

# Education and *income distribution* in urban Brazil, *1976-1996*

---

**Francisco H.G. Ferreira**

*Professor, Department of  
Economics,  
Catholic University of  
Rio de Janeiro  
fferreira@econ.puc-rio.br*

**Ricardo Paes de Barros**

*Institute of Applied  
Economic Research (IPEA),  
Rio de Janeiro  
rbarros@ipea.gov.br*

Despite tremendous macroeconomic instability, Brazil's urban income distributions in 1976 and 1996 appear, at first glance, deceptively similar. Mean household income per capita was stagnant, with a minute accumulated growth of 4.3% over the two decades. The Gini coefficient hovered just above 0.59 in both years, and the incidence of poverty (with respect to a poverty line of R\$60/month at 1996 prices) was effectively unchanged at 22%. Yet, behind this apparent stability, a powerful combination of labour market, demographic and educational dynamics were at work, one effect of which was to generate a substantial increase in extreme urban poverty. Using a micro-simulation-based decomposition methodology which endogenizes labour incomes, individual occupational choices and education decisions, we show that the distribution of incomes was being affected, on the one hand, by a decline in average returns to both education and experience and by impoverishing changes in the structure of occupations and labour force participation (all of which tended to increase poverty), and on the other hand by an increase in educational endowments across the distribution and a progressive reduction in dependency ratios (both of which tended to reduce poverty).

# I

## Introduction

Both by the standards of its own previous growth record, during the ‘Brazilian miracle’ years of 1968-1973, and by those of other leading developing countries thereafter, notably in Asia, the two decades between 1974 and 1994 (i.e., between the first oil shock and the return of stability with the Plano Real) were dismal for Brazil. First and foremost, they were characterized by persistent macroeconomic disequilibrium, the main symptoms of which were stubbornly high and accelerating inflation and a GDP time-series marked by unusual volatility and a very low positive trend.

The macroeconomic upheavals during this period involved three price and wage freezes (during the Cruzado Plan of 1986, the Bresser Plan of 1987 and the Verão Plan of 1989), all of which were followed by higher inflation rates; one temporary financial asset freeze (with the Collor Plan of 1990); and finally a successful currency reform followed by the adoption of a nominal anchor in 1994 (the Plano Real). The national currency changed its name four times.<sup>1</sup> Throughout the period, macroeconomic policy was almost without exception characterized by relative fiscal laxity and growing monetary stringency.

In addition, substantial structural changes were taking place. Brazil’s population grew by 46.6% between 1976 and 1996 and also became more urban (the urbanization rate rose from 68% to 77%). The average level of education of the population over 10 years of age rose

from 3.2 to 5.3 effective years of schooling.<sup>2</sup> Open unemployment grew steadily more prevalent. The sectoral composition of the labour force changed away from agriculture and manufacturing, and towards services. The degree of formalization of the labour force declined substantially: the proportion of formal workers (wage workers with formal documentation) was almost halved, from just under 60% to just over 30% of all workers. And yet, despite the macroeconomic turmoil and continuing structural changes, a casual glance at the indicators of inequality and headline poverty given in table 1 might suggest that little had changed in Brazilian urban income distribution between 1976 and 1996.

As is often the case, however, casual glances may turn out to be misleading. This apparent distributional stability belies a number of powerful, and often countervailing, changes in four areas: the returns to education in the labour markets; the distribution of educational achievements over the population; the pattern of occupational choices; and the demographic structure resulting from household fertility choices. In this paper, we note two ‘puzzles’ about the evolution of Brazil’s urban income distribution in the 1976-1996 period, and suggest explanations for them.

The first puzzle is posed by the combination of (slow) growth in mean incomes and stable or slightly declining inequality, on the one hand, and rising extreme poverty on the other. We argue that this can only be explained by the growth in the size of a group of very poor households, who appear to be effectively excluded both from the labour markets and from the system of formal safety nets. This group is trapped in indigence at the very bottom of the urban Brazilian income distribution, and contributes to the rises in poverty measures which are particularly sensitive to the depth –P(1)– and severity –P(2)– of poverty, particularly when poverty is defined with respect to a low poverty line.<sup>3</sup> This is starkly cap-

□ This is a summarized version of “The slippery slope: Explaining the increase in extreme poverty in urban Brazil, 1976-1996”, which was published in *The Brazilian Review of Econometrics* in November 1999. We are very grateful to François Bourguignon for his guidance and support and to James Heckman, Nora Lustig, Naércio Menezes Filho and participants at the LACEA 1998 conference in Buenos Aires, the AEA 2000 conference in Boston, and a seminar at Cornell University for helpful comments. We are also grateful to the World Bank for financial support. We owe a special debt of gratitude to Philippe George Leite, Roberta Barreto, Carlos Henrique Corseuil, Sérgio Firpo, Luis Eduardo Guedes, Cristiana Lopes, Vanessa Moreira, Daniele Reis and Alinne Veiga for their superb research assistance.

<sup>1</sup> The changes were from Cruzeiro to Cruzado in 1986; from Cruzado to Novo Cruzado in 1989; from Novo Cruzado back to Cruzeiro in 1990, and from Cruzeiro to Real in 1994.

<sup>2</sup> ‘Effective’ years of schooling are based on the last grade completed, and are thus net of repetition. All the figures are from Ferreira and Paes de Barros, 1999.

<sup>3</sup> All poverty measures reported in this paper are the P( $\alpha$ ) class of decomposable measures developed by Foster, Greer and Thorbecke (1984). An increase in  $\alpha$  implies an increase in the weight given to the distance between people’s incomes and the poverty line.

TABLE 1

## Brazil: General economic indicators, 1976, 1981, 1985 and 1996

	1976	1981	1985	1996
Annual per capita GNP (in constant 1996 Reals)	4,040	4,442	4,540	4,945
Annual inflation rate <sup>a</sup>	42%	84%	190%	9%
Open unemployment <sup>b</sup>	1.82%	4.26%	3.38%	6.95%
Average years of schooling <sup>c,d</sup>	3.23	4.01	4.36	5.32
Rate of urbanization <sup>d</sup>	67.8%	77.3%	77.3%	77.0%
Self-employed as share of labour force <sup>d</sup>	27.03%	26.20%	26.19%	27.21%
Share of formal employment <sup>d,e</sup>	57.76%	37.97%	36.41%	31.51%
Mean (urban) household per capita income <sup>d</sup>	265.10	239.08	243.15	276.46
Inequality (Gini) <sup>d</sup>	0.595	0.561	0.576	0.591
Inequality (Theil - T) <sup>d</sup>	0.760	0.610	0.657	0.694
Poverty incidence (R\$ 30/month) <sup>f</sup>	0.0681	0.0727	0.0758	0.0922
Poverty incidence (R\$ 60/month) <sup>f</sup>	0.2209	0.2149	0.2274	0.2176

<sup>a</sup> Percent, from January to December. Based on the IGP-DI for 1976, and on the INPC-R for all other years.

<sup>b</sup> Based on the IBGE Metropolitan Unemployment Index.

<sup>c</sup> For all individuals 10 years of age or older, in urban areas.

<sup>d</sup> Calculated by the authors from the urban national household survey samples.

<sup>e</sup> Defined as the number of employees 'com carteira' as a fraction of the sum of all wage employees and self-employed workers.

<sup>f</sup> Urban only, monthly and spatially deflated. Expressed in constant 1996 Reals.

tured by figure 1 in section II below, which plots the observed (truncated) Pen parades for the four years being studied.<sup>4</sup> The main endogenous channel through which the marginalization of this group is captured in our model is a shift in their occupational 'decisions' away from either wage or self-employment towards unemployment or exit from the labour force.

The evidence which we will examine in section IV below reveals downward shifts in the earnings-education profile, controlling for age and gender, in both the wage-earning and self-employment sectors.<sup>5</sup> Although the profile is slightly convex, the magnitude of the shift implies a decline in the (average) rate of return to education for all relevant educational levels. Similarly, average returns to experience also fell unambiguously for 0-50 years of experience (see figure 5). The combined effect of these changes –the 'price effect'– was an increase in simulated poverty for all measures and for both lines. Simulated inequality also rose, albeit much more mildly. Both effects were exacerbated when the changes (up to 1996) in the determinants of labour force participation decisions were also taken into account.

The second puzzle, then, is what forces counterbalance these price and occupational choice effects, so as

to explain the observed stability in inequality and 'headline' poverty.<sup>6</sup> We find that they were fundamentally the combination of increased education endowments, which move workers up along the flattening earnings-education slope, together with an increase in the correlation between family income and family size, caused by a more than proportional reduction in dependency ratios and family sizes for the poor. This demographic factor had both direct effects on per capita income –through a reduction in the denominator– and indirect effects, through participation decisions and higher incomes.

We addressed these issues by means of a micro-simulation-based decomposition of distributional changes, developed by Bourguignon, Ferreira and Lustig (1998), which itself builds upon the work of Almeida dos Reis and Paes de Barros (1991) and Juhn, Murphy and Pierce (1993). The approach has two distinguishing features. First, unlike other dynamic inequality decompositions, such as that proposed by Mookherjee and Shorrocks (1982), it decomposes the effects of changes on an entire distribution, rather than on a scalar summary statistic (such as the mean log deviation). This allows for much greater versatility: within the same framework, a wide range of simulations can be performed to investigate the effects of changes in specific parameters on any number of inequality or poverty measures (and

<sup>4</sup> The idea of "parades" of income recipients was developed by Jan Pen. Pen parades are the mathematical inverse of distribution functions; that is: they plot the incomes earned by each person (or group of persons), when these are ranked by income.

<sup>5</sup> This shift is from 1976 to 1996, and takes place after upward shifts in the 1980s. See Figure 4.

<sup>6</sup> By 'headline' poverty, we mean the incidence of poverty computed with respect to the R\$60/month poverty line.

then for any number of poverty lines or assumptions about equivalence scales). Second, the evolving distribution which it decomposes is a distribution of household incomes per capita (with the recipient unit generally being the individual). Therefore, moving beyond pure labour market studies, the effect of household composition on living standards and participation decisions is explicitly taken into account. As it turns out, these are of great importance for a fuller understanding of the dynamics involved.

The remainder of the paper is organized as follows. Section II briefly reviews the main findings of the litera-

ture on income distribution in Brazil over the period studied and presents summary statistics and dominance comparisons for the four observed distributions we analyze: 1976, 1981, 1985 and 1996. Section III outlines how the basic model of Bourguignon, Ferreira and Lustig (1998) was adapted to the case of Brazil. Section IV presents the results of the estimation stage and discusses some of their implications. Section V presents the main results of the simulation stage and decomposes the observed changes in poverty and inequality. Finally, section VI presents the conclusions and draws some policy implications.

## II

### Income distribution in Brazil from 1976 to 1996: a brief review of the literature and of our data set

There is little disagreement in the existing literature about the broad trends in Brazilian inequality since reasonably reliable data first became available in the 1960s. The Gini coefficient rose substantially during that decade, from around 0.500 in 1960 to 0.565 in 1970 (see Bonelli and Sedlacek, 1989). There was a debate over the causes of this increase, spearheaded by Albert Fishlow on the one hand, and Carlos Langoni on the other, but there was general agreement that the 1960s had seen substantially increased dispersion in the Brazilian income distribution.<sup>7</sup>

The 1970s displayed a more complex evolution. Income inequality rose between 1970 and 1976, reached a peak in that year, but then fell – both for the distribution of total individual incomes in the economically active population (EAP) and for the complete distribution of household per capita incomes – from 1977 to 1981 (see Bonelli and Sedlacek (1989); Hoffman (1989) and Ramos (1993)). The recession year of 1981 was a local minimum in the inequality series, whether measured by the Gini or the Theil-T index. After 1981, inequality rose during the recession years of 1982 and 1983. Some authors report small declines in some indices in 1984, but the increase was resumed in 1985. 1986, the year of the Cruzado Plan, saw a break in the series, caused both by

a sudden (if short-lived) decline in inflation and a large increase in reported household incomes. Stability and economic growth led to a decline in measured inequality, according to all authors. Thereafter, however, with the failure of the Cruzado Plan stabilization attempt and the return to stagflation, inequality resumed its upward trend, with the Gini finishing the decade at 0.606.

The general trends identified in the existing literature are mirrored in the statistics for the years with which we concern ourselves in this paper, namely 1976, 1981, 1985 and 1996. The distributions for each of these years come from the Pesquisa Nacional por Amostra de Domicílios (National Household Survey –PNAD), run by the Brazilian Geographical and Statistical Institute (IBGE). Except where otherwise explicitly specified, we deal with distributions for urban areas only, where the welfare concept is total household income per capita (in constant 1996 Reals, spatially deflated to adjust for regional differences in average cost of living) and the unit of analysis is the individual. Details of the PNAD sampling coverage and methodology, sample sizes, the definition of key income variables, spatial and temporal deflation issues, and adjustments with respect to the National Accounts baseline are discussed in Ferreira and Paes de Barros (1999).

Table 2 below presents a number of summary statistics for these distributions, in addition to the mean, which was given in table 1 above. The four inequality indices used throughout this paper are the Gini coefficient and three members of the Generalized Entropy

<sup>7</sup> The Fishlow-Langoni debate concerned the importance of education vis-a-vis repressive labour market policies in determining the high level of Brazilian inequality. See, for example, Fishlow (1972), Langoni (1973) and Bacha and Taylor (1980).

TABLE 2

**Brazil: Basic distributional statistics for different degrees of household economies of scale**

	1976	1981	1985	1996
Median (1996 R\$) <sup>a</sup>	127.98	124.04	120.83	132.94
<b>Inequality</b>				
Gini - $\theta = 1.0$	0.595	0.561	0.576	0.591
Gini - $\theta = 0.5$	0.566	0.529	0.548	0.567
E(0) - $\theta = 1.0$	0.648	0.542	0.588	0.586
E(0) - $\theta = 0.5$	0.569	0.472	0.524	0.534
E(1) - $\theta = 1.0$	0.760	0.610	0.657	0.694
E(1) - $\theta = 0.5$	0.687	0.527	0.580	0.622
E(2) - $\theta = 1.0$	2.657	1.191	1.435	1.523
E(2) - $\theta = 0.5$	2.254	0.918	1.134	1.242
<b>Poverty - R\$30/ month</b>				
P(0) - $\theta = 1.0$	0.0681	0.0727	0.0758	0.0922
P(0) - $\theta = 0.5$	0.0713	0.0707	0.0721	0.0847
P(1) - $\theta = 1.0$	0.0211	0.0337	0.0326	0.0520
P(1) - $\theta = 0.5$	0.0235	0.0315	0.0303	0.0442
P(2) - $\theta = 1.0$	0.0105	0.0246	0.0224	0.0434
P(2) - $\theta = 0.5$	0.0132	0.0226	0.0204	0.0357
<b>Poverty - R\$60/ month</b>				
P(0) - $\theta = 1.0$	0.2209	0.2149	0.2274	0.2176
P(0) - $\theta = 0.5$	0.2407	0.2229	0.2382	0.2179
P(1) - $\theta = 1.0$	0.0830	0.0879	0.0920	0.1029
P(1) - $\theta = 0.5$	0.0901	0.0875	0.0927	0.0960
P(2) - $\theta = 1.0$	0.0428	0.0525	0.0534	0.0703
P(2) - $\theta = 0.5$	0.0471	0.0508	0.0521	0.0625

<sup>a</sup> For urban areas only, spatially deflated.

Class of inequality indices,  $E(\phi)$ . Specifically, we have chosen  $E(0)$ , also known as the mean log deviation or the Theil–L index;  $E(1)$ , more commonly known as the Theil–T index, and  $E(2)$ , which is half of the square of the coefficient of variation. These provide a useful range of sensitivities to different parts of the distribution.  $E(0)$  is more sensitive to the bottom of the distribution, while  $E(2)$  is more sensitive to higher incomes.  $E(1)$  is somewhere in between, whereas the Gini places greater weight around the mean.

We also present three poverty indices from the Foster-Greer-Thorbecke additively decomposable class  $P(\alpha)$ .  $P(0)$ , also known as the headcount index, measures poverty incidence,  $P(1)$  is the normalized poverty deficit, and  $P(2)$  is an average of squared normalized deficits, thus placing greater weight on incomes furthest from the poverty line. We calculate each of these with respect to two poverty lines, representing R\$1 and R\$2 per day at 1996 prices.<sup>8</sup>

<sup>8</sup> At 1996 market exchange rates, this was roughly equal to US\$1 and US\$2. In real terms, this would be slightly less than the conventional poverty lines of PPP US\$ 1 and 2 valued at 1985 prices which the World Bank often uses for international comparisons, due to US inflation in the intervening decade.

Each of these poverty and inequality indices is presented both for the (individual) distribution of total household incomes per capita and for an equivalized distribution using the Buhmann, Rainwater, Schmaus and Smeeding (1988) parametric class of equivalence scales (with  $\theta = 0.5$ ). This provides a rough test that the trends we describe are robust to different assumptions about the degree of economies of scale in consumption within households. As usual, per capita incomes generate an upper limit for inequality measures, whereas allowing for some proportion of local public goods within households raises the income of (predominantly poor) very large households and lowers inequality. In the case of the poverty measures, the poverty lines were adjusted as follows:  $z^* = z[\mu(n)]^{1-\theta}$ , where  $\mu(n)$  is the mean household size in the distribution (see Deaton and Paxson, 1997).

Table 2 also confirms that the evolution of inequality over the period is marked by a decline from 1976 to 1981 but a subsequent deterioration over the remaining two sub-periods. This trend is robust to the choice of equivalence scale, proxied here by two different values for  $\theta$ , although the inequality levels are always lower when we allow for economies of scale within households.

TABLE 3

**Brazil: Stochastic dominance results**

	1976	1981	1985	1996
1976			F	
1981				L
1985				L
1996				

It is also robust to the choice of inequality measure, at least as regards the inequality increases from 1981 to 1996 and from 1985 to 1996, as the Lorenz dominance results identified in table 3 indicate.

The results for poverty are more ambiguous. With respect to the higher poverty line, the incidence is effectively unchanged throughout the period (and even displays a slight decline for the equalized distribution).  $P(1)$  and  $P(2)$ , however, show increases over the period, and these become both more pronounced and more robust with respect to  $\theta$  as the concavity of the poverty measure increases. This suggests that the depth and severity of poverty, affected mostly by falling incomes at the very bottom of the distribution, were on the rise.

These results are reflected in table 3, where a letter in cell  $(i, j)$  indicates that the distribution for year  $i$  second order or Lorenz stochastically dominates (L) or first-order

der stochastically dominates (F) over that for year  $j$ . 1981 and 1985 both display Lorenz dominance over 1996, as suggested above. There is only one case of first-order welfare dominance throughout the period: social welfare measured in terms of money was unambiguously higher in 1976 than in 1985. Indeed, all poverty measures reported for both of our lines (and for  $\theta = 1.0$ ) are higher in 1985 than in 1976.<sup>9</sup> This is conspicuously not the case for a comparison between 1976 and 1996. Whereas poverty measures very sensitive to the poorest are higher for 1996, poverty incidence for the 'higher' lines (R\$ 60 per month) falls from 1976 to 1996, suggesting a crossing of the distribution functions. Figure 1 shows this crossing, by plotting the Pen parades ( $F^{-1}(y)$ )—truncated at the 60<sup>th</sup> percentile—for all four years analyzed. Note that whereas 1976 lies in all cases above 1985, all other pairs cross. In particular, 1976 and 1996 cross somewhere near the 17<sup>th</sup> percentile.

Before we turn to the model used to decompose changes in the distribution of household incomes, which will shed some light on all of these changes, it will prove

<sup>9</sup> Note that this first-order welfare dominance is not robust to a change in  $\theta$  to 0.5.

FIGURE 1

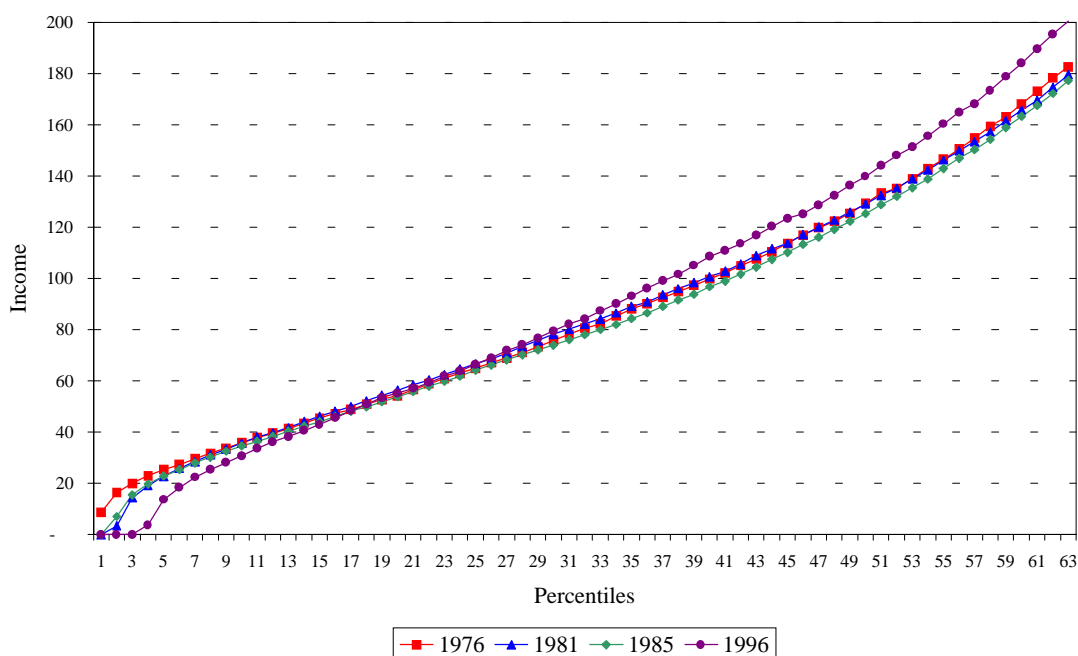
**Brazil: Truncated Pen parades, 1976-1996**

TABLE 4

**Brazil: Educational and labour force participation statistics, by gender and race<sup>a</sup>**

	1976	1981	1985	1996
Average years of schooling (males)	3.32	4.04	4.36	5.20
Average years of schooling (females)	3.14	3.99	4.37	5.43
Average years of schooling (blacks and mixed race)	–	–	–	4.20
Average years of schooling (whites)	–	–	–	6.16
Average years of schooling (Asians)	–	–	–	8.13
Labour force participation (males)	73.36%	74.63%	76.04%	71.31%
Labour force participation (females)	28.62%	32.87%	36.87%	42.00%
Labour force participation (blacks and mixed race)	–	–	–	55.92%
Labour force participation (whites)	–	–	–	56.41%
Labour force participation (Asians)	–	–	–	54.88%

<sup>a</sup> Average 'effective' years of schooling for persons ten years of age or older, in urban areas. Labour force participation refers to urban areas only.

helpful to gather some evidence on the evolution of educational attainment (as measured by average effective years of schooling) and labour force participation, for different groups in the Brazilian population, partitioned by gender and ethnicity. Table 4 presents these statistics.

As may be seen, there was some progress in average educational attainment in urban Brazil over this period. Average effective years of schooling for all individuals ten years or older, as reported in table 1, rose from 3.2 to 5.3. In fact, this piece of good news will prove of vital importance for having prevented a more pronounced increase in poverty. Table 4 reveals that the male-female educational gap has now been eliminated, with females older than ten being on average slightly more educated than males. Clearly, this must imply a large disparity in

favour of girls in recent cohorts. While a cohort analysis of educational trends is beyond the scope of this paper,<sup>10</sup> such a rapid reversal may in fact warrant a shift in public policy towards programmes aimed at keeping boys in school, without in any way discouraging the growth in female schooling. Finally, the remarkable disparity in educational attainment across ethnic groups should be noted, with Asians substantially above average and blacks and those of mixed race below it.

As for labour force participation, the persistent and substantial increase in female participation—from 29% to 42% over the two decades—was partly mitigated by a decline in male participation rates. Those trends notwithstanding, the male-female participation gap remains high, at around 30 percentage points. There is little evidence of differential labour force participation across ethnic groups.

### III

## The model and the decomposition methodology

Let us now turn to the Brazilian version of the general semi-reduced-form model for household income and labour supply developed by Bourguignon, Ferreira and Lustig (1998). It is used here to investigate the evolution of the distribution of household incomes per capita over the two decades from the mid-1970s to the mid-1990s. Specifically, we analyze the distributions for 1976, 1981, 1985 and 1996 and simulate changes between them. As noted above, the paper covers only Brazil's urban areas

(which account for some three-quarters of its population). The general model therefore collapses to two occupational sectors: wage earners and self-employed in urban areas.<sup>11</sup>

<sup>10</sup> See Duryea and Szekely (1998) for such an educational cohort analysis of Brazil and other Latin American countries.

<sup>11</sup> In Brazil, wage earners include employees with or without formal documentation ('com ou sem carteira'). The self-employed are termed own-account workers ('conta propria').

Total household income is given by:

$$Y_h = \sum_{i=1}^n w_i L_i^w + \sum_{i=1}^n \pi_i L_i^{se} + Y_{oh} \quad [1]$$

where  $w_i$  are the total wage earnings of individual  $i$ ,  $L^w$  is a dummy variable that takes a value of 1 if individual  $i$  is a wage earner (and zero otherwise);  $\pi_i$  are the self-employment profits of individual  $i$ ;  $L^{se}$  is a dummy that takes the value of 1 if individual  $i$  is self-employed (and zero otherwise); and  $Y_o$  is income from any other sources, such as transfers or capital incomes. Equation [1] is not estimated econometrically. It aggregates information on the first right-hand-side term from equations [2] and [4], on the second term from equations [3] and [4] and on the third term directly from the household data set.

The wage-earnings equation is given by:

$$\text{Log} w_i = X_i^P \beta^w + \varepsilon_i^w \quad [2]$$

where  $X_i^P = (ed, ed^2, exp, exp^2, D_g)$ .  $ed$  denotes completed effective years of schooling. Experience ( $exp$ ) is defined simply as: age - education - 6, though a more desirable definition would require the age when a person first entered employment, a variable which is not available for 1976.<sup>12</sup>  $D_g$  is a gender dummy, which takes the value 1 for females and zero for males.  $w_i$  are the monthly earnings of individual  $i$ .  $\varepsilon_i$  is a residual term which captures any other determinant of earnings, including any unobserved individual characteristics, such as innate talent. This extremely simple specification was chosen so as to make the simulation stage of the decomposition feasible, as described below. Analogously, the self-employed earnings equation is given by:

$$\text{Log} \pi_i = X_i^P \beta^{se} + \varepsilon_i^{se} \quad [3]$$

Equations [2] and [3] are estimated by simple ordinary least squares (OLS). Equation [2] is estimated for all employees, whether or not heads of household and whether or not they have formal sector documentation ('com or sem carteira'). Equation [3] is estimated for all self-employed individuals (whether or not heads of households). Because the errors  $\varepsilon$  are unlikely to be independent of the exogenous variables, a sample selec-

tion bias correction procedure might be used. However, the standard Heckman procedure for sample selection bias correction requires equally strong assumptions about the orthogonality between the error terms  $\varepsilon$  and  $\xi$  (from the occupational choice multinomial *logit* below). The assumptions required to validate OLS estimation of [2] and [3] are no more demanding than those required to validate the results of the Heckman procedure. We assume, therefore, that all errors are independently distributed, and do not correct for sample selection bias in the earnings regressions.

We now turn to the labour force participation model. Because we have a two-sector labour market (segmented into the wage employment and self-employment sectors), labour force participation and the choice of sector (occupational choice) could be treated in two different ways. One could assume that the choices are sequential, with a participation decision independent of the occupational choice, and the latter conditional on the former. This approach, which would be compatible with a sequential *probit* estimation, was deemed less satisfactory than one in which individuals face a single three-way choice, between staying out of the labour force, working as employees, or engaging in self-employment. Such a choice can be estimated by a multinomial logit model. According to that specification, the probability of being in state  $s$  ( $= 0, w, se$ ) is given by:

$$P_i^s = \frac{e^{Z_i^s \gamma_s}}{e^{Z_i^0 \gamma_0} + \sum_{j \neq s} e^{Z_i^j \gamma_j}} \quad \text{where } s, j = (0, w, se) \quad [4]$$

where the explanatory variables differ for household heads and other household members, by assumption, as follows.

For household heads:

$$Z_i^h = \left( \begin{array}{c} X_i^P; n_{0-13}, n_{14-65}, n_{>65}, \frac{1}{n_{14-65}} \sum_{-1} D_{14-65} ed, \left[ \frac{1}{n_{14-65}} \sum_{-1} D_{14-65} ed \right]^2 \\ \frac{1}{n_{14-65}} \sum_{-1} D_{14-65} age, \left[ \frac{1}{n_{14-65}} \sum_{-1} D_{14-65} age \right]^2, \frac{1}{n_{14-65}} \sum_{-1} D_{14-65} Gd, D \end{array} \right)$$

Note that this is essentially a reduced-form model of the labour supply, where own earnings are replaced by the variables that determine them, according to [2] or [3].

For other members of the household:

$$Z_i^h = \left( \begin{array}{c} X_i^P; n_{0-13}, n_{14-65}, n_{>65}, \frac{1}{n_{14-65}} \sum_{-1} D_{14-65} ed, \left[ \frac{1}{n_{14-65}} \sum_{-1} D_{14-65} ed \right]^2 \\ \frac{1}{n_{14-65}} \sum_{-1} D_{14-65} age, \left[ \frac{1}{n_{14-65}} \sum_{-1} D_{14-65} age \right]^2, \frac{1}{n_{14-65}} \sum_{-1} D_{14-65} Gd, D_1^{se}, L_1^w, D \end{array} \right)$$

<sup>12</sup> Since education is given by the last grade completed, and is thus net of repetition, this definition will overestimate the experience of those who repeated grades at school and hence bias the experience coefficient downwards. The numbers involved are not so large as to alter any conclusions on trends, however.



Where  $n_{k-m}$  is the number of persons in the household whose age falls between  $k$  and  $m$ ;  $D_{14-65}$  is a dummy that takes the value 1 for individuals whose age is between 14 and 65;  $D^{se}$  is a dummy for a self-employed head, and the penultimate term is the earnings of a wage-earning head. These last two variables establish a direct conduit for the effect of the head's occupational choice (and possibly income) on the participation decisions of other members.  $D$  is a dummy variable that takes the value of 1 if there are no individuals aged 14-65 in the household. The sums defined over  $\{j\}$  are sums over  $\{\forall i \in h/j\}$ .

The multinomial logit model in [4] corresponds to the following discrete choice process:

$$s = \underset{j}{\text{Argmax}} \{U_j = Z_i^h \gamma_j + \xi_j, j = (0, w, se)\} \quad [5]$$

where  $Z$  is given above, separately for household heads and other members; the expressions  $\xi_j$  are random variables with a double exponential density function, and  $U_j$  may be interpreted as the utility of alternative  $j$ . Once the vector  $\gamma_j$  is estimated by [4] and a random term  $\xi$  is drawn, each individual chooses an occupation  $j$  so as to maximize the above utility function.

Once equations [2], [3] and [4] have been estimated, we have two vectors of parameters for each of the four years in our sample ( $t \in \{1976, 1981, 1985, 1996\}$ ):  $\beta_t$  from the earnings equations for both wage earners and the self-employed (including constant terms  $\alpha_t$ ), and  $\gamma_t$  from the participation equation. In addition, from equation [1] we have  $Y_{oh}$  and  $Y_{ht}$ . Let  $X_{ht} = \{X_i^p, Z_i^h \mid \forall i \in h\}$  and  $\Omega_{ht} = \{\varepsilon_{ip}^w, \varepsilon_{ip}^{se}, \xi_i^j \mid i \in h\}$ . We can then write the total income of household  $h$  at time  $t$  as follows:

$$Y_{ht} = H(X_{ht}, Y_{oh}, \Omega_{ht}, \beta_t, \gamma_t) \quad h = 1, \dots, m \quad [6]$$

Based on this representation, the distribution of household incomes:

$$D_t = \{Y_{1t}, Y_{2t}, \dots, Y_{mt}\} \quad [7]$$

can be rewritten as:

$$D_t = D[\{Y_{ht}, Y_{oh}, \Omega_{ht}\}, \beta_t, \gamma_t] \quad [8]$$

Where  $\{.\}$  refers to the joint distribution of the corresponding variables over the whole population.

We are interested in understanding the evolution of  $D_t$  over time, or possibly that of a set of alternative summary poverty or inequality measures defined on the basis of it. Based on the representation of a distribution given by [8], changes in the distribution of incomes can

be decomposed into price effects ( $\beta$ ), occupational choice effects ( $\gamma$ ), endowment effects ( $X, Y_0$ ) and residual effects ( $\Omega$ ), as outlined in Ferreira and Paes de Barros (1999). The simplest decomposition applies to those arguments which are exogenous to the household: that is,  $\beta$ s,  $\gamma$ s, and the variance of the various residual terms. Changing the values of  $\beta$ s amounts to assuming a change in the rate of return on human capital variables in equation [2] and [3]. We refer to this as a "price effect".

Calculating the price and occupational choice effects is reasonably straightforward, once the relevant exogenous parameters have been estimated. Estimating individual endowment effects requires a further step, since elements of the  $X$  and  $Y$  vectors are jointly distributed, and a change in the value of any one variable must be understood as being conditional on all the other observable characteristics.

Specifically, if we are interested in the effect of a change in the distribution of a single specific variable  $X_k$  on the distribution of household incomes between times  $t$  and  $t'$ , it is first necessary to identify the distribution of  $X_k$  conditional on other relevant characteristics  $X_{-k}$  (and possibly other incomes  $Y_0$ ). This can be done by regressing  $X_k$  on  $X_{-k}$  at dates  $t$  and  $t'$ , as follows:

$$X_{kit} = X_{-kit} \mu_t + u_{kit} \quad [9]$$

where  $k$  is the variable,  $i$  is the individual, and  $t$  is the date. The vector of residuals  $u_{kit}$  represents the effects of unobservable characteristics (assumed to be orthogonal to  $X_{-k}$ ) on  $X_k$ . The vector  $\mu_t$  is a vector of coefficients capturing the dependency of  $X_k$  on the true exogenous variables  $X_{-k}$ , at time  $t$ . For the sake of simplicity, let us assume that the error terms  $u$  are normally distributed with a mean of zero and a common standard deviation  $\sigma_t$ .

The same equation can, of course, be estimated at date  $t'$ , generating a corresponding vector of coefficients  $\mu_{t'}$ , and a standard error of the residuals given by  $\sigma_{t'}$ . We are then ready to simulate the effect of a change in the conditional distribution of  $X_k$  from  $t$  to  $t'$ , by replacing the values of  $X_{kit}$  observed in the sample at time  $t$ , with:

$$X_{kit}^* = X_{-kit} \mu_{t'} + u_{kit} \frac{\sigma_{t'}}{\sigma_t} \quad [10]$$

The contribution of the change in the distribution of the variable  $X_k$  to the change in the distribution of incomes between  $t$  and  $t'$  may now be written as:

$$R_{it}^{x*} = D[\{X_{kit}^*, X_{-kit}, Y_{oh}, \Omega_{ht}\}, \beta_t, \gamma_t] - D[\{X_{kit}, X_{-kit}, Y_{oh}, \Omega_{ht}\}, \beta_t, \gamma_t] \quad [11]$$

In this paper, we perform four regression estimations such as [9], and hence four simulations such as [10]. The four variables estimated are  $X_k = \{n_{0-13}, n_{14-65}, n_{>65}, ed\}$ . In the case of the education regression, the vector of explanatory variables  $X_{-kit}$  was (1, age, age<sup>2</sup>,  $G_d$ , regional dummies).<sup>13</sup> In the case of the regressions with the numbers of household members in certain age intervals as dependent variables, the vector  $X_{-kit}$  was (1, age, age<sup>2</sup>,

ed, ed<sup>2</sup>, regional dummies), where age and education are those of the household head. The simulations permitted by these estimations allow us to investigate the effects of the evolution of the distribution of educational attainment and of demographic structure on the distribution of income. We now turn to the results of the estimation stage of the model.

## IV

### Estimating the model

The results of the OLS estimation of equation [2] for wage earners (formal and informal) are shown in table 5. The static results are not surprising. All variables are significant and have the expected signs. The coefficients of education and their squares are positive and significant. The effect of experience (defined as [age – education – 6]), is positive but concave. The gender dummy (female = 1) is negative, significant and large.

The dynamics are more interesting. Between 1976 and 1996, the earnings-education profile changed shape.

TABLE 5

**Brazil: Wage earnings regression for wage-employees (equation [2])<sup>a</sup>**

Year	1976	1981	1985	1996
Intercept	4.350 (0.0001)	4.104 (0.0001)	3.877 (0.0001)	4.256 (0.0001)
Education	0.123 (0.0001)	0.136 (0.0001)	0.129 (0.0001)	0.080 (0.0001)
Education <sup>2</sup> (x 100)	0.225 (0.0001)	0.181 (0.0001)	0.283 (0.0001)	0.438 (0.0001)
Experience	0.075 (0.0001)	0.085 (0.0001)	0.087 (0.0001)	0.062 (0.0001)
Experience <sup>2</sup> (x 100)	-0.105 (0.0001)	-0.119 (0.0001)	-0.121 (0.0001)	-0.080 (0.0001)
Gender (1 = female)	-0.638 (0.0001)	-0.590 (0.0001)	-0.635 (0.0001)	-0.493 (0.0001)
R <sup>2</sup>	0.525	0.538	0.547	0.474

Source: Calculations by the authors, based on the national household survey - PNAD (IBGE, 1976, 1981, 1985, 1996).

<sup>a</sup> The values in parentheses are P-values.

<sup>13</sup> The regional dummies were for the five Brazilian regions: North, Northeast, Centre-West and South, with Southeast being the default category.

After rising in the late 1970s, the linear component fell substantially from 1981 to 1996. Meanwhile, the coefficient of squared years of schooling fell up to 1981 but then more than doubled up to 1996, ending the period substantially above its initial 1976 value. Overall, the relationship became more convex, suggesting a steepening of marginal returns to education at high levels. However, when we plot the parabola which models the partial earnings-education relationship from equation [2], the lowering of the linear term dominates. The profile shifts upward from 1976 to 1981, and again to 1985, before falling precipitously (while becoming more convex) up to 1996 (see figure 2). The net effect across the entire period was a fall in the cumulative returns to education (from zero to  $t$  years) for the entire range, despite increasing marginal returns at high levels of education. The implications for poverty and inequality are clear, with the education price effect leading, ceteris paribus, to an increase in the former and a decline in the latter.

Returns to experience also increased from 1976 to 1981, and from 1981 to 1985, with a concave pattern and a maximum at around 35 years of experience (figure 3). From 1985 to 1996, however, there was a substantial decline in cumulative returns to experience, even with respect to 1976, up to 50 years of experience. The relationship became less concave, and the maximum returns moved up to around 40 years. Over the entire period, the experience price effect was mildly unequalizing (it contributed to increases in inequality until 1985, after which the effect was reversed) and had serious repercussions in terms of increased poverty.

The one piece of good news is the reduction in the male-female earnings disparity. While female earnings, controlling for both education and experience, remained substantially lower in all four years, suggesting that there

FIGURE 2

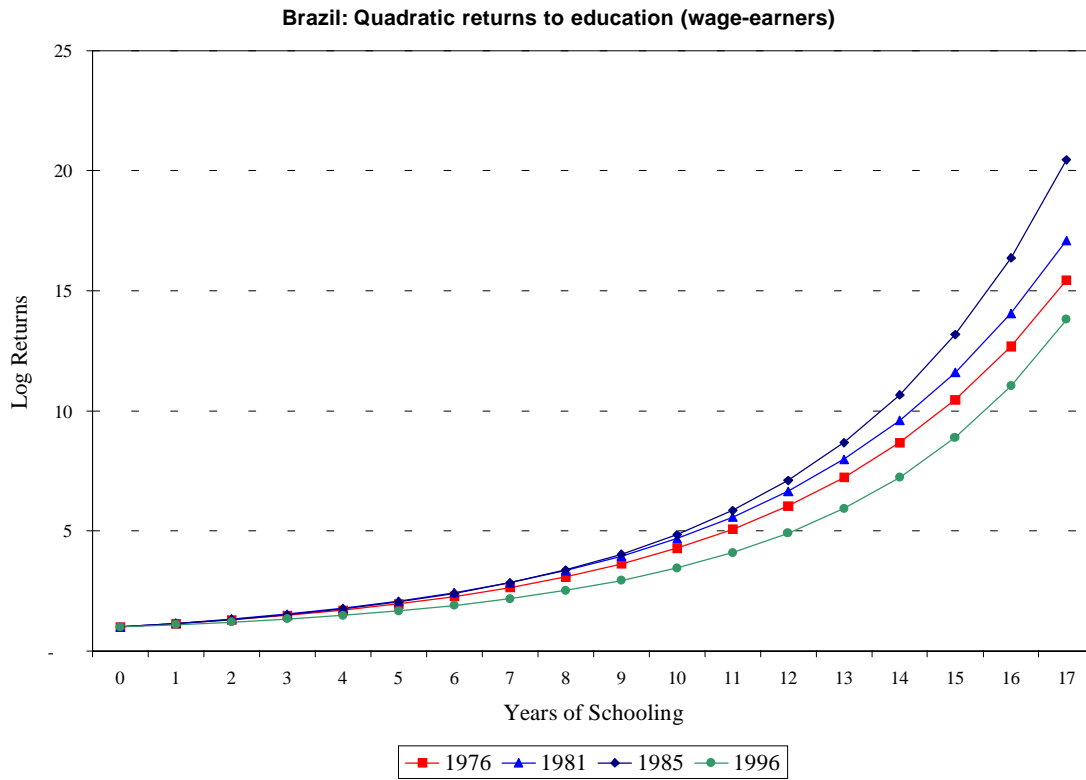
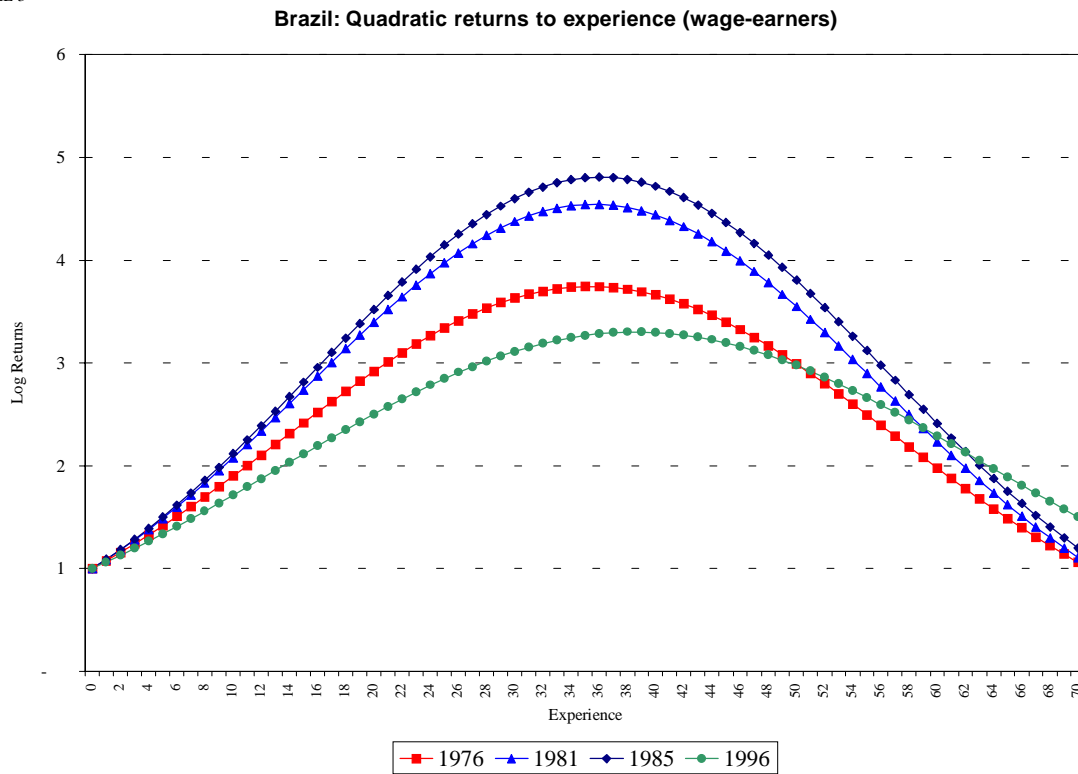


FIGURE 3



may be some labour market discrimination, there was nevertheless a decline in this respect between 1976 and 1996. The effect of this, as we will see from the simulations reported in section V, was both mildly equalizing and poverty-reducing. Let us now turn to equation [3], which seeks to explain the earnings of the self-employed with the same set of independent variables as equation [2]. The results are reported in table 6, which shows that education is also an important determinant of incomes in the self-employed sector. The coefficient of the linear term has a higher value in all years than for wage-earners, but the quadratic term is lower. This implies that, *ceteris paribus*, the return to low levels of education might be higher in self-employment than in wage work, but would eventually become lower as years of schooling increase. This will have an impact on occupational choice, estimated through equation [4]. Dynamically, the same trend was observed as for wage-earners: the coefficient of the linear term fell over time, but the relationship became more convex.<sup>14</sup> The coefficients of experience and experience squared follow a similar pattern to that observed for wage earners. Once again, the cumulative return to experience fell over the bulk of the range from 1976 to 1996, contributing to the observed increase in poverty. The effect of being female (other things being equal) is even more markedly negative in this sector than in the wage-earning sector, though it too fell over the 1976-1996 period, despite a temporary increase in disparity in the 1980s.

A word of caution is in order before proceeding. All of the estimation results reported above refer to equations with total earnings as dependent variables. The changes in coefficients will therefore reflect changes not only in the hourly returns to a given characteristic but also in any supply responses that may have taken place. The analysis should be understood in this light.

Let us now turn to the estimation of the multinomial *logit* in equation [4]. This was estimated separately for household heads and for others, since the set of explanatory variables was slightly different in each case (see the description of vectors  $Z_j$  and  $Z_i$  in section III above).<sup>15</sup>

For household heads, education was not significantly related to the likelihood of choosing to work in the wage sector *vis-à-vis* staying out of the labour force, at any time. The dominant effect on the occupational choices of urban household heads over this period was a sub-

TABLE 6

**Brazil: Total earnings regression for the self employed (equation [3])<sup>a</sup>**

Year	1976	1981	1985	1996
Intercept	4.319 (0.0001)	4.192 (0.0001)	3.853 (0.0001)	4.250 (0.0001)
Education	0.196 (0.0001)	0.148 (0.0001)	0.165 (0.0001)	0.114 (0.0001)
Education <sup>2</sup> (x 100)	-0.206 (0.0001)	0.021 (0.4892)	0.012 (0.6545)	0.219 (0.0001)
Experience	0.074 (0.0001)	0.079 (0.0001)	0.084 (0.0001)	0.063 (0.0001)
Experience <sup>2</sup> (x 100)	-0.101 (0.0001)	-0.108 (0.0001)	-0.111 (0.0001)	-0.082 (0.0001)
Gender	-1.092 (0.0001)	-1.148 (0.0001)	-1.131 (0.0001)	-0.714 (0.0001)
R <sup>2</sup>	0.431	0.434	0.438	0.336

Source: Calculations by the authors, based on the national household survey - PNAD (IBGE, 1976, 1981, 1985 and 1996).

<sup>a</sup> The values in parentheses are P-values.

stantial decline in the constant term affecting the probability of participating in either productive sector, as opposed to remaining outside the labour force or being unemployed. Since it is captured by the constant, this effect is not related to the educational or experience characteristics of the head, or to the endowments of his or her household. We interpret it, instead, as the effect of labour market demand-side conditions, leading to reduced participation in paid work.<sup>16</sup> This effect will be shown, in the occupational choice simulations reported in the next section, to increase both inequality and poverty.

For other members of the household, education did seem to raise the probability of choosing wage work *vis-à-vis* staying out of the labour force, with the relationship changing from concave to convex (albeit weakly so) over the period. It also enhanced the probability of engaging in own-account activities rather than staying outside the labour force in both periods, although this relationship remained concave. The number of children in the household significantly discouraged participation in both sectors, although more so in the wage-earning one. The change in the constant term was much smaller than for household heads, suggesting that negative labour market conditions hurt primary earners to a greater extent. Consequently, the effect of the occupational choices of other household members on poverty and inequality will be seen to be much milder than in the case of heads of household.

<sup>14</sup> In this case, it actually switched from concave to convex.

<sup>15</sup> Space constraints prevent the presentation of the tables reporting these estimations. They may be found in Appendix 3 to Ferreira and Paes de Barros (1999).

<sup>16</sup> In terms of the 'occupational choice' framework, these are changes in the constraints with respect to which those choices are made.

The results of the estimation of equation [9], with the educational level of individuals ten years old or older as the dependent variable, regressed against the vector (1, age, age<sup>2</sup>,  $G_d$ , regional dummies), is also given in Ferreira and Paes de Barros (1999). Over time, there is a considerable increase in the value of the intercept, which will yield higher predicted values for educational attainment, controlling for age, gender and regional location. Furthermore, the gender dummy went from large and negative to positive and significant, suggesting that women have more than caught up with males in educational attainment in Brazil over the last twenty years. The effect of individual age is stable, and regional dis-

parities, with the South and Southeast ahead of the three central and northern regions, persist.

Regressing the number of household members in the age groups 0-13, 14-65 and over 65 (respectively) on the vector (1, ed, ed<sup>2</sup>, age, age<sup>2</sup>, regional dummies) yields the finding that the schooling of the head has a large, negative and significant effect on the demand for children, so that as education levels rise, family sizes would tend to fall, other things being equal. Additionally, some degree of convergence across regions in family size can be inferred, with the positive 1976 regional dummy coefficients for all regions (with respect to the Southeast) declining over time and more than halving in value by 1996.

## V

### The simulation results

Having estimated earnings equations for both sectors of the model (wage-earners [2] and the self-employed [3]); participation equations for both household heads and non-heads [4]; and 'endowment' equations [9] for the exogenous determination of education and family composition, we are now in a position to carry out the decompositions described by Bourguignon, Ferreira and Lustig (1998). These simulations, as discussed earlier, are carried out for the entire distribution. The results are summarized in table 7, in terms of the mean household per capita income  $\mu(y)$ , four inequality indices (the Gini coefficient, the Theil-L index ( $E(0)$ ), the Theil-T index ( $E(1)$ ) and ( $E(2)$ )), and the standard three members of the Foster-Greer-Thorbecke class of poverty measures,  $P(\alpha)$ , with  $\alpha = 0, 1, 2$ , computed with respect to two poverty lines: an indigence line of R\$30 per month and a poverty line of R\$60 per month (both expressed at 1996 São Paulo metropolitan region prices).<sup>17</sup>

Table 7 contains information about a large number of simulated economic changes, always by applying combinations of 1996 coefficients to the 1976 population. In order to address the two puzzles posed in the Introduc-

tion –namely the increase in extreme urban poverty between 1976 and 1996 despite (sluggish) growth and (mild) reduction in inequality and the coexistence of a deteriorating labour market with stable 'headline' poverty– we now plot differences in the (logarithms) of incomes between the simulated distribution (of household incomes per capita) and that observed for 1976, for a number of the simulations in table 7.

Figure 4 plots the combined price effects ( $\alpha$  and  $\beta$ ) separately for wage-earners and the self-employed. As can be seen, these effects were negative (i.e., would have implied lower income in 1976) for all percentiles. The losses were greater for wage earners than for the self-employed, and for the latter they were regressive. These losses are exactly what one would have expected from the downward shifts of the partial earnings-education and earnings-experience profiles given in figures 2 and 3.

In figure 5, we adopt a different approach to the price effects, by plotting the income differences for each individual price effect simulation (for both sectors combined) and then aggregating them. As we would expect from figures 2 and 3, the returns to education and experience both tend to increase poverty. Only the change in partial returns to education is mildly equalizing (as may be seen from table 7). The change in the partial returns to experience tends to increase both inequality and poverty. The change in the intercept, calculated at the mean values of the independent variables, was also negative throughout. This proxies for a 'pure growth' effect, capturing the effects on earnings of processes not captured by edu-

<sup>17</sup> Table 7 and the remaining figures in this chapter refer to the simulation of applying the coefficients estimated for 1996 to 1976. Similar exercises were conducted for 1981 and 1985, and are reported in Ferreira and Paes de Barros (1999). Likewise, the 'return' simulation of applying the 1976 coefficients to 1996 was conducted, and the directions and broad magnitudes of the changes confirmed the results presented here.

TABLE 7

**Brazil: Simulated poverty and inequality for 1976,  
using 1996 coefficients**

	Mean per capita income	Inequality				Poverty <sup>a</sup>					
		Gini	E(0)	E(1)	E(2)	Z = R\$30 / month			Z = R\$ 60 / month		
						P(0)	P(1)	P(2)	P(0)	P(1)	P(2)
1976 observed	265.101	0.595	0.648	0.760	2.657	0.0681	0.0211	0.0105	0.2209	0.0830	0.0428
1996 observed	276.460	0.591	0.586	0.694	1.523	0.0922	0.0530	0.0434	0.2176	0.1029	0.0703
<i>Price effects</i>											
$\alpha, \beta$ for wage earners	218.786	0.598	0.656	0.752	2.161	0.0984	0.0304	0.0141	0.2876	0.1129	0.0596
$\alpha, \beta$ for self-employed	250.446	0.597	0.658	0.770	2.787	0.0788	0.0250	0.0121	0.2399	0.0932	0.0490
$\alpha, \beta$ for both	204.071	0.598	0.655	0.754	2.190	0.1114	0.0357	0.0169	0.3084	0.1249	0.0673
$\alpha$ only, for both	233.837	0.601	0.664	0.774	2.691	0.0897	0.0275	0.0129	0.2688	0.1040	0.0545
All $\beta$ (but not $\alpha$ ) for both	216.876	0.593	0.644	0.736	2.055	0.0972	0.0303	0.0143	0.2837	0.1114	0.0590
Education $\beta$ for both	232.830	0.593	0.639	0.759	2.691	0.0779	0.0234	0.0110	0.2531	0.0953	0.0488
Experience $\beta$ for both	240.618	0.600	0.664	0.771	2.694	0.0851	0.0265	0.0125	0.2592	0.1000	0.0525
Gender $\beta$ for both	270.259	0.595	0.649	0.751	2.590	0.0650	0.0191	0.0090	0.2160	0.0797	0.0404
<i>Occupational choice effects</i>											
$\gamma$ for both sectors (both heads and other members)	260.323	0.609	0.650	0.788	2.633	0.0944	0.0451	0.0331	0.2471	0.1082	0.0671
$\gamma$ for both sectors (only for other members)	265.643	0.598	0.657	0.757	2.482	0.0721	0.0231	0.0119	0.2274	0.0867	0.0454
$\gamma, \alpha, \beta$ for both sectors	202.325	0.610	0.649	0.788	2.401	0.1352	0.0597	0.0402	0.3248	0.1466	0.0902
<i>Demographic patterns</i>											
$\mu d$ only, for all	277.028	0.574	0.585	0.704	2.432	0.0365	0.0113	0.0063	0.1711	0.0554	0.0264
$\mu d, \gamma, \alpha, \beta$ for all	210.995	0.587	0.577	0.727	2.177	0.0931	0.0433	0.0321	0.2724	0.1129	0.0677
<i>Education endowment effects</i>											
$\mu e$ only, for all	339.753	0.594	0.650	0.740	2.485	0.0424	0.0136	0.0073	0.1593	0.0567	0.0287
$\mu d, \mu e$ for all	353.248	0.571	0.584	0.688	2.320	0.0225	0.0078	0.0049	0.1131	0.0359	0.0173
$\mu e, \mu d, \gamma, \alpha, \beta$ for all	263.676	0.594	0.600	0.727	1.896	0.0735	0.0374	0.0296	0.2204	0.0913	0.0561

Source: Based on the national household survey - PNAD (IBGE, 1976 and 1996).

<sup>a</sup> Z = poverty line.

cation, experience, gender, or the unobserved characteristics of individual workers. It is intended to capture the effects of capital accumulation, managerial and technical innovation, macroeconomic policy conditions, and other factors which are likely to determine economic growth but are not included explicitly in the Mincerian equation. Its negative effect in this simulation suggests that these factors increased poverty in urban Brazil over the period studied.

Once again, the only piece of good news comes from the gender simulation, which reports a poverty-reducing effect, as a result of the decline in male-female earnings differentials shown in tables 5 and 6. This effect, however, was far from being sufficient to offset the combined negative effects of the other price effects. As the thick line at the bottom of figure 5 indicates, the combined effect of imposing the 1996 parameters of the two Mincerian equations on the 1976 population was to substantially increase poverty.

Figure 6 plots the (logarithm) of the income differences between the distribution given by imposing the 1996 occupational choice parameters (the  $\gamma$  vector from

the multinomial logit in equation [4]) on the 1976 population, and the observed 1976 distribution. It does so both for all individuals (the lower line), and for non-heads (the upper line). The effect of this simulated change in occupational choice and labour force participation behaviour is to greatly increase poverty and inequality, as the relevant indices in table 7 confirm. It suggests the existence of a group of people who, by voluntarily or involuntarily leaving the labour force, swelling the ranks of the unemployed, or being relegated to very ill-remunerated occupations, probably in the informal sector, are becoming increasingly impoverished.

By combining the negative price and occupational choice effects, we get an idea of the overall effect of Brazil's urban labour market conditions over this period. This is shown in figure 7, where the lowest curve plots the differences between the household per capita incomes from a distribution in which all  $\alpha$ s,  $\beta$ s and  $\gamma$ s change, and the observed 1976 distribution. It shows the substantially poverty-augmenting (and unequalizing) combined effect of changes in labour market prices and occupational choice parameters on the 1976 distribution.

FIGURE 4

Brazil: Combined price effects, by sector

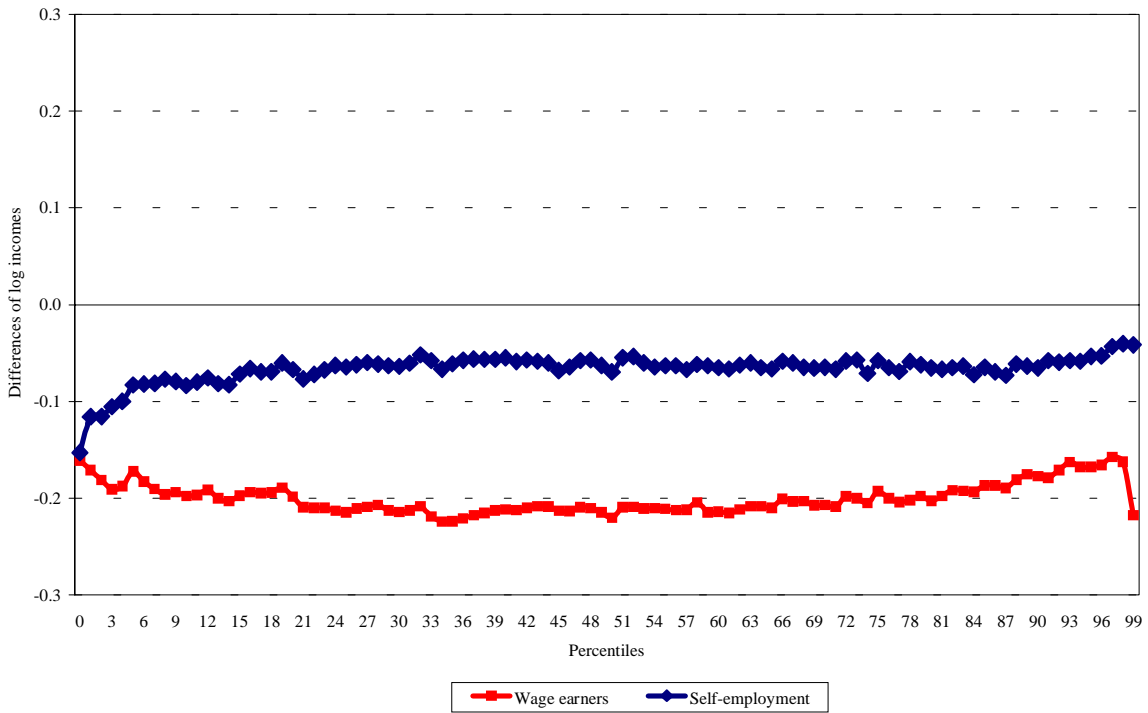
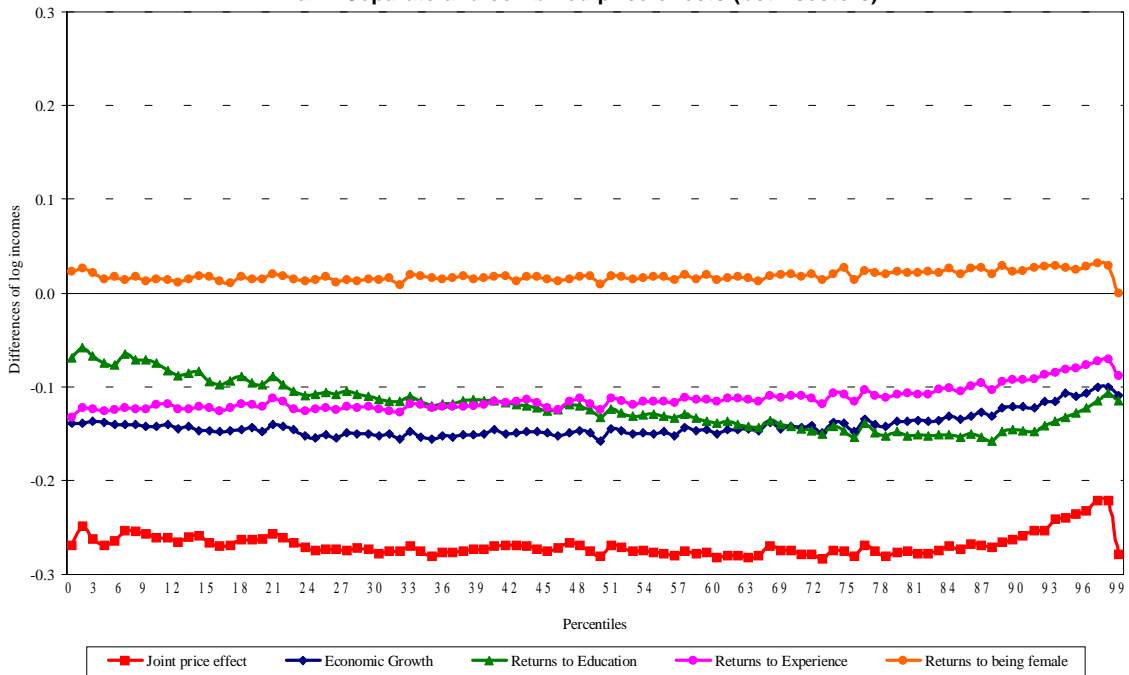


FIGURE 5

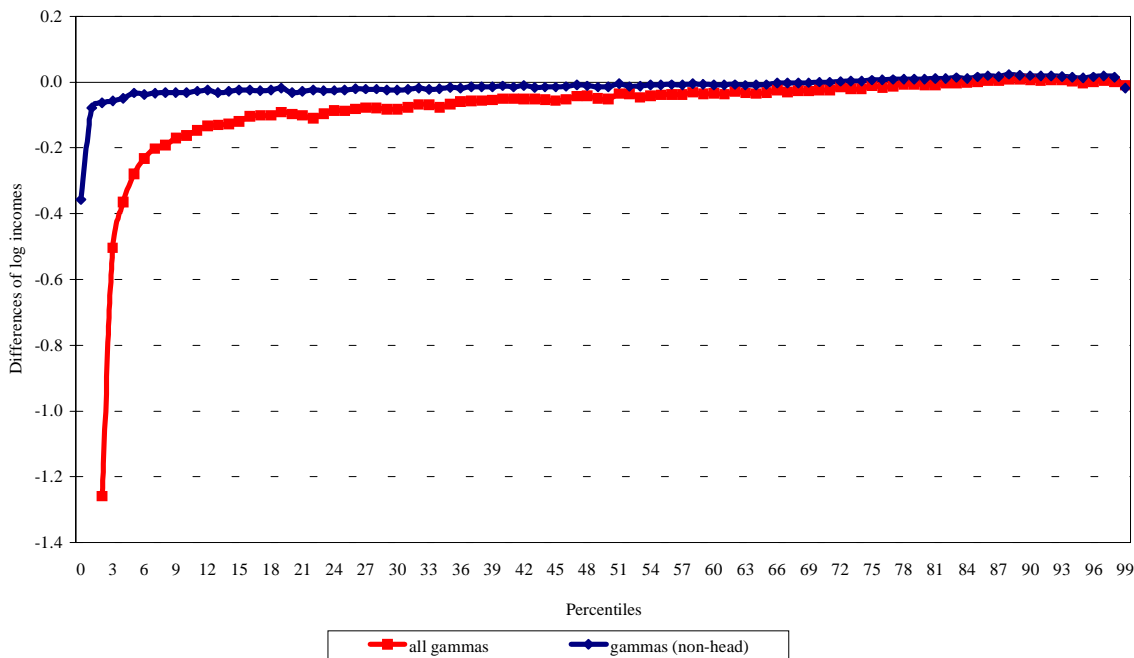
Brazil: Separate and combined price effects (both sectors)



Source: IBGE, 1976 and 1996.

FIGURE 6

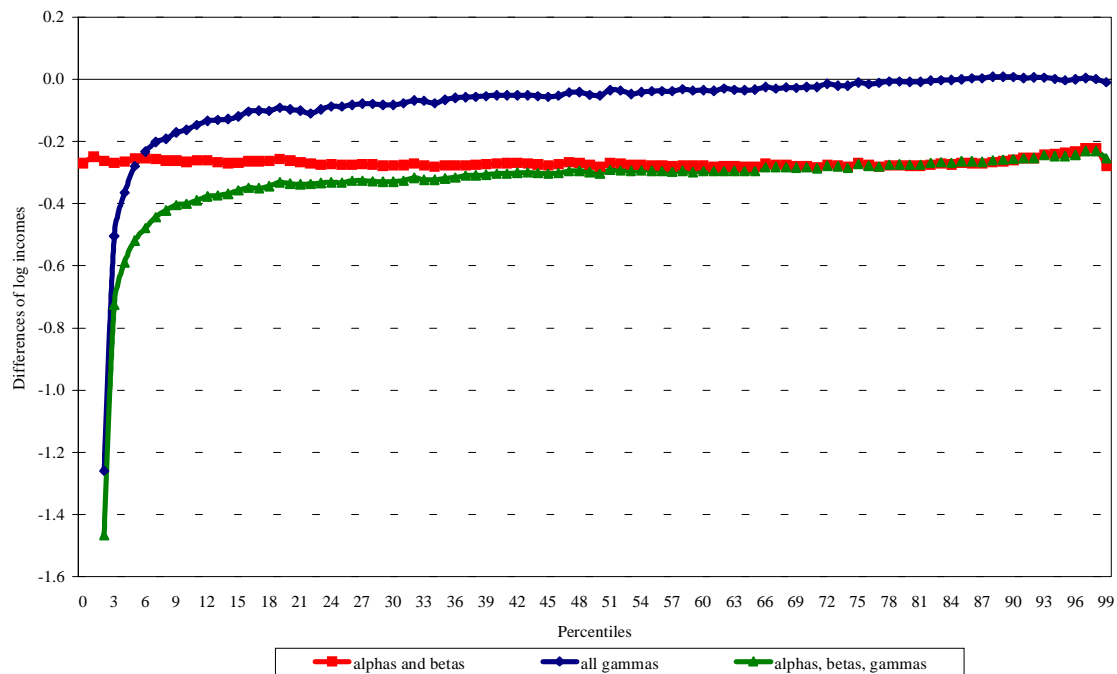
**Brazil: Effects of occupational choice**



Source: IBGE, 1976 and 1996.

FIGURE 7

**Brazil, labour market: Combined price and occupational choice effects**



Source: IBGE, 1976 and 1996.



At this point, the second puzzle can be stated clearly: given these labour market circumstances, what factors can explain why mean incomes rose, headline poverty did not rise, and inequality appears to have fallen slightly? The first part of the answer is shown graphically in figure 8, where the upper line plots the differences between the (log) incomes from a distribution given by imposing on the 1976 population the transformation (10) for the demographic structure of the population. The changes in the parameters  $\mu_d$  (and in the variances of the residuals in the corresponding regression) have a positive effect on incomes for all percentiles, and in an equalizing manner. However, when combined in a simulation in which the values of all  $\alpha$ s,  $\beta$ s and  $\gamma$ s also change, it can be seen that the positive demographic effect is still overwhelmed. Nevertheless, it is clear that the reduction in dependency ratios, and hence in family sizes, in urban Brazil over this period had an important mitigating effect on the distribution of incomes.

There remains one final piece of the puzzle which is needed in order to explain why the deterioration in labour market conditions did not have a worse impact on poverty. This, as should be evident from the increase in mean years of effective schooling registered in table 1, is the shift to the right in the distribution function of

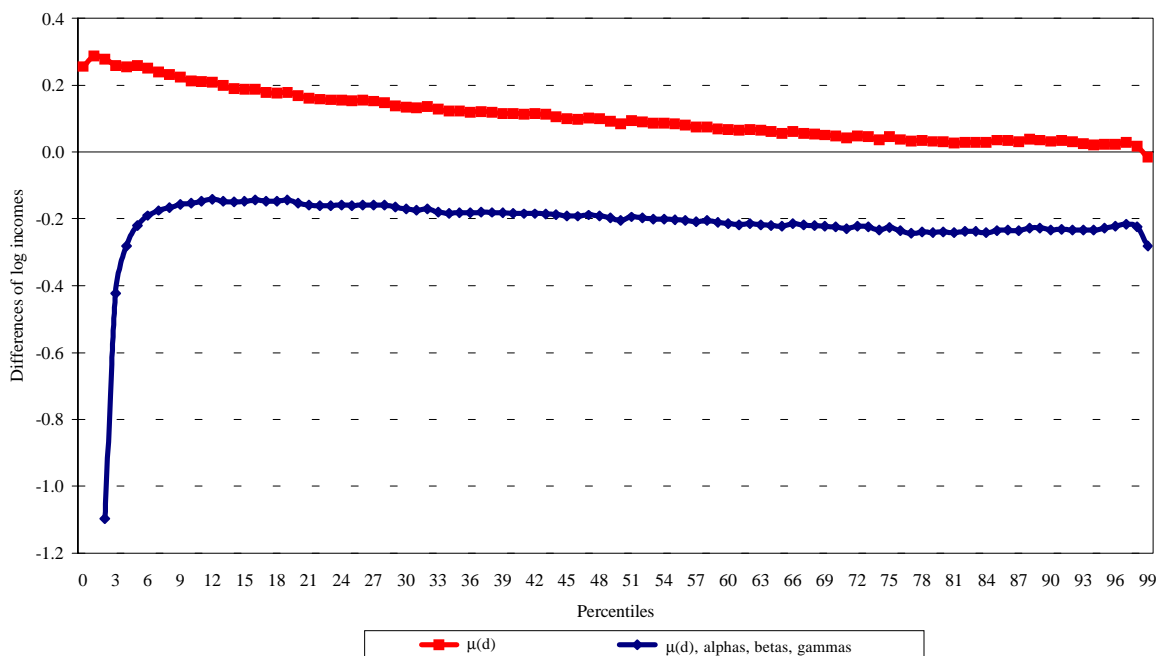
education. This is shown in figure 9, which reveals that gains in educational attainment were particularly pronounced at lower levels of education, and thus presumably among the poor.

A gain in educational endowments across the income distribution, but particularly among the poor, has both direct and indirect effects on incomes. The direct effects are through equations [2] and [3], where earnings are positive functions of schooling. The indirect effects are both through the occupational choices that individuals make, and through the further impact that education has on reducing the demand for children, and hence family size. A simulation of the effect of education is thus quite complex.<sup>18</sup> After it is completed, one observes (figure 10) a rather flat improvement in (log) incomes across the distribution (i.e., a scaling effect). When this is combined with changes in the parameters of the demographic equations, however, the effect becomes more marked, and is not only more poverty-reducing, but also mildly equalizing. The bottom line

<sup>18</sup> Note that the different effects are not simply being summed. The effect of greater educational endowments is simulated through every equation where it appears in the model, thereby affecting fertility choices and occupational statuses, as well as earnings.

FIGURE 8

**Brazil: Demographic effects**



Source: IBGE, 1976 and 1996.

FIGURE 9

Brazil: Shift in distribution of schooling, 1976-1996

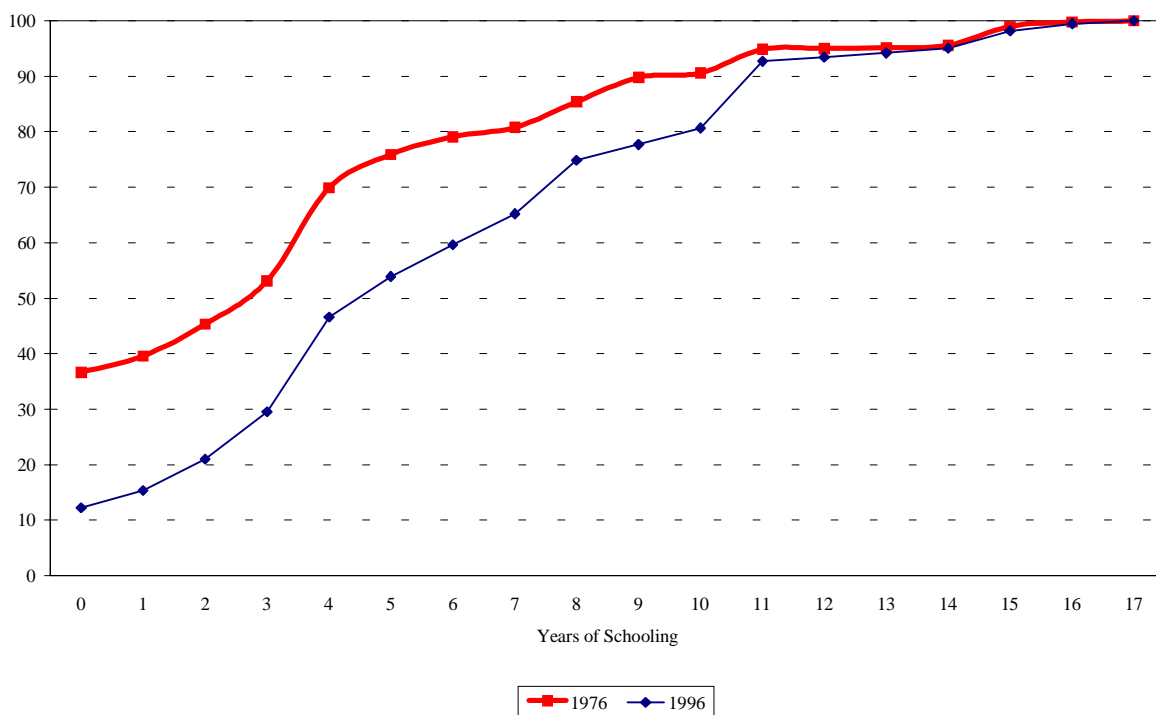
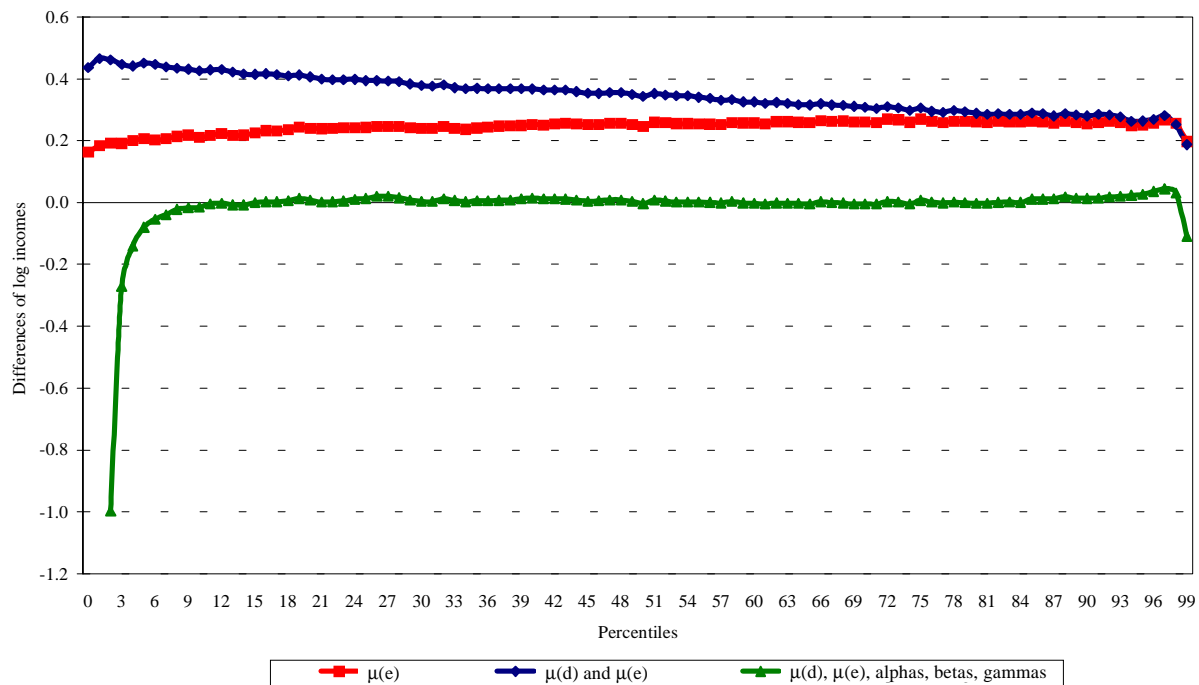


FIGURE 10

Brazil: Educational endowments and demographic effects



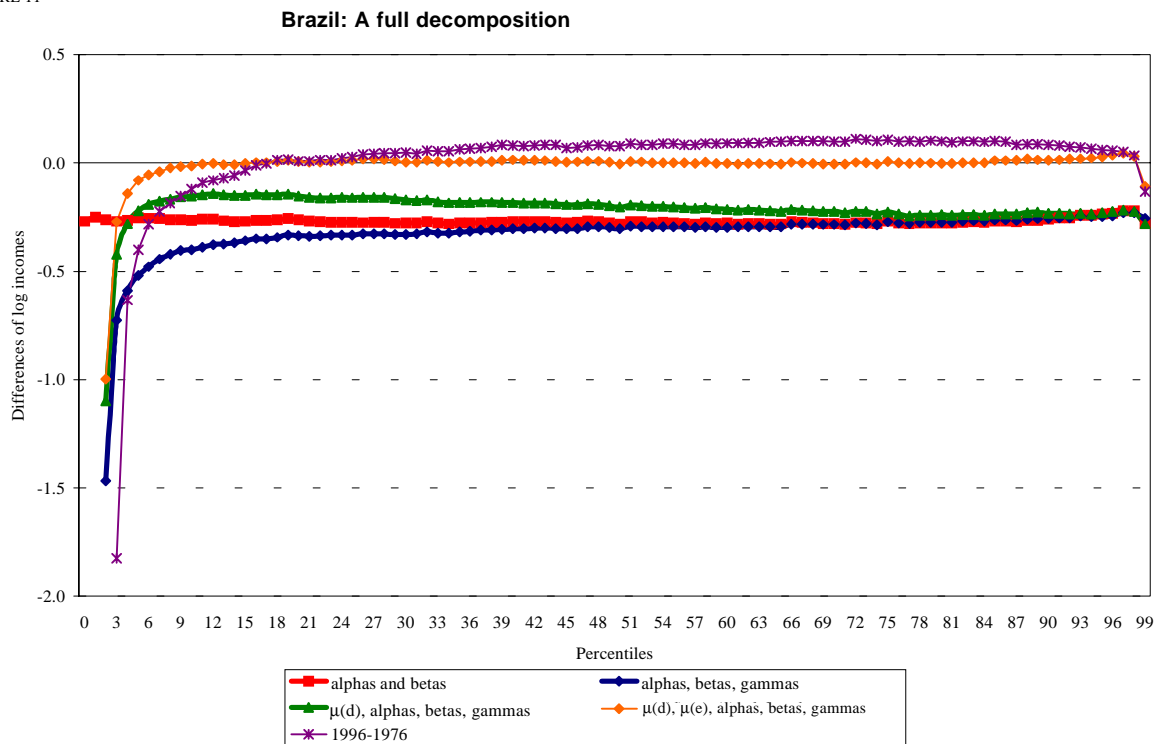
Source: IBGE, 1976 and 1996.

in figure 10, in keeping with this pattern, combines both of these effects with the changing  $\alpha$ s,  $\beta$ s and  $\gamma$ s. The result is striking: this complex combined simulation suggests that all of these effects, over twenty turbulent years, cancel each other out almost exactly from the 15<sup>th</sup> percentile up. This explains the smallness of the changes in headline poverty. From around the 12<sup>th</sup> percentile down, however, the simulation suggests a prevalence of the negative occupational choice (and to a lesser extent, price) effects, with substantial income losses. These account for the rise in indigence captured by the R\$30/month poverty line.

The bottom line in figure 10 is, in a sense, the final attempt by this methodology to simulate the various changes leading from the 1976 to 1996 distribution. Figure 11 is a graphical test of the approach. Here, the line denoted “1996-1976” plots the differences in actual (log) incomes between the observed 1996 and the observed

1976 distributions. Along with this, we also plotted every (cumulative) stage of our simulations. First the impoverishing (but roughly equal) price effects; then these combined with the highly impoverishing occupational choice effects; then the slightly less bleak picture arising from a combination of the latter with the parameters of the family size equations, and finally, the curve plotting the differences between the incomes from the simulation with all parameters changing, and the observed 1976 figures. As may be seen, the last line appears to replicate the actual differences fairly well. Of course, the point of the exercise is not to replicate the actual changes perfectly, but rather to learn the different effects of different parameters, and possibly to infer any policy implications from them. Nevertheless, the success of the last simulation in approximately matching the actual changes does give some extra confidence in the methodology and in any lessons we may derive from it.

FIGURE 11



Source: IBGE, 1976 and 1996.

## VI

### Conclusions

In the final analysis, does this exercise help us improve our understanding of the evolution of Brazil's urban income distribution over this turbulent twenty-year period? Whereas many traditional analysts of income distribution dynamics might have inferred, from the small changes in mean income, in various inequality indices, and in the incidence of poverty,<sup>19</sup> that there was little –if anything– to investigate, digging a little deeper has unearthed a wealth of economic factors interacting to determine substantial changes in the environment faced by individuals and families, and in their responses.

In particular, we have found that, despite a small fall in measured inequality (although the Lorenz curves cross) and a small increase in mean income, extreme poverty has increased, whether measured by sufficiently low poverty lines or sufficiently high poverty aversion parameters. This seems to have been caused by outcomes related to participation decisions and occupational choices, in combination with declines in the labour market returns to education and experience. These changes are associated with greater unemployment and informality, as one would expect, but more research into them seems necessary. While we seem to have identified the existence of a group excluded from both the productive labour markets and any substantive form of safety net,

we have not been able to fully interpret the determinants of their occupational choices. Issues of mobility – exacerbated by the fact that the welfare indicator is based on current monthly income– also call for further research in this context. The policy implications would seem to indicate the desirability of targeted labour programmes, or other safety nets, but it would be foolhardy to go into greater detail until the profile of the group which seems to have fallen into extreme poverty in 1996 is better understood.

Secondly, we have found that, even above the 15<sup>th</sup> percentile, where urban Brazilians have essentially 'stayed put', this was the result of some hard climbing along a slippery slope. They had to gain an average of two extra years of schooling (which still leaves them under-educated for the country's per capita income level), and substantially reduce fertility, in order to counteract falling returns in both the formal labour market and in self-employment.

It may well be, as many now claim, that an investigation of non-monetary indicators –such as access to services, or life expectancy at birth– should lead us to consider the epithet of 'a lost decade' as too harsh for the 1980s. Unfortunately, however, we find that if one is sufficiently narrow-minded to consider only welfare measured in terms of money, urban Brazil has in fact experienced not just one, but two, lost decades.

(Original: English)

<sup>19</sup> With respect to the already low R\$ 60/month poverty line, by historical standards for Brazil.

#### Bibliography

- Almeida dos Reis, J. G. and R. Paes de Barros (1991): Wage inequality and the distribution of education: A study of the evolution of regional differences in inequality in metropolitan Brazil, *Journal of Development Economics*, vol. 36, N° 1, Amsterdam, The Netherlands, Elsevier Science Publishers, B.V.
- Bacha, E. and L. Taylor (1980): Brazilian income distribution in the 1960s: "Facts", model results, and the controversy, in L. Taylor, E. Bacha, E. Cardoso and F. Lysy, *Models of Growth and Distribution for Brazil*, Washington, D.C., World Bank.
- Bonelli, R. and G. L. Sedlacek (1989): Distribuição da renda: evolução no último quarto de século, in G.L. Sedlacek and R. Paes de Barros, *Mercado de trabalho e distribuição da renda: uma coletânea*, Serie monográfica, N° 35, Rio de Janeiro, Institute of Applied Economic Research (IPEA).
- Bourguignon, F., F. Ferreira and N. Lustig (1998): The microeconomics of income distribution dynamics in East Asia and Latin America, *World Bank Research Proposal*, Washington, D.C., World Bank, April.
- Buhmann, B., L. Rainwater, G. Schmaus and T. Smeeding (1988): Equivalence scales, well-being, inequality and poverty: Sensitivity estimates across ten countries using the Luxembourg income study database, *Review of Income and Wealth*, vol. 34, N° 2, New Haven, Connecticut, International Association for Research on Income and Wealth.

- Deaton, A. and C. Paxson (1997): *Poverty Among Children and the Elderly in Developing Countries*, Working paper, N° 179, Princeton, New Jersey, Princeton University, Research Program in Development Studies.
- Duryea, S. and M. Szekely (1998): *Labor markets in Latin America: A supply-side story*, paper presented at the IDB/ IIC Annual Meeting, Cartagena de Indias, Colombia, Inter-American Development Bank (IDB)/Inter-American Investment Corporation (IIC).
- Ferreira, F. and R. Paes de Barros (1999): The slippery slope: Explaining the increase in extreme poverty in urban Brazil, 1976-1996, *The Brazilian Review of Econometrics*, vol. 19, N° 2, Rio de Janeiro, Brazilian Econometrics Society, November.
- Fishlow, A. (1972): Brazilian size distribution of income, *The American Economic Review*, vol. 62, N° 2, Nashville, Tennessee, American Economic Association.
- Foster, J., J. Greer and E. Thorbecke (1984): A class of decomposable poverty measures, *Econometrica*, vol. 52, N° 3, Evanston, Illinois, Econometric Society.
- Hoffman, R. (1989): Evolução da distribuição da renda no Brasil, entre pessoas e entre famílias, 1979/86, in G.L. Sedlacek and R. Paes de Barros, *Mercado de trabalho e distribuição da renda: uma coletânea*, Serie monográfica, N° 35, Rio de Janeiro, Institute of Applied Economic Research (IPEA).
- IBGE (Brazilian Geographical and Statistical Institute) (1976): *Pesquisa Nacional por Amostra de Domicílios*, Brasília, Fundação IBGE.
- (1981): *Pesquisa Nacional por Amostra de Domicílios*, Brasília, Fundação IBGE.
- (1985): *Pesquisa Nacional por Amostra de Domicílios*, Brasília, Fundação IBGE.
- (1996): *Pesquisa Nacional por Amostra de Domicílios*, Brasília, Fundação IBGE.
- Juhn, C., K. Murphy and B. Pierce (1993): Wage inequality and the rise in returns to skill, *Journal of Political Economy*, vol. 101, N° 3, Chicago, Illinois, The University of Chicago Press.
- Langoni, C.G. (1973): *Distribuição da renda e desenvolvimento econômico do Brasil*, Rio de Janeiro, Expressão e Cultura.
- Mookherjee, D. and A. F. Shorrocks (1982): A decomposition analysis of the trend in UK income inequality, *The Economic Journal*, vol. 92, N° 368, Oxford, U.K., The Royal Economic Society.
- Ramos, L. (1993): *A distribuição de rendimentos no Brasil: 1976/85*, Rio de Janeiro, Institute of Applied Economic Research (IPEA).