1968). While both of these views have similar predictions for experienced workers, the firm-specific view sharply distinguishes new entrants from experienced workers. If experienced workers carry sunk investments, new entrants imply extra (training) costs, and the analogy to capital is precise: experienced workers are to fixed capital as new entrants are to investments in capital. In investments models it is the fixity of capital that destabilizes demand for investment, and for firm-specific skills it is the fixity of experienced workers that destabilizes demand for new entrants.

In a recent paper James Smith and I (Smith and Welch 1978) examined earnings of new-entrant high school and college graduates using data from the March 1968–75 *CPS*'s. Our finding is that, within industries, business cycles are neutral between high school and college graduates. Any aggregate nonneutrality for them seems to be an artifact of differences in the industrial composition of employment. But, college/high school graduate differences in employment patterns are large and there is much room for compositional effects. For example, 43 percent of all new-entrant college graduates work in service industries (largely health, education, and professional services), while only 11 percent of high school graduates work in these industries. In contrast, 39 percent of high school graduates and only 24 percent of college graduates work in manufacturing. High school graduates work in industries that are disproportionately vulnerable to business cycles and, for "normal recessions," they are more affected than college graduates.

In fact, 1971 appears to have been a normal recession, and as tables 1 and 2 show, relative earnings of college graduates were unusually high that year. But 1973 and 1974 were atypical. In both years industrial employment patterns were mixed with some industries lying above long-run trend and others below trend and in the aggregate employment was relatively depressed in industries disproportionately employing college graduates (Smith and Welch 1978, table 4). Even so, Smith and I could attribute little of the 1969–74 trend in relative earnings of college graduates to business cycles. Either our approach was faulty or the explanation of this change lies elsewhere.

I consider the bulge of entering cohorts a likely alternative. Table 3 shows, for each schooling group, the percentage of all workers in their first 5 years on the job. If experience distributions within schooling group are regular, the fraction of new entrants in the total signals the rate of growth of the total. Higher entry rates for college imply that average levels of schooling are rising. In this vein, convergence between 1967 and 1975 in differences among schooling groups suggests that the rate of growth in average schooling levels is itself falling. This points to what is well known, that rates of progression into college fell after 1969. Although cohorts entering since then

TABLE 3

PERCENTAGE OF WORK FORCE WITH LESS THAN 5 YEARS OF WORK EXPERIENCE BY YEARS OF SCHOOL COMPLETED, 1967–75

YEARS OF SCHOOL COMPLETED	YEAR								
	1967	1968	1969	1970	1971	1972	1973	1974	1975
Grade school:									
8–11 years	8.1	8.5	10.0	12.0	12.3	13.7	15.0	15.1	16.0
12 years	14.4	15.3	15.2	16.5	18.0	19.1	20.7	20.8	20.9
College:									
1–3 years	19.2	19.1	18.7	22.0	23.9	25.3	26.7	25.3	23.5
4 or more years	17.8	19.4	18.9	19.7	21.1	21.1	22.6	22.6	23.5

have more-than-national-average levels of schooling, so that their arrival increases the average, they are not as highly schooled as their recent predecessors. In any case, the changes in entering cohort shares are large for all levels of schooling, and unless new entrants and experienced workers are perfect substitutes, the entry bulge must have depressed new-entrant wages. Table 4 addresses this point. It compares weekly earnings of those with less than 5 years of experience to those with 23–27 years. In every schooling group new-entrant wages fell, and 1967–75 changes swamp anything shown in table 1 or 2. For example, relative to peak earners, weekly wages of new-entrant college graduates dropped 14 percent. Either for new entrants or peak earners, college/high school graduate wage ratios changed by less than 3 percent. Clearly, the most dramatic changes of this period were not in relative wages among schooling groups but in wage structures within schooling group, and, I think, increased new-entrant cohort size is the most likely explanation of this change.

TABLE 4

WEEKLY WAGES OF NEW ENTRANTS RELATIVE TO PEAK EARNERS BY YEARS OF SCHOOLING, 1967–75

YEARS OF SCHOOL COMPLETED	YEAR								
	1967	1968	1969	1970	1971	1972	1973	1974	1975
Grade school:									
8–11 years	.56	.52	.50	.46	.41	.45	.44	.47	.48
12 years	.65	.63	.62	.58	.56	.55	.54	.55	.55
College:									
1–3 years	.59	.62	.58	.53	.50	.51	.49	.53	.55
4 or more years	.63	.63	.63	.59	.58	.55	.54	.54	.54

How Does Cohort Size Affect Earnings?

Only if all workers, regardless of experience, substitute perfectly for each other is the structure of earnings independent of cohort size, of the number of workers at a particular experience level. If perfect substitution were to hold, then the only feasible interpretation of life-cycle profiles would be one of purely physical aging. The investment views, that age profiles are products of learning and depreciation, all suggest that people at different stages of the career do different things. If so, then the value of each thing would reasonably depend on the number of people doing it, and cohort size matters.

But as quickly as we admit this possibility, we are left with the thorny problem of specifying such an effect. At issue is the question of the way the number of workers in one particular class affects their own productivity and that of all other classes.[5] This is a question of substitutability between groups, and we know too little about the nature of these relationships for easy inference.

With workers classified by schooling and experience or date of entering the work force, the number of specifications of substitution patterns is too large to expect unstructured data to be able to sort through them. For example, there may be asymmetries. Young high school and college graduates may be better substitutes in jobs ordinarily performed by high school graduates than in jobs ordinarily performed by college graduates. If so, large entering classes of college graduates may partly ratchet into typical high school jobs when options for entering cohorts of high school jobs do not include switching into jobs ordinarily performed by college graduates.

If job experience congeals initially malleable skills, then substitutability across activities may erode as experience accrues. At the other extreme, if what distinguishes workers is the skills acquired in school, then as these skills erode with work experience, substitutability among schooling groups may increase.

The approach I follow here is to ignore substitution between schooling classes in favor of a sharpened focus within classes. The structure is as though each school-completion group forms a separable branch of an aggregate production process and substitution among schooling classes is independent of levels of experience. Within each schooling group relative wages (productivities) are de-

[5] There have, of course, been several attempts to estimate substitution relationships for worker groups segregated on the basis of age and other things. Recent papers by Anderson (1977) and Freeman (1978) are examples. I avoid this kind of specification for two reasons. First, I do not know how many age groups to specify, but I do know that many inputs are generally not empirically manageable. Second, the abrupt now-you're-in-now-you're-out implications of rigid age demarcations that are usually imposed are unappealing.

termined by numbers of workers at each experience level. I assume that these productivity effects depend only on ratios of number of workers, so that the structure of relative wages across experience levels (the wage profile) depends only on the experience distribution within the group.

The empirical specification I use is to construct a measure of own-cohort size and to estimate own-wage elasticities for this measure. Estimated elasticities are interacted with work experience in an attempt to see if effects dilute as time since entry increases. I think it likely that if larger cohorts have a more difficult start, these effects will erode as the career unfolds.

Career Phases

Perhaps the simplest view of the way cohort size affects earnings follows from the notion that work careers consist of a series of more or less distinct phases. A new job-market entrant arrives fresh from school and enters his profession as a trainee, apprentice, or learner. Only rarely, if his school training is narrowly applied or if the tasks of his profession are very simple, will he immediately become a full-fledged member of the profession. More generally, he enters as a raw recruit or learner, first achieves junior membership, and somewhat later senior membership in the profession.

Just how many phases there are does not really matter. What does matter is that at each phase members of the profession do different things and, vis-à-vis aggregate product, these things are not perfect substitutes. Each activity is productive and marginal productivities are determined, as for any factor, by numbers of workers engaged in all activities. In this view, a profession is an ordered series of worker types. At any moment in the career, a member of the profession is in transit between two of these types and can be viewed as a (convex) combination of them.

This view is essentially identical to the optimal life-cycle configuration suggested by Rosen (1972). In his statement of the problem, a career consists of a continuum of occupations and a worker solves for an optimal occupational sequence by recognizing that each occupation corresponds to learning options that affect performance in subsequent occupations. Rosen allowed productivity in each occupation to depend on number of workers in that and other occupations, and it is clear that had he considered cohort size, the theory would have predicted adverse effects on earnings. To highlight effects of cohort size, I abstract from questions of optimal rates of progression, of transit between career phases, and take them as exogenous. I also abstract from depreciation or skill obsolescence. Presumably the skills possessed on job-market entry are more conducive to learning (think of schooling as learning to learn) and are depleted as the career progresses, so that comparative advantage switches from learning to more directly productive activities. As such, progression is toward higher realized wage activities.

Consider an aggregate production function of the form:

$$y = f(N,Z), \tag{1}$$

where N refers to the productive effort of persons in a given profession (characterized by a given level of schooling). Those things contained in Z include other schooling composites and, like N, each is assumed to form a (weakly) separable branch of the aggregate process, $f(\cdot)$, so that N is only illustrative.

The total effort, N, is itself a function of numbers of workers in each of several worker types,

$$N = g(N_1, N_2, \ldots), \tag{2}$$

where the number of workers of each type is the number of members of the profession devoting their effort to that type of activity. A career phase involves transition between two types such that members of the profession in their first x_1 years on the job are in transition between the first two activities or worker types, in their next $x_2 - x_1$ years transition is between activities two and three, and so on. The ith career phase is passed in experience year, x_i.

If $n(x)$ refers to the number of members of the profession with x years of work experience, then the number of workers of type j is given by:

$$N_j = \int_{x_{j-1}}^{x_j} [1 - p_{j-1}(x)]n(x)dx + \int_{x_j}^{x_{j+1}} p_j(x)n(x)dx. \tag{3}$$

where $p_j(x)$ refers to the part of a worker's time spent in activity j and

$$p_j'(x) > 0;\ x_j \leq x \leq x_{j+1};\ \text{with } p_j(x_j) = 1 \text{ and } p_j(x_{j+1}) = 0.^{6}$$

That is, as a worker enters the ith career phase he initially devotes full time to the ith activity and at that moment begins transition into the $i + 1$st activity. As the ith phase progresses the proportion of time spent in the ith activity decreases until at the end of the phase all of his time is devoted to the next activity and a new phase begins.

The wage of those with x years of experience, their marginal product, is

$$w(x) = \frac{\partial f}{\partial n(x)} = f_1[p_i g_i + (1 - p_i)g_{i+1}], \tag{4}$$

[6] In expression (3), the first (RHS) term is omitted for the first career phase since negative experience has no meaning.

where x falls in the ith career phase. Notice in this phase,

$$\frac{\partial N_i}{\partial n(x)} = p_i \text{ and } \frac{\partial N_{i+1}}{\partial n(x)} = 1 - p_i. \tag{5}$$

Since $w_i = f_1 g_i$ is the wage of activity i workers, the wage of persons in the ith career phase is the combination,

$$w(x) = p_i(x)w_i + [1 - p_i(x)]w_{i+1}, \tag{6}$$

and wage growth during this phase, $[\partial w(x)]/\partial x$, follows from the presumption that the fraction of time allocated to activity i is falling and that realized wages rise with activity level, $w_{i+1} > w_i$.

An individual's wage consists of two parts, the price or marginal product of the profession, f_1, and the individual's own contribution to the level of effort of his profession,

$$\frac{\partial N}{\partial n(x)} = w(x)/f_1 = p_i g_i + (1 - p_i)g_{i+1}. \tag{7}$$

Changes in f_1 are neutral across experience groups and f_1 determines levels but not shapes of wage profiles. In analyzing cohort size effects, I abstract from f_1 which is determined, among other things, by aggregate factor ratios and in empirical analysis over fairly short periods is unlikely to be distinguished from pure trend.

Effects of cohort size on (own) wages are given by the quadratic form,

$$\frac{\partial w(x)/f_i}{\partial n(x)} = \frac{\partial^2 N}{\partial n(x)^2} = (p_i, 1 - p_i)\{G_{jk}\}(p_i, 1 - p_i)' < 0, \tag{8}$$

where

$$G_{jk} = \begin{pmatrix} g_{ii} & g_{i,i+1} \\ g_{i,i+1} & g_{i+1,i+1} \end{pmatrix}$$

and, not surprisingly, if $g(\cdot)$ is (quasi-) concave, the effect of larger cohort size is to reduce wages if there are more than two factors in $g(\cdot)$. 347

For simplicity consider the two-factor constant elasticity of substitution case,

$$g = (\delta_1 N_1^{-\beta} + \delta_2 N_2^{-\beta})^{-1/\beta}, \tag{9}$$

where $\sigma = 1/(1 + \beta)$ is the elasticity of substitution between N_1 and N_2. In this case there are only two activities, learner (N_1) and worker (N_2), and the life cycle can be viewed first as one of transition from learner to worker followed by a period as a fully vested worker. In this example,

$$\frac{\partial w(x)/f_1}{\partial n(x)} = -\frac{1}{\sigma}\theta N_1 N_2 \left(\frac{p}{N_1} - \frac{(1-p)}{N_2}\right)^2, \tag{10}$$

where $\theta = g_1 g_2/N$ and $p = p(x)$ is the fraction of time, at x, spent as a learner.

The profile implied by equation (10) is illustrated in figure 1. The difference between the normal wage profile and that for an illustratively large cohort is never positive. There is, however, a neutrality point where the cohort's division between worker and learner is the same as for all members of the group, and since at that point, the example cohort has no effect on factor ratios, it has no effect on wages.[7]

In this example, new entrants are exclusively learners, and at point of entry they not only draw all their wages as learners, but have the greatest depressing effect on learner wages. As experience accrues, the cohort transits to the worker phase, drawing an increasingly larger share of a wage being depressed by their arrival. In the early career as the cohort is disproportionately involved in the learning phase, transition implies faster-than-normal wage growth to the neutrality point. Afterward, the depressant effect on worker productivity dominates and wage growth is slower than normal. Finally at the point indicated as $p = 0$ in figure 1, when the cohort is fully vested in worker status, the process is completed. Thereafter, wages are depressed and the extent of the depression remains constant.

Several points deserve note. Effects of increased cohort size are inversely proportional to the elasticity of substitution. The substitution elasticity indexes worker-learner differences in the nature of jobs performed. Greater similarity of activities implies greater substitutability. It is likely that the substitution elasticity is related to the transition function, $p(x)$. Rapid transition from learner to worker status implies that learners can easily adapt to worker tasks. I expect that when transition occurs easily, worker-learner tasks are more similar, that is, workers and learners are better substitutes.

This leads immediately to predictions across schooling groups of differences in worker-learner substitution elasticities. It is a near

[7] This is, of course, an artifact of the two-factor assumption. If there were three or more factors and if $g(\cdot)$ were linear homogeneous with a two-factor branch like that of eq. (9), eq. (10) would partition into two effects, one always negative and the other never positive. Equation (10) captures this second effect only, and it vanishes when the cohort being considered divides its time such that it has no effect on the N_1/N_2 ratio. With more than two factors, the cohort being considered would necessarily increase both N_1 and N_2 relative to other factors (regardless of whether it changed the N_1/N_2 ratio) and would therefore reduce wages for the N_1-N_2 composite.

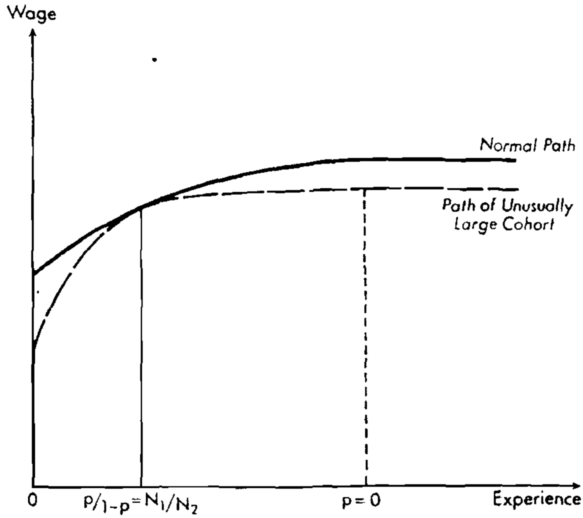


FIG. 1.—Hypothetical contrast between career wage paths of normal and unusually large cohort.

tautology that those who acquire more schooling ultimately perform tasks requiring more training. As total training increases, a balance is reached between learning in the isolation and abstraction of schools and learning on the job. If so, then those having more schooling transit less rapidly from learner to worker status after beginning work, and it is likely that worker-learner substitution elasticities are smaller than for those with less schooling. I will not pretend that I have not seen the data; I do contend, however, that the story the data tell of cohort-size effects increasing with schooling is highly plausible.

Worker/learner wage ratios are determined by three factors in this model. Two of these are treated as parameters in the intermediate process forming the schooling composite. The first refers to the distribution parameter, δ_2/δ_1, and the second refers to the elasticity of substitution. The third is the worker/learner ratio, N_2/N_1.

The normal wage path depicted in figure 1 holds these three factors constant and only describes transition from learner to worker status. If the experience distribution were uniform, that is, if $n(x) = n$ for all x, then the worker/learner ratio is the ratio of the fraction spent in learner status. But a uniform experience distribution implies no growth. Each year the number of new entrants equals the number of retirees. If steady growth occurs, the experience distribution is a negative exponential of the form $q(x) \propto e^{-\rho x}$, such that the ratio, $\ln [n(x - 1)/n(x)] = \rho$, is constant. If growth remains steady, the ratio N_2/N_1 is constant, and the more rapid the growth, the lower the ratio.

TABLE 5

ESTIMATED AVERAGE ANNUAL GROWTH RATES IN NUMBERS OF WHITE MALE WORKERS BY SCHOOLING CLASS, 1967–75

	YEARS OF SCHOOL COMPLETED			
	8–11	12	13–15	16 or More
Estimated growth rate as a percentage	−.31	2.89	4.59	6.25
SE	.07	.09	.11	.17

NOTE.—Numbers of people observed are scaled to take account of the declining size of *Current Population Surveys* relative to the total U.S. population.

Thus, in a regime of steady growth, for other things equal, the wage ratio W_2/W_1 is a positive function of the growth rate.

This is an obvious point which has been ignored in the literature of income returns to schooling. A common finding of such estimates is that life-cycle earnings profiles are more concave for higher levels of schooling, which has been interpreted as implying that early career investments in training (investments in human capital), expressed as a fraction of income potential, increase with the level of schooling. Such a view has obvious intuitive appeal, but greater concavity is just another way of saying that new-entrant wages are reduced relative to those of peak earners. And the reported positive association between concavity of the wage profile and schooling may partly be an artifact of rising average education levels: this, because growing school-completion levels imply positive association between schooling and the learner/worker employment ratio within schooling level which, in turn, implies lower learner/worker wage ratios.

As a crude approximation to average growth rates among schooling groups, I calculated regressions of the form:

$$\ln n(x)_{it} = a_i + \rho_i(t - x) + u_{xit}, \qquad (11)$$

where t indicates survey year, x is experience, and i refers to schooling group.[8]

Implied growth rates are summarized in table 5 and tell an unsurprising story of rising average educational levels. What may be surprising is that the number of persons with 8–11 years of schooling is actually falling while the number of college graduates is averaging an annual growth of over 6 percent.[9]

[8] There are nine surveys and 44 experience levels for a total of 396 observations for each schooling group.

[9] Recall that table 3 shows a rising new-entrant share of the work force with 8–11 years of schooling. The observation in table 5 of negative average growth for this group

Before examining the evidence of cohort size effects on earnings, a few other features of the career-phase model should be considered. First, recall the difference between the normal wage path and that of a large cohort illustrated in figure 1. Relative to the normal path, earnings for the large cohort grow more rapidly before the neutrality point is reached and less rapidly thereafter. Although this model ignores individual optimization concerning on-the-job training, its predictions are similar to what we might expect from optimizing behavior. If large entering cohorts depress wages, then the opportunity cost of on-the-job training is depressed on entry and cost incentives are to speed learning.

In this model, the depressant effect on entry wages diminishes as the N_1/N_2 ratio rises.[10] This follows from the observation that the higher is N_1/N_2, the smaller is the effect of a given size cohort, $n(x)$ on factor ratios at point of entry, $x = 0$.

The level of experience when wage neutrality is reached occurs earlier, the higher is N_1/N_2 and later in the career as the transition process is itself delayed. The existence of a fully neutral point is an artifact of the simplification that the career consists of only two phases, a transitional one and a worker phase. If the career structure is fully hierarchical such that individuals are continuously transiting between adjacent stages, then there would be a series of neutrality points or points of minimum depressant effect, but at no point would wage depression vanish altogether.

Regardless of whether a point of full neutrality exists, any career-phase model will predict that wage depressant effects on job-market entry erode in the early career as transition spreads the cohort bulge among more than one activity. This kind of result is intuitively appealing and (largely) supported by the data.

The Regression Analysis

Before turning to regressions it may be useful to review the underlying data. There are nine surveys referring to earnings years 1967 to

only demonstrates that in the cross-section number of workers increases with experience *on average* and this effect dominates shifts between years. Actually, by 1975 the experience distribution in this group is bimodal. Contrary to earlier trend, the baby boom entry bulge has its impact on high school dropouts as well as for those with more schooling.

[10] Actually, there is some ambiguity on this score. There is none if the worker share of the composite product exceeds one-half or if the worker-learner substitution elasticity exceeds two. Otherwise, if $\delta_2 < 1/2 - [\sigma/(1 - \sigma/2)]$ (where δ refers to worker share) the entry effect increases, in absolute value, as N_1/N_2 increases. There is no ambiguity about the size of the entry effect ($p = 1$) relative to the persistent effect once transition is complete, $p = 0$. That ratio is $(N_2/N_1)^2$.

1975 and, although there is some overlap between adjacent years resulting from the *CPS*'s in 4-out to 8-in 4-month rotation, overlapping observations are not identified. Thus the surveys are treated as though they are statistically independent. All observations refer to white males who presumably are out of school and not retired and are between 14 and 65 years old. Individuals are placed in four school-completion groups.

For each individual, j, work experience is estimated in the form of a full density, $p_{ij}:i = 1{,}45$, where p_{ij} corresponds to the individual's estimated probability of being in the ith year on the job and p_{45} refers to the open-end class, more than 44 years. Experience probabilities are conditional on age, schooling, and year of birth as described in Welch and Gould (1976). These data are aggregated (within schooling group) based on single years of job-market experience, and the total number of persons with i years of experience in a particular survey year is simply the aggregate of the individual probabilities, that is, $n_i = \Sigma_j p_{ij}$.

Within each experience interval two proportions are calculated. One refers to the fraction of observations lost due to failure to report earnings for those who worked (i.e., those individuals in families for which *CPS* imputes earnings) and the other refers to the fraction of observations who did not work and therefore have no reported earnings. Conditional on observations of earnings, (geometric) means of annual and average weekly earnings are then computed within experience interval.

The regression observations refer to means computed within survey year for single-year experience intervals. The open-end interval (more than 44 years of experience) is omitted and observations are pooled across surveys. There are therefore 396 (44 experience levels × 9 surveys) observations for each schooling group.

The objective is to estimate effects of own cohort size on earnings. The first step in computing cohort size involves normalization by the size of the work force. It seems likely that wages of a particular cohort are affected both by its own size and by the size of surrounding cohorts as well. Furthermore, calculated proportions include sampling errors, so for estimation the proportion of group members at each experience level is smoothed by computing a moving average with inverted V weights. Cohort size is then defined as

$$c(x) = \sum_{i=-2}^{2} \alpha_i n_{x+i}, \tag{12}$$

where n_x is the fraction of those in the group who are in their xth year of work experience. The α weights are: $\underline{\alpha} = \frac{1}{3}\left(\frac{1}{3}, \frac{2}{3}, 1, \frac{2}{3}, \frac{1}{3}\right)$, except for recent entrants where succeeding cohort fractions are not defined.

In this case, the α distribution is truncated and remaining weights are scaled accordingly to sum to one.

The objective is to estimate wage elasticities with respect to cohort size and to allow these elasticities to vary over the career. In fact, only one form of experience interaction is examined. It is a spline of the form:

$$\frac{Ew(x)}{Ec(x)} = \gamma_0 S + \gamma_1, \tag{13}$$

where $EZ = \partial Z/Z$ and where $S = \max\{0, 1 - x/\bar{x}\}$. The wage elasticity begins at an entry ($x = 0$) value of $\gamma_0 + \gamma_1$ and then declines linearly (at a rate $-\gamma_0/\bar{x}$) to a permanent level, γ_1 in $\bar{x}$ years, and remains at γ_1 thereafter. Experimentation with alternative values of $\bar{x}$ (the duration of the unique initial effect) over the four schooling groups suggests that it increases with schooling. The estimates reported impose $\bar{x}$ = (6, 7, 8, and 9 years), respectively, for those with 8–11, 12, 13–15, and 16 or more years of schooling.

One characteristic noted in Smith and Welch (1978) is that for the *CPS* data, wage profiles are not adequately described by the commonly used quadratic, experience and experience squared. When estimated profiles are restricted to this form, they consistently overestimate early career earnings. Although we suggested an alternative specification, in this case I have added S (as defined above) to experience and experience squared to allow more flexibility to the wage path. This (splined) experience profile is itself continuous but its derivative is discontinuous at $\bar{x}$.

In addition to the cohort and experience variables, a linear trend term is included along with the national aggregate unemployment rate for white males. The two variables for proportions of observations excluded (see the discussion of eqq. [12] and [13]) also appear as explanatory variables. One is called the exclusion rate due to income imputation. It is the fraction of observations lost in computing mean earnings because of the imputation problem. Inclusion of this variable represents a straightforward attempt to control for selectivity bias in reporting income. If those not reporting have above-average earnings, then the higher the proportion of those not reporting, the lower will be the mean wage of those who do.

The other is called the exclusion rate due to nonwork. It refers to those without valid income observations. Most of those excluded on this basis either are not in the labor force or did not work. Others who appear to have worked were coded as having no earnings. Still others show reported earnings that are so low that I presume they are coded incorrectly. Therefore I exclude from earnings comparisons all who either did not work or had calculated weekly earnings of less than

$10. The variable describing proportions of observations within each experience and survey year lost on this basis is viewed as a composite of selectivity and cyclical effects. Presumably those who are excluded on this basis have below-average earnings so that as this proportion increases, the mean wage of those included should also increase. But, to the extent that the proxy variable used for business cycles (the national average unemployment rate for white males which varies between but not within surveys) controls imperfectly for them, it seems likely that cyclical effects which drive this measure of exclusion may imply that as conditions worsen to increase proportions not working, they also reduce mean earnings of those who do work.

Tables 6, 7, and 8 summarize regression results for mean earnings within schooling class. The regressions reported are weighted by numbers of earners in each experience cell.

Table 6 contains regressions for (ln) annual earnings, table 7 refers to weekly wages, and table 8 refers to weekly wage with the list of explanatory variables being augmented by a part-time variable indicating the proportion of persons in an observational cell who "usually" worked less than 35 hours per week. The part-time variable is viewed as partial control for hours worked. It is not perfect (and if hourly wages were observed, I would have used them), but it is more than "just a dummy" variable. With individual observations, a dummy for less than 35 hours only controls for variation between those working less and those working more than 35 hours leaving within-group variance uncontrolled. But with grouped data, this variable is the proportion of people working less than 35 hours, and with it, there is more effective control for average hours. Consider, for example, an hours-worked distribution over a limited range of, say, 0–70 hours. As work experience changes or as cyclic conditions change, the hours density shifts, and as it does the frequency of hours less than 35 changes. In general, for well-behaved functions we would presume that average hours and the frequency of hours less than 35 move closely together, and if they do, the part-time variable is a "good" control for average hours.[11]

In examining the estimates of tables 6, 7, and 8, notice that calculated standard errors for coefficients understate estimation error.

[11] The type of variation that this measure misses stems from compensated shifts in the density that hold the part-time frequency constant and changes the mean. Such motion seems unlikely. More importantly, by forcing the part-time measure to have only one parameter it captures only average covariation between part-time frequency and mean hours. Since covariance from changes in levels of economic activity is probably different than covariance resulting from changes, say, in potential experience, the forced average effect is incorrectly specified. In any case, in these data the part-time variable is the best available adjustment for hours worked short of resorting to an instrument between it and the hours worked last week for the March survey.

TABLE 6

Regression Estimates of Determinants of Annual Earnings of White Males, 1967–75

Independent Variables	Years of School Completed: Grade School 8–11	Grade School 12	College 1–3	College 4 or More
Cohort size:				
Main effect	−.252	−.120	−.194	−.204
	(12.3)	(6.7)	(13.5)	(11.3)
Interaction with early career spline	.0140	−.302	−.365	−.791
	(.33)	(6.7)	(7.3)	(12.0)
Experience:				
Early career spline	−.690	−1.46	−1.52	−2.71
	(4.2)	(9.7)	(9.9)	(12.7)
Experience	.0428	.0393	.0374	.0600
	(43.2)	(32.7)	(21.7)	(36.9)
Experience squared	−.00073	−.00083	−.00086	−.00146
	(31.8)	(28.3)	(21.6)	(28.0)
Exclusion rate due to nonwork	−.647	.472	−.745	.622
	(4.7)	(2.2)	(2.2)	(2.1)
Exclusion rate due to income imputation	−.231	−.430	−.249	−.855
	(3.1)	(5.6)	(2.8)	(12.7)
Unemployment rate	−.0270	−.0360	−.0287	−.0176
	(10.7)	(16.9)	(13.4)	(8.4)
Trend	.0137	.0137	.0108	.0131
	(10.5)	(11.1)	(7.8)	(12.1)
Intercept	7.3	8.25	8.14	8.23
R^2	.989	.983	.984	.985
SE of estimate	.034	.033	.035	.030

Note.—Number of observations = 396; the dependent and cohort size variables are in (natural) logarithms; the exclusion and unemployment rates are expressed as fractions; absolute values of t-statistics are in parentheses.

With the data arrayed across experience levels and across years, there is probably nontrivial positive serial correlation which I ignored in the calculations reported. Furthermore, residuals are probably (positively) correlated across equations (schooling groups) so that estimates should not be viewed as independent. A preferred estimation procedure to gain efficiency would have recognized both contemporaneous and serial correlations, but software for this kind of procedure is not available to me. As is always true in such cases, estimates may be unbiased but are inefficient, and computed standard errors are biased.

357

Notice that the three sets of regressions in tables 6, 7, and 8 can be

TABLE 7

REGRESSION ESTIMATES OF DETERMINANTS OF WEEKLY EARNINGS OF WHITE MALES, 1967–75

	YEARS OF SCHOOL COMPLETED			
	Grade School		College	
INDEPENDENT VARIABLES	8–11	12	1–3	4 or More
Cohort size:				
Main effect	−.161	−.080	−.163	−.194
	(9.4)	(6.5)	(12.4)	(11.6)
Interaction with early career spline	−.072	−.237	−.278	−.591
	(2.0)	(7.0)	(6.1)	(9.6)
Experience:				
Early career spline	−.849	−1.156	−1.180	2.010
	(6.2)	(10.2)	(8.4)	(10.1)
Experience	.033	.033	.037	.050
	(40.2)	(36.9)	(23.5)	(39.4)
Experience squared	−.00055	−.00067	−.00080	−.00143
	(28.8)	(30.3)	(22.1)	(29.4)
Exclusion rate due to nonwork	−.615	+.103	−.751	+.744
	(5.3)	(.64)	(2.5)	(2.7)
Exclusion rate due to income imputation	−.338	−.441	−.366	−.924
	(5.4)	(7.7)	(4.5)	(14.7)
Unemployment rate	−.0117	−.0184	−.0197	−.0121
	(5.6)	(11.5)	(10.2)	(6.2)
Trend	.0172	.0147	.0129	.0130
	(15.8)	(15.9)	(10.2)	(12.9)
Intercept	3.90	4.47	4.33	4.33
R^2	.988	.986	.984	.985
SE of estimate	.028	.025	.032	.028

NOTE.—Number of observations = 396; the dependent and cohort size variables are in (natural) logarithms; the exclusion and unemployment rates are expressed as fractions; absolute values of t-statistics are in parentheses.

contrasted for implications on hourly wages, hours per week, and weeks per year. The dependent variable in table 6 is the logarithm of annual earnings and is therefore the sum of the logarithms of the weekly wage and of weeks worked per year. Coefficient differences, table 6 less 7, are then coefficients for weeks worked. Similarly, if the part-time variable effectively controls for hours, differences between tables 7 and 8 refer to hours worked; while table 8 coefficients refer to the hourly wage.

These distinctions are useful in contrasting cohort effects, since if larger entering cohorts imply increased job competition, effects

TABLE 8

Regression Estimates of Determinants (including Part-Time Status) of Weekly Earnings of White Males, 1967–75

Independent Variables	Years of School Completed: Grade School 8–11	Grade School 12	College 1–3	College 4 or More
Cohort size:				
Main effect	−.181	−.096	−.168	−.218
	(12.6)	(8.7)	(14.0)	(13.7)
Interaction with early career spline	.118	−.193	−.141	−.503
	(3.5)	(6.4)	(3.2)	(8.6)
Experience:				
Early career spline	.184	−.860	−.596	−1.620
	(1.3)	(8.3)	(4.1)	(8.5)
Experience	.035	.033	.0367	.0581
	(49.4)	(41.9)	(25.6)	(40.8)
Experience squared	−.00058	−.00067	−.00080	−.00141
	(35.9)	(34.5)	(24.1)	(31.2)
Exclusion rate due to nonwork	−.274	+.266	−.517	.608
	(2.7)	(1.9)	(1.9)	(2.4)
Exclusion rate due to income imputation	−.516	−.498	−.457	−.915
	(9.5)	(9.8)	(6.1)	(15.7)
Unemployment rate	−.0139	−.0163	−.0170	−.0095
	(7.9)	(11.5)	(9.4)	(5.1)
Part-time	−.905	−1.113	−.869	−1.400
	(12.8)	(10.7)	(8.8)	(7.8)
Trend	.0172	.0149	.0124	.0136
	(18.9)	(18.3)	(10.8)	(14.4)
Intercept	3.84	4.41	4.32	4.27
R^2	.99	.989	.987	.987
SE of estimate	.024	.022	.029	.026

Note.—Number of observations = 396; the dependent and cohort size variables are in (natural) logarithms; the exclusion and unemployment rates are expressed as fractions; absolute values of t-statistics are in parentheses.

352

should emerge not only for wages but also for labor-utilization rates (weeks and hours worked). Before examining cohort elasticities, consider other implications of these estimates. As is common for this kind of analysis, there is evidence that early-career wage growth is more rapid for those with more schooling. The splined experience coefficients generally imply more rapidly early-career earnings growth than can be captured in the experience quadratic alone.

Evidently, earnings are very sensitive to aggregate unemployment. This perhaps is not surprising for annual earnings where weeks worked are free to vary but may be for weekly earnings holding part-time status constant (table 8). There is at least some support for notions of wage flexibility. In general, wage sensitivity to aggregate unemployment falls with increased schooling, so that these data continue to support the idea that more schooled workers are less affected by swings in aggregate labor demand.

Coefficients on the exclusion rate due to income imputation refer to adjustments in mean earnings of those reporting income associated with variations in proportions reporting. These coefficients are consistently negative to suggest that those with the highest wages are the most likely to not report earnings. In fact, mean rates for income not reported together with these coefficient estimates imply average annual earnings for the combined total of those who do and do not report that exceed the average of those reporting by 3–5 percent in the first and fourth schooling groups and by 6 percent for high school and 14 percent for college graduates. This is important especially for income comparisons across schooling groups, as, for example, in rate-of-return calculations. The implication of table 6 is that annual earnings of college relative to high school graduates are 8 percent higher in the total population than in the reporting population. This could result in underestimates of rates of return that as an order of magnitude would approach 2 percentage points, about 20 percent of most current estimates.

Notice also that magnitudes of coefficients on this variable increase as control for labor utilization increases (in tables 7 and 8), so that the problem of understatement due to nonreporting seems more severe in weekly wage comparisons.

There is an often-stated belief that the annual *Current Population Surveys* are more reliable than the decennial census data mainly because of greater effort to ensure reporting. Evidently, nonreporting remains a serious problem even for *CPS*. Furthermore, analysts using *CPS* data should be aware that when missing observations are filled with imputations, the assumption of the procedure used is that nonreporters are representative of reporters. Unless my estimates are completely wrong, they are not.[12]

[12] To impute missing income observations, the census bureau uses what it calls a "hot deck" procedure. Briefly, this procedure fills a matrix stratified on various characteristics with the "last" valid observation and uses that observation to fill in the blanks. Aside from convenience there is little to recommend this approach, especially in light of the emerging literature on selectivity biases. The number of strata on which imputations are based was expanded for the 1976 survey to include schooling, labor-force status of spouse, marital status, region, and type of residence, and more detail was introduced for age, family relationship, occupation, class of worker, weeks worked, race, etc. For a description of this procedure and for contrasts with earlier imputations, see U.S. Department of Commerce, Bureau of the Census (1976).

Coefficients on exclusions due to nonwork do not have consistent sign patterns across schooling groups but are consistent across income definition. As such, unambiguous interpretation is not easy. In any case the overall average exclusion rate due to nonwork is less than 3 percent, so that this correction has little effect on estimated mean earnings.

In these data, dependent variables are measured in constant (CPI) dollars and the evidence is of very low rates of growth in real earnings or wages. Since the unemployment rate is held constant, it is tempting to argue that growth implied by the year variable reflects cyclic corrected longer-run trends, but in my opinion it would be a mistake to do so. Recall that the observation period is itself strongly trended vis-à-vis aggregate economic activity. The first 3 years, 1967–69, were more robust than any to follow, and 1975 was by far the most depressed. The question relevant to trend comparisons is whether the unemployment rate alongside the nonwork exclusion rate adequately controls for cycles so that the residual trend can be taken as a secular one.

In these data, there are a variety of indexes of labor utilization rates which are themselves indexes of labor demand. In addition to weeks worked and the part-time variable, the most obvious is weeks unemployed or looking for work and the exclusion rate itself. In auxiliary calculations (not reported here), I regressed each of these variables on the same set of independent variables as those of tables 6 and 7.[13] The result in every schooling group is that even though the aggregate unemployment rate is taken into account, weeks worked are negatively and weeks looking for work are positively trended through the period. Except for college graduates, the exclusion rate due to not working is also positively trended. Part-time status alone seems devoid of trend. Evidently, the unemployment rate is too imperfect an instrumental predictor of labor demand for the implied trend rates of earnings to be taken as estimates of long-run patterns. In any case, estimated trend growth rates show little evidence of differences among schooling groups.

353

Estimated Cohort Effects

Table 9 summarizes cohort elasticities from regressions reported in tables 6–8. The entry elasticity is the respective wage or earning elasticity evaluated for those in their first year on the job ($x = 1$), and the persistent elasticity is the main effect, copied from tables 6–8.

[13] Of course when the exclusion rate is the dependent variable, it is not included as an independent variable. It is included, however, for other dependent variables.

TABLE 9

ESTIMATED ELASTICITIES OF ANNUAL AND WEEKLY EARNINGS WITH RESPECT TO COHORT SIZE

	Years of School Completed			
	Grade School		College	
Estimated Elasticity	8–11	12	1–3	4 or More
A. Annual earnings:				
Entry	−.240	−.369	−.514	−.907
	(8.8)	(10.9)	(12.7)	(12.1)
Persistent	−.252	−.080	−.194	−.204
	(112.3)	(6.6)	(13.5)	(11.3)
B. Weekly wages, part-time status excluded:				
Entry	−.221	−.283	−.406	−.720
	(9.8)	(11.2)	(11.1)	(14.8)
Persistent	−.161	−.080	−.163	−.194
	(9.4)	(6.6)	(12.4)	(11.6)
C. Weekly wages, part-time status included:				
Entry	−.082	−.261	−.291	−.665
	(3.8)	(11.6)	(8.1)	(14.5)
Persistent	−.181	−.096	−.168	−.218
	(12.6)	(8.7)	(14.0)	(13.7)

NOTE.—Absolute values of t-statistics in parentheses.

Only those with 8–11 years of schooling show persistent elasticities exceeding entry elasticities (in parts *A* and *C*). This result is anomalous, and I have no explanation for it. Each of the other schooling groups shows that entry effects exceed persistent ones.

Aside from high school dropouts, wage and earning elasticities rise with level of schooling and most of the discrepancy occurs in the early career. The career-phase model predicts that effects are inversely proportionate to worker-learner substitution elasticities. Thus the finding of effects increasing with schooling level suggests that worker-learner substitution elasticities fall with increased schooling.

One of the most interesting patterns exhibited in table 9 is that entry elasticities fall between part *A*, where weeks worked are free to vary, and part *B*, where they are not. These elasticities also fall between part *B*, where hours worked are variable, and part *C*, where they are partially controlled. The evidence then is that large entry cohorts depress not only wages but also hours and weeks worked. In contrast, persistent elasticities are much less sensitive to control for weeks and hours. Evidently, although new-entrant labor-utilization rates are depressed, this effect is transient.

It may be surprising that wage effects have as large a persistent component as they do. I, for one, would not have been surprised to find no persistent elasticity, an indication that cohort size effects are restricted to congestion at point of entry. Nonetheless, career-phase models suggest that although entry effects may exceed persistent ones, it is imperfect substitution between work activities performed at different phases of the career that generates both entry and persistent effects. From this perspective, the existence of an entry effect implies a persistent effect and vice versa. It is important, however, in assessing the size of persistent effects to keep in mind the operational definition of cohort size as the share of the current total work force. Through time as the working population expands, the share accounted for by a particular cohort declines.

How Large Are the Estimated Effects?

Are the estimated effects presented in table 9 large enough for those interested in earning patterns to view fluctuations in cohort size as a legitimate concern? We know from table 3 that the age or experience composition of the work force changed significantly in an 8-year interval. And table 4 shows falling relative earnings of new entrants that are coincident with increasing new-entrant shares of the total work force.

There are two obvious ways to try to get an intuitive feel for orders of magnitude of earnings changes implied by the parameter estimates in table 9. The first is to examine predicted changes in the range of the data used here, that is, for the 1967–75 interval. The second is to examine the historical record for longer periods to see what patterns of shifting cohort size have been and to combine them with parameter estimates for implications of effects on lifetime earnings.

For the *CPS* data, among new entrants (those in their first 5 years on the job) 1967–75 growth in cohort size yields predicted reductions in weekly wages ranging from a low of 6 percent for those with 1–3 years of college to a high of 13 percent for college graduates. Implied reductions are 8 percent for high school graduates and 12 percent for those with 8–11 years of schooling.

354

For all but college graduates, relative cohort size of peak earners (23–27 years of work experience) fell over the period, and resultant predictions of wage growth range between 3 and 5 percent. For college graduates, relative cohort size of peak earners did not change. Thus, new-entrant weekly earnings are predicted to have fallen relative to wages of peak earners by 15 percent for those with 8–11 years of schooling, by 10 percent both for high school graduates and for those who attended but did not finish college, and by 13 percent for college graduates. These predictions, which ignore changes other than cohort size, are reasonably close to the observations reported in table 4, especially for those with 8–11 years of grade school and for college graduates. Predicted changes understate observations for high school graduates and for those with 1–3 years of college.

In comparing wages across schooling groups, between 1967 and 1975 the observed college/high school graduate new entrant weekly wage ratio fell something less than 3 percent. The change predicted by cohort growth with larger wage responses for college graduates is 5 percent. If these estimates are correct, then what has passed for a new depression in higher education may be unique to the entrants of the early 1970s. For them, as effects erode over the career, the future is brighter, and for subsequent arrivals who themselves will be members of smaller cohorts, the future is also brighter.

The story of large swings in total numbers of births and of secular drift in average schooling levels is well known. To get a rough idea of how these changes affect cohort shares of the work force, and therefore lifetime earnings, I have merged the historical record of number of births with what evidence I could find for school-completion distributions to construct cohort shares similar to those used in the regressions described above.

I have proceeded under the assumption that after 1975, the number of births in the United States will stabilize at 2.5 million per year (roughly the level maintained from 1971–75) and that average levels of school completion will continue their upward drift for a few years and then stabilize. For each level of school completion, I assume that all members of a birth cohort who attain that level begin work at the same age and retire after 45 years. With these assumptions, I construct a "guesstimated" work experience × schooling distribution for each calendar year, 1940–2035. The final year corresponds to retirement for new entrants of 1990, so that the constructed distributions permit estimates of lifetime earnings for cohorts joining the work force over the half-century 1940–90.

For each market-entering cohort, work-force shares are calculated for every year from entry to retirement, and these shares are the raw input for calculations of cohort size effects on lifetime earnings. Aggregation of a particular entering cohort's shares over its career gives a crude basis for measuring total cohort size.

As expected, these aggregates show extreme swings in cohort size. For example, the group with 8–11 years of schooling has a minimum average cohort share for 1953 entrants and a maximum for 1977 entrants, and the average share of the largest cohort exceeds that of the smallest by 62 percent. For high school graduates, the smallest cohort entered in 1954 and the largest in 1975, and the largest accounts for a 55 percent greater lifetime average work-force share than the smallest. Those with 1–3 years of college have even greater variation in lifetime cohort shares. The minimum for 1956 entrants is only 45 percent of the maximum (for 1967 entrants). For college graduates, the smallest cohort entered in 1962 and the largest in 1970, and the largest has a lifetime average share 51 percent bigger than the smallest. In general, these patterns reflect the depression-related trough in number of births and the late 1950s culmination of the postwar baby boom. Timing differentials reflect delayed entry for those with more schooling and swings about trend in school completion.

It is of course obvious that cohort size effects on life earnings do not correspond perfectly to the average lifetime shares just described. Cohorts that are large as they enter the work force and small as they leave it are penalized relative to those with similar average shares but with more uniform career representation. In a sense, something analogous to double-discounting is involved. First there is time preference, and I assume a 5 percent discount rate to represent it. Second, there is the fact that, aside from high school dropouts, estimated cohort size effects erode over the career.

In view of the extraordinarily large swings in cohort size, the estimates for effects on life earnings are not massive but, to the cohorts most affected, they may not be trivial. Using coefficient estimates for part C of table 9, which refer to my best estimates of hourly wage effects and presumably, therefore, correspond more closely to full incomes, variations for high school graduates are restricted to a 4 percent range. Not surprisingly, there is more action for college graduates where a 10 percent range is implied.

Calculations for college graduates suggest that, in terms of pure cohort size alone, the most favored generation entered the work force in 1962 and the least favored entered only 8 years later. As noted, life earnings for 1970 entrants (ignoring secular real wage growth) are estimated to be about 90 percent of the level for 1962 entrants. This focus on extremes omits much that is of interest in comparisons across entering cohorts.

For example, if the presumptions of future birthrates and school-attainment levels are at all accurate, then the wage depression for college graduates implied by the arrival of baby boom cohorts that starts in the mid-1960s will begin to erode around 1980, and the estimates for 1990 entrants suggest career earnings equal to those of the "most favored" classes who enter between the mid-1950s and early 1960s. Although the baby boom cohorts seem to have the poorest career income prospects, my simulations suggest that their lifetime incomes will not be significantly below the life earnings of cohorts entering in the mid-1940s. These earlier entrants joined the work force during a period of rapid acceleration in numbers of college graduates. As such, they began their careers with very large cohort shares, and although these shares fell rapidly under continuing growth in numbers of entrants, the early career effects swamped the (double-discounted) late career effects. In comparison to the baby boom groups, the mid-1940 entrants have smaller average work-force shares but the shape of the career profile nullifies what would otherwise be an advantage.

It is not necessary to enumerate the heroic nature of these simulations. They are only an exercise to derive a "feel" for orders of magnitude of the estimated effects involved and to examine the historical record of swings in the experience composition of the work force. But, if these machinations reasonably describe cohort patterns in life earnings, then one final observation is in order: The depression in early career earnings of college graduates witnessed in the first part of this decade is not unique, and just as earlier depressions eroded into higher incomes, the recent depression will be followed by higher college-graduate earnings.[14]

Summary and Conclusions

Anyone working with wage and earnings data for the period 1967 to 1975 must be impressed by variability. Real annual earnings fell in 1971, in 1974, and again in 1975 for experienced workers. For new entrants, real earnings fell in 1970 and in 1972 and again in 1974 and 1975. College graduates are, however, somewhat atypical. Although they shared in the 1970–72 recession, effects were less severe. However, the foundering economy of late 1973 and 1974 affected college graduates more than others, and in 1975 when real income fell sharply for most schooling groups, income of college graduates, especially those with considerable work experience, actually rose.

This kind of variability challenges those concerned with wage behavior and gives much room for statistical exploration, yet so much happened during the period that any set of relatively simple explanations is likely to be confounded by competing explanations. Clearly business cycles should be taken into account, but the evidence is that the composition of the cycles that occurred was different enough to raise questions of "how."

Relative incomes changed among schooling groups, especially for

[14] In a recent paper, T. W. Schultz (1977) exhibits a variety of relative wage calculations for different times in the twentieth century. Working with four groups: unskilled, manufacturing production workers, teachers, and associate professors, Schultz describes four periods of rapidly declining relative wages of skilled workers (1915–19, 1932–35, 1940–42, and 1969–75). Since only the last two of these periods correspond to rapid cohort expansion, there would seem to be a real advantage to further exploration of the historical record. In particular, Schultz stresses that both inflation and levels of economic activity are candidates for further investigation.

new entrants. Why? Is it cycles or a longer-run depression in the market for college graduates? The hypothesis put forward here is that whatever it is, it is not long run. The 1975 data challenge any notions of persistence.

The most dramatic change in earnings structures witnessed during these 9 years was the drop in earnings of new entrants relative to more experienced workers, and this change coincided with the arrival of the peak-sized cohorts spawned by the post–World War II baby boom. How does a market digest an increasingly large number of new arrivals? As a device for describing effects of cohort size on earnings, I have explored a career-phase model in which activities performed at various phases are not perfect substitutes for activities performed at other phases. I then attempted to fit a (loosely specified) version of this model to earnings data taken from the March *Current Population Surveys* for 1968–76.

By conventional statistical standards, there is fairly strong evidence that large cohorts do depress earnings and that most of the effect comes early in the career. The evidence also suggests that cohort size effects increase with level of schooling. Whether these effects are real or just statistical illusions depends on what the future brings. Population age structures suggest that entering cohort sizes are probably already falling for lower levels of schooling and will do so soon for higher levels. The prediction of the estimates presented is that when this happens, wages of new entrants will rise relative to experienced workers and that, given similar reductions in cohort size, wages of new-entrant college graduates will rise relative to those of workers with less schooling.

Appendix

TABLE A1

NUMBERS OF INDIVIDUAL RECORDS INCLUDED FOR ESTIMATION OF EARNINGS PROFILES AND COHORT SIZES BY YEARS OF SCHOOL COMPLETED, 1968–76

Years of School Completed	Year								
	1968	1969	1970	1971	1972	1973	1974	1975	1976
Grade school:									
8–11 years	6,823	6,466	6,344	6,159	5,892	5,453	4,997	4,578	4,406
12 years	7,817	7,934	8,123	8,487	8,352	8,280	7,994	7,763	7,462
College:									
1–3 years	2,373	2,489	2,584	2,699	2,651	2,712	2,731	2,729	2,772
4 or more years	3,065	3,061	3,097	3,312	3,550	3,383	3,546	3,518	3,477

SOURCE.—*Current Population Surveys* (March 1968–76).

TABLE A2

PERCENTAGES OF OTHERWISE VALID OBSERVATIONS LOST DUE TO INCOME IMPUTATIONS BY YEARS OF SCHOOL COMPLETED AND SURVEY YEARS, 1968–76

YEARS OF SCHOOL COMPLETED	YEAR								
	1968	1969	1970	1971	1972	1973	1974	1975	1976
Grade school:									
8–11 years	10.6	13.7	10.0	10.9	10.1	11.6	14.2	15.8	18.8
12 years	10.9	13.8	10.8	9.9	10.2	11.9	14.0	15.0	20.0
College:									
1–3 years	12.0	15.0	11.4	11.7	11.6	12.7	14.3	16.4	19.6
4 or more years	12.4	15.3	12.4	12.6	7.3	13.3	14.4	16.2	22.4

TABLE A3

ESTIMATED AVERAGE EXPERIENCE PROFILES BY AGE (17–30) FOR HIGH SCHOOL GRADUATES, 1967–75

	YEAR AT WORK							
AGE (Years)	1	2	3	4	5	6	7	8
17	.592	.276	.098	.029	.005	...	...	...
18	.535	.275	.128	.046	.014	.002	...	...
19	.458	.287	.151	.071	.025	.007	.001	...
20	.359	.290	.186	.097	.046	.016	.005	.001
21	.244	.259	.220	.146	.077	.036	.013	.004
22	.153	.201	.217	.189	.127	.067	.031	.011
23	.086	.136	.181	.198	.174	.119	.062	.029
24	.045	.083	.129	.173	.189	.166	.113	.060
25	.017	.044	.082	.127	.169	.186	.164	.112
26	.005	.017	.043	.080	.125	.168	.185	.164
27	.003	.007	.108	.043	.080	.124	.166	.183
28	.002	.004	.009	.020	.004	.079	.123	.165
29	.001	.002	.004	.010	.021	.044	.079	.122
30	.001	.002	.003	.005	.011	.021	.043	.077

References

Anderson, Joseph. "An Economic Demographic Model of the U.S. Labor Market." Ph.D. dissertation, Harvard Univ., 1977.

Becker, Gary S. *Human Capital*. New York: Columbia Univ. Press (for Nat. Bur. Econ. Res.), 1964.

Berndt, E. R., and Christensen, L. R. "Testing for the Existence of a Consistent Aggregate Index of Labor Input." *A.E.R.* 64, no. 3 (June 1974): 391–404.

Butz, William P., and Ward, Michael P. "The Emergence of Countercyclical U.S. Fertility." *A.E.R.* 69, no. 3 (June 1979): 318–28.

Easterlin, Richard A. "What Will 1984 Be Like? Socioeconomic Implications of Recent Twists in Age Structure." Presidential address to the Population Association of America, Atlanta, April 1978.

Easterlin, Richard A.; Wachter, Michael L.; and Wachter, Susan M. "Demographic Influences on Economic Stability: The United States Experience." *Population and Development Rev.* 4, no. 1 (March 1978): 1–23.

Freeman, Richard B. "Overinvestment in College Training." *J. Human Resources* 10, no. 3 (Summer 1975): 287–311.

———. *The Overeducated American.* New York: Academic Press, 1976.

———. "The Decline in Economic Rewards to College Education." *Rev. Econ. and Statis.* 59, no. 1 (February 1977): 18–29.

———. "The Effect of Demographic Factors on the Age-Earnings Profile in the U.S." Unpublished paper. Harvard Univ., Dept. Econ., August 1978.

Griliches, Zvi. "Capital-Skill Complementarity." *Rev. Econ. and Statis.* 51, no. 4 (November 1969): 465–68.

Oi, Walter. "Labor as a Quasi-fixed Factor." *J.P.E.* 70, no. 6 (December 1962): 538–55.

Rosen, Sherwin. "Short-Run Employment Variations on Class-I Railroads in the U.S., 1947–63." *Econometrica* 36 (July/October 1968): 511–29.

———. "Learning, Experience and the Labor Market." *J. Human Resources* 7, no. 3 (Summer 1972): 326–42.

Schultz, T. W. "Education in an Unstable Economy." Human Capital Paper no. 77:1, Univ. Chicago, Dept. Econ., January 1977.

Smith, James P., and Welch, F. "Local Labor Markets and Cyclic Components in Demand for College Trained Manpower." *Annales de l'INSEE* 30–31 (April-September 1978): 599–630.

U.S. Department of Commerce, Bureau of the Census. "Money Income and Poverty Status of Families and Persons in the United States: 1975 and 1974 Revisions (Advance Report)." *Current Population Reports.* Consumer Income Series P-60, no. 103. Washington: Government Printing Office, September 1976.

U.S. Department of Labor, Bureau of Labor Statistics. *Special Labor Force Report No. 186.* Washington: Government Printing Office, March 1975.

Wachter, Michael L. "Intermediate Swings in Labor-Force Participation." *Brookings Papers on Economic Activity*, no. 2 (1977), pp. 545–74.

Welch, F. "Education in Production." *J.P.E.* 78, no. 1 (January/February 1970): 35–59.

Welch, F., and Gould, W. "An Experience Imputation or an Imputation Experience." Mimeographed. RAND Corp., 1976.

357

to 23.0 million. The ratio of participants less than 35 years of age to participants 35 and over jumped from .46 to .67.[2]

What are the consequences for the wage structure of such sizable changes in the age composition of the labor force? Does an increase in the relative number of young workers alter their wages relative to the wages of older workers? To what extent do cross-sectional age-earnings profiles respond to exogenous demographic shifts? What are the possible implications of the experience of the late 1960s and 1970s for the decade of the 1980s, when the number of young workers is expected to decline?

The answers to these questions depend on the degree of flexibility of relative wages by age or, in the context of standard labor demand analysis, on elasticities of complementarity [10, 16] which link changes in factor prices to changes in the supply of inputs. If the relevant elasticities are large, changes in the age composition of the labor force can substantially alter age-earnings profiles, whereas if elasticities are small, profiles will be relatively independent of the number of workers in different age groups.

This paper studies the effect of changes in the age structure of the workforce on age-earnings profiles in the U.S. and provides estimates of the relevant elasticities of complementarity for young and older workers. It focuses in particular on developments in the late 1960s and in the 1970s, when the relative number of young workers increased sufficiently rapidly as to provide a strong "test" of the potential dependence of the profile on exogenous shifts in the age structure of the labor supply. Section I documents the magnitude of the change in the age structure of the workforce which constitutes the "experimental" variation under study. Section II presents evidence from the Current Population Surveys of the U.S. Bureau of the Census that the age-earnings profile of male workers changed in the period, with the ratio of the earnings of older men to the earnings of younger men rising sharply. Section III develops the relevant labor demand and production function models needed to analyze the link between changes in the age structure of the workforce and changes in age-earnings profiles. Section IV estimates the extent to which shifts in age-earnings profiles can be attributed to demographically induced movements along labor demand schedules whose magnitude is reflected in the relevant elasticities of complementarity.

The principal finding is that the age-earnings profile of male workers, which has traditionally been viewed as a stable economic relation determined by human capital investment decisions, appears to be significantly influenced by the age composition of the workforce. Apparently because

2 All of the data in this paragraph are from [20], Table A-2, p. 137, and from [21], various editions.

younger and older male workers are imperfect substitutes in production, changes in the number of young male workers relative to older male workers substantially influence the ratio of the earnings of younger men to the earnings of older men. The effect of changes in the relative numbers of workers of different ages on age-earnings profiles is especially marked among college graduates. By contrast, the age-earnings profile of female workers, which tends to be quite flat, appears to be little influenced by the age composition of the female workforce, possibly because the intermittent work experience of women makes younger women and older women closer substitutes in production.

Whether the sizable decline in the earnings of the large cohort of young workers entering the market in the 1970s relative to the earnings of older cohorts will persist, creating a lifetime "size of cohort" earnings effect, or whether the new entrants of the 1970s will significantly catch up in earnings in future years remains to be seen.

The dependence of the age-earnings profile on demographically induced movements along a relative demand schedule suggests that standard human capital models of the profile, which posit that earnings rise with age or experience *solely* as a result of individual investment behavior, are incomplete. If, as found in this study, elasticities of substitution or complementarity among age groups are not infinite, human capital cannot be treated as a homogeneous input with a single rental price, whose "units" of investment determine the age-earnings profiles. Differences in the activities of younger and older workers and in the demand for those activities decisively influence the shape of the profile. To understand the relation between earnings and age, it is necessary to analyze the demand for workers by age and employer personnel policies as well as to analyze human capital investment decisions.

I. THE CHANGING AGE STRUCTURE OF THE WORKFORCE

The broad outlines of the remarkable change in the age structure of the workforce of the United States under investigation is examined in Table 1, which records the absolute number of workers aged 20–24 and 25–34 and 359
the number aged 20–24 and 25–34 relative to the number of workers aged 35 and over for the period 1966 to 1976. The table treats male and female workers and college graduates and high school graduates separately. These patterns of change stand out in the table: a remarkable increase in the number of young workers in total, the result of the "baby boom" that followed World War II; an even greater percentage increase in the number of young college graduate workers, the result of the unprecedented proportion of young persons choosing to enroll in college in the 1960s; and an

THE EFFECT OF DEMOGRAPHIC FACTORS ON AGE-EARNINGS PROFILES*

RICHARD B. FREEMAN

ABSTRACT

The age-earnings profile of male workers is significantly influenced by the age composition of the workforce. When the number of young workers increased sharply in the 1970s, the profile "twisted" against them, apparently because younger and older male workers are imperfect substitutes in production. The effect is especially marked among college graduates. By contrast, the age-earnings profile of female workers appears to be little influenced by the age composition of the female workforce, possibly because the intermittent work experience of women makes younger and older women closer substitutes in production. The dependence of the age-earnings profile on demographically induced movements along a relative demand schedule suggests that standard human capital models of the profile, which posit that earnings rise with age and experience *solely* as a result of individual investment behavior, are incomplete.

The age distribution of the labor force varies greatly over time in response to long-term demographic changes. Historically, high birth rates produce a relatively large number of younger labor force participants and an age distribution skewed toward younger ages, while low birth rates have the opposite impact.[1] Because of the "baby boom" that followed World War II and peaked in 1955–60 [18, p. 56], there was an especially significant change in the age structure of the U.S. workforce in the late 1960s and early 1970s, when the number of young persons increased extremely rapidly. From 1966 to 1976, the number of labor force participants aged 20–24 grew from 9.7 million to 16.7 million, while the number aged 25–34 grew from 16.8 million

The author is Professor of Economics, Harvard University.

* The research was supported by U.S. Department of Labor Grant 21-25-78-19. I have benefited from the comments of Richard Layard and the referees. [Manuscript received August 1978; accepted January 1979.]

1 For models of the link between birth rates and the age distribution of workers, see Keyfitz [13].

THE CHANGED STRUCTURE OF THE WORKFORCE, BY SEX AND EDUCATION

	Workforce, by Age				College Graduate Workforce, by Age				High School Graduate Workforce, by Age			
	Numbers (in 000s)		Relative to Workers 35+		Numbers (in 000s)		Relative to College Workers 35+		Numbers (in 000s)		Relative to 35+	
	20–24	25–34	20–24	25–34	20–24	25–34	20–24	25–34	20–24	25–34	20–24	25–34
A. *Male Workers*												
1966	6,139	10,761	.201	.352	280	1,067	.129	.490	2,057	3,929	.249	.475
1968	6,788	11,376	.221	.371	290	1,137	.131	.514	2,066	4,220	.232	.473
1970	7,378	11,974	.241	.391	376	1,200	.161	.514	2,324	4,529	.248	.483
1972	7,795	12,806	.257	.423	567	1,393	.232	.569	2,772	4,792	.288	.497
1974	8,105	13,993	.270	.465	585	1,725	.222	.655	3,011	5,110	.289	.507
1976	8,421	14,990	.282	.502	686	2,203	.242	.779	3,334	6,309	.330	.525
% change 1966–76	37.2	39.3	40.2	42.6	145.0	106.5	84.7	59.0	62.1	60.5	42.2	11.0
B. *Female Workers*												
1966	3,601	4,516	.220	.276	331	444	.352	.472	1,879	2,063	.311	.342
1968	4,251	5,104	.251	.301	436	526	.397	.479	2,104	2,358	.317	.356
1970	4,893	5,704	.276	.322	515	614	.483	.576	2,400	2,764	.322	.371
1972	5,337	6,525	.298	.365	620	801	.540	.698	2,593	3,068	.331	.392
1974	5,867	7,826	.322	.430	699	1,101	.565	.890	2,598	3,452	.319	.424
1976	6,339	9,183	.336	.487	757	1,388	.565	1.04	2,935	3,958	.346	.466
% change 1966–76	76.0	103.3	52.7	76.4	128.8	121.6	60.5	120.3	56.2	91.9	11.3	36.3

Sources: Workforce data compiled from *Employment and Training Report of the President* [20], transmitted to Congress 1977, Table A-2, p. 137. Workers by education data compiled from "Educational Attainment of Workers" [21], Special Labor Force Reports, various editions.

especially marked increase in the number of young female workers, the result of a sizable jump in the labor force participation rate of young women in the late 1960s and 1970s.

Among male workers, the numbers aged 20–24 and 25–34 increased by over one-third from 1966 to 1976, while the number of college graduates aged 20–24 and aged 25–34 more than doubled. As a result of these changes, the ratio of all male workers 20–34 to those 35 and over rose from .55 in 1966 to .78 in 1976, while the ratio of male college graduate workers 20–34 years of age to male college graduate workers 35 and over increased from .62 (1966) to 1.02 (1976). According to the figures in Panel B, the number of female workers aged 20–24 increased by 76 percent, while the number aged 25–34 increased by 103 percent. These gains outstripped the rate of increase of the older female workforce by sufficient magnitudes to raise the ratio of 20–24 to 35+-year-old female workers by 53 percent and the ratio of 25–34 to 35+-year-old female workers by 76 percent. Among college graduate workers, the number of young women aged 20–24 more than doubled, while the number aged 25–34 more than tripled. While less dramatic, the number of young female high school graduate workers also increased, particularly in the 25–34 bracket.

Data on the number of new high school and college graduates and on the number of new graduates entering the labor market tell a similar story about the influx of young persons into the workforce, though the timing of changes necessarily differs from that in Table 1. Between 1960 and 1970, the number of new high school graduates per thousand persons in the civilian labor force rose from 26.8 to 35.0.[3] From 1960 to 1972, the number of high school graduates and dropouts entering the job market grew from 13.2 per thousand members of the civilian labor force to 17.1 per thousand members of the civilian labor force.[4] For college graduates, the picture is more complex, as the tendency to enroll for graduate studies in the 1960s delayed the labor market entrance of the large classes of the sixties until the following decade. From the late 1960s to the mid-1970s, however, the ratio of new bachelor's graduates on the market to the civilian labor force increased sharply.[5]

Because data on graduates refer to flows rather than stocks of persons in a wide age grouping, they reveal further the beginning of the decline in the 360

3 Figures on high school graduates are from [18], p. 153. Figures on the civilian labor force are from [20], Table A-3, p. 139.

4 Figures on high school graduates and dropouts in the labor force are from [22], "Employment of June 1970 High School Graduates," Special Labor Force Report No. 15, Table 2, and from *Handbook of Labor Statistics 1977*, Table 31, p. 77. Data on the civilian labor force are from [20], Table A-3.

5 See the estimates given in Freeman [7], pp. 67–68.

number of young workers which will mark the 1980s. From 1972 to 1977, the number of persons graduating from high school grew by a bare 4 percent [18, p. 153]. From 1974 to 1976, the number of persons graduating from college fell by 3.0 percent [18, p. 153]. The beginning of the demographic change from the large cohorts of young persons of the late 1960s and 1970s to the smaller cohorts of the 1980s can be seen in these figures.

While not a "classical experiment" with all other factors held fixed, the sizable increase in the number of young workers found in Table 1 and in the graduate data cited above and the concurrent change in the age composition of the workforce constitutes the type of exogenous shift in the age composition of the workforce that should provide a reasonably strong test of whether the age-earnings relation depends on factors beyond investments in training. Was the change in supply accompanied by changes in relative wages?

II. CHANGES IN THE RELATIVE EARNINGS OF YOUNG WORKERS

Evidence on earnings by age from the Current Population Survey of the Bureau of the Census provides an answer to this question. The CPS data show that for male workers, who traditionally have had steep cross-sectional age-earnings profiles, the demographic changes of the late 1960s and of the 1970s were accompanied by a substantial "twist" in the age-earnings profile against the young. By contrast, among female workers, who have traditionally had flat age-earnings profiles, the demographic changes do not appear to have altered the relative wages of the young. While other factors (to be explored in Section IV) may have also been at work, the concatenation of increases in relative numbers and decreases in relative wages in the period is highly suggestive of movement along a negatively sloped demand curve, with a moderate elasticity of substitution between workers of different ages.

Table 2 summarizes the CPS evidence on age-earnings profiles in terms of the ratio of the income or earnings of 45–54 or 45–49-year-old workers to the incomes or earnings of 20–24, 25–29, or 25–34-year-old workers, in toto and for high school graduates and college graduates taken separately. The figures in lines 1–8 are taken from the U.S. Bureau of the Census's *Current Population Reports,* with the data for 1975 adjusted to take account of changes in the imputation procedure used by the Census. As discussed in *Current Population Reports,* Consumer Income Series P-60, No. 105, the Census made major changes in its method of computing incomes in 1975 when it introduced a new imputation procedure to estimate missing records. The new procedure tends to raise the average earnings of more educated and older workers relative to what would have resulted from the previous imputation procedure. For comparability over time, the 1975 data in the table are adjusted to a pre-1975 basis by multiplying the reported 1975 figures by the ratio of incomes in 1974 calculated from the old imputation procedure to incomes in 1974 calculated from the new procedure, using unpublished Census tabulations.[6]

Lines 1–4 treat the mean incomes of all workers who report greater than zero values for the year. These incomes are likely to be sensitive to cyclical changes in unemployment rates. Lines 5–8 are based on the incomes of year-round full-time workers, which should be less sensitive to the cycle and which should offer a better measure of rates of pay than the incomes of all workers. For women, the sizable number of part-time workers and significant nonwage incomes makes interpretation of the incomes for all workers complex, suggesting that attention be focused on year-round full-time employees. The final lines of the table record ratios of usual weekly earnings. These figures have the advantage of referring to labor market earnings rather than total incomes and of covering narrower age groups. They suffer from lack of information on years of schooling.

The figures for male workers in the table show a substantial change in age-earnings profiles, with the income or earnings of older men rising sharply relative to the income or earnings of younger men. For all men, the data in lines 1 and 2 show an increase of 20 percentage points (9 percent) in the ratio of the incomes of 45–54-year-olds to the income of 20–24-year-olds and an increase of 5 percentage points (4 percent) in the ratio of the incomes of 45–54-year-olds relative to the incomes of 25–34-year-olds. The comparable changes for year-round and full-time workers in lines 5 and 6 are 24 points (6 percent) and 8 points (7 percent). With the best measures of rates of pay, median usual weekly earnings, there is a rise in the ratio of the earnings of 45–49-year-olds to the earnings of 20–24-year-olds not enrolled in school of 33 points (25 percent) from 1967 to 1977 and a rise in the ratio of the earnings of 45–49-year-olds to the earnings of 25–29-year-olds of 15 points (14 percent).

The figures for college and high school graduates show further that the "twist" in male age-earnings profiles was most pronounced for college graduates. Between 1968 and 1975 the ratio of the incomes of 45–54-year-old year-round and full-time male college graduates to the incomes of 25–34-year-old year-round and full-time male college graduates rose by 25 percentage points (18 percent), while the comparable ratio for male high school graduates rose by 11 points (9 percent). Similarly, for all male college graduates, the ratio of the incomes of 45–54-year-olds to the incomes of 25–34-year-olds rose by 19 percentage points (13 percent), while the ratio of

6 Let W'_{1974} = income of 1974 calculated by old imputation procedure, W_{1974} = income in 1974 calculated by new imputation procedure, and W_{1975} = income in 1975 calculated by new imputation procedure. Then the adjustment is $(W'_{1974}/W_{1974})(W_{1975})$.

362

TABLE 2
RATIOS OF THE INCOMES OR EARNINGS OF OLDER TO YOUNGER WORKERS
BY SEX AND EDUCATION, 1967–77

Income Measure and Group	Males 1967–68[a]	Males 1975[b]	Males 1977	Females 1967–68[a]	Females 1975[b]	Females 1977
Mean incomes of all workers						
1. Ratio of the income of persons aged 45–54 to the income of persons aged 20–24	2.30	2.50	—	1.38	1.48	—
2. Ratio of the income of persons aged 45–54 to the income of persons aged 25–34	1.18	1.23	—	1.16	1.05	—
3. Ratio of the income of persons aged 45–54 to the income of persons aged 25–34, high school graduates	1.21	1.30	—	1.24	1.13	—
4. Ratio of the income of persons aged 45–54 to the income of persons aged 25–34, college graduates	1.43	1.62	—	1.35	1.22	—
Mean incomes of year-round and full-time workers						
5. Ratio of the income of persons aged 45–54 to the income of persons aged 20–24	1.74	2.00	—	1.21	1.35	—
6. Ratio of the income of persons aged 45–54 to the income of persons aged 25–34	1.18	1.26	—	1.00	1.01	—
7. Ratio of the income of persons aged 45–54 to the income of persons aged 25–34, high school graduates	1.20	1.31	—	1.07	1.08	—
8. Ratio of the income of persons aged 45–54 to the income of persons aged 25–34, college graduates	1.38	1.63	—	1.05	1.14	—
Mean usual weekly earnings, full-time white workers						
9. Ratio of the earnings of persons aged 45–49 to the earnings of persons 20–24 out of school workers	1.27	1.57	1.60	—	—	—
10. Ratio of the earnings of persons aged 45–49 to the earnings of persons 25–29	1.06	1.19	1.21	—	—	—

Sources: Current Population Reports [17], No. 66, Tables 39, 41; No. 101, Tables 53, 59; No. 105, Table 4. U.S. Department of Labor, Bureau of Labor Statistics: unpublished tabulations from May 1967, 1975, and 1977 *Current Population Surveys.*

a 1968 in lines 1–8; 1967 in lines 9–10.

b Figures for 1975 in lines 1–8 based on unpublished Census data which give incomes in 1974 and 1975 on a comparable basis. As discussed in Series P-60, No. 105, the Census used a different inputation procedure for estimating incomes in 1975 The new procedure tends to bias upward the earnings of more educated and older workers.

the incomes of high school graduates 45–54 years old to the incomes of high school graduates 25–34 years old rose by just 9 percentage points (7 percent).

The data for women tell a different story. The mean incomes for all workers show a rise in the income of women aged 45–54 relative to the income of women aged 20–24, but do not show a rise in the income of women aged 45–54 to the income of women aged 25–34 and depict a drop in the income of older high school or college graduate women relative to younger high school or college graduate women (lines 3 and 4). On the other hand, the mean incomes for year-round and full-time workers show an increase in the ratio of the incomes of 45–54-year-olds compared to the incomes of 25–34-year-olds. The unclear pattern in the data may reflect the general flatness of the age-earnings profile for women, particularly for year-round and full-time workers. If the flat profiles result from high substitutability between workers of different ages in production,[7] large changes in the age structure of the population would have little or no impact on the profiles.

Because age-earnings profiles differ markedly between men and women and appear to have changed only for men in the period covered, the remainder of this study will focus solely on male age-earnings relations. The flat and apparently stable profiles among women suggest very different economic processes at work, which merit separate detailed study beyond the scope of the current inquiry.

Analysis of Male Earnings Using CPS Data Tapes

More precise estimates of the extent of change in the age-earnings profiles of male workers can be obtained by linear regression analysis of the effect of age on earnings using the CPS data on individuals that underlie the published aggregates. Regression analysis of the data for individuals has several advantages over comparisons of published means: it permits investigation of labor market earnings (rather than of incomes, as given in *Current Population Reports* Series P-60); it permits calculation of weekly earnings (yearly earnings over weeks worked) as an indicator of rate of pay; it allows for greater disaggregation of workers by age and education; and it can be used more readily to make statistical tests of the significance of observed changes.[8]

To estimate changes in age-earnings profiles among male workers in the

7 This statement is simply a speculative hypotheses. Groups of workers could have similar earnings for any number of reasons unrelated to elastiticies of substitution. Moreover, a high elasticity of substitution need not produce similar wages for groups of workers.

8 The sampling errors reported in *Current Population Reports* [17] can be used to make statistical tests of changes in incomes, but are not well suited to test changes in ratios of incomes.

period under study using CPS data on individuals, the following linear regression model was fit with the March 1969 and March 1978 tapes[9] (which give earnings for the preceding year):

$$\ln W_{ij} = a + \Sigma_{ij} b_{ij} E_i A_j + cR_{ij} + \mu_{ij} \tag{1}$$

where W_{ij} = weekly or annual earnings of workers in the ith education and jth age groups; E_i = dichotomous dummy variable that takes the value 1 for persons in the ith education group and the value 0 otherwise, with i covering seven education groups: 0–8 years of schooling, 9–11 years, 12 years, 13–15 years, 16 years, 17 years, and more than 17 years; A_j = dichotomous dummy variable that takes the value 1 for persons in the jth age group and 0 otherwise, with j covering six age groups: 18–24, 25–29, 30–34, 35–44, 45–54, and 55–64; R_{ij} = dummy variable which takes the value 1 for blacks; and μ_{ij} is assumed $N(0,\sigma^2)$. One interaction term is omitted to define the base group.

The results of the calculations are summarized in Table 3 in terms of *differences* between the coefficients on the log earnings of workers aged 45 to 54 and the coefficients on the log earnings of workers in younger age brackets for persons with 12 years of schooling (e.g., high school graduates) and for persons with 16 years of schooling (e.g., college graduates), taken separately. Comparable estimates for other education groups are also available, but for simplicity are not given in the table. In the left side of the table the dependent variable is the log of weekly earnings, the best measure of rates of pay on the March CPS files. The right-hand side of the table treats the log of annual earnings. The computations confirm the finding of a sizable decline in the earnings of young workers relative to the earnings of older workers, particularly among the more educated. Among high school graduates, the log weekly earnings of 45–54-year-olds is .18 log points higher relative to the log weekly earnings of 18–24-year-olds in 1977 than in 1968; it is .17 points higher relative to the log earnings of 25–29-year-olds in 1977 than in 1968; is .09 points higher relative to the log earnings of 30–34-year-olds in 1977 than in 1968; and is .10 points higher relative to the log

9 Unfortunately, there is no way to deal directly with the problem of changes in imputation procedures between the 1969 and 1978 CPS tapes. Prior to 1976 the public use tapes do not contain "flags" for imputed earnings, making comparable regressions impossible. As the new imputation procedure appears to have greatly affected the earnings of workers by education but not by age, however, the results are unlikely to be seriously marred by this problem. Regressions with the March 1975 tape (which used the old imputation procedure) give, in any case, results much like those found with the March 1978 tape. For a detailed discussion of the imputation problem, see *Current Population Reports* [17], No. 105. That report shows that age comparisons are only modestly biased by the change in imputation procedure, but that education comparisons are seriously biased. There is a marked jump in the relative earnings of college graduates in the March 1976 tape due to the new imputation procedure.

REGRESSION ESTIMATES AND STANDARD ERRORS ON THE DIFFERENCE BETWEEN THE LOG OF EARNINGS OF MEN AGED 45–54 AND OF YOUNGER MEN, FOR COLLEGE AND HIGH SCHOOL GRADUATES, 1968 AND 1977

Education and Age Group	Coefficient and Standard Errors for Earnings of 45–54-Year-Olds Relative to Other Group Weekly Earnings		Change in Relative Earnings	t-Test of Change[a]	Coefficient and Standard Errors for Earnings of 45–54-Year-Olds Relative to Other Group Yearly Earnings		Change in Relative Earnings	t-Test of Change[a]
	1968	1977			1968	1977		
12 years of schooling								
18–24	.60(.03)	.78(.06)	.18	2.57	.95(.03)	1.02(.08)	.17	1.89
25–29	.13(.03)	.30(.06)	.17	2.43	.17(.03)	.37(.08)	.20	2.22
30–34	.02(.03)	.11(.06)	.09	1.29	.02(.03)	.16(.08)	.14	1.56
35–44	−.03(.03)	.07(.06)	.10	1.43	.03(.03)	.08(.08)	.06	0.67
45–54	.00	.00	—	—	.00	.00	—	—
16 years of schooling								
18–24	.83(.05)	1.13(.06)	.30	3.75	1.18(.06)	1.48(.11)	.30	2.31
25–29	.28(.05)	.58(.07)	.30	3.33	.37(.05)	.70(.10)	.33	3.00
30–34	.10(.05)	.40(.07)	.30	3.33	.10(.05)	.47(.10)	.37	3.36
35–44	.01(.04)	.18(.07)	.17	2.14	.02(.05)	.21(.10)	.19	1.73
45–54	.00	.00	—	—	.00	.00	—	—
Number of observations	30,231	29,842	—	—	30,231	29,842	—	—
R^2	.248	.272	—	—	.297	.284	—	—

Source: Obtained by regression of the log of earnings on dummy variable for race and education by age, as described in the text. The regressions included age-education dummy variables for the following education groups: 0–8 years of education, 9–11 years of education, 13–15 years of education, 17 and 18+ years of education and for 55–64-year-olds in addition to the groups in the table. The March 1969 and March 1978 Current Population Survey Tapes were used in the analysis.

a The significance for the t-tests are: 5 percent level, 1.65; 1 percent level, 2.33.

earnings of 35–44-year-olds in 1977 than in 1968. Among men with four years of college, the changes are greater, with increases of .30 log points in the difference between the log earnings of 45–54-year-olds and the log earnings of 18–24, 25–29, and 30–34-year-olds, respectively, from 1968 to 1977. In other education groups, the results are comparable. While not all of the changes in the difference in log earnings are significant at the 5 percent level, those for the more educated group are highly significant.

The calculations using log annual earnings tell a similar story, with the earnings of the older workers rising relative to those of younger workers, especially among the college educated.

We conclude that there was, in fact, a substantial change in the age-earnings profiles for male workers during the period when the number of young workers increased relative to the number of older workers. Are these two changes causally linked? To what extent can the twist in the age-earnings profile against the young be attributed to the sharp increase in the relative number of young workers?

III. LABOR DEMAND AND AGE-EARNINGS PROFILES

In the framework of standard labor demand analysis, the impact of changes in the relative supply of workers by age depends on the substitutability of inputs. When labor of different ages is readily substitutable for other inputs, large changes in supply will cause only modest changes in wages by age. Conversely, when substitutability is limited, sizable changes in age-earnings profiles are needed for demand to adjust to changes in the age composition of the workforce. Since the number of workers of different ages is taken as exogenously determined by demographic changes, the demand for labor schedule becomes a wage-determination schedule, linking factor prices to factor quantitites. In such a model, Hicks's elasticity of complementarity, which relates factor prices to inputs (see [10, 16]) is the appropriate elasticity concept, rather than the standard Allen elasticity of substitution. In this study I will concentrate on the inverse of the Hicks elasticity, defined as: 324

(2) $$\dot{S}_{ij} = (\dot{W}_i/\dot{L}_j)\ 1/\alpha_j$$

where W_i = wage of factor i; L_j = amount of factor j; α_j = share of factor j in cost; and where dots above variables represent log changes (i.e., $\dot{W}_i = d\ln W_i$) and where the elasticity S_{ij} is taken with other inputs held fixed but with output allowed to vary.[10] The "own elasticity" of complementarity S_{ii} is implicitly defined as

10 Output must be allowed to vary in this experiment because the quantity of inputs is changing, which will change output.

$$\text{(3)} \qquad \Sigma_j \alpha_j S_{ij} = 0 \text{ so that } S_{ii} = -\Sigma_{j \neq i}(a_j/a_i)S_{ij}$$

S_{ii} is negative for normal production functions.

When all inputs vary, the factor price determination equation of factor i can be written as:[11]

$$\text{(4)} \qquad \dot{W}_i = \Sigma_{j \neq i}\alpha_j S_{ij}\dot{L}_j + \alpha_i S_{ii}\dot{L}_i = \Sigma_{j \neq i} a_i S_{ij}(\dot{L}_j - \dot{L}_i)$$

Equation (4) shows how changes in the quantities of all inputs or changes in the quantities of inputs relative to the input i alter the wage of input i. To see how changes in inputs alter relative wages, let i index one input (say, older workers) and k index another input (say, younger workers), then subtract the factor price equation based on (4) for input k from that for input i to obtain:

$$\text{(5)} \qquad \dot{W}_i - \dot{W}_k = \Sigma_{j \neq i \neq k}[\alpha_j(S_{ij} - S_{kj})\dot{L}_j] + \alpha_i[S_{ii} - S_{ki}]\dot{L}_i - \alpha_k[S_{ik} - S_{kk}]\dot{L}_k$$

When factors other than k and i do not affect the ratio of their marginal products, the elasticity of complementarity is equal to one divided by the standard Allen elasticity of substitution (see [10, 16]), giving the following relative wage determination equation:

$$\text{(6)} \qquad \dot{W}_i - \dot{W}_k = -(1/\sigma_{ik})(\dot{L}_i - \dot{L}_k)$$

where σ_{ik} = elasticity of substitution between i and k, and $\sigma_{ik} > 0$.

Production Functions and Elasticities

The relevant elasticities of complementarity or substitution for workers in different age groups are estimated in this study assuming one of two functional forms for the production process: the constant elasticity of substitution (CES) form and the translogarithmic (TL) form.

The CES form has both desirable and undesirable features for empirical analysis of the effects of changes in the relative number of workers by age on age-earnings profiles. On the positive side, since the CES has only one elasticity of substitution, the appropriate relative wage equation is of the form of (6), which is reasonably simple to estimate and which provides direct information on the impact of relative quantities on relative earnings. The possibility that changes in relative wages, of the type observed in the 1970s, are due largely to cyclical rather than demographic changes can be readily examined with equation (6) by addition of variables measuring the business

11 To derive (4), rewrite (2) as $\dot{W}_i = a_j S_{ij} \dot{L}_j$. This gives the change in $\dot{W}_i$ due to changes in $\dot{L}_j$, all else the same. To get the change in $\dot{W}_i$ due to changes in all other factors, we sum $\dot{W}_i = a_j S_{ij} \dot{L}_j$ across all j, which yields the expression (4) in the text.

cycle. The major disadvantage of the CES is that it cannot be used to test the possibility that in a consistent production function framework changes in other factors, such as capital, may be influencing the relative demand for labor of different ages.

The translogarithmic (TL) production function provides an appropriate system for examining the effect of changes in the supply of several inputs on age-earnings profiles. The TL form yields a consistent system of demand equations with potentially different elasticities of complementarity (or substitution) between any pair of inputs. In the translog production system, the production function is:

$$\ln Y = A + \Sigma_i \alpha_i \ln L_i + 1/2\, \Sigma_i \Sigma_j \gamma_{ij} \ln L_i \ln L_j \tag{7}$$

The derived demand equations are:

$$\alpha_i = a_i + \Sigma_j \gamma_{ij} \ln L_j \tag{8}$$

subject to cross-equation and within-equation restrictions on the parameters

$$\Sigma a_i = 1 \tag{9}$$

$$\Sigma_j \gamma_{ij} = 0 = \Sigma_i \gamma_{ij} \tag{10}$$

where α_i = share of input *i* in cost, and L_j = amount of input *j*. Equation (8) relates the share of each input in cost to the quantities of inputs; since quantities are given, the equations are in effect relative wage equations, dependent on how the wage components of cost change when quantities change. The Hicks elasticities of complementarity can be readily derived from the translog system by differentiation with respect to the relevant L_i (see [16, p. 47]). The resultant equations are:[12]

$$S_{ij} = (\gamma_{ij})/(\alpha_i \alpha_j) + 1 \tag{11}$$

$$S_{ii} = (1/\alpha_i)^2 [\gamma_{ii} + \alpha_i^2 - \alpha_i] \tag{12}$$

Equations (11) and (12) relate the elasticities of complementarity to the parameters of the equation and factor shares.

The TL system has one major advantage as a model of demand: it provides estimates of elasticities of complementarity for more than two inputs in a consistent production function framework. It has two disadvantages. First, specification or measurement error in the equation for a factor of only marginal concern, such as for capital in the case at hand, can greatly impact estimates of the demand equations for other factors. Second, the TL

12 Since $\alpha_i = L_i f_i / f$ (where f is the production function), we rewrite (8) as $f_i = f(a_i + \Sigma\gamma_{ij}\ln L_j)/L_i$. Then $f_{ij} = (f\gamma_{ij}/L_iL_j) + f_j(a_i + \Sigma\gamma_{ij}\ln L_j)L_i$. Since $f_i/f = (a_i + \Sigma\gamma_{ij}\ln L_j)/L_i$, the right-hand side of the expressions can be further simplified to $(f\gamma_{ij}/L_iL_j) + f_jf_i/f$. Multiplying both sides by F/f_if_j yields the expression in the text for S_{ij}. Derivation of S_{ii} is similar.

model is an equilibrium model that cannot be readily modified to allow for the effect of cyclical factors on relative demands.

Since neither the CES nor TL models of production or demand are without problems, the effect of the age composition on relatives wages by age will be estimated using both models. By using the two functional forms, each of which has weaknesses and strengths, I hope to obtain a better fix on the key demand relation under study than could otherwise be done.

IV. DETERMINANTS OF CHANGE IN MALE AGE-EARNINGS PROFILES

Table 4 presents estimates of the impact of the ratio of the number of younger workers to the number of older workers, the state of the business cycle, and a general trend on the curvature of male age-earnings profiles. The dependent variables in the calculations are the log of the income of 45–54-year-old men relative to the income of the 25–34-year-old men (lines 1–5) and the log of the income of 45–54-year-old men relative to the income of 20–24-year-old men (lines 6–8). Lines 1–2 and 6–7 refer to the incomes of year-round full-time workers. As noted earlier, these incomes provide a better measure of wages than do the incomes of all workers; unfortunately they are not available until 1955. Lines 3 and 8 treat the ratio of the mean income of all men with nonzero incomes in the relevant age groups over the period 1947–74. These figures provide additional time-series variation in the data at the expense of potential confounding of changes in utilization of labor with changes in wages paid. Lines 4 and 5 consider the age-earnings profiles of college graduates and high school graduates separately for the limited number of years for which such figures are available.

There are three independent variables. The changing age structure of the workforce is measured by the log of the ratio of the number of male workers aged 45–54 to the number aged 25–34 (lines 1–5) or to the number aged 20–24 (lines 6–8). To test the possibility that changes in relative earnings by age are dominated by cyclical rather than demographic factors, the calculations also include a measure of the state of the business cycle, the deviation of the log of real gross national product (GNP) from its trend level: positive deviations reflect a strong economy, while negative deviations are indicative of recession conditions. This measure of the cycle is highly correlated with such alternative business cycle indicators as the rate of unemployment or the difference between actual and potential gross national product (see [8]). Because of the operation of formal seniority systems and because older workers will have cumulated specific human capital while young workers will not, labor demand is likely to be more cyclical for young than for older workers, suggesting that the deviation of real GNP from trend will be negatively related to the ratio of income of older to younger workers. The third independent variable, a time trend, is introduced to control for any of a variety of factors (greater education of the young, gradual shifts in industrial structure) which might influence the demand for young as opposed to older male workers and thus the relevant income ratios.

The regression results suggest that, with trend and cyclical factors fixed, the relative number of older to younger workers has a significant impact on male age-earnings profiles. In the calculations focusing on the income of 45–54-year-olds relative to 25–34-year-olds, the measure of relative numbers by age obtains a significant negative coefficient, except among high school graduates (line 4). In line 1, which analyzes the incomes of year-round and full-time workers with the trend variable omitted, the log of the ratio of the number of 45–54-year-olds to the number of 25–34-year-old workers obtains a coefficient of −.27, three times its standard error. Inclusion of the trend in line 2 reduces the coefficient on relative numbers substantively, but still leaves a significant relation between relative numbers by age and relative incomes by age. In line 3, which treats the income of all men over the period 1947–74, the coefficient on the relative number is a highly significant −.20.

The regressions for the income of year-round full-time workers aged 45–54 relative to the income of year-round full-time workers aged 20–24 in lines 6 and 7 yield sizable significant coefficients for the relative number of workers by age, though in these cases addition of the time trend raises rather than reduces the estimated coefficients. In line 8, by contrast, the ratio of the number of 45–54-year-old workers to the number of 20–24-year-olds has no apparent impact on relative incomes. This is due largely to the trend term, which is highly collinear with the relative numbers variable over the period: with trend deleted, the coefficient on relative numbers in regressions explaining the ratio of the income of 45–54-year-olds to the income of 20–24-year-olds is −.31 with a standard error of .09.

The business cycle indicator, the deviation of real GNP from its trend, has little impact on the income ratios of year-round and full-time male workers in lines 1 and 2, suggesting that the wages of 25–34-year-old men are no more sensitive to the cycle than are those of 45–54-year-old men. In the regression for all male workers in line 3, however, the cyclical variable is accorded a sizable effect, which presumably reflects the greater cyclical sensitivity of the time worked of the younger men. Much stronger cyclical effects are obtained in the calculations for the income of 45–54-year-old men relative to the income of 20–24-year-olds in lines 6–8, indicating that cyclical ups-and-downs have their greatest impact on the incomes of 20–24-year-olds.

367

TABLE 4
DEMOGRAPHIC AND TREND DETERMINANTS OF THE MEAN INCOME OF OLDER TO YOUNGER MALE WORKERS, 1947–74

Dependent Variables (in ln units)	Coefficients and Standard Errors: Constant	Cycle[a]	Relative Number of Workers by Age	Time	R^2	Durbin-Watson Statistic
Income of 45–54-year-olds relative to the income of 25–34-year-olds						
1. Year-round and full-time workers, 1955–74	.04	.08 (.26)	−.27 (.09)		.34	0.33
2. Year-round and full-time workers, 1955–74	.00	.01 (.11)	−.14 (.04)	.005 (.001)	.89	1.79
3. All workers, 1947–74	.09	−.29 (.14)	−.20 (.05)	.003 (.001)	.54	1.70
4. Four-year college graduate workers, 1956–74[b]	.09	−.24 (.64)	−.51 (.28)	.006 (.005)	.33	b
5. High school graduate workers, 1956–74	.15	−.26 (.12)	−.07 (.11)	.003 (.003)	.48	b
Income of 45–54-year-olds relative to income of 20–24-year-olds						
6. Year-round and full-time workers, 1955–74	.60	−.72 (.24)	−.32 (.04)		.84	1.04
7. Year-round and full-time workers, 1955–74	.69	−.70 (.25)	−.41 (.19)	−.003 (.006)	.85	1.05
8. All workers, 1947–74	.59	−.71 (.26)	−.01 (.05)	.011 (.001)	.86	1.92

Source: Income data and numbers from *Current Population Reports* [17]. The figures of mean incomes for the period 1947–60 are from U.S. Bureau of the Census, *Trends in Income of Families and Persons in the United States: 1947–1960,* Technical Paper No. 7. Due to the absence of mean figures for the years 1961–64, mean incomes for those years were obtained by interpolation using least squares regression of the log of the mean incomes on the log of median incomes.

College incomes: *Current Population Reports* [17], No. 92, Table A, p. 2, with 1965–67 ungrouped data spliced for consistency with 1967–72 grouped data using the 1967 overlap year. The missing year 1965 was obtained by applying percentage changes in median incomes from 1964 to 1965. 1973 and 1974 are taken from Series P-60, Nos. 97 and 101. Because the 1965 income figures were published only for men with 4+ years of college, I estimated the income of 4-year graduates in 1956 by regressing income for 4-year graduates on income for graduates with 4 or more years of college, using all the years after 1956, and extrapolating with the equation.

a The measure of the cycle in lines 1–2 and 4–7 was obtained as the residual from the following regression of real GNP on time for the period 1955–74:

$$\log GNP = 8.32 + \underset{(.001)}{.039T} \qquad R^2 = .983$$

The measure of the cycle in lines 3 and 8 was obtained as the residual from the following regression of real GNP on time for the period 1947–74:

$$\log GNP = 8.02 + \underset{(.001)}{.037T} \qquad R^2 = .989$$

b The years 1957, 1959–61, and 1963 were omitted due to absence of data. Because of the omission, no Durbin-Watson statistic was calculated.

The regressions in lines 4 and 5, which focus on the age-earnings profiles of college graduates and high school graduates, respectively, suggest that the increased number of young male workers relative to the number of older male workers had an especially large impact on college graduates' profiles compared to only a modest impact on high school graduates' profiles. Similar results based on different models and data have also been obtained by other scholars. Using the Michigan Panel Survey on Income Dynamics, William Johnson found that the relative size of a cohort significantly reduced the impact of the earnings of college graduates but had little effect on the earnings of less educated workers [11, p. 13]. Using CPS Demographic files, Finis Welch [24] found a significant effect of cohort size on earnings in all school groups, with, however, much larger elasticities for college graduates. All of these results suggest that younger and older workers are better substitutes in the lower schooling groups than in the college graduate group. This is consistent with an explanation of the much heralded fall in the rate of return to investment in college education in terms of the increased supply of young college workers.

The regression estimates in Table 4 can be used to gauge the impact of relative income of young workers by multiplying the coefficients on the explanatory variables by the observed changes in the variables.[13] For year-round and full-time workers whose incomes are closest to the wages of concern, such an analysis attributes most of the change to demographic factors:

	45–54-year-olds/ 25–34-year-olds	45–54-year-olds/ 20–24-year-olds
Actual change in log income ratios, year-round and full-time workers, 1968–74	.06	.18
Change predicted by regression model and actual change in independent variables	.06	.15
Due to change in relative number of workers by age	.03	.11
Due to cyclical change	.00	.06
Due to trend	.03	−.02

Source: Calculated from regressions 2 and 7 of Table 4 and actual changes in explanatory variables.

According to these calculations, the regression model explains much of the

13 Let a_i be the estimated effect of variable S_i on relative earnings and let ΔX_i be the change in X_i in the period. Then the contribution of the change in X_i to the change in relative earnings is $\hat{a}_i \Delta X_i$.

1968–74 increased curvature of cross-sectional age-earnings profiles in terms of the increased number of younger workers relative to the number of older workers. Half of the .06 increase in the log of the ratio of the income of 45–54-year-old to the income of 25–34-year-old workers is attributed to the increase in the number of 25–34-year-old workers relative to the number of 45–54-year-old workers, while 60 percent (.11/.18) of the increase in the log of the income of 45–54-year-olds relative to the income of 20–24-year-olds is explained by the increased number of 20–24-year-olds relative to the number of 45–54-year-old workers. The trend variable accounts for the remainder of the change in the income ratio of 45–54 and 25–34-year-olds, while the cyclical variable explains most of the remaining change in the 45–54 to 20–24-year-olds income ratio.

In sum, the estimated effects of demographic, cyclical, and trend factors using relative income equations based on CES-type functional forms suggest that the observed twist in the age-earnings profile against young men can be attributed in large part to the changed age structure of the workforce.[14]

Translog System Estimates

The effect of changes in the ratio of the number of younger to the number of older male workers and of changes in the amount of other inputs on male age-earnings profiles can be analyzed with the translog system of derived demand equations (9)–(11). Such an analysis enables us to check on the robustness of the results of Table 4 under a different functional specification and to test two competing hypotheses about the causes of the observed twist in the male age-earnings profiles: (1) That the twist results from increases in the amount of capital, which shifts demand for labor toward older as opposed to younger workers. Increases in capital would have such an effect on demand if, as seems reasonable, capital is relatively more complementary (less substitutable) with older than with younger workers. One reason for expecting relatively greater complementarity is that skilled labor has been found to be relatively more complementary with capital than unskilled labor (Griliches [9]). Since older workers tend to be more skilled than younger

14 While the regression calculations "explain" the bulk of the variation in the income of older men relative to the income of younger men, it is important to recognize that some observations do not appear to fit the model. In particular, despite a relatively large number of younger workers in the late 1940s/early 1950s, the age-income profile was not as steeply sloped as the analysis would lead one to expect. This may be the result of the effect of the Great Depression and World War II on the employment experiences of workers or the result of the reduction in female participation at the end of the war. What is needed is a detailed study of the labor market for younger and older workers in that period to determine why relatively large numbers of young workers were "absorbed" into the workforce with little reduction in the earnings of young workers relative to older workers.

workers, demand for the former could be expected to be relatively more complementary with capital. Another is that older workers presumably have also accumulated greater specific human capital, linked to physical capital. (2) That the twist in the male age-earnings profiles is due to the influx of female workers, who are better substitutes for young male workers than for older male workers. The reason for expecting greater substitutability between women and younger men than between women and older men is that women are more likely to be competing for entry-level or early career jobs than for more senior positions in firms. If women are better substitutes for the young, then increased numbers of women workers would raise the earnings of older men relative to younger men.

A translog derived demand system was estimated with the number of full-time-equivalent female workers aged 20–64, the quantity of capital, and number of full-time-equivalent male workers aged 20–35, and the number of full-time-equivalent male workers aged 35–64 as inputs. Workers below the age of 20 were eliminated due to serious problems in estimating their earnings. Male workers were divided into two groups because of the problems of estimating translog systems with more than four inputs using a limited number of observations. The younger of the two groups, 20–34-year-olds, covers the 20–24 and 25–34-year-olds treated in earlier tables.

The inclusion of women and capital in the analysis creates considerable data problems due to inadequate information on the amount and rewards of these inputs. The problem for women is that published CPS figures have historically reported incomes and the number of persons with income rather than labor market earnings and workers with earnings. Since many women have income but do not work in the labor market, it is imperative to adjust the data to take account of this difference. Comparison of the mean *earnings* of women to the mean income of women in recent years when both figures were published in the *Current Population Reports* shows small differences (less than 1 percent in 1974, for example),[15] suggesting that the mean incomes provide reasonable estimates of mean earnings. Comparisons of the number of persons with income and with earnings, however, reveal a sizable differential. In 1974, for example, 38 percent more women were reported as having income than were reported as having earnings.[16] To take account of this problem and of the part-time work done by many women employees, the number of full-time-equivalent women working in the U.S.

15 In 1974 the mean income of all women was $4142. The mean earnings of women with no other income was $4101. See *Current Population Reports* [17], No. 101, Table 71, p. 145.

16 In 1974 there were 59.2 million women reporting income and 42.9 million reporting labor market earnings in the *Current Population Reports* [17], No. 101, Table 71, p. 145. These figures are similar to those in other years.

was estimated by adjusting data from the *Employment and Training Report of the President 1977* by estimates of weeks worked from the *Work Experience of the Population* reports of the BLS, as described in the table note. Comparable estimates of employment of men aged 20–34 and 35–64 were made from the same sources.

The problem with data on capital is that we lack good information on both the capital stock and user cost. An estimate of private capital was made from data on corporate and noncorporate investment in nonresidential business capital, provided by the U.S. Department of Commerce (Musgrave [14]) and unpublished data on government capital stock from the same source. The user cost of capital is based on methodology developed by Jorgenson and associates [12]. The particular measure used takes into account the differential tax treatment of equipment and structures and the presence of both equity and debt finance, and adjusts the cost of capital for depreciation and expected capital gains. It is based on a weighted average of the user cost of equipment and the user cost of structures, with weights taken as the share of each in total investment from the National Income and Product Accounts. The precise estimating procedure is described in detail in Clark and Freeman [6].

The share of each of the inputs in cost was estimated by dividing the total factor payments to each by the sum of factor payments. In the sample the mean share of national income going to men aged 35–64 was .39; the share going to men aged 20–24 was .18; the share going to women aged 20–64 was .12; and the share going to capital was .31. Though the data leave much to be desired, they provide at least a crude means of examining in the context of the translog specification the effect of the increased supply of women and capital on the age-earnings profiles of men.

Estimates of the constrained translog system were made with an iterative version of Zellner's minimum distance estimator. Given the constraints, one of the factor share equations is redundant and can be dropped without affecting results: the capital share equation was deleted.

Table 5 records the estimated parameters of the system. Table 6 presents the elasticities of complementarity and of factor prices with respect to changes in quantities derived from the parameters at the *mean* value of the shares. In all cases the elasticities of input prices with respect to the quantity of the same input are negative, as required by the model. All of the labor inputs are complementary with capital (i.e., have positive elasticities of complementarity). As hypothesized, however, older men are relatively more complementary with capital and have a larger elasticity of factor price with respect to capital than do younger men (the relevant S_{ij}s are 1.76 for men aged 35–64 and 1.47 for men aged 20–34; the relevant factor price elasticities are .55 for men aged 35–64 and .46 for men aged 20–34). Contrary to expectation, the number of women is estimated to have a slight

TABLE 5
ESTIMATES OF DERIVED DEMAND
EQUATIONS OF THE TRANSLOG COST SYSTEM, 1950–74[a]

	Share of Inputs			
	Male Workers 20–34	Male Workers 35+	Female Workers 20+	[Capital][b]
1. Constant	.173(.001)	.404(.003)	.112(.002)	.301
2. ln male workers 20–34	.080(.006)	−.088(.006)	−.018(.006)	.026(.009)
3. ln male workers 35+	−.088(.006)	.051(.011)	−.054(.007)	.092(.011)
4. ln female workers 20+	−.018(.006)	−.054(.007)	.021(.008)	.052(.012)
5. ln capital	.026(.009)	.092(.011)	.052(.012)	−.170(.020)
6. R^2	.767	.674	.861	—
7. SEE	.0035	.0110	.0008	—
8. D.W.	1.52	0.49	0.69	—

Source: Income share calculated using incomes of workers from Current Population Survey, *Consumer Income Reports,* various editions, and number of workers employed from *Employment and Training Report of the President* [20], Table A-14, p. 232. Because men and women workers are employed different numbers of hours over the year, the numbers used as explanatory variables were adjusted to reflect approximate hours worked. Data on weeks worked by part-time and full-time employees by sex and age were obtained from *Work Experience of the Population* [23], Table A-1a, p. A-9; full-time-equivalents were calculated by estimating the mean weeks worked by full-time workers and part-time workers, with the former assumed to work 40 hours per week and the latter 20 hours per week. Estimates were made for a single year 1969 for men aged 20–34 and 25–34, men aged 35–64, and women aged 20–64. The mean weeks for the groups were: men 20–24, 35; men 25–34 and 35–64, 48; women 20–64, 37. Accordingly, the number of female employees was adjusted downward by 37/48 and the number of men 20–24 by 35/48. The figures were obtained from U.S. Bureau of Labor Statistics, capital from Musgrave [14] and unpublished U.S. Department of Commerce data. The price of capital was obtained from Clark and Freeman [6], as described in the text. Estimates were made by LSQ part of *Time Series Processor* program.

a Numbers in parentheses are standard errors. Dependent variables are shares of inputs.

b Capital equation omitted; estimated values obtained from constraints.

positive effect on the earnings of young men compared to a slight negative effect on the earnings of older men. The magnitudes of those elasticities of complementarity and of the elasticities of factor prices between the female workforce and the two age groups of men are, however, sufficiently slight as to suggest that male and female workers operate in essentially separate production processes. Given the problems noted above in estimating the input of female labor, however, little weight ought to be given to the link between the female workforce and demand for other factors.

TABLE 6
ESTIMATES OF ELASTICITIES OF COMPLEMENTARITY AND OF THE ELASTICITY OF FACTOR PRICES TO CHANGES IN THE QUANTITY OF INPUTS, EVALUATED AT THE MEAN SHARES OF FACTORS IN COST

	Change in the Quantity of:			
	Men 20–34	Men 35–64	Women 20–64	Capital
The elasticity of complementarity S_{ij}, by group				
Men, aged 20–34	−.209			
Men, aged 35–64	−.25	−1.25		
Women, aged 20–64	.17	.15	−5.88	
Capital	1.47	1.76	2.40	−3.95
Elasticities of factor prices with respect to changes in quantities, by group				
Men, aged 20–24	−.38	−.10	.02	.46
Men, aged 35–64	−.05	−.49	−.02	.55
Women, aged 20–64	.03	−.06	−.71	.73
Capital	.26	.69	−.29	−1.23

Source: Elasticity of complementarities calculated from the estimates of γ_{ij} and γ_{ii} in Table 6 using the formula

$$S_{ij} = (1/\alpha_i\alpha_j)\gamma_{ij} + 1$$

$$S_{ii} = 1/\alpha_i^2(\gamma_{ii} + \alpha_i^2 - \alpha_i)$$

Elasticities of factor prices computed as $\alpha_j S_{ij}$ and $\alpha_i S_{ii}$.

The most important finding from the TL system regressions is that, consistent with our previous results, the earnings of men aged 20–34 depend critically on the number of young male workers. The own elasticity of complementarity for the number of young men is sizably negative and far in excess of the cross-elasticity between young and older men, implying that an increase in the number of young men would reduce their wage relative to the wage of older men. Most importantly, the estimated factor price elasticities indicate that changes in the numbers of male workers of different ages will substantively influence the earnings of younger and older men and thus are likely to alter male age-earnings profiles.

The quantitative contribution of changes in the various inputs on male age-earnings profiles in the period when the profiles changed sharply can be evaluated by substituting the estimated elasticities of factor prices with respect to quantities from Table 6 into the relative wage determination equation (5) derived from Section III and multiplying the resultant coefficients by the actual changes in the inputs. Let $\dot{L}_w$ = the log change in the input of female labor aged 20–64, and $\dot{K}$ = the log change in the capital input. Then plugging the factor price elasticities from the table into equation (5) yields:

$$(14) \qquad \dot{W}_{35\text{-}64} - \dot{W}_{25\text{-}34} = .33\dot{L}_{25\text{-}34} - .39\dot{L}_{35\text{-}64} - .04\dot{L}_w + .09\dot{K}$$

as the appropriate equation determining the earnings of older men relative to younger men. From 1968 to 1974, the four inputs changed as follows: $L_{25\text{-}34} = .24$; $\dot{L}_{35\text{-}64} = -.02$; $\dot{L}_w = .17$; $\dot{K} = .11$; while $\dot{W}_{35\text{-}64} - \dot{W}_{25\text{-}34} = .07$. This yields the following decomposition of the change in relative incomes:

Actual change ($\dot{W}_{35\text{-}64} - \dot{W}_{25\text{-}34}$), 1968–78	.07
Change predicted by translog model and actual change in independent variables	.09
Due to changed number of young men ($\dot{L}_{25-34}$)	.08
Due to changed number of older men ($\dot{L}_{35-64}$)	.01
Due to changed number of females ($\dot{L}_w$)	−.01
Due to changed stock of capital ($\dot{K}$)	.01

Source: Calculated using (14) and actual changes in explanatory variables.

According to these calculations, the increased supply of young men was *the* principal factor depressing their earnings relative to the earnings of older men from the late 1960s to the mid-1970s.[17]

V. CONCLUSION

The analysis in this paper has shown that from the late 1960s through the mid-1970s when the number of young workers increased rapidly, the earnings of young male workers fell relative to the earnings of older male workers, altering male age-earnings profiles, particularly for college graduates. Demand for labor equations based on the constant elasticity of substitution production function and on the translogarithmic production function suggest that the increased number of young male workers was the

17 Using a different data set, with different definitions of age groups and a different time period, Joseph Anderson [1] also has found that a large fraction of the change in the relative wages of the young is due to their increased numbers. However, he also attributes a large portion of the change to greater complementarity between capital and middle aged (25–54-year-old) workers than between capital and younger workers, those aged 14–24. While both Anderson's results and those in this study show a depressant effect of the increased relative number on the wages of workers, they differ in the estimated effect of capital, which may highlight the problems of measuring that variable. Differences in the age groups covered may also explain the moderately different results obtained between Anderson's data set and the one used here.

major causal force underlying the increased earnings of older men relative to the earnings of younger men. Alternative factors that might explain the observed twist in male age-earnings profiles—the business cycle, general trends, the increased supply of women, and growth of capital—were found to have much smaller effects on age-earnings profiles.

The late 1960s to mid-1970s twist in male age-earnings profiles raises important questions about the effect of the size of a cohort on the earnings of the cohort and about the economic determinants of cross-sectional differences in earnings by age. Will the relatively low earnings of the large young cohorts of the 1970s be maintained in the future? Will age-earnings profiles change in favor of the prospective small youth cohorts of the 1980s? Do demographic swings produce significant intergenerational income inequality? How important are differences in the relative size of cohorts in determining the shape of cross-sectional age-earnings profiles?

Arguments can be advanced for and against the likely permanence of a "cohort effect" on the earnings of the young (20–34-year-old) male workers of the 1970s.

Three basic forces would appear to operate to create permanent cohort effects. First is the tendency of many firms to promote individuals by seniority along well-defined job ladders and to raise pay through "standard" increases over starting rates. With this type of pay policy, a group that enters the market with low earnings compared to others will never catch up to the position it would have had had it entered with higher initial earnings. Second is the possibility that the large young worker cohort of the 1970s has been "pushed" into jobs with flatter longitudinal age-earnings profiles than are normally chosen by young workers (as may have occurred among college graduates, some of whom has been unable to obtain "college type" jobs). While some of the young may be able to switch to jobs with steeper earnings trajectories in the future, the overall rate of increase in pay may still fall short of what would have occurred for a smaller cohort that obtained more desirable jobs at the outset. Third, to the extent that future promotions and raises will be depressed by competition from a large pool of persons in the same age groups, the large youth cohorts of the 1970s may actually lose ground in the job market relative to other cohorts in the future.

On the other side of the spectrum, if the low initial earnings of persons in the 1970s are given a strict investment interpretation, they suggest that young workers are making greater investments in on-the-job training than in the past, presumably through selection of "learning jobs." This will show up in especially steep longitudinal profiles for this cohort in the future and thus in significant catching up. Perhaps more realistically, the likelihood that substitution among workers of different ages increases with age suggests at least some diminution in the effect of cohort size over time (Welch [24]).

Limited evidence on cohort earnings profiles in the past and on the jobs held by the new entrants of the 1970s suggests that some permanent effect of cohort size is likely. Ruggles and Ruggles's [15] analyses of the social security (LEED) data file reveal a noticeable drop in the longitudinal profile for the cohort that entered the labor force in about 1930, apparently because "the labor market conditions at the time of their entry had a significant depressing effect on their earnings relative to those who preceded them and were already established in the labor market," whereas the earnings of the cohort born a decade later were "high relative to the cohorts surrounding them and seem to have enjoyed this advantage continuously" [15, p. 124]. Chamberlain's [5] evidence on the return to schooling addresses directly the possibility that the relatively depressed position of young college graduates compared to high school graduates is a temporary phenomenon due to greater investments in on-the-job training. His estimates of the return to schooling at the "overtaking point" (when earnings are no longer depressed by investments) suggest a drop from 12 percent in 1969 to 7 percent in 1973. Finally, the marked decrease in the proportion of new college graduates (see Freeman [7, 8]) in professional and managerial jobs, where the profiles are traditionally the steepest, suggests that it will be difficult for young college workers to "catch up" significantly in the future. Overall, while some catch-up is likely, the large youth cohorts of the 1970s can be expected to continue to have lower earnings than the smaller cohorts that preceded them and the smaller cohorts that will enter the job market in the 1980s.

Developments in the 1980s should provide, at the least, an interesting "test" of alternative interpretations of the determinants of changes in male age-earnings profiles. If relative wages by age shift in favor of young workers when the number of young workers declines in the mid- to late-1980s, the importance of demographic factors as determinants of male age-earnings profiles will be given further support. If the large youth cohorts of the 1970s do not experience especially rapid longitudinal gains in earnings in the 1980s, the incompleteness of the standard interpretation of cross-section profiles in terms of investments in training will be further demonstrated. Opposite patterns will, of course, have contrary implications.

Finally, while the 1980s will provide further information concerning the importance of demographic and demand factors relative to investments in human capital as determinants of age-earnings profiles, the evidence presented in this study does suggest that greater consideration be given to demand for labor by age in analyzing male age-earnings profiles.

REFERENCES

1. Joseph Anderson. "Substitution Among Age Groups in the United States Labor Force." Williams College, December 1978.

2. E. Berndt and L. Christensen. "The Translog Function and the Substitution of Equipment, Structures and Labor in U.S. Manufacturing, 1929–1968." *Journal of Econometrics* 1 (March 1973): 81–114.
3. Hans Binswanger. "A Cost Function Approach to the Measurement of Elasticities of Factor Demand and Elasticities of Substitution." *American Journal of Agricultural Economics* 56 (May 1974): 377–86.
4. Alan Cartter. *Ph.D.'s and the Academic Labor Market.* New York: Carnegie Commission and McGraw-Hill, 1976.
5. G. Chamberlain. "Omitted Variable Bias in Panel Data: Estimating the Returns to Schooling." Harvard Institute of Economic Research Discussion Paper 618, May 1978.
6. Kim Clark and Richard Freeman. "How Elastic Is the Demand for Labor?" National Bureau of Economic Research Working Paper 309, 1979.
7. Richard Freeman. *The Overeducated American.* New York: Academic Press, 1976.
8. ———. "Changes in the Labor Market for Black Labor." Brookings Papers, Summer 1973. Reprinted in *Cases and Materials on Employment Discrimination Law,* ed. A. Smith. Indianapolis: Bobbs-Merrill, 1977.
9. Zvi Griliches. "Capital-Skill Complementarity." *Review of Economics and Statistics* 51 (November 1969): 465–68.
10. J. Hicks. "Elasticity of Substitution Once Again: Substitutes and Complements." *Oxford Economic Papers* 22 (November 1970): 289–96.
11. William R. Johnson. "Vintage Effects in the Earnings of White American Men." University of Virginia, October 1978.
12. Dale Jorgenson. "Capital Theory and Investment Behavior." *American* 372
Economic Review 53 (May 1963): 247–59.
13. N. Keyfitz. *Applied Mathematical Demography.* New York: John Wiley & Sons, 1977.
14. John Musgrave. "Fixed Nonresidential Business and Residential Capital in the United States, 1925–1975." *Survey of Current Business* 56 (June 1976).
15. Nancy D. Ruggles and Richard Ruggles. "The Anatomy of Earnings Behavior." In *The Distribution of Economic Well-Being,* ed. F. T. Juster. New York: National Bureau of Economic Research, 1977.
16. Ryuzo Sato and Tetsunori Koizumi, "On the Elasticities of Substitution and Complementarity." *Oxford Economic Papers* 25 (March 1973): 44–56.
17. U.S. Bureau of the Census. *Current Population Reports,* Consumer Income Series P-60, various editions.
18. ———. *Statistical Abstract of the United States 1977, Trends in Income of Families and Persons in the U.S. 1947–1960.* Washington: U.S. Government Printing Office, 1977.
19. U.S. Department of Health, Education, and Welfare. *The Condition of Education,* Vol. 3, Pt. 1. Washington: U.S. Government Printing Office, 1977.
20. U.S. Department of Labor. *Employment and Training Report of the President 1976.*
21. U.S. Department of Labor, Bureau of Labor Statistics. "Educational Attainment of Workers." Special Labor Force Reports, various editions.

22. ———. "Employment of High School Graduates and Dropouts." Special Labor Force Reports, various editions.
23. ———. *Work Experience of the Population*, 1969.
24. Finis Welch. "Effects of Cohort Size on Earnings: The Baby Boom Babies' Financial Bust." Cambridge University Conference on Income Distribution, 1978.

374

The Economic Journal n° 324 Dec. 1971 Vol. 81

ECONOMIC CONSEQUENCES OF RAPID POPULATION GROWTH[1]

I. Introduction

"Population" was a major concern of the early classical economists. But subsequently it was forgotten by the profession for over half a century. Today, when the consequences of population growth are so especially important for the Less Developed Countries (L.D.C.s), barely a dozen economists are writing articles on population issues.[2]

Described below are the principal conceptual findings of a small group of economists that has been working together during the past few years on various projects concerning interactions between population growth and economic development.[3] None of their conclusions has hitherto appeared in an economic journal.[4] The time has come to present to economists the more important conclusions of this team research as described in various TEMPO publications.[5]

These conclusions for Less Developed Countries (L.D.C.s) relate to (*a*) distinctions between size, growth and fertility of population; (*b*) the impact of fertility reduction on income per capita; and (*c*) the international consequences of fertility differentials among countries.[6]

II. Distinguishing Population Size, Growth and Fertility

It is necessary for economic analysts to distinguish among the economic incidence of population size, population growth, and population fertility.

Whether a country has a "large" or "small" absolute population usually should refer to the size of its total population (or labour force) relative

[1] The author is Manager of Economic Development Programs, TEMPO, General Electric's Center for Advanced Studies, located in Santa Barbara, California. He wishes to acknowledge the assistance of Richard A. Brown of the TEMPO staff. He also has an obvious indebtedness to present and former colleagues as listed below, in footnote 3.

[2] Including J. J. Spengler, James E. Meade, Goran Ohlin, Colin Clark, Henry Leibenstein, Ansley J. Coale and the author.

[3] Most of the analyses presented here were done under contract to the United States Agency for International Development. Colleagues have included Richard G. Zind, James P. Bennett, William E. McFarland, Donald J. O'Hara, Ross D. Eckert, Arthur S. DeVany, David N. Holmes and Richard A. Brown. However, the author is alone responsible for the views expressed here.

[4] Instead they have appeared in such non-economic journals as *Science* [3], the *Journal of Biosocial Sciences*, [6], and *Policy Sciences* [5].

[5] See [1, 2, 4, 12, 13, 14 and 15], all published by TEMPO, Santa Barbara, California, and available through it or U.S.A.I.D., Office of Population, in Washington, D.C.

[6] TEMPO studies in population have also concerned the incidence of zero population growth in the United States, hardly attainable before A.D. 2040, upon different industries and factor incomes [10].

to the availabilities of usable natural resources and/or produced domestic capital. This contrasts with population densities per square kilometer, which by themselves have little economic meaning. An economy with little capital per worker is likely to have a low level of consumption, and this situation will be worsened if " land " (natural resources) is also scarce. A population in this sense may be too large or small regardless of its rate of growth. In terms of available capital, Mauritius has a " large " population compared to that of Japan, for example.

Where an already " large " population is combined with rapid natural increase and high fertility, as in lands of ancient culture such as China and India, the demographic–economic situation is at its worst.

The economic danger of rapid population growth lies in the consequent inability of a country both to increase its stock of capital and to improve its state of art rapidly enough for its *per capita* income not to be less than it otherwise would be. If the rate of technological innovation cannot be forced, and is not advanced by faster population growth, a rapid proportionate growth in population can cause an actual reduction in income *per capita*. Rapid population growth inhibits an increase in capital per worker, especially if associated with high crude birth rates that make for a very young age distribution. This is regardless of population densities. Although Brazil has a low population density in terms of " land," its population growth rate appears uneconomically high for adequate capital accumulation, and more babies cannot usefully populate its " empty " lands.[1]

High fertility rates have the demographic effect of increasing the proportionate number of children.[2] A country with a crude birth rate of over 40/1000 a year is likely to have over 40% of its population under 15 years of age. Youngsters under 15 years of age are significant consumers but insignificant producers. Large families including many children, with consequently low incomes per family member, are comparatively poor contributors to domestic saving. Low savings *per capita* are associated with " young " populations and high fertility.[3]

[1] Many less developed countries (L.D.C.s) include large areas of unpeopled " empty " land that superficially seem to invite extra population for their development—as for instance the Amazon Basin of Brazil. Unfortunately for this easy analysis, high fertility means more babies born, not in the Amazon Basin but where their mothers are in São Paulo, Rio, Recife, etc. Even assuming that high fertility eventually causes a migration of adult workers into the Amazon Basin, they would have very low productivity without accompanying capital. And it is no accident that capital does not flow into this area, bringing people with it as during the rubber boom days, but instead is profitably invested elsewhere. The " empty " land argument for high fertility proves invalid on analysis for most L.D.C.s.

[2] Age distribution is far more sensitive to age-specific fertility rates than to age-specific mortality rates.

[3] A very young age distribution, resulting from high fertility rates, ordinarily reduces the absolute value of domestic savings and investment. This is because there are disproportionately few adults of working age, so that G.N.P. is less than otherwise. This lower output effect is not fully offset by the fact that a population with disproportionately many children usually consumes less from a given G.N.P., even allowing in less developed countries for increasing expenditures on schooling.

Public health measures have led to rather dramatic reductions in age specific death rates, especially during the past 25 to 35 years, in poor and backward countries. The "killing" epidemics such as cholera have been far more successfully controlled than "crippling" diseases such as bilharzia. Continued high fertility rates have meanwhile led to natural population increases of 2–3% a year that double population every 35 to 23 years respectively. Indeed, although often ignored, one of the characteristics most distinguishing backward from advanced countries is a high birth rate of over 35 per 1,000 a year (*e.g.*, Indonesia as contrasted with Japan). Among countries as within countries, high fertility seems to cause relative poverty, besides often being a consequence of it.

If a nation's population is to increase naturally at $X\%$ a year, it is better that this result from low rather than high crude birth and crude death rates. A 1% annual increase for example, resulting from birth and death rates of 45 and 35 per 1,000, will be associated with more brutish living than if it were the outcome respectively of rates of 25 and 15 per 1,000 a year. In the latter case the ratio of children to work age adults will be lower, investment from domestic savings should be absolutely greater, and the income per equivalent consumer will be greater.[1]

In human terms, and perhaps far more important, low birth and death rates mean that there are fewer unwanted infants born too soon and fewer premature deaths. "Balanced" public health programmes that include birth as well as death "control" could give each family a little more security. Perhaps the true essence of economic development is that it gives families and individuals more command over their lives.[2]

III. Economic Development Through Reducing Fertility

The economic development of L.D.C.s has many facets, but most of these, such as levels of education and health, availability of capital per worker and adequacy of infrastructure, tend to be associated with output and hence also income *per capita*. Thus higher ratios of G.N.P. to population can usually serve as a surrogate for economic improvement. Moreover, although

[1] The "equivalent consumer" concept accounts for the fact that relative consumption varies with age and sex. Thus children typically consume less private and public sector goods and services than do adult males of working age. Hence, the increase in output (or income) per head that ordinarily follows in an L.D.C. from a reduction in fertility somewhat overstates the improvement in individual welfare, simply because there are now relatively fewer children and more adults. One solution is over time to divide G.N.P. not by absolute population but by the estimated number of equivalent consumers. (See [6], pp. 48, 49.)

[2] It is sometimes erroneously supposed that, because income *per capita* can arithmetically be increased by having a smaller population, public health programmes should concentrate more on preventing births and less on postponing deaths. However, families are more likely to save, invest and innovate, making and following plans for their own financial advancement, if uncertainties regarding deaths can be reduced. In circumstances of frequent, unpredictable and premature deaths in a family, planning and executing courses of action are inhibited, except that of having more births to replace deaths.

governments have customarily concentrated on accelerating the G.N.P. growth numerator, an increasing number of L.D.C.s are now also concerned with slowing the population growth denominator.

Rationally, if a government intends to spend $\$X$ over, say, 10 years to increase output per head, it should estimate whether it could achieve a greater increase in this ratio through expenditures on birth control than on investments in physical capital. Certainly, where the marginal product of labour approaches zero, a reduction in births for, say, a decade must raise *per capita* income later. In this case, expenditures for contraception must be many times more effective per dollar in raising *per capita* income than expenditures for plant and equipment.[1]

Where labour has a very low marginal product relative to that of capital, which is reputedly the case in most backward as compared with most advanced countries, practically all economic-demographic models indicate that a gradual halving of fertility over several decades raises income per head substantially. The loss of labour force after 15 years is more than offset by the more immediate increase in *per capita* income and in aggregate saving. After 15 years there is less labour but more investment than otherwise, with more capital and output per worker, a lower underemployment rate, and fewer consumers to share in a G.N.P. that has grown about as rapidly as it would have done with unchanged fertility.

This has been shown in several analyses, both for an abstract country called Developa, and for Guatemala, Turkey and most recently Chile.[2]

The main elements of the dynamic model used are:

V/P, Gross domestic product per head, which by wiser public and/or private actions one hopes to see rising faster than it would otherwise;

B, births, which depend on initial and changing age distributions and on projected age-specific fertility rates;

D, deaths, which depend on projected age-specific mortalities, age distributions, etc.;

P, population, which is arithmetically last year's population plus B minus D; [3]

[1] Where the marginal product of labour is zero, or when the analysis is for a 10-year period during which prevented babies do not reach 15 years to become lost workers, a reduction in births cannot reduce G.N.P. and must raise income per head above what it otherwise would be. This is also because the cost of the contraceptives needed to prevent a birth is so much less than the discounted cost of the investment otherwise needed to provide a typical annual flow of goods and services to an extra person. Such a benefit/cost comparison is repugnant to some, but no less valid on that account. See also [9].

[2] See [11, 14 and 15].

[3] International migration is here ignored. For countries with the worst population densities and growths, there is almost no in-migration and net out-migration is trivial in percentage terms. Emigration can seldom provide relief for population pressure either. A 1% emigration from India would mean 5 million people (net) moving permanently abroad each year. In itself this would constitute a major transportation job, greater than current air travel across the North Atlantic. More seriously perhaps, there are not enough countries willing and able to receive such a flow of

V, gross domestic product, which is a function of employed labour force, domestic capital stock and state of technology;
K, stock of capital, which increases according to aggregate domestic saving plus capital borrowings from abroad; [1]
S, aggregate domestic saving, which is positively related to V and negatively to P;
E, employed labour force, positively related to L and K, with E/L being monotonically related to K/E;
L, available labour force, a function of population size and age distribution;
T, technology, with improvements in the state of art resulting in more V from a given E and K.

In the models usually employed to date, V has been determined by a modified Cobb–Douglas function, in which the output elasticities of E and K have sometimes summed to less than unity to reflect scarcity of natural resources. The influence of technology has been compounded at a fixed annual rate and is incorporated in the aggregate production function as a shift factor.[2] Births and deaths have influenced V indirectly through L, K and hence E, while affecting P directly.[3]

The models have been used to examine *contrasting* fertility projections upon projected V and P, and hence V/P, over the next 35 years or so. Calculations are year by year. Results are usually tabulated at 5-year intervals.

In Table I, for an abstract country named Developa, a constant G.R.R. (gross reproduction rate) of 3·025 is contrasted with the case of a G.R.R. that falls by arithmetic retrogression from 3·025 to 1·479 over 25 years.[4] Subsequent V is hardly affected, the decline in fertility raising K enough to compensate for the fall in L. This increase in K results from a " release " of consumption, part of which is additional saving, because the number of

people. The few small countries that do receive a considerable population growth from immigration—over 5% in Kuwait's case—are atypical. The migration that does affect economic development is the internal flow to city from countryside. These effects are now being incorporated into a two sector economic-demographic model being developed at TEMPO.

[1] The basic model assumes no net international transfers of capital. It is programmed however to allow for any year to year exogenous capital movements that the analyst cares to assume. A modified programme also provides for enough inflow of capital in each year to maintain a stipulated constant annual improvement in V/P.

[2] This reflects an agnostic uncertainty as to whether technology is especially associated with, say, increased capital stock, improved worker education, or general level of welfare. In effect " technology " here is the residual source of all increased output that cannot be attributed to capital or labour increments. The main impact of a high rate of " technology " improvement in this model is to reduce the comparative importance of reducing fertility in raising projected V/P.

[3] In this formulation the demographic " side " of the model affects the economic side, but not conversely. Conceptually it could be supposed for instance that a rising V/P after some lag would reduce age specific fertilities. However, while this relation is often asserted, it has still to be demonstrated.

[4] The gross reproduction rate is the number of live female births a typical woman will have during her child-bearing years.

379

children declines relatively. There is less unemployment, and more output per worker, because the K/E ratio is higher. With similar V, but a smaller-than-otherwise P, V/P is higher.[1]

Specifically, taking the case summarised in Table I, V/P rises from \$200 a year in 1970 to \$419 with low fertility as against \$293 with high fertility by A.D. 2000. After thirty years, with low fertility the capital stock is larger

TABLE I

Effects of Declining Fertility on Output and Per Capita Income in " Developa "

Item.	1970	1985 High fertility.	1985 Low fertility.	2000 High fertility.	2000 Low fertility.
P, Population (10^6)	10·0	15·9	14·4	25·7	18·8
V, Output (\$$10^9$)	2·00	3·54	3·63	7·53	7·87
V/P, Income per head (\$)	200	223	251	293	419
L, Available labour (10^6)	3·61	5·69	5·69	9·10	8·32
Unemployment rate (%)	15	18	16	10	6
K, Capital stock (\$$10^9$)	5·00	7·06	7·28	13·20	15·57
K/E, Capital per worker (\$)	1,626	1,509	1,521	1,637	1,984
S/V, Savings from income (%)	4·7	6·4	7·9	9·6	12·5
Earnings per worker (\$)	325	378	378	459	501
Return on capital (%)	16	20	20	23	20
Children/Population (%)	44	44	39	45	32
$G.R.R.$, gross reproduction rate	3·025	3·025	2·092	3·025	1·479
Female life expectancy (years)	55·0	58·0	58·0	61·0	61·0

N.B. These numerical results, employing the initial conditions of 1970 and the economic parameters listed in the text, were developed from the TEMPO demographic-economic model, described in detail in Reference [1, 12].

(\$15·57 billion as against \$13·20 billion), capital per worker is higher(\$1,984 as against \$1,637), and the unemployment rate is lower (6% as against 10%). The relative scarcity of labour has increased after 30 years, with annual earnings per full-time equivalent worker of \$501 as against \$459, while the return on capital is 3 percentile points lower than with high fertility. One basic reason for better economic performance is that by A.D. 2000 children are 32% of the population with low fertility as compared with 45% with unchanged high fertility.

The outcome of a higher-than-otherwise V/P with declining fertility has been shown to be most insensitive to labour and capital output elasticities, technology improvement rates, the savings equation, the employment of labour function, or the projected exogenous decline in mortality rates. For each single comparison of projected fertility differences, these assumptions, and of course the initial conditions of population size, age distribution and

[1] The tables are for a case where technology improves 0·015 a year compounded, the output elasticities of employment and capital are 0·5 and 0·4 respectively, and aggregate domestic savings are 0·8V-\$35P.

capital stock, are always similar. In every case, if fertility is declining faster, V/P is nevertheless rising faster.[1]

A related point is that a lower fertility means slower population growth and hence more time for domestic capital to accumulate and the state of art to improve. This is shown in Table II, based on the same case as Table I,

TABLE II

Contrasting Attainment of Same Population in Different Years: Unfavourable Economic Consequences of Fast Population Growth in "Developa"

Item.	1985 (High fertility).	1990 (Low fertility).
P, Population (10^6)	15·9	15·9
V, Output ($\$10^9$)	3·54	4·68
V/P, Income per head ($)	223	295
L, Available labour (10^6)	5·69	6·57
Unemployment rate (%)	18	18
K, Capital stock ($\$10^9$)	7·06	8·98
K/E, Capital per worker ($)	1,509	1,543
S/V, Savings from income (%)	6·4	9·4
Earnings per worker ($)	378	402
Return on capital (%)	20	21
Gross reproduction rate	3·025	1·817
Female life expectancy (years)	58·0	59·0

Source: Same as Table I.

the difference being that Developa attains a population of 15·9 million in A.D. 1985 with unchanged fertility but only in A.D. 1990 with declining fertility. By waiting 5 more years for its population of 15·9 million, Developa can provide this size of population with a yearly V/P of $295 instead of $223, having a larger labour force (6·57 million as against 5·69 million) *and* a larger capital stock ($8·98 billion as against $7·06 billion). The argument is not that Developa should never have a much larger population. It is rather that population growth must be slow, regardless of "empty" lands waiting to be populated.

Dynamic models of this kind can also be used to sense the "return"

[1] See [13] for a detailed account of the sensitivity analysis. However, some feeling for the insensitivity of the main conclusion of the analysis, namely that a more rapid decline in fertility hardly affects V while significantly lowering P below what it would otherwise be, can be gained from the following ratios given in the cited document (which used slightly different parameters for its production function.) In that case, by A.D. 2000, the "low fertility P" was 27% below the "high fertility P" in all the calculations, and for the standard set of parameters, the ratio of "high fertility V" to "low fertility V" was 1·005. If the rate of annual technological improvement was 2·5% instead of 1·5%, this ratio was 1·022. If the savings function were not $0{\cdot}2V$-$\$30P$ but rather $0{\cdot}07V$, this ratio was 1·036. If the respective output elasticities of capital and labour were not 0·4 and 0·6 but rather 0·5 and 0·5, this ratio was not 1·005 but 0·988. The results are also very insensitive to economies or diseconomies of scale: thus, if the output elasticities sum not to unity but to 0·7 and 1·25 respectively, the A.D. 2000 ratios of V were not 1·005 but respectively 0·987 and 1·016. Altogether, the interactions of the model seem to dampen the effect of altering assumed parameters and initial conditions, a state of affairs that strengthens the credibility of the conclusions.

from " investments " in contraception. This of course requires an assumption as to the annual cost per effective contraceptive user and the age distribution of these voluntary " acceptors." The " return " or gain can reasonably be defined as the gain in income per head times the population enjoying it. Over a period of 35 years this benefit-to-cost ratio ranges between 50 and 150 to 1.[1] But for any historic period this arithmetic ratio must understate the return, especially for shorter periods, for even with no future costs there will always be future and generally increasing gains from past contraceptive expenditures.[2]

In these cases initial " size " of population is of minor importance. If natural resources are very scarce, so that there are markedly diminishing returns to labour and capital taken together, it is true that reduced labour employment because of reduced fertility occasions a smaller loss in V attributable to E. But by the same token the increase in V attributable to more K with fewer births is also smaller.

More important than size of population is the changing rate of its growth. An increasing growth rate, especially when due to a declining death rate, is economically disastrous. A decreasing population growth rate, because of declining fertility rates, is a major source of economic development.[3]

IV. International Consequences of Fertility Differences

High fertility rates tend to limit what L.D.C.s can export and in addition make them less creditworthy as international borrowers.

International trade theory has always emphasised that what countries export and import depends largely on relative factor prices. Countries with high fertility rates have a comparative advantage in labour intensive products, because comparatively their labour's marginal productivity is low and their capital's marginal productivity is high. However, because high fertility countries have low *per capita* incomes, there are additional trade consequences.

High-fertility, low-income countries generally export primary agricultural commodities, except for those few and fortunate nations possessed of

[1] In the Table I case this benefit/cost ratio is 116 to 1 by A.D. 2000. The annual cost of practicing contraception is assumed to be \$5 a year per user and effectiveness is supposed to be 0·8. The distribution of users by age is proportionate to the reduction in age specific fertilities assumed for the lower fertility case in the comparison. Thus the benefit/cost ratios obtained from these dynamic economic-demographic models are very similar in magnitude to the benefit/cost ratios, estimated in 1966 by far simpler and static means and published in this Journal. [9]

[2] See References [6] and [8].

[3] Of course an L.D.C. can instead raise its annual rate of *per capita* income improvement by faster innovating or saving. It is simple to calculate for any given case what these " trade-offs " are among fertility, savings, and innovating rate changes. (See Reference [6].) But calculating arithmetical equivalences in terms of V/P improvement does not create operational alternatives. Families do not save or innovate more *because* they increase their fertility. In fact a more plausible argument might be that the sort of families which practice birth control effectively are likely to be exceptional savers and innovators.

valuable mineral assets such as petroleum. Poverty makes for a consumption pattern of basic " necessaries," and hence a domestic production pattern of limited variety. Poor countries cannot afford the technological education and do not have the high-income market needed for products that are technically advanced, superior in performance or which incorporate high styling. With less capital, the workers of these countries must compete for foreign exchange largely through muscle power, usually applied to exporting sugar, coffee and other products of tropical agriculture. Such a worsening of the terms of trade between backward and advanced countries as may in fact have occurred [1] is probably as much due to continued fertility differentials among nations as to any other single cause. Those L.D.C.s that are too much dependent on agricultural exports to finance adequate industrial product imports should usually blame their own excessive fertility.

The same high fertility rates that increase the " need " for assistance of L.D.C.s also make them less creditworthy as borrowers and hence more dependent on grants. Aspirations for G.N.P. increase are often several percentile points higher than is realistic because of the typically expected 3% annual growth in population. This in turn " requires " a larger yearly increment in capital stock. But the low ratio of work age population to children caused by high fertility reduces output and aggregate savings for investment. Unfortunately, the very inability to save that " requires " external borrowing also makes subsequent repayment difficult or impossible. Savings are after all the ultimate source of repayment.

These interactions can also be explored by the TEMPO economic-demographic model. A " required " annual improvement in *per capita* income can be stipulated. The computer can be programmed to assume an inflow or outflow of capital in each year depending on whether domestic savings are respectively insufficient or excessive to occasion precisely the stipulated improvement in *per capita* income. Ordinarily, if an L.D.C. stipulates an unrealistic annual improvement, it will never be able to repay its borrowings with interest. Alternatively, an L.D.C. with a lower fertility may be able to realise a higher annual improvement in *per capita* income, *and* eventually service all borrowings from abroad, than can an L.D.C. with higher fertility.

Considering Developa again, Table III indicates some borrowing and repayment consequences of aspiring to alternative constant improvements in *per capita* income, contrasting again the two economic projections based on the same high and low fertility projections. Thus the highest sustainable annual improvement in *per capita* income is under 2·0% (actually 1·8%) with high (unchanged) fertility and over 2·5% (actually 2·9%) with low fertility if external borrowings are ever to be repaid (including a 5% annual

[1] It is by no means clear that the barter terms of trade have generally worsened for backward countries when unquantified improvements in industrial products exported from advanced countries are taken into account.

interest rate on the outstanding balance). Even at 1·5% yearly improvement, high-fertility Developa does not commence repayment before 30 years, but a low-fertility Developa can aspire to a 2·5% annual improvement and start repayment after 21 years. Alternatively, for a 1·5% sustained annual improvement in V/P, completed debt service takes 42 and 16 years

TABLE III

Effects of Fertility on Developa's Ability to Service Debts while realising a Stipulated Annual Improvement in Income Per Capita.

Stipulated annual improvement in G.N.P. *per capita*.	Future year when repayment of principal begins. (*b*)		Future year when loans are completely repaid.		Debt outstanding when repayment begins (*c*) (millions of dollars).	
	Fertility High	Fertility Low	Fertility High	Fertility Low	Fertility High	Fertility Low
1·0	14	5	23	10	482	130
1·5	30	11	42	16	2643	415
2·0	(a)	14	(a)	23	(a)	1066
2·5	(a)	21	(a)	33	(a)	2743

(*a*) Debt never repaid.
(*b*) Repayment begins in the year that domestic saving first becomes greater than is required to realise the stipulated annual improvement in *per capita* income.
(*c*) To be compared with initial year G.N.P. of $2,000 million.
Source: Same as Table I, plus Reference [7].

respectively with high and low fertility, with a respective maximum outstanding debt of $2,643 and $415 million.[1]

The credit unworthiness associated with high fertility may give international lending agencies a powerful and more acceptable means of inducing certain L.D.C.s to undertake vigorous birth control programmes. It may be politically impossible for, say, the World Bank Group to make a loan for infrastructure conditional upon the borrowing government's promoting contraception. But development assistance agencies when making loans are certainly entitled to consider all factors influencing a borrowing L.D.C.'s ability or inability to repay. It is, after all, a normal practice of borrowers to accept loan conditions. One problem throughout the 'Fifties and early 'Sixties was that international lending agencies were not prepared to do anything about perhaps the most important single cause of poverty in L.D.C.s—excessive population growth.

[1] Table III is based on the assumption that each year Developa borrows exactly enough from abroad to increase its domestic investment sufficiently to maintain the stipulated X% annual improvement in V/P. Alternatively, if domestic saving is more than enough to maintain this X% improvement, the excess domestic saving is used to repay international borrowings. The external liability for debt service includes a 5% interest charge on the current year's outstanding debt. See Reference [7].

384

V. Comment

Studies of political economy cannot logically fail to study people and hence populations. Economic development largely concerns the development of people, which means investments in education and health as well as in physical capital, both of which are encouraged when lower birth rates make labour more scarce relative to capital. One of the distinguishing and surely significant characteristics of Less Developed Countries is their high fertility rates and their consequently high proportion of unproductive children. Reductions in fertility enable domestic capital to be accumulated more rapidly. Greater future savings of L.D.C.'s because of effective birth control programmes should render them more credit-worthy. The poverty induced by high fertility also affects commodity flows and terms of trade. The influences of population growth are so pervasive throughout all macro-economic relations that they should surely become a major concern of the economics profession.

Stephen Enke

General Electric Company-TEMPO
Santa Barbara, California.

References

1. A. DeVany and S. Enke, *Population Growth and Economic Development: Background and Guide*, Santa Barbara, TEMPO, 1968, No. 119.
2. R. Eckert and D. O'Hara, *Manual for the Calculation of Government Expenditures for Selected Social Services*, Santa Barbara, TEMPO, 1968, No. 121.
3. S. Enke, "Birth Control for Economic Development," *Science*, May 1969, No. 164, pp. 798–802.
4. S. Enke, *Economic Benefits of Slowing Population Growth: Charts and Notes*, Santa Barbara, TEMPO, 1968, No. 122.
5. S. Enke, "The Economics of Having Children," *Policy Sciences*, June 1970, No. 1, Vol. 1, pp. 15–29.
6. S. Enke and R. G. Zind, "Effects of Fewer Births on Average Income," *Journal of Biosocial Sciences*, January 1969, Vol. 1, pp. 41–55.
7. S. Enke, "High Fertility Impairs Credit Worthiness of Developing Nations," (*Festschrift* in honour of Professor Edgar Hoover, title not chosen) (Gordon and Breach, 1971).
8. S. Enke, "Politico-Economic Global Systems," *Macrosystems: Analysis, Instrumentation, and Synthesis of Complicated Systems*, Spring, 1971, Holt, (Reinhart and Winston).
9. S. Enke, "Some Aspects of Slowing Population Growth," Economic Journal, March 1966, No. 76, pp. 44–56.
10. S. Enke, *Zero US Population Growth—When, How and Why*, Santa Barbara, TEMPO, 1970, No. 35.

11. Bruce Herrick and R. Moran, " Economic Effects of Chilean Fertility Decline,' Santiago, Chile, Centro de Investigaciones Economicas, Universidad Catolica de Chile (to be published).
12. W. E. McFarland, *Description of the Economic-Demographic Model*, Santa Barbara, TEMPO, 1968, No. 120.
13. W. E. McFarland, *Sensitivity Analysis of the Economic-Demographic Model*, Santa Barbara, TEMPO, 1969, No. 52.
14. W. E. McFarland and D. O'Hara, *Guatemala: The Effects of Declining Fertility* Santa Barbara, TEMPO, 1969, No. 50, Vol. II.
15. W. E. McFarland and D. O'Hara, *Turkey: The Effects of Falling Fertility* Santa Barbara, TEMPO, 1969, No. 50, Vol. I.

386

The Economic Journal June 1974 Vol. 84

REDUCING FERTILITY TO ACCELERATE DEVELOPMENT[1]

I. Introduction

This article describes a disaggregated economic–demographic model, including a government sector, that estimates the impacts of publicly-financed fertility reduction programmes on various indices of development over a 30-year time period. The new model is a substantial evolution from one previously described in this Journal.[2] Unlike the earlier and highly aggregated model, it can be used, by officials of a national planning office, for instance, to estimate the development consequences of reallocating government funds among education, health, social overhead capital, direct public investment, welfare (including other transfers), defence, other general administration, and family planning programmes. Development planners can, of course, use this simulation model whether or not they fund programmes to reduce fertility. The hope, however, is that they will on occasion use it to estimate the results of diverting funds to family planning from education, health, etc. They will find that, in addition to a more rapid improvement in *per capita* output and income, they can thereby *increase* the number of school children, hospital beds, etc., *relative* to population. Using the model, with its time horizon of 30 years or more, means that fertility and mortality changes can be taken into account as *policy* variables in a way that they can not in traditional development plans. The governments of developing countries also need some such model to provide a longer term framework for their usual five year plans.

II. Comparison of Present and Initial Models

The present model (TEMPO II) has different purposes and uses from the earlier model (TEMPO I).

The sole purpose of TEMPO I was to demonstrate that a reduction in fertility would result in higher outputs (and hence incomes) per head of population in future years than would occur if fertility rates remained unchanged. This almost inevitable result, insensitive to reasonable variations

[1] The author is Consulting Economist to TEMPO, the General Electric Company's Center for Advanced Studies, located at 777 14th St. NW, Washington, D.C. 20005. He wishes to acknowledge the contributions of various present and past members of his staff including Dr. William E. McFarland, Dr. Douglas L. Maxwell, Professor Donald J. O'Hara, and Mr. Richard Brown, and also the help of the editor and two referees. Much of the research described in this article was financed by the U.S. Agency for International Development.

[2] Enke, Stephen, "Economic Consequences of Rapid Population Growth," Economic Journal, December 1971.

in parameter values, is readily explainable. Primarily, a reduction in fertility alters a population's age distribution, so that the ratio of work-age to non-work age populations increases. Increased income and savings *per capita* result. G.D.P.s, even *after* 15 years, are only slightly affected, an increase in capital stock largely offsetting a decrease in labour force. But this practically unaffected G.D.P. is of course shared among fewer people when there has been a fertility reduction. Moreover, *while* fertility is declining, there is an additional reason for extra improvements in output per head: prevented children are not lost to the labour force until after 15 years or so, and meanwhile their absence means increased saving and investment per head.[1] At least these are the outcomes of specific country applications to date.[2]

The earlier model, although it served its purpose, was criticised on various grounds. It could not have been more aggregative. There were no sectors, such as rural and urban, modern and traditional, domestic and international, or private and public. Alternative fertility projections could not be influenced within the model by government budget decisions but were instead stipulated. There was no role for government planning officials in the model because it distinguished no government, no budget to be re-allocated, and no family planning programme to reduce fertility.

The subsequent TEMPO II model, of which there are several modifications already, was designed for developing country officials employed perhaps in a national planning commission. They can with the model trace the long-run consequences of reallocating the budget among the eight programmes already listed (including family planning). The model also enables official users to make decisions as to the way resources come under governmental command, alternative sources being taxes, borrowing private savings, foreign assistance, and lastly inflation. The present model, besides distinguishing between government and private activities, includes a private *subsistence* sector and a private *modern* sector. Within the labour force of the private modern sector a distinction is made between educated and uneducated labour. There is however no foreign sector as such, although most consequences of external official assistance, foreign private investment, or immigration, can readily be incorporated. In other respects, despite the

[1] Even if the saving rate remains constant, savings *per capita* will rise; actually, the savings rate is expected to rise somewhat, consistent with fewer children per adult.

[2] TEMPO I has so far been applied to 17 countries, sometimes by TEMPO, sometimes by local governments, sometimes by private local researchers; countries of application now include Bolivia, Brazil, Chile, Colombia, Costa Rica, Guatemala, India, Indonesia, Jamaica, Mexico, Nepal, Nigeria, Peru, Taiwan, Tanganyika, Turkey and Venezuela. In these applications, country specific parameters have been estimated whenever feasible, notably in Chile. In other cases plausible values have been used. Ordinarily, age specific mortality and fertility rates can be derived from either existing data or by using crude birth and death rates with standard life tables of the kind generated by Coale and Demeny at Princeton. Savings functions can only be guesstimated. Capital and labour elasticities of output have to be based on impressions as to the share of G.D.P. of different productive factors.

pleas of some potential users for more " realism," the TEMPO II model has been designed to avoid all complications that are not considered essential to the purpose at hand.[1]

III. Description of Model

Fig. 1 is a simplified flow chart of the TEMPO II model. Basically, the model includes a private *modern* economy and a private *subsistence* economy, each with its own demographic and economic equations and initial conditions. There is assumed to be migration between these two private sectors. The modern sector's output is generated by a modified Cobb-Douglas type production function that includes, in addition to a shift factor representing technological advances, three factors of production in the form of capital, uneducated labour, and educated labour. There is also a government sector that taxes, borrows, and inflates, and which allocates resources thus controlled among education, health, social overhead capital, direct public investment, welfare (including other transfers), defence, other general administration, and family planning programmes.

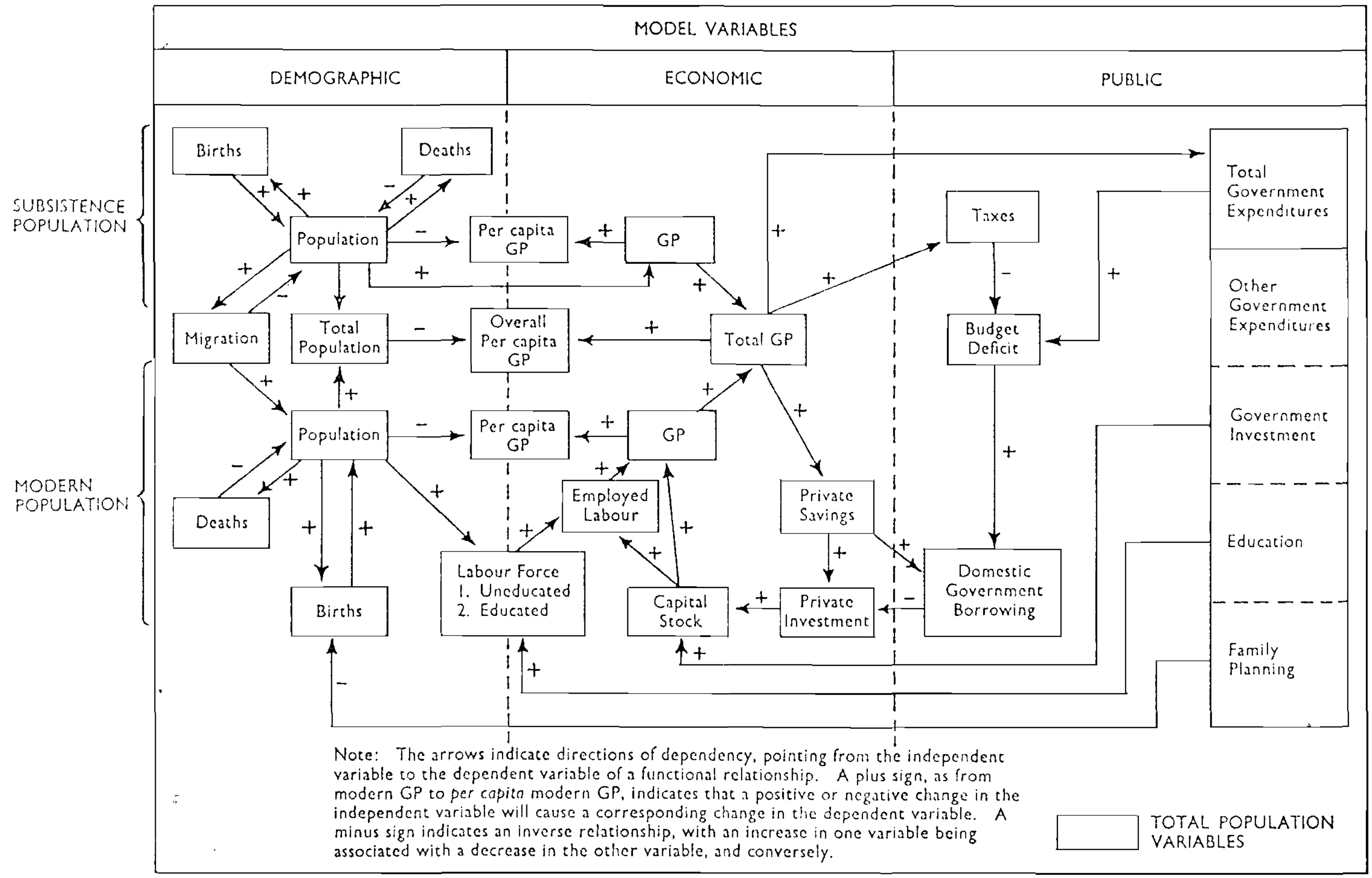
The Modern Sector

The modern sector has a population that is growing from natural increase and from net in-migration from the subsistence sector. International migration can be taken into account. Internal in-migration is assumed to vary with the ratio of *per capita* uneducated labour earnings in the modern sector to subsistence sector *per capita* income from all sources. This implies an essentially economic reason for migration, and the ratio should in theory allow for differences in employment and costs of living between the two sectors. (One alternative would have been to use a behavioural–empirical rule for migration, based in part on age and sex.)

The total output of the modern economy is assumed to depend upon the application of educated and uneducated labour, capital, and advancing technology. Its gross investment is the sum of private investment (which can include private lending from abroad) and government investment (which includes direct productive investment and social overhead capital). Net investment is total investment minus depreciation on the total capital stock in the modern sector.

The modern sector's labour force is divided into two kinds of labour on the basis of education—" uneducated " labour (less than 6 years schooling) and " educated " labour (6 or more years schooling). Each kind of labour is treated as a distinct factor of production. The school system, a major

[1] There is a tendency among some officials to place credence on the results of a model only if it incorporates a great deal of real world detail. Much of this detail adds almost nothing however to the ability of such a model to simulate the effects of official programmes to reduce fertility. Not only does such detail increase programming and running time exponentially, but it makes parameter estimation more difficult, costly, and often unreliable.



Note: The arrows indicate directions of dependency, pointing from the independent variable to the dependent variable of a functional relationship. A plus sign, as from modern GP to *per capita* modern GP, indicates that a positive or negative change in the independent variable will cause a corresponding change in the dependent variable. A minus sign indicates an inverse relationship, with an increase in one variable being associated with a decrease in the other variable, and conversely.

Fig. 1

object of expenditure by government, converts uneducated labour into more productive educated labour.[1]

The employment rate among the modern sector's labour force increases with the stock of capital and decreases with the size of the labour force itself. As the capital stock to labour stock ratio increases, both the employment rate and the capital-to-employed labour force ratio increase. Thus there is simultaneous capital widening and deepening.

The private savings that add to the modern economy's stock of capital come from the disposable income remaining in the sector after government taxation. Private saving is assumed to vary positively with such disposable income and negatively with the sector's population.[2] It is assumed that the government can borrow up to one-fourth of private savings, but at the cost of reduced private investment, so that government borrowing for non-investment purposes reduces G.D.P.

The Subsistence Sector

The subsistence sector is primitive. Capital accumulates from family savings and is invested as more livestock, new terracing, enlarged irrigation ditches, etc. Because capital is associated so directly with each family, it is not treated as a separate factor of production in the subsistence sector but is assumed to grow proportionally with the subsistence population. There is no inter-sector capital flow. More important in determining subsistence output per head is the amount of usable agricultural land available *per capita*. Technology is assumed constant. There is underemployment but no unemployment of labour in the sector. Output increases as a function of the subsistence population, but less than proportionally, regardless of age distribution.[3]

Out-migration increases as subsistence real income per adult falls relative to the *per capita* wage income of the uneducated population in the modern economy. The population of the subsistence economy steadily declines relative to that of the other sector because the capital investment and technological improvement of the modern sector constantly operate to increase its uneducated workers' earnings. This tendency is further accentuated when there is a lessened scarcity of educated relatively to uneducated labour in the modern sector because of high educational expenditures.

Residents of the subsistence sector use money and suffer from inflation as do those of the modern sector. They pay taxes. But they do not receive family planning or more than very elementary educational services. They

[1] By comparing the varying marginal products of educated and uneducated labour with their costs of production, the apparent rate of return on extra education " investment " can of course be calculated.

[2] This assumption is less favourable to arguments in favour of fertility reduction than are savings functions using *changes* in *per capita* income as an independent variable.

[3] In this sector it is assumed that no technological innovation occurs, the ratio of capital per worker is constant, and that the best land is scarce.

are not direct beneficiaries of direct productive investment by government. There is no savings function for the subsistence economy.[1]

The Public Sector

It is decisions on allocations within the government's budget, among categories of disbursements *and* receipts, that shape the public sector and " drive " the model. These decisions are stipulated by the analyst. It is through such decisions that he " exercises " the model.

Each of the eight expenditure categories competes with the other seven for funds. Hence an increase in defence expenditures from a given overall budget, even though it does not affect current G.D.P., can reduce government expenditures for education and investment and hence G.D.P. in later years. Government expenditures for direct productive investment or social overhead capital increase the capital stock and hence the modern sector's output.[2] Expenditures on education increase the productivity of the modern sector's labour force and may contribute indirectly to sector inmigration. Health expenditures have no direct feedback through increased output.[3] Expenditures for " other general administration " are a function of G.D.P. and have no feedback.[4] Transfer payments for welfare and for interest on the national debt—but not for family planning inducements—also are without feedback.

Family planning outlays, although comparatively small, are treated separately because a major thrust of the model is to estimate " payoffs " from reductions in fertility. Two classes of expenditures are distinguished: one is to pay for the supply of contraceptive services (clinics, doctors, nurses, devices, and information) while the other is for incentives (bonuses that compensate effective acceptors). The first expenditures result in true resource costs, but the latter are transfer payments that do not occasion the same economic loss.[5]

The size of the government budget, and hence the share of G.D.P. that the government controls, can affect G.D.P. Increased taxes reduce the

[1] The model's modern–subsistence dichotomy is not an urban–rural dichotomy. The modern sector includes urban areas plus those economic activities in rural areas that are " modern," in the sense that they use modern technologies, outside capital and management, etc. Thus a foreign owned and operated mine or rubber plantation, although in a remote geographic area, is part of the modern economy.

[2] S.O.C. expenditures need not be considered an equivalent gross investment. A reduction factor may be applied. Some prestige public works add little to output.

[3] Increased health in developing countries depends not only on health *expenditures* but on rising standards of living reflected partly in better nutrition. The expected outcome is increased effectiveness of employed labour, reduced mortality rates, and possibly increased fertility. However, as these are tenuous interactions, they have not been explicitly incorporated.

[4] Some have argued that such expenditures should be a function of both G.D.P. *and* population.

[5] This distinction between resource costs and transfer payments may seem unimportant to a director of an official family planning programme or to a minister of finance who must find the funds. Money is money to them. However, from the viewpoint of a national planning commission, this distinction is important.

disposable income from which the modern private sector saves to invest. Hence large government expenditures for defence can reduce G.D.P. It is assumed that the government borrows up to one-fourth of private savings, to the extent necessary after allowing for tax receipts, and that any remaining deficit is covered by new money creation. The resultant inflation is in proportion to the increase in the stock of money relative to real output increases. This serves to ration all excess demand to the same degree (except for funds and resources used to meet pre-specified family planning and education performance goals).

Some Special Features

One feature of the TEMPO II model is that, having a time horizon of 30 years or so, it is supply-constrained. The output is limited by the state of art and the availability of productive factors, with labour employment rates in the modern sector depending partly upon the capital to labour ratio. This contrasts with many shorter-run macro-economic models, often applied to industrialised countries, in which an increase in government spending induces increased employment and output. In the present model the effect of increased government spending is currently to reduce private consumption and private investment.[1]

The present model, although it has no explicit foreign sector, is not necessarily a closed economy model. If the country has an "import" current account, it is presumably receiving net resources, either from private or public sources abroad. External private lending and investment can be handled through an exogenous increase in the capital stock. Official external assistance can be treated as a supplement to G.D.P. with an exogenous determination as to the proportions that become a net increase in domestic consumption or investment. The model does not subsume that local consumption excludes imports or that local production excludes exports. The production equations relate to G.D.P. and not G.N.P. But if G.D.P. exceeds G.N.P., possibly because of debt service on former borrowings from abroad, this will be reflected in a more active or less passive current account (which in turn can be incorporated as explained above).

A number of probable but uncertain feedbacks have been omitted. The possible but incorporated impact of health expenditures on labour productivity and mortality rates has been mentioned; so has the influence of improved nutrition. Another omission is increased female employment in the modern sector because of reduced fertility. Still another is reduced fertility associated with increased female education. Inclusion of these last two interactions would of course strengthen the policy implication of the model, namely that reduced fertility increases output per head. As it is, the

[1] However, increased government expenditures for education or investment increase productivity, employment' and output in subsequent years.

computations of the model permit an *a fortiori* argument, for the economic benefits of reducing births would be even greater otherwise.

Family planning is shown by TEMPO II to result basically in much the same total output produced with a smaller total population. Pollution and the use of exhaustible resources depend upon total output more than on population size. Accordingly, for the "environmentalists," it should be stressed that fertility reduction provides a higher real income per individual without increased pollution or the increased use of exhaustible resources.

IV. Application of Model

Use of the model can best be understood through its application to a hypothetical less-developed country again called Developa.

Developa's output per head is initially $250 a year, a little lower than some Latin American countries and a little higher than some Asian and African countries. At the outset, 60% of the population resides in the subsistence sector, 40% in the modern. Migration from subsistence to modern sector in the 1975 base year is 1·5% of the former's population. Initial crude birth and death rates are respectively 4·4 and 1·6% a year, giving a 2·8% annual population growth (or a doubling of population in 25 years if sustained). Initial population is 10 million: that no developing nation may have this exact population is unimportant, because it is the economic and social importance of a country *relative* to its population size that matters.

Table II lists in detail, both by sector and *in toto*, many other initial conditions for 1975. This detail is subdivided into demographic, economic, and government categories. For example, initially the children to work-age population ratio is 0·81, the inflation rate is 8·5%, and hospital beds per 100,000 population are 19·5.

The first empirical task in the case of a specific country application is to estimate a considerable number of statistical inputs. Obvious ones are age specific fertility and mortality rates by sex for each year of the period analysed. Among the more important economic parameters to be estimated, sometimes from shares of productive factors in G.D.P., are the output elasticity of capital (here assumed to be 0·35), of educated labour (0·30), of uneducated labour (0·25) in the modern sector.[1] The annual improvement in the state of arts in this section is assumed to be 0·01. Among the three most important functions requiring parameter estimation are those that relate absolute disposable income and population size to private savings, subsistence real income and modern uneducated annual earnings to intersector migration, and labour force and capital to employment in the modern sector. Sometimes these functions can be based on better-known relations in "similar" countries. (These functions and certain other basic equations are described in the Appendix.)

[1] These yield initial marginal products of 24·5%, $3,790 and $617, respectively.

The crucial policy assumptions that the model's user must make refer to Developa's government budget. The user must stipulate government expenditures year by year, as percentages of G.D.P. according to the present computer programme, for health, social overhead capital, direct investment, welfare transfers, defence, etc.[1] The user must also describe how these government expenditures are to be financed; that is, he must

TABLE I

Developa budget projections as a percentage of gross domestic product [a]

Expenditures and Revenues.[b]	1975	1985	1995	2005
Education [c]	4·0	5·8	7·1	8·3
Transfers	1·9	2·3	2·6	2·8
Health	1·0	1·1	1·3	1·4
Social Overhead Capital	2·9	2·9	2·9	2·9
Direct Investment	1·9	2·3	2·6	2·8
Defence	1·9	1·9	1·9	1·9
General	4·9	4·9	4·9	4·9
Total Expenditures	18·6	21·1	23·1	24·9
Total Revenues [d]	16·5	18·7	20·4	21·4
Deficit	2·1	2·4	2·7	3·5

Notes:

[a] In the initial reference projection, it is assumed that the government has no family planning programme.

[b] Parts do not sum to totals due to rounding.

[c] Education expenditures are not directly stipulated as a percentage of G.D.P. in advance but are determined by the assumed attainment of previously stipulated enrolment ratios for primary, secondary, tertiary, and professional schools.

[d] Taxes plus borrowing from domestic and foreign sources.

determine the changing t coefficient in the tax receipts equation (see Appendix), the extent of borrowing from private savings each year, the probable amount of foreign assistance received,[2] and hence the amount of money creation (with its attendant inflation) for each year. In practice, given projected expenditures and receipts, he may seek to limit inflation to a "tolerable" rate. To ensure that related deficits after borrowing do not exceed what is "tolerable," he must over time increase tax receipts as a percentage of G.D.P., because social aspirations for education and welfare are increasing the costs of these programmes in G.D.P. terms.

In this example, using the expenditures over time denoted by Table I,

[1] Table I gives these stipulations for the example described below. An option for education (used here) is to *stipulate* school enrolments as a ratio of school age children (instead of stipulating expenditures relative to G.D.P.). Similarly, family planning expenditures may be in terms of the percentages of eligible women (age 15 to 49 years) or their spouses that it is planned to have as regular acceptors. "Transfers" does not include financial inducements to practise (accept) family planning, but interest on national debt, pensions, etc.

[2] It is assumed here that foreign assistance declines slowly over 30 years from 0·5 to 0·3 of one per cent of G.D.P.

TABLE II

Developa demographic, economic, and government programme variables, 1975 to 2005 (with and without family planning)

Variables.	Year.			
	1975 (Both Projections.)	2005 (No. F.P.)	2005 (With F.P.)	1999 (No. F.P.)
Demographic				
Population (millions)	10·0	23·2	19·5	19·6
Age 0–14	4·3	10·0	6·9	8·5
Age 15–64	5·4	12·4	11·8	10·4
Age 65+	0·3	0·8	0·8	0·6
Modern sector	4·0	15·5	12·0	12·2
Subsistence sector	6·0	7·7	7·5	7·4
Total births (thousands)	444	950	624	819
Crude birth rate (%)	4·4	4·1	3·2	4·2
Crude death rate (%)	1·6	1·3	1·3	1·4
Gross reproduction rate, modern (%)	2·90	2·64	1·32	2·69
Gross reproduction rate, subsistence (%)	3·12	3·12	3·12	3·12
Modern Sector In-migration (thousands)	90·0	156·7	170·1	141·2
Migration to subsistence population ratio (%)	1·5	2·1	2·3	1·9
Children to work-age population ratio (%)	81·0	80·6	58·8	81·2
Total dependency ratio (%)	86·9	86·8	65·3	87·3
Labour force, modern (millions)	1·47	5·79	5·50	4·51
Population growth rate (%)	2·8	2·8	1·9	2·8
Economic [a]				
G.D.P. *per capita* ($)	250	460	530	396
Modern sector	312	588	732	513
Subsistence sector	208	204	205	205
G.D.P. ($ billions)	2·50	10·67	10·30	7·75
Modern sector	1·25	9·10	8·78	6·23
Subsistence sector	1·25	1·57	1·53	1·52
Gross investment ($ millions)	317	1,476	1,436	1,052
Gross investment to G.D.P. ratio (%)	13·0	14·0	14·0	14·0
Capital per worker, modern ($)	2,948	3,993	4,148	3,725
Capital *per capita* ($)	357	873	1,027	714
Return on capital (%)	24·5	18·5	18·1	18·7
Marginal product, modern educated ($)	3,790	2,130	2,291	2,301
Marginal product, modern uneducated ($)	617	749	743	672
Unemployment rate, modern uneducated (%)	20·0	15·4	15·0	16·3
Annual inflation (%)	8·5	14·5	11·3	12·5
Cumulative general price index (1975 = 100)	100	1,054	1,632	949
G.D.P. growth rate (%)	4·2	5·5	5·0	5·4
G.D.P. *per capita* growth rate (%)	1·3	2·6	3·0	2·4
Government Programmes [a]				
Government expenditures to G.D.P. ratio (%)	18·6	25·0	24·0	24·0
Education expenditures ($ millions)	100	884	772	589
Primary enrolment rate (%)	45·4	90·7 [b]	90·5 [b]	81·9 [b]
Secondary enrolment rate (%)	23·4	46·7 [b]	46·4 [b]	42·1 [b]
Health expenditures ($ millions)	24	151	148	103
Hospital beds per 100,000 population	19·5	52·3	60·7	42·3
S.O.C. *per capita* ($)	7·30	13·20	14·40	11·40
Educated labour force, modern (thousands)	208	1,582	1,420	1,064
Uneducated labour force to total labour force ratio, modern (%)	86·0	73·0	74·0	76·0
F.P. expenditures ($ millions)	—	—	94·7	—

Notes: [a] All dollar amounts are given in base year 1975 prices.
[b] Enrolment rates are stipulated goals.

projections were made for 30 years (or nominally from 1975 to 2005).[1] In the reference projection there are no government expenditures for family planning programmes. The main outputs for this run appear in Table II, for the years 1999 and 2005. These are later compared with what happens by 2005 with publicly financed family planning (F.P.). The comparison of 1999 without F.P. and 2005 with F.P. are of interest because the total population is almost the same in both cases.

The outputs of the " No F.P." reference case are described only briefly here because the real interest is in comparing it with the " F.P." run. The projection shows substantial growth and development for Developa over the 30 years. This is due primarily to a steady advance in technology, a rapid growth in educated workers, and an increasing share of economic activity in the modern sector. Gross annual investment increases almost five-fold. Government expenditures increase from 19% to 25% of G.D.P.[2] The population is growing rapidly too, at 2·8% a year. Therefore, although G.D.P. increases 330% by 2005 (in 1975 prices), G.D.P. *per capita* increases only 85%.

V. Development Projections With Family Planning

A major purpose of the model is of course to project the consequences for economic development—in its various dimensions—of diverting a small fraction of government funds from traditional activities to officially supported programmes of family planning. In this case the analyst stipulates a proportion of women aged 15–49 (or their spouses) in the modern sector that he can reasonably expect and want to be voluntarily practising some type of effective birth control in each of the next 30 years. Having made this stipulation the government analyst estimates what will be the public cost, both in terms of supply (devices, clinics, etc.) and in terms of demand creation (advertising, inducements, etc.). He must then decide from which other government programmes he will divert funds so that total government expenditures are unchanged.

Here it is supposed that the family planning acceptance rate *goal* is 50% of eligible women aged 14–49 (or their spouse) by 1999. The assumed cost to the government for *supplying* devices and information is $20 per acceptor a year (which is high). It is supposed that acceptance rates beyond 15% will not be achieved without various kinds of financial compensations, such as bonuses for women who remain non-pregnant, payments to men for vasectomies, subsidies to commercial distributors of contraceptive pills and condoms, etc.[3] The cost of creating acceptor demand in such ways is supposed to increase linearly from $0 at 15% to $40 per acceptor-year at

[1] A complete run on an IBM 360-65 takes approximately 1 minute and costs about $15.

[2] This is partly because government has set school enrolment goals of 90% for elementary school children, a doubling since 1975, and the school age population is increasing rapidly.

[3] See " Offering Bonuses for Reduced Fertility," Stephen Enke and Bryan Hickman, *Journal of Biosocial Science*, July 1973.

50% acceptance.[1] Such a massive programme to reduce fertility would by 2005 absorb 0·5% of Developa's *output* and 4·5% of its government *budget*. The number of acceptors is by then 1·58 million and 278 thousand births are prevented that year. The budget cost is $94·7 millions and the resource cost is $44·2 billion.[2]

In the family planning projection, compared with the reference projection (without family planning), it is assumed that the annually increasing cost of the supply of family planning services and of increasing the demand for them through inducements is funded by drawing proportionately on all other government programmes (health, defence, etc.) except education (for which enrolment goals are stipulated).

Table II indicates the results of this increasing government programme to reduce fertility through two comparisons. The first is the "2005 (with F.P.)" and "2005 (No F.P.)" comparison. The second is a comparison at the same population level of approximately 19·5 million, achieved either by 1999 without such a programme or not until 2005 with a government contraceptive programme.[3]

In the 2005 comparison, the main demographic results include a gross reproduction rate that is half as large by then with government contraceptive programmes, with a rate of 1·32 instead of 2·64 in the modern sector. The population growth rate is 1·9% instead of 2·8% a year. The age distribution has become older so that the total dependency rate is 65 instead of 87%.

The main economic comparisons include a G.D.P. *per capita* of $530 a year instead of $460 without government programmes of family planning. This difference is due to the difference in population (19·5 instead of 23·2 million) outweighing the small differences in G.D.P. ($10.30 billion with and $10.67 billion without government programmes). This small difference in G.D.P. is due to the offsetting effects of a greater capital to worker ratio ($4,148 instead of $3,993) and a smaller labour force (5·50 million instead of 5·79 million) in the modern sector.

The main "social" comparison involves education and health services. School enrolment rates are necessarily the same because they are so stipulated for these runs, but educational expenditures are $772 instead of $884 million, because there are fewer children of school age; it is largely for this reason that the inflation rate is 11·3% instead of 14·5% a year. Health expenditures are slightly less also ($148 million instead of $151 million), but higher on a *per capita* basis; the supply of hospital beds and ancillary

[1] This $40 includes financial compensation (a transfer payment) and administration cost (a resource cost). Administration cost is presumed to be one quarter of the compensation costs. Thus by 2005 the demand-creating cost per acceptor year is $32 of transfer payment and $8 of real resource cost.

[2] Realistically, the supply schedule of acceptors (relative to eligible women and spouses) might be assumed to shift rightwards over time, in which case estimates of cost under present conditions may overstate costs in later decades.

[3] Even in the "No F.P." case there is an exogenous reduction in fertility of 10% over the 30 years attributable to private organisations and personal decisions uninfluenced by government.

services is also better (60·7 instead of 52·3 per 100,000 population). Social overhead capital per person is also higher ($14.40 instead of $13.20).

The comparison of " 1999 (No F.P.) " and " 2005 (With F.P.) " is included in Table II because government leaders in many developing nations seem to aspire to larger populations for their countries.[1] However, assuming that larger populations are eventually needed by such countries, it is still more economical to attain a given population increase later rather than sooner. One reason is that this provides more time for technical advances to be incorporated in production and for annual investment to increase the capital to population ratio.

VI. Significance of Models

It is difficult to demonstrate the apparent significance for development of reductions in fertility, and of programmes to reduce fertility more rapidly, on the basis of comparing only two simulated projections. Although those that have been responsible for evolving these models have had the benefit of making dozens of runs with varying parameters, these cannot be presented in detail here. The case presented here has been selected because it probably understates the argument.[2]

Nevertheless, certain conclusions do emerge, and are worth mentioning. First, there is always a substantial increase in output *per capita* over several decades if one assumes improvements in the state of art and in capital accumulation relative to work-age population: all this occurs with or without government programmes to reduce fertility. Second, reductions in fertility improve output to population ratios more rapidly: this improvement is by under one-fifth for different population sizes in 2005 or by over one-half for a 19·6 million population achieved in either 2005 (with family planning) or 1999 (without family planning). Third, G.D.P. projections are very insensitive to whether or not there are massive government programmes of family planning, the real difference being that there are fewer people to share what is approximately the same G.D.P. when there are substantial fertility reductions.[3]

[1] A frequently mentioned reason is the supposed need to populate and develop now " empty " regions in their nations. It can be shown however that a policy of developing such " empty " lands more rapidly than otherwise, through continued high fertility, can be most costly in terms of *per capita* income. Extra people can develop little without extra capital and there are other more productive uses for new investment. Moreover, many " empty " lands are covered by dense jungle that is costly to clear, even to construct necessary infrastructure. A useful analysis of this issue is contained in a forthcoming TEMPO publication by Dr. William E. McFarland.

[2] The influence of population on saving is not strongly represented in these two runs for instance. The cost per acceptor of supplying contraceptives is probably high. Also there are omitted " feedbacks " (*e.g.*, between fertility and female employment) as already mentioned.

[3] Leaders of some developing nations seem to fear that government programmes to reduce fertility will result in their nations being less strong economically and militarily. Both kinds of strength tend to depend more on G.D.P. size than population size however. Moreover, as between two countries with the same G.D.P., the one with higher living standards will probably be able better to acquire, maintain, and operate more sophisticated kinds of military equipment.

The real significance of demographic–economic models such as TEMPO II may be that they enable planners in developing countries to sense the impact of their decisions over a future many times longer than is encompassed by the usual five year plan. When planners have longer time horizons they cannot fail to become more conscious of demographic impacts on economic development. A fertility rate that may be a datum for a five year plan can become a policy variable for a 30 year projection. Delaying births through public health agencies can then be perceived as the only responsible supplement to their delaying deaths. It becomes apparent that what matters for human welfare is not absolute numbers of children in school or beds in hospitals but rather their number relative to school age children and total population respectively. Government planners can better understand through such models that the way to increase income per head is not simply to maximise G.D.P. increments but also to slow down increments in the population denominator of this ratio.

STEPHEN ENKE

Washington D.C.

Date of receipt of final typescript: November 1973.

APPENDIX

Partial Technical Description of Model

A complete technical description of TEMPO II is contained in " Description of the TEMPO II Budget Allocation and Human Resources Model," April 1973.[1] A partial description, relating only to some of the equations of more importance and interest, is given below. The meanings of the variables used in these equations are listed at the end.

Production

Subsistence Sector. Gross product in the subsistence sector, *GPS*, depends only on the size of the subsistence population *PS*. Production is made dependent on total population because of the ambiguity of the conventional labour force/non-labour force dichotomy in traditional rural populations. Algebraically,

$$GPS = x \,.\, [PS(T - 1)]^y$$

An increase in population of 1% results in an increase in output of y per cent. Since the subsistence population is working with little capital and a fixed quantity of land, y is normally less than one. That is, an increase in population does not increase output proportionately and *per capita* production falls.

Modern Sector. Gross product in the modern sector, *GPM*, is given by a Cobb-Douglas type production function that includes technological progress, capital stock, and two types of labour, *i.e.*,

$$GPM = z \,.\, (1 + q)^T \,.\, [K(T - 1)]^u \,.\, [NE(T - 1)]^v \,.\, [NU(T - 1)]^w$$

[1] The authors are Dr. William E. McFarland, Dr. James E. Bennett, and Mr. Richard A. Brown. This document is available on request to TEMPO, 777-14th St. N.W., Washington, D.C. 20005, U.S.A.

The parameter q measures technological progress, or the extent to which improvements in technique and organisation permit more output to be obtained from given inputs. K is the stock of capital measured in base year prices. NE is employed educated labour, and NU is employed uneducated labour. The exponents u, v, and w are output elasticities. Gross product is the sum of the gross products of the two sectors:

$$GP = GPM + GPS$$

Disposable Income and Consumption

Preliminary consumption C^p is a function of disposable income DI^p and population. Preliminary disposable income is gross product less taxes TAX plus total transfer payments $TRFP^p$. (Total transfer payments are the sum of general transfers, such as direct welfare payments, and family planning incentive payments.) Because corporations are not separated from households in the accounting framework of the model, disposable income includes disposable household income and the gross earnings (including depreciation but net of taxes) of private businesses. Algebraically,

$$DP^p = GP - TAX + TRFP^p$$
$$C^p = c(1) \, . \, DI^p + \delta \, . \, P$$

where

$$\delta = c(2) \, . \left[\frac{DI/P}{(DI/P)(0)}\right]^{c(3)} . \quad 0 \leqslant c(3) \leqslant 1$$

and $(DI/P)(0)$ is the initial value of *per capita* disposable income.

The parameter $c(1)$ is the lower bound of the average propensity to consume (*i.e.*, C/DI). The index number for *per capita* income modifies the effect of population growth by introducing an upward shift of the consumption rate as *per capita* income increases. Note that if $c(3)$ has a value of zero, total consumption becomes a linear function of disposable income and population. If $c(3)$ has a value of unity, population is eliminated from the consumption function and consumption becomes proportional to disposable income.

As noted above, " real " consumption will generally be some fraction of preliminary consumption, or $C^a = \text{adj} \, . \, C^p$.

Private Savings, Investment, and the Capital Stock

Preliminary private savings S^p is equal to preliminary disposable income minus preliminary consumption. Preliminary investment by the private sector $PINV^p$ is preliminary private savings minus the preliminary amount of government borrowing BOR^p:

$$S^p = DI^p - C^p$$

and

$$PINV^p = S^p - BOR^p$$

Again, " real " private investment $PINV^a$ is some fraction of preliminary private investment; $PINV^a = \text{adj} \, . \, PINV^p$.

The capital stock K at time T is the previous period's capital stock less depreciation, *i.e.*, $(1{\cdot}0 - d) \, . \, K(T-1)$, plus real private investment $PINV$ and increments resulting from direct government investment DGI^a and government investment in social overhead capital SOC^a.

$$K = (1\cdot0 - d) \,.\, K(T-1) + f(1) \,.\, DGI^a + f(2) \,.\, SOC^a + PINV^a$$

Capital stock is measured in constant base year prices. The coefficients $f(1)$ and $f(2)$ are applied to GDI^a and SOC^a to adjust for expenditures in these areas that do not add to the productive capacity of the economy.

Employed Labour Force in the Modern Sector

Employed educated labour NE is assumed to be a constant fraction ne of the educated labour force LFE:

$$NE = ne \,.\, LFE$$

Employed uneducated labour NU at time T is some fraction of the adjusted uneducated labour force $LFUA$. This fraction increases as the capital-labour ratio increases, approaching asymptotically some maximum rate of employment, $k(1)$.

$$NU = \left\{k(1) - k(2) \,.\, \left[\frac{LFU(T-1)}{K(T-1)}\right]\right\} \,.\, LFUA$$

The second parameter, $k(2)$, is chosen so as to obtain the appropriate initial rate of employment.

The employment function is based on the assumption that unemployment results primarily from a growth of the labour force too rapid to be utilised (or " absorbed ") by the available stock of capital, *i.e.*, all unemployment is structural. A more rapid growth of capital relative to the growth of labour therefore results in a higher rate of employment.

Migration

Migration from the subsistence sector to the modern sector is assumed to depend on the ratio of *per capita* wage income of the uneducated modern population to *per capita* income of the subsistence population. Specifically, total net migration MIG is

$$MIG = m(1) \,.\, R^{m(2)} \,.\, PS$$

where

$$R = \frac{w \,.\, GPM(T-1)}{PMU(T-1)} \div \frac{GPS(T-1)}{PS(T-1)}$$

The uneducated modern population PMU is assumed to receive as wage income a share of the modern gross product GPM equal to w, the output elasticity of uneducated labour in the modern sector production function.

The parameter $m(2)$ measures the sensitivity of the subsistence population to changes in the ratio of incomes. It implicitly takes into consideration such factors as the cost of migration, differences in the cost of living, cultural differences, and imperfect information and perceptions. By setting $m(2)$ equal to zero, net migration becomes a constant proportion of the subsistence population—*i.e.*, migration becomes totally insensitive to the income ratio. The other parameter, $m(1)$, is adjusted so as to obtain the appropriate initial rate of migration.

402

The age-sex composition of the migrants is determined by a set of exogenously specified weights.

Government Sources of Funds

The government finances its outlays from tax revenue TAX, government borrowing from the private sector BOR^p, and the creation of new money. At the option of the user, the government may also receive funds from foreign loans and grants GFB, specified as a percentage of gross product.

Target tax revenues TAX are proportional to gross product,

$$TAX = t \cdot GP$$

where t is the average tax rate. This proportion can be specified to change linearly through time.

The difference between total government outlays G and taxes is the deficit DEF to be financed,

$$DEF^p = G^p - TAX$$

This deficit is financed through a combination of domestic borrowing from the private sector BOR and the creation of new money or borrowing from the central bank. It is assumed that the government finances the deficit as much as possible through domestic borrowing, but that this borrowing cannot exceed some fixed portion of private savings S^p. Algebraically,

$$BOR^p = \text{Minimum of } (DEP^p \text{ or } a \cdot S^p)$$

where " a " is the maximum proportion of private savings that can be borrowed by the government. The remainder of the deficit is assumed financed by the creation of new money or borrowing from the central bank.

PARTIAL LIST OF VARIABLES AND SYMBOLS

Demographic Variables

Description	*Symbol*
Population	
total	P
modern, total	PM
subsistence, total	PS
modern, uneducated, total	PMU

Economic Variables

Gross product	
total	GP
modern	GPM
subsistence	GPS
Labour force (modern only)	
educated	
uneducated	
uneducated, adjusted for mor	

PARTIAL LIST OF VARIABLES AND SYMBOLS—*Contd.*

Economic Variables

Description	*Symbols*
Employed labour (modern only)	
educated	*NE*
uneducated	*NU*
Capital stock	*K*
Consumption, private	*C*
Savings, private	*S*
Investment, private	*PINV*
Disposable income	*DI*

Government Programme Variables

Government outlays	
total	*G*
social overhead capital	*SOC*
direct government investment	*DGI*
Tax revenue	*TAX*
Borrowing (domestic)	*BOR*
Government deficit	*DEF*
Total transfer payments	*TRFP*

The age-sex composition of the migrants is determined by a set of exogenously specified weights.

Government Sources of Funds

The government finances its outlays from tax revenue TAX, government borrowing from the private sector BOR^p, and the creation of new money. At the option of the user, the government may also receive funds from foreign loans and grants GFB, specified as a percentage of gross product.

Target tax revenues TAX are proportional to gross product,

$$TAX = t \,.\, GP$$

where t is the average tax rate. This proportion can be specified to change linearly through time.

The difference between total government outlays G and taxes is the deficit DEF to be financed,

$$DEF^p = G^p - TAX$$

This deficit is financed through a combination of domestic borrowing from the private sector BOR and the creation of new money or borrowing from the central bank. It is assumed that the government finances the deficit as much as possible through domestic borrowing, but that this borrowing cannot exceed some fixed portion of private savings S^p. Algebraically,

$$BOR^p = \text{Minimum of } (DEP^p \text{ or } a \,.\, S^p)$$

where " a " is the maximum proportion of private savings that can be borrowed by the government. The remainder of the deficit is assumed financed by the creation of new money or borrowing from the central bank.

PARTIAL LIST OF VARIABLES AND SYMBOLS

Demographic Variables

Description	*Symbol*
Population	
total	P
modern, total	PM
subsistence, total	PS
modern, uneducated, total	PMU

Economic Variables

Gross product	
total	GP
modern	GPM
subsistence	GPS
Labour force (modern only)	
educated	LFE
uneducated	LFU
uneducated, adjusted for morbidity	$LFUA$

PARTIAL LIST OF VARIABLES AND SYMBOLS—*Contd.*

Economic Variables

Description	*Symbols*
Employed labour (modern only)	
educated	*NE*
uneducated	*NU*
Capital stock	*K*
Consumption, private	*C*
Savings, private	*S*
Investment, private	*PINV*
Disposable income	*DI*

Government Programme Variables

Government outlays	
total	*G*
social overhead capital	*SOC*
direct government investment	*DGI*
Tax revenue	*TAX*
Borrowing (domestic)	*BOR*
Government deficit	*DEF*
Total transfer payments	*TRFP*

PRINCIPAL ASSISTANTS

William J. Barber, Professor of Economics,
Wesleyan University, Middletown, Connecticut, U.S.A.

Altti Majava, Demographer, National Planning Office,
Helsinki, Finland

Alva Myrdal, Cabinet Member, Minister for
Disarmament, Stockholm, Sweden

Paul P. Streeten, Professor of Economics,
University of Sussex, Brighton, England

David Wightman, Reader in International Economic History,
University of Birmingham, England

George W. Wilson, Chairman, Department of Economics and Division of
Economic Research, Indiana University, Bloomington, Indiana, U.S.A.

A TWENTIETH CENTURY FUND STUDY

ASIAN DRAMA

An Inquiry Into the Poverty of Nations

by GUNNAR MYRDAL

VOLUME III

PANTHEON
A Division of Random House
NEW YORK

Third printing

Published in New York in hard cover by the Twentieth Century Fund and in paper cover by Pantheon Books, a division of Random House, Inc. Published simultaneously in Toronto, Canada, by Random House of Canada Limited.

Printed in the United States of America
Library of Congress Catalog Card Number 67–19179

Outline of the Book

Volume I

Foreword by August Heckscher — *page* v

Preface — vii

Prologue The Beam in Our Eyes — 5

PART ONE. INTRODUCTION — 37

Chapter 1 Scope and Direction of the Study — 39
2 The Value Premises Chosen — 49
3 The Wider Field of Valuations — 71

PART TWO. POLITICAL PROBLEMS — 127

Chapter 4 The Coming of Independence — 129
5 The Frontiers of Independence — 175
6 Partition of Imperial India and Its Aftermath — 231
7 India — 257
8 Pakistan — 305
9 Ceylon and Southeast Asia — 343

PART THREE. ECONOMIC REALITIES — 411

Chapter 10 Population and the Development of Resources — 413
11 National Output and Structure of Economy — 473
12 Levels of Living and Inequality — 529
13 Foreign Trade and Capital Flows — 581
14 Differences in Initial Conditions — 673

405

Volume II

PART FOUR. A THIRD WORLD OF PLANNING 707

Chapter 15 Spread and Impact of the Ideology of Planning 709
16 Equality and Democracy 741
17 Concepts and Practice of Socialism 799
18 Concepts and Practice of Democratic Planning 849
19 Operational Controls Over the Private Sector 901
20 Corruption – Its Causes and Effects 937

PART FIVE. PROBLEMS OF LABOR UTILIZATION 959

Chapter 21 "Unemployment" and "Underemployment" 961
22 Labor Utilization in Traditional Agriculture 1029
23 Labor Utilization Outside Traditional Agriculture 1093
24 The Industrialization Issue 1149
25 The Case for Crafts and "Small-Scale Industry" 1207
26 Agricultural Policy 1241

PART SIX. PROBLEMS OF POPULATION SIZE 1385

Chapter 27 Population Prospects 1387
28 Population Policy 1463

Volume III

PART SEVEN. PROBLEMS OF POPULATION QUALITY 1531

Chapter 29 "Investment in Man" 1533
30 Health 1553
31 Education: The Legacy 1621
32 Literacy and Adult Education 1651
33 The School System 1697

Postscript 1831

APPENDICES

Appendix 1 Diplomacy by Terminology 1839

2 The Mechanism of Underdevelopment and Development and a Sketch of an Elementary Theory of Planning for Development 1843

3 Economic Models and Their Usefulness for Planning in South Asia 1941

4 The Structure of the Plans 2005

5 A Note on "Accounting Prices" and the Role of the Price Mechanism in Planning for Development 2031

6 A Critical Appraisal of the Concept and Theory of Underemployment 2041

7 A Note on Inadequate Approaches to the Economic Effects of Population Changes 2063

8 A Note on Positive Operational Controls 2077

9 A Comparison of European and South Asian Nationalism 2109

10 Climate and Its Economic Consequences in South Asia 2121

11 Notes on Migration 2139

12 Family Planning Policy in India 2151

13 Computation of Income Per Head 2165

14 Estimation of Over-All Income Inequality 2181

15 The Instability of Export Earnings 2187

16 A Critical Appraisal of Selected Studies on "Unemployment" and "Underemployment" 2203

Author Index 2225

Subject Index 2249

APPENDIX 7

A NOTE ON INADEQUATE APPROACHES TO THE ECONOMIC EFFECTS OF POPULATION CHANGES

To clear the deck for the analysis in Chapter 28 (Part I) of the economic effects of population growth in South Asia, this brief note will seek to demonstrate the limitations of some theoretical approaches to this problem that are met frequently in both popular and scientific discussions.

1 *Population Optimum and Similar Concepts*

One such approach makes use of the concept of *population optimum.* Although this concept is heard of less often than it used to be, it still underlies much of the discussion of population problems. The postulation of an optimum size of population for a particular country represents an attempt to define "objectively" an ideal. In modern times the basis of this definition has been the maximization of income or output per head.[1]

[1] To cite but two examples: "The concept, 'optimum population,' is employed to signify that under given conditions in a country there is a population size that is preferable to any other larger or smaller size. What this size is depends upon the conditions that supposedly are given and upon what it is that a population wants. For purposes of the present discussion it will be supposed that what a population seeks to maximize is per capita income over the relatively long run, or the present value of future per capita income, or some other such indicator of economic 'welfare.'" (Joseph J. Spengler, "The Economics of Population Growth," in *The Population Crisis and the Use of World Resources,* Stuart Mudd, ed., W. Junk Publishers, The Hague, 1964, p. 87.) ". . . there will be at least one point in which productivity is a maximum. This point of density is called the optimum density or optimum population point." (Jan Tinbergen, *International Economic Integration,* Elsevier, Amsterdam, 1954, p. 37.)

The normative arbitrariness of this supposedly "objective" definition of population optimum is illustrated by the fact that in earlier days an alternative was occasionally put forward. Henry Sidgwick, for instance, postulated maximum total against maximum average satisfaction; see Gunnar Myrdal, *The Political Element in the Development of Economic Theory,* Routledge & Kegan Paul Ltd., London, 1953, pp. 38ff., especially p. 38, f.n. 6. Also, J. E. Meade, in Chapter VI of his *Trade and Welfare* (Oxford University Press, London, 1955), proposes, against current fashion, the maximization of

The idea of an optimum population not only objectifies implicit valuations, as does all welfare theory, but also illustrates the illegitimate *ceteris paribus* assumption which leads to logical contradictions. It implies a comparison between the present size of a population and its optimum size, in terms of maximum income (or output) per head. In this comparison all conditions relevant to income and output, other than population size, are assumed to remain unchanged, including the amount and kind of capital equipment, the state of the arts (technical knowledge) and all production functions, and the amount and terms of foreign trade. In addition to these conditions, conventionally held constant in short-term, static economic analysis,[1] all attitudes and institutions affecting productivity must also be assumed to remain constant.

The change-over from the present size of a population to its optimal size cannot occur instantaneously. It must involve changes in one or several of the determinants of population change: natality, mortality, migration. Changes in these occur gradually and are interrelated through their combined effects on age distribution. Different combinations of constancy and changeability in natality, mortality, and migration and different rates of change in these determinants will result in differences in the time period required to achieve a change in population size. During the transition period the age distribution and the other conditions mentioned in the previous paragraph relevant to output and income are bound to change. The assumption of constancy is both unrealistic (for autonomous changes and changes induced by future policies) and logically invalid (for changes resulting from the population changes themselves).

---

total welfare as the ultimate criterion of the optimum population (p. 83). (Like most modern welfare theorists, Meade ignores the old discussions, in which the utilitarian theory was worked out with greater conviction and in clearer terms; a closer reading of the exchanges between Sidgwick and Edgeworth and some of J. S. Mill's writings would have given his abstract normative reasoning greater depth and a wider historical perspective.) Recently P. T. Bauer has also shown leanings toward the alternative norm, though without expressing his thoughts very clearly; see P. T. Bauer, *Some Economic Aspects and Problems of Under-Developed Countries*, Forum of Free Enterprise, Bombay, 1959, p. 15, and P. T. Bauer and B. S. Yamey, *The Economics of Under-Developed Countries*, University Press, Cambridge, 1957, pp. 63, 152 *et passim*.

Concerning the doctrinal origin of the concept of "population optimum" in classical economic theory and its relation to the philosophies of natural right and utilitarianism, and for a general criticism of the concept, see the writer's *The Political Element in the Development of Economic Theory*, pp. 38ff. *et passim;* and *Population, A Problem for Democracy*, Harvard University Press, Cambridge, Mass., 1940, pp. 131ff., especially pp. 139ff.

[1] A host of specific assumptions, rarely spelled out, are necessary if the analysis is to have precise meaning. Since income (output) per head can be defined unambiguously only in a one-commodity world, the index number problem arises, and with it the question of how the heterogeneous collection of products is to be weighted. There are also the questions of how income is and "should be" distributed, of the direction of demand, the homogeneity and quality of the labor force, and so on. If inequalities of income distribution are substantial and if some factors of production are specific, it is impossible, over a wide range of outputs, to state without ambiguity what constitutes an increase in output per head. These implicit assumptions, like those mentioned in the text, acquire importance because the comparison is of the relatively long-run variety (see the first footnote).

Here is a particularly striking instance of the inadmissibility of the *ceteris paribus* assumption. Since changes in population size cannot be instantaneous, we must specify what population determinants are to be changed and at what speed. The choice between the different possibilities involves valuations. During the period of transition, "other things" cannot be assumed to stay constant and the process by which and the time in which the optimum size is reached are themselves bound to affect the conditions that would have to be assumed constant in order to make sense of the "optimum."[1]

Apparently in order to avoid the logical difficulties – felt but never properly spelled out – of the static *ceteris paribus* assumption, resort is sometimes had to the idea of *optimum population growth*. But this does not solve the problem. It implies, of course, the concept of a population optimum at a particular point of time. If in the initial situation the actual size of the population is other than optimum, one would perhaps be tempted to conclude that optimum population growth would be the rate that would most speedily change the actual population to an optimum. But that problem cannot be discussed without considering the factors discussed above in connection with the concept of optimum population size. It might seem, then, that the way to salvage the concept of optimum population growth would be to assume that the initial population size corresponds to the optimum size. Its optimum growth would then, however, be determined by all the other social and economic conditions as they change autonomously and through population growth. In other words, the optimum remains indeterminate until changes in other conditions and inter-relationships are made explicit. Moreover, to assume that the population initially is of the optimum size and thereafter moves along an optimum curve is to beg the question; all the difficulties mentioned in the foregoing paragraphs arise again.

We should not be surprised that in spite of the ubiquity of these concepts, it has never been possible to apply them effectively in analyzing a real situation or a process. A critical analysis would have exposed the futility of such an attempt. Indeed, the very idea that the population problem in the extremely poor countries in South Asia can be analyzed with practical benefit in terms of optimum population size or optimum population growth is fanciful and attests the sterility of social and economic speculation that has no contact with reality. The approach is as absurd as would be a discussion of the housing problem in South Asia in terms of "optimum housing accommodation." The practical problem for population policy, as for other issues, is how to achieve an *improvement*, not an *optimum*.[2] The valuations involved should be stated and the desired situation should be compared with a specified situation, either present or potential, with full account taken of the consequences of the process of transition – normally through a reduction in fertility rates – to the improved situation.

The concepts of *overpopulation* and *underpopulation*, when related to a concept of population optimum, are indeterminate for the reasons given. But we

[1] In this book we have repeatedly found it necessary to discard concepts and theories that use static comparisons. The grounds for their rejection have always been logical inconsistency and inadequacy to reality. See, in particular, Appendix 6, Section 8.

[2] Appendix 2, Sections 7 and 13.

have also found "overpopulation" used as a synonym for "underemployment,"[1] a concept we have discarded as logically invalid and not adequate to reality in South Asia.[2] It is not easy to give a clear meaning to the concept of overpopulation even if it is freed from associations with concepts of population optimum or underemployment. When related to the practical problem of a population policy based on an analysis of the facts and postulated value premises, "overpopulation" simply means that a reduction in fertility rates is desirable, because it would improve development prospects. As we show in Chapter 28 (Part I), all South Asian countries are in this sense "overpopulated" in various degrees – as are, indeed, most poor countries. It is difficult to see the scientific purpose of the term, given this very wide, even if clear, meaning. It is, in any case, the *degree* of "overpopulation" that is of interest. In practice, when the term is not related to the logically faulty and unrealistic concepts of population optimum or underemployment, it is used to indicate in a broad and general way that a population is very much "too large" to allow "decent" levels of living and "reasonable" prospects for development. We have no use for the term or for the related term "population pressure."

All these concepts and theories, stated in terms of "population," are also misleading because they abstract from changes in age distribution, which inevitably accompany a primary change in fertility rates (and most non-imaginary cases of migration). One service of the study of population growth by Coale and Hoover[3] is that it stresses this point and illustrates it by realistic projections. Too often, birth control in South Asia is advocated only because of the difficulties foreseen in *creating employment* for the growing population, without recognizing that for a long time to come the rate of increase in the labor force will be determined by births that occurred in an earlier period. Those who do recognize this fact sometimes draw the opposite conclusion that the spread of birth control is *not an urgent concern* because a lower birth rate cannot give much economic relief in the next few decades. This reasoning, of course, puts things upside down. By reducing the dependency burden, a reduction in the birth rate will immediately begin to improve levels of income and living. In many direct and indirect ways it will also make it easier to accelerate development policies by improving labor utilization and productivity. A lower current birth rate will, in addition, have important effects in the next generation, both on the reproduction potential and on the rate of increase in the labor force.[4]

2 *The Simple Model*

Most of the approaches to the problems of population growth – whether they are embellished by the concepts exemplified above or not – have usually this in common: the economic effects of a population increase are considered in terms

[1] Appendix 6, *Sections* 1, 3, esp. pp. 2042 (*note* 1), 2045; Chapter 21, Section 14. This is an aberration of the classical doctrine that assumed full employment in the long run when there would be equilibrium between supply and demand. See Chapter 21, p. 974, footnote 1.

[2] Chapter 21, Sections 12–14; Appendix 6.

[3] Ansley J. Coale and Edgar M. Hoover, *Population Growth and Economic Development in Low-Income Countries*, Princeton University Press, Princeton, 1958.

[4] Chapter 28, Section 4.

of the man/land ratio and the static law of diminishing returns. They recognize the possibility that the "population pressure" could be reduced by a movement of labor from agriculture to industry, where increasing returns are assumed to prevail, and by investments and technical improvements in agriculture. The principal innovation in the modern approach has been to reduce to one – capital – the scarce factors cooperating with labor and to assume a definite relationship between capital input and total product, both considered in aggregate terms. An increase in the labor force raises the demand for a certain amount of capital investment – often called, after Sauvy,[1] the purely "demographic investment" – simply in order to keep total product per head unchanged. This model is now reproduced in almost every scientific[2] or popular discussion of economic development in underdeveloped countries; as we shall point out in the next section, it is implied in the more complicated model that takes age distribution into account as well. With some variation in the numerical values chosen, it is expressed in this standard form:

> If population increases by 2 percent annually and if the marginal capital/output ratio is 3 to 1, 6 percent of the national income must be saved and invested per year in order to maintain the present level of income per head. If it is desired to increase income per head by 2 percent a year, another 6 percent of the national income must be saved and invested.

A more careful formulation would assume, among other things, a constant ratio of labor force to population. This would imply an unchanged age distribution and therefore a constant birth rate over a long period and an equal incidence of the decline in the mortality rate in all age groups, if there is such a decline.

Let us, for the sake of the argument, make the assumption of a constant ratio of labor force to population and concentrate on the economic effects of an increase in the labor force. The model shows that the increase in the labor force has an adverse effect on average income, because, in order to provide the labor force with capital at unchanged levels of average productivity, a certain amount of net savings is used that would otherwise be available to invest for development. Assuming, further, *ceteris paribus*, constant return, and a given capital/output ratio, the model provides a very simple measure of this adverse economic effect of an increase in the labor force: the extra capital required merely to keep average income at the initial level.

469

This kind of analysis, in terms of capital investment and output, is typical of the modern approach to the economic problems of the South Asian countries. In our opinion it is too mechanistic and schematic. It gives the appearance of knowledge where none exists, and an illusory precision to this pretended knowledge. By overemphasizing investment – usually physical investment – it tends to direct research along lines that are largely unrealistic for South Asia. Because this modern approach represents a major bias in planning and policy we have criticized it in several contexts in this book.[3] A few of these criticisms are worth

[1] Alfred Sauvy, *Théorie générale de la population*, Vol. I. *Économie et Population*, Presses Universitaires de France, Paris, 1952, pp. 288–290.

[2] For an example see Jan Tinbergen, *The Design of Development*, Johns Hopkins Press, Baltimore, 1958, p. 14.

[3] See in particular Appendix 2, Sections 19–21, and Appendix 3.

noting here because they are of special importance for the study of population growth.

As we show in Part Five of the book, the present situation in the South Asian region is one of severe underutilization of the labor force, though not to the same degree in every country. Economic development must be thought of primarily as the achievement of greater labor input and labor efficiency (for agriculture, see particularly Chapter 26, Section 1). Economic development means, in the first place, that more people will work, that they will work longer and more efficiently, and that they will cooperate in order to create institutions that make this possible and rewarding. Changes in these respects will increase the productivity of the labor force. The availability of more capital can, of course, also increase labor productivity; if wisely invested according to a plan designed for that purpose, it may even facilitate and hasten fuller and more intensive labor utilization. The same effect can also be expected from many specific policies: for instance, policies directed at effective land and tenancy reforms; at rising levels of nutrition, health, and education; providing a larger and better trained cadre of managerial and supervisory personnel at all levels of responsibility; improving efficiency and honesty in public office; and, generally, promoting a consolidated nation ruled by an informed and determined government.

These policies cannot meaningfully be expressed in terms of capital invested. Neither can a realistic analysis proceed from the assumption that the conditions these policies seek to change are constant. The assumption of *ceteris paribus* would, indeed, be logically impossible, since some of these conditions will be altered by the increase in the labor force – the primary change whose economic effects on income per head the model is intended to measure. Besides, except for capital inflow from abroad, the actual limits of savings and investment are set by the same factors that limit the scope and effectiveness of the policies mentioned for increasing labor input and efficiency – levels of living, attitudes, and institutions, precisely the things from which the modern approach, as exemplified by the model quoted above, abstracts. Also, the economic effects of investment, as of any other policy, cannot be analyzed in isolation; they depend on the combination of policies applied. Finally, the age distribution cannot be disregarded when, as ordinarily, the model is used to motivate a policy of spreading birth control.

We cannot regard the modern theoretical approach as an advance.

3 *The More Inclusive Model*[1]

Increased attention has recently been paid to the economic effects of the change in age distribution resulting from reduced fertility rates, and in consequence more complicated models have been constructed.[2] Because the demographic determi-

[1] George Peterson kindly made suggestions for clarifying and strengthening the argument of this section and helped draft certain parts of it.

[2] Coale and Hoover have presented such models (*Population Growth and Economic Development in Low-Income Countries*, pp. 259ff.), and so have others, particularly among the French demographers (see, for instance, George Balandier, ed., *Le Tiers Monde*, Institut National d'Etudes Demographiques, Cahiers No. 27, Paris, 1956). See

nants – births, deaths, and migration – and the resulting population size, population growth, and age distribution lend themselves to the construction of models with determinate and predictable relationships, it has been tempting to graft onto this demographic mechanism those economic relationships that are thought to be determined by the demographic changes, particularly by the changes in age structure and their effect on the labor force and the dependency rate. By tracing the effects not only of the change in size of the labor force, which for a generation is almost a datum, but also of the changing dependency rates on consumption, saving, and investment and on the allocation of investment, it is hoped to show how income per head changes over time, largely as a result of different assumptions about fertility rates and the consequent modification of the age distribution. The behavior of mortality rates, although determined in part by the age structure, has a much less important effect on it than do changes in fertility rates (Chapter 27, Section 5). International migration can be disregarded as insignificant for the South Asian countries (Chapter 28, Section 14).

In projecting the effects of reduced fertility on Indian economic growth in the period 1956–86, Coale and Hoover assume (1) that the rate of growth of output is determined largely by public development outlays and private investment; (2) that the amount of funds available for these expenditures depends both on national income and on income per head; (3) that these expenditures fall into two categories – "direct growth" investment and public "welfare" outlays – the former contributing more to output than the latter (the direct growth investment contributes directly to output whereas the public welfare outlays do so by enhancing the incentives and energy of the labor force); (4) that welfare outlays can be subdivided into those providing facilities for the existing population and those adding to facilities because of population growth, and that these two categories make different contributions to output; (5) that only welfare outlays that assist the existing labor force increase present output, while there is a lag of fifteen years before those assisting the rest of the population affect output.

Coale and Hoover justify the assumption that the rate of growth of output depends largely on (1) public developmental outlays and public and private productive investment that affect output directly and (2) the indirect effects of welfare outlays on the incentives and energy of the present and future labor force on the grounds that these two types of determinants are the ones most clearly affected when alternative assumptions are made about rates of population growth; other determinants, such as government policies, technical assistance, and administration, are considered to be relatively unaffected. Coale and Hoover are aware that to ignore the effects of consumption levels on the vigor, efficiency, and adaptability of the labor force is a serious omission, but plead ignorance in defense. They stress that they are not making forecasts, but projections intended only to show the effects of different assumptions about fertility. Their projections contain substantial refinements of the simpler model, criticized in the previous section. Like savings in the simpler model, investment and public

---

also "Growth Models for Illustrating the Effects of Alternative Employment and Investment Policies," United Nations, ECAFE, *Economic Bulletin for Asia and the Far East*, Vol. IX, No. 1, New York, 1958.

developmental outlays, including governmental welfare expenditures, determine the growth of national income according to an "incremental developmental outlay/output" ratio – a sophistication of the more familiar incremental capital/output ratio. This coefficient is first assumed to rise from 3:1 in 1956 at the rate of 2 percent per annum. Finally, Coale and Hoover vary the parameters in several ways, changing the coefficient of investment and developmental outlays to incremental consumer income, changing the weights of the growth contribution made by welfare outlays, and changing the assumptions about the rise of the developmental outlay/output ratio. Although these changes yield widely varying *incomes* in 1986, the percentage *difference in income* associated with reduced fertility is said to be persistent and stable. In 1986 this differential ranges only from 38 to 48 percent.[1]

The main feature of Coale and Hoover's model is its quantitative illustration that a decrease in fertility accelerates the rise in total national income and, even more, in income per head, because (1) a larger proportion of income is available for growth exenditure and (2) a smaller proportion of this larger ratio goes into low-yielding and late-yielding welfare expenditure on extra population. Coale and Hoover are careful to point out that "This array of results should suffice to show the inapplicability of any of these projections as a forecast of likely economic growth, since we have no real basis for defending any one of the seven combinations of underlying assumptions as the best." They claim, however, that "The main body of the table . . . does give a positively significant result. It shows that through this whole gamut of projections, despite the wide variation in rates of progress that they imply, the *differential associated with reduced fertility* is remarkably persistent and stable."[2] In view of their modest claims and pioneering effort, it seems churlish to find fault with this conclusion. But despite their disclaimers, they create a false air of precision and a false confidence in their model. A critical examination is therefore justified. For the sake of argument, we shall assume that several aggregative concepts, and in particular, income, savings, and development outlay, can be given a clear meaning.[3]

First, Coale and Hoover are fully aware of leaving out of consideration a number of important variables, simply because they are difficult or impossible to quantify. Since the largest variation is associated with the distribution of imputed growth effects among different types of outlays, it is plausible to assume that the growth effects of policies and of changes in other conditions left out of consideration would make a substantial difference. We shall return to this point at the end of this section.

Secondly, the apparent stability of the income differential associated with reduced fertility is largely spurious. It is the result of their altering the parameters, which they list one at a time rather than in groups or all together. Thus if one inserts simultaneously all the values that lower the differential, or, *a*

[1] Coale and Hoover, *Population Growth and Economic Development in Low-Income Countries*, p. 281.

[2] *Ibid.*

[3] See Chapter 11, Section 1; Chapter 12, Section 2; and see Appendix 4, Section 3 *et passim*.

fortiori, all values that raise it, the differential is greatly increased. Coale and Hoover themselves say that there is "no real basis" for preferring one of the listed values of their parameters to any other. It follows that there is no real basis for choosing any single combination of parameters over any other single combination. Assume that the ratio Y_L/Y_H, where L indicates low and H high fertility, is a function of three parameters A, B, and C and that for each parameter there are three possible values, denoted by subscripts 1, 2, and 3. Assume that when the other two parameters are held constant, the ratio Y_L/Y_H is greatest when the variable parameter is evaluated at 1, second greatest at 2, and least at 3. What Coale and Hoover show is that if we start with a function of the form

$$Y_L/Y_H = f(A_2, B_2, C_2)$$

and substitute new values for the parameters A, B, C, *one at a time*, the ratio does not change much. That is, $f(A_2, B_2, C_2)$ is approximately equal to $f(A_1, B_2, C_2)$, which again is approximately equal to $f(A_2, B_1, C_2)$, and so on. But they have *not* shown that $f(A_1, B_1, C_1)$ is approximately equal to $f(A_3, B_3, C_3)$. It is obvious that this difference is much greater than the differences obtained by changing the parameters one at a time. Coale and Hoover give the impression that their variation of parameters is representative of all possible variations, for they speak of not computing ratios for all 27 possible combinations because this would take too much time and space. But the 7 computations they present cluster around the minimum variations. The apparent stability of the income differential when the parameters are varied is therefore due to a highly selective procedure.

Thirdly, the variations in their assumptions are arbitrarily confined to a narrow area. It can easily be shown that widely varying income differentials will result if quite plausible assumptions are made about, say, different capital/output ratios for the low-fertility than for the high-fertility projection, or about the savings function, or the welfare effects. A higher fertility rate necessitates a larger diversion of funds into welfare expenditure. Not only does this reduce the funds available for direct growth expenditure, but it may also raise the capital/output ratio for the direct growth investment that does take place. The use of identical capital/output ratios for the alternative projections is therefore not the only plausible procedure. For a number of reasons, spelled out elsewhere,[1] the amount of investment will influence the size of the capital/output ratio. If we now compare the two projections that use the same capital/output ratios with those assuming different ratios, the differential Y_L/Y_H is seen to be quite different for the two cases. The developmental outlay/output ratio may differ even more. It thus appears that a slight variation in the assumptions is sufficient to cast doubt on Coale and Hoover's principal conclusion. Their argument is that economists do not know precisely the functional relationships between outlays and income, but that whichever of several assumptions about these is made, the percentage by which low-fertility income exceeds high-fertility income is roughly the same. But if we consider the possibility that different capital/output ratios and developmental outlay/output ratios for the

[1] Appendix 3, Sections 13, 17 (iii), 18, and 22.

members of each pair of projections are as plausible as are the same capital/output ratios, the conclusion no longer holds. We arrive at considerable differences in the differential.

Since even when based on the authors' own highly restricted views of the determinants of growth, the stability of the differential for low- and high-fertility projections is largely spurious, it follows that if we include the possibility of institutional and attitudinal changes and of different policies, there is little reason to assume stability of the differential. We have seen in our discussion of the importance of coordinating policies,[1] especially investment projects, that development contributions depend on a set of complementary and supplementary measures. The evidence from the Coale and Hoover projections, which impute different growth effects to different outlay distributions, supports the view that the range of the differential would be widened if other determinants of growth were considered.

Fourthly, insofar as there is a constant differential when the purely economic parameters are varied, it should be noted that labor force size does not respond to these variations, since all those persons are already alive who will enter the labor force in fifteen years or who will bear children in the next fifteen years; that is, the potential increment to the labor force over the next thirty years has an upper limit. The scope for economic changes over three decades is therefore strictly circumscribed, despite wide variations in the economic parameters. Changes in the age structure of the population due to variations in fertility rates are slow compared with the quasi-constancy of the size of the labor force, which dominates Coale and Hoover's thirty-year model. Even over a period longer than thirty years, the effects of the different fertility rates dominate over the effects of varying the purely economic parameters. For although it is true that after thirty years the increase in the labor force will add substantially to total production, this will not be sufficient to compensate for the depressing legacy of the high fertility. This depressing influence is due to three factors. (1) Output per worker will be lower than would be the case in the low-fertility projection, where, instead of duplicating facilities, each worker would be equipped with more capital. (2) The greater burden of dependency reduces the resources available for raising productivity per worker. (3) The lower level of income reached after thirty years will reduce the saving and investment ratio below what it would be on the low-fertility assumption. For all three reasons the gap in output per worker and income per consumer between the two projections will continue to widen, even if economic parameters were varied within a wide range.

Fifthly, the striking thing about Coale and Hoover's calculations, even under their arbitrary assumptions that minimize the range of the differential, is not the stability of the differential but its change from stability over the first ten years to increasing variation in the two succeeding decades. The relative superiority of total income under the low-fertility assumption over that under the high-fertility assumption is the same for projections 4 and 5 at the end of the first decade; at the end of the second decade, projection 4 exceeds projection 5 by 3 percent; and at the end of the third decade, by 10 percent. This suggests, as

[1] Appendix 2, Section 19; Appendix 3, Section 4 and Part II.

one would expect, that changes in economic parameters, while slow in affecting the differential associated with reduced fertility, work with accelerating speed at later stages.

We turn now from the alleged stability of the income differential associated with reduced fertility to the basic formula by which the size of growth funds is determined, which is open to certain objections. The formula is fundamental because government outlays plus private investment are assumed in the model to be the necessary and sufficient generators of growth. The formula is

$$F = C\left[\frac{F_o}{C_o} + a\left(\frac{Y}{C} - \frac{Y_o}{C_o}\right)\right] = aY - \left(\frac{aY_o - F_o}{C_o}\right)C$$

where F is public outlays plus private investment, a is the incremental propensity to save for the economy, Y is national income, C the number of consumers, and the subscript o denotes the base year. The authors assume that a is 30 percent. The $\frac{aY_o - F_o}{C_o}$ term is estimated to be 49.27 for 1956. The formula for the funds available for government outlay plus private investment then becomes

$$F = .3Y - 49.27\ C$$

The figure for India's incremental propensity to save has subsequently been shown to be too high. But the real question is not what figure to put into the formula, but whether in a country like India savings can reasonably be made a function exclusively of total income levels and of income per head. There are three main savings sectors in the Indian economy: government, individuals and small businesses, and corporations. Coale and Hoover exclude non-monetized savings and investment — a serious omission in the Indian setting. The personal sector accounts for somewhat less than three-sevenths of the total monetized savings, government for most of the rest, and corporations for a small but growing share. For each of these sectors the Coale and Hoover assumption that savings become an increasing proportion of rising income per head is open to grave doubts.

412

The assumption of a growing savings ratio for individuals is retained without any attempt to render it plausible, and despite evidence of a tendency to constancy of this ratio in other countries. Coale and Hoover reject this evidence on the ground that their projections are concerned not so much with change over time as with the income differential at any one time, according to different population assumptions. But their assumptions imply at least a doubling, and in some cases a tripling, of the average propensity to save between 1956 and 1986. If, instead, we were to assume a constant average propensity to save, total savings by individuals and small businesses would be the same for any given total income, irrespective of population size and income per head. Such an assumption would greatly reduce the superiority of the low-fertility projection over the high-fertility one.

The assumption that government savings increase as income per head increases is also questionable. The size of government savings, primarily through taxes but also through the surpluses of public enterprises, depends on many factors other than income per head. As the Indian Taxation Enquiry Commission and the Second and Third Five Year Plans have made clear, India's prevailing tax structure does not guarantee that tax receipts will increase even in proportion to the rise in income per head in the next thirty years. Progressive income tax rates play a much less important role than they do in Western countries and are not effectively enforced. The future proportion of national income saved by the Indian government depends much more on the kind of tax laws, the quality of the tax administration, and the political acceptability of the laws than on income per head.

Similarly, the amount of savings originating in the corporation sector depends largely on the opportunities for profitable investment, which in turn depend on a large number of forces besides growth of income per head. If, in view of these criticisms, a recalculation is made, for instance, on the assumption of a constant ratio of savings to income, it reduces the differential in total income between the low- and high-fertility assumption by some two-thirds.

Nothing of what has been said must be understood to suggest that savings and investment cannot or should not constitute an increasing proportion of rising income per head. Of course they can and they should. But it is important to remember that this does not happen automatically. In a model that abstracts from government policy and other "non-economic" determinants of growth, it would be more plausible to make savings depend on total income than on both total income and income per head. Again, it may be that rapidly growing income per head contributes to the breakdown of traditional attitudes, to a change in the pattern of demand, and to changes in the attitudes of active and potential businessmen toward innovation. Moreover, it is easier for governments to raise tax rates if incomes per head are rising. But none of this is consistent with assuming that these other conditions will remain constant while the average savings propensity will rise. Nor can one assume, as Coale and Hoover do, that all "non-economic" factors, such as government policy, will automatically adjust to rising income per head, or that government policies to raise the savings level cannot, or should not, precede such growth in income per head. Savings in general and taxes in particular are functions of public policy and of many interrelated conditions and cannot be relied on to rise without changes in these. The abstraction from government policies and institutional changes cannot yield the required rise in the savings ratio.

We have concentrated on the assumptions regarding the savings ratio. Similar considerations apply to the proportion of funds devoted to welfare expenditures. Policies in this field, especially those relating to the capital-intensive sectors of housing and construction and the dispersal or concentration of towns and industries, can have widely varying effects on economic growth. These variations influence the capital/output ratio of the welfare expenditures and, therefore, for given welfare requirements, the proportion of funds that can be allocated to direct growth investment. Once again a choice between different types of policies opens up a substantially wider range in the income differential for different assumptions.

To conclude, it is, of course, a considerable advance to use different incremental capital/output ratios for different types of investment and different types of developmental expenditure, including expenditure that improves the quality of the labor force, and to vary these ratios over time. Nevertheless, the arbitrary selection of parameters yielding the minimum range in the differential, the setting of the analysis in the frame of the relationship between capital input and product output – with its inherent bias toward overemphasis on investment and toward abstraction from other conditions, and thus toward isolation from all other policies and the consequent ignoring of the importance of how policies are combined – and the abstraction from effects of different levels of consumption on labor input and labor efficiency create the illusion that the demographic model so amplified can yield more inferences about the economy than its limitations permit. And even so we have not spelled out our objections to the main aggregative concepts used.

In our analysis of the economic consequences of population trends in Part I of Chapter 28, we began by experimenting with the construction of numerical models of the type criticized here. When finally we gave up the attempt and found ourselves restricted to much more general conclusions,[1] this was due not only to the lack of empirical data but also to the logical imperfections described briefly in this appendix.

Indeed, the point of our criticism of both the simple model and the more inclusive one is not the lack or inadequacy of empirical data for the estimation of the parameters intended to trace the economic effects of population movements. Were this the only deficiency, either model could still serve as a "theory" in our sense: a logically coordinated system of questions to be answered by further research. Such systems have their use in the organization and direction of research. The criticism is rather that these models are not appropriate systems of questions and so are not adequate to the South Asian reality. Their conclusions do not add to our knowledge since they are contained in the questionable simplifications of their premises.

We would not exclude the possibility of constructing models more adequate for this purpose. But such models would have to contain many more parameters and account for many more interrelationships. They would have to be very much more complex in order to be logically consistent and adequate to reality. With the present dearth of empirical data, indulging in this type of preparatory macroanalysis does not seem to be a rewarding endeavor.

[1] See, in particular, Chapter 28, Section 2.