CEPAL REVIEW

Nº 123
DECEMBER • 2017

ECONOMIC COMMISSION FOR LATIN AMERICA AND THE CARIBBEAN

UNITED NATIONS

ECLAC

ISSN 0251-2920

CEPAL

REVIEW

ECONOMIC
COMMISSION FOR
LATIN AMERICA AND
THE CARIBBEAN

UNITED NATIONS

ECLAC

N° 123
DECEMBER • 2017

CEPAL REVIEW

Alicia Bárcena
Executive Secretary

Mario Cimoli
Deputy Executive Secretary a.i.

ECONOMIC
COMMISSION FOR
LATIN AMERICA AND
THE CARIBBEAN

Osvaldo Sunkel
Chairman of the Editorial Board

Miguel Torres
Technical Editor

United Nations

ECLAC

ISSN 0251-2920

The *CEPAL Review* was founded in 1976, along with the corresponding Spanish version, *Revista CEPAL*, and it is published three times a year by the Economic Commission for Latin America and the Caribbean (ECLAC), which has its headquarters in Santiago. The *Review* has full editorial independence and follows the usual academic procedures and criteria, including the review of articles by independent external referees. The purpose of the *Review* is to contribute to the discussion of socioeconomic development issues in the region by offering analytical and policy approaches and articles by economists and other social scientists working both within and outside the United Nations. The Review is distributed to universities, research institutes and other international organizations, as well as to individual subscribers.

The opinions expressed in the articles are those of the authors and do not necessarily reflect the views of ECLAC.

The designations employed and the way in which data are presented do not imply the expression of any opinion whatsoever on the part of the United Nations concerning the legal status of any country, territory, city or area or its authorities, or concerning the delimitation of its frontiers or boundaries.

To subscribe, please visit the Web page: http://ebiz.turpin-distribution.com/products/197587-cepal-review.aspx

The complete text of the *Review* can also be downloaded free of charge from the ECLAC website (www.eclac.org).

This publication, entitled CEPAL Review, is covered in the Social Sciences Citation Index (SSCI), published by Thomson Reuters, and in the Journal of Economic Literature (JEL), published by the American Economic Association

United Nations publication
ISSN: 0251-2920
ISBN: 978-92-1-121972-2 (print)
ISBN: 978-92-1-058610-8 (pdf)
LC/PUB.2017/24-P
Distribution: General

Printed at United Nations, Santiago
S.17-00682

Contents

Explanatory notes

- Three dots (...) indicate that data are not available or are not separately reported.
- A dash (-) indicates that the amount is nil or negligible.
- A full stop (.) is used to indicate decimals.
- The word “dollars” refers to United States dollars, unless otherwise specified.
- A slash (/) between years (e.g. 2013/2014) indicates a 12-month period falling between the two years.
- Individual figures and percentages in tables may not always add up to the corresponding total because of rounding.

Effects of internal migration on the human settlements system in Latin America and the Caribbean

Jorge Rodríguez Vignoli

Abstract

The gradual abatement of rural-urban migration in Latin America and the Caribbean is leading to growing prevalence of migration between cities, a phenomenon about which little theory or empirical studies have been developed in the region. Accordingly, this work uses census microdata —the only source available in the region for estimating migration between cities— from a dozen countries to: (i) estimate the recent evolution of this migration using categories based on cities' population size (including a residual category that groups municipalities without cities); (ii) estimate the effect of this migration on the composition by sex, age, and education level of these categories of city, and (iii) evaluate in a general and preliminary manner the two-way links between the socioeconomic conditions of cities and the magnitude and effects of migration.

Keywords

Internal migration, human settlements, cities, population composition, quality of life, population censuses, migration statistics, Latin America and the Caribbean

JEL classification

R23, P25, O54

Author

Jorge Rodríguez Vignoli is a research assistant at the Latin American and Caribbean Demographic Centre (CELADE)-Population Division of the Economic Commission for Latin America and the Caribbean (ECLAC). Email: jorge.rodriguez@cepal.org.

I. Introduction[1]

Urbanization is part of the long-term structural processes that interact to generate economic and social modernization and cultural modernity. A few years ago, urbanization reached a milestone, when the global urban population passed 50% of the total. All existing projections suggest that this percentage will continue to rise and will likely follow a logistic trajectory, on the assumption that a proportion of the population will remain in rural areas, either because of individual preference or because of economic and social needs.

Today, 80% of the Latin American and Caribbean population lives in urban areas (United Nations, 2015). This being so, the most frequent kind of internal migration is almost certainly between urban areas. However, inter-city migration has been much less studied than rural-urban migration, which tends to remain the focus, despite its gradual and inevitable depletion. Accordingly, a first objective of this work is to provide a quantitative approach to migration flows between categories and nodes of the human settlements system, with an emphasis on inter-city migration. A second objective is to show that internal migration —especially between cities— has important effects on human settlements and on cities in particular. The analysis will focus on a small fraction of those effects; specifically, the impact of migration on the sex, age and educational structure of settlements.

Following this introduction is a section presenting the main research ideas and the links between them. Next comes a methodological section, which defines the concepts used and lists and discusses the sources, procedures and indicators employed. This is followed by the results, concerning first the magnitude of migration between cities and settlements and later the effect of migration on the composition of the population in cities and settlements. Lastly, the results are discussed, ordering the main conclusions by the three hierarchical levels of the human settlements system: large, medium-sized and small cities (also includes a residual category, called “rest”, which groups all the municipalities which do not have cities).

II. Urbanization, city systems, concentration and internal migration

Urbanization can be based on very different city systems: from a single metropolis —that is, a highly concentrated system, or a primate system in the technical jargon, such as a city-State or a country with a single large city that coexists with the rest of the rural territory— to a myriad of cities of different sizes. The diversity of levels of primacy in specific city systems suggests that a broad and complex range of factors are involved in determining this, as well as idiosyncratic conditions (Pacione, 2009; Fujita and Krugman, 2004; Hall, 1996; Romero, 1976). The bulk of the literature suggests that primate systems tend to be dysfunctional for development (Williamson, 1965; Henderson, 2003; Atienza and Aroca, 2012).

Homeostatic or self-regulating visions of society —whether neoclassical in economics, functionalist or systems theory in sociology or evolutionist in development theory— assume that the forces of deconcentration will ultimately prevail and large cities will lose economic power and pull, such that the primacy indexes of city systems will fall (Cunha and Rodríguez, 2009; World Bank, 2009; Fujita and

[1] The author would like to thank Daniela González, Research Assistant with CELADE-Population Division of ECLAC, for her assistance in the definition of cities and obtaining socioeconomic indicators on them, and Mario Acuña, Luis Rodríguez and David Candia, consultants with CELADE-Population Division of ECLAC, for their help with data processing. The author is also grateful for the comments of two anonymous referees, which helped to improve the article. Any remaining limitations and errors in the text are entirely the author’s own.

Krugman, 2004; Henderson, 2003; Cuervo and González, 1997). Evolutionist approaches also propose deconcentration as a —sometimes final— "counter-urbanization" phase. During this phase, growth and migratory pull shifts from the large cities to medium-sized and small ones, in response to the problems that emerge in the large cities (Pacione, 2009; Geyer and Kontuly, 1993; Berg and others, 1982).

Counter-urbanization and, in general, the hypothesis of an inevitable tendency towards the deconcentration of city systems, have been questioned from various perspectives. One of these is "concentrated deconcentration" (Cunha, 2015; ECLAC, 2012; Cunha and Rodríguez, 2009; Villa and Rodríguez, 1997) or the formation of city-regions (Sassen, 2007), whereby metropolitan areas lose demographic (and economic) gravitational pull as their surroundings, especially nearby cities, gain pull of their own; here, what is actually occurring is the expansion of the geographical scale of the metropolitan area. Counter-urbanization is also questioned from a number of evolutionist perspectives, in particular those which include a final stage of reconcentration or metropolitan resurgence (Cunha, 2015; Pacione, 2009). The structuralist approach developed by the Economic commission for Latin America and the Caribbean (ECLAC) adopts mainly the centre-periphery perspective, which is normally associated with mechanisms and processes that exacerbate or at least tend to reproduce concentration, asymmetry and inequality. At the global level, peripheries may be characterized as follows: they are productive structures that have few dynamic activities, being associated instead with exports of primary or semi-processed goods with few linkages, or subsistence activities; they are structurally highly uneven in terms of productivity levels between sectors and firm sizes, which leads to acute labour segmentation and steep income inequalities; and they are slow to absorb technical progress, which tends to be concentrated in just a few sectors of the economy. These features tend to be reproduced within countries, except that there is a centre or centres formed by the large cities, and the periphery consists of the rest of the human settlements system. The structuralist approach is not limited to the effects of centrifugal and centripetal forces, but also underlines the cumulative role of history in the formation of the Latin American States and economies of today. In this regard, it highlights the role played by the social and political-institutional matrix in the concentration of the population since the Spanish conquest (or even before). The difference between diversified production structures like those of the metropolis, and structures that are specialized —particularly in raw materials— like those of the rest of the country or the periphery, will in principle fuel inequalities and concentration. To this is added the effects of value chains, which transfer resources from the periphery to the centre, as well as selective migration, which also favours the centre. Be this as it may, these basically concentrating mechanisms can be offset by public policies or they can be disabled by structural shifts in the economy. Accordingly, the structuralist approach does not necessarily project an inevitable rise in concentration in the large cities either (ECLAC, 2015, pp. 18-24).

Migration is key dimension of the discussion on the trend and future of demographic and economic concentration in large cities and their influence on national development. In the case of migration between cities, the differentiation between origin and destination, so evident in the case of rural-urban migration, is blurred given that both are urban. Although there are socioeconomic and other disparities between cities that prompt people's decisions to migrate, the differences tend to be smaller and more subtle, nuanced and complex than the distinction between the countryside and the city. The classic models of rural-urban migration based on labour market differentials —i.e. basically disparities in unemployment and income— are thus somewhat limited in their ability to explain inter-city migration (Rodríguez and Busso, 2009; Aroca, 2004; Brown, 1991; Martine, 1979; Villa and Alberts, 1980), as they tend to disregard the factors related to area of residence, culture, education, living standards and cost of living that appear to motivate decisions to move from one city to another and can sometimes be dissociated from levels and trajectories of income and employment.

Consequently, the petering out of migration from the countryside and the more complex and diverse fabric of the city system may be changing the direction of migratory flows. It may also be changing the traditional effects of internal migration on the composition of the population, which in Latin America was broadly to feminize and rejuvenate[2] large cities, as well as to reduce their education level, and to masculinize and age small cities and rural areas (Elizaga, 1970; Camisa, 1972; Elizaga and Macisco, 1975; Alberts, 1977; Rodríguez, 2013a and 2013b).

This work proceeds, first, with an updated quantification of migration between categories of cities and settlements. Second, it gives an initial description of relations as they exist today between demographic size, living standards and migratory pull of cities and settlements. Lastly, it evaluates empirically the effects of migration on composition by age, sex and level of education of the population in cities and settlements.

III. Methodological framework

The results presented in this article come from the processing of census databases from 10 countries of the region —the Bolivarian Republic of Venezuela, Brazil, Costa Rica, the Dominican Republic, Ecuador, Honduras, Mexico, Panama, the Plurinational State of Bolivia and Uruguay— two of them (Brazil and Mexico) the most populous in Latin America. The census data were used to build matrices of origin and destination of recent migration between cities, understood as all localities and urban and metropolitan agglomerations of at least 20,000 inhabitants. In order to capture the whole of the human settlements system in the calculations and the analysis, the matrices include a residual category grouping all those municipalities without cities.

The geographical definition of cities comes from the spatial distribution and urbanization in Latin America and the Caribbean (DEPUALC) database (www.cepal.org/celade/depualc/). This recently updated database uses an urban areas and conglomerates approach to avoid limiting cities strictly to the urban sprawl. In general, the geographical definition of cities in DEPUALC is based on the urban area of the minor administrative division (MIAD)[3] where the city is located or of the MIADs that contain or make up the city (in the case of cities which exceed the boundaries of a single MIAD). All cities which meet the minimum size requirement (i.e. 20,000 inhabitants or more) form part of the list of cities.

MIADs are used to create the origin-destination matrices because they offer the most disaggregated geographical scale at which migration is captured in most countries. The questions about the MIAD of current residence and the MIAD of residence five years earlier are used to build up all the cities with 20,000 or more inhabitants in each country, either as complete MIADs (the MIAD where the city is located, in the case that the city does not exceed it), or as groupings of MIADs (those which make up the city or across which the city extends). These new entities corresponding to cities are then used as origins and destinations in the respective destinations, generating inter-city migration matrices which yield all the standard indicators on volume and intensity of migration.

The matrices include a column entitled “rest”, which groups all the MIADs which do not have a city, from which estimates can be obtained on net migration exchange between the cities system and the rest of the settlements in the country. In other words, the “rest” category permits an innovative approach

[2] “Rejuvenate” is not used in the demographic sense of increasing the child population, but in the literal sense of increasing the proportion of young people, defined as the population aged between 15 and 29 years.

[3] A minor administrative division (MIAD) can refer to a municipality, delegation, department, among others, depending on the country.

to directly estimating rural-urban migration. The threshold for defining rural is quite strict in this case, since these are MIADs which do not have or do not form part of a city of at least 20,000 inhabitants. Even so, this category includes some MIADs whose population is entirely rural and dispersed, as well as others that have small urban localities or areas that are still rural but in the early or intermediate stages of suburbanization.

The migration matrices by city —of some 800 by 800 in countries such as Brazil— are available in the Database on Internal Migration in Latin America and the Caribbean (MIALC) (www.cepal.org/celade/migracion/migracion_interna/). In this article, in order to conduct a standardized analysis and for obvious reasons of space, the results are presented by cities grouped by population size. The groupings are:[4] (i) 1,000,000 inhabitants or more (large cities); (ii) between 500,000 and 999,999 inhabitants (upper-intermediate cities); (iii) between 10,000 and 499,999 inhabitants (lower-intermediate cities); (iv) between 50,000 and 99,999 inhabitants (upper-small cities); (v) between 20,000 and 49,999 inhabitants (lower-small cities); (vi) less than 20,000 inhabitants ("special" lower-small category)[5], and (vii) the rest.

The calculations derived from the matrices are: (i) population resident at the date of the census; (ii) population resident five years before the census; (iii) non-migrants; (iv) immigrants; (v) emigrants; (vi) net migration; (vii) gross migration; (viii) in-migration rate; (ix) out-migration rate; and (x) net migration rate.[6] These data are accompanied by the proportion of cities with positive balances or rates; otherwise the overview of the migratory situation of the human settlements system could derive solely from the total, which could be shaped mainly by the main city or cities or by extreme cases.

To estimate the effect of internal migration on the composition of the population, the procedure developed by the Latin American and Caribbean Demographic Centre (CELADE)-Population Division of ECLAC is used (Rodríguez, 2013a and 2013b; CEPAL, 2012; Rodríguez and Busso, 2009). The procedure is based on comparison of the marginals of the flow indicator matrix (from the migration matrix five years before the census), one of which is the value of the attribute at the time of the census (factual value), i.e. with migration, and the other, the attribute as it would be if migration had not occurred (counterfactual value). The difference between the two is the effect (net and exclusive) of migration on the attribute. The ratio between the effect and counterfactual is the relative effect of migration on the composition of the population in terms of that particular attribute.

This procedure was used to estimate the effect of migration on the structure by sex, age and education of the cities of the region, grouped into the size categories mentioned, including the category "rest", which assimilates the rural or semirural sphere. To calculate the effect on population structure by sex, the masculinity ratio was used, that is, the ratio of men to women; in the case of age structure, the percentage represented by selected age groups —5-14 years, 15-29 years, 30-44 years, 45-59 years and 60 years or over— within the total population (strictly speaking, the population including the migration

[4] Except in the case of the term "large cities", used to refer to cities whose size makes them in fact large in any national context, the use of terms such as "intermediate" or "small" is exclusively pragmatic. Whether a city is "intermediate" in a given country depends on the city system in that country. Cities that are intermediate in Brazil or Mexico would be the second largest in Uruguay and Panama, for example, by number of inhabitants. Although the DEPUALC and MIALC databases permit a definition of intermediate cities associated with the characteristics of the city system in each country, the author of this work has opted for a common criterion for all the countries in order to standardize the analysis, even though this may involve the risk of grouping different cases together.

[5] The population corresponding to the category "less than 20,000 inhabitants" is a derivation of the method and not a value for all the cities of less than 20,000 inhabitants. The category is those which had less than 20,000 inhabitants in 2010 but not in 2000 and therefore do not fall within the 20,000-49,999 group in 2000, and have to be added to the "rest" category, rather than being treated as a special case. There are exceptional cases of this sort in 2010, corresponding to cities which that year exceeded 20,000 inhabitants in the matrix but not in DEPUALC. This category is marginal and can be added to the "rest" category in the corresponding years.

[6] For further details on the calculation of these indicators, see Rodríguez (2013a and 2011), Welti (1997) and Villa (1991).

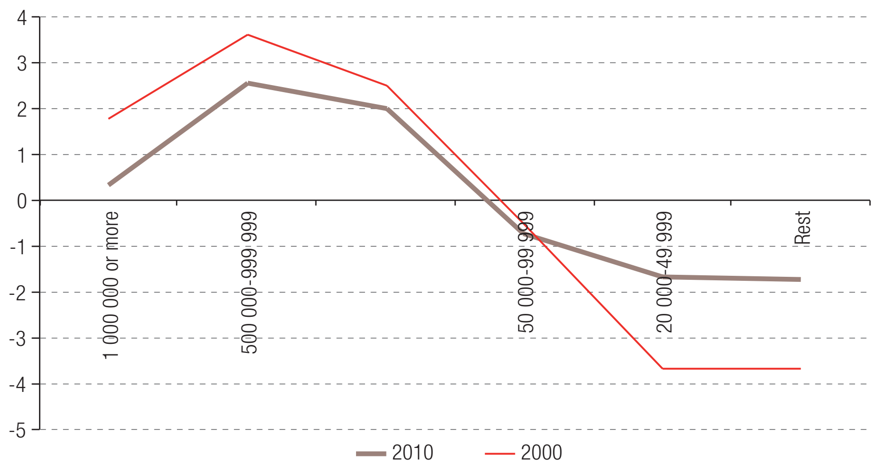
matrix, which excludes children under age 5 and recently international immigrants); and in the case of education, average years of study of the population aged 25 years and over and the group aged 45-59 years, to try to control for the distortion generated by the age structure. These effects can, of course, be broken down into the impacts of in-migration and out-migration. The first is obtained as the difference between the factual value and the value for non-immigrants for each place. The second is obtained as the difference between the value for non-migrants and the counter-factual value for each place.

IV. Results and analysis

1. City systems and internal migration: continuity and change in migratory pull and the general growth effect

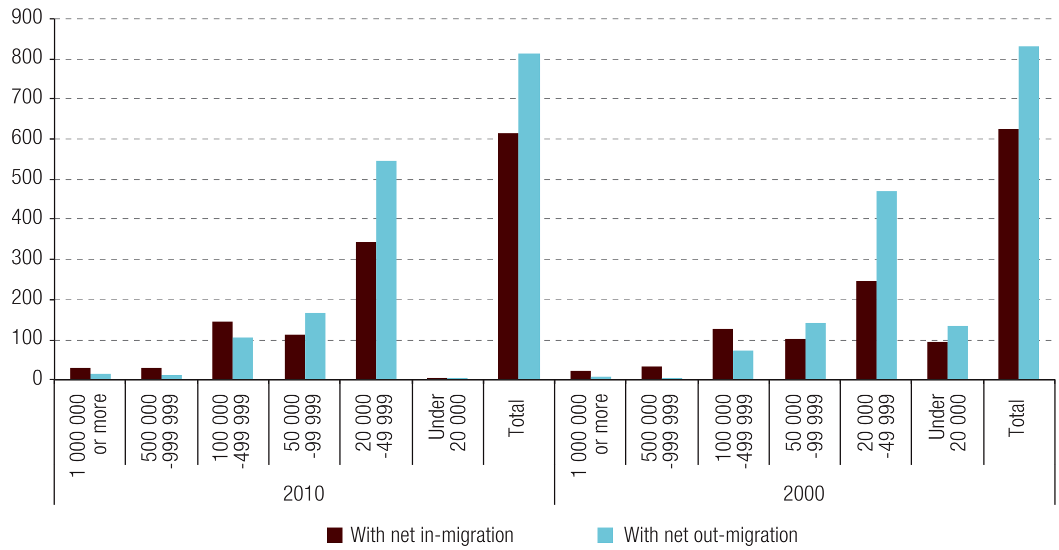
The results presented confirm the findings of Rodríguez (2011), insofar as the lower bands of the city system show net loss of population, the intermediate bands are tending to pull in population and the upper bands are still exerting pull as well (see figure 1). The fact that the lower bands of the city system are losing population leads a situation that is surprising and even paradoxical, given the well-documented march of urbanization in the past few decades: because small cities far outnumber larger ones, the fact that they are losing population means that most cities are showing net out-migration (see figure 2).

Figure 1
Latin America and the Caribbean (selected countries): net internal migration rate by segment of the human settlements system grouped by population size, population aged 5 years and over[a]
(Number of persons per 1,000)



Source: Prepared by the author.
Note: Excludes the category “less than 20,000 inhabitants”.
[a] Includes 10 countries with censuses and information available from the 2010 census round (Bolivarian Republic of Venezuela (2011), Brazil (2010), Costa Rica (2011), Dominican Republic (2010), Ecuador (2010), Honduras (2013), Mexico (2010), Panama (2010), Plurinational State of Bolivia (2012) and Uruguay (2011)) and eight from the 2000 census round (Bolivarian Republic of Venezuela (2001), Brazil (2000), Costa Rica (2000), Dominican Republic (2002), Ecuador (2001), Honduras (2001), Mexico (2000) and Panama (2000)).

Figure 2
Latin America and the Caribbean (selected countries): cities by net migration sign, by range of population size of the city[a]
(Number of cities)



Source: Prepared by the author, on the basis of information from the Database on Internal Migration in Latin America and the Caribbean (MIALC).

Note: The number of cities in 2000 includes almost 300 localities which were not cities in that year's census, but had become cities by the census of 2010. Their inclusion facilitates the diachronic comparison and explains the very similar number of cities at the two points in time.

[a] Includes 10 countries with censuses and information available from the 2010 census round (Bolivarian Republic of Venezuela (2011), Brazil (2010), Costa Rica (2011), Dominican Republic (2010), Ecuador (2010), Honduras (2013), Mexico (2010), Panama (2010), Plurinational State of Bolivia (2012) and Uruguay (2011)) and eight from the 2000 census round (Bolivarian Republic of Venezuela (2001), Brazil (2000), Costa Rica (2000), Dominican Republic (2002), Ecuador (2001), Honduras (2001), Mexico (2000) and Panama (2000)).

Table 1 offers another way to approach the direct estimation of rural-urban migration, by calculating the migratory balance of the category termed "rest" —bearing in mind the caveats set forth in the methodological framework, particularly with respect to the heterogeneity within that category. The "rest" category is systematically losing population, which suggests that rural-urban migration is continuing and showing no signs of reversal, although it is slowing.

It can also be seen in table 1 that most migrants are urban-urban. Of the 14.4 million migrants registered in the 2010 census round, 11.2 million (78%) were immigrants into cities and 10.6 million (73.5%) were emigrants from cities, which means that three of every four migrants moved between cities. This percentage is around 80% in the case of the intra-category flows shown in the second component of table 1, and could be even higher in the case of intra-metropolitan migration —not included in the calculations presented here— which is substantial in the more urbanized and "metropolized" countries. The calculations also omit intrarural migration (which occurs within the "rest" segment), which precludes obtaining a measure of rural-rural flows.

Also striking in table 1 is that the large cities show quite a small migratory balance (a rate of barely 0.3 per 1,000 according to the 2010 census round), much lower than in the 2000 census round. Notably, inspection of the matrices of migration between cities —available in the MIALC database— shows that this group is internally diverse, since all the megacities or supercities (cities of 10 million or more inhabitants) show net out-migration, while most of the other large cities still show net in-migration. Medium-sized cities were the most attractive during the period examined, with a slightly falling net in-migration rate between the two census rounds. In any case, these are still moderate rates (of around 3 per 1,000), far short of the annual rates of upwards of 20 per 1,000 that were common until the 1980s (Alberts, 1977).

Table 1

Latin America and the Caribbean (selected countries): international migration indicators by city groupings by population size, excluding and including intra-category migratory movements, population aged 5 years and over[a]

	Option 1: excluding intra-category migratory movements										
Census round	Groups of cities by size of population	Population resident in 2010	Population resident in 2005	Non-migrants	Immigrants	Emigrants	Net migration	Gross migration	In-migration rate (per 1 000)	Out-migration rate (per 1 000)	Net migration rate (per 1 000)
2010 census round	1. 1 000 000 or more	130 957 264	130 757 276	127 202 365	3 754 900	3 554 911	199 988	7 309 811	5.7	5.4	0.3
	2. 500 000-999 999	27 406 682	27 056 232	25 962 344	1 444 338	1 093 889	350 449	2 538 226	10.6	8.0	2.6
	3. 100 000-499 999	51 970 165	51 451 091	49 160 957	2 809 207	2 290 134	519 073	5 099 341	10.9	8.9	2.0
	4. 50 000-99 999	22 172 936	22 256 688	20 871 167	1 301 769	1 385 521	-83 752	2 687 290	11.7	12.5	-0.8
	5. 20 000-49 999	35 997 837	36 297 085	34 021 489	1 976 348	2 275 596	-299 249	4 251 944	10.9	12.6	-1.7
	6. Under 20 000	114 506	116 831	104 718	9 788	12 112	-2 324	21 901	16.9	20.9	-4.0
	7. Rest	78 073 209	78 757 395	74 954 991	3 118 218	3 802 405	-684 186	6 920 623	8.0	9.7	-1.7
	Total for the human settlements system	346 692 599	346 692 599	332 278 031	14 414 568	14 414 568	0	28 829 136	8.3	8.3	0.0
2000 census round	1. 1 000 000 or more	99 306 010	98 419 025	95 171 096	4 134 913	3 247 929	886 985	7 382 842	8.4	6.6	1.8
	2. 500 000-999 999	25 189 355	24 735 987	23 572 789	1 616 566	1 163 197	453 368	2 779 763	13.0	9.3	3.6
	3. 100 000-499 999	41 343 343	40 825 305	38 482 860	2 860 483	2 342 444	518 038	5 202 927	13.9	11.4	2.5
	4. 50 000-99 999	18 736 768	18 786 657	17 343 752	1 393 016	1 442 905	-49 889	2 835 921	14.8	15.4	-0.5
	5. 20 000-49 999	28 553 605	29 084 249	26 740 465	1 813 140	2 343 783	-530 643	4 156 924	12.6	16.3	-3.7
	6. Under 20 000	6 066 723	6 110 868	5 568 626	498 097	542 242	-44 145	1 040 340	16.4	17.8	-1.5
	7. Rest	66 417 807	67 651 520	63 481 708	2 936 099	4 169 813	-1 233 713	7 105 912	8.8	12.4	-3.7
	Total for the human settlements system	285 613 611	285 613 611	270 361 297	15 252 314	15 252 314	0	30 504 628	10.7	10.7	0.0

Table 1 (concluded)

Census round	Groups of cities by size of population	Option 2: including intra-category migratory movements									
		Population resident in 2010	Population resident in 2005	Non-migrants	Immigrants	Emigrants	Net migration	Gross migration	In-migration rate (per 1 000)	Out-migration (per 1 000)	Net migration rate (per 1 000)
2010 census round	1. 1 000 000 or more	130 957 264	130 757 276	126 049 248	4 908 016	4 708 028	199 988	9 616 043	7.5	7.2	0.3
	2. 500 000-999 999	27 406 682	27 056 232	25 812 021	1 594 661	1 244 211	350 449	2 838 872	11.7	9.1	2.6
	3. 100 000-499 999	51 970 165	51 451 091	48 626 464	3 343 700	2 824 627	519 073	6 168 328	12.9	10.9	2.0
	4. 50 000-99 999	22 172 936	22 256 688	20 767 434	1 405 503	1 489 254	-83 752	2 894 757	12.7	13.4	-0.8
	5. 20 000-49 999	35 997 837	36 297 085	33 730 438	2 267 398	2 566 647	-299 249	4 834 045	12.5	14.2	-1.7
	6. Under 20 000	114 506	116 831	104 718	9 788	12 112	-2 324	21 901	16.9	20.9	-4.0
	7. Rest	78 073 209	78 757 395	74 954 991	3 118 218	3 802 405	-684 186	6 920 623	8.0	9.7	-1.7
	Total for the human settlements system	346 692 599	346 692 599	330 045 315	16 647 284	16 647 284	0	33 294 569	9.6	9.6	0.0
2000 census round	1. 1 000 000 or more	99 306 010	98 419 025	94 225 768	5 080 242	4 193 257	886 985	9 273 499	10.3	8.5	1.8
	2. 500 000-999 999	25 189 355	24 735 987	23 463 233	1 726 122	1 272 754	453 368	2 998 876	13.8	10.2	3.6
	3. 100 000-499 999	41 343 343	40 825 305	37 980 943	3 362 400	2 844 362	518 038	6 206 762	16.4	13.8	2.5
	4. 50 000-99 999	18 736 768	18 786 657	17 232 333	1 504 435	1 554 324	-49 889	3 058 759	16.0	16.6	-0.5
	5. 20 000-49 999	28 553 605	29 084 249	26 486 306	2 067 299	2 597 943	-530 643	4 665 242	14.3	18.0	-3.7
	6. Under 20 000	6 066 723	6 110 868	5 548 557	518 166	562 311	-44 145	1 080 477	17.0	18.5	-1.5
	7. Rest	66 417 807	67 651 520	63 481 708	2 936 099	4 169 813	-1 233 713	7 105 912	8.8	12.4	-3.7
	Total for the human settlements system	285 613 611	285 613 611	268 418 848	17 194 763	17 194 763	0	34 389 525	12.0	12.0	0.0

Source: Prepared by the author, on the basis of information from the Database on Internal Migration in Latin America and the Caribbean (MIALC).

[a] Includes 10 countries with censuses and information available from the 2010 census round (Bolivarian Republic of Venezuela (2011), Brazil (2010), Costa Rica (2011), Dominican Republic (2010), Ecuador (2010), Honduras (2013), Mexico (2010), Panama (2010), Plurinational State of Bolivia (2012) and Uruguay (2011)) and eight from the 2000 census round (Bolivarian Republic of Venezuela (2001), Brazil (2000), Costa Rica (2000), Dominican Republic (2002), Ecuador (2001), Honduras (2001), Mexico (2000) and Panama (2000)).

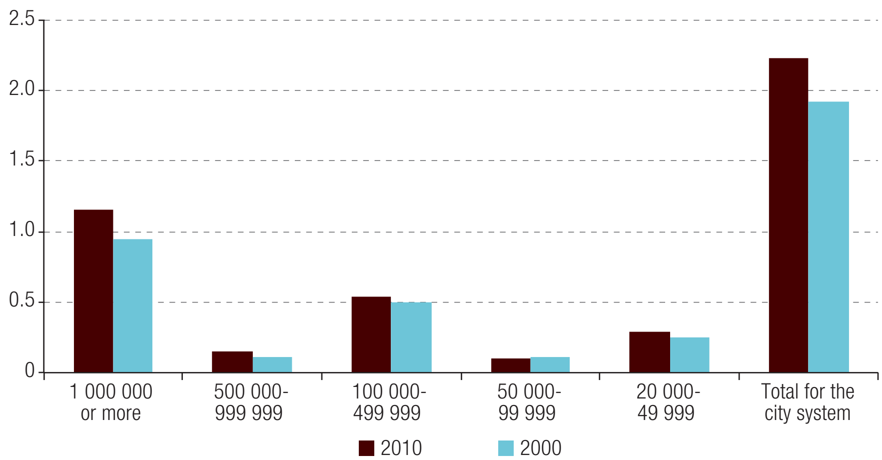
The combination of data provided in table 1 suggests that internal migration is tending to produce a certain deconcentration of the population, although this process is limited to the medium-sized cities, extending to some extent to smaller cities but less to the rural sphere.

Comparison between the two components of table 1 allows another discussion, given that the difference between the two corresponds, as explained in the methodological section, to the amount of migration between cities in the same category. This migration alters the quantity of immigrants and emigrants in each category of city and their respective rates, but not the net migration in the category —precisely because it is intra-category migration, which does not imply exchanges with other categories—so this value is identical in the two components of the table.[7] The results (see figure 3) show that migration within each category of the city system tended to increase in the last inter-census period, at least in absolute terms, which reflects a growing horizontal migratory exchange that contrasts with the overall reduction in internal migration shown in previous figures and studies (ECLAC, 2012; Bell and Salut, 2009). This warrants further research.

Figure 3

Latin America and the Caribbean (selected countries): number of intra-category migrants in the city system, population aged 5 years and over[a]

(Millions of persons)



Source: Prepared by the author, on the basis of information from the Database on Internal Migration in Latin America and the Caribbean (MIALC).

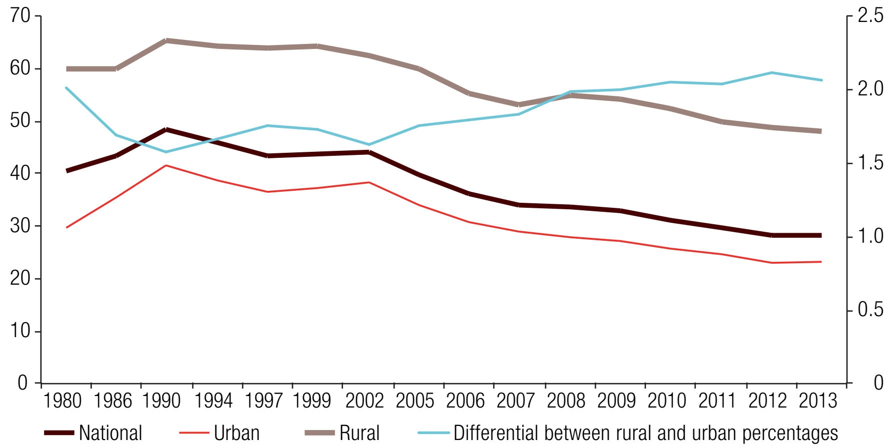
Note: Excludes the category "less than 20,000 inhabitants".

[a] Includes 10 countries with censuses and information available from the 2010 census round (Bolivarian Republic of Venezuela (2011), Brazil (2010), Costa Rica (2011), Dominican Republic (2010), Ecuador (2010), Honduras (2013), Mexico (2010), Panama (2010), Plurinational State of Bolivia (2012) and Uruguay (2011)) and eight from the 2000 census round (Bolivarian Republic of Venezuela (2001), Brazil (2000), Costa Rica (2000), Dominican Republic (2002), Ecuador (2001), Honduras (2001), Mexico (2000) and Panama (2000)).

Unlike what has been observed in the case of rural-urban migration, which is explained by natural structural reasons, such as persistent sharp inequalities between urban and rural areas —corroborated both in figure 4, which shows the enormous, stubborn and even growing urban-rural poverty gap in Latin America, and in recent publications (Srinivasan and Rodríguez, 2016)—, in the case of cities grouped by population size, the inequalities are less systematic (see table 2).

[7] Unfortunately, migration in the "rest" category is not registered, because is it a residual category treated as a unit without distinguishing one locality (municipality, strictly speaking) from another.

Figure 4
Latin America and the Caribbean: poverty by area of residence and differential between rural and urban areas, 1980-2013
(Percentages and rural/urban ratio)



Source: Economic Commission for Latin America and the Caribbean (ECLAC), CEPALSTAT, on the basis of special tabulations from household surveys of the respective countries, 2015.

Table 2
Latin America and the Caribbean (6 countries): indicators of living conditions (Millennium Development Goals) by city groupings by population size[a]

Cities	Average years of schooling			Net enrolment in primary school	Primary education completion rate[b]	Literacy rate	Male/female ratio			Literacy rate
	Both sexes	Men	Women				Primary education	Primary education	Tertiary education	
1 000 000 or more	10.0	10.3	9.7	80.6	98.3	98.9	1.02	0.99	0.94	98.3
500 000-999 999	10.2	10.5	9.9	74.6	96.8	98.7	1.02	0.97	0.91	97.0
100 000-499 999	9.7	9.8	9.5	82.5	97.2	98.8	1.02	0.99	0.88	97.2
50 000-99 999	8.6	8.9	8.3	78.2	97.8	98.4	1.02	0.98	0.94	96.6
20 000-49 999	8.2	8.5	8.0	78.7	96.5	98.0	1.02	0.96	0.90	96.1
Cities	Proportion of the population with access to drinking water[b]	Proportion of the population with access to sanitation	Proportion of the population with access to electricity	Telephone in the household	Mobile telephone in the household	Computer in the household	Internet in the household	Masculinity ratio	Youth ratio	Old-age ratio
1 000 000 or more	84.6	96.2	99.5	64.4	75.8	43.9	31.9	94.2	41.8	14.6
500 000-999 999	93.4	79.8	99.0	54.1	82.7	42.8	33.4	93.7	41.7	13.9
100 000-499 999	83.7	94.9	90.2	49.6	78.4	38.7	25.9	93.8	44.8	14.3
50 000-99 999	83.9	84.5	93.0	41.2	69.6	29.9	18.6	94.3	49.1	14.0
20 000-49 999	82.6	79.6	93.0	36.9	67.2	25.7	15.4	94.4	50.3	15.4

Source: Prepared by the author, on the basis of information from the database Spatial distribution and urbanization in Latin America and the Caribbean (DEPUALC).

[a] Includes six countries with census information available from the 2010 census round: Costa Rica, Dominican Republic, Ecuador, Mexico, Plurinational State of Bolivia and Uruguay.

[b] More details on the indicators can be found in the DEPUALC database.

The main pattern observed is that small cities generally tend to have lower standards of living, which pushes the population to emigrate towards the higher levels of the city system, certainly not towards the rural sphere which, as shown in figure 4, has much lower indicators.

2. City systems and internal migration: continuity and change in the migratory pull and growth effect for particular population subgroups

The tables below present the best known traditional synthetic indicators of migration —migratory balance and migration rates— for the two variables that define the demographic structure of the population: sex and age. In the case of migratory balance, since these are absolute numbers, the table is a summary of the regional totals for each category of the human settlements system, obtained from the sum of the migratory balances of all the cities included in the analysis, so it offers a sort of regional migratory balance. To this are added indicators of the quantity and percentage of the cities of net in-migration and out-migration, giving a first picture of the diversity behind the regional totals.

The migration rates, conversely, are shown disaggregated by country, because: (i) they are relative figures so can be used for comparison between countries, and (ii) this makes it possible to control for the dominant effect exerted by Brazil and Mexico on the regional averages, owing to their demographic weight and number of cities. The rates are disaggregated by sex only, since disaggregations by age would take up too much space (although these are available upon request). The presentation and analysis of these indicators are also a preamble to the following section, which analyses the results of an innovative procedure developed by CELADE-Population Division of ECLAC to estimate the effect of internal migration on the sex and age composition of the different categories of cities, as well as their educational level.

Tables 3 and 4 show that large cities are still the most attractive for women, and that the lower band of the human settlements system, especially rural areas, expels more women than men. Although the migratory balance is falling for both sexes, in the case of men this is leading to migratory equilibrium, whereas for women the balance is still positive by around 200,000 migrants. By contrast, the migratory balance of medium-sized cities shows no great difference by sex, quantity or trend, although the balance is slightly larger for women. The migratory balances of the lower part of the human settlements system maintain their traditional tendency to expel both sexes, although this is clearly more marked in the case of women, in whose case these areas show a negative balance of over 400,000 migrants.

With regard to age, there is a pattern that stands out for its persistence, universality and magnitude. Young people (aged 15-29 years) are strongly attracted to large cities and, conversely, leave small cities and rural areas (represented by the category “rest” in table 5) in large numbers. This behaviour is very marked, as table 5 shows that in both censuses large cities lost population from all the other age groups, but the gain in the youth population offsets this loss to such an extent that the large cities show a positive migratory balance at the regional level. Although the youth group has been not unaffected by the generalized fall in the migratory balance of the large cities, in this case the reduction is nowhere near a plunge.

Disaggregating by age also reveals a new —and, to a point, unexpected— trend. In the 2010 census round, the “rest” category showed positive balances in three age groups —under 15 years, 30-44 years and 45-59 years— which marks a departure from the observations of the 2000 census round. In the classic models of migration by age (Moultrie and others, 2013; Rogers and Castro, 1982), this combination is usually associated with what is termed family migration, i.e. families of adults and children migrating together.

Table 3

Latin America (selected countries): migratory balance by sex and range of population size of cities and settlements, population aged 5 years and over[a]

(Number of persons)

	On the basis of data from the 10 countries selected					
Population size of cities and the rest of the human settlements system	Men		Women		Total	
	2000	2010	2000	2010	2000	2010
1 000 000 or more	303 500	1 641	583 485	198 347	886 985	199 988
500 000-999 999	192 562	151 478	260 806	198 972	453 368	350 449
100 000-499 999	242 468	261 659	275 570	257 414	518 038	519 073
50 000-99 999	-20 212	-33 148	-29 677	-50 603	-49 889	-83 752
20 000-49 999	-224 751	-123 632	-305 892	-175 616	-530 643	-299 249
Under 20 000	-10 802	-1 125	-33 343	-1 199	-44 145	-2 324
Rest	-482 766	-256 872	-750 948	-427 314	-1 233 713	-684 186
	On the basis of data from the 8 countries with census information for both years					
Population size of cities and the rest of the human settlements system	Men		Women		Total	
	2000	2010	2000	2010	2000	2010
1 000 000 or more	303 500	4 534	583 485	182 308	886 985	186 842
500 000-999 999	192 562	151 478	260 806	198 972	453 368	350 449
100 000-499 999	242 468	264 312	275 570	257 986	518 038	522 298
50 000-99 999	-20 212	-34 311	-29 677	-51 529	-49 889	-85 841
20 000-49 999	-224 751	-123 414	-305 892	-174 314	-530 643	-297 729
Under 20 000	-10 802	-1 234	-33 343	-1 273	-44 145	-2 507
Rest	-482 766	-261 364	-750 948	-412 149	-1 233 713	-673 513

Source: Prepared by the author, on the basis of information from the Database on Internal Migration in Latin America and the Caribbean (MIALC).

[a] Includes 10 countries with information available from the 2010 census round: Bolivarian Republic of Venezuela (2011), Brazil (2010), Costa Rica (2011), Dominican Republic (2010), Ecuador (2010), Honduras (2013), Mexico (2010), Panama (2010), Plurinational State of Bolivia (2012) and Uruguay (2011); and 8 with information available from the 2000 census round: Bolivarian Republic of Venezuela (2001), Brazil (2000), Costa Rica (2000), Dominican Republic (2002), Ecuador (2001), Honduras (2001), Mexico (2000) and Panama (2000).

Table 4

Latin America (selected countries): annual average rate of net migration by sex and range of population size of cities and settlements, population aged 5 years and over

(Number of persons per 1,000)

Population size of cities and the rest of the human settlements system	Bolivia (Plurinational State of), 2012		Brazil, 2000		Brazil, 2010		Costa Rica, 1984		Costa Rica, 2000		Costa Rica, 2010		Ecuador, 1990		Ecuador, 2001		Ecuador, 2010		Honduras, 2001		Honduras, 2013	
	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women
1 000 000 or more	-0.2	1.5	1.3	2.5	0.3	0.8			-1.8	-1.8	-4.7	-3.6	5.5	8.2	5.8	7.2	0.7	1.6			-0.9	0.9
500 000-999 999			4.8	5.8	4.2	4.5	-1.0	3.0											4.9	6.5	0.6	1.7
100 000-499 999	-0.8	0.1	2.8	3.0	2.7	2.7			6.9	7.2	3.9	3.7	-0.1	1.5	0.3	1.0	-1.0	-0.5	1.1	2.6	4.7	5.5
50 000-99 999	1.9	2.7	-0.5	-0.4	-0.3	-0.6	4.3	5.5	-5.4	-5.5	-3.6	-4.1	-12.3	-11.1	-8.9	-7.9	0.1	-0.2	1.4	0.9	-0.7	-0.7
20 000-49 999	3.0	2.4	-4.0	-5.1	-1.9	-2.4	0.5	-2.0	-1.6	-1.6	0.6	0.4	-7.3	-8.6	-7.4	-8.7	-0.4	-0.8	2.1	2.1	-0.8	-1.5
Under 20 000	2.1	1.5	0.4	-1.5	-5.1	-5.4	0.5	-1.4	1.3	1.6			-0.6	-3.5	-1.3	-3.4			1.2	-0.7		
Rest	-0.2	-2.6	-2.7	-4.8	-1.8	-2.7	-1.7	-5.3	-1.8	-2.1	3.4	2.9	-0.4	-2.8	-2.4	-4.0	-0.1	-1.1	-3.6	-4.8	-0.1	-1.1

Population size of cities and the rest of the human settlements system	Mexico, 2000		Mexico, 2010		Panama, 1990		Panama, 2000		Panama, 2010		Dominican Republic, 2002		Dominican Republic, 2010		Uruguay, 1996		Uruguay, 2011		Venezuela (Bolivarian Republic of), 2001		Venezuela (Bolivarian Republic of), 2011	
	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men	Women
1 000 000 or more	1.2	1.7	-0.9	-0.4			15.4	16.0	10.9	10.9	5.1	6.6	7.8	9.3	1.6	2.1	-0.4	0.3	-5.9	-4.6	-2.6	-2.2
500 000-999 999	1.5	2.1	1.6	2.3	3.1	5.7					3.6	4.6	-1.5	0.4					2.4	3.2	-0.2	-0.2
100 000-499 999	2.7	2.7	2.4	1.9			1.4	2.8	-1.6	-2.0	-1.1	-1.8	-3.5	-4.3			-2.7	-2.9	1.3	1.6	1.7	1.7
50 000-99 999	0.7	0.4	-0.4	-0.6	-1.4	-0.6			4.3	5.6	-10.4	-12.5	-13.6	-16.5	4.7	6.9	-0.5	-1.3	3.4	2.7	1.5	1.5
20 000-49 999	-2.6	-3.5	-0.7	-1.1	-0.2	1.0	-4.6	-5.0	-5.4	-5.6	-9.0	-11.2	-8.2	-11.5	-4.3	-2.1	-4.6	-4.4	2.1	0.7	1.2	0.6
Under 20 000	-4.0	-3.8			2.7	-0.7	-13.3	-16.2			-12.2	-16.5			0.0	-3.6			6.7	6.0		
Rest	-3.7	-4.5	-1.2	-1.8	-4.0	-7.8	-17.8	-21.5	-13.2	-14.7	1.8	1.5	-0.2	-1.6	-2.4	-6.6	3.9	2.9	1.3	-0.6	0.5	0.3

Source: Prepared by the author, on the basis of information from the Database on Internal Migration in Latin America and the Caribbean (MIALC).

Table 5

Latin America (selected countries): migratory balance by age group and range of population size of cities and settlements, five years before the census[a]

(Number of persons)

Population size of cities and the rest of the human settlements system	5-14 years		15-29 years		30-44 years		45-59 years		60 years or over		Total	
	2000	2010	2000	2010	2000	2010	2000	2010	2000	2010	2000	2010
1 000 000 or more	17 050	-184 862	1 037 373	768 358	-69 788	-199 330	-70 093	-126 394	-27 557	-57 785	886 985	199 988
500 000-999 999	74 752	26 689	192 136	200 555	105 870	57 612	47 262	34 683	33 348	30 910	453 368	350 449
100 000-499 999	95 177	73 510	206 924	235 858	128 414	122 030	51 791	52 482	35 731	35 192	518 038	519 073
50 000-99 999	3 862	-3 966	-85 225	-88 743	9 968	4 366	10 947	15	10 559	4 576	-49 889	-83 752
20 000-49 999	-85 829	227	-365 159	-293 978	-57 458	-3 653	-14 661	-2 315	-7 535	470	-530 643	-299 249
Under 20 000	-863	-159	-54 953	-1 877	5 686	-448	4 330	-16	1 655	175	-44 145	-2 324
Rest	-104 148	88 560	-931 096	-820 173	-122 692	19 423	-29 577	41 544	-46 200	-13 540	-1 233 713	-684 186

Source: Prepared by the author, on the basis of information from the Database on Internal Migration in Latin America and the Caribbean (MIALC).

[a] Includes 10 countries with information available from the 2010 census round: Bolivarian Republic of Venezuela (2011), Brazil (2010), Costa Rica (2011), Dominican Republic (2010), Ecuador (2010), Honduras (2013), Mexico (2010), Panama (2010), Plurinational State of Bolivia (2012) and Uruguay (2011); and 8 with information available from the 2000 census round: Bolivarian Republic of Venezuela (2001), Brazil (2000), Costa Rica (2000), Dominican Republic (2002), Ecuador (2001), Honduras (2001), Mexico (2000) and Panama (2000).

3. Net and exclusive effects of internal migration on population structure by sex and age

The main results of the application of the innovative procedure developed by CELADE-Population Division of ECLAC for estimating the effect of internal migration on the composition of the population may be summarized as follows:

(i) In almost all the countries, internal migration continues to reduce the masculinity index of large cities, with the exception of a few cities such as San José or Panama City, where this effect has dissipated as net migration rates have converged between the two sexes. In general, this "feminizing" effect has lessened, although with (probably circumstantial) fluctuations in some countries. The strongest feminizing effect is seen in the large cities of Ecuador, between 1985 and 1990, when migration led to a fall of 1.4% in the masculinity ratio. Although at first sight this does not appear to be a large figure, in comparative demographic terms it does represent an exceptional shift, because falls of that magnitude in the masculinity ratio of national populations or cities in just five years are normally the result of highly gender-biased mortality events, such as wars. The reverse side of the sharp feminization of the large cities is the continued masculinization of small cities and the rural environment. In some countries, these segments show rises of around 1% in the masculinity ratio with respect to the counterfactual scenario of no migration in the reference period.

(ii) The "rejuvenating effect" of migration on large cities is fully confirmed. Almost all the countries saw rises of over 1% in the proportion of youth with respect to the non-migration (counterfactual) scenario in the last five years. This figure exceeded 3% in several countries and came close to 5% in the most extreme cases, such as Panama, in the 2000 census round (see table 6). Given the importance of the calculations of this effect, and the possibility of using them to illustrate inputs, results, potentials and limitations of the procedure used, table 6 also shows other results obtained. The first three columns contain the factual, counterfactual and non-migrant values for the percentage of young people.[8] There is an obvious disparity between the large and medium-sized cities, and the small cities and the "rest", which is particularly pronounced in the Dominican Republic, Ecuador, Honduras, Panama, the Plurinational State of Bolivia and Uruguay, where the difference between the factual value (which includes migration that has actually occurred) between large cities and the "rest" is 2 percentage points in Uruguay and as much as 5 percentage points in the Plurinational State of Bolivia.[9] In principle, these differences should occur in the opposite direction, because the more advanced stage of the already long-standing and rapid demographic transition in the large cities generates an age structure with a smaller proportion of young people and larger proportion of mature adults and older persons. In this regard, the disparities observed, which cannot be attributed to inertial population dynamics, can only be produced by the cumulative effect of internal migration. The second and third columns show the percentage young people would represent in the absence of internal migration between city categories and the percentage of young people among non-migrants, which are inputs for subsequent calculations. The fourth column shows the absolute difference between the factual and counterfactual values, which is the net and exclusive effect of migration on the percentage of young people. This is referred to as an absolute effect, because it is a subtraction of original values (in this case, percentages). In all the countries, this effect is positive

8 It will be recalled that "percentage of young people" refers to the population aged between 15 and 29 years within the total population, and does not include those aged under 5 or recent international migrants.

9 Only the Bolivarian Republic of Venezuela and Costa Rica show differences in the other direction, i.e. a larger percentage of young people elsewhere than in the large cities.

in the case of large cities and negative in the case of small cities and the rest. The greatest effect occurs in Panama, where migration raises the percentage of young people in large cities by 1.2 percentage points and reduces it by 1.6 percentage points in the “rest” category. The fifth column shows the data used for comparative purposes, because these standardize, in the form of a ratio, the absolute effect with respect to the counterfactual value of the attribute. Again, the largest relative effect occurs in Panama, where internal migration raises the percentage of young people in large cities by 4.5% and lowers it by 5.7% in the “rest” category. Finally, the last two columns show the absolute effect of in-migration and out-migration, the sum of which gives the total absolute effect. This breakdown is essential to properly interpret the processes underlying the effect of migration on age structure. In fact, the case shown in table 6 is very illustrative in this respect. Why does net migration raise the percentage of young people in cities? Let us take the case of the large cities of the Plurinational State of Bolivia, according to the 2012 census, to follow the arguments with figures. First, young people are more highly represented in the age structure of immigrants than in the age structure of non-migrants, as deduced from comparison between the percentage of young people among non-migrants (33.3% in the case of Bolivia’s large cities according to the 2012 census) and the factual percentage (34.1% in the same case), which has only two components: the percentage of non-migrants and the percentage of immigrants. Second, emigrants too have an age structure more heavily concentrated in youth than non-migrants, as deduced from comparison between the percentage of young people among non-migrants and the counterfactual percentage (33.9% in this case), which has only two components: the percentage of non-migrants and the percentage of emigrants. Third, the effect of net migration reflects the fact that the rejuvenating effect of in-migration outweighs the “counter-rejuvenating” effect of out-migration, either because immigrants have an age structure with a larger proportion of young people than emigrants, or because the number of immigrants far exceeds the number of emigrants in this age group. Whatever the case, the procedure estimates each effect precisely; thus, the absolute effect of net migration —a rise of 0.4 percentage points in the youth population— arises from the 0.9-percentage-point elevating effect of in-migration combined with the 0.4-percentage-point lowering effect of out-migration (the sum of these two effects does not coincide with the respective total, owing to the round of decimals).

(iii) Internal migration tends to reduce, sometimes sharply, the proportion of young people in small cities and in the “rest” category, which in extreme cases falls as much as 8% with respect to the non-migration (counterfactual) scenario over the past five years. This fall is certainly being caused by the still massive out-migration of young people from these categories of the human settlements system, as shown in table 5.

(iv) The reverse side of this contraction of the youth segment of the population owing to migration away from small cities and rural areas is the rise in the relative weight of the other age groups, particularly those aged under 15 and over 59 years. So, although migration tends to increase the relative weight of adults aged 30-59 years in the lower bands of the city system, the end result is nevertheless a rise in the dependency ratio, which dilutes and shortens the demographic dividend in these places.

Finally, the effects on education levels are fairly tenuous —or at least not as sharp as the effects on the age structure— and do not have an obvious counterpart like the other two variables. This remains the case after controlling for age. Both in large cities and in small cities and the rest, migration tends to produce a fall in the level of schooling. Only medium-sized cities register positive effects, although very slight ones in most cases.

Table 6

Latin America (selected countries): net and exclusive effect of net migration (absolute and relative), in-migration (absolute) and out-migration (absolute) on the percentage of the population aged 15–29 years, by population size of cities and settlements

Categories of size of the human settlements system	Bolivia (Plurinational State of), 2012 (2007-2012)							Brazil, 2010 (2005-2010)						
	Factual (*per 100*)	Counterfactual (*per 100*)	Non-migrants (*per 100*)	Absolute effect (*percentage points*)	Relative effect (*per 100*)	Effect of in-migration (*percentage points*)	Effect of out-migration (*percentage points*)	Factual (*per 100*)	Counterfactual (*per 100*)	Non-migrants (*per 100*)	Absolute effect (*percentage points*)	Relative effect (*per 100*)	Effect of in-migration (percentage points)	Effect of out-migration (*percentage points*)
1 000 000 or more	34.1	33.7	33.3	0.4	1.3	0.9	-0.4	29.8	29.2	28.9	0.6	2.0	0.8	-0.3
500 000-999 999				None				29.1	28.4	28.0	0.7	2.3	1.0	-0.4
100 000-499 999	35.3	35.1	33.9	0.2	0.7	1.4	-1.2	29.1	28.9	28.2	0.2	0.8	0.9	-0.7
50 000-99 999	33.6	33.9	32.8	-0.2	-0.7	0.9	-1.1	29.1	29.3	28.2	-0.2	-0.6	0.9	-1.1
20 000-49 999	32.6	32.7	31.0	-0.2	-0.6	1.5	-1.7	29.2	29.8	28.5	-0.6	-1.9	0.6	-1.2
Rest	29.1	29.7	28.6	-0.6	-2.0	0.5	-1.1	28.7	29.8	28.5	-1.0	-3.5	0.2	-1.2
Categories of size of the human settlements system	Costa Rica, 2011 (2006-2011)							Ecuador, 2010 (2005-2010)						
	Factual (*per 100*)	Counterfactual (*per 100*)	Non-migrants (*per 100*)	Absolute effect (*percentage points*)	Relative effect (per 100)	Effect of in-migration (*percentage points*)	Effect of out-migration (*percentage points*)	Factual (*per 100*)	Counterfactual (*per 100*)	Non-migrants (*per 100*)	Absolute effect (*percentage points*)	Relative effect (*per 100*)	Effect of in-migration (*percentage points*)	Effect of out-migration (*percentage points*)
1 000 000 or more	29.3	28.9	28.7	0.5	1.6	0.6	-0.1	30.5	29.7	29.4	0.8	2.8	1.1	-0.3
500 000-999 999				None							None			
100 000-499 999	29.9	29.5	29.3	0.4	1.4	0.6	-0.1	31.0	31.1	29.9	-0.1	-0.3	1.1	-1.2
50 000-99 999	30.4	31.0	30.1	-0.7	-2.1	0.3	-0.9	29.4	29.9	28.6	-0.5	-1.8	0.8	-1.3
20 000-49 999	30.7	31.1	30.4	-0.4	-1.2	0.4	-0.7	30.4	31.0	29.5	-0.6	-1.9	0.9	-1.5
Rest	29.9	30.4	29.6	-0.5	-1.6	0.3	-0.8	29.0	29.5	28.5	-0.5	-1.8	0.4	-1.0
Categories of size of the human settlements system	Honduras, 2013 (2008-2013)							Mexico, 2010 (2005-2010)						
	Factual (*per 100*)	Counterfactual (*per 100*)	Non-migrants (*per 100*)	Absolute effect (*percentage points*)	Relative effect (per 100)	Effect of in-migration (*percentage points*)	Effect of out-migration (*percentage points*)	Factual (*per 100*)	Counterfactual (*per 100*)	Non-migrants (*per 100*)	Absolute effect (*percentage points*)	Relative effect (*per 100*)	Effect of in-migration (*percentage points*)	Effect of out-migration (*percentage points*)
1 000 000 or more	35.1	34.5	34.3	0.6	1.8	0.8	-0.2	29.8	29.2	28.9	0.6	2.0	0.8	-0.3
500 000-999 999	36.1	35.0	34.9	1.1	3.1	1.2	-0.1	29.1	28.4	28.0	0.7	2.3	1.0	-0.4
100 000-499 999	35.0	34.5	34.1	0.5	1.5	0.9	-0.4	29.1	28.9	28.2	0.2	0.8	0.9	-0.7
50 000-99 999	34.2	34.4	33.7	-0.2	-0.7	0.4	-0.7	29.1	29.3	28.2	-0.2	-0.6	0.9	-1.1
20 000-49 999	33.6	34.0	33.0	-0.4	-1.2	0.6	-1.0	29.2	29.8	28.5	-0.6	-1.9	0.6	-1.2
Rest	32.6	33.0	32.4	-0.4	-1.2	0.2	-0.6	28.7	29.8	28.5	-1.0	-3.5	0.2	-1.2

Table 6 (concluded)

Categories of size of the human settlements system	Panama, 2010 (2005-2010)							Dominican Republic, 2010 (2005-2010)						
	Factual (*per 100*)	Counterfactual (*per 100*)	Non-migrants (*per 100*)	Absolute effect (*percentage points*)	Relative effect (*per 100*)	Effect of in-migration (*percentage points*)	Effect of out-migration (*percentage points*)	Factual (*per 100*)	Counterfactual (*per 100*)	Non-migrants (*per 100*)	Absolute effect (*percentage points*)	Relative effect (*per 100*)	Effect of in-migration (*percentage points*)	Effect of out-migration (*percentage points*)
1 000 000 or more	28.3	27.1	27.0	1.2	4.5	1.3	-0.1	31.2	30.5	30.4	0.7	2.4	0.8	0.0
500 000-999 999				None				30.5	29.9	29.6	0.6	1.9	0.9	-0.3
100 000-499 999	28.3	28.3	27.6	-0.1	-0.4	0.7	-0.8	30.5	30.6	30.0	-0.1	-0.3	0.5	-0.6
50 000-99 999	28.8	28.3	26.4	0.5	1.6	2.3	-1.9	29.5	31.2	29.3	-1.6	-5.2	0.2	-1.8
20 000-49 999	26.8	27.7	26.0	-0.9	-3.2	0.9	-1.8	28.9	29.3	28.6	-0.4	-1.4	0.3	-0.7
Rest	25.9	27.5	25.6	-1.6	-5.7	0.2	-1.8	28.9	29.3	28.6	-0.4	-1.4	0.3	-0.7
Categories of size of the human settlements system	Venezuela (Bolivarian Republic of), 2011 (2006-2011)							Uruguay, 2011 (2006-2011)						
	Factual (*per 100*)	Counterfactual (*per 100*)	Non-migrants (*per 100*)	Absolute effect (*percentage points*)	Relative effect (*per 100*)	Effect of in-migration (*percentage points*)	Effect of out-migration (*percentage points*)	Factual (*per 100*)	Counterfactual (*per 100*)	Non-migrants (*per 100*)	Absolute effect (*percentage points*)	Relative effect (*per 100*)	Effect of in-migration (*percentage points*)	Effect of out-migration (*percentage points*)
1 000 000 or more	29.5	29.5	29.3	0.1	0.2	0.2	-0.1	24.5	23.5	23.4	1.0	4.4	1.1	-0.1
500 000-999 999	31.2	31.2	31.0	0.1	0.2	0.2	-0.1				None			
100 000-499 999	30.7	30.7	30.5	0.0	0.0	0.2	-0.2	26.3	27.1	25.7	-0.7	-2.8	0.7	-1.4
50 000-99 999	31.7	31.7	31.5	0.0	-0.1	0.2	-0.2	23.9	24.9	23.3	-0.9	-3.8	0.7	-1.6
20 000-49 999	30.9	31.0	30.7	-0.1	-0.3	0.2	-0.3	22.6	24.0	22.1	-1.4	-5.7	0.5	-1.9
Rest	30.8	31.0	30.6	-0.2	-0.7	0.2	-0.4	22.4	23.3	22.1	-0.9	-3.8	0.4	-1.2

Source: Prepared by the author, on the basis of information from the Database on Internal Migration in Latin America and the Caribbean (MIALC).

V. Discussion and conclusions

Three of every four migrants originate and end up in urban areas, according to the 2010 census round; a slightly higher proportion than in the 2000 census round. The average for the region —strictly speaking, for the countries included in the calculations— does not mean that there are no countries where rural-urban migration still prevails. However, these are the exception and the process is dwindling inexorably.

These results contrast with the great deficit of theory, policy, data and research on inter-city migration. The results shown here show that this oversight is unwise, because what happens in its cities is crucial for the country overall and migration reveals both strengths and weaknesses in cities, the city system or the human settlements system, and it poses challenges for policymaking and public action in general. In particular, migration is showing mixed symptoms in the case of the large cities and signs of stagnation at the base of human settlements system. At the same time, the apparent strengths of the intermediate segment will probably encounter limits and major challenges in the near future.

The slowing of in-migration to the large cities is good news, in principle, bearing in mind the complications caused by mass migration in the past. Although past migration endowed cities with much-needed labour force at a time of strong manufacturing growth, the economic system was unable to fully absorb it. The influx also exerted heavy pressure on the infrastructure, services and governance of large cities, which national and local governments were not able to manage sustainably. The lost decade of the 1980s rapidly turned these complications into severe problems, with steep rises in poverty, high unemployment, poor public safety and urban deficits in general, including in governance. In fact, the living standards indicators obtained from censuses show that large cities no longer enjoy any obvious advantages in this regard, which has led them to slowly but steadily lose their migratory pull. Indeed, large cities today are at virtual equilibrium in terms of migration, with a still-positive rate of around 0.3 per 1,000 and subdued population growth, given that not only migration, but also fertility has fallen. Relentless migratory pressure has thus ceased to be a challenge for these cities.

Nor does this migratory quasi-equilibrium in large cities mask anomalies akin to the high rates seen at the height of the rural-urban in-migration of the 1950s to 1980s. The highest rates do not exceed a 1% annual average, and these correspond to cities with specific attractions —like Brasilia or Santa Cruz (Plurinational State of Bolivia), which combine public investment and government employment with the momentum provided by dynamic export activities in the first decade of the twenty-first century— or cities in the process of forming “metropolitan regions” with existing supercities —like Campinas or Santos in Brazil, close to the megacity of São Paulo (Cunha, 2015). Conversely, the migratory quasi-equilibrium does hide a large and growing number of cases of net out-migration. Although still a minority of cases, this group includes all the supercities (those with upwards of 10 million inhabitants) covered in the study. Thus, without there being a causal relationship, reaching the threshold of 10 million inhabitants is associated with a migratory turning point at which a city becomes an expeller. In general, this is not good news, not so much because of the direct demographic effect involved, but because of the implications: the predominance of push factors likely associated with urban and governance problems, as well as diseconomies and rising costs in these supercities. Be this as it may, net out-migration avoids migration increasing the population in these cities, which already have large number of inhabitants and serious governance issues, partly owing to their large demographic and territorial size.

Although the differential migration of women to large cities has lessened, net out-migration cities still attract women more and expel them less. The opportunities structure in the cities still offers women more options or, from the opposite perspective, the net-loss sectors of the human settlements system —that is, the lower bands— lack options for women, at least in relative terms.

In the case of age-differentiated migration, even supercities that are losing population continue to attract young people, as their opportunities structure appears to be particularly attractive for this age group. This phenomenon has been little studied as yet in the region, although it has been quite well documented in other parts of the world (Williamson, 1988; Florida, 2005; Pacione 2009). Its underlying causes may readily be divined: more and better options for work, study and life plans in general (including cultural consumption, finding a partner and the use of leisure time) for this age group in large cities. The concentration of tertiary education institutions in large cities increases this attraction, as do standards and paces of living that suit young people well although they may not be ideal for other age groups. The range of services and housing operations in large cities is also more suited to young people, by type if not price. Net youth in-migration contrasts with the net out-migration of other age groups, which amplifies the rejuvenating effect of internal migration on the age structure of large cities. The procedure developed by CELADE-Population Division of ECLAC shows that this last effect reinforces and extends the demographic dividend in large cities, owing to the downward pressure it exerts on the dependency index. The persistent, mass influx of young people also has other social and economic effects that are more difficult to estimate, but that various authors argue boost the economy and culture of the cities in question (Florida, 2005). Not all the effects of migration on the population composition of large cities are beneficial, however. In particular, the procedure applied in this work shows that migration reduces average levels of education in large cities, something that is not attributable to the age structure of migrants. The loss is not a significant one, but the fact that it exists at all warrants further research, at least to identify whether it reflects the in-migration of population with a low level of education, or the out-migration of population with a high level of education, a question which will be addressed in future work.

What do the data on migration reveal about medium-sized cities? Clearly, these have become consolidated as the most attractive segment in the human settlements system. This had already been noted in previous research (Rodríguez, 2011) and had been suggested by other work on the basis of population growth indices (Villa and Rodríguez, 1998). But it had not yet been demonstrated with the most recent census figures available, with concrete data that suggest a moderate pull factor not comparable to the high figures of past decades. In principle, this should ease demographic pressure on infrastructure, equipment and the various services and amenities for which cities are responsible, even though these need to keep expanding to facilitate migrants' social and labour integration and guarantee respect for their rights. The attraction of medium-sized cities is consistent with their indicators of living standards, which exceed those of large cities and far exceed those of small cities. In addition, the size of medium cities offers some governance and quality of life, which are powerful pull factors.

This said, the segment of medium-sized cities shows great diversity. In fact, there is a sharp contrast, especially among cities with between 100,000 and 499,999 inhabitants, many of which register net out-migration. It may be because they offer an immediate alternative to large cities, that the upper-medium cities (between 500,000 and 999,999 inhabitants) are the group with the highest rate of net migration and the smallest proportion of expelling cities. This is different from the case of cities with 100,000 to 499,999 inhabitants, which in the larger countries may form part of the lower segment of the cities system and show similar trends to those of small cities.

These cities also resemble the large cities in that they receive higher percentage of female and youth in-migration. In fact, in some countries, the percentage of young people has grown more in medium-sized cities than large cities as a result of migration. In addition, and unlike large cities, medium-sized cities, tend to gain in average schooling levels with migration. However, these "positive" effects are more systematic and noticeable in the case of the upper-medium cities.

In sum, their migratory pull suggests that medium-sized cities offer better socioeconomic conditions and living standards, which is borne out by the limited census data available in this respect. The effects of migration on the composition of the population tend to make these cities —especially upper-medium cities— more competitive and innovative. This is therefore a segment that benefits from in-migration and faces the challenge of administering its dividends in order to progress towards sustainable development.

What do the data on migration reveal about the lower levels of the human settlements system, the small cities and rural areas? All three categories of the lower segment of the system used in this research (cities with between 50,000 and 99,999 inhabitants, cities with between 20,000 and 49,999 inhabitants and the "rest", a category which groups the MIADs where there are no cities of 20,000 or more inhabitants) were found to show population loss and the proportion of cities by migratory pull or push showed that the great majority registered net out-migration. In fact, this explains the paradox of having a majority of population-losing cities in a region where the process of urbanization continues exclusively owing to the —albeit decreasing— persistence of rural-urban migration.

The net out-migration registered by these parts of the settlements system is both a matter for concern and a problem. The fact that this segment does not experience the pressure of rapid growth from internal migration hides structural lags that generate net out-migration. These lags can be seen at a basic level in the indicators of living conditions calculated from census data, as well as in the poverty indices derived from the surveys, which are much higher in rural areas. Despite decades of rural exodus —which eroded the supply of migrants— and despite industrial policies aimed at renovating the primary activities that tend to be concentrated in this segment of the human settlements system, and a broad range of strategies aimed at improving and upgrading (including decentralization, local development and rural development), the indicators of well-being and access to services in this segment remain well below the rest of the human settlements system. As well, job creation is still insufficient and often not aimed at the local population, but at outside workers whose jobs do not require them to settle permanently in the area, and with income much lower than can be earned in the larger cities.

The problem lies in the fact that the bulk of the out-migration-related population loss consists of people who are younger and/or more educated than those remaining behind. As a result, this segment of the cities system is more aged and its demographic dependency ratios are much higher than natural population trends would suggest they should be. The demographic dividend is smaller and lasts for a shorter period. Out-migration is also biased towards women, whose skills still do not seem to have the space to fully develop in these localities. The only aspect of population composition that does not show stylized adverse effects is average schooling, which does not appear to depend on the age structure of migrants, since the results vary little when age is controlled for.

The "rest" segment shows a turning point in net migration (from negative to positive) in several age ranges other than young people (those aged 15-29 years). This is an interesting phenomenon, which could be interpreted as a sign of families with children returning to the rural or semirural environment. However, owing to the diversity of the "rest" category in the methodology used —which include from MIADs consisting entirely of scattered rural populations to MIADs that are in the midst of "rururbanization" but have not yet integrated with the nearest city— it is likely that this category is capturing municipalities in the process of "rururbanization" around nearby cities (Aguilar and Escanilla, 2011; Ávila, 2009; Pacione, 2009; Champion, 2008; Arroyo, 2001), especially in view of the family slant this process usually takes. This is something that should be assessed in future research, which would need to open the black box of the "rest" segment and, possibly, differentiate between distinct types of municipalities, for example between fully rural isolated municipalities and "rururban" municipalities. This would challenge the procedure used here and the databases employed for this work, which currently lack the information that would be needed to make such a distinction. More research is thus needed to identify the reasons for this emerging migratory pull.

In short, at the time of import substitution industrialization, all forces tended to favour migration towards large cities, the hubs of demand for industrialization-related labour, the best wages and supply of education, the highest levels of basic services and access to goods and services, the lowest levels of poverty, and a range of technological and cultural innovations that fed expectations of a better quality of life. Those times have changed, however, and the large cities present more contrasts than in the past. They still have pull factors, such as supply of education, skilled employment, positions of power and access to cutting-edge technology, but also clear push factors, such as labour informality and precarious living conditions, falling standards —and rising costs—of living and a build-up of urban deficits (ECLAC, 2012). In this context, the post-Fordist model of production and technological innovations facilitate the deconcentration of employment, at least towards the medium-sized cities, and other nodes of the cities system become economically competitive in relation to large cities. Thus they receive public and private investment which brings their infrastructure, facilities and services closer to those of the large cities, and they gain major relative advantages in terms of governance and quality of life. Nevertheless, they still lag behind in key aspects such as education, culture and recreation, where the large cities retain the lead, at least in Latin America and the Caribbean.[10] Large cities also still exert a significant social and economic pull and in fact remain attractive despite their multiple problems, showing a resilience that could help them retain their key role in the future.

The competition between large and medium-sized cities takes on a different hue in the comparison with the rest of the cities system and especially with the rural segment of the human settlements system. Small cities and rural areas still suffer much more from poverty, industrial lag and lack of services and infrastructure. Factors such as the higher cost of social investment, the limited capacities and resources of local governments, the almost complete absence of higher education institutions and centres of educational excellence, the lack of skilled human resources (partly due to out-migration) and the vicious cycle of deficits that hinder the emergence of social mobility opportunities for the population still outweigh the possible advantages in terms of standard of living, safety and governance. Out-migration can thus further slow the development of the most underdeveloped segments of the human settlements system, sharpening social inequalities instead of reducing them as the prevailing theories contend (ECLAC, 2015 and 2012; Kanbur and Rappoport, 2005). Admittedly, this is only one effect and not necessarily the dominant one, because the advantages of migration for both territories —including whole countries— and for individuals are well documented (ECLAC, 2012; World Bank, 2009; UNFPA, 2007; Aroca, 2004; Williamson, 1988). What is more, the emergence of a suburban or “rururban” segment composed of areas and localities with characteristics that are formally and scenically rural, but fully urban in terms of lifestyle and day-to-day links with the city, could shift the negative effects of internal migration on rural areas and small cities, as younger, better-off families move into them. This does not represent a return to the countryside, however, but the urbanization of the rural milieu.

[10] In this respect, the location of universities still has a markedly metropolitan bias in almost all the countries of the region and is one of the main forces pulling young people into large cities. The incipient attempts to change this pattern have yet to be evaluated in terms of migratory impact (Rodríguez and others, 2017; Fusco and Ojima, 2016).

Bibliography

Aguilar, A. and I. Escanilla (coords.) (2011), *Periurbanización y sustentabilidad en grandes ciudades*, Mexico City, Editorial M.A. Porrúa.

Alberts, J. (1977), "Migración hacia áreas metropolitanas de América Latina. Un estudio comparativo", *Series E*, No. 24, Santiago, Latin American Demographic Centre (CELADE).

Aroca, P. (2004), "Migración intrarregional en Chile. Modelos y resultados 1987-2002", *Notas de Población*, No. 78 (LC/G.2229-P), Santiago, Economic Commission for Latin America and the Caribbean (ECLAC).

Arroyo, M. (2001), "La contraurbanización: un debate metodológico y conceptual sobre la dinámica de las áreas metropolitanas", *Papeles de Población*, year 7, No. 30, Toluca, Center for Research and Advanced Studies of the Population, Autonomous University of the State of Mexico.

Atienza, M. and P. Aroca (2012), "Concentración y crecimiento en Chile: una relación negativa ignorada", *EURE*, vol. 38, No. 114, Santiago, Catholic University of Chile.

Ávila, H. (2009), "Periurbanización y espacios rurales en la periferia de las ciudades", *Estudios Agrarios*, No. 41, Mexico City.

Bell, M. and M. Salut (2009), "Cross-national comparisons of internal migration, human development", *Technical Paper*, No. 2009/30, New York, United Nations.

Berg, L. van den and others (1982), *Urban Europe: A Study of Growth and Decline*, vol. 1, Oxford, Pergamon Press.

Brown, L. (1991), *Place, Migration and Development in the Third World*, London, Routledge.

Camisa, Z. (1972), "Efecto de la migración en el crecimiento y la estructura de la población de las ciudades de la América Latina", *Series C*, No. 139, Santiago, Latin American Demographic Centre (CELADE).

Champion, A. (2008), "The changing nature of urban and rural areas in the UK and other European countries" (UN/POP/EGM-URB/2008/07), New York, United Nations.

Cuervo, L. and J. González (1997), *Industria y ciudades en la era de la mundialización. Un enfoque socioespacial*, Bogota, Tercer Mundo Editores.

Cunha, J. (2015), "Dinâmica demográfica e migratória 1991-2010: realidades e mitos", *A metrópole de São Paulo no século XXI: espaços, heterogeneidades e desigualdades*, E. Marques (org.), Editora Unesp.

Cunha, J. and J. Rodríguez (2009), "Crecimiento urbano y movilidad en América Latina", *Revista Latinoamericana de Población*, No. 4-5, Latin American Population Association.

ECLAC (Economic Commission for Latin America and the Caribbean) (2015), "Panorama del desarrollo territorial en América Latina y el Caribe, 2015. Pactos para la igualdad territorial" (LC/W.671), Santiago.

___(2012), *Population, Territory and Sustainable Development* (LC/L.3474(CEP.2/3)), Santiago.

Elizaga, J. C. (1970), "Migraciones a las áreas metropolitanas de América Latina", *Series E*, No. 6, Santiago, Latin American Demographic Centre (CELADE).

Elizaga, J. C. and J. Macisco (1975), "Migraciones internas. Teoría, método y factores sociológicos, *Series E*, No. 19, Santiago, Latin American Demographic Centre (CELADE).

Florida, R. (2005), *Cities and the Creative Class*, New York, Routledge.

Fujita, M. and P. Krugman (2004), "The new economic geography: past, present and the future", *Papers in Regional Science*, vol. 83, No. 1, Wiley.

Fusco, W. and R. Ojima (2016), "Nordeste do Brasil: interiorização do ensino superior e mobilidade pendular" [online] http://187.45.187.130/~abeporgb/xxencontro/files/paper/305-117.pdf.

Geyer, M. and T. Kontuly (1993), "A theoretical foundation for the concept of differential urbanization", *International Regional Science Review*, vol. 15, No. 2, Sage.

Hall, P. (1996), *Ciudades del mañana. Historia del urbanismo en el siglo XX*, Barcelona, Ediciones del Serbal.

Henderson, J. V. (2003), "The urbanization process and economic growth: the so-what question", *Journal of Economic Growth*, vol. 8, No. 1, Springer.

Kanbur R. and H. Rapoport (2005), "Migration selectivity and the evolution of spatial inequality", *Journal of Economic Geography*, vol. 5, No. 1, Oxford University Press.

Martine, G. (1979), "Migraciones internas: ¿investigación para qué?", *Notas de Población*, No. 19, Santiago, Economic Commission for Latin America and the Caribbean (ECLAC).

Moultrie, T. and others (2013), *Tools for Demographic Estimation*, Paris, International Union for the Scientific Study of Population (IUSSP).

Pacione, M. (2009), *Urban Geography. A Global Perspective*, New York, Routledge.

Rodríguez, J. (2013a), "La migración interna en las grandes ciudades en América Latina: efectos sobre el crecimiento demográfico y la composición de la población", *Notas de Población*, No. 96 (LC/G.2573-P), Santiago, Economic Commission for Latin America and the Caribbean (ECLAC).

___(2013b), "How Is Internal Migration Reshaping Metropolitan Populations in Latin America? New Methodologies and New Evidence" [online] https://iussp.org/sites/default/files/event_call_for_papers/IUSSP2013-JR-MigrationandCompositionEffectnLatinAamericaMetropolis-23-08-2013-Final.pdf.

___(2011), "Migración interna y sistema de ciudades en América Latina: intensidad, patrones, efectos y potenciales determinantes, censos de la década de 2000", *Población y Desarrollo series*, No. 105 (LC/L.3351), Santiago, Economic Commission for Latin America and the Caribbean (ECLAC).

Rodríguez, J. and G. Busso (2009), *Migración interna y desarrollo en América Latina entre 1980 y 2005. Un estudio comparativo con perspectiva regional basado en siete países*, Libros de la CEPAL, No. 102 (LC/G.2397–P), Santiago, Economic Commission for Latin America and the Caribbean (ECLAC).

Rodríguez, J. and others (2017), "¿Perdió el Área Metropolitana del Gran Santiago su atractivo? Sí, pero no. Un examen basado en datos y procedimientos novedosos para la estimación de la migración interna y sus efectos durante el periodo 1977-2013", *EURE*, vol. 43, No. 128, Santiago, Catholic University of Chile.

Rogers, A. and L. Castro (1982), "Patrones modelo de migración", *Demografía y Economía*, vol. 16, No. 3, Mexico City.

Romero, J. L. (1976), *Latinoamérica: las ciudades y las ideas*, Buenos Aires, Siglo XXI.

Sassen, S. (2007), "El reposicionamiento de las ciudades y regiones urbanas en una economía global: ampliando las opciones de políticas y gobernanza", *Revista Latinoamericana de Estudios Urbano Regionales (EURE)*, vol. 33, No. 100, Santiago, Catholic University of Chile.

Srinivasan, S. and A. Rodríguez (2016), "Pobreza y desigualdades rurales. Perspectivas de género, juventud y mercado de trabajo", *Desarrollo Productivo series*, No. 206 (LC/L.4206), Santiago, Economic Commission for Latin America and the Caribbean (ECLAC).

UNFPA (United Nations Population Fund) (2007), *State of World Population 2007. Unleashing the Potential of Urban Growth*, New York.

United Nations (2015), *World Urbanization Prospects: The 2014 Revision* (ST/ESA/SER.A/366), New York.

Villa, M. (1991), "Introducción al análisis de la migración: apuntes de clase: notas preliminares", *Series B*, No. 91 (LC/DEM/R.164), Santiago, Latin American and Caribbean Demographic Centre (CELADE)-Population Division of ECLAC.

Villa, M. and J. Alberts (1980), "Redistribución espacial de la población en América Latina", *Series E*, No. 28, Santiago, Latin American Demographic Centre (CELADE).

Villa, M. and J. Rodríguez (1998), "Distribución espacial de la población, urbanización y ciudades intermedias: hechos en su contexto", *Ciudades intermedias de América Latina y el Caribe: propuesta para la gestión urbana* (LC/L.1117), R. Jordán and D. Simioni, Santiago, Economic Commission for Latin America and the Caribbean (ECLAC).

___(1997), "Dinámica sociodemográfica de las metrópolis latinoamericanas durante la segunda mitad del siglo XX", *Notas de Población*, No. 65 (LC/DEM/G.177), Santiago, Economic Commission for Latin America and the Caribbean (ECLAC).

Welti, C. (ed.) (1997), *Demografía I*, Mexico City, Latin American Programme of Population Activities/National Autonomous University of Mexico (UNAM).

Williamson, J. (1988), "Migrant selectivity, urbanization, and industrial revolutions", *Population and Development Review*, vol. 14, No. 2, New York, Population Council.

___(1965), "Regional inequality and the process of national development: a description of the patterns", *Economic Development and Cultural Change*, vol. 13, No. 4, Chicago, The University of Chicago Press.

World Bank (2009), *World Development Report 2009: Reshaping Economic Geography. Overview*, Washington, D.C.

Economic growth and income concentration and their effects on poverty in Brazil

Jair Andrade Araujo, Emerson Marinho
and Guaracyane Lima Campêlo

Abstract

We use panel data for Brazilian states from 1995 to 2009 to analyse the impact of economic growth and income inequality on poverty change in Brazil, seeking to evaluate the Bourguignon (2003) hypothesis that the more unequal a country is, the less effective economic growth will be at reducing poverty. To this end, we estimate poverty elasticities relative to income and inequality, specifying two dynamic econometric models estimated via the generalized method of moments (GMM) system developed by Arellano and Bond (1991), Arellano and Bover (1995) and Blundell and Bond (1998). The model-estimated results prompt the conclusion that the income growth effect on poverty reduction is smaller when the initial development level is low. The same is found when the initial inequality level is high. Therefore, regions with a low initial development level, high initial inequality or both present less favourable conditions for reducing poverty through income growth.

Keywords

Economic growth, income distribution, poverty, poverty mitigation, econometric models, Brazil

JEL classification

D60, D63, C33

Authors

Jair Andrade Araujo is a professor on the Rural Economy master's degree course and Head of the Department of Agricultural Economics at the Federal University of Ceará, Brazil. Email: jairandrade@ufc.br.

Emerson Marinho is a professor on the Economics postgraduate programme at the Federal University of Ceará, Brazil. Email: emarinho@ufc.br.

Guaracyane Lima Campêlo is a professor with the Department of Economics and Finance at the Sobral campus of the Federal University of Ceará, Brazil. Email: cyanelc@gmail.com.

I. Introduction

Many countries' development policies are mainly focused on improving the well-being of the population. Of the many goals this entails, poverty reduction is fundamental, especially considering that poverty often proves persistent despite the economy's rising capacity to generate wealth.

According to Rocha (2006), it is clear that economic expansion does not necessarily benefit everyone in a given society even when achieved successfully. Rich countries are evidently struggling to eliminate residual poverty, while economic growth in developing countries has exacerbated social inequalities.

The question of how changes in income and inequality affect poverty reduction has been brought to the fore in recent years by the proven fact that economic growth has yet to resolve poverty issues in many countries.

For example, Cline (2004) studied several countries during the 1990s and concluded that poverty did fall in many of them as a consequence of economic growth. However, some countries that did not enjoy significant economic growth were also successful in reducing poverty. This shows that economic growth by itself cannot explain poverty changes and that income inequality is of paramount importance as a poverty reduction factor.

Chen and Wang (2001) investigated poverty and inequality in China during the 1990s. After decomposing the poverty changes caused by economic growth and by shifts in inequality, they concluded that it was the rich who profited most from economic growth, as the average income of the wealthiest 20% increased by more than the average income overall. This highlights the importance of income inequality as a variable in poverty reduction.

Barros, Henriques and Mendonça (2001) found that the incidence of poverty was higher in Brazil than in most countries with similar per capita incomes. They concluded that income inequality was the reason why economic growth was relatively inefficient at reducing poverty. In other words, the effect of economic growth on poverty reduction was smaller in Brazil than in other countries with the same income level.

If any change in poverty is the consequence of income redistribution or economic growth (or both), the importance of each effect should be identified. Recent studies have sought to explain which factors affect the income-poverty and inequality-poverty elasticities.

For example, Ravallion and Chen (1997) used a sample of developing countries to demonstrate income-poverty elasticity (measured by the number of people with incomes below the poverty line of one dollar a day), finding this elasticity to be -3. This means that for every 1% rise in the average income, the proportion of individuals with incomes below the poverty line falls by 3%. However, there are countries that have been able to reduce household poverty without significant economic growth.

Bourguignon (2003) estimated income-poverty elasticity for a set of countries, using a log-normal distribution to approximate income distribution. He demonstrated that the higher the average income and the lower its concentration, the higher the resulting elasticity.

Empirical evidence in the case of Brazil has been provided by Marinho and Soares (2003), who estimated the average income elasticity of poverty in the Brazilian states from 1985 to 1999. They concluded that a higher average income resulted in a higher absolute elasticity and that higher income concentration led to lower absolute elasticity. The highest income-poverty elasticity values were found in the states of São Paulo and Rio de Janeiro.

In another study carried out for Brazil, Hoffman (2004) used a different methodology from that of Marinho and Soares (2003) to estimate these same elasticities. He found that both estimations showed very similar standard variations across the different states.

Salvato and Araujo Junior (2007) used data from Brazilian municipalities to investigate the relationship between growth, poverty and inequality, measuring the elasticity of poverty relative to economic growth and changes in income inequality. They also tested for the existence of a non-linear interaction effect between growth and initial inequality, seeking to evaluate the hypothesis that higher inequality was associated with a lessening of the poverty reduction efficiency of growth. They found that of the major regions, the south-east boasted the highest elasticity, while São Paulo was the state that achieved the best results. They also noticed a negative correlation between the elasticity module and initial inequality, which implies that higher initial inequality means a diminution of the poverty reduction brought about by economic growth, corroborating the Bourguignon (2003) hypothesis. The results also suggest a negative correlation between redistribution elasticity and initial inequality.

However, these issues have not been fully clarified, since, according to Barreto (2005), there is still no consensus about the relationship between poverty, growth and inequality. Thus, it is extremely important to determine the effects that each of these factors exerts on poverty.

Measuring these elasticities is a vital part of planning for income growth and redistribution policies, considering that poverty reduction is affected both by shifts in economic growth and by inequality reduction, as Cline (2004) points out.

Taking these facts into consideration, the aim of this work is to analyse the impact of economic growth and shifts in income inequality on poverty changes in Brazil. Since growth alone cannot explain alterations in poverty levels, we treat income inequality as a factor in these, seeking to evaluate the hypothesis that the more unequal a country is, the less effective economic growth will be at reducing poverty (Bourguignon, 2003).

In order to verify these effects, we estimate the elasticity of poverty with respect to income and inequality. These latter variables are estimated by applying a dynamic econometric panel data model developed by Arellano and Bond (1991), Arellano and Bover (1995) and Blundel and Bond (1998). In the panel, the units of analysis are the Brazilian states during the period from 1995 to 2009.

The present article is composed of six sections besides this introduction. Section II reviews the Brazilian and international literature on the triangular relationship between poverty, economic growth and inequality, and offers a brief history of inequality in Brazil. Section III provides theoretical definitions of income-poverty and inequality-poverty elasticities. Section IV discusses the database, while section V introduces the econometric model and its estimation methods and presents an analysis of the results. Section VI contains analysis of the results and lastly, section VII draws the final conclusions.

II. The triangular relationship between poverty, economic growth and inequality

This section presents a review of the literature on the triangular relationship between poverty, economic growth and income inequality. The interaction of these three variables provides a sufficient basis for diagnosing the extent to which income growth or a decrease in inequality affects poverty reduction.

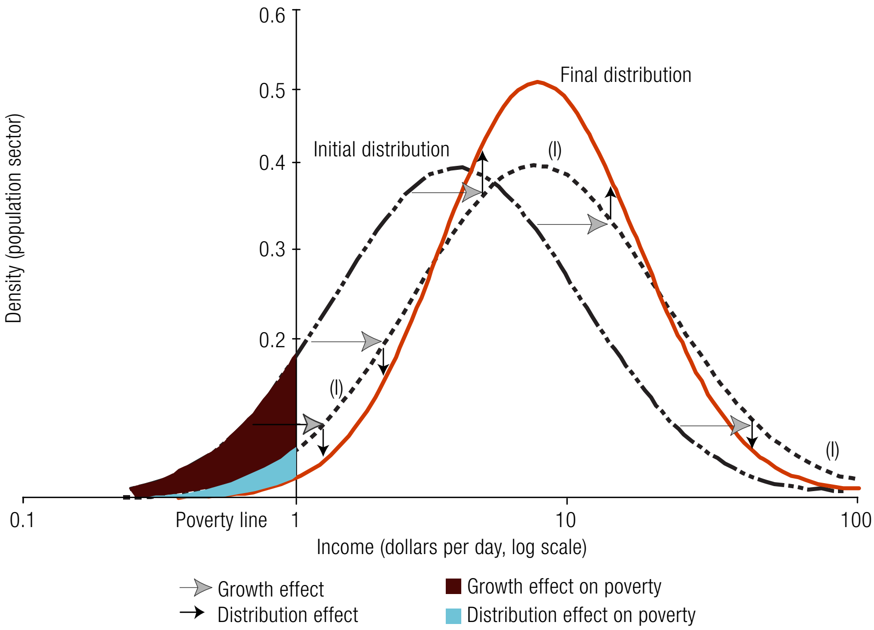
Bourguignon (2003) clearly describes what he calls the poverty-inequality-growth triangle, arguing that the three variables interact. In his article, he assumes the log-normality of income distribution and assigns changes in poverty levels to two different causes: (i) the growth effect, which arises when there is a proportional change in income levels not necessarily accompanied by any change in relative incomes, and (ii) the distribution effect, meaning a change in the distribution of relative income.

This approach can be used to demonstrate that changes in poverty levels may result from either economic growth (typified by increases in average income) or a lessening of income inequality.

The curves in figure 1 show income distribution densities, represented on the horizontal axis by a logarithmic scale. The displacement from the initial distribution to the final one occurs through an intermediate step, which is the horizontal transposition of the initial distribution to curve (I). This change represents a proportional increase in all incomes across the population due to the growth effect.

Figure 1

Decomposition of poverty changes due to economic growth and income distribution



Source: F. Bourguignon, "The growth elasticity of poverty reduction: explaining heterogeneity across countries and time periods", *Inequality and Growth: Theory and Policy Implications*, T. Eicher and S. Turnovsky (eds.), Cambridge, Massachusetts, The MIT Press, 2003.

Thus, the change involves displacement of the income distribution density to the right. If z is taken as the poverty line, it can be seen that there is a reduction in the number of poor individuals. This decline in the percentage of poor people is exclusively a consequence of the growth effect.

The shift of curve (I) towards the final distribution comes about with constant average income and a change in relative income distribution, this being the distribution effect. Thus, the poverty level falls without any alteration in the average income of the population because of a decrease in income inequality. In other words, a decrease in income concentration leads to a reduction in the number of people with incomes below the poverty line.

Bourguignon (2004) called this relationship the "poverty-inequality-growth triangle". Setting out from this, many studies have been undertaken to identify and measure the relationship between the effects of growth and changes in inequality on poverty reduction.

1. Poverty versus economic growth

Several empirical studies in Brazil and internationally have analysed the interactions between economic growth and poverty. There seems to be a broad consensus among researchers that two factors are fundamental in the study of poverty reduction: the average growth rate and initial income inequality.

For example, Kraay (2004) carried out a poverty variance decomposition to ascertain the importance of economic growth in poverty reduction, using a sample of developing countries during the 1980s and 1990s. His analysis concluded that changes in poverty indices were due to average income growth, the implication being that policies to promote economic growth were essential for the well-being of the poor.

The relationship between growth and poverty reduction can be measured by the income elasticity and growth elasticity equations. If elasticity is high, public anti-poverty policies based on economic growth are more efficient. If elasticity is low, however, poverty reduction strategies should include a combination of economic growth and some type of income redistribution. Ravallion and Chen (1997) estimated income-poverty and inequality-poverty elasticities for 45 countries. The results for low-inequality countries showed that if income increased by 1%, poverty was reduced by 4.3%. In countries where inequality was high, however, the decline in poverty with the same growth would be only 0.6%. The conclusion was that growth in itself had little effect on poverty, but that if inequality tended to decline because of the growth, the effect on poverty was much stronger.

In another study, Ravallion (2001) found that growth-poverty elasticity was much higher in countries that combined growth with some kind of inequality reduction policy. This suggests that the growth-poverty elasticity methodology needs to be controlled for the income redistribution component.

As an example, Ravallion (2005) estimated pro-poor growth in China and India in the 1990s and found that changes in income distribution disadvantaged the poor, as the income growth rate for this population sector was lower than overall ordinary income growth during the period analysed. This result yielded a growth incidence curve with a positive slope for the highest income levels. However, pro-poor income growth was still positive, suggesting a decline in absolute poverty.

Chen and Wang (2001) studied the relationship between poverty, income and inequality in China during the 1990s. They concluded that while economic growth tended to reduce poverty, income concentration tended to increase it. They also found that the growth in average income was more beneficial for the rich, as only the average income of the richest 20% grew by more than average income overall, showing that income concentration reduces the positive effect of economic growth on poverty levels.

Stewart (2000) estimated that a gross domestic product (GDP) growth rate of 1.0% prompted a 0.21% decline in poverty in Zambia, while the same change led to a 3.4% decline in poverty in Malaysia, the discrepancy being due to the differences in income inequality between the two countries.

Bearing out these findings, Deininger and Squire (1996) analysed the potential effect of economic growth in reducing inequality for a sample of several countries, finding that different inequality levels had different consequences for economic growth and that these levels were negatively related to poverty.

Using a sample of 84 countries between 1996 and 2000, Son (2004) showed that economic growth contributed to poverty reduction in 95% of cases. In the others, either the growth rate was negative or it was not possible to draw conclusions because of sampling ambiguities.

The Brazilian literature includes some studies on these subjects. For example, Hoffmann (1995) found a reduction in poverty levels during the 1970s, with high income growth and relatively unchanged inequality. In the 1980s, poverty and inequality increased due to extremely high inflation.

In another study, Hoffmann (2005) found that a 1% increase in per capita household income in Brazil led to a reduction of 0.84% in the proportion of poor people and that the absolute value of this elasticity increased with income and decreased with a reduction in inequality.

Marinho and Soares (2003), using data from 26 Brazilian states between 1985 and 1999, applied a methodological procedure that allowed poverty changes resulting from alterations in average income and income concentration to be decomposed. The results showed that income concentration had a greater influence than income levels in all the northern states. Likewise, income growth was of strategic importance in fighting poverty. Generally speaking, their study showed that higher average rates resulted in a higher absolute elasticity while greater income concentration meant a lower absolute elasticity.

Manso, Barreto and Tebaldi (2005) searched National Household Survey (PNAD) data from 1995 to 2004 for evidence of the interactions between income growth, poverty reduction and the wealth distribution profile. In their study, these authors widened the discussion on the problem of regional imbalance in Brazil by evaluating the impact of economic growth on poverty rates. Their analysis allowed economic growth and income inequality effects in each region of the country to be isolated. The results suggest that average income growth and income distribution components are sufficient to explain most of the variations in poverty in Brazilian states. Findings so far reinforce the evidence that policies aimed at fighting poverty through economic growth are more efficient when accompanied by positive income redistribution.

2. Poverty versus inequality

This subsection investigates the relationship between poverty and inequality in the literature. In general, many authors affirm that the percentage of poor people in a given region decreases when there is a policy of economic growth combined with income redistribution.

Income inequality is an important component in the debate on poverty. Poverty is a worldwide problem that affects modern society and has been discussed in numerous studies, persisting alongside the growing stock of material wealth around the world. Its extent and severity are demonstrated by the number of poor people across the planet, and inequality and poverty go hand in hand.

Ravallion (2005) observed a non-linear relationship between growth-poverty elasticity and the initial inequality level in a set of underdeveloped countries and argued that economic growth had little effect on the poor if inequality was not thereby reduced. It is estimated that growth of 1.0% can reduce poverty by up to 4.3% in countries with low inequality. In countries that suffer from high income inequality, however, the same 1.0% growth only yields a 0.6% decrease in the percentage of poor people.

Therefore, one of the factors affecting poverty reduction rates for a given growth rate is the change in income distribution. This was also found by Datt and Ravallion (2002) when they measured poverty changes resulting from the growth-income distribution effect. Models designed to estimate the elasticity of poverty reduction must incorporate income inequality as an explanatory variable to prevent changes in income distribution being subsumed into growth elasticity.

According to Bourguignon (2004), the reduction of income inequality is an important tool in reducing poverty, and economic growth may not be such a necessary element. Similar results were found in Brazil by Barros, Henriques and Mendonça (2001). These authors emphasized that inequality in income distribution was responsible for economic growth being less efficient than it might be for poverty reduction purposes. In other words, the effect of growth on poverty reduction was smaller in Brazil than in other countries with the same income level.

Only a few studies have sought to explain the connection between poverty, growth and inequality in Brazil. According to Rocha (2006), although poverty in Brazil has persisted for decades, it was only after the inflation problem was solved that social problems started to be treated as a priority, the reduction of inequality being an example. This may account for the small output of articles studying the poverty-growth-inequality triangle in Brazil.

Barreto (2005) considers that poverty reduction can be quickly achieved when a growing country achieves a less unequal income distribution. Therefore, the implementation of public policies aimed at the reduction of inequality, besides solving this problem, may also indirectly help towards other economic policy goals such as increased growth rates and poverty reduction. In general, the literature suggests that a significant reduction in poverty rates is a consequence of economic growth combined with policies to reduce income inequality. The joint result of these two effects is that poverty reduction feeds through directly into improvements in the average income of the poorest.

Rocha (2006), for instance, indicates that the proportion of poor people in Brazil was reduced by nearly two percentage points between 2001 and 2004. According to this author, the reduction that occurred in the early years of the decade was a consequence of several factors that had differing regional impacts, including distributive changes in labour yield and the expansion of welfare benefits.

Rocha (2006) argues that poverty in Brazil is highly persistent mainly because of inequality and that it can be reduced through both income growth and better distribution, the consensus being, however, that the emphasis should be on reducing income inequality, since relying on income growth without progress on inequality may result in the goal of poverty elimination being deferred to a distant future.

Poverty reduction rates in Brazil from 1995 to 2009 are shown in table 1. The proportion of poor people (P_0) fell from 38.70% in 1995 to 23.50% in 2009, a figure that, while still high, represents a reduction of 15.20 percentage points.

Table 1
Poverty rates in Brazil, 1995-2009
(Percentages)

Year	P_0	Year	P_0
1995	38.7	2003	39.1
1996	38.1	2004	37.0
1997	38.5	2005	34.1
1998	37.2	2006	29.6
1999	39.0	2007	28.0
2001	38.3	2008	25.4
2002	38.2	2009	23.5
Difference -15.20			

Source: Prepared by the authors, on the basis of data from the National Household Survey (PNAD).

Thus, analysis of this indicator shows that there was significant poverty reduction in Brazil during the 1995-2009 period.

3. Growth versus inequality

The literature has analysed the economic growth versus inequality relationship, taking into consideration the causalities behind these two variables. Many of the issues relating to them have to do with the way inequality is generated and reproduced over time and the way inequality and economic development processes are connected.

The Kuznets (1955) inverted-U hypothesis is the starting point for this theory. According to this, inequality should first increase with the beginnings of economic development as the economy moves from the rural sectors towards industrialization (transfer of labour from the less productive to the more productive sector). Later on, once the bulk of the labour force is working in the industrial sector, inequality should diminish.

Thus, a development policy could be summarized as the promotion of economic growth in a way that also serves to reduce inequality. With higher and better-distributed incomes, the problem of poverty should be solved.

According to Barreto (2005), a number of studies have analysed the impact of inequality on economic growth. Different models show that inequality may either retard or stimulate growth.

For example, Alesina and Rodrik (1994) established that the causality between growth and inequality was based on three things: (i) government spending and redistributive tax policies should be negatively related to growth because of their perverse effects on capital accumulation; (ii) tax aliquots should tend to be proportional to income, and the benefits of public expenditure should be available to all individuals alike, with the implication that spending levels should be inversely related to income; and (iii) the tax burden adopted by the government should be the one desired by the median voter, implying less capital accumulation and therefore less growth.

Stewart (2000) takes the approach that inequality should be negative for growth, arguing that: (i) high inequality means political instability, uncertainty, less investment and less growth; (ii) high inequality results in a populist redistributive taxation policy, disincentive effects and lower growth rates; and (iii) high inequality affects the behaviour of the richest, who press for preferential tax treatment, leading to overinvestment in certain areas and therefore to lower growth.

There are others, though, who consider that inequality may stimulate economic growth. Bourguignon (1981) argues that the tendency to save is different for the rich and the poor, being higher among the former than the latter, which would imply a tendency for investment to be greater in more unequal economies, with the corollary of potentially faster growth. Conversely, authors such as Barro (2000) and López (2004) do not find any relationship whatsoever between inequality and economic growth and maintain that investment levels do not significantly depend on inequality. Some income inequality data for Brazil between 1995 and 2009 are shown in table 2. These data are the Gini index, the ratio between the income of the richest 10% and the poorest 40%, the ratio between the income of the richest 20% and the poorest 20%, the ratio between the income of the richest 10% and the poorest 10%, the average per capita household income and the percentage of people below the poverty line.

Table 2 shows that the income of the richest 10% in Brazil was 23.7 times as great as that of the poorest 40% in 1995, falling to 16.3 times by 2009. The ratios of the incomes of the richest 10% relative to the poorest 10% and of the richest 20% relative to the poorest 20% also declined significantly during the period, especially the former, which moved down from 67.0 to 43.8. Meanwhile, per capita family income grew by approximately 22.5%. The last column shows that the proportion of people below the poverty line stabilized at around 19%, then fell to 12% in 2009.

Table 2

Main per capita family income distribution statistics in Brazil, 1995-2009

Year	Gini	Income of richest 10% as multiple of poorest 40%	Income of richest 20% as multiple of poorest 20%	Income of richest 10% as multiple of poorest 10%	Average family income per capita (*reais per month*)	People below the poverty line (*percentages*)
1995	0.601	23.7	27.4	67.0	520.6	19.7
1996	0.602	24.2	29.3	74.9	529.7	19.5
1997	0.602	24.2	28.7	72.3	529.0	19.8
1998	0.601	23.6	27.5	67.2	534.5	19.1
1999	0.595	22.7	26.2	63.2	504.4	19.9
2001	0.597	22.9	26.9	68.4	511.9	19.7
2002	0.590	21.9	24.7	59.2	511.9	19.5
2003	0.585	21.1	24.3	59.4	481.9	20.1
2004	0.575	19.5	22.0	51.7	497.9	18.9
2005	0.572	19.2	21.3	49.7	528.4	17.5
2006	0.560	18.3	20.4	47.5	577.5	15.2
2007	0.550	17.7	20.2	49.0	592.5	14.4
2008	0.540	16.8	18.9	44.0	622.6	12.9
2009	0.540	16.3	18.6	43.8	637.4	12.2

Source: Institute for Studies on Labour and Society (IETS).
Note: The 2009 poverty line of 196.00 reais is taken, deflated by the national consumer price index (INPC).

These data show that income inequality in Brazil has declined in recent years, corroborating results obtained by Neri (2006), Barros and others (2007) and Hoffmann (2007). Similarly, Manso, Barreto and Tebaldi (2005) found a significant reduction of income inequality in Brazil after the Real Plan was implemented. From 1995 to 2004, there was a 2.71% reduction in the Gini index.

III. The elasticity of poverty to inequality and income

The purpose of establishing income-poverty and inequality-poverty elasticities is to analyse the poverty impact of growth and changes in income inequality. This methodology was originally proposed by Bourguignon (2003).

That author follows the classic definition proposed by Foster, Greer and Thorbecke (1984), whereby poverty is measured by the percentage of poor people. On that basis, the proportion of persons with a per capita income below the poverty line is given by:

$$H_t = Pr\left(y_t < z\right) \equiv F_t\left(z\right) \tag{1}$$

where the function $F_t(z)$ is given by the income distribution function.

Therefore, the proportion of the population with an income below the absolute poverty line z at time t is equal to the probability that income y_t is below the poverty line. The change in the percentage of poor people between two time periods t and t' is accordingly:

$$\Delta H = H_{t'} - H_t = F_{t'}\left(z\right) - F_t\left(z\right) \tag{2}$$

Assuming that the income distribution curve is log-normal, Bourguignon (2003) defines the original curve displacement shown in figure 1, with the final distribution curve in respect of poverty variation being as follows:

$$\Delta H = H_{t'} - H_t \approx \left[F_t\left(\tfrac{z}{\bar{y}_{t'}}\right) - F_t\left(\tfrac{z}{\bar{y}_t}\right)\right] + \left[F_{t'}\left(\tfrac{z}{\bar{y}_{t'}}\right) - F_t\left(\tfrac{z}{\bar{y}_{t'}}\right)\right] \tag{3}$$

The first expression in brackets corresponds to the growth effect, while the relative income distribution F_t is kept constant. The second expression is for the inequality effect, and there is a shift in the distribution of relative income, which remains constant.

Thus considered, changes in poverty are influenced by two effects: the first is due to income growth and the second is a consequence of the inequality of income distribution.

According to Epaulard (2003), the relative change in poverty resulting from income growth and the redistribution effect may be decomposed as follows:

$$\frac{dH}{dt} = \frac{\partial H_t}{\partial \bar{y}_t}\frac{d\bar{y}_t}{dt} + \frac{\partial H_t}{\partial G_t}\frac{dG_t}{dt} \tag{4}$$

In terms of elasticity we have:

$$\frac{dH}{dt} = \varepsilon_y^H \frac{d\bar{y}_t}{dt}\frac{H_t}{\bar{y}_t} + \varepsilon_G^H \frac{H_t}{G_t}\frac{dG_t}{dt} \tag{5}$$

where the Gini coefficient is defined as $G = 2\Phi\left(\frac{\sigma_t}{\sqrt{2}}\right) - 1$. The term $\Phi(\cdot)$ is the cumulative density function for the standard normal distribution (the normal distribution with a mean of zero and a standard deviation of 1) and σ_t is the standard deviation of the income logarithm. Thus, Epaulard (2003) proves that the income-poverty elasticity ε_y^H and the inequality-poverty elasticity ε_G^H are defined by the following expressions:

$$\varepsilon_{\bar{y}}^{H} = \frac{\partial H_t}{\partial y_t}\frac{\bar{y}_t}{H_t} \equiv -\frac{1}{\sigma_t}\frac{\text{ø}\left(\frac{log\left(z/\bar{y}_t\right)}{\sigma_t} + \frac{1}{2}\sigma_t\right)}{\Phi\left(\frac{log\left(z/\bar{y}_t\right)}{\sigma_t} + \frac{1}{2}\sigma_t\right)} \leq 0 \tag{6}$$

$$\varepsilon_{G}^{H} = \frac{\partial H_t}{\partial \sigma_t}\frac{\sigma_t}{H_t} \equiv \frac{1}{\sigma_t}\frac{\text{ø}\left(\frac{log\left(z/\bar{y}_t\right)}{\sigma_t} + \frac{1}{2}\sigma_t\right)}{\Phi\left(\frac{log\left(z/\bar{y}_t\right)}{\sigma_t} + \frac{1}{2}\sigma_t\right)}\left(\frac{log\left(z/\bar{y}_t\right)}{\sigma_t} + \frac{1}{2}\sigma_t\right) \gtreqless 0 \tag{7}$$

The author referred to also shows that the income-poverty elasticity ($\varepsilon_{\bar{y}}^{H}$) and the inequality-poverty elasticity (ε_G^H) decrease in absolute terms at the poverty line and average income ratios $\left(z/\bar{y}_t\right)$ and with the standard deviation of the income logarithm (σ_t). Income-poverty elasticity is always positive or null. On the other hand, inequality-poverty elasticity may be higher or lower than zero.[1]

Consequently, the effect of changes in income distribution on poverty reduction is the function of income growth and the level of inequality. This means that poverty changes may result both from economic growth (typified by the increase of average income) and from a decline in income inequality. However, the poverty reduction effect is much stronger when both factors combine.

IV. The database

The data used in the estimation of the econometric models described in the next section were obtained from the National Household Survey (PNAD) published by the Brazilian Geographical and Statistical Institute (IBGE). The sample is composed of all Brazilian states for the years from 1995 to 2009.[2]

The family income per capita variable is calculated by dividing total family income by the number of family members. The arithmetic average of this variable is then established and average income values thus obtained for the states in the sample. We expect to find a negative relationship between poverty and this variable. It is worth noting that the Brazilian economy showed growth in income per capita over the period 1995-2009.

In this article, families living on a per capita family income insufficient to meet their basic needs are classified as poor. Thus, the absolute poverty indicator used is the proportion of poor people (P_0).

1 According to Epaulard (2003), inequality-poverty elasticity should be positive unless a country has a very low average income. This elasticity will be positive as long as $\bar{y}_t < z\,exp\left(-\frac{1}{2}\sigma_t^2\right)$.

2 The PNAD was not carried out in 2000. To fill this gap, we have taken arithmetic averages of variables from 1999 to 2001. The old states of the northern region were not included in the sample owing to the non-availability of data from rural areas before 2004.

The poverty line adopted to construct this indicator was half the monthly minimum wage.[3] The poverty indicator P_0 is defined as $P_0 = \frac{q}{n}$, where n is the total number of individuals and q is the number of people with a per capita family income below the poverty line.

The inequality measurement method used is the Gini coefficient, calculated on the basis of per capita family income as extracted from the PNAD. This index is frequently used to express income inequality and may be linked to the so-called Lorenz curve, which is defined by the set of points obtained by plotting income shares against population shares in ascending order. On the basis of this curve, we then calculate Gini coefficients for each of the states between 1995 and 2009. As discussed in the previous section, the relationship between the Gini coefficient and poverty must be positive. In other words, the higher the inequality, the more poverty there is.

It is important to emphasize that all monetary variables have been adjusted to real 2009 values using the national consumer price index (INPC) prepared with 2009 data.

V. The econometric model

The econometric specification of the model is based on the contribution of economic growth and shifts in income distribution to changes in poverty. We admit as a hypothesis that the current poverty trend tends to perpetuate itself, affect future poverty performance or both.[4] To explore this, the relationship between changes in poverty and their determinants is investigated using a dynamic panel data regression model defined as follows:[5]

$$\Delta\, ln\left[P_{0,it}\right] = \beta_0 + \beta_1 \Delta\, ln\left[P_{0,it-1}\right] + \beta_2 \Delta\, ln\left[\bar{Y}_{it}\right] + \beta_3 \Delta\, ln\left[Gini_{it}\right] + \eta_t + \mu_{it} \tag{8}$$

The variables of this model are defined as: $\Delta\, ln\, P_{0,it} = ln\, P_{0;it} - ln\, P_{0;it-1}$, representing the change in the proportion of people who are poor between two periods of time; $\Delta\, ln\, \bar{Y}_{it} = ln\, \bar{Y}_{it} - ln\, \bar{Y}_{it-1}$, the change in the average family income per capita; and $\Delta\, ln\, Gini_{it} = ln\, Gini_{it} - ln\, Gini_{it-1}$, the change in income concentration as measured by the Gini coefficient i; while η_i represents the non-observable random effects of individuals and μ_{it} represents random disturbances. The model variables are defined using a natural logarithm in which subscript i represents the state and t the time period. Thus, parameters β_2 and β_3 are income-poverty elasticity $\varepsilon_y^{P_0}$ and inequality-poverty elasticity $\varepsilon_G^{P_0}$. Note that these elasticities do not change over time.

An expansion of this model introduced by Kalwij and Verschoor (2004) allows the income and inequality elasticities to change over time depending on the inverse initial development level (poverty line divided by initial family income per capita and the initial inequality level).[6] By entering these variables in the model, we intend to evaluate the hypothesis that growth is less effective at reducing poverty when initial inequality is higher (Bourguignon hypothesis). This dynamic model is described as follows:

[3] This line was also used by Rocha (2006), Barreto (2005) and Marinho and Soares (2003). However, the shift line may alter the outcome.

[4] Ribas, Machado and Golgher (2006) found evidence of poverty persistence in Brazil.

[5] This model may be seen in Bourguignon (2003) and in Kalwij and Verschoor (2004). However, those authors do not think that poverty can develop dynamic behaviour.

[6] Kalwij and Verschoor (2004) likewise do not give consideration to poverty behaving dynamically over time.

$$\Delta\, ln\left[P_{0,\,it}\right] = \beta_0 + \beta_1 \Delta\, ln\left[P_{0,\,it-1}\right] + \beta_2 \Delta\, ln\left[\bar{y}_{it}\right] + \beta_3 \Delta\, ln\left[\bar{y}\right] ln\left[G_{i0}\right] + \beta_4 \Delta\, ln\left[\bar{y}_{it}\right] ln\left[\frac{z_{it}}{y_{i0}}\right]$$

$$\beta_5 \Delta\, ln\left[Gini_{it}\right] + \beta_6 \Delta\, ln\left[Gini_{it}\right] ln\left[Gini_{i0}\right] + \beta_7 \Delta\, ln\left[Gini_{it}\right] ln\left[\frac{z_{it}}{\bar{y}_{i0}}\right] + \beta_8\, ln\left[G_{i0}\right] \tag{9}$$

$$\beta_9\, ln\left[\frac{z_{it}}{\bar{y}_{i0}}\right] + \eta_i + \mu_{it}$$

where in addition to the variables $\Delta\, ln\left[P_{it}\right]$, $\Delta\, ln\left[\bar{y}_{it}\right]$ and $\Delta\, ln\left[Gini_{it}\right]$, which follow the formulations previously described, we have $\Delta\, ln\left[\bar{y}_{it}\right] ln\left[G_{i0}\right]$ and $\Delta\, ln\left[\bar{y}_{it}\right] ln\left[\frac{z_{it}}{y_{i0}}\right]$, representing the interactions between average family income per capita and the initial Gini index at state i (G_{i0}) and the inverse initial development level $\frac{z_{it}}{y_{i0}}$ (poverty line divided by initial family income per capita). Likewise, the variables $\Delta\, ln\left[Gini_{it}\right] ln\left[G_{i0}\right]$ and $\Delta\, ln\left[Gini_{it}\right] ln\left[\frac{z_{it}}{\bar{y}_{i0}}\right]$ represent the interactions between the Gini inequality index and the initial inequality index associated with state i and the inverse initial development level.

The hypotheses adopted in these models are $E[\eta_i] = E[\mu_{it}] = E[\eta_i \mu_{it}] = 0$ and $E[\mu_{it}\mu_{is}] = 0$ for $i=1.2,.....,N$ e $\forall t \neq s$. Additionally, there is a standard hypothesis relating to the initial conditions: ΔlnP_{it}: $E[\,\Delta lnP_{it\text{-}1}\,\mu_{it}] = 0$ for $i=1.2,....,N$ and $t=1.2,.....,T$ (Ahn and Schmidt, 1995).

The second model specification therefore takes into account that the poverty elasticities of average family income per capita and inequality depend on the initial inequality and the ratio between the poverty line and the initial average family income per capita.

Naturally, coefficients β_2 and β_5 are not interpreted as income elasticity and inequality elasticity. To calculate these elasticities, it is necessary to consider the interaction terms. Thus, the income-poverty and inequality-poverty elasticities are now respectively defined as:

$$\varepsilon^{P_0}_{\bar{y}_{it}} = \beta_2 + \beta_3\, ln\left[G_{i0}\right] + \beta_4\, ln\left[\frac{z_{it}}{\bar{y}_{i0}}\right] \tag{10}$$

$$\varepsilon^{P_0}_{G_{it}} = \beta_5 + \beta_6\, ln\left[G_{i0}\right] + \beta_7\, ln\left[\frac{z_{it}}{\bar{y}_{i0}}\right] \tag{11}$$

We can now observe that the income-poverty and inequality-poverty elasticities do change through time.

The traditional estimation techniques are inappropriate for the two models shown because of two main econometric problems. The first is the presence of unobservable effects of individuals η_i, together with the lagged dependent variable $\Delta lnP_{k,it\text{-}1}$, on the right-hand side of the equation. In this case, omitting the individual fixed effects in the dynamic panel model makes the ordinary least square (OLS) estimators biased and inconsistent.

For example, the provable positive correlation between the lagged dependent variable and the fixed effects means that the coefficient β_1 estimator is upward-biased. On the other hand, the within-groups estimator (which corrects for the presence of fixed effects) generates a downward-biased estimation of β_1 in panels with a small temporal dimension (Judson and Owen, 1999).

Seeking to correct these problems, Arellano and Bond (1991) proposed a GMM-differentiated estimator. This method consists in eliminating fixed effects through the first-difference equation. Therefore, for the two models we have:

$$\Delta\left[\Delta\, ln\left[P_{0,it}\right]\right] = \beta_1 \Delta\left[\Delta\, ln\left[P_{0,it-1}\right]\right] + \beta_2 \Delta\left[\Delta\, ln\left[\bar{Y}_{it}\right]\right] + \beta_3 \Delta\Delta\left[ln\left[Gini_{it}\right]\right] + \Delta\mu_{it} \tag{12}$$

$$\begin{aligned}\Delta\left[\Delta\, ln\left[P_{0,it}\right]\right] &= \beta_0 + \beta_1 \Delta\left[\Delta\, ln\left[P_{0,it-1}\right]\right] + \beta_2 \Delta\left[\Delta\, ln\left[\bar{Y}_{it}\right]\right] + \beta_3 \Delta\left[\Delta\, ln\left[\bar{Y}\right]\left[G_{i0}\right]\right] \\ &+ \beta_4 \Delta\left[\Delta\, ln\left[\bar{y}_{it}\right] ln\left[\tfrac{z_{it}}{\bar{y}_{i0}}\right]\right] + \beta_5 \Delta\left[\Delta\, ln\left[Gini_{it}\right]\right] + \beta_6 \Delta\left[\Delta\, ln\left[Gini_{it}\right] ln\left[G_{i0}\right]\right] \\ &+ \beta_7 \Delta\left[\Delta\, ln\left[Gini_{it}\right] ln\left[\tfrac{z_{it}}{\bar{y}_{i0}}\right]\right] + \beta_8 \Delta\left[\Delta\, ln\left[G_{i0}\right]\right] + \beta_9 \Delta\left[ln\left[\tfrac{z_{it}}{\bar{y}_{i0}}\right]\right] + \Delta\mu_{it}\end{aligned} \tag{13}$$

where for a variable w_{it}, any $\Delta\, ln\left[w_{it}\right] = ln\left[w_{it}\right] - ln\left[w_{it-1}\right]$. Through the construction of equations (12) and (13), $\Delta\left[\Delta\, ln\left[P_{0,it-1}\right]\right]$ and $\Delta\mu_{it}$ are correlated, and consequently OLS estimators for their coefficients will be biased and inconsistent. Accordingly, it is necessary to implement instrumental variables for $\Delta\left[\Delta\, ln\left[P_{0,it-1}\right]\right]$ in this case. The set of hypotheses adopted for equations (8) and (9) implies that the moment condition $E\left[\Delta\left[\Delta\, ln\, P_{0,it-S}\right]\Delta\mu_{it}\right] = 0$ for t=3.4,....T and $s \geq 2$ are valid. On the basis of these moments, Arellano and Bond (1991) suggest using $\Delta ln\left[P_{0,it-s}\right]$ for t=3.4,....T and $s \geq 2$ as instruments for equations (12) and (13).

With regard to the other explanatory variables, there are three possible situations. An explanatory variable x_{it} may be classified as (i) strictly exogenous if it is not correlated to the past, present and future error terms, (ii) frankly exogenous if it is only correlated to past error terms and (iii) endogenous if it is correlated to the past, present and future error terms. In the second case, x_{it} lagged values for one or more periods are valid instruments for the estimation of equations (12) and (13). In the last case, x_{it} lagged values for two or more periods are valid instruments for the estimations of these same equations.

On the other hand, Arellano and Bover (1995) and Blundell and Bond (1998) affirm that these instruments are weak when the dependent and explanatory variables show a strong persistence or the relative variance of fixed effects increases, or both. This produces a differentiated GMM estimator that is inconsistent and biased for small T panels.

Therefore, the aforementioned authors suggest an estimation using a system that combines the set of level equations (8) and (9) and difference equations (12) and (13) as a way of reducing the bias and imprecision problems. This is where the generalized method of moments comes from (GMM system). For difference equations, the set of instruments is the same as described above. For a level regression, the appropriate instruments are the lagged differences of the respective variables. For instance, if it is assumed that the differences in the explanatory variables are not correlated to the individual fixed effects for t=3.4,....T and $E[\Delta[\Delta lnP_{0,i2}]\ \eta_i]=0$ for i = 1.2.3,...,N, then the different explanatory variables and $\Delta[\Delta lnP_{k,it-1}]\ \eta_i]$, if they are either exogenous or frankly exogenous, are still valid instruments for level equations. The same happens if they are endogenous, but with instruments that are the explanatory variables in one-period lagged difference and $\Delta[\Delta lnP_{k,it-1}]$.

Finally, and as a means of testing model robustness and consistency, Arellano and Bond (1991) suggest two different test types: the Hansen and Sargan tests, which check whether the instruments used and the additional instruments required by the GMM system are valid. Lastly, the Arellano and Bond statistical tests verify whether the error μ_{it} has a first-order serial correlation and whether $\Delta\mu_{it}$ shows second-order correlation. For the purposes of estimator consistency, μ_{it} is expected to result in a first-order correlation while the $\Delta\mu_{it}$ series should not be second-order-correlated.

It is worth stressing that the GMM system introduced in the next section derives from the estimation performed with the estimator as corrected by the Windmeijer (2005) method, with a view to preventing the variance estimator from underestimating the true variances in a finite sample. The estimator applied was suggested by Arellano and Bond (1991) in two steps. In the first step, the error terms are assumed to be independent and homoscedastic through time in their respective states. In the second stage, residuals obtained in the first stage are used to build a consistent estimation of a variance-covariance matrix, thus relaxing the independence and homoscedasticity hypothesis. The second-stage estimator is asymptotically more efficient than the first-stage estimator.

VI. Econometric model results

This section introduces the results of estimations for the parameters of the two models, which will be used to calculate the income-poverty and inequality-poverty elasticities.

The results estimated for the first model using the OLS, within-groups and GMM system methods can be seen in table 3.

Table 3
Results of regression models for $\Delta ln[P_{0,it}]$: model 1

	Ordinary least squares [a]		Within-group [b]		Generalized method of moments system [c]	
	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
$\Delta ln[P_{0,it-1}]$	0.1840 (0.0672)	0.00	0.1529 (0.0686)	0.02	0.1139 (0.0239)	0.00
$\Delta ln[\bar{y}_{it}]$	-0.7654 (0.0651)	0.00	-0.7886 (0.0658)	0.00	-0.6899 (0.0507)	0.00
$\Delta ln[Gini_{it}]$	0.8785 (0.1451)	0.00	0.9046 (0.1464)	0.00	0.7799 (0.1385)	0.02
Constant	-0.0079 (0.0049)	0.11	-0.0080 (0.0050)	0.10	-0.0114 (0.0007)	0.00
	F(3.269) = 53.11 Prob > F = 0.0000 R^2 = 0.37		F(3.249) = 53.21 Prob > F = 0.0000		F(2. 20) = 124.30 Prob > F = 0.0000	
H_0: Absence of autocorrelation in first-order residuals			P-value		0.001	
H_0: Absence of autocorrelation in second-order residuals			P-value		0.101	
Hansen test			Prob > chi^2		0.288	
Sargan test			Prob > chi^2		0.262	

Source: Prepared by the authors.
Note: (i) the values in parentheses are the standard deviations corrected by the Windmeijer method (2005); (ii) the Hansen test values are the p-values for the null hypothesis that the instruments are valid; (iii) the Sargan test values are the p-values for the validity of the additional instruments required by the GMM system, taking the explanatory variables in lagged differences as instruments for the GMM system and lagging $\Delta ln[P_{0,it-1}]$ and $\Delta ln[\bar{y}_{it}]$ by one period; (iv) there were 273 observations, 21 groups and 17 instruments.

In this table, the value of the coefficient estimated for the variable $\Delta ln[P_{it-1}]$ in column [c] by applying the GMM system method is within the values of the coefficients estimated for this same variable (columns [a] and [b] using the OLS and within-group methods). The GMM system thus reduces the problem of estimation bias, as there is a one-period lagged dependent variable on the right-hand side of equation (8). Note that in column [c] the statistical significance of the estimated coefficient of $\Delta ln[P_{0,it-1}]$ confirms the initial hypothesis that poverty variation is persistent.

The results estimated for the income elasticity and inequality elasticity parameters were -0.69 and 0.78, respectively, as shown in column [c]. Thus, a 1.0% increase in per capita income results in a decrease of 0.68% in the percentage of poor individuals. An increase of 1.0% in the inequality index leads to growth of 0.78% in poverty levels. It is worth remarking that the estimated elasticity values agree with the theoretical elasticity introduced in section III. They also corroborate results obtained in international papers such as those of Kalwij and Verschoor (2004), Bourguignon (2004) and Marinho and Soares (2003) and Hoffmann (2004) in Brazil. This indicates that policies aimed at reducing inequality are more effective at fighting poverty than those solely aimed at improving average income.

Estimated results for the parameters of equation (9) can be seen in table 4 below. Again, the value of the estimated parameter for the variable $\Delta ln[P_{it\text{-}1}]$ is within the range of the values estimated for this same variable (columns [a] and [b]) using the OLS and within-group methods. When estimated by the GMM system, this parameter is not statistically significant.

Table 4

Results of regression models for $\Delta ln[P_{0,it}]$: model 2

	Ordinary least squares [a]		Within-group [b]		Generalized method of moments system [c]	
	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
$\Delta\, ln\, P_{0,\,it\,-\,1}$	0.1463 (0.0676)	0.03	0.0425 (0.0720)	0.55	0.1301 (0.0711)	0.08
$\Delta\, ln[\bar{Y}_{it}]$	-0.3675 (0.1485)	0.01	-0.4137 (0.1516)	0.00	-1.0806 (0.2936)	0.00
$\Delta\, ln[\bar{Y}_{it}]ln[G_{i0}]$	0.4371 (0.2629)	0.09	0.5238 (0.2687)	0.05	1.6851 (0.5050)	0.00
$\Delta\, ln[\bar{Y}_{it}]ln\left[\frac{z_{it}}{\bar{y}_{i0}}\right]$	1.064 (0.4832)	0.02	1.0820 (0.4801)	0.05	1.1565 (0.3860)	0.00
$\Delta\, ln[Gini_{it}]$	0.4209 (0.3479)	0.22	0.4865 (0.3507)	0.16	3.4064 (0.8328)	0.00
$\Delta\, ln[Gini_{it}]ln[G_{i0}]$	-0.3783 (2.6100)	0.53	-0.5010 (0.6166)	0.41	-5.6068 (1.4515)	0.00
$\Delta\, ln[Gini_{it}]ln\left[\frac{z_{it}}{\bar{y}_{i0}}\right]$	-2.998 (0.8709)	0.00	-3.0703 (0.8771)	0.00	-1.2865 (0.6120)	0.05
$ln[G_{i0}]$	0.1283 (0.1121)	0.25		-	1.1980 (0.5580)	0.04
$ln\left[\frac{z_{it}}{\bar{y}_{i0}}\right]$	-0.0931 (0.0377)	0.01	-0.2159 (0.0479)	0.00	0.2876 (0.6176)	0.00
Constant	0.0851 (0.0650)	0.19	0.0491 (0.1344)	0.00	0.7001 (0.3002)	0.03
	F(9.63) = 21.93 Prob > F = 0.0000 R^2 = 0.43		F(8.244) = 26.63 Prob > F = 0.0000		F(8.20) = 16.24 Prob > F = 0.0000	
H_0: Absence of autocorrelation in first-order residuals			P-value		0.002	
H_0: Absence of autocorrelation in second-order residuals			P-value		0.829	
Hansen test			Prob > chi^2		0.360	
Sargan test			Prob > chi^2		0.269	

Source: Prepared by the authors.

Note: (i) the values in parentheses are the standard deviations corrected by the Windmeijer method (2005); (ii) the Hansen test values are the p-values for the null hypothesis that the instruments are valid; (iii) the Sargan test values are the p-values for the validity of the additional instruments required by the GMM system, the instruments used in this system being the explanatory variables in lagged differences and $\Delta[\Delta ln[P_{0,it\text{-}1}]]$ and $\Delta[\Delta ln[\bar{y}_{it}]]$ lagged one period; (iv) there were 273 observations, 21 groups and 17 instruments.

Among the isolated factors that significantly contribute to poverty growth, mention may be made of the following, in order of increasing importance: the interaction between income changes and the inverse initial development level, the interaction between income changes and initial income inequality, and income inequality in the present. Column [c] in table 4 shows positive and significant values for these variables.

The interaction term between income changes and the inverse initial development level yields a positive and statistically significant estimated coefficient, and the same is true of the coefficient of interaction between income changes and the initial inequality level, as demonstrated by the values in column [c].

As the isolated average income effect on poverty is negative, the effect of an income increase on poverty reduction is smaller than when the initial development level is low. The same happens when the initial inequality index is high. These results confirm conclusions reached by Medina and Galván (2014b), who computed poverty elasticities and identified how poverty indicators, changes in income and the Lorenz curve were modified. They used household survey databases from 18 Latin American countries from 1997 to 2000 and the five years from 2002 to 2007. The aim of the study is to measure the ability of income and inequality to influence poverty reduction on the basis of simulated counterfactual scenarios that take account of the sensitivity of poverty indices to changes in income and inequality levels. The analysis is carried out using all income distribution information available in each country. Decomposition methods designed to separate out changes in poverty into income growth and inequality effects are applied to simulate counterfactual scenarios that provide insight into the importance of changes in income inequality on the basis of the marginal proportional rate of substitution (MPRS) proposed by Kakwani. The results suggest that it would be wrong to propose the same policy options to all countries, as the sensitivity of the poverty rate depends on its initial level and the degree of inequality.

Accordingly, we can affirm that in regions with low initial development levels or high initial inequality, or both, the conditions for reducing poverty through income growth are relatively unfavourable. This being so, we can conclude that the high inequality and low initial development level of most Brazilian states are impediments to alleviating poverty by improving incomes.

As for the estimated coefficient in column [c] relating to the interaction between the change in inequality and the inverse initial development level, it is negative and statistically significant. The same is true of the interaction between the inequality variable and its initial level. Thus, the effect of the change in inequality on poverty reduction is smaller when the initial development level is low or when the initial inequality level is high. In other words, a decline in income inequality may be less effective at decreasing poverty in regions that suffer from a low initial development level, high initial inequality or both.

The low initial development level and high initial income inequality in Brazil therefore pose difficulties for poverty reduction, regardless of whether this is pursued by efforts to boost economic growth or to reduce income inequality.

This is borne out by the findings of Medina and Galván (2014a), who employed different econometric methodologies to analyse the contribution of economic growth and inequality to the evolution of poverty using a database of household surveys conducted in a set of 18 Latin American countries. The progress made on inequality, together with an increase in per capita income, explains the drop in poverty seen during the period 2002-2007. The results suggest that it is possible to reduce poverty by means of policies primarily designed to diminish the inequality of income distribution, especially in more developed countries, while in poorer economies there is also a need to increase the incomes of disadvantaged families as a necessary condition for reducing poverty. These authors found that an appropriate mix of policies to increase incomes and improve income distribution would generate a virtuous circle of rapid and sustained poverty reduction. Clearly, the sensitivity of poverty indicators to changes in average family income is correlated with the level of inequality, so that reducing income inequality improves the decline in poverty.

The inverse initial development coefficient has a positive and statistically significant relationship with the percentage of poor people (values in column [c]). Thus, the higher the inverse initial development level, the greater the incidence of poverty. Or to put it another way, the lower the initial level of family income per capita, the higher the incidence of poverty.

The last rows of table 4 introduce the Arellano and Bond (1991) test results for the first- and second-order residual autocorrelations and the Hansen and Sargan tests for instrument validity. Going by the p-values in column [c], the results of the Arellano and Bond tests suggest that we can reject the null hypothesis of absence of first-order autocorrelation and accept the existence of second-order residual autocorrelation. The p-values for the Hansen and Sargan tests allow us to accept the hypothesis that the instruments used in the model estimations are valid.

1. Income-poverty and inequality-poverty elasticities in Brazilian states

The income-poverty and inequality-poverty elasticities for Brazilian states were calculated in accordance with expressions (10) and (11). The parameters for these two expressions were obtained by estimating the second model, on the basis that this was more appropriate for determining these elasticities because it took account of income distribution characteristics, the inequality level and the initial development level. Table 5 gives average elasticities for the Brazilian states and regions from 1995 to 2009, and shows the standard deviations estimated.

Table 5
Average income-poverty and inequality-poverty elasticities in Brazilian states

State	Income-poverty elasticity	Standard deviation	Inequality-poverty elasticity	Standard deviation
Maranhão	-1.61		2.33	
Piauí	-1.52		2.30	
Ceará	-1.50		2.34	
Rio Grande do Norte	-1.54		2.30	
Paraíba	-1.53		2.40	
Pernambuco	-1.56		2.52	
Alagoas	-1.55		2.50	
Sergipe	-1.56		2.47	
Bahia	-1.57		2.40	
North-east	-1.54	0.03	2.39	0.08
Minas Gerais	-1.58		2.48	
Espírito Santo	-1.61		2.51	
Rio de Janeiro	-1.63		2.49	
São Paulo	-1.61		2.42	
South-east	-1.60	0.02	2.47	0.03
Paraná	-1.59		2.49	
Santa Catarina	-1.65		2.38	
Rio Grande do Sul	-1.64		2.47	
South	-1.62	0.03	2.44	0.05
Mato Grosso do Sul	-1.58		2.49	
Mato Grosso	-1.61		2.49	
Goiás	-1.59		2.51	
Federal District	-1.62		2.50	
Mid-west	-1.60	0.01	2.49	0.009

Source: Prepared by the authors.

As expected from the signs of the theoretical income-poverty elasticity and inequality-poverty elasticity introduced in section III, the former is negative and the second positive for all Brazilian states and regions. In other words, growth in average income and the reduction of income inequality led to a fall in the number of poor individuals.

However, the values of these elasticities as shown in table 5 reveal that changes in income inequality have had a greater effect on poverty than average income growth. This matches the findings of Kakwani (1990) and Marinho and Soares (2003).

At the regional level, absolute values for income-poverty elasticity prove to be lower in the north-east than in the other regions. This result confirms the theoretical hypothesis that income-poverty elasticity is lower in economies with lower average incomes. In richer regions, average income growth has more influence on poverty reduction. These results confirm conclusions reached by Marinho and Soares (2003) and Hoffmann (2004). Consequently, less developed regions such as the Brazilian north-east have more difficulty reducing poverty through income growth. Likewise, inequality-poverty elasticity is also lower in the north-east than in the other regions, but changes in inequality have a greater impact on poverty than average income growth.

In general terms, these results suggest that inequality reduction policies are the most effective way of fighting poverty in Brazil.

VII. Final considerations

The aim of this article is to estimate poverty elasticities relative to income and inequality in Brazil in an effort to analyse the determinants of poverty reduction. More specifically, it assesses whether changes in poverty are the consequence of income redistribution, economic growth or both, bringing out the influence of each effect on poverty changes.

Estimation results for the first model showed that the income-poverty and inequality-poverty elasticities were -0.68 and 0.77, respectively. This means that a 1.0% increase in per capita income results in a reduction of 0.68% in the proportion of poor people. Likewise, 1.0% growth in inequality leads to growth of 0.77% in poverty. It is important to note that these results corroborate findings in international studies such as those by Kalwij and Verschoor (2004) and Bourguignon (2004) as well as those by Marinho and Soares (2003) and Hoffmann (2004) for Brazil. The implication of these findings is that policies aimed at reducing inequalities are more effective at fighting poverty than those solely concerned with boosting average income.

Estimated results for the second model, which allows elasticities to change through time, showed that the factors contributing to poverty growth were, in order of increasing importance: the interaction between income changes and the inverse initial development level, initial income inequality, the interaction between income changes and initial income inequality, and income inequality in the present.

The impact of income growth on poverty reduction is smaller when the initial development level is low, and also when initial inequality is high. Thus, we can conclude that regions with a low initial development level or high initial inequality, or both, are less favourably positioned to reduce poverty through income growth. It is accordingly clear that the high inequality and low initial development of most Brazilian states are obstacles to reducing poverty by raising incomes.

The effect of changes in inequality on poverty reduction is likewise smaller when the initial development level is low or when initial inequality is high. Thus, attempting to fight poverty by reducing income inequality in Brazilian states or regions that suffer from low initial development levels, high initial inequality or both may not have the expected outcome.

As already pointed out, the low initial development level and high initial inequality of Brazil are barriers to reducing poverty regardless of whether this is addressed by boosting economic growth or reducing income inequalities.

With regard to the income-poverty and inequality-poverty elasticities, it transpired that the impact of income inequality on poverty was greater than that of average income growth. This was also observed by Kakwani (1990) and Marinho and Soares (2003).

At the regional level, the absolute value of poverty-income elasticity is lower in the north-east than in the other Brazilian regions. This result confirms the theoretical hypothesis that poverty-income elasticity is lower in economies with low average incomes. In richer regions, the effect of average income growth on poverty reduction is stronger. The results obtained in this article agree with the findings of Marinho and Soares (2003) and Hoffmann (2004). In short, less developed regions like the Brazilian north-east have more difficulty reducing poverty through income growth.

Likewise, inequality-poverty elasticity is lower in the north-east than in other regions, but the impact of inequality on poverty is higher than the impact of average income growth. Overall, these results suggest that inequality reduction policies are most effective when it comes to fighting poverty in Brazil.

Bibliography

Ahn, S. C. and P. Schmidt (1995), "Efficient estimation of models for dynamic panel data", *Journal of Econometrics*, vol. 68, No. 1, Amsterdam, Elsevier.

Alesina, A. and D. Rodrik (1994), "Distributive policies and economic growth", *Quarterly Journal of Economics*, vol. 109, No. 2, Oxford University Press.

Arellano, M. and O. Bover (1995), "Another look at the instrumental variable estimation of error-components model", *Journal of Econometrics*, vol. 68, No. 1, Amsterdam, Elsevier.

Arellano, M. and S. Bond (1991), "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations", *The Review of Economic Studies*, vol. 58, No. 2, Oxford University Press.

Barreto, F. A. F. D. (2005), "Crescimento econômico, pobreza e desigualdade: o que sabemos sobre eles?", *Série Ensaios sobre Pobreza*, No. 1, Fortaleza, Federal University of Ceará.

Barro, R. (2000), "Inequality and growth in a panel of countries", *Journal of Economic Growth*, vol. 5, No. 1.

Barros, R. P. de and others (2007), "Determinantes da queda da desigualdade de renda Brasileira", *Texto para Discussão*, No. 1253, Rio de Janeiro, Institute of Economic Applied Research.

Barros, R. P. de, R. Henriques and R. Mendonça (2001), "A estabilidade inaceitável: desigualdade e pobreza no Brasil", *Texto para Discussão*, No. 800, Rio de Janeiro, Institute of Economic Applied Research.

Blundell, R. and S. Bond (1998), "Initial conditions and moment restrictions in dynamic panel data models", *Journal of Econometrics*, vol. 87, No. 1, Amsterdam, Elsevier.

Bourguignon, F. (2004), "The Poverty-Growth-Inequality Triangle", Washington, D.C., World Bank [online] http://siteresources.worldbank.org/INTPGI/Resources/342674-1206111890151/15185_ICRIER_paper-final.pdf.

___(2003), "The growth elasticity of poverty reduction: explaining heterogeneity across countries and time periods", *Inequality and Growth: Theory and Policy Implications*, T. Eicher and S. Turnovsky (eds.), Cambridge, Massachusetts, The MIT Press.

___(1981), "Pareto superiority of unegalitarian equilibria in Stiglitz' model of wealth distribution with convex saving function", *Econometrica*, vol. 49, No. 6, New York, Econometric Society.

Chen, S. and Y. Wang (2001), "China's growth and poverty reduction: recent trends between 1990 and 1999", *Policy Research Working Paper*, No. 2651, Washington, D.C., World Bank.

Cline, W. R. (2004), *Trade Policy and Global Poverty*, Washington, D.C., Center for Global Development/Institute for International Economics.

Datt, G. and M. Ravallion (2002), "Is India's economic growth leaving the poor behind?", *Journal of Economic Perspectives*, vol. 16, No. 3, Nashville, Tennessee, American Economic Association.

Deininger, K. and L. Squire (1996), "A new data set measuring income inequality", *World Bank Economic Review*, vol. 10, No. 3, Washington, D.C., World Bank.

Epaulard, A. (2003), "Macroeconomic performance and poverty reduction", *IMF Working Paper*, No. 03/72, Washington, D.C., International Monetary Fund (IMF).

Foster, J., J. Greer and E. Thorbecke (1984), "A class of decomposable poverty measures", *Econometrica*, vol. 52, No. 3, New York, Econometric Society.

Hoffmann, R. (2007), "Transferência de renda e a redução da desigualdade no Brasil e em cinco regiões entre 1997 e 2005", *Desigualdade de renda no Brasil: uma análise da queda recente*, vol. 2, R. Paes de Barros, M. N. Foguel and G. Ulyssea (orgs.), Brasilia, Institute of Applied Economic Research.

___(2005), "Elasticidade da pobreza em relação à renda média e à desigualdade no Brasil e nas unidades da federação", *Economia*, vol. 6, No. 2, National Association of Centers for Post-graduation in Economics (ANPEC).

___(2004), "Elasticidade da pobreza em relação à renda média e à desigualdade" [online] http://www.anpec.org.br/encontro2004/artigos/A04A054.pdf.

___(1995), "Desigualdade e pobreza no Brasil no período 1970-1990", *Revista Brasileira de Economia*, vol. 49, No. 2.

Kakwani, N. (1990), "Poverty and economic growth: with application to Côte d'Ivoire", *Living Standards Measurement Study Working Paper*, No. 63, Washington, D.C., World Bank.

Kalwij, A. and A. Verschoor (2004), "How good is growth for the poor? The role of the initial income distribution in regional diversity in poverty trends", *CentER Discussion Paper Series*, No. 115, CentER-Tilburg University.

Kraay, A. (2004), "When is growth pro-poor? Evidence from a panel of countries", Washington, D.C., *Policy Research Working Paper*, No. 3225, Washington, D.C., World Bank.

Kuznets, S. (1955), "Economic growth and income inequality", *American Economic Review*, vol. 45, No. 1, Nashville, Tennessee, American Economic Association.

Judson, R. A. and A. L. Owen (1999), "Estimating dynamic panel data models: a guide for macroeconomists", *Economics Letters*, vol. 65, No. 1, Amsterdam, Elsevier.

Lopez, J. H. (2004), "Pro-growth, pro-poor: is there a trade-off?", *Policy Research Working Paper Series*, No. 3378, Washington, D.C., World Bank.

Manso, C. A., F. A. Barreto and E. Tebaldi (2005), "O desequilíbrio regional brasileiro: novas perspectivas a partir das fontes de crescimento pró-pobre", *Série Ensaios sobre Pobreza*, No. 6, Fortaleza, Federal University of Ceará.

Marinho, E. and F. Soares (2003), "Impacto do crescimento econômico e da concentração de renda sobre a redução da pobreza nos estados brasileiros" [online] http://www.anpec.org.br/encontro2003/artigos/B36.pdf.

Medina, F. and M. Galván (2014a), "Crecimiento económico, pobreza y distribución del ingreso. Fundamentos teóricos y evidencia empírica para América Latina, 1997-2007", *Estudios Estadísticos series*, No. 82 (LC/L.3689), Santiago, Economic Commission for Latin America and the Caribbean (ECLAC).

___(2014b), "Sensibilidad de los índices de pobreza a los cambios en el ingreso y la desigualdad. Lecciones para el diseño de políticas en América Latina, 1997-2008", *Estudios Estadísticos series*, No. 87 (LC/L.3823), Santiago, Economic Commission for Latin America and the Caribbean (ECLAC).

Neri, M. (2006), "Desigualdade, estabilidade e bem-estar social", *Ensaios Econômicos*, No. 637, Rio de Janeiro, Getulio Vargas Foundation.

Ravallion, M. (2005), "Pro-poor growth: a primer", *Policy Research Working Paper*, No. 3242, Washington, D.C., World Bank.

___(2001), "Growth, inequality and poverty: looking beyond averages", *World Development*, vol. 29, No. 11, Amsterdam, Elsevier.

Ravallion, M. and S. Chen (1997), "What can new survey data tell us about recent changes in distribution and poverty?", *World Bank Economic Review*, vol. 11, No. 2, Washington, D.C., World Bank.

Ribas, R. P., A. F. Machado and A. B. Golgher (2006), "Fluctuations and persistence in poverty: a transient-chronic decomposition model for pseudo-panel data", *Texto para Discussão*, No. 290, Federal University of Minas Gerais.

Rocha, S. (2006), *Pobreza no Brasil. Afinal, de que se trata?*, Rio de Janeiro, FGV Editora.

Salvato, M. A. and A. F. de Araujo Junior (2007), "Crescimento pró-pobre no Brasil: uma avaliação empírica da década de 1990", *Texto para Discussão*, No. 02, Minas Gerais, Pontifical Catholic University of Minas Gerais.

Stewart, F. (2000), *Distribuição de renda e desenvolvimento*, Brasilia, NEAD.

Son, H. H. (2004), "A note on pro-poor growth", *Economics Letters*, vol. 82, No. 3, Amsterdam, Elsevier.

Windmeijer, F. (2005), "A finite sample correction for the variance of linear efficient two-step GMM estimators", *Journal of Econometrics*, vol. 126, No. 1, Amsterdam, Elsevier.

Personal income tax and income inequality in Ecuador between 2007 and 2011

Liliana Cano

Abstract

This paper uses data from individual income tax returns to explore the redistributive effect of personal income tax in Ecuador between 2007 and 2011. Following common practice in tax incidence analysis, we first compute indices of income tax progressivity and redistributive impact. We then mobilize microsimulation techniques to simulate the redistributive effect of personal income tax under different taxable income scenarios. Finally, we calculate the effective tax rates paid by top income groups and derive a range of optimal income taxes for the top 1% income group. We obtain two main empirical results. First, although Ecuador's personal income tax is highly progressive, its redistributive capacity is low: our findings show that high-income individuals are more likely to reduce their taxable income through legal tax deductions than low-income individuals. Second, while the effective tax rates paid by high-income individuals are relatively low, optimal tax rates could be as high as 63%.

Keywords

Income tax, fiscal policy, income distribution, mathematical models, simulation methods, Ecuador

JEL classification

D31, H24, O54

Author

Liliana Cano is a postdoctoral researcher with the Research Laboratory on Economics, Policy and Social Systems (LEREPS) of the University of Toulouse 1 Capitole, France. Email: liliana.cano@ut-capitole.fr.

I. Introduction

There has been increasing interest in the study of income inequality over recent years in the fields of both research and politics. Since the seminal contributions of Piketty (2001 and 2014), Piketty and Saez (2003) and Atkinson and Piketty (2007 and 2010) on the long-run evolution of income and wealth inequality in most industrialized countries, recent public debate has primarily focused on the role of income taxes in reducing inequality (Atkinson, 2014; Piketty, 2015). Personal income tax is the public policy instrument that is often considered when the main objective is to modify the post-tax income distribution (Poterba, 2007). However, the prospects of reducing income inequality through taxation crucially depend on how progressive a country's taxes are. Thus, the redistributive effect of income taxes has increasingly become a central issue in both developed and developing countries.

This paper casts new light on the redistributive effect of personal income tax in Ecuador between 2007 and 2011. Following common practice in the public policy evaluation literature, we first compute different indices of tax progressivity and redistributive effect, namely the Kakwani, Suits and Reynolds-Smolensky indices. Second, we employ microsimulation techniques to simulate the redistributive impact of Ecuador's personal income tax under different taxable income scenarios. We allow different definitions of income tax deductions and we present alternative scenarios that could potentially improve the redistributive effect of income taxes. Third, drawing on the top incomes literature (Piketty, 2001; Piketty and Saez, 2003; Atkinson and Piketty, 2010), we use homogenous series of top income shares in Ecuador from 2007 to 2011 (Cano, 2015) to compute the effective tax rates paid by top income groups. Finally, guided by the taxable income elasticity literature (Lindsey, 1987; Feldstein, 1999; Auten and Carroll, 1999; Gruber and Saez, 2002; Saez, 2001; Chetty, 2009; Saez, Slemrod and Giertz, 2012) and using different compensated and uncompensated elasticity values, we derive a range of optimal tax rates for the top 1% income group. Our results rely on individual income tax returns data compiled annually by the Ecuadorian Internal Revenue Service.

We have two main motivations for studying the redistributive effect of personal income tax in Ecuador. First, several recent studies have documented the decline of income inequality in most Latin American countries since the early 2000s (Gasparini and others, 2009; Cornia, 2010; López-Calva and Lustig, 2010; Lustig, López-Calva and Ortiz-Juárez, 2013; Cornia, 2014; ECLAC, 2012, 2013 and 2014), mainly owing to (i) a decline in the skill premium, (ii) a decrease in the urban-rural wage gap, (iii) public transfer programmes such as conditional cash transfer programmes and (iv) favourable external conditions. In addition, while taxation had a negligible effect on income inequality in previous decades (Cornia, Gómez Sabaini and Martorano, 2011), a growing body of literature has documented the positive effect of the tax reforms of the 2000s in reducing income inequality in the region (Jiménez, Gómez Sabaini and Podestá, 2010; Roca, 2009; Cetrángolo and Gómez Sabaini, 2006; Cornia, Gómez Sabaini and Martorano, 2011; Hanni, Martner and Podestá, 2015). The present paper aims to contribute to this growing literature by assessing the redistributive impact of personal income tax in Ecuador.

Alongside the emergence in the late 1990s of conditional cash transfer programmes, which promote long-term human capital accumulation among the less well-off (Rawlings and Rubio, 2005), policymakers in Latin America and the Caribbean have mainly focused on improving the way public transfer programmes are used to achieve fairer income distribution (e.g. through better targeting and coverage). The role of taxes in tackling income inequality has probably been treated as a secondary issue because Latin American tax systems have been based mainly on indirect taxes (value added tax and trade taxes), even though progressive income taxation began to be introduced in the region in the early twentieth century (Gómez Sabaini, 2006; Cornia, Gómez Sabaini and Martorano, 2011). In recent years, however, debate in Latin America has increasingly turned on the need to reinforce tax progressivity and to increase tax revenue and thereby improve public transfers. Indeed, social, economic and institutional reforms in most of Latin America and the Caribbean since the early 2000s have influenced a new wave of tax reform.[1] Following Fairfield (2010) and Gómez Sabaini and Moran (2014), this may be characterized as a "second generation" of tax reforms aimed at: (i) improving tax revenues and the fairness of the tax system, (ii) increasing tax progressivity and (iii) strengthening the ability of tax agencies to fight evasion. Thus, progressive taxation and income redistribution have been the primary goals of Latin American tax policy over recent years, at least in theory (Cornia, Gómez Sabaini and Martorano, 2011).

Although significant progress has been made towards these objectives,[2] challenges remain and further work is needed, since tax systems are still based on indirect taxes, making the entire system highly regressive. Goñi, López and Servén (2011) show, for instance, that income inequality after taxes and transfers has declined by less in Latin America and the Caribbean than in the countries of the Organization for Economic Cooperation and Development (OECD), whereas the Gini coefficient before fiscal policy is nearly the same in both groupings. The authors also show that the reduction in income inequality in Latin America and the Caribbean has mainly been achieved by public transfers, with taxes being of little help in tackling inequality. Ecuador's tax system is no exception.[3] The 2008 tax reform was enacted precisely to increase the progressivity of personal income tax, promote tax equity and increase tax revenue. This involved, firstly, creating two additional income tax brackets and a top marginal tax rate of 35% and, secondly, introducing new personal income tax deductions for spending on housing, education, health, clothing and food.[4] As yet, though, very little is known about the impact of this tax reform on income inequality in Ecuador.

A second motivation for this paper was the growing interest in the study of top income shares and the effective tax rates paid by high-income individuals. Following the seminal contributions of Piketty (2001 and 2014) and Piketty and Saez (2003) on the long-run distribution of top incomes in France and the United States, the relationship between income concentration and tax policy has received considerable attention in the fields of both research and politics. Following this literature, Cano (2015) constructed top annual income share series for Ecuador from 2004 to 2011 out of individual income tax return data, using population and income totals based on census estimates and national accounts, respectively, as external controls. In 2010, the share of total income accruing to the richest 1% of the population was between 14% and 17%, depending on how the income used as the numerator of the share is defined.[5] We employ these estimates in the present paper to calculate the effective tax rates paid by top income groups. A further motivation was the increasing interest in the question of how progressive income taxes ought to be. Using different estimates of elasticity and following the taxable income elasticity literature (Saez, 2001; Gruber and Saez, 2002), we derive a range of optimal tax rates for the top 1% income group.

The remainder of this paper is organized as follows. Section II describes data and methodology, section III presents results and section IV offers conclusions and discusses policy implications.

[1] A detailed review of Latin American tax reforms and tax policy patterns is provided in Gómez Sabaini (2006), Cetrángolo and Gómez Sabaini (2006), Gonzalez and Martner (2009), Jiménez, Gómez Sabaini and Podestá (2010), Gómez Sabaini and Jiménez (2012), Tanzi (2013) and Gómez Sabaini and Moran (2013).

[2] Technical reports prepared by the Economic Commission for Latin America and the Caribbean (ECLAC) have well documented the major achievements of the latest tax reforms in Latin America and the Caribbean, as has the latest report on revenue statistics in Latin America jointly produced by the Organization for Economic Cooperation and Development (OECD), the Inter-American Centre of Tax Administrations (CIAT), the Inter-American Development Bank (IDB) and ECLAC.

[3] In Ecuador, revenue raised from indirect taxes such as value added tax (VAT) averaged 4.9% of GDP over the period 1990-2000, increasing to nearly 8% of GDP in 2001-2013. In 2013, moreover, 53% of total tax revenue came from indirect taxes and just under 21% from taxes on income, profits and capital gains (OECD/ECLAC/CIAT/IDB, 2015).

[4] Personal tax deductions were introduced in the 2008 tax reform to promote tax equity. The main objective was precisely to allow low-income taxpayers to benefit from larger tax deductions.

[5] In order to provide an accurate picture of incomes at the top of the distribution, Cano (2015) constructed top income share series for Ecuador using different criteria (e.g., net income, gross income) and different percentages of intermediate costs and deductions as the numerators of the shares. In 2011, the income accruing to the top 1% was between 12% and 15%.

II. Data and methods

1. Data

Our estimates rely on individual income tax returns compiled yearly by the Ecuadorian tax administration.[6] The tax returns database provides information on individuals that includes: (i) employee earnings in the form of wages and salaries, (ii) capital income (dividends, interest and other capital income), (iii) business income, (iv) self-employment income, (v) income from other sources and (vi) tax deductions, tax liabilities and taxes paid by tax filers.

The data used for the analysis come from three different tax forms: (i) form 102, used to report information on wages, self-employment income, business income, capital income and other possible sources of income from tax filers required to keep accounts (e.g. individuals engaging in commercial activities); (ii) form 102A, used to report information on wages, self-employment income, capital income and other possible sources of income from tax filers not required to keep accounts; (iii) form 107, used to report information on formal employees' wages and salaries. Personal income is taxed at progressive marginal tax rates ranging from 0% to 35%. Individuals whose sole source of income is wages (i.e. employees) are not required to fill in a tax return because income tax is automatically withheld by employers. Nevertheless, employees earning income from sources other than wages (e.g. dividends, interest, rents) are required to consolidate all sources of income (e.g. wages and capital income) in a single annual tax return (tax form 102 or 102A). The same progressive rate schedule of between 0% and 35% is applied to total income. We identified the tax filers whose incomes were reported in both form 107 and form 102 or 102A, in order not to duplicate data on wages and salaries. In all such cases, we have only worked with the data from form 102 or 102A. Regarding capital income, before 2010 distributed dividends that had already been subject to corporation taxes were exempted from personal income tax to avoid double taxation. Since 2010, dividends received by individuals living in Ecuador have formed part of the personal income tax base. Income taxes in Ecuador are declared in United States dollars and are assessed at the individual level and not at the household level as they are, for instance, in the United States and in some European countries such as France, Germany and the United Kingdom.

By Latin American standards, the Ecuadorian income tax returns database is at roughly the midpoint in terms of the number of individuals covered. Almost 27% of the adult population (aged 20 and over) reported income to the Fiscal Administration in 2011, including those whose income was below the tax threshold.[7] Comparison with other countries for which estimates of top income shares and tax incidence are available shows that this was lower than the adult population coverage in Chile (67% in 2009) and Uruguay (74% in 2012), but higher than that in Colombia (4% in 2010) and Argentina (3% in 2004).[8]

[6] The Ecuadorian Internal Revenue Service (SRI) kindly agreed to give us access to the entire personal income tax returns database from 2004 to 2011. The database is composed of nearly 1.9 million observations on average (2.3 million in 2011). In this paper we work with the 2007-2011 period.

[7] Data for the population aged 20 and over come from the Ecuadorian National Survey of Employment, Unemployment and Underemployment (ENEMDU) of December 2011.

[8] The studies on top income shares and tax incidence are Fairfield and Jorratt (2016) for Chile, Burdín, Esponda and Vigorito (2014) for Uruguay, Alvaredo and Londoño (2013) for Colombia, and Alvaredo (2010) for Argentina. The low population coverage in Argentina and Colombia is probably due to the fact that their estimates exclude taxpayers whose sole source of income is wages.

2. Methods

We rely on four complementary methods to assess the relationship between personal income tax and income inequality. First, to analyse the redistributive effect of income taxes, we compute different progressivity and redistribution indices commonly proposed by the impact evaluation literature: the Kakwani and Suits indices to measure the level of progressivity and the Reynolds-Smolensky index to measure the redistributive effect of Ecuador's personal income tax. At the same time, we draw the concentration curves for personal income tax deductions and tax liabilities, along with the pre-tax income Lorenz curve, for the years 2008 and 2010. These progressivity and redistribution indicators allow us to study the distributional effect of income taxes while comparing the pre-tax and post-tax income distributions.

The concentration curves and the Kakwani, Suits and Reynolds-Smolensky synthetic indices are derived from the classical approach of the Lorenz curve and the Gini coefficient. The Lorenz curve (L_p) plots the cumulative percentage of income on the vertical axis against the cumulative percentage of individuals, ranked by income from poorest to richest, on the horizontal axis. The Gini coefficient (G_p) compares the area between the Lorenz curve (L_p) and the diagonal of perfect equality (45) against the total area under the diagonal, taking values of between 0 (absolute equality) and 1 (absolute inequality).

The Lorenz curve and Gini coefficient can be reformulated to assess changes in income distribution, notably when taxes are included. The concentration curve plots the cumulative percentage of tax liabilities on the vertical axis against the cumulative percentage of individuals, ranked by income, on the horizontal axis. If the tax concentration curve lies farther than the Lorenz curve from the diagonal of perfect equality, taxes are more unequally distributed than income and thus are progressive. The corresponding concentration coefficient C_t, also known as the quasi-Gini coefficient, is interpreted analogously to the Gini coefficient.

The Kakwani progressivity index is calculated by comparing the tax concentration curve and the pre-tax income Lorenz curve. The Kakwani index is defined as twice the area between the pre-tax income Lorenz curve and the tax concentration curve. Thus, it is equivalent to the difference between the concentration coefficient of taxes (or quasi-Gini coefficient) and the Gini coefficient of pre-tax income distribution (Kakwani, 1977):

$$K = C_t - G_y \tag{1}$$

where C_t is the concentration coefficient of taxes (or quasi-Gini coefficient) and G_y is the Gini coefficient of pre-tax income distribution. A positive Kakwani index indicates that taxes are progressive (i.e. tax liabilities increase with income), a negative Kakwani index that taxes are regressive (i.e. tax liabilities decrease with income) and a Kakwani index of zero that taxes are proportional to income.

Another popular index of progressivity is the one proposed by Suits (1977), which measures the departure from proportionality by comparing the pre-tax income Lorenz curve with the diagonal line of perfect equality. The Suits index is an adaptation of the Gini coefficient and is constructed by plotting the cumulative percentage of taxes on the vertical axis against the cumulative percentage of income on the horizontal axis. As suggested by Amarante and others (2011), the Suits index can be formulated as:

$$S = 2\int_0^1 \left(i - C_f\left(i\right)\right)di \tag{2}$$

If taxes are proportional, the concentration curve of taxes coincides with the diagonal line of perfect equality (45°), and the Suits index will take the value of 0. If taxes are progressive, the concentration curve will be below the line of perfect equality, and the Suits index will be positive. And if taxes are regressive, the concentration curve will be above the line of perfect equality, and the Suits index will be negative. For instance, if only high-income individuals paid taxes, the Suits index would take a value of 1. Conversely, if only the poorest individuals paid taxes, the Suits index would take a value of -1. Although the Kakwani and Suits indices are quite similar in design, there are some differences between them. As pointed out by Amarante and others (2011), while the Kakwani index integrates with respect to the population, the Suits index integrates with respect to income.

To measure the redistributive effect of income taxes, we make use of the Reynolds-Smolensky index, which measures how taxes affect after-tax income distribution, capturing the difference between the pre-tax and post-tax income Ginis as follows:

$$RS = G_y - G_{y-t} \tag{3}$$

where G_y is the Gini coefficient before taxes and G_{y-t} is the Gini coefficient after taxes (Reynolds and Smolensky, 1977). A positive Reynolds-Smolensky index indicates that taxes are progressive, because the post-tax income distribution is more equal than the pre-tax income distribution. A high Reynolds-Smolensky index value suggests that taxes have great redistributive potential.

Second, we use static microsimulation techniques to simulate the redistributive impact of personal income tax in Ecuador on different definitions of taxable income. For each observation we arithmetically simulate pre-tax income under two different counterfactual scenarios: (i) a 50% threshold for costs and deductions and (ii) elimination of all income tax deductions.[9] Counterfactual scenarios show what would occur if changes in income tax deductions were implemented. We then apply tax rates and income tax schedules in both counterfactual scenarios and for each observation to calculate tax liabilities and post-tax income. Finally, we calculate tax progressivity and redistribution indices for the scenarios simulated.

Third, to cast further light on the factors influencing the redistributive capacity of income tax policy in Ecuador, we compute effective tax rates paid by top income groups, using the top income share series constructed by Cano (2015) to calculate the income tax rates actually paid by high-income individuals.

Fourth, we draw on the taxable income elasticity literature and employ different international values for compensated and uncompensated elasticity (e.g. 0.2 and 0.5), as proposed by Saez (2001), to derive a range of optimal tax rates for the top 1% income group.

A word of caution is needed here. Our analysis relies on individual income tax returns and therefore does not take account of workers operating in the informal sector or individuals whose income does not exceed the standard personal allowance (US$ 9,210 in 2011), with the exception of formal sector workers who are present in the database even though they do not earn enough to pay taxes. Our data were also affected by tax evasion and avoidance issues. Because of the methodological differences, our results are likely to differ from other studies on the redistributive impact of personal income tax in Ecuador (Roca, 2009; Hanni, Martner and Podestá, 2015), based on data from household surveys.

[9] There are two different types of income tax deductions in Ecuador: (i) all costs and deductions that are mandatory by law; and (ii) personal income tax deductions for spending on housing, education, health, clothing and food allowed under the tax reform of 2008.

III. Results

1. Measuring tax progressivity and the redistributive effect of personal income tax

This subsection assesses the progressivity and redistributive capacity of personal income tax in Ecuador over the period 2007-2011 (i.e. before and after the 2008 tax reform). As mentioned in the section on methods above, we compute the Kakwani and Suits indices to analyse the progressivity of personal income tax and the Reynolds-Smolensky index to measure its redistributive capacity.[10] Table 1 presents the results.

Table 1
Ecuador: personal income tax progressivity and redistribution indicators

Indicator	2007	2008	2009	2010	2011
Pre-tax Gini	0.6006	0.6558	0.6441	0.6378	0.5938
Post-tax Gini	0.5921	0.6483	0.6377	0.6307	0.5844
Average tax rate	0.0249	0.0267	0.0218	0.0234	0.0291
Reynolds-Smolensky index	0.0085	0.0075	0.0064	0.0071	0.0093
Kakwani progressivity index	0.3388	0.2756	0.2915	0.3023	0.3145
Suits progressivity index	0.5711	0.4503	0.4752	0.528	0.4623

Source: Prepared by the author, on the basis of the individual income tax returns database of the Ecuadorian Internal Revenue Service (SRI).

The Kakwani progressivity index K is calculated as the difference between the tax concentration coefficient[11] and the Gini coefficient for pre-tax income. If K>0, income taxes are progressive, while if K<0 income taxes are regressive and do not contribute to a reduction in income inequality. Results from the Kakwani index K for 2007 (before the tax reform) and 2008 onward (after the tax reform), displayed in table 1, suggest that personal income tax is progressive in Ecuador, as it is positive throughout the period studied: 0.34 in 2007, 0.28 in 2008, 0.29 in 2009, 0.30 in 2010 and 0.32 in 2011. This shows that the cumulative percentage of income taxes paid by wealthy taxpayers is higher than the cumulative percentage of income taxes paid by the poorest individuals.

The same pattern emerges from the Suits index, interpreted analogously to the Gini coefficient. For instance, a Suits index value of 1 corresponds to the hypothetical situation in which high-income individuals alone pay the totality of income taxes (extreme progressivity), while a Suits index value of -1 corresponds to the extreme situation in which only the poorest individuals pay the totality of income taxes (extreme regressivity). The results displayed in table 1 show that the Suits index is positive over the period 2007-2011 (0.57 in 2007, 0.48 in 2009 and 0.46 in 2011), indicating that personal income tax is progressive in Ecuador. In addition, we follow Roca (2009) and Hanni, Martner and Podestá (2015) in calculating the share of total income taxes paid by each income decile. Like these authors, we find

[10] Progressivity and redistribution indices are computed using the PROGRES module developed for Stata by Philippe Van Kerm and Andreas Peichl of the Luxembourg Institute of Socio-Economic Research (LISER) and the Centre for European Economic Research (Van Kerm and Peichl, 2007).

[11] It will be recalled that the tax concentration coefficient, also known as the quasi-Gini coefficient, is the coefficient corresponding to the tax concentration curve.

that nearly 90% of Ecuador's personal income tax is paid by the richest decile. This result shows that personal income tax in the country is highly progressive.[12]

As pointed out by Roca (2009, pp. 53-55), measurement of the progressivity and redistributive capacity of income taxes in Latin American countries should be seen as a theoretical exercise because of the high levels of tax evasion and avoidance, and also because of personal tax allowances.[13] Moreover, Latin American countries are unfortunately characterized by large disparities of income. Indeed, Hanni, Martner and Podestá (2015) show that the share of total income accruing to the richest decile in Latin American countries averaged 32% in 2012. Most probably because of the high levels of income concentration, tax evasion and tax avoidance, personal income tax in Ecuador is actually less progressive than documented in this section. For this reason, and to cast new light on the progressivity of income taxes, we compute effective tax rates paid by top income groups in section III.3.

In sum, the results from the Kakwani and Suits indices and those for the personal income tax burden suggest that personal income tax is generally progressive in Ecuador. These findings are in line with trends found by Roca (2009) and Hanni, Martner and Podestá (2015), which are based on household surveys.

From a theoretical point of view, progressive income taxes "push" the pre-tax income Lorenz curve towards the diagonal, making the after-tax income distribution less unequal. The magnitude of this movement is captured by the Reynolds-Smolensky index, which measures the difference between the Gini coefficient for pre-tax income and the Gini coefficient for post-tax income. The results of the Reynolds-Smolensky index for Ecuador are also displayed in table 1. At first glance, it can be seen that the pre-tax Gini coefficient is higher that the post-tax Gini coefficient over the period studied, suggesting that personal income tax effectively reduces income inequality. However, this reduction is so small as to be almost unobservable, with no significant change in the Reynolds-Smolensky index in any year. The Gini index after income taxes declines from 0.655 to 0.648 in 2008 and from 0.637 to 0.630 in 2010. These results reveal that although Ecuador's personal income tax is progressive, as shown by the Kakwani and Suits indices, its redistributive capacity is very weak. Moreover, no significant changes are observable between 2007 and 2008, i.e. before and after the tax reform which was actually enacted to increase tax progressivity and therefore income redistribution.

To cast further light on the factors that erode the redistributive capacity of personal income tax, we study the new personal income tax deductions introduced by the 2008 tax reform over the period 2008-2010.[14] Figures 1 and 2 plot the pre-tax income Lorenz curve for 2008 and 2010, respectively (dark red lines), along with concentration curves for each type of personal income tax deduction and for income tax liabilities.[15]

Several observations arise from figures 1 and 2. First, the tax concentration curve (plotted in green) lies below the pre-tax Lorenz curve in both figures, showing that tax liabilities are more unequally distributed than income.[16] This result is in line with previous evidence from the Kakwani and Suits indices showing personal income tax to be progressive.

---

[12] This result probably reflects the fact that most tax filers present in the tax returns database, especially workers, earn less than the threshold above which income tax is levied.

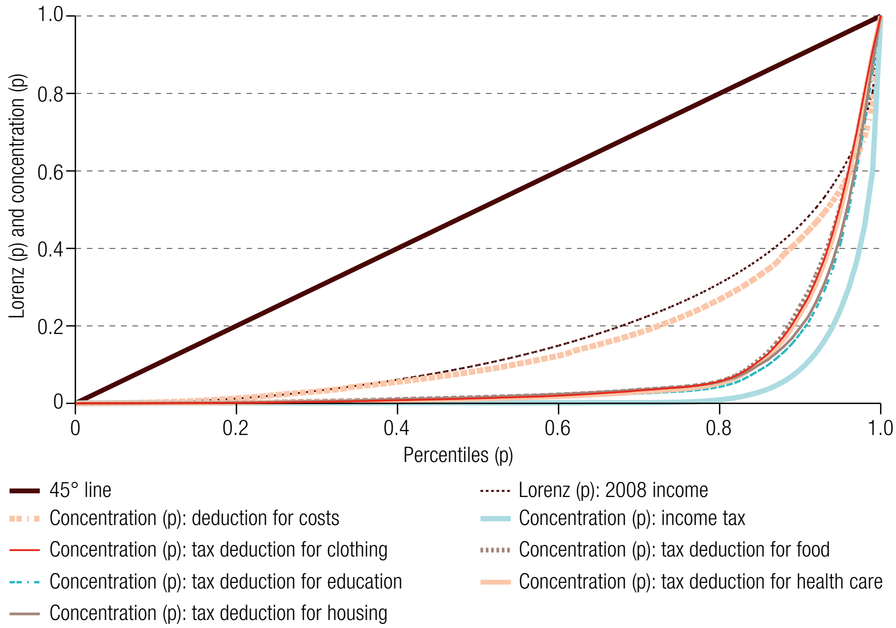
[13] Roca (2009) suggests that nearly 70% of "lowest-income" individuals in Ecuador do not earn enough to pay taxes, and therefore tax evasion mostly concerns high-income individuals.

[14] Unfortunately, we did not have access to microdata on personal income tax deductions for 2011.

[15] Concentration curves plot the cumulative percentage of different personal income tax deductions and tax liabilities against the cumulative percentage of population ranked by total income (C_p).

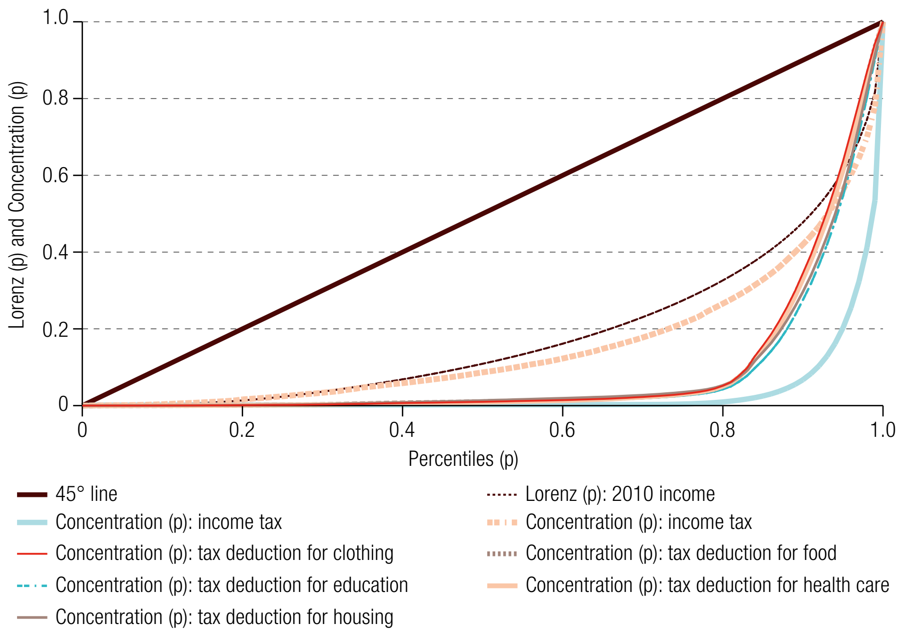
[16] In other words, the share of tax liabilities accruing to low-income taxpayers is smaller than the share of tax liabilities accruing to wealthy taxpayers.

Figure 1
Ecuador: Lorenz and concentration curves for income tax deductions and tax liabilities, 2008



Source: Prepared by the author, on the basis of the individual income tax returns database of the Ecuadorian Internal Revenue Service (SRI).
Note: The chart plots the gross pre-tax income Lorenz curve alongside different concentration curves for income tax deductions and tax liabilities in 2008.

Figure 2
Ecuador: Lorenz and concentration curves for income tax deductions and tax liabilities, 2010



Source: Prepared by the author, on the basis of the individual income tax returns database of the Ecuadorian Internal Revenue Service (SRI).
Note: The chart plots the gross pre-tax income Lorenz curve alongside different concentration curves for income tax deductions and tax liabilities in 2010.

Second, figures 1 and 2 both clearly show that the taxpayers who have benefited most from personal income tax deductions are those in the eightieth percentile and above. Similarly interpreting the tax liability concentration curves, we observe that the share of personal income tax deductions accruing to high-income taxpayers is greater than the share of personal income tax deductions accruing to the poorest taxpayers. Put differently, high-income individuals (from the eightieth percentile upward) are the beneficiaries of most of the new personal income tax deductions introduced by the 2008 tax reform. Cano (2015) shows that the top 5% income share in Ecuador, controlled for the total income and total population variables, represented the top 22 centiles of the tax income distribution in 2008 and the top 19 centiles in 2011 (without the total population control variable). Thus, figures 1 and 2 show that the richest 5% have benefited the most from personal income tax deductions.

While all indicators suggest that the design of personal income tax in Ecuador is broadly progressive, a detailed analysis of income tax deductions shows that high-income individuals are more likely to reduce their taxable income through personal income tax deductions than middle- and low-income individuals, eroding the redistributive effect of progressive taxation. This result is in line with the results of the Reynolds-Smolensky index presented in table 1, leading us to confirm that the redistributive capacity of personal income tax is weak in Ecuador, as in most Latin American countries (Roca, 2009). Tax simulations carried out by ECLAC have also highlighted the fact that targeted public expenditure using revenue generated by stronger personal income tax is generally more redistributive than improvements associated with personal income tax itself.

2. Microsimulation and income tax incidence: alternative scenarios

To shed more light on the redistributive effect of personal income tax, this section employs microsimulation techniques at the level of the individual to capture how the pre-tax income distribution changes when legally mandated personal income tax deductions are reduced or eliminated. Following on from the preceding section, where we empirically demonstrated that the "winners" from personal income tax deductions were taxpayers belonging to the top 5% income group,[17] since it is their taxable income that tax deductions reduce the most, the main objective of this subsection is to test whether the redistributive effect of personal income tax increases when tax deductions are modified.

Microsimulation models are often characterized as static, dynamic and behavioural.[18] Static microsimulation models are traditionally employed to evaluate the immediate impact of policy modifications on individuals. In them, microunits of observation remain constant, and they take no account of ongoing changes in demographic composition or of behavioural changes due to policy shifts or the time dimension. By contrast, dynamic microsimulation models contain specific parameters that estimate individuals' responses to policy changes, and are generally employed to analyse long-term outcomes, intergenerational impacts, and socioeconomic and demographic projections under current policies, among other things (Bourguignon and Spadaro, 2006). Behavioural models, meanwhile, use microeconometric tools to estimate the effects of policy modifications on individuals' behaviour, usually in respect of the labour supply (Figari, Paulus and Sutherland, 2015).

We employ a static microsimulation model to analyse pre-tax and post-tax income distributions and tax liabilities under different simulated scenarios. Our main objective is to give a bird's eye view of possible redistributive effects if personal income tax deductions were modified. As our microsimulation

[17] It will be recalled that income and population totals are used as external controls to construct top income share groups.

[18] Spadaro (2007) and Figari, Paulus and Sutherland (2015) offer an excellent overview of microsimulation approaches and different microsimulation models typically employed to evaluate the impact of public policies.

approach is static, we do not incorporate changes in individuals' behaviour resulting from tax modifications. Drawing on our micro level data, for each tax filer we arithmetically simulate pre-tax income for two different counterfactual scenarios.[19] In the first scenario, we allow 50% of costs and deductions and remove personal income tax deductions. In the second, we remove income tax deductions of all kinds. We then employ statutory tax rates and tax tables for 2008, 2009 and 2010 to simulate the new tax liabilities and after-tax income for each individual under both scenarios.

The baseline and simulated scenarios are constructed as follows:

Baseline scenario 0: $Y_0 - t_0$, where Y_0 is gross income and t_0 are tax liabilities reported by tax filers.

Scenario 1: $Y_0 - t_1$, where Y_0 is gross income and t_1 are simulated tax liabilities when 50% of costs and deductions are allowed and all personal tax deductions are eliminated.

Scenario 2: $Y_0 - t_2$, where Y_0 is gross income and t_2 are simulated tax liabilities when all tax deductions are eliminated.

The methodology described in subsection 1 is then used to compute inequality, progressivity and redistribution indicators for the simulated scenarios.

Table 2 presents the results of this tax policy evaluation over three years. In 2008, first, the Kakwani and Suits progressivity indices are positive in all the simulated scenarios, meaning that personal income tax remains progressive, as it is in the baseline. This result was to be expected, since no new marginal tax rates or new income tax brackets are created in either scenario 1 or scenario 2, with the simulations in fact employing the same tax rates as scenario 0. However, the Reynolds-Smolensky index, which captures the difference in the Gini coefficient before and after personal income tax, changes more substantially in both simulated scenarios.

Table 2

Ecuador: static microsimulation of the distributional effects of personal income tax, 2008-2011

Indicator	2008			2009			2010		
	Baseline	Scenario 1	Scenario 2	Baseline	Scenario 1	Scenario 2	Baseline	Scenario 1	Scenario 2
Pre-tax Gini	0.6558	0.6123	0.6504	0.6441	0.5993	0.6396	0.6378	0.5911	0.6331
Post-tax Gini	0.6483	0.581	0.6140	0.6377	0.5679	0.6027	0.6307	0.5607	0.5969
Average tax rate	0.0267	0.0872	0.1106	0.0218	0.0847	0.1086	0.0234	0.0814	0.1060
Reynolds-Smolensky index	0.0075	0.0313	0.0364	0.0064	0.0313	0.0369	0.0071	0.0303	0.0362
Kakwani progressivity index	0.2756	0.3274	0.2922	0.2915	0.3385	0.3025	0.3023	0.3425	0.3058
Suits progressivity index	0.4503	0.5797	0.5343	0.4752	0.5875	0.5421	0.528	0.5851	0.5409

Source: Prepared by the author, on the basis of the individual income tax returns database of the Ecuadorian Internal Revenue Service (SRI).

Note: This table presents progressivity and redistribution indicators under different simulated scenarios. The baseline is the actual situation. In scenario 1, the personal income tax deductions mandated by the 2008 tax reform are eliminated and other legal tax deductions are allowed to the extent of 50%. In scenario 2, both tax deductions and personal income tax deductions are eliminated.

Whereas the redistributive effect of personal income tax is very weak in scenario 0 (baseline), with the Gini coefficient declining from 0.656 before tax to 0.648 after tax, simulated scenario 1, in which the new personal income tax deductions are eliminated, presents a larger redistributive effect.

Indeed, the Reynolds-Smolensky index rises from 0.0075 (baseline) to 0.0313 (scenario 1), with income inequality as measured by the Gini coefficient would decline by nearly 3 points, from 0.61

[19] A counterfactual scenario shows what would happen if a policy change were implemented.

to 0.58, after tax policy. Put differently, simulated results from scenarios 1 and 2 suggest that the redistributive effect of personal income tax would be greater if personal income tax deductions had not been introduced by the 2008 tax reform. The same pattern is found for 2009 and 2010. Table 2 shows that personal income tax remains progressive in all simulated scenarios. Moreover, when personal income tax deductions are eliminated (simulated scenario 1), the redistributive effect of direct taxation increases.

The results for scenario 2, where all tax deductions (personal tax deductions and other legal deductions) are eliminated, are quite similar to those already observed for scenario 1 in all years. The redistributive effect of personal income tax, as measured by the Reynolds-Smolensky index, increases by nearly 4 points, with the Gini coefficient declining from 0.65 before income tax to 0.61 after tax in 2008. Similar patterns are observed in 2009 and 2010.

To summarize, the results of the simulated scenarios confirm that the personal income tax system is progressive. Moreover, these scenarios show that the redistributive effect of income taxes could be greater if tax deductions in general were regulated. We have simulated extreme cases in which all tax deductions are eliminated, but we know that some kinds of deductions are needed to increase income tax fairness. In Ecuador, the introduction of personal tax deductions has created incentives for individuals, especially those belonging to the top 5%, to reduce their taxable income and therefore tax liabilities. As suggested earlier, this cancels out the progressivity effect theoretically sought by the tax reform. It is therefore imperative to properly regulate tax allowances and tax deductions in the country.

To throw more light on other factors that erode the redistributive effect of personal income tax, the next section analyses the effective tax rate paid by high-income individuals over the period 2007-2011. We then draw on the literature dealing with the elasticity of taxable income to compute optimal tax rates for high-income groups, and propose policy recommendations.

3. Effective tax rates paid by top income groups

(a) Constructing top income share series

We start our analysis by constructing series of top income shares for the period 2007-2011. As in Cano (2015), we follow the standard top incomes literature (Piketty, 2001; Piketty and Saez, 2003; Atkinson and Piketty, 2007; Atkinson, Piketty and Saez, 2011), constructing first the top 1% income series (denoted P99-100) and then series for a number of finer fractiles: P99.5-100 (the top 0.5%), P99.9-100 (the top 0.1%), and P99.99-100 (the top 0.01%).

Each fractile is constructed relative to the total number of potential tax filers in the entire Ecuadorian population of adults aged 20 and over, which, as pointed out by Piketty (2003), should not be confused with the actual number of tax returns filed. Income is defined as being prior to personal income taxes. This definition includes all items reported on tax returns: salaries and wages, pensions, self-employment income, net income from unincorporated businesses, dividends, interest, other capital and rental income and other income items. We then compute income shares by dividing the income amounts accruing to each fractile (P99-100, P99.9-100, P99.99-100) by a control total for income. Atkinson, Piketty and Saez (2011) propose two different methods for computing this control. The first starts from income tax data and adds the income of non-filers, while the second computes an external control total derived from national accounts. In this paper, as in Cano (2015), we follow the second approach, also employed by Piketty (2001) and Piketty and Saez (2003), and construct a control total for income from national accounts as follows: balance of primary household income, plus social benefits other than social transfers in kind, minus employers' actual social contributions, minus employers' imputed social contributions, minus attributed property income of insurance policyholders, minus imputed rentals for owner-occupied housing, minus fixed capital consumption. This approach generates a control variable for total income of about 59% to 66% of gross domestic product (GDP), depending on the year. The results show the share of total income accruing to the richest 1% income group in 2010 ranging from 14% to 17%, depending on the definition of income employed as the numerator of the share.[20]

Once top income share series have been constructed, we analyse the effective tax rates paid by these groups.[21] This paper is a contribution to the growing literature analysing not only top income dynamics but also their tax policy implications (Piketty and Saez, 2003; Saez and Veall, 2005; Landais, Piketty and Saez, 2011; Alvaredo and Londoño, 2013; Burdín, Esponda and Vigorito, 2014; Fairfield and Jorratt, 2016).

Table 3 presents the results for effective tax rates paid by top income groups. While personal income in Ecuador is taxed at progressive marginal rates rising from 0% to 35%, the effective tax rates paid by very high-income individuals are lower because of tax deductions, tax exemptions and probably tax evasion. Over the period 2007-2011, the top 1% income group paid an average effective tax rate of about 7% and the top 0.1% an average effective tax rate of 9.4%. Our results show that, other than in 2010, average effective tax rates decreased within top income groups, especially for the top 0.1%, 0.05%, 0.01% and 0.001% income groups. For instance, while the average effective tax rate of the top 1% income group was nearly 6.3% in 2009, the average tax rate of the top 0.001% income group was 1.9%.

This situation is connected with the composition of income in top groups and with the fact that capital income is taxed less than wage income. Cano (2015) analysed the composition of income in top groups and found that in 2011 the income of the top 1% was mainly composed of salaries (45%) and income from business activities (27%) or self-employment (21%), and only to a lesser extent of capital income (7%). Meanwhile, the income of the top 0.01% comprised a smaller share of salaries (16%) but larger ones of income from business activities (29%), self-employment income (28%) and capital income (27%). It is interesting to note how much the composition of income varied within the top decile (P90-100). While the share of salary income clearly decreased up the scale, that from self-employment and business activities increased up to the top 0.1%-0.01% before decreasing again towards the top 0.01%. By contrast, the share of capital income increased at the very top. For instance, while the share of capital was 4% in the top 10% and just under 1.7% in the top 10%-5% in 2011, it was 50% in the top 0.001%. The rising share of capital income in very high-income groups suggests that those whose income comes from this source are the largest group at the top of the income distribution in Ecuador.

It should be noted that table 3 displays a marked difference between the situation in 2007 and that from 2008 onwards, as average tax rates for very high-income groups dropped from their 2007 level, although they recovered significantly in 2010. In our view, the fall-off between 2007 and 2009 is probably explained by high-income individuals reducing their taxable income through tax deductions, especially the new personal income deductions allowed by the 2008 tax law. In this subsection, we have shown graphically that the new personal income deductions were very much concentrated at the upper end of the distribution (i.e. from the eightieth percentile upward or the top 5% when a control total for population is used). The results presented in this section suggest that the richest 0.1% have probably benefited most from this kind of tax deduction.

[20] In order to provide an accurate picture of incomes at the top of the distribution, Cano (2015) constructed top income share series for Ecuador under different definitions of income as the numerator of the share.

[21] Effective tax rates are defined as individual income taxes paid divided by declared earned income.

Table 3

Ecuador: average effective income tax rates in top income groups, 2007-2011

Year	Top 10%	Top 5%	Top 1%	Top 0.5%	Top 0.1%	Top 0.05%	Top 0.01%	Top 0.001%	Top 1.0%-0.5%	Top 0.5%-0.1%	Top 0.1%-0.05%	Top 0.05%-0.01%	Top 0.01%-0.001%
2007	3.1	4.0	7.4	9.0	12.4	13.3	12.2	6.8	3.6	6.4	10.3	14.2	15.1
2008	3.7	4.7	7.2	8.0	8.5	7.7	5.8	2.4	5.1	7.5	10.3	9.4	7.6
2009	3.1	3.9	6.3	6.9	7.0	6.2	4.4	1.9	4.4	6.9	9.1	8.0	5.9
2010	3.3	4.2	7.0	7.9	9.5	9.7	9.6	8.8	4.3	6.5	8.9	9.9	9.9
2011	4.7	6.2	8.7	9.3	9.7	9.2	6.6	3.7	7.3	9.0	10.9	10.7	8.9

Source: Prepared by the author, on the basis of the individual income tax returns database of the Ecuadorian Internal Revenue Service (SRI).

Note: This table displays average effective tax rates in top income groups, using the gross income criterion (business income is net of costs and deductions).

These results lead us to believe that despite the high level of income concentration, high-income individuals pay very low effective income tax rates. In addition, although all the synthetic indicators (which are more sensitive to changes in the middle of the distribution than at the tails) suggest that personal income tax in Ecuador is progressive, a more thorough analysis of tax rates actually paid by high-income individuals shows that tax rates decrease within top income groups.

4. Using elasticities to derive optimal tax rates for high-income individuals

This subsection draws on the taxable income elasticity literature to derive optimal income tax rates for top income groups. We employ different compensated and uncompensated elasticity values as proposed by Saez (2001) and derive a range of optimal tax rates for the top 1% income group. Unfortunately, we were not able to calculate the elasticity of taxable income with respect to marginal rates using the 2008 tax reform as a natural experiment, because the results were heavily influenced by changes in population coverage between 2007 and 2008, with the percentage of adults included in the tax returns database increasing from 18% in 2007 to nearly 27% in 2011 (see chart in annex A1). Since the changes in population coverage distorted our elasticity results, we decided to employ standard elasticity values proposed by the literature. Fairfield (2010) proceeded in the same way to estimate optimal tax rates for high-income individuals in Chile and Argentina.[22]

We proceed as follows. First, following the top incomes literature (Atkinson, Piketty and Saez, 2011), we calculate the inverted Pareto coefficient (β) for the period 2007-2011 as follows:

$$\alpha = \frac{1}{1 - \left[\dfrac{\log \frac{S_1}{S_{0.1}}}{\log 10} \right]}$$

$$\beta = \frac{\alpha}{\alpha - 1}$$

where S_1 is the top 1% income share and the top 0.1% income share. As pointed out by Atkinson, Piketty and Saez (2011), a higher inverted Pareto coefficient β yields a flatter upper tail in the distribution. Borrowing the explanation proposed by these authors, if β=2 then the average income of individuals

[22] We computed the elasticity of taxable income with respect to marginal tax rates for high-income individuals by following Riihela, Sullstrom and Tuomala (2014), thus: $\epsilon = \dfrac{\left(LogS_1 - LogS_0\right)}{Log\left(1 - t_1\right) - Log\left(1 - t_0\right)}$, where S_1 is the top 1% income share after a tax reform, S_0 the top 1% before the reform, t_1 the top marginal tax rate after reform and t_0 the marginal tax rate before reform. Unfortunately, our results were heavily influenced by the increase in adult population coverage between 2007 and 2008.

with incomes above US$ 100,000 is US$ 200,000 and the average income of individuals with incomes above US$ 1 million is US$ 2 million. Once our β Pareto parameters have been constructed, we follow Saez (2001) and derive optimal tax rates by applying his formula as follows:

$$\tau = \frac{1 - g}{1 - g + \zeta^u + \zeta^c (a - 1)}$$

where τ is the optimal tax rate, g the redistributive goal of government,[23] α–1 the Pareto parameter, ξ^u the uncompensated elasticity and ξ^c the compensated elasticity. As suggested by Saez, the case g=0 corresponds to one where the government does not value the marginal consumption of high-income individuals and sets the highest top rate it can to maximize its tax revenue from them. The formula proposed by Saez is mostly specialized for g=0. Moreover, as suggested by Saez, our estimates of optimal tax rates are adjusted for the presence of a value added tax (VAT) of 12% as follows: (1–t)*, where τ is the consumption tax rate.

Since there is little agreement in the taxable income elasticity literature when it comes to fixed values for the elasticity of top income individuals' taxable income, we follow Saez (2001) and Fairfield (2010) in deriving optimal tax rates by testing different elasticity values as follows: uncompensated elasticities of 0, 0.2 and 0.5 and compensated elasticities of 0.2, 0.5 and 0.8. Additionally, we present both results when g=0 and g=0.25.

Table 4

Ecuador: elasticity-derived optimal tax rates for high-income individuals, 2007-2011

Uncompensated elasticity			0			0.2			0.5
Compensated elasticity		0.2	0.5	0.8	0.2	0.5	0.8	0.5	0.8
g=0									
Year	β								
2007	1.979	74%	59%	49%	63%	52%	44%	44%	39%
2008	2.023	73%	58%	48%	63%	51%	44%	44%	38%
2009	2.078	72%	57%	47%	62%	51%	43%	43%	37%
2010	1.961	74%	59%	50%	63%	52%	45%	44%	39%
2011	1.953	74%	60%	50%	63%	52%	45%	45%	39%
g=0.25									
Year	β								
2007	1.979	70%	53%	43%	58%	46%	38%	38%	32%
2008	2.023	69%	52%	42%	57%	45%	37%	37%	32%
2009	2.078	68%	51%	41%	57%	44%	36%	37%	31%
2010	1.961	70%	54%	43%	58%	46%	38%	38%	33%
2011	1.953	70%	54%	44%	58%	46%	39%	38%	33%

Source: Prepared by the author, on the basis of the individual income tax returns database of the Ecuadorian Internal Revenue Service (SRI).

Note: Following Saez (2001), this table displays optimal tax rates for high-income individuals, with g being the ratio of social marginal utility with infinite income over the marginal value of public funds, β the Pareto parameter, ξ^U the uncompensated elasticity and ξ^C the compensated elasticity. The results are based on income tax returns data. The β Pareto parameter was computed using top income share series. Optimal tax rates are adjusted for the presence of VAT at 12%.

Table 4 displays the optimal top marginal tax rates computed for Ecuador. Assuming that elasticities are the same as in the United States at nearly 0.2 (both compensated and uncompensated), the optimal top marginal tax rate for high-income individuals is between 57% and 63%. The proposed top marginal tax rates are naturally higher than the current top marginal tax rate of 35%. Our results

[23] The concept of a social welfare function in the computation of top marginal tax rates is very well documented by Saez (2001) and Diamond and Saez (2011).

are in line with Fairfield's for Chile and Argentina, as that author proposes optimal tax rates of between 55% and 64% for Chile and between 56% and 59% for Argentina, assuming elasticities similar to those for the United States. For France, Saez (2001) used taxable income elasticities to propose an optimal tax rate for high-income individuals of 75%.

IV. Conclusions

In this paper, we have studied the relationship between tax policy and income inequality in Ecuador, obtaining several empirical results. First, personal income tax in Ecuador is broadly progressive, with the Kakwani and Suits progressivity indices and concentration curve analysis showing that the richest individuals generally pay more income tax than the poorest ones. When we analysed the effective tax rates paid by very high-income individuals, however, we found that these decreased within top income groups, especially within the top 0.1%, 0.05%, 0.01% and 0.001%.

Second, the redistributive capacity of personal income tax is very weak. The Reynolds-Smolensky redistribution index suggests that income inequality is only one point lower after income tax than before it, declining from 0.66 to 0.65 in 2008 and from 0.64 to 0.63 in 2010. That redistributive capacity is so slight is mainly due to personal income tax deductions, which are mostly employed by high-income individuals.

Third, results obtained from the concentration curve approach showed that the tax filers benefiting most from personal income tax deductions during the period 2008-2010 were those from the eightieth percentile of the tax database upward (corresponding to the top 5% when a control variable for total population is employed). High-income individuals are more likely to reduce their taxable income through legal income deductions, thereby eroding the tax base and cancelling out the distributive effect sought by the 2008 tax reform. The results of our static microsimulation exercise showed that the redistributive effect of personal income tax could be greater if tax deductions were better targeted and controlled.

Fourth, despite the reversal of the trend in 2010, average effective tax rates paid by high-income individuals are very low when dividends are an integral part of the personal income tax base, with the top 1% income group paying an average effective tax rate of 7% and the top 0.1% a rate of 9.4% over the entire period.

Fifth, we have used the taxable income elasticity literature to derive optimal tax rates for high-income individuals. While the current top marginal tax rate in Ecuador is 35%, our results suggest that the optimal top rate could be between 57% and 63%.

Finally, it is important to stress the vital need to close tax loopholes, notwithstanding the efforts made by the Ecuadorian Internal Revenue Service in recent years.

The new personal income tax deductions, which were introduced to improve tax equity, are too much of a blunt instrument, and policymakers should consider other ways of modifying the tax burden of low- and middle-income taxpayers. Policymakers should also consider improving the targeting of income tax deductions at the top of the income distribution.

Moreover, our findings suggest that it would be possible to create a more progressive rate structure for personal income tax in Ecuador, along with a higher top marginal tax rate. Naturally, there is the concern that increasing top marginal tax rates could affect work rates and business creation and therefore economic growth. Nevertheless, empirical research (Piketty, Saez and Stantcheva, 2014) has proved that certain countries such as the United States and the United Kingdom, where top marginal tax rates have been greatly reduced across time, have not grown faster than countries which have maintained high tax rates. Top income inequality has increased, however. Thus, greater tax progressivity in Ecuador would lead to an increase in tax revenue and public investment, notably in education, skills and health care.

Bibliography

Alvaredo, F. (2010), "The rich in Argentina over the Twentieth Century, 1932-2004", *Top Incomes: a Global Perspective*, A. B. Atkinson and T. Piketty, Oxford, Oxford University Press.

Alvaredo, F. and J. Londoño (2013), "High incomes and personal taxation in a developing economy: Colombia 1993-2010", *CEQ Working Paper Series*, No. 12, Commitment to Equity.

Amarante, V. and others (2011), "Distributive impacts of alternative tax structures. The case of Uruguay", *Working Paper*, No. 09/11, Montevideo, University of the Republic.

Atkinson, A. B. (2014), "After Piketty?", *The British Journal of Sociology*, vol. 65, No. 4, Wiley.

Atkinson, A. B. and T. Piketty (2010), *Top Incomes: a Global Perspective*, Oxford, Oxford University Press.

___(2007), *Top Incomes over the Twentieth Century. A Contrast between Continental European and English-speaking Countries*, Oxford, Oxford University Press.

Atkinson, A. B., T. Piketty and E. Saez (2011), "Top incomes in the long run of history", *Journal of Economic Literature*, vol. 49, No. 1, Nashville, Tennessee, American Economic Association.

Auten, G. and R. Carroll (1999), "The effect of income taxes on household income", *Review of Economics and Statistics*, vol. 81, No. 4, Cambridge, Massachusetts, The MIT Press.

Bourguignon, F. and A. Spadaro (2006), "Microsimulation as a tool for evaluating redistribution policies", *Journal of Economic Inequality*, vol. 4, No. 1, Springer.

Burdín, G., F. Esponda and A. Vigorito (2014), "Inequality and top incomes in Uruguay: a comparison between household surveys and income tax micro-data", *CEQ Working Paper Series*, No. 1, Commitment to Equity.

Cano, L. (2015), "Income Inequality, Top Income Shares and Economic Mobility: Ecuador 2004-2011", thesis, Toulouse, University of Toulouse 1 Capitole.

Cetrángolo, O. and J. Gómez Sabaini (2006), *Tributación en América Latina: en busca de una nueva agenda de reformas*, ECLAC Books, No. 93 (LC/G.2324-P), Santiago, Economic Commission for Latin America and the Caribbean (ECLAC).

Chetty, R. (2009), "Is the taxable income elasticity sufficient to calculate deadweight loss? The implications of evasion and avoidance", *American Economic Journal: Economic Policy*, vol. 1, No. 2, Nashville, Tennessee, American Economic Association.

Cornia, A. (ed.) (2014), *Falling Inequality in Latin America. Policy Change and Lessons*, Oxford University Press.

___(2010), "Income distribution under Latin America's new left regimes", *Journal of Human Development and Capabilities*, vol. 11, No. 1, Taylor and Francis.

Cornia, A., J. Gómez Sabaini and B. Martorano (2011), "A new fiscal pact, tax policy changes and income inequality: Latin America during the last decade", *UNU-WIDER Working Paper*, No. 70, Helsinki, World Institute for Development Economics Research (UNU-WIDER).

Diamond, P. and E. Saez (2011), "The case for a progressive tax: from basic research to policy recommendations", *Journal of Economic Perspectives*, vol. 25, No. 4, Nashville, Tennessee, American Economic Association.

ECLAC (Economic Commission for Latin America and the Caribbean) (2014), *Social Panorama of Latin America 2014* (LC/G.2635-P), Santiago.

___(2013), *Social Panorama of Latin America 2013* (LC/G.2580), Santiago.

___(2012), *Social Panorama of Latin America 2012* (LC/G.2557-P), Santiago.

Fairfield, T. (2010), *The Politics of Revenue-Raising Tax Reform in Latin America*, Berkeley, University of California.

Fairfield, T. and M. Jorratt (2016), "Top income shares, business profits, and effective tax rates in contemporary Chile", *Review of Income and Wealth*, vol. 62, No. S1, Wiley.

Feldstein, M. (1999), "Tax avoidance and the deadweight loss of the income tax", *Review of Economics and Statistics*, vol. 81, No. 4, Cambridge, Massachusetts, The MIT Press.

Figari, F., A. Paulus and H. Sutherland (2015), "Microsimulation and policy analysis", *Handbook of Income Distribution*, vol. 2, A. B. Atkinson and F. Bourguignon, Amsterdam, Elsevier.

Gasparini, L. and others (2009), "A turning point? Recent developments on inequality in Latin America and the Caribbean", *CEDLAS Working Papers*, No. 081, Universidad Nacional de la Plata.

Gómez Sabaini, J. (2006), "Cohesión social, equidad y tributación. Análisis y perspectivas para América Latina", *Políticas Sociales series*, No. 127 (LC/L.2641-P), Santiago, Economic Commission for Latin America and the Caribbean (ECLAC).

Gómez Sabaini, J. and J. Jiménez (2012), "Tax structure and tax evasion in Latin America", *Macroeconomía del Desarrollo series*, No. 118 (LC/L.3455), Santiago, Economic Commission for Latin America and the Caribbean (ECLAC).

Gómez Sabaini, J. and D. Moran (2013), "Política tributaria en América Latina: agenda para una segunda generación de reformas", *Macroeconomía del Desarrollo series*, No. 133 (LC/L.3632), Santiago, Economic Commission for Latin America and the Caribbean (ECLAC).

González, I. and R. Martner (2009), "Del síndrome del casillero vacío al desarrollo inclusivo: buscando los determinantes de la distribución del ingreso en América Latina", *Working Paper*, Santiago, Economic Commission for Latin America and the Caribbean (ECLAC).

Goñi, E., H. López and L. Servén (2011), "Fiscal redistribution and income inequality in Latin America", *World Development*, vol. 39, No. 9, Amsterdam, Elsevier.

Gruber, J. and E. Saez (2002), "The elasticity of taxable income: evidence and implications", *Journal of Public Economics*, vol. 84, No. 2, Amsterdam, Elsevier.

Hanni, M., R. Martner and A. Podestá (2015), "The redistributive potential of taxation in Latin America", *CEPAL Review*, No. 116 (LC/G.2643-P), Santiago, Economic Commission for Latin America and the Caribbean (ECLAC).

Jiménez, J. P., J. C. Gómez Sabaini and A. Podestá (2010), "Tax gap and equity in Latin America and the Caribbean", *Fiscal Studies*, No. 16, Santiago, Economic Commission for Latin America and the Caribbean (ECLAC).

Kakwani, N. (1977), "Measurement of tax progressivity: an international comparison", *The Economic Journal*, vol. 87, No. 345, Royal Economic Society.

Landais, C., T. Piketty and E. Saez (2011), *Pour une révolution fiscale: un impôt sur le revenu pour le XXIème siècle*, Paris, Seuil.

Lindsey, L. (1987), "Individual taxpayer response to tax cuts: 1982-1984, with implications for the revenue maximizing tax rate", *Journal of Public Economics*, vol. 33, No. 2, Amsterdam, Elsevier.

López-Calva, L. and N. Lustig (2010), *Declining Inequality in Latin America: A Decade of Progress*, Washington, D.C., Brookings Institution Press.

Lustig, N., L. López-Calva and E. Ortiz-Juárez (2013), "Deconstructing the decline in inequality in Latin America", *Policy Research Working Paper*, No. 6552, Washington, D.C., World Bank.

OECD/ECLAC/CIAT/IDB (Organization for Economic Cooperation and Development/Economic Commission for Latin America and the Caribbean/Inter-American Centre for Labour Administration/Inter-American Development Bank) (2015), *Revenue Statistics in Latin America and the Caribbean, 1990-2013*, Paris, OECD Publishing.

Piketty, T. (2015), "Putting distribution back at the center of economics: reflections on capital in the Twenty-First Century", *The Journal of Economic Perspectives*, vol. 29, No. 1, Nashville, Tennessee, American Economic Association.

Piketty, T. (2014), *Capital in the Twenty-First Century*, Harvard University Press.

___(2003), "Income inequality in France, 1901-1998", *Journal of Political Economy*, vol. 111, No. 5, Chicago, The University of Chicago Press.

___(2001), *Les hauts revenus en France au XXe siècle: inégalités et redistributions, 1901-1998*, Paris, Bernard Grasset.

Piketty, T. and E. Saez (2003), "Income inequality in the United States, 1993-1998", *Quarterly Journal of Economics*, vol. 118, No. 1, Oxford University Press.

Piketty, T., E. Saez e I. Stantcheva (2014), "Optimal taxation of top labor incomes: a tale of three elasticities", *American Economic Journal*, vol. 6, No. 1, Nashville, Tennessee, American Economic Association.

Poterba, J. (2007), "Income inequality and income taxation", *Journal of Policy Modeling*, vol. 29, No. 4, Amsterdam, Elsevier.

Rawlings, L. and G. Rubio (2005), "Evaluating the impact of conditional cash transfer programs", *The World Bank Economic Review*, vol. 20, No. 1, Oxford, Oxford University Press.

Reynolds, M. and E. Smolensky (1977), *Public Expenditures, Taxes and the Distribution of Income: the United States 1950-1961*, New York, Academic Press.

Riihela, M., R. Sullstrom and M. Tuomala (2014), "Top incomes and top tax rates: implications for optimal taxation of top incomes in Finland", *Tampere Economic Working Papers Net Series*, No. 88, University of Tampere.

Roca, J. (2009), "Tributación directa en Ecuador: evasión, equidad y desafíos de diseño", *Macroeconomía del Desarrollo series*, No. 85 (LC/L.3057-P), Santiago, Economic Commission for Latin America and the Caribbean (ECLAC).

Saez, E. (2001), "Using elasticities to derive optimal income tax rates", *The Review of Economic Studies*, vol. 68, No. 1, Oxford, Oxford University Press.

Saez, E., J. Slemrod and S. Giertz (2012), "The elasticity of taxable income with respect to marginal tax rates: a critical review", *Journal of Economic Literature*, vol. 50, No. 1, Nashville, Tennessee, American Economic Association.

Saez, E. and M.R. Veall (2005), "The evolution of high incomes in Northern America: lessons from Canadian evidence", *The American Economic Review*, vol. 95, No. 3, Nashville, Tennessee, American Economic Association.

Spadaro, A. (2007), *Microsimulation as a Tool for the Evaluation of Public Policies. Methods and Applications*, BBVA Foundation.

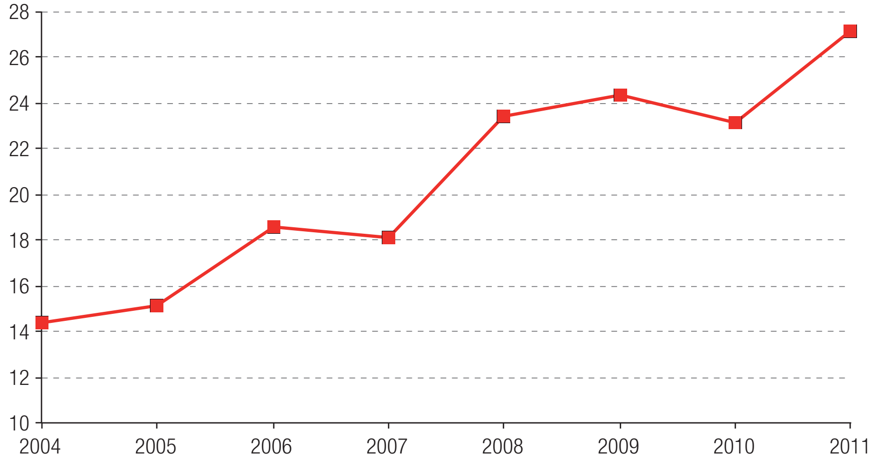
Suits, D. (1977), "Measurement of tax progressivity", *American Economic Review*, vol. 67, No. 4, Nashville, Tennessee, American Economic Association.

Tanzi, V. (2013), *Tax Reform in Latin America: a Long Term Assessment*", Washington, D.C., Woodrow Wilson International for Scholars.

Van Kerm, P. and A. Peichl (2007), "Progress: Stata module to measure distributive effects of an income tax", *Statistical Software Components*, No. S456867, Boston College.

Annex A1

Figure A1.1
Ecuador: tax filers as a proportion of total tax units (adults aged 20 and over), 2004-2011
(Percentages)



Source: Prepared by the author, on the basis of the individual income tax returns database of the Ecuadorian Internal Revenue Service (SRI) and population estimates provided by the National Institute of Statistics and Censuses (INEC).

Analysis of formal-informal transitions in the Ecuadorian labour market

Adriana Patricia Vega Núñez

Abstract

The study analyses the transitions between the formal and informal labour markets using longitudinal data for Ecuador. First, we use transition matrices to characterize the short-run dynamics among the different labour statuses in Ecuador. Next, multinomial logit models are used to identify the factors that determine the probability of remaining in or moving across the formal and informal sectors. Education level, years of experience and wage differentials by sector have a significant effect on worker transitions, showing that benefits and costs vary depending on the individual's preferences and skills.

Keywords

Employment, labour market, informal sector, labour mobility, mathematical analysis, employment statistics, Ecuador

JEL classification

O170, R230, J420

Author

Adriana Patricia Vega Núñez is a doctoral candidate in the Department of Econometrics, Statistics and Applied Economics at the University of Barcelona, Spain. Email: adriana.vega.nunez@gmail.com.

I. Introduction

Widespread informality is a characteristic of developing and transition economies.[1] A diverse range of people are involved in informal economic activities in these countries and the effects on them are manifold. This persistent phenomenon remains a major challenge in the many countries seeking to reduce it.[2] Charmes (2009) shows that informal employment has existed since it was first defined in the mid-1970s. Until the 2000s, informal employment trended upwards in several regions around the world.[3] On average, over the last 30 years, informal employment has accounted for more than 47% of total non-agricultural employment in developing regions, approximately 24% in transition economies and more than 50%, on average, in Latin America.

While researchers' definitions of the term vary, informality is usually associated with negative phenomena: unprotected workers, tax evasion, illegal activities, low productivity, low investment rates and so on. However, different schools of thought have emerged in the literature on the causes of informal employment. The main question is whether individuals or businesses exit the formal sector voluntarily, or whether they are excluded from it. There are two dominant schools of thought discussed in the World Bank publication, *Informality: Exit and Exclusion* (Perry and others, 2007).

The exclusionary view, which reflects the traditional thinking, points to a segmented market where people choose informal employment because it is the only alternative (De Soto, 2000). In that sense, informal workers would prefer formal jobs that offer higher wages and labour protection.

By contrast, the concept of the voluntary informal sector is diametrically opposed, in that workers choose from among the various jobs on offer and join the informal sector willingly. Maloney (2004) supports this, drawing on evidence from Latin America to posit that the informal sector is an unregulated micro-entrepreneurial sector and not a disadvantaged residual of segmented labour markets. Furthermore, Bosch and Maloney (2005) suggest that a considerable portion of workers in the informal sector, particularly the self-employed, enter the sector voluntarily. Accordingly, many forms of informal employment can offer desirable advantages, such as independence or training opportunities.

The aim of this study is to examine labour market dynamics across the different employment statuses and observe the characteristics that determine the probability of transition or continuance across the formal and informal sectors in Ecuador. The empirical analysis is based on the Ecuadorian Survey of Urban Employment, Unemployment and Underemployment (ENEMDU) for the period 2011-2012. The study was done by examining the flows across the following labour sectors: formal, informal, unemployed and outside the labour force. It contributes to the existing literature in two ways: first, the panel survey used, which tracks both formal and informal workers over two consecutive years, allows us to identify changes in the labour force over time; second, the study adds to the scarce data available for Ecuador. To this end, we created a transition matrix to map the dynamics of the market and interactions across the sectors. Additionally, since the transition matrix does not reflect the observable characteristics of workers, a multinomial logit analysis was used to identify these characteristics and determine the probability of choosing each labour sector.

[1] Irrespective of the methods for measuring informality, it is high in Latin America (Perry and others, 2007). Informality has increased over time in Sub-Saharan Africa and Asia (Jütting and others, 2008).

[2] According to Baccheta, Ernst and Bustamante (2009), on the basis of broad definitions of informality, countries in Africa, Asia and Latin America present persistent informality rates.

[3] There has been a rising trend in informal employment in Sub-Saharan Africa, South and East Asia, Latin America, West Asia, North Africa and transition countries (Charmes, 2009).

The article is divided into five sections. Following this introduction, section II provides a brief summary of empirical literature on the dynamics in the formal and informal labour market. Section III presents the data and definitions of informality and other main variables, as well as the econometric methodology and models used in the study. Section IV contains the results of the study. Lastly, in section V, we discuss the main findings and present concluding remarks.

II. Literature review

As mentioned previously, two perspectives are used to explain the link between informality and segmentation in the labour market. The first assumes that workers would prefer formal wage employment, but are forced into the informal sector by restrictions and limited availability of formal jobs. In the model devised by Harris and Todaro (1970), setting a minimum wage above the equilibrium wage resulted in a limitation of formal jobs, leading to labour market segmentation.

In line with the second viewpoint, Maloney (1999) regards the formal and informal sectors as an integrated market in which workers choose their occupations from among the different jobs available on the basis of their preferences, abilities and needs. Thus, workers who prefer informal to formal employment do so because the former offers more desirable characteristics.

These two schools of thought examine the diverse flows in the labour market and the corresponding sectoral wage differentials in different ways. In a segmented market, flows from informal to formal jobs should exceed reverse flows. By comparison, in an integrated market, the flows between formal and informal jobs should occur in both directions and be similar in volume (Fields, 2009).

Recent evidence points to significant labour market mobility in emerging countries. The movement of workers across jobs, from unemployed to employed and into or out of the labour force indicate the mobility in a sector. In this regard, Maloney (1999) analysed worker transitions between sectors using panel data from Mexico, finding that patterns of labour market mobility indicate that much of the informal sector is a desirable destination and that the distinct modalities of work, formal and informal, are well integrated. Duryea and others (2006) examined evidence on worker flows across sectors in the labour market in three Latin American countries: Argentina, Mexico and the Bolivarian Republic of Venezuela. The authors found high mobility not only into and out of the labour market, but also across different types of jobs. They also found that, on average, workers who moved from formal wage to informal wage employment experienced a decline in earnings, while switching from informal to formal wage jobs produced the opposite effect. Cea and Contreras (2008) used panel data to provide evidence for Chile. The results revealed a strong tendency towards continuance in individuals' labour status. The study also suggests that age, schooling, and non-labour income are significant determinants of the probability of having a particular employment status. In a study on Argentina, Jiménez (2011) found evidence supporting segmentation of the formal sector of the labour market. In this case, one particular group of workers —unregistered wage earners— remained under unfavourable working conditions.

In their analysis of panel data from Ukraine, Lehmann and Pignatti (2008) found evidence of segmentation in the labour market such that moves by informal wage workers are largely involuntary. Slonimczyk and Gimpelson (2013), using a multinomial logit model and allowing for individual heterogeneity in preferences, point to the existence of an integrated labour market in the Russian Federation.

In general, informality is a prominent feature of transition and emerging economies, as is the mobility of workers across labour market sectors. These topics have been the subject of extensive research worldwide over the years, with diverse results. However, evidence from Ecuador is limited, as previous studies of the Ecuadorian labour market have mainly focused on wage distribution as opposed to effects on employment or transitions between the formal and informal sector. A recent study by

Canelas (2014) examines whether changes in the minimum wage have an impact on formality and informality rates and wage levels in Ecuador. The results suggest that changes in the minimum wage do not affect employment rates and wages.

III. Data and methodology

The source of the data used in this study is the national survey (ENEMDU), a rotating household panel survey conducted by the National Statistics and Census Institute of Ecuador (INEC). The panel does not follow individuals continuously, but is constructed from four reports, spread over two consecutive years. Households are interviewed in two consecutive quarters, replaced by a new sample unit in the next two consecutive quarters and then returned one final time to the sample in two successive quarters. This national survey includes both urban and rural populations.

The panel analysed is the fourth quarter of 2011 to the fourth quarter of 2012. The analysis uses only one year's data because the results obtained from the separate transition tables calculated for previous years were quite similar to those in the panel evaluated in this study (see annex tables A1.1 and A1.2). The weights used reflect the expansion factors specified by INEC. The analysis includes workers aged 15 years and older.

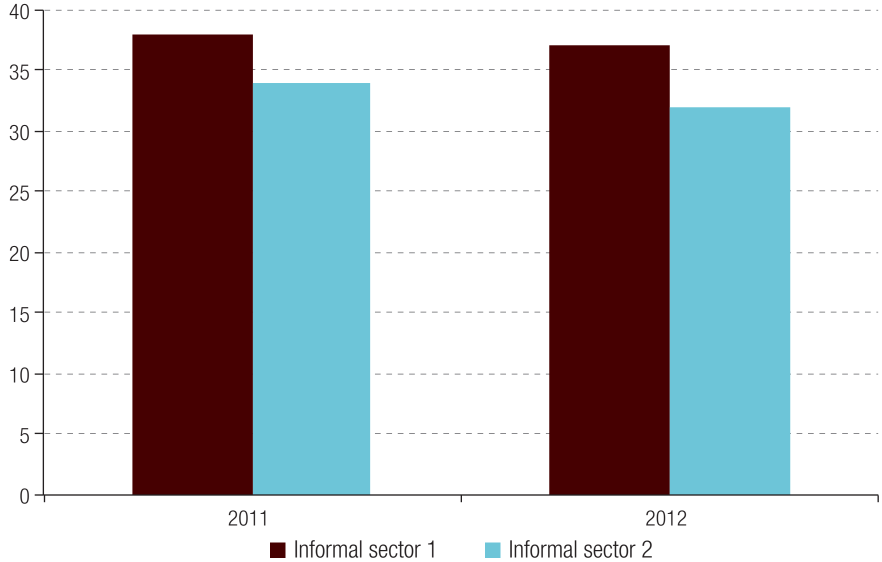
In order to identify the different labour statuses that are used in the analysis, we can split the labour force into three major categories: employed, unemployed and outside the labour force. In addition, the employed category was divided into two further subdivisions: individuals in the formal sector and individuals in the informal sector, which included wage workers (in private companies or domestic work) and independent (self-employed) workers. We thus considered five different statuses in the labour market: (i) formal waged, (ii) informal waged, (iii) self-employed, (iv) unemployed and (v) outside the labour force. These divisions are made in the working-age population and are mutually exclusive. Individuals are classified as unemployed if they did not work in the reference week but had searched for a job. Outside the labour force refers to individuals who are not working or seeking some form of employment.

There are diverse definitions of informality, based on various criteria such as the size of production units or firms, or the number of employees and the type of employment relationship. The International Labour Organization (2013) defines the informal sector as a set of units engaged in the production of goods or services with the primary objective of generating employment and incomes for the persons concerned. These units typically operate at a low level of organization and on a small scale. Labour relations are based mostly on casual employment, kinship or personal and social relations rather than contractual arrangements with formal guarantees. In this context, and considering the data from the national survey, two definitions of the informal sector are used in this study. The first is related to the characteristics of the enterprise and encompasses wage employment in small, unregistered establishments with less than 10 employees, plus all independent and own-account work. The second is based on social security coverage: workers are considered to be in the informal sector if they are not covered by social security.

Figure 1 plots the informality rate based on the two above-mentioned definitions for the observation period 2011-2012. As the figure shows, the share of informal workers in the Ecuadorian labour force was between 32% and 38% in the observation period. In general, although Ecuador's informality rate may be considered high, it is within the average for informal employment in most Latin American countries.[4]

[4] Between the 1970s and the 2000s, on average, informal employment accounted for more than 47% of total non-agricultural employment in West Asia and in North Africa, more than 70% in sub-Saharan Africa, more than 50% in Latin America, nearly 70% in South and South-East Asia and 24% in transition economies (Charmes, 2009).

Figure 1
Ecuador: informal workers as a proportion of the total labour force, 2011-2012
(Percentages)



Source: Author's calculations, on the basis of the Ecuadorian Survey of Urban Employment, Unemployment and Underemployment, 2011-2012.
Note: Informal sector 1 is made up of wage workers employed in small unregistered establishments with less than 10 employees, plus all independent and self-employed workers. Informal sector 2 is made up of workers who have no social security coverage.

Considering the five labour statuses, 20 different transitions were identified to represent the flows across labour market sectors during the period analysed. We also included five states of continuance which indicate no change in labour status from the previous period. The analysis concentrates primarily on the movements from the formal and the informal sector (wage employment and self-employment) to each of the other labour statuses, in order to identify the characteristics of workers making these transitions. Annex table A1.3 shows the mean values for the sample using the first definition of informality, respectively, for age, level of schooling, work experience and the initial real wage for workers transitioning between or remaining in sectors. The initial and final real wages denote the monthly real wage of workers for December 2011 and December 2012, respectively (treating earnings from the main job as the only source of labour income). Finally, the earnings differential corresponds to the difference between the final and the initial real wage for workers moving to or remaining in a labour sector.

In the rest of this section, we describe the econometric techniques used in the study to provide empirical evidence of the mobility patterns across the aforementioned sectors of the Ecuadorian labour market. To identify the directions and volume of the various labour market flows, two methods were used: transition matrices and a multinomial logit analysis of movements between sectors.

1. Transition matrices

The transition matrices allow us to determine the flows of workers between the selected labour sectors by calculating the conditional probability of finding a worker in state j at the end of the period, given that the worker began in state i; this is denoted by P_{ij}. The sum of each row of the transition matrix is equal to 100% and the totals at the end of the columns and rows represent the share of workers in each category at the end of the period P_i and P_j. The components in the main diagonal reveal the share of workers who remained in the same labour category at the end of the period. The information provided by the transition matrices gives us a first idea of the different movements of workers between sectors.

2. Multinomial logit analysis of transitions between labour sectors

Since the previous methodology is merely descriptive, we also use a multinomial logit analysis to determine which characteristics affect workers' probability of choosing to move to sector j relative to the probability of staying in sector i. The multinomial logit model is a probabilistic discrete model which can explain tendencies to move or remain in the different sectors of the labour market. We can find the effect for each worker characteristic on the probability of transition into a sector. We model flows among five different labour market states: formal waged (j=1), informal waged (j=2), self-employed (j=3), unemployed (j=4), and outside the labour force (j=5).

We use the standard exponential form for the multinomial logit analysis:

$$P_{ij} = \frac{exp\left(X_i \beta_j\right)}{\Sigma_{l=1}^{m} exp\left(X_i \beta_j\right)} \tag{1}$$

where the vector β_j measures the degree to which an increase in worker characteristic X_i increases the probability of a worker moving to state j relative to the probability of staying in state i. The worker characteristics are age, gender, marital status, level of schooling, years of experience, region and earnings differential logarithms. We can calculate the probability of transition where the explanatory variables determine the increase or decrease in probabilities.

IV. Results

1. Transition matrices

In the first analysis with transition matrices, we describe labour mobility by calculating the conditional probabilities of finding a worker in state j at the end of the period, conditional on being in state i initially. We therefore obtain a 5x5 annual matrix for the period analysed, December 2011 to December 2012 (see table 1). Since the purpose of this study is to identify the patterns of worker transitions out of and into the formal and informal sectors, we focus mainly on the transition from the formal sector to each of the other labour statuses and from the informal sector (both waged and self-employed) to the other sectors. The matrix also reveals some facts about the dynamics of the market and the relationships between the sectors defined.

Table 1
Ecuador: worker mobility across sectors of the labour market, 2011-2012
(Percentages)

	Final status						
Initial status	Formal waged	Informal waged	Informal self-employed	Unemployed	Outside the labour force	Total	P_i
Formal waged	79	6	6	2	6	100	25
Informal waged	21	46	18	4	12	100	10
Informal self-employed	8	7	67	1	16	100	25
Unemployed	23	11	25	19	22	100	4
Outside the labour force	4	4	8	3	81	100	36
P_j	26	10	24	3	37		

Source: Author's calculations on the basis of the Ecuadorian Survey of Urban Employment, Unemployment and Underemployment, 2011-2012.
Note: P_i. is the relative size of a sector in the initial period; P_j is the relative size of a sector in the final period.

With regard to the movements of workers initially in the formal sector, we found a 79% continuance rate. Interestingly, the largest outflow from this sector is towards the informal category (12% in total), in two distinct flows. The first is the movement into the informal wage sector (6%) and reveals the deterioration in job conditions for workers in terms of formality. This particular type of labour transition could be explained by the rationalization of formal employment opportunities (Fields, 1972; Perry and others, 2007). The second is toward informal self-employment (6%), which may be associated with voluntary decisions of individuals seeking more autonomy or the result of inefficiencies in formal sector protections and low levels of labour productivity (Maloney, 1999). Additionally, formal sector workers leaving their jobs could transition to unemployment (2%) or out of the labour force (6%).

Moves by informal wage workers paint a different picture. First, the probability of staying in this sector is around 46%, meaning that these individuals prefer to seek other employment opportunities rather than remain in these jobs. If we look at the outflows from this segment of the informal sector, we find that 21% of workers obtained better quality jobs by switching to the formal sector. This type of mobility may be construed as workers improving their job conditions, given that workers in the informal sector typically earn less than similar workers in the formal sector (Günther and Launov, 2006). The other subdivision of the informal sector accounts for around 18% of outflows, highlighting the fact that wage workers in the informal sector seek not only formal-sector employment, but also self-employment. There are characteristics of the formal sector and self-employment —including, for the former, labour conditions and, the latter, independence— that make them somewhat more attractive than informal wage employment. Therefore, the probability of transition from informal wage employment to either formal or self-employment is higher than the probability of transition to unemployment. The flow from informal wage employment to outside the labour force is 12%.

The third row of table 1 illustrates the transitions from self-employment. A first observation is that the continuance rate in this sector shows that a significant proportion of individuals remain self-employed. A second is that the flows from self-employment to both the formal wage sector and the informal wage sector are quite similar (8% and 7%, respectively), while transition to unemployment is 1%. Mandelman and Montes-Rojas (2009) posit two possible reasons for transitions to the self-employed sector: the sector attracts individuals with outstanding entrepreneurial abilities, or it serves as a refuge for the unemployed. Either of these could explain the substantial continuance rate in this sector and the similar flows to the formal sector and to the informal wage sector.

We observed high mobility out of the unemployed sector, as only 19% of individuals remained unemployed during the observation period. Most of the unemployed transitioned to self-employment, followed by the formal sector and informal wage employment, at rates of 25%, 23%, and 11%, respectively.

Most of those who found jobs did so in the informal sector (wage and self-employment), which may be attributed to insufficient human capital or other individual worker characteristics or preferences. In the traditional view of unemployment as the lowest state in the labour market, individuals in this situation may have difficulty in entering the formal sector and thus be more likely to be propelled into accepting employment with inferior labour conditions and lower earnings. Jütting and others (2008) argue that it is also possible that working in the informal sector provides more advantages than the formal sector. After all, informality cannot be considered the last resort of a worker. Workers may voluntarily choose the informal sector, where they might have the opportunity to accumulate experience or training, particularly in the case of young low-skilled workers or unskilled older individuals, or find greater flexibility and autonomy. The World Bank (2012) highlights that entrepreneurs value the independence and flexibility of having their own business and may see being autonomous as the main reason to do so.

Examination of the outside the labour force status shows a high continuance rate. Transitions into the self-employment, formal sector, informal wage and unemployed categories represented just 8%, 4%, 4% and 3%, respectively. Here again, the total outflows to the informal sector are higher than those to the formal sector. In this context, it is apparent that the informal sector has some attractive characteristics for workers.

2. Multinomial logit analysis

The main objective of this section is to obtain an overview of the labour force dynamics in Ecuador. Thus, the informal sector, comprising both wage employment and self-employment, is the largest source of employment. As we mentioned before, this segment of the labour force is not necessarily inferior to the formal sector, given the significant amount of worker transitions from formal to informal employment.

The multinomial logit analysis of movements between labour sectors allows us to determine, in statistical terms, the likelihood of workers to move from the initial sector to another sector, on the basis of their specific characteristics (see tables 2 and 3).

Table 2
Ecuador: multinomial logit analysis of transitions out of the formal sector, 2011-2012
(Estimated coefficients)

Transitions / Variables	From the formal sector to			
	Informal waged	Informal self-employed	Unemployed	Out of the labour force
Constant	1.391	-2.305	-3.387	-1.548
	(3.15)***	(5.07)***	(2.90)***	(2.48)**
Age	-0.022	0.026	-0.018	0.006
	(2.56)**	(3.52)***	(0.90)	(0.52)
Male[a]	-0.144	0.026	-0.014	0.811
	(0.78)	(0.15)	(0.04)	(3.42)***
Married[a]	-0.365	0.206	-1.003	-0.469
	(1.93)*	(1.20)	(1.89)**	(1.72)*
Schooling	-0.397	-0.153	-0.277	-0.441
	(8.93)***	(3.69)***	(2.66)***	(7.02)***
Experience	-0.005	-0.042	-0.077	-0.008
	(0.40)	(4.11)***	(1.93)**	(0.56)
Coast	-0.050	0.204	1.522	-0.436
	(0.23)	(0.86)	(1.97)***	(1.42)
Centre	-0.069	0.388	1.503	-0.618
	(0.22)	(1.34)	(1.68)**	(1.51)
South	-0.257	0.318	1.221	0.227
	(0.87)	(1.13)	(1.33)	(0.59)
Log. Diff. earnings	-0.017	-0.176	-0.778	-0.887
	(0.30)	(4.13)***	(10.17)***	(17.91)***

Source: Author's calculations, on the basis of the Ecuadorian Survey of Urban Employment, Unemployment and Underemployment, 2011-2012.

Note: The coefficients reflect how the different worker characteristics and the percentage change in real wage affect the probability of moving from the initial sector to the final sector relative to the probability of staying in the initial sector. The informal sector is made up of wage workers employed in small, unregistered establishments with less than 10 employees, plus all independent and self-employed workers. Z-statistics are in parentheses; *p<0.1, **p<0.05, ***p<0.01.

[a] These are dummy variables; default categories are: male, marital status excluding married, and north region.

Table 3
Ecuador: multinomial logit analysis of transitions out of the informal sector, 2011-2012
(Estimated coefficients)

Variables \ Transitions	From informal waged to				From informal self-employed to			
	Formal waged	Informal self -employed	Unemployed	Out of the labour force	Formal waged	Informal waged	Unemployed	Out of the labour force
Constant	-1.774	-3.022	-5.078	-5.392	-3.084	0.404	-3.691	-2.908
	(3.50)***	(5.65)***	(4.21)***	(5.37)***	(6.61)***	(0.88)	(3.17)***	(6.04)***
Age	-0.004	0.034	0.004	0.024	-0.016	-0.027	-0.061	0.004
	(0.43)	(4.13)***	(0.23)	(1.63)	(2.15)**	(3.95)***	(3.21)***	(0.62)
Male[a]	-0.334	-0.154	-0.594	0.744	-0.436	-0.589	-0.346	1.078
	(1.68)**	(0.77)	(1.28)	(2.07)**	(2.67)***	(3.52)***	(0.87)	(6.44)***
Married[a]	0.063	0.022	-0.848	-0.932	0.464	-0.411	-0.984	-0.391
	(0.29)	(0.11)	(1.58)	(2.30)**	(2.70)***	(2.40)**	(1.95)**	(2.45)**
Schooling	0.295	0.157	0.246	0.168	0.303	-0.150	0.319	-0.064
	(5.21)***	(2.73)***	(1.91)**	(1.63)	(7.35)***	(3.23)***	(2.92)***	(1.49)
Experience	-0.018	-0.005	0.004	-0.011	-0.005	-0.001	-0.104	-0.017
	(1.47)	(0.48)	(0.19)	(0.59)	(0.53)	(0.15)	(02.48)**	(2.41)**
Coast	-0.565	0.187	0.135	-0.037	-0.332	-0.165	0.577	0.430
	(2.39)**	(0.75)	(0.24)	(0.08)	(1.52)**	(0.75)	(0.87)	(1.93)**
Centre	0.095	-0.086	0.334	-0.270	-0.040	-0.168	0.325	-0.164
	(0.32)	(0.26)	(0.47)	(0.47)	(0.15)	(0.59)	(0.37)	(0.56)
South	-0.499	-0.466	0.167	0.099	-0.132	-0.110	1.318	0.406
	(1.64)	(1.33)	(0.22)	(0.16)	(0.50)	(0.39)	(1.74)**	(1.45)
Log. Diff. earnings	0.034	-0.283	-0.942	-1.146	0.126	0.174	-0.636	-0.661
	(0.44)	(4.28)***	(9.89)***	(13.22)***	(3.42)***	(4.38)***	(7.69)***	(19.63)***

Source: Author's calculations on the basis of the Ecuadorian Survey of Urban Employment, Unemployment and Underemployment, 2011-2012.

Note: The coefficients reflect how the different worker characteristics and the percentage change in real wage affect the probability of moving from the initial sector to the final sector relative to the probability of staying in the initial sector. The informal sector is made up of wage workers employed in small, unregistered establishments with less than 10 employees, plus all independent and self-employed workers. Z-statistics are in parentheses; *p<0.1, **p<0.05, ***p<0.01.

[a] Dummy variables; default categories are: male, marital status excluding married, and north region.

The results obtained show that workers become less likely to move from formal status to informal status as their level of education increases. On average, workers who transitioned to informal wage employment and self-employment had completed six and seven years of studies, respectively.

With regard to experience, the probability of moving into the informal sector, specifically to self-employment, decreases as the worker's experience increases. It is worth noting that the mean years of experience for individuals who started their own business is 8, which represents a considerable number of years of training. Aroca and Maloney (1998) found that the informal self-employed sector is a desirable destination for workers, but it requires an accumulation of financial and human capital. Thus, the mean years of experience suggests that workers first accumulate savings and knowledge, which can then be used to start a business. The characteristics of each sector of the labour market and the specificities of a job determine the differences in wages between sectors. Tansel and Oznur (2012) examine whether informal workers are paid less than similar workers in the formal sector. The authors found that unobserved fixed effects combined with observable worker characteristics explained the pay differentials between formal and informal employment.

Thus, unobserved characteristics such as personal production capacity, character traits and management quality could affect workers' productivity and, therefore, the wage differential. As expected, the logit results show that the probability of moving from formal employment to self-employment decreases as the percentage difference, in real wages, between the initial sector and final sector rises. The sample mean difference in real earnings showed that the individuals who moved in this direction earned less. Thus, if this difference increases, workers would be less likely to move into the informal sector.

Similarly, we follow the transitions between formal employment and the two statuses in which there is no participation in the labour market: unemployed and out of the labour force. In both cases, the probability of these transitions decreases as the level of education increases. It is also important to note that the probability of workers moving from the formal sector to unemployment decreases as the years of experience increase.

Moving to the second multinomial logit analysis, an interesting pattern arises in the flows from informal wage employment to the formal sector. If we consider that the number of workers with higher levels of education raises the mean years of schooling in the formal sector, annex table A1.3 shows that the mean years of schooling of individuals who remained as informal wage workers is 5, while that of individuals who moved to the formal status is 6. The results suggest that individuals with higher levels of schooling are more likely to enter formal sector employment. This correlation may imply that workers start out in the informal sector —seeing it as an employment option while they continue their education— and after raising their level of instruction and skills, they seek better labour conditions in the formal sector.

The findings also show that education has a positive effect on the transitions from informal wage employment to self-employment, in other words, movements within the informal sector. However, the percentage difference of the real wage between the initial and final sector has a negative impact on this flow.

The third section of annex table A1.3 presents the transitions from self-employment to the other labour sectors. The probability of moving from self-employment to the formal sector is associated with two variables: education and the percentage difference in real wages. Regarding the first, individuals with higher levels of education tend to move to better jobs in the formal sector. The mean differences in the real wage of the workers who moved to the formal sector is positive and significant, thus if this difference increases, so will the probability of moving to the formal sector.

The results suggest that individuals with higher levels of education are less likely to move from self-employment to informal wage employment; indicating that workers prefer the autonomy that the latter probably would not offer. The difference in real wage in this instance is positive.

Lastly, with regard to the transition from the informal sector (both waged and self-employed) to unemployed and out of the labour force, the main finding is that the percentage real-wage differential negatively influences these two flows.

In order to have a better idea of the economic significance of the effect of observable worker characteristics on the different transitions across the labour sectors, we use the two multinomial logit models to run simulations, assigning specific worker characteristics.[5] We first analyse the effect of years of experience on the probability of moving from the formal sector to other sectors and the probability of remaining in the same sector. For this estimation, we set the age at 38 years and education level at secondary school and we then compared the probabilities of transition or continuance for both men and women (annex figures A1.1A-A1.1E). For both sexes, more years of experience led to a significant increase in the probability of continuance in formal wage employment. The transition probabilities show that exit rates to the informal and unemployed sectors decline with years of experience.

[5] Simulation results for other characteristics are omitted to save space, but are available from the author on request.

We also analysed the effect of different levels of education for the second and third logit models. Here, we again set the age at 38 years and assigned 10 years of experience (see annex figures A1.2A to A1.2I). Interestingly, as the level of education increases, so does the probability of moving from the informal sector (both wage employment and self-employment) to the formal sector. Analysis of the probability of continuance as an informal wage worker shows that this decreases as the level of education rises. Using the same worker characteristics, the probability of transition from wage employment to self-employment, unemployment and out of the labour force increases as the level of education increases. Given the above-mentioned worker characteristics, the pattern in the probability of continuance in self-employment is interesting: this probability increases as workers attain secondary education then declines thereafter. Lastly, the probability of transition from self-employment to wage employment or out of the labour force decreases as the level of education increases.

V. Concluding remarks

Informality is a widespread phenomenon which involves a significant share of the labour force in many developing and transition economies. The study provides an overview of the dynamics of the formal and informal sectors and some specific transition patterns within the labour sectors. We specify a transition matrix and a multinomial logit model to identify the movements across the sectors and the effect of each worker characteristic on the probability of transition to or from a sector.

The results observed in the transition matrix suggest dynamic flows not only in and out of the labour market, but also across the sectors. Nonetheless, these results can be considered to be within the normal percentage range of movements found in similar analyses of labour mobility in Latin America. The sectoral flows involving employed individuals suggest that job opportunities are sought in both the formal and informal sectors, as illustrated by flows in the transition matrix towards the formal sector, informal wage employment and self-employment.

The multinomial logit analysis was applied in view of the fact that transition matrices do not consider observable characteristics that can affect workers' choice of sector. This approach indicated that education, years of experience, and other characteristics influence the type of employment selected and, consequently, the transitions to or continuance in the various sectors.

The findings presented in the study point to significant interaction between the formal and informal sectors. The patterns of mobility imply that informal employment should be viewed as a desirable destination, as is the formal sector. The three employment statuses present various advantages and disadvantages, depending on the individual's preferences and abilities.

Lastly, the possible labour market flows from formal employment to informal wage and self-employment have important policy implications: policies should vary depending on whether individuals are in formal or informal employment. Consequently, it may be worthwhile to question the importance and effectiveness of existing labour market, productivity and social protection policies. If informal workers enter this sector involuntarily, policymakers must focus on aspects such as wage rigidities and be observant of companies' social protection obligations. Another important issue is the impact of informal employment on the economy: while informality is favourable from an individual standpoint, a certain formalization of the economy is necessary from a societal perspective. In most cases, entry into the informal sector is the optimal decision taken by workers based on their preferences, level of education and the restrictions or limitations of the formal sector. However, a large informal workforce is not necessarily the best option for the society as a whole, since formalization is the easiest way to organize the labour force and the associated legal framework.

In the case of voluntary entry into the informal sector, policymakers must be aware of the heterogeneity and complexities of jobs and working conditions within the informal sector and take into account the inefficiencies in labour legislation and low levels of formal sector productivity.[6] By this reckoning, workers who are in the informal sector voluntarily should find ways to substitute the protection offered by formal institutions; this would call for an in-depth assessment of the quality of various jobs rather than legal protections. It is often thought self-evident that, given the low levels of human capital in the informal sector, improving the accessibility and quality of education will increase worker productivity and pay in the formal sector. However, when choosing the sector in which they prefer to work, individuals take not only their potential earnings into account, but also all other elements and advantages associated with each job or sector.

Bibliography

Aroca, P. and W. Maloney (1998), "Logit analysis in a rotating panel context and an application to self-employment decisions", *Policy Research Working Paper*, No. 2069, Washington, D.C., World Bank.

Bacheta, M., E. Ernst and J. Bustamante (2009), "Globalization and informality in time of crisis", *Globalization and Informal Jobs in Developing Countries*, Geneva, International Labour Organization (ILO)/World Trade Organization (WTO).

Bosch, M. and W. Maloney (2005), "Labor market dynamics in developing countries. Comparative analysis using continuous time Markov processes", *Policy Research Working Paper*, No. 3583, Washington, D.C.

Canelas, C. (2014), "Minimum wage and informality in Ecuador", *Wider Working Paper*, No. 2014/006, Helsinki, World Institute for Development Economics Research (UNU-WIDER).

Cea, S. and M. I. Contreras (2008), "Transiciones laborales: evidencia para datos de panel", thesis, Santiago, University of Chile.

Charmes, J. (2009), "Concepts, measurement and trends", *Is Informal Normal? Towards More and Better Jobs in Developing Countries*, J. Jütting and J. Laiglesia (eds.), Paris, Organization for Economic Cooperation and Development (OECD).

Duryea, S. and othes (2006), "For better or for worse? Job and earnings mobility in nine middle- and low-income countries", *Brookings Trade Forum. Global Labor Markets?*, Washington, D.C., Brookings Institution Press.

Fields, G. (2009), "Segmented labor market models in developing countries", *The Oxford Handbook of Philosophy of Economics*, D. Ross and H. Kincaid (eds.), Oxford University Press.

___(1972), "Rural-urban migration, urban unemployment and underemployment, and job-search activity in LDC's", *Journal of Development Economics*, vol. 2, No. 2, Amsterdam, Elsevier.

Günther, I. and A. Launov (2006), "Competitive and segmented informal labor markets", *IZA Discussion Papers*, No. 2349, Bonn, Institute for the Study of Labor (IZA).

Harris, J. and M. Todaro (1970), "Migration, unemployment and development: a two-sector analysis", *American Economic Review*, vol. 60, No. 1, Nashville, Tennessee, American Economic Association.

ILO (International Labour Organization) (2013), *Measuring Informality: A Statistical Manual on the Informal Sector and Informal Employment*, Geneva.

Jiménez, M. (2011), "La economía informal y el mercado laboral en la Argentina: un análisis desde la perspectiva del trabajo decente", *Working Paper*, No. 116, La Plata, Center for Distributive, Labor and Social Studies (CEDLAS).

Jütting, J., J. Parlevliet and T. Xenogiani (2008), "Informal employment re-loaded", *IDS Bulletin*, vol. 39, No. 2, Institute of Development Studies.

Lehmann, H. and N. Pignatti (2008), "Informal employment relationships and labor market segmentation in transition economies: evidence from Ukraine", *IZA Discussion Paper*, No. 3269, Bonn, Institute for the Study of Labor (IZA).

Maloney, W. (2004), "Informality revisited", *World Development*, vol. 32, No. 7, Amsterdam, Elsevier.

6 See Maloney (2004).

___(1999), "Does informality imply segmentation in urban labor markets? Evidence from sectorial transitions in Mexico", *The World Bank Economic Review*, vol. 13, No. 2, Oxford University Press.

Mandelman, F. and G. Montes-Rojas (2009), "Is self-employment and micro-entrepreneurship a desired outcome?", *World Development*, vol. 37, No. 12, Amsterdam, Elsevier.

Perry, G. and others (2007), "The *razón de ser* of the informal sector", *Informality: Exit and Exclusion*, Washington, D.C., World Bank.

Slonimczyk, F. and V. Gimpelson (2013), "Informality and mobility: evidence from Russian panel data", *IZA Discussion Paper*, No. 7703, Bonn, Institute for the Study of Labor (IZA).

Soto, H. de (2000), *The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else*, New York, Basic Books.

Tansel, A. and E. Oznur (2012), "The formal/informal employment earnings gap: evidence from Turkey", *IZA Discussion Paper*, No. 6556, Bonn, Institute for the Study of Labor (IZA).

World Bank (2012), *Ecuador: the Faces of Informality*, Report No. 67808-EC, Washington, D.C.

Annex A1

Table A1.1
Ecuador: worker mobility across sectors of the labour market, 2007-2008
(Percentages)

Initial sector	Final status						P_i
	Formal waged	Informal waged	Informal self-employed	Unemployed	Outside the labour force	Total	
Formal waged	79	6	6	2	6	100	27
Informal waged	21	46	18	4	12	100	13
Informal self-employed	8	7	67	1	16	100	22
Unemployed	23	11	25	19	22	100	4
Outside the labour force	4	4	8	3	81	100	34
P_j	26	10	24	3	37	500	

Source: Author's calculations on the basis of the Ecuadorian Survey of Urban Employment, Unemployment and Underemployment, 2011-2012.
Note: P_i is the relative size of a sector in the initial period; P_j is the relative size of a sector in the final period.

Table A1.2
Ecuador: worker mobility across sectors of the labour market, 2009-2010
(Percentages)

Initial sector	Final status						P_i
	Formal waged	Informal waged	Informal self-employed	Unemployed	Outside the labour force	Total	
Formal waged	75	8	8	2	6	100	25
Informal waged	19	47	17	4	13	100	12
Informal self-employed	10	9	61	3	18	100	23
Unemployed	24	14	16	16	30	100	5
Outside the labour force	5	4	7	3	81	100	35
P_j	27	12	21	4	37		

Source: Author's calculations on the basis of the Ecuadorian Survey of Urban Employment, Unemployment and Underemployment, 2011-2012.
Note: P_i is the relative size of a sector in the initial period; P_j is the relative size of a sector in the final period.

Table A1.3

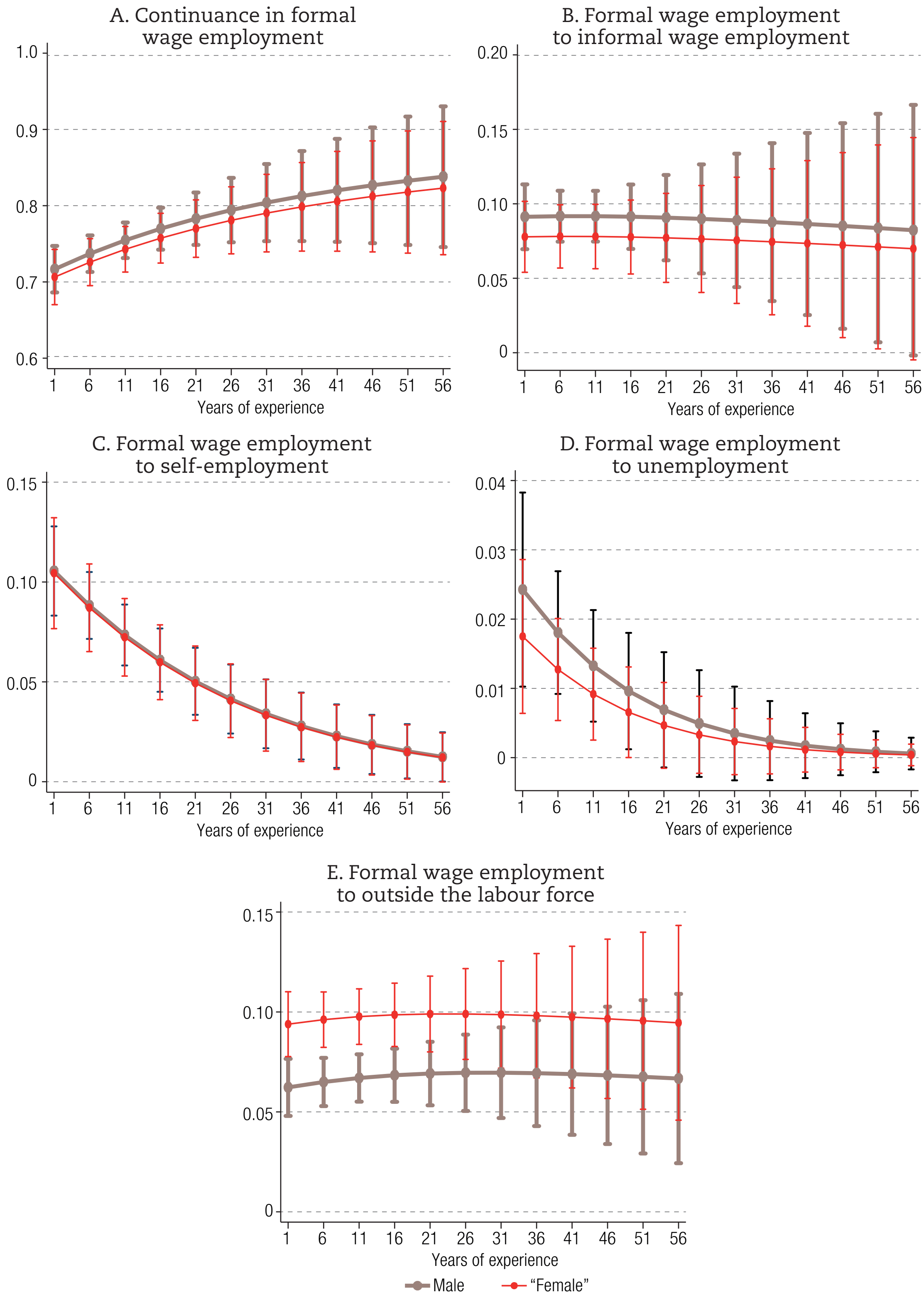
Ecuador: summary statistics using the first definition of the informal sector, 2011-2012[a]

Worker transitions between sectors	Mean				
	Number of observations	Age (*years*)	Schooling (*years*)	Experience (*years*)	Initial real wage (*USD*)
From the formal sector to					
Formal sector	1 841	39.2	7.4	9.8	466.2
		(12.91)	(1.93)	(10.38)	(505.70)
Informal sector (wage employment)	171	35.6	5.9	7.3	257.4
		(12.95)	(1.73)	(9.02)	(221.01)
Informal sector (self-employment)	178	41.6	6.8	8.1	366.9
		(12.55)	(1.99)	(8.97)	(392.70)
Unemployed	38	29.7	7.1	3.3	338.3
		(9.67)	(1.91)	(5.03)	(296.44)
Outside the labour force	156	40.9	6.7	10.7	375.1
		(20.14)	(2.02)	(13.65)	(466.04)
From the informal sector (wage employment) to					
Formal sector	196	34.0	6.1	6.2	209.5
		(12.69)	(1.73)	(8.19)	(114.56)
Informal sector (wage employment)	438	37.3	5.2	8.66	192.6
		(14.30)	(1.61)	(10.34)	(81.24)
Informal sector (self-employment)	190	42.4	5.3	10.5	180.1
		(12.50)	(1.76)	(12.31)	(98.84)
Unemployed	38	32.2	5.9	7.6	178.2
		(12.69)	(1.66)	(12.22)	(96.84)
Outside the labour force	106	36.4	5.5	6.8	129.5
		(17.52)	(1.93)	(12.10)	(78.03)
From the informal sector (self-employment) to					
Formal sector	200	42.6	6.8	11.3	290.8
		(14.06)	(1.97)	(11.08)	(394.28)
Informal sector (wage employment)	191	42.1	5.32	11.4	180.7
		(14.94)	(1.86)	(12.91)	(324.39)
Informal sector (self-employment)	1 591	48.0	5.5	13.6	215.0
		(13.53)	(1.90)	(12.03)	(302.08)
Unemployed	34	30.4	7.2	4.0	107.9
		(10.62)	(1.65)	(4.12)	(164.50)
Outside the labour force	350	48.0	5.4	10.7	90.7
		(19.53)	(1.90)	(13.73)	(137.73)

Source: Author's calculations on the basis of the Ecuadorian Survey of Urban Employment, Unemployment and Underemployment, 2011-2012.

[a] This definition of the informal sector refers to wage workers employed by small establishments with less than 10 employees and are not registered, plus all independent and self-employed workers. Standard errors are in parentheses.

Figure A1.1
Effect of years of experience on various labour market transitions, by gender[a]
(Probabilities)



Source: Author's calculations on the basis of estimation results.
[a] Predictive margins with 95% confidence intervals.

Figure A1.2
Effect of level of education on various labour market transitions, by gender[a]
(Probabilities)

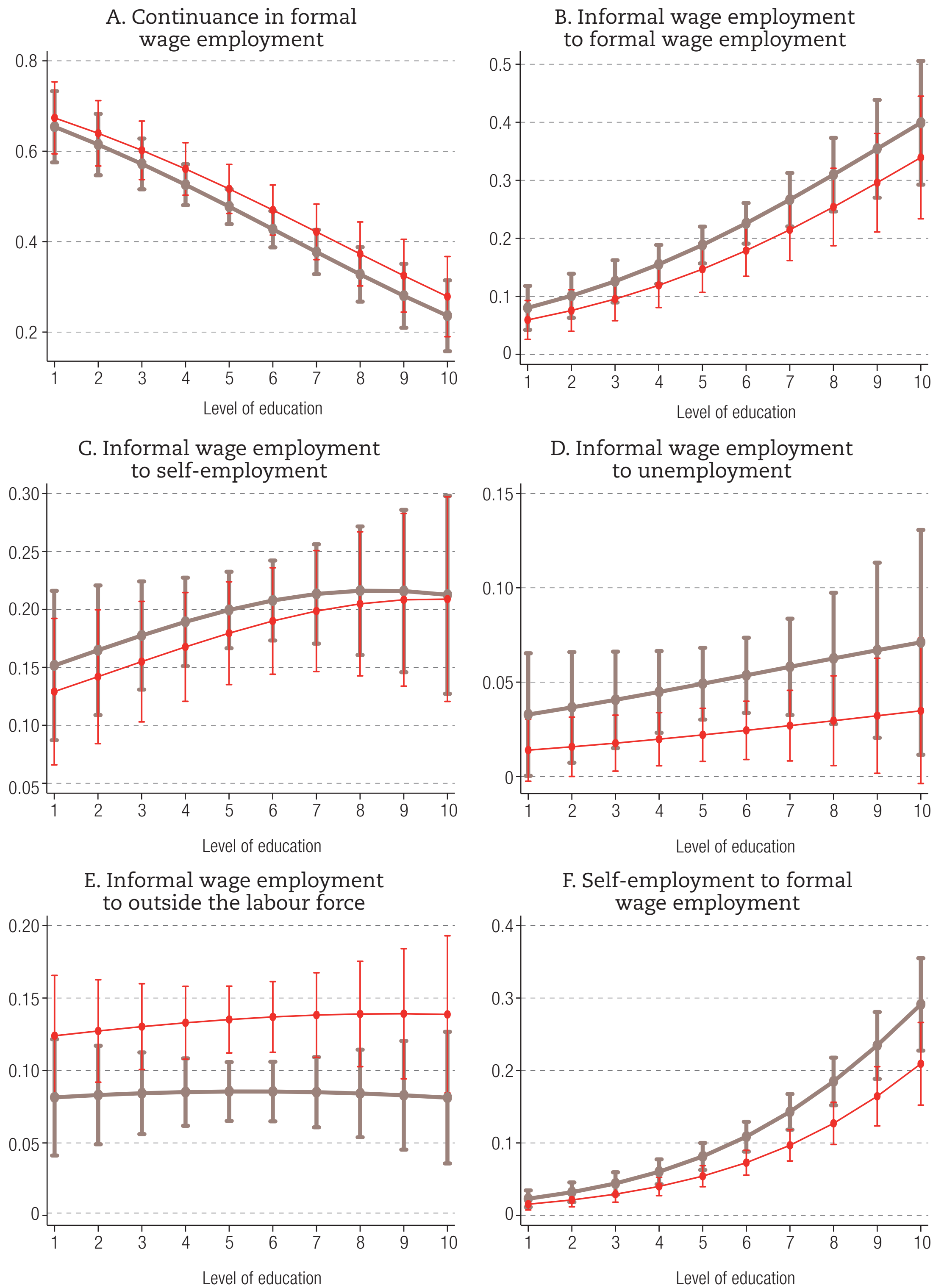
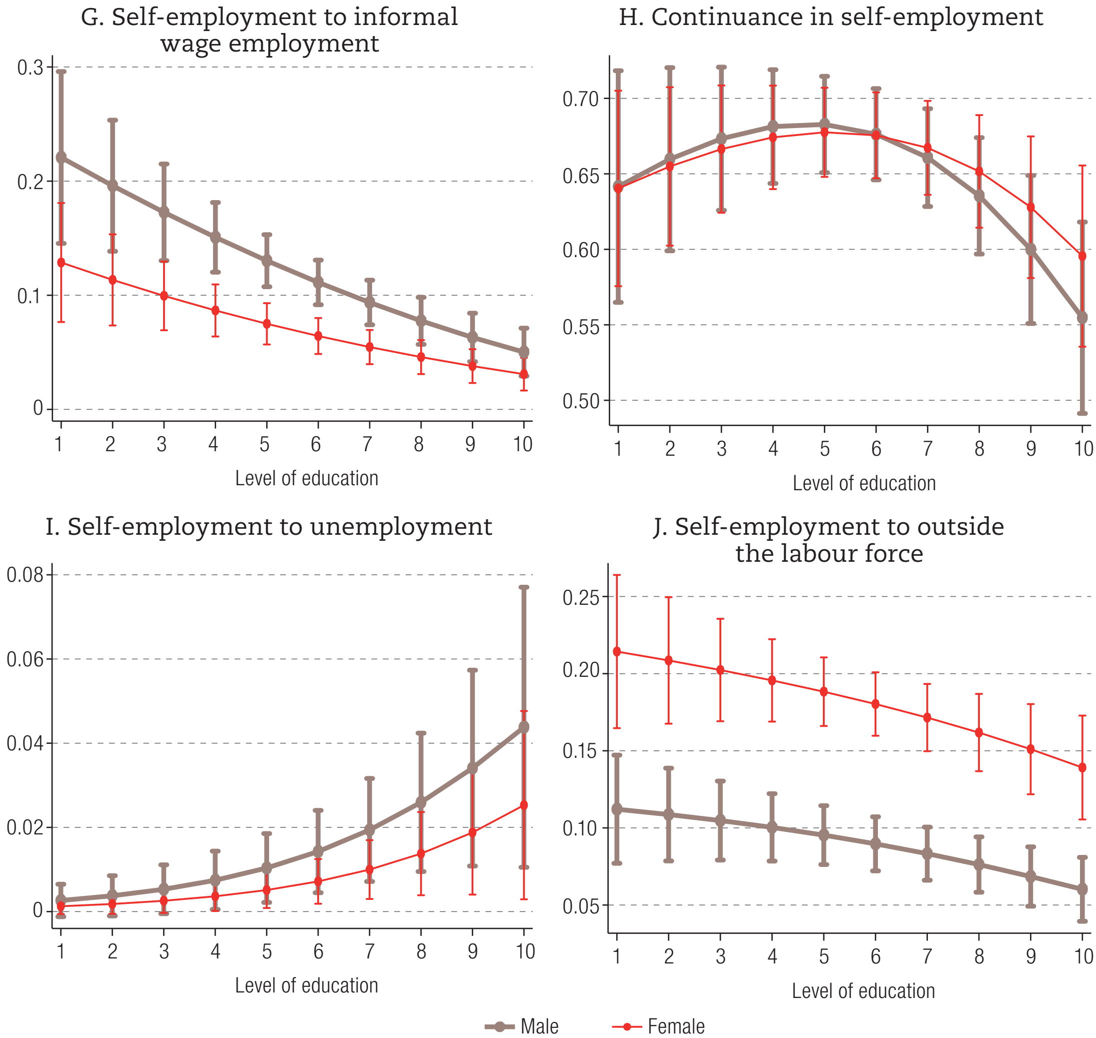


Figure A1.2 (concluded)



Source: Author's calculations on the basis of estimation results.

[a] Predictive margins with 95% confidence intervals.

The impact on wages, employment and exports of backward linkages between multinational companies and SMEs

Juan Carlos Leiva, Ricardo Monge-González and Juan Antonio Rodríguez-Álvarez[1]

Abstract

Policymakers often look for ways to attract foreign direct investment (FDI) by multinational corporations (MNCs). This paper estimates the impact of a programme, *Costa Rica Provee*, that seeks to increase backward linkages between small and medium-sized enterprises (SMEs) and MNCs in Costa Rica. The impacts were measured by reference to real average wages, employment demand and the probability of exporting, using a combination of fixed effects and propensity score matching with panel data on treated and untreated firms for 2001-2011. Programme beneficiaries evinced higher average wages, labour demand and export probabilities than untreated firms, with dose and duration also having a major influence.

Keywords

Transnational corporations, foreign direct investment, small enterprises, medium enterprises, employment, wages, exports, case studies, Costa Rica

JEL classification

G28, L53, O25

Authors

Juan Carlos Leiva is Professor at the School of Business Administration of the Technological Institute of Costa Rica. Email: jleiva@itcr.ac.cr.

Ricardo Monge-González is the Executive Director of the High-Technology Advisory Committee Foundation (CAATEC) and a Professor at the School of Business Administration of the Technological Institute of Costa Rica. Email: rmonge@caatec.org.

Juan Antonio Rodríguez-Álvarez is Professor at the School of Statistics of the University of Costa Rica and at the School of Business Administration of the Technological Institute of Costa Rica. Email: jara107@gmail.com.

[1] The authors wish to acknowledge the support provided for this research by the Inter-American Development Bank (IDB) and the Technological Institute of Costa Rica.

I. Introduction

The literature indicates that the impact of foreign direct investment (FDI) by multinational corporations (MNCs) on host-country economic development can materialize in a number of ways (Spencer, 2008).

Backward linkages are one of these ways, and they are important for host economies because they provide a channel through which knowledge, skills and technology can spread to local firms. The empirical evidence is mixed (Liu, 2011), but Giroud, Jindra and Marek (2012) find evidence that foreign affiliates' technological capability, embeddedness and autonomy are positively related to knowledge transfer via backward linkages.

Taking a different approach, Prashantham and Dhanaraj (2014) find that MNC relational capital is positively associated with the internationalization capability of new local ventures. For their part, Kim and Li (2014) investigate whether inward FDI contributes to entrepreneurial activity within the host country and find that the positive benefits for business creation are strongest in regions with weak institutional infrastructures and low overall educational attainment.

Domestic firms' absorptive capacity is an enabler of positive externalities through backward FDI linkages (Crespo and Fontoura, 2007; Ferragina and Mazzotta, 2014). Specifically, knowledge spillovers from foreign firms substantially depend on the absorptive capacity and productivity levels of individual firms in the host country (Damijan and others, 2013).

In sum, the empirical evidence on FDI-induced knowledge spillovers is mixed (Liu, 2011). It seems that the way backward linkages are measured matters greatly when it comes to establishing whether or not FDI is beneficial to host countries (Barrios, Görg and Strobl, 2011). Lastly, the heterogeneity of firms in terms of absorptive capacity, size, productivity and technology levels affects the results (Damijan and others, 2013).

In pursuit of the positive effects of FDI, policymakers in many developing countries have developed strategies to tempt MNCs to invest. Their main goal has often been to increase exports and FDI inflows, leading them to neglect backward linkages and technology transfer and absorption (Padilla-Pérez and Gaudin, 2014). For this reason, some developing countries, including Chile, Colombia, Costa Rica and Mexico, have created specific programmes to promote backward linkages.

However, the emergence of policies to foster linkages, knowledge spillovers and technology transfer has not been matched by actions to measure their impact. For example, López-Acevedo and Tan (2010) identified 23 rigorous impact evaluation studies, of which only 3 looked at measures to promote backward linkages from investment.

In these cases, the authors found that participation in any programme produced statistically significant and generally positive impacts on several firm performance measures, although some programmes were more effective than others. In Chile, for example, the greatest impact on final outcomes was found to come from technical assistance programmes, followed by cluster programmes and those to promote technology development and adoption.

This paper aims to help close the knowledge gap by assessing the impact of a specific programme, *Costa Rica Provee*, which promotes backward linkages between multinationals and local small and medium-sized enterprises (SMEs) in Costa Rica, a developing country. Specifically, the research focuses on the identification of timing or dynamic effects (how long it takes for results to come through) and treatment intensity (dosage effects).

Costa Rica Provee, a name that translates as "Costa Rica Supplies", was created in 1998 when the country's authorities recognized the need to develop suppliers, given that MNCs operating in export processing zones (EPZs) were poorly integrated with local companies, and to improve Costa Rica's investment climate. After several revisions and adjustments, *Costa Rica Provee* is now more driven by MNC demand, identifying the main input and raw material requirements of MNCs and matching these with local suppliers.

Both public and private organizations have influence over *Costa Rica Provee*. The institutional framework includes actions for policy implementation, monitoring, accountability and consultation or lobbying.

Our results indicate that *Costa Rica Provee* has had positive and significant impacts on the performance of beneficiary firms, specifically on their real average wages, labour demand and probability of exporting.

The article is structured as follows. Following this introduction, section II presents a literature review, section III provides an overview of the *Costa Rica Provee* programme, section IV describes the methodology employed to estimate its impacts, section V discusses the findings and section VI concludes.

II. Literature review

Knowledge spillover at the firm level is defined as knowledge created by one firm (in this case an MNC) and used by a second firm (a host-country firm) which does not (fully) compensate the MNC for this. It is thus a positive externality. In the case of technology transfer, conversely, the MNC is compensated. It also entails different kinds of costs (Smeets, 2008; Javorcik, 2004).

As noted earlier, policymakers in many countries have devised strategies to persuade MNCs to make investments in the form of FDI. MNCs have the potential to contribute to increased productivity among local firms, and the literature indicates that the impact of FDI on host-country economic development can materialize in various ways, such as through knowledge spillovers and technology transfer, although the evidence on this is mixed.

Some previous studies have shown negative results, with limited linkage effects and a low level of integration into host economies (Liu, 2011). By contrast, Damijan and others (2013) suggest that horizontal spillovers have become increasingly important over the past decade.

Amendolagine and others (2013) investigate the determinants of backward linkages between foreign manufacturing firms and domestic suppliers in 19 sub-Saharan African countries, finding that the time since entry of the foreign firm, the presence of a local partner in the ownership structure and a final market orientation are associated with stronger local linkages. They distinguish between the backward linkages associated with different types of FDI: resource-seeking, efficiency-seeking and market-seeking.

Damijan and others (2013) present a comparative study of the importance of direct technology transfer and spillovers via FDI for a group of 10 transition economies, using a dataset of over 90,000 firms. This study concludes that outcomes are affected by absorptive capacity, size, productivity and technology levels. The spillover effects produced by foreign firms depend substantially on the absorptive capacity and productivity of individual local firms, with only the most productive and absorptive being able to benefit from knowledge spillovers.

Giroud, Jindra and Marek (2012) analyse backward linkages between foreign affiliates and local suppliers in five transition economies using survey data on 809 foreign affiliates. They find a non-linear relationship between the extent of local sourcing and knowledge transfer to domestic suppliers, but their evidence shows that foreign affiliates' technological capability, embeddedness and autonomy are positively related to knowledge transfer via backward linkages.

Kiyota and others (2008) examine determinants of the backward vertical linkages of Japanese foreign affiliates in manufacturing for the period 1994-2000, finding that unobserved affiliate-specific characteristics explain much of the variation in backward linkages among foreign affiliates and that an affiliate's experience has positive and sometimes non-linear effects on local procurement.

Finally, Javorcik (2004) finds evidence of positive productivity spillovers from FDI for local suppliers when investment projects are co-owned by domestic and foreign firms.

From a different perspective, Alfaro and others (2004) provide evidence that only countries with well-developed financial markets gain adequately from FDI in terms of growth rates.

The main point, then, is that success in attracting FDI does not automatically generate benefits through backward linkages. For a host country, there may be a case for government intervention to remove the obstacles limiting the interaction of foreign firms with local suppliers and buyers, particularly SMEs. Backward linkage development must be approached from both the demand side (MNCs) and the supply side (local firms), because success depends both on the interest of MNCs in sourcing inputs in the host country and on the host's domestic linkage capability.

On the demand side, there are various points to consider, beginning with the sophistication of the MNC subsidiary's production processes: more advanced operations could create more and higher-value local linkages. Secondly, the chief executive officers of new MNC subsidiaries do not necessarily pursue linkages with local firms as part of corporate policy, since building facilities and launching operations are the main initial priorities. In their procurement policy, local managers often look to global suppliers rather than local firms for security reasons (robust production processes). In addition, newly arrived local procurement managers usually lack knowledge of local capabilities. The high costs associated with identifying local suppliers represent an information asymmetry that limits local linkages (i.e. a market failure) (Wanga and others, 2012; Hallin and Holmström, 2012; Liu, 2011; Zhang and others, 2010; Smeets, 2008; Saggi, 2002).

On the supply side, local firms are not necessarily capable of delivering goods and services to multinationals because of a lack of firm-level capabilities in areas such as entrepreneurship, technology, production scale, risk management and financing. Even when local firms are successful in becoming MNC suppliers, the absorptive capacity of the host country also depends on systemic learning infrastructure, institutions and government policies (Wanga and others, 2012; Zhang and others, 2010; Paus and Gallagher, 2008).

Local firms, especially SMEs, face significant obstacles in pursuing and identifying better business opportunities with more advanced companies (incomplete information). Potential high-value transactions and contracts with advanced MNCs are often out of reach for SMEs, even if they have basic production skills that could be enhanced through specific investments. They may also find it too costly to engage in identification of market opportunities (coordination failures). The need to carry out and finance the required investment in technological upgrading to comply with MNC requirements can be yet another structural obstacle to the cluster development of local suppliers.

On the basis of the foregoing, a national plan to promote productive linkages between MNCs and local firms may be seen as a response to specific market failures (coordination failures among local companies) and externalities (from FDI). Thus, there are arguments in favour of government action. For this reason, some developing countries, including Chile, Colombia, Costa Rica and Mexico, have created specific programmes to promote backward linkages.

However, the emergence of policies to foster linkages, knowledge spillovers and technology transfer activities has not been matched by action to measure their impacts. For example, López-Acevedo and Tan (2010) identified 23 rigorous impact evaluation studies, of which only 3 looked at measures to promote backward linkages from investment.

Generally speaking, although earlier evaluations of SME programmes were pessimistic about their impacts, recent studies have found positive impacts on intermediate outcomes such as research and development expenditures, worker training, new production processes and quality control programmes, as well as networking with other firms and with different information and funding sources. However, evaluations continue to yield mixed results for impacts on firm performance (López-Acevedo and Tan, 2010).

Specifically, programmes involving backward linkages are few. In Chile, Tan (2010) evaluated 603 firms, of which 207 reported having participated in one or more programmes (the treatment group) and 396 stated that they had never participated in any (the control group). The programme categories were: supplier development; technical assistance; support for business clusters; technology development; technology transfer; working capital; debt rescheduling; and other types.

Using propensity score matching combined with difference-in-difference models, Tan found evidence that programme participation was causally related to improvements in a range of intermediate outcomes (training, adoption of new technology and organizational practices) as well as positive gains in sales, labour productivity, wages and, to a lesser extent, employment. Positive treatment effects were also found by type of programme. However, only 2% of the firms evaluated had participated in a supplier development programme.

In Colombia, Duque and Muñoz (2010) evaluated the impact of the Colombian Fund for Modernization and Technological Development for Micro, Small, and Medium-sized Enterprises (FOMIPYME). FOMIPYME promotes modernization and technological change, including new business start-ups, entrepreneurship training, innovation and technology upgrading, marketing strategies, network-building, product commercialization, export promotion, productivity improvements and mini-clusters. One line is supplier development, but the impact evaluation did not analyse this separately.

Using a fixed effects model that controlled for programme impacts over time, Duque and Muñoz (2010) detected a generally positive effect on wages in the first two years of treatment, turning negative thereafter. Total factor productivity also showed a large and positive effect, diminishing in the second and third year after treatment but increasing significantly in the fifth year.

In Mexico, López-Acevedo and Tinajero (2010) found a huge array of programmes, numbering about 150. One of them, the Fund for the Promotion of Production Chains (FIDECAP), sought to encourage and strengthen the vertical and horizontal linkages of SMEs with other firms. Their findings indicate that participation in certain types of programme is associated with higher value added, sales, exports and employment. However, the assessment does not individualize each programme.

In another context, Görg, Hanley and Strobl (2011) investigate whether government subsidies encourage MNCs to create linkages with domestic suppliers in Ireland. Their results indicate that while foreign plants from Europe and the United States develop backward linkages independently of grant receipt, multinationals from other parts of the world respond positively to government support. They conclude, therefore, that governments should avoid taking a one-size-fits-all approach to incentivizing MNCs to develop local linkages.

III. The *Costa Rica Provee* programme

Since the creation of the export processing zone (EPZ) regime in the early 1980s, the promotion of productive linkages has been a subject of public concern in Costa Rica because of the weakness of vertical integration in industry, itself the result of an inward-looking development strategy based on import substitution during the 1960s and 1970s, which promoted the manufacture of final goods rather than raw materials and intermediate goods.

The National Programme of Science and Technology 1986-1990 made reference to this topic. Notwithstanding the interest in the public sector, the first efforts to develop local suppliers came from the private sector. Baxter Health Care, Inc., one of the first major MNCs established in Costa Rica, created a technical assistance programme to develop local suppliers in the mid-1990s as part of the firm's business strategy for the country.

In 1998, a group of public and private organizations created the Local Industry Improvement Programme to help local companies do more business with high-technology MNCs. Later, in 1999, the Supplier Development Project for High-Technology Multinational Companies was created. The next step was the creation of *Costa Rica Provee*, a national supplier development office legally constituted in early 2002. The programme was delayed for almost two years because of organizational and administrative difficulties. In 2004, the Executive Committee transferred *Costa Rica Provee* to the Foreign Trade Corporation of Costa Rica (PROCOMER) in order to provide continuity for the programme by consolidating it within a well-funded organization and to strengthen indirect exports to MNCs.

Costa Rica Provee explores the needs of multinational companies, identifies business opportunities and recommends partner suppliers that comply with the production, technical and quality specifications and characteristics required by MNCs. The programme has directed its services toward three strategic business areas: (i) the information and communications technology, electrical and electronics and metallurgical industries, (ii) the medical, chemical and pharmaceutical industries and (iii) agribusiness and textiles.

Costa Rica Provee has become more oriented towards demand from MNCs, identifying their main input and raw material requirements and matching these with local suppliers. It has also created business opportunities by means of small projects between SMEs and MNCs aimed at helping local suppliers to rise in the value chain and ultimately become global suppliers.

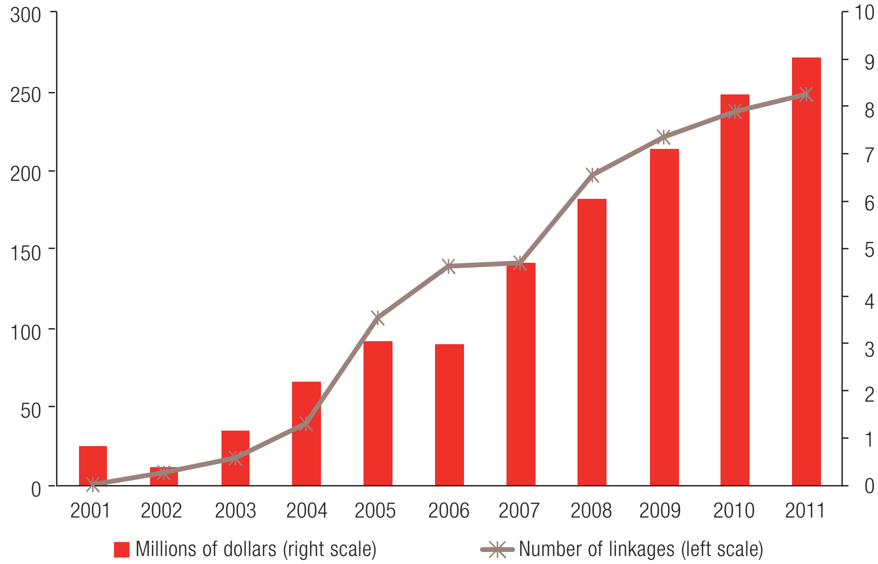
Costa Rica Provee was not created by law, but its activities are influenced by the EPZ Act and its regulations, particularly as regards customs procedures. In fact, the Act regulates commercial relations between EPZ firms and local companies through two mechanisms: (i) direct purchase, when an EPZ company buys a final good or service from a local firm without the MNC contributing any raw materials, machinery or equipment to the local supplier, and (ii) outsourcing, when the EPZ company provides the local supplier with raw materials and even machinery and equipment to produce the final goods.

The EPZ Act has undergone three major linkage-related reforms (in December 1999, June 2006 and August 2008) that have made these mechanisms more flexible. The most recent made substantial changes to outsourcing procedures: the outsourcing cap was raised from 25% to 50% of total MNC value added, simultaneous contracting with different suppliers was permitted, and the one-year limit on contract duration was eliminated. In addition, the movement of machinery and equipment outside EPZs was allowed, enabling local suppliers to integrate them into the production process. Red tape and burdensome administrative procedures were also eliminated: the number of steps involved in registration was reduced from 10 to 2, and approval time was cut from between 15 and 20 days to just 3.

Both public and private organizations have influence over *Costa Rica Provee*. The institutional framework is as follows. PROCOMER is responsible for the design and reform of *Costa Rica Provee*, influenced by Ministry of Foreign Trade regulatory actions applying to EPZs. Implementation, monitoring and accountability are also a PROCOMER responsibility. A sizable group of private and public organizations with an interest in promoting production linkages between MNCs and local suppliers are connected to *Costa Rica Provee*.

A summary of the programme's results is shown in figure 1. Between 2001 and 2011, the number of backward linkages registered each year by *Costa Rica Provee* increased from 1 to 248, representing a rise in sales from US$ 0.8 million in 2001 to US$ 9.0 million in 2011. Groote (2005) found that only 17.3% of the linkages created by *Costa Rica Provee* were incorporated into MNCs' high-technology final products, meaning that most linkages were associated with non-specialized inputs. There was a significant jump in the number of backward linkages created each year over the period 2007-2009, from 141 in 2007 to 197 in 2008 and 220 in 2009. Over the whole period 2001-2011, the programme generated a total of 1,355 linkages between local firms and MNCs.

Figure 1
Productive linkages created yearly by *Costa Rica Provee*, 2001-2011
(Number of linkages and millions of dollars)



Source: Prepared by the authors, on the basis of data from the Ministry of Foreign Trade (COMEX) and the Foreign Trade Corporation of Costa Rica.

Although the results are positive and show that linkages are increasing, the operations involved are of a small magnitude in relation to the size of the Costa Rican economy and MNC purchases. For instance, local purchases by MNCs in Costa Rica totalled US$ 591.1 million in 2007, while those made under *Costa Rica Provee* auspices in the same year were just US$ 4.7 million, or less than 1% of the total. According to data from the Ministry of Finance, approximately 9,654 local companies supplied different types of goods and services to MNCs operating under the EPZ regime from 2001 to 2011. This number of local suppliers contrasts with the small number of *Costa Rica Provee* beneficiary firms (just 403) during the same period. Thus, *Costa Rica Provee* beneficiary firms made up only 4% of all MNCs' local suppliers.

Flores (2011) investigates whether *Costa Rica Provee* has helped to develop backward linkages between high-technology MNCs and local firms. He uses empirical evidence to evaluate the relationship between being part of *Costa Rica Provee* and having achieved linkages involving higher asset specificity, and he estimates some econometric models with data from a panel of 94 high-technology MNCs from 2001 to 2008. The empirical results provide no robust evidence of a positive effect of *Costa Rica Provee* on the generation of backward linkages between high-technology MNCs and local suppliers. Paus and Gallagher (2008) claim that Costa Rica has not fully realized the potential of FDI for economic development, since backward linkages between MNCs and local firms are not as robust as they should be.

IV. Methodology

We follow a quasi-experimental approach that requires specific data on the programme under consideration, including data on firms affected by the measure or participating in the programme and data on a control group of similar firms which are not affected or participating. A panel of companies treated and untreated by *Costa Rica Provee* between 2001 and 2011 was built for this purpose.

Since *Costa Rica Provee* beneficiaries are not randomly selected, the participation or selection of firms in the treatment and control groups should be based on observable and unobservable characteristics that can be controlled for (quasi-experimental design). The technique we use in carrying out the impact evaluation is a combination of regression methods and propensity score matching that explicitly controls for differences in observable variables between groups and fixed effects models, which use data from before and after the programme (treated and control groups) to account for certain types of unobserved heterogeneity.

As is well known, the challenge of impact evaluation is to be able to compare a firm's performance after programme intervention to what would have happened if the firm had not participated. Since the hypothetical scenario cannot actually be studied, it is necessary to identify a group of firms that are similar to the group receiving the treatment (programme beneficiaries) in all respects except for their participation in the programme. The way this control group is selected is critical because any difference in performance between the control group and the treatment group, in terms of observed or unobserved attributes, affects the accuracy of the estimates of the programme's net impact. For this reason, it is important to explain the strategy used to correct potential selection biases and thus provide an assurance that the results obtained from the impact evaluation are actually attributable to the programme intervention under analysis.

1. The strategy for identifying the control group

Since none of the firms in the panel received support from *Costa Rica Provee* between 2001 and 2003, these are taken as the baseline or pre-treatment years for the purposes of this analysis.

To estimate the impact of support on SME performance, we combine propensity score matching with a fixed effects model. Propensity score matching makes it possible to control for selection bias attributable to firms' observable characteristics, while the fixed effects method controls for unobservable attributes which are considered to be fixed over time (time-invariant firm characteristics) and which may affect a firm's decision to seek support from *Costa Rica Provee* or its performance over time.

Selecting the control group involves an analysis of the variables characterizing all the firms before they became programme beneficiaries, i.e. between 2001 and 2003. Since beneficiary firms received support from *Costa Rica Provee* at different times during the period studied, estimation of the propensity score matching for the panel data requires calculation of a dummy variable D taking the value 1 if a firm was a beneficiary of *Costa Rica Provee* at least once during the period 2004-2011 and 0 if it was never a beneficiary. In other words, D takes a value of 1 when a firm starts participating in the programme.

Propensity score matching estimates the probability of a firm participating in *Costa Rica Provee* as a function of a set of observed variables. The first estimation is of the probability of participation as the matching criterion between beneficiary firms (treatment) and non-beneficiaries (control). Given the large number of variables characterizing firms, their values have to be reduced to a scalar $p(x)$, defined below, to allow matching. As pointed out by Bernal and Peña (2011), it is important not to omit any variable or to overspecify the model. Careful attention must be paid to the choice of variables to include.

The propensity score is defined as the conditional probability that a firm will become a beneficiary of *Costa Rica Provee*, given the values of a set of observed variables X, which is expressed as:

$$px = PD = 1|X = x = ED|X = x \tag{1}$$

where X is a vector of individual characteristics or variables of the firm and its environment.

Rosenbaum and Rubin (1983) show that if the fact of being a beneficiary or not is the result of a random selection process in the neighbourhood defined by the multidimensional vector X, this selection is also random in the region defined by the scalar $p(x)$. Therefore, the average treatment effect on the treated (ATT) in the case of the treatment given by *Costa Rica Provee* to beneficiary firms may be specified by the equations:

$$ATT = E[Y_1 - Y_0] = E[E[Y_1 - Y_0 | p(x)]] \tag{2}$$

and

$$E[Y_1 | p(x),\ D=1] - E[Y_0 | p(x),\ D=0] = E[Y_1 - Y_0 | p(x)] \tag{3}$$

where Y_i is the outcome variable on which the impact of the *Costa Rica Provee* programme is being measured, while subscript i indicates the year of observation of the outcome variable.

The impact of *Costa Rica Provee* may then be estimated as the difference between the average of the outcome variable for the treatment group (beneficiaries) and that for the control group in the area of common support (where the data show an overlap in the characteristics of beneficiaries and non-beneficiaries) defined by the propensity score matching.

A problem with the estimation (ATT) is that it does not take into account the possibility of selection bias due to unobserved variables, and this is compounded by the fact that, according to the panel data, treatment did not occur within the same year for all firms and was not continuous once a business entered the programme. We therefore estimate the programme's impact by using the propensity score matching results to define the treatment and control groups in a way that meets the common support condition, while estimating the impact equations by means of a regression method that follows the fixed effects approach.

2. Specification of the models and estimation procedure

To estimate the impact of *Costa Rica Provee* on SME performance, we applied a set of regression models to one set of panel data from 2004 to 2011, relating the outcome variable (wages, employment or exports) to a set of covariates, including a dummy variable (D) to indicate whether or not the firm was a beneficiary of the programme at some time in that period. For the case of wages and employment, we derived the model specifications on the assumption that Costa Rican SMEs displayed profit-maximizing behaviour.

The estimation was conducted using the ordinary least squares (OLS), propensity score matching and fixed effects approaches. In the case of exports, a linear probability model was used to estimate the impact of the programme on the probability that a firm exported at some time between 2004 and 2011. In this latter case, both the fixed effects and propensity score matching approaches were also used.

In short, the following three equations were estimated:

$$(w-p)_{it} = \beta_0 + \beta_1 (PREM{*}SE)_{it} + \beta_2 D_{it} + \beta_3 D_{it\text{-}1} + \beta_4 D_{it\text{-}2} + \beta_5 X_{it} + \varepsilon_{it} \qquad (4)^2$$

$$l_{it} = \gamma_0 + \gamma_1 D_{it} + \gamma_2 D_{it\text{-}1} + \gamma_3 D_{it\text{-}2} + \gamma_4 X_{it} + \sigma_{it} \qquad (5)^3$$

$$exp_{it} = \delta_0 + \delta_1 D_{it} + \delta_2 D_{it\text{-}1} + \delta_3 D_{it\text{-}2} + \delta_4 X_{it} + \rho_{it} \tag{6}$$

[2] See annex A1 for the derivation of equation (4).

[3] See annex A2 for the derivation of equation (5).

where $(w-p)$ is the average real wage paid by the firm (in logs), $PREM * SE$ the wage premium received by skilled workers, L the number of workers hired by the firm (in logs), exp a dummy variable equal to 1 if the firm exported in year t and 0 otherwise, and X the covariates. Each error term in equations (4), (5) and (6) is a two-component term, with one component relating to an unobserved specific effect of the firm which does not vary over time (production sector, managerial capacity, etc.) but which may have an impact on the outcome variable, and another component which is purely stochastic.

We estimated another specification of equation (3) that included lag values of the dependent variable. This was because a firm's exports in year t are explained by its export performance in years $t-1$, $t-2$ and $t-3$. Thus, a dynamic linear probability model was estimated. Following the literature, we did not use the fixed effect approach in the estimation of this new specification.

In addition to estimating the three equations mentioned above, we explored the timing of the effects and whether dosage was really important, following Crespi and others (2011). For these purposes, we modified the above three equations, substituting for the impact variable D another dummy called D_timing that takes a value of 1 in every year from the first intervention and 0 if there has been no intervention. For the dosage effect, we substituted for the impact variable D another variable called D_dosage that takes a value of 1 in every year from the year the firm was first treated up to the year before the second treatment, 2 from the year the firm was treated for the second time up to the year before the third treatment, and so on successively, and 0 if there has been no treatment. In other words, this covers cases in which a firm was a beneficiary in more than one year.

3. Data

We collected information on beneficiaries and linked this to social security and export data in order to obtain microdata on final outcomes (total employment, average wages and exports) and on each firm's industry sector, location and legal status. These are official data from various government sources (ministries, the social security institute and foreign trade figures). We were thus able to construct the panel of companies with and without treatment under the *Costa Rica Provee* programme between 2001 and 2011.

V. Results

Before presenting the results of the *Costa Rica Provee* impact evaluation, we shall present those from the propensity score matching technique used to identify the firms belonging to the control group, specifically the common support.

1. Estimation of the propensity score and construction of the common support

Table 1 shows the variables used to estimate the propensity score for the sample firms and the results of the estimation. We estimate the probability of firms participating in the programme between 2004 and 2011 with reference to their characteristics between 2001 and 2003, i.e. before any of the firms in the sample participated in the programme.

Table 1
Estimation of the probit function for propensity score matching measured for the period 2001-2003
(Coefficients and p-values)

Variable	Coefficient
Firm is located in San José	0.6280***
	(0.1544)
Firm is located in Cartago	0.8718***
	(0.1894)
Firm is located in Heredia	0.7408***
	(0.1846)
Lithographic process	0.8663***
	(0.2429)
Firm exported in 2002	0.5457***
	(0.1312)
Real wage in 2001 (log)	0.1369***
	(0.0323)
Workforce growth from 2001 to 2003	0.2010*
	(0.1058)
Constant	-4.2812***
	(0.5299)
Number of observations	1 670
Wald chi-squared (7)	100.20
Prob>chi-squared	0.0000
Pseudo R-squared	0.1058

Source: Prepared by the authors.
Note: *Statistically significant at 10%; **statistically significant at 5%; ***statistically significant at 1%.

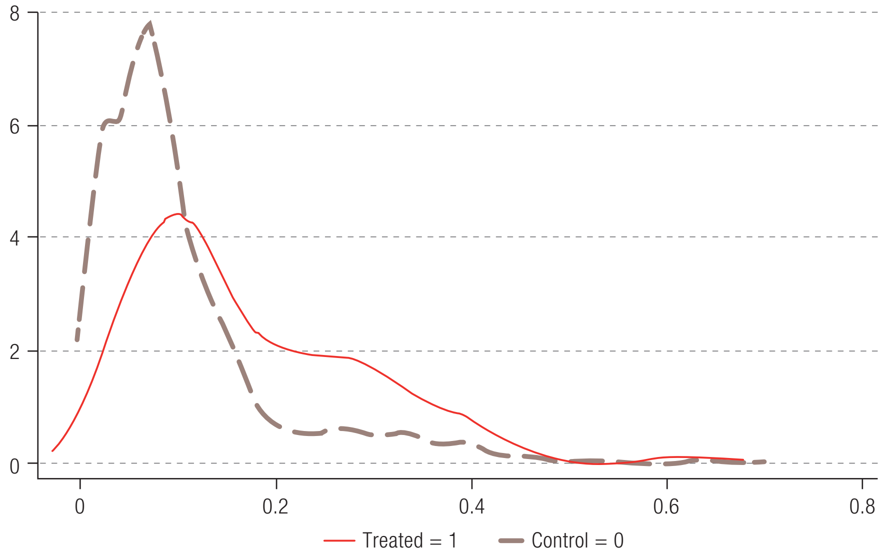
Propensity scores estimated by means of the participation model presented in table 1 are used to identify firms that did not participate in the *Costa Rica Provee* programme but that have the closest propensity score values to firms that did. Participation model variables include: geographic location (the three provinces where MNCs have the most local suppliers), sector of economic activity (the lithographic process, since this is the most common input provided by local suppliers to MNCs) and some firm characteristics such as the number of workers, average wage and a dummy variable indicating whether the firm exported in 2002.

All the coefficients included in the equation are significant. In addition, the model as a whole is significant, which means it is appropriate for estimating the probability of firms participating in the *Costa Rica Provee* programme and allocating them to either the treatment group or the control group. To achieve common support, it is necessary to eliminate the 20% of observations with the lowest density in the participation probability.

Figure 2 shows the distribution of the propensity scores after the matching of firms, i.e. it shows the propensity score matching results for firms in the treatment and control groups previously selected within the common support.

After identifying the firms to be included in the control group, i.e. firms with similar propensity score values, it was necessary to check that the characteristics of the firms in the control group were equal to those of the firms participating in the programme (Rosenbaum and Rubin, 1983). We did this by analysing t-tests for equality of means in the treated and untreated groups before and after matching (t-tests are based on a regression of each variable on the treatment indicator).

Figure 2
Density of treated and untreated firms resulting from propensity score matching in the common support for the *Costa Rica Provee* impact evaluation
(Probabilities)



Source: Prepared by the authors.

Table 2 shows the balance in the observable variables before and after matching for the firms in the common support. After matching, it is not possible to reject the null hypothesis that differences in means between firms in the programme and in the control group are zero for all the variables simultaneously. Therefore, the treated and untreated groups in the sample after the matching procedure are statistically comparable for the observable variables included in the participation model.

Table 2
Balance in observable variables before and after matching for the *Costa Rica Provee* impact evaluation

Variable	Sample	Treated	Control	Difference	Standard error	T-statistic
Firm is located in San José	Unmatched	0.61074	0.53386	0.07688	0.04276	1.80
	Matched	0.61667	0.62500	-0.00833	0.06786	-0.12
Firm is located in Cartago	Unmatched	0.14765	0.09599	0.05166	0.02581	2.00
	Matched	0.15000	0.14167	0.00833	0.04945	0.17
Firm is located in Heredia	Unmatched	0.18121	0.12032	0.06089	0.02844	2.14
	Matched	0.15833	0.20000	-0.04167	0.05338	-0.78
Firm exported in 2002	Unmatched	0.26846	0.08284	0.18562	0.02529	7.34
	Matched	0.17500	0.15833	0.01667	0.04846	0.34
Lithographic process	Unmatched	0.07383	0.01512	0.05870	0.01204	4.87
	Matched	0.00833	0.00000	0.00833	0.00833	1.00
Real wages in 2001	Unmatched	16.72404	15.81788	0.90617	0.13314	6.81
	Matched	16.59053	16.59801	-0.00749	0.20257	-0.04
Workforce growth from 2001 to 2003	Unmatched	0.10335	0.08817	0.01518	0.03800	0.40
	Matched	0.12301	0.05486	0.06815	0.05586	1.22

Source: Prepared by the authors.

2. The impact of *Costa Rica Provee* on real average wages

For the impact of the *Costa Rica Provee* programme to be correctly estimated, as mentioned previously, it is necessary to control not only for firms' participation but also for observable and unobservable variables whose behaviour may affect the result variable. Since the beneficiary firms of *Costa Rica Provee* are SMEs, the study sample for both beneficiary and control group firms was limited to businesses with up to 100 employees.

Results for real wages (equation (4) in the methodology) are presented in table 3. The second column shows a positive and significant result for the treatment variable, D_t (0.1212), which suggests that firms' participation in *Costa Rica Provee* has a positive and significant impact on the real wages they pay their employees. A comparison of the first and second columns of table 3 also shows that the wage premium for differences in labour categories ($Prem*SE$) has a positive and significant coefficient (0.0775).

It is also interesting to note in column 3 that the impact of participating in *Costa Rica Provee* is experienced not only during the actual year the treatment is applied, but also one and two years later (the coefficients associated with D_{t-1} and D_{t-2} are positive and significant at 0.1462 and 0.1473, respectively).

Table 3
Impact of *Costa Rica Provee* on real average wages
(Fixed effects and cluster-robust standard errors)

Variable	(1) fixed effects	(2) fixed effects	(3) fixed effects	(4) fixed effects	(5) fixed effects
D_t (dummy equal to 1 if firm was treated in year t and 0 if never treated)	0.0780***	0.1212***	0.1304***		
	(0.0181)	(0.0157)	(0.0155)		
D_{t-1} (treatment dummy lagged one year)			0.1462***		
			(0.0173)		
D_{t-2} (treatment dummy lagged two years)			0.1473***		
			(0.0207)		
$Prem*SE$ (wage premium for different labour categories)		0.0775***	0.0781***	0.0793***	0.0792***
		(0.0052)	(0.0052)	(0.0053)	(0.0053)
D_timing_t (dummy equal to 1 from the first year the firm was treated and so on successively, 0 if never treated)				0.3300***	
				(0.0201)	
D_dosage_t (dummy equal to 1 for all years from the first year the firm was treated until the year before the second treatment, 2 from the year the firm was treated until the year before the third treatment and so on, 0 if never treated)					0.1659***
					(0.0139)
Constant	13.8197***	13.4258***	13.4188***	13.4057***	13.4080***
	(0.0003)	(0.0263)	(0.0267)	(0.0275)	(0.0274)
Observations	26 082	26 082	26 082	26 082	26 082
R-squared	0.0006	0.1627	0.1658	0.1716	0.1712
Number of observations	4 628	4 628	4 628	4 628	4 628

Source: Prepared by the authors.
Note: *Statistically significant at 10%; **statistically significant at 5%; ***statistically significant at 1%.

When the treatment variable is replaced by D_timing and the dynamic (non-linear) effects of participation in *Costa Rica Provee* are analysed (column 4), the results indicate that the longer a firm is treated, the greater the impact. In fact, the coefficient associated with D_timing is positive and significant (0.3300). This finding may suggest that *Costa Rica Provee* beneficiary firms continue to take advantage of the knowledge acquired from their commercial relationship with MNCs, with a permanent impact on their performance.

Finally, the results in column 5 for treatment dosage (D_dosage) suggest that for a *Costa Rica Provee* beneficiary firm to have been treated several times during the period analysed (2004-2011) helps it to increase real wages for its employees (the coefficient associated with D_dosage is positive and significant at 0.1659). A possible interpretation of this result is that the more commercial relationships (linkages) with multinational corporations the SMEs participating in the *Costa Rica Provee* programme have, the greater the knowledge they acquire, the result being a positive impact on their future performance.

Bearing in mind that the comparison group taken for the table 3 estimations can be improved by using firms whose probabilities of participating in the programme are similar to those of firms in the control group, the table 3 models were estimated again, controlling for fixed effects, but using only the common support firms. These new estimations are considered to be more robust because firms which are not good "clones" of beneficiary firms were eliminated from the control group by means of propensity score matching. Table 4 shows the impact of *Costa Rica Provee* on real wages, using the fixed effects and propensity score matching approaches.

The results presented in table 4 are consistent with those in table 3, indicating that the participation of SMEs in the *Costa Rica Provee* programme certainly has a positive and significant impact on real wages at beneficiary firms (columns 1 to 5). However, when we tested whether the parallel pre-treatment trend assumption was valid, we found that using fixed effects was an invalid approach in this case. In fact, the results for all the coefficients associated with pre-treatment variables ($PD\_$) in column 6 are significant and different from one another. For this reason, we carried out propensity score matching with a least squares dynamic model, the results of which are presented in column 7. The coefficient associated with the treatment variable (D) in this specification is positive and significant (0.0377), so we can conclude that the participation of firms in *Costa Rica Provee* has a positive and significant impact on real wages.

Finally, all the coefficients associated with the pre-treatment variables are negative. One possible interpretation of this result is that firms facing a negative shock before the treatment were the ones that sought to participate in the *Costa Rica Provee* programme.

Table 4

Impact of the *Costa Rica Provee* programme on real average wages

(Propensity score matching, fixed effects, dynamic least squares and cluster-robust standard errors)

Variable	(1) propensity score matching and fixed effects	(2) propensity score matching and fixed effects	(3) propensity score matching and fixed effects	(4) propensity score matching and fixed effects	(5) propensity score matching and fixed effects	(6) parallel pre-treatment trends test	(7) propensity score matching and dynamic least squares
D_t (dummy equal to 1 if firm was treated in year t and 0 if never treated)	0.0355***	0.0619***	0.0641***			0.0409***	0.0377***
	(0.0244)	(0.0210)	(0.0212)			(0.0213)	(0.0146)
D_{t-1} (treatment dummy lagged one year)			0.0596***				
			(0.0211)				
D_{t-2} (treatment dummy lagged two years)			0.0758***				
			(0.0240)				
$Prem*SE$ (wage premium for different labour categories)		0.0313***	0.0317***	0.0327***	0.0326***	0.0314***	0.0115***
		(0.0030)	(0.0030)	(0.0031)	(0.0031)	(0.0030)	(0.0027)
D_timing_t (dummy equal to 1 from the first year the firm was treated and so on successively, 0 if never treated)				0.1909***			
				(0.0278)			
D_dosage_t (dummy equal to 1 for all years from the first year the firm was treated until the year before the second treatment, 2 from the second year the firm was treated until the year before the third treatment and so on, 0 if never treated)					0.0938***		
					(0.0159)		
PD_1 (pre-treatment dummy equal to 1 for the first year before the firm was treated and 0 if never treated)						-0.0708***	
						(0.0245)	
PD_2 (pre-treatment dummy equal to 1 for the second year before the firm was treated and 0 if never treated)						-0.0642***	
						(0.0206)	
PD_3 (pre-treatment dummy equal to 1 for the third year before the firm was treated and 0 if never treated)						-0.1147***	
						(0.0227)	
$(w-p)\_{t-1}$ (real wages variable lagged one year)							0.7567***
							(0.0390)
$(w-p)\_{t-2}$ (real wages variable lagged two years)							0.0451***
							(0.0321)
$(w-p)\_{t-3}$ (real wages variable lagged three years)							0.0883***
							(0.0107)
Constant	13.9604***	13.7796***	13.7754***	13.7647***	13.7664***	13.7801***	1.4811***
	(0.0004)	(0.0171)	(0.0176)	(0.0182)	(0.0180)	(0.0172)	(0.1903)
Observations	12 450	12 450	12 450	12 450	12 450	12 450	12 349
R-squared	0.0003	0.0798	0.0816	0.0883	0.0889	0.0814	0.1712
Number of observations	1 626	1 626	1 626	1 626	1 626	1 626	1 620

Source: Prepared by the authors.
Note: *Statistically significant at 10%; **statistically significant at 5%; ***statistically significant at 1%.

3. The impact of *Costa Rica Provee* on labour demand

The results for the impact of *Costa Rica Provee* on labour demand (number of workers) are presented in tables 5 and 6. The table 5 results are for estimations using only the fixed effects method. From column 1 of table 5 it can be concluded that participation in *Costa Rica Provee* has a positive and significant impact on labour demand at beneficiary firms, given that the coefficient associated with the treatment variable, D_t, is positive and significant (0.1124).

Table 5

Impact of *Costa Rica Provee* programme on labour demand

(Fixed effects and cluster-robust standard errors)

Variable	(1) fixed effects	(2) fixed effects	(3) fixed effects	(4) fixed effects
D_t (dummy equal to 1 if firm was treated in year t and 0 if never treated)	0.1124***	0.1208***		
	(0.0264)	(0.0256)		
D_{t-1} (treatment dummy lagged one year)		0.1429***		
		(0.0269)		
D_{t-2} (treatment dummy lagged two years)		0.1398***		
		(0.0316)		
D_timing_t (dummy equal to 1 from the first year the firm was treated and so on successively, 0 if never treated)			0.0293***	
			(0.0347)	
D_dosage_t (dummy equal to 1 for all years from the first year the firm was treated until the year before the second treatment, 2 from the second year the firm was treated until the year before the third treatment and so on, 0 if never treated)				0.1616***
				(0.0209)
Constant	2.0003***	1.9965***	1.9916***	1.9915***
	(0.0005)	(0.0009)	(0.0014)	(0.0014)
Observations	26 082	26 082	26 082	26 082
R-squared	0.0009	0.0032	0.0054	0.0074
Number of observations	4 628	4 628	4 628	4 628

Source: Prepared by the authors.

Note: *Statistically significant at 10%; **statistically significant at 5%; ***statistically significant at 1%.

The data in column 2 indicate that the impact of participating in *Costa Rica Provee* arises during the initial year of treatment, as well as one and two years afterwards; the coefficients associated with the treatment variables D_t, $D_t - 1$ and $D_t - 2$ are positive and significant (0.1208, 0.1429 and 0.1398, respectively). On the other hand, when D_timing is substituted for the treatment variable and the dynamic (non-linear) effects of participation in *Costa Rica Provee* are analysed (column 3), it may be concluded that the longer a firm is treated, the greater the impact on labour demand. The coefficient associated with D_timing is positive and significant (0.2693).

Lastly, the results in column 4 for the treatment dosage (D_dosage) suggest that the more times a firm is treated under the *Costa Rica Provee* programme during the period analysed (2004-2011), the greater its labour demand. The coefficient associated with D_dosage is positive and significant (0.1616).

The results for the impact of *Costa Rica Provee* on labour demand, using only common support firms and controlling for fixed effects, are presented in table 6. These estimations are considered to be stronger because firms which are not good "clones" of beneficiary firms are eliminated from the control group by means of propensity score matching.

The results in table 6 confirm that *Costa Rica Provee* has a positive and significant impact on labour demand at beneficiary firms, as the coefficient associated with the treatment variable (D_t) is positive and significant (0.0958, column 1). The impact is observed in the year the treatment is applied and one and two years later. The values of the coefficients associated with treatment variables D_t, $D_t - 1$ and $D_t - 2$ are positive and significant (0.0984, 0.1117 and 0.0829, respectively).

When the dynamic results of treatment (D_timing) are analysed, a positive and significant coefficient is obtained (0.2081), indicating that a longer period of treatment has a greater impact on labour demand. In addition, the coefficient associated with dosage (D_dosage) is positive and significant (0.1062), indicating that successive treatments have a greater impact on the performance of beneficiary firms than a single treatment.

When it comes to testing whether the parallel pre-treatment trends assumption holds, the results in column 5 show that using fixed effects is a valid approach in this case. In fact, the results for all the coefficients associated with pre-treatment variables are non-significant, except for PD_3. But given that the significance of this last coefficient is very low and the first two pre-treatment variable coefficients are not significant, we consider there is strong evidence for accepting the parallel pre-treatment trends assumption.

Table 6

The impact of the *Costa Rica Provee* programme on labour demand

(Propensity score matching, fixed effects and cluster-robust standard errors)

Variable	(1) propensity score matching and fixed effects	(2) propensity score matching and fixed effects	(3) propensity score matching and fixed effects	(4) propensity score matching and fixed effects	(5) parallel pre-treatment trends test
D_t (dummy equal to 1 if firm was treated in year t and 0 if never treated)	0.0958***	0.0984***			0.0913***
	(0.0305)	(0.0299)			(0.0328)
$D_t - 1$ (treatment dummy lagged one year)		0.1117***			
		(0.0343)			
$D_t - 2$ (treatment dummy lagged two years)		0.0829**			
		(0.0350)			
D_timing_t (dummy equal to 1 from the first year the firm was treated and so on successively, 0 if never treated)			0.2081***		
			(0.0461)		
D_dosage_t (dummy equal to 1 for all years from the first year the firm was treated until the year before the second treatment, 2 from the second year the firm was treated until the year before the third treatment and so on, 0 if never treated)				0.1062***	
				(0.0217)	
PD_1 (pre-treatment dummy equal to 1 for the first year before the firm was treated and 0 if never treated)					0.0058
					(0.0520)
PD_2 (pre-treatment dummy equal to 1 for the second year before the firm was treated and 0 if never treated)					0.0023
					(0.0381)
PD_3 (pre-treatment dummy equal to 1 for the third year before the firm was treated and 0 if never treated)					-0.0770*
					(0.0409)
Constant	2.3123***	2.3097***	2.3054***	2.3065***	2.3127***
	(0.0005)	(0.0011)	(0.0019)	(0.0015)	(0.0009)
Observations	12 450	12 450	12 450	12 450	12 450
R-squared	0.0009	0.0025	0.0045	0.0052	0.0011
Number of observations	1 626	1 626	1 626	1 626	1 626

Source: Prepared by the authors.

Note: *Statistically significant at 10%; **statistically significant at 5%; ***statistically significant at 1%.

4. The impact of *Costa Rica Provee* on the probability of exporting

The results obtained with equation (6) for the impact of *Costa Rica Provee* on the probability of exporting are presented in tables 7 and 8. Table 7 shows the results of the analysis using a linear probability model with only the fixed effects approach. The data in column 1 of table 7 show that the coefficient associated with the treatment dummy (D_t) is positive and significant (0.0315), indicating that the participation of SMEs in the *Costa Rica Provee* programme increases the probability of exporting for beneficiary firms as compared to firms in the control group. In addition, participation in *Costa Rica Provee* seems to have an impact on the export performance of beneficiary firms not only in the year they receive the treatment but also two years later. The coefficients associated with these effects are positive and significant (0.0372 and 0.0942, respectively), as shown in column 2 of table 7.

Table 7

Impact of the *Costa Rica Provee* programme on the probability of exporting: linear probability model

(Fixed effects and cluster-robust standard errors)

Variable	(1) fixed effects	(2) fixed effects	(3) fixed effects	(4) fixed effects
D_t (dummy equal to 1 if firm was treated in year t and 0 if never treated)	0.0315*	0.0372**		
	(0.0171)	(0.0170)		
$D_t - 1$ (treatment dummy lagged one year)		0.0046		
		(0.0164)		
$D_t - 2$ (treatment dummy lagged two years)		0.0942***		
		(0.0191)		
D_timing_t (dummy equal to 1 from the first year the firm was treated and so on successively, 0 if never treated)			0.0613***	
			(0.0174)	
D_dosage_t (dummy equal to 1 for all years from the first year the firm was treated until the year before the second treatment, 2 from the second year the firm was treated until the year before the third treatment and so on, 0 if never treated)				0.0585***
				(0.0123)
Constant	0.0981***	0.0969***	0.0962***	0.0948***
	(0.0003)	(0.0005)	(0.0007)	(0.0008)
Observations	26 062	26 062	26 062	26 062
R-squared	0.0005	0.0035	0.0020	0.0068
Number of observations	4 625	4 625	4 625	4 625

Source: Prepared by the authors.
Note: *Statistically significant at 10%; **statistically significant at 5%; ***statistically significant at 1%.

Another interesting result from this exercise is that the longer a firm has been treated, the greater the impact on its probability of exporting. The coefficient associated with the dynamic effect of intervention (D_timing), shown in the third column of table 7, is positive and significant (0.0613). In addition, it seems that the more times a firm participates in the *Costa Rica Provee* programme, the more its probability of exporting increases. The coefficient associated with the dosage treatment variable (D_dosage) is positive and significant (0.0585). In other words, it seems that the greater the number of linkages with multinational corporations created by this programme for beneficiary firms, the greater their probability of placing their products in international markets.

When the propensity score matching and fixed effects approaches are used together to estimate the impact of *Costa Rica Provee* on the probability of exporting by beneficiary firms, the results are similar to those obtained when only the fixed effects approach is used, although these new results are more robust than those shown in table 7. Thus, as shown in table 8, all the coefficients associated with treatment variables (D_t, D_timing and D_dosage) turn out to be positive and significant in this case (0.0485, 0.0891 and 0.0676, respectively), confirming the importance of SMEs' participation in the *Costa Rica Provee* programme for improving their probability of exporting.

Finally, we tested whether the parallel pre-treatment trend assumption was valid and found that using fixed effects was an invalid approach in this case. In fact, the results for all the coefficients associated with pre-treatment variables ($PD\_$) in column 5 are significant, except in the case of PD_3, where the coefficient is very significant. For this reason, we estimate a propensity score matching with a least squares dynamic model, the results of which are presented in column 6. The coefficient associated with the treatment variable (D) in this specification is positive and significant (0.0586), so we can conclude that firms' participation in *Costa Rica Provee* has a positive and significant impact on the probability of exporting, with the probability of exporting being 5.9 percentage points higher for a treated firm than for an untreated firm.

Table 8

The impact of the *Costa Rica Provee* programme on exports: linear probability model

(Propensity score matching, fixed effects, least squares dynamic and cluster-robust standard errors)

Variable	(1) propensity score matching plus fixed effects	(2) propensity score matching plus fixed effects	(3) propensity score matching plus fixed effects	(4) propensity score matching plus fixed effects	(5) parallel pre-treatment trends test	(6) propensity score matching plus dynamic least squares
D_t (dummy equal to 1 if firm was treated in year t and 0 if never treated)	0.0471	0.0485*			0.0428	0.0586**
	(0.0290)	(0.0285)			(0.0295)	(0.0233)
$D_t - 1$ (treatment dummy lagged one year)		-0.0071				
		(0.0271)				
$D_t - 2$ (treatment dummy lagged two years)		0.1111***				
		(0.0263)				
D_timing_t (dummy equal to 1 from the first year the firm was treated and so on successively, 0 if never treated)			0.0891***			
			(0.0291)			
D_dosage_t (dummy equal to 1 for all years from the first year the firm was treated until the year before the second treatment, 2 from the second year the firm was treated until the year before the third treatment and so on, 0 if never treated)				0.0676***		
				(0.0190)		
PD_1 (pre-treatment dummy equal to 1 for the first year before the firm was treated and 0 if never treated)					0.0138	
					(0.0344)	
PD_2 (pre-treatment dummy equal to 1 for the second year before the firm was treated and 0 if never treated)					0.0039	
					(0.0277)	
PD_3 (pre-treatment dummy equal to 1 for the third year before the firm was treated and 0 if never treated)					-0.0876***	
					(0.0310)	

Table 8 (concluded)

Variable	(1) propensity score matching plus fixed effects	(2) propensity score matching plus fixed effects	(3) propensity score matching plus fixed effects	(4) propensity score matching plus fixed effects	(5) parallel pre-treatment trends test	(6) propensity score matching plus dynamic least squares
$exp\__{t}-1$ (export variable lagged one year)						0.5454***
						(0.0236)
$exp\__{t}-2$ (export variable lagged two years)						0.2561***
						(0.0299)
$exp\__{t}-3$ (export variable lagged three years)						0.1158***
						(0.0212)
Constant	0.1166***	0.1154***	0.1138***	0.1127***	0.1170***	0.0172***
	(0.0005)	(0.0009)	(0.0012)	(0.0013)	(0.0006)	(0.0014)
Observations	12 450	12 450	12 450	12 450	12 450	12 450
R-squared	0.0009	0.0048	0.0036	0.0091	0.0023	0.0533
Number of observations	1 626	1 626	1 626	1 626	1 626	1 626

Source: Prepared by the authors.
Note: *Statistically significant at 10%; **statistically significant at 5%; ***statistically significant at 1%.

VI. Analysis and conclusions

This study has attempted to help close the knowledge gap by assessing the impact of one specific programme (*Costa Rica Provee*) that promotes backward linkages between multinationals and local firms (SMEs) in a developing country, namely Costa Rica. Impacts were estimated on the assumption that beneficiary firms were trying to maximize their profits and that *Costa Rica Provee* aimed to increase these firms' productivity, and three performance variables were taken: real average wages, labour demand and the probability of exporting.

The *Costa Rica Provee* programme was found to have positive and significant impacts on the performance of beneficiary firms, specifically on their real average wages, labour demand and probability of exporting. Average wages paid by firms treated by *Costa Rica Provee* were found to be higher than those paid by untreated firms (0.04) and labour demand was found to be higher than at untreated firms (0.10). These benefits were observed up to two years beyond the initial year in which firms participated in the programme. The amount of time elapsing from the initial participation in *Costa Rica Provee* also had a positive impact on the performance of beneficiary firms, as did the number of times that SMEs were able to generate linkages with MNCs. The probability of exporting was found to be about 5.9 percentage points higher for a treated firm than for an untreated firm.

While earlier evaluations of SME programmes were pessimistic about their impacts, recent studies have found positive impacts of programme participation on intermediate outcomes but mixed results for impacts on firm performance (López-Acevedo and Tan, 2010). Our study points to positive impacts on firm performance.

This finding is in line with the result obtained by Tan (2010) in Chile, Duque and Muñoz (2010) in Colombia and López-Acevedo and Tinajero (2010) in Mexico, as they all found positive impacts on firm performance from SME programmes that included backward linkages. The contribution of our research is the evaluation of one specific backward linkages programme. Additionally, our evidence on positive programme impacts over time adds to existing evidence from Duque and Muñoz (2010) in Colombia.

The limitations of this study are a stimulus to future research. On the demand side, it would be interesting to explore MNC characteristics such as the size and age of firms, their mode of establishment and the type and nature of their production processes. On the supply side, absorptive capacity is a matter that could be worth investigating. This combination of factors influences the nature (in terms of type, depth and quality) of linkages and can be expected to impact the outcomes of *Costa Rica Provee*. The way backward linkages are measured also matters.

Our study also suggests policy implications. As mentioned previously, success in attracting FDI does not automatically generate the benefits of backward linkages. Costa Rica has been successful in attracting FDI (World Economic Forum, 2013), but its development model has weaknesses, including limited success in tying the new economy (dynamic sectors such as MNCs) to what can be a sluggish old economy, as in the case of some SMEs. Programmes like *Costa Rica Provee* can help with this, and it is important to maintain an MNC demand-driven programme and help local suppliers rise in the value chain. Indeed, the scope of the programme could be expanded.

Bibliography

Alfaro, L. and others (2004), “FDI and economic growth, the role of local financial markets”, *Journal of International Economics*, vol. 64, No. 1, Amsterdam, Elsevier.

Amendolagine, V. and others (2013), “FDI and local linkages in developing countries: evidence from Sub-Saharan Africa”, *World Development*, vol. 50, Amsterdam, Elsevier.

Barrios, S., H. Görg and E. Strobl (2011), “Spillovers through backward linkages from multinationals: measurement matters!”, *European Economic Review*, vol. 55, No. 6, Amsterdam, Elsevier.

Bernal, R. and X. Peña (2011), *Guía práctica para la evaluación de impacto*, Bogota, University of the Andes.

Crespi, G. and others (2011), “Evaluating the impact of science and technology programs: a methodological toolkit”, *Technical Notes*, No. IDB-TN-333, Washington, D.C., Inter-American Development Bank [online] http://publications.iadb.org/handle/11319/5396?locale-attribute=en.

Crespo, N. and M. P. Fontoura (2007), “Determinant factors of FDI spillovers – what do we really know?”, *World Development*, vol. 35, No. 3, Amsterdam, Elsevier.

Damijan, J. P. and others (2013), “Impact of firm heterogeneity on direct and spillover effects of FDI: micro-evidence from ten transition countries”, *Journal of Comparative Economics*, vol. 41, No. 3, Amsterdam, Elsevier.

Duque, J. F. and M. Muñoz (2010), “Evaluating SME support programs in Colombia”, *Impact Evaluation of SME Programs in Latin America and Caribbean*, G. López-Acevedo and H. Tan (eds.), Washington, D.C., World Bank.

Ferragina, A. M. and F. Mazzotta (2014), “FDI spillovers on firm survival in Italy: absorptive capacity matters!”, *Journal of Technology Transfer*, vol. 39, No. 6, Springer.

Flores, J. J. (2011), “Backward linkages in Costa Rica: the role of the linkages promotion agency. Evidence from 2001 to 2008”, thesis, Rotterdam, Erasmus University [online] http://thesis.eur.nl/pub/10139/.

Giroud, A., B. Jindra and P. Marek (2012), “Heterogeneous FDI in transition economies – A novel approach to assess the developmental impact of backward linkages”, *World Development*, vol. 40, No. 11, Amsterdam, Elsevier.

Görg, H., A. Hanley and E. Strobl (2011), “Creating backward linkages from multinationals: is there a role for financial incentives?”, *Review of International Economics*, vol. 19, No. 2, Wiley.

Groote, R. (2005), “Costa Rica. Proyecto de Desarrollo de Proveedores para Empresas Multinacionales de Alta Tecnología. ATN/ME-6751-CR. Final Evaluation”, Washington, D.C., Inter-American Development Bank (IDB).

Hallin, C. and L.C. Holmström (2012), “Revisiting the external impact of MNCs: an empirical study of the mechanisms behind knowledge spillovers from MNC subsidiaries”, *International Business Review*, vol. 21, No. 2, Amsterdam, Elsevier.

Javorcik, B. (2004), “Does foreign direct investment increase the productivity of domestic firms? In search of spillovers through backward linkages”, *American Economic Review*, vol. 94, No. 3, Nashville, Tennessee, American Economic Review.

Kim, P. H. and M. Li (2014), "Injecting demand through spillovers: foreign direct investment, domestic socio-political conditions, and host-country entrepreneurial activity", *Journal of Business Venturing*, vol. 29, No. 2, Amsterdam, Elsevier.
Kiyota, K. and others (2008), "Reconsidering the backward vertical linkages of foreign affiliates: evidence from Japanese multinationals", *World Development*, vol. 36, No. 8, Amsterdam, Elsevier.
Liu, B. J. (2011), "MNEs and local linkages: evidence from Taiwanese affiliates", *World Development*, vol. 39, No. 4, Amsterdam, Elsevier.
López-Acevedo, G. and H. W. Tan (eds.) (2010), *Impact Evaluation of SME Programs in Latin America and Caribbean*, Washington, D.C., World Bank [online] http://documents.worldbank.org/curated/en/319161468337915156/pdf/526680ENGLISH010evaluation01PUBLIC1.pdf.
López-Acevedo, G. and M. Tinajero (2010), "Evaluating SME support programs in Mexico", *Impact Evaluation of SME Programs in Latin America and Caribbean*, G. López-Acevedo and H. Tan (eds.), Washington, D.C., World Bank.
Padilla-Pérez, R. and Y. Gaudin (2014), "Science, technology and innovation policies in small and developing economies: the case of Central America", *Research Policy*, vol. 43, No. 4, Amsterdam, Elsevier.
Paus, E. and K. Gallagher (2008), "Missing links: foreign investment and industrial development in Costa Rica and Mexico", *Studies in Comparative International Development*, vol. 43, No. 1.
Prashantham, S. and C. Dhanaraj (2014), "MNE ties and new venture internationalization: exploratory insights from India", *Asia Pacific Journal of Management*, vol. 32, No. 4, Springer.
Rosenbaum, P. and D. Rubin (1983), "The central role of the propensity score in observational studies for causal effects", *Biometrika*, vol. 70, No. 1, Oxford, Oxford University Press.
Saggi, K. (2002), "Trade, foreign direct investment and international technology transfer: a survey", *The World Bank Research Observer*, vol. 17, No. 1, Washington, D.C., World Bank.
Smeets, R. (2008), "Collecting the pieces of the FDI knowledge spillovers puzzle", *The World Bank Research Observer*, vol. 23, No. 2, Washington, D.C., World Bank.
Spencer, J. W. (2008), "The impact of multinational enterprise strategy on indigenous enterprises: horizontal spillovers and crowding out in developing countries", *Academy of Management Review*, vol. 33, No. 2, Academy of Management.
Tan, H. (2010), "Evaluating SME support programs in Chile", *Impact Evaluation of SME Programs in Latin America and Caribbean*, G. López-Acevedo and H. Tan (eds.), Washington, D.C., World Bank.
Wanga, C. and others (2012), "Reconceptualizing the spillover effects of foreign direct investment: a process-dependent approach", *International Business Review*, vol. 21, No. 3, Amsterdam, Elsevier.
World Economic Forum (2013), "Foreign Direct Investment as a Key Driver for Trade, Growth and Prosperity: The Case for a Multilateral Agreement on Investment" [online] http://www3.weforum.org/docs/GAC13/WEF_GAC_GlobalTradeFDI_FDIKeyDriver_Report_2013.pdf.
Zhang, Y. and others (2010), "FDI spillovers in an emerging market: the role of foreign firms´ country origin diversity and domestic firms´ absorptive capacity", *Strategic Management Journal*, vol. 31, No. 9, Wiley.

Annex A1

Estimating the impact of programme participation on real wages

Let us assume a modified Cobb-Douglas production function:

$$Y = K^{\alpha} L^{1\text{-}\alpha} exp(\delta D + \varepsilon) \tag{1}$$

where Y is output, K capital, L the number of workers and D a dummy variable equal to 1 if the firm participated in the programme and 0 otherwise. The coefficient on this dummy variable allows us to test whether participation in an innovation or linkage development programme such as PROPYME or *Costa Rica Provee* affects total factor productivity.

Under a profit maximization assumption, the first-order condition tells us that:

$$PK^{\alpha}(1-\alpha)L^{\text{-}\alpha} exp(\delta D + \varepsilon) - W = 0 \tag{2}$$

Taking logs and arranging terms, we get:

$$w - p = ln(1-^{\alpha}) + ^{\alpha}(k - l) + \delta D + \varepsilon \tag{3}$$

Thus, real wages depend on $(k-l)$ and total factor productivity $(\delta D + \varepsilon)$.

On the other hand, if we add a mix of workers with different qualities, we get:

$$L^{*} = L_1 + \theta_2 L_2 + \ldots = L(1+q) \tag{4}$$

But it is likely that labour input is the result of the services provided by workers of different qualities. Let us replace L by effective labour L^{*}. Then equation (2) can be expressed as follows:

$$PK^{\alpha}(1-\alpha)(L)^{\text{-}\alpha}(1+q)^{1\text{-}\alpha} exp(\delta D + \varepsilon) - W = 0 \tag{5}$$

Taking logs:

$$w - p = ln(1-\alpha) + \alpha(k - l) - (1-\alpha)ln(1-q) + \delta D + \varepsilon \tag{6}$$

Supposing that there are only two types of workers, skilled and unskilled, we have a difference in productivity θ and a premium $PREM$, so that $\theta - 1 = PREM$.

Let WO and WE be the average wages of unskilled and skilled workers and LO and LE the number of unskilled and skilled workers, respectively. The average wage of the firm can be written as:

$$W = (WO^{*}LO + WE^{*}LE)/(LO + LE) \tag{7}$$

This expression is equal to:

$$W = WO^{*}(1 - LE/(LO + LE)) + WE^{*}(LE/(LO + LE)) = WO^{*}(1 + (WE + WO - 1)(LE/(LO + LE)) \tag{8}$$

Let $PREM$=(1+(WE+ WO–1) be the skilled worker premium. Additionally, let us define $(LE/(LO+LE))$ as the share of skilled workers and use SE as the abbreviation for this term (i.e. $SE=(LE/(LO+LE))$). Taking logs on (7) and substituting terms, we get:

$$lnW \approx lnWO + (PREM^{*}SE) \tag{9}$$

The principal idea here is that the average wage of the firm is equal to the wage of unskilled workers plus a term that takes into account the premium charged by skilled workers times the share of this type of worker in the total number of workers.

Let $L^{*}= LO+\theta^{*}LE=L(LE/L+\theta\ LE/L)$

where $L=LO+LE$

given that $LO/L=1–LE/L$, then:

$$L^{*}= LO(1+(\theta-1)SE) \tag{10}$$

Taking logs:

$$l^{*}=(\theta-1)SE \tag{11}$$

Therefore, from (4) we know that $(\theta-1)SE=q$.

From the discussion above, we know that when the firm is maximizing profits, we also have a premium called $PREM=\theta-1$.

$$q=(\theta-1)SE=PREM^{*}SE \tag{12}$$

This means that differences in productivity are equal to the wage premium. Substituting (12) in (6), we have:

$$w-p=ln(1-\alpha)+\alpha(k-l)-(1-\alpha)(PREM^{*}SE)+\delta D+\varepsilon \tag{13}$$

To operationalize (13), we can write:

$$(w-p)=\beta_0+\beta_1(k-l)+\beta_2(PREM^{*}SE)+\beta_3 D+\tau \tag{14}$$

As discussed in the body of this paper, equation (14) can be estimated using a combination of two techniques: fixed effects and propensity score matching. Due to problems with data availability for capital (K), we assume that fixed effects let us control the effect of $(K-l)$. Although we are assuming that the effect of the programme occurs in the same year as the intervention, we will test whether that is actually the case or whether we have to wait one or two years after the firm received the treatment to see any effect. This is the reason for including lags for variable D.

For the purposes of estimation, equation (14) can be expressed as follows:

$$(w-p)_{it}=\beta_0+\beta_1(PREM^{*}SE)_{it}+\beta_2 D_{it}+\beta_3 D_{it\text{-}1}+\beta_4 D_{it\text{-}2}+\varepsilon_{it} \tag{15}$$

Finally, given data availability constraints, we use as a proxy for $PREM^{*}SE$ the ratio between the firm's average wages and the industry's average wages for each year included in the analysis, all of them in nominal values.

Annex A2

Estimating the impact of programme participation on labour demand

Let us assume a modified Cobb-Douglas production function:

$$Y = K^{\alpha} L^{1-\alpha} exp(\delta D + \varepsilon) \tag{1}$$

where Y is output, K capital, L the number of workers and D a dummy variable equal to 1 if the firm participated in the programme and 0 otherwise. In this formulation, participation in an innovation or linkage development programme such as PROPYME or *Costa Rica Provee* might affect total factor productivity.

From the first-order conditions of profit maximization, and taking logs, we get:

$$p + \alpha k + ln(1-\alpha) - \alpha l + \delta D + \varepsilon - w = 0 \tag{2}$$

where p is the price of the output produced by the firm (in logs).

Arranging terms, we have:

$$\begin{gathered} l = \frac{p}{\alpha} + \frac{\alpha k}{\alpha} + \frac{ln(1-\alpha)}{\alpha} + \frac{\delta D + \varepsilon}{\alpha} + \frac{w}{\alpha} \\ l = \frac{1}{\alpha} p + k + \frac{1}{\alpha} ln(1-\alpha) + \frac{1}{\alpha}(\delta D + \varepsilon) - \frac{1}{\alpha} w \\ l = \frac{1}{\alpha} ln(1-\alpha) + k - \frac{1}{\alpha}(\delta D + \varepsilon) \end{gathered} \tag{3}$$

As discussed in the main body of this paper, equation (3) can be estimated using a combination of two techniques: propensity score matching and fixed effects. Due to problems with data availability for capital (K), we assume that fixed effects let us control the effect of this variable.Once again, we assume that the effect of the programme arises in the same year as the intervention, but also test to determine whether this is actually the case or whether we have to wait one or two years after the firm received the treatment to see any effect. This is the reason for including lags for variable D. Thus, equation (3) can be expressed as follows:

$$l_{it} = \gamma_0 - \gamma_1 (w-p)_{it} + \gamma_2 D_{it} + \gamma_3 D_{it-1} + \gamma_4 D_{it-2} + \sigma_{it} \tag{4}$$

Annex A3

Variable definitions (in alphabetical order)

Chemicals: Dummy variable taking the value 1 if the firm has economic activity in chemicals and 0 otherwise.

Employment: Number of employees hired by the firm per year.

Exports (t): Dummy variable taking the value 1 if the firm has exported during year t and 0 otherwise.

Geographic location: Dummy variable taking the value 1 if the firm is located in province i of Costa Rica (San José, Cartago or Heredia) and 0 otherwise. We take six of the seven provinces in Costa Rica (Alajuela, Cartago, Guanacaste, Heredia, Puntarenas and San José).

Legal status: Dummy variable taking the value 1 if the firm is legally registered as a commercial entity and 0 otherwise.

Lithographic: Dummy variable taking the value 1 if the firm engages in lithographic processes and 0 otherwise.

Manufacturing: Dummy variable taking the value 1 if the firm engages in manufacturing and 0 otherwise.

Wages or salaries in real terms: Total amount of wages and salaries paid by the firm per year, deflated by the industrial price index at the two-digit level of the Standard Industrial Classification (SIC) in the case of manufacturing firms and by the consumer price index otherwise to obtain real wages.

Job satisfaction in Chile: geographic determinants and differences

Luz María Ferrada

Abstract

Logit, binary and multinomial models are used in this study to determine the impact of objective and perceived working conditions on workers' job satisfaction. Possible differences between job satisfaction in the Metropolitan Region and in other areas of Chile are also explored. The data used in this analysis are drawn from the first National Survey on Employment, Work, Health and Quality of Life of Workers in Chile (ENETS). Wage levels were found to have a positive impact across the board, while residence in an area other than the Metropolitan Region also had a significantly positive effect. These results were corroborated using matching techniques. The finding that subjective perceptions have a great deal of explanatory power and that their impact outweighs the influence of objective conditions may be of interest in the areas of both public policy and business administration.

Keywords

Employment, working conditions, wages, job satisfaction, measurement, surveys, econometric models, Chile

JEL classification

J280, M540, R230

Author

Luz María Ferrada is an academic with the Administrative and Economic Sciences Department of the Universidad de Los Lagos, Chile. Email: lferrada@ulagos.cl.

I. Introduction

The steady growth of Chile's gross domestic product (GDP) since the 1990s has been coupled with an expanding demand for labour and rising real wages. Data compiled by the National Institute of Statistics for 1993-2009 show increases of 1,802,064 in the number of jobs (INE, n/d) and of 1%, on average, in annual wages over the variation in the consumer price index (CPI) (INE, n/d). A number of studies indicate, however, that these trends have not narrowed the wage gaps between different regions within the country and that job quality remains very poor. This situation raises questions as to how satisfied with their jobs Chilean workers are, what factors influence their level of job satisfaction the most and whether the results differ substantially across regions.

The international literature indicates that the way in which workers perceive their jobs has a strong impact on various aspects of their lives. This may be the most convincing justification for the large number of international studies that have been conducted in this field. Economic research in this field is quite recent, however, and especially so in Chile. Most of the research in this area has been conducted by psychologists, although studies focusing on human resource management —mainly in specific industries— have also been carried out. The chief contribution made by this study is that it deals with a subject that has been studied very little in Chile and incorporates a geographic dimension, thereby breaking new ground in this field of research in Chile.

Job satisfaction is defined in various ways in different fields of study. One generic approach that has been used in many studies involves conceptualizing it as an emotional state produced by a subject's subjective perception of his or her on-the-job experiences; this emotional state has been described as the level of well-being that a person derives from his or her job.[1] It has also been defined as a reflection of the level of utility that a job provides, i.e., how well it meets a worker's needs or expectations and contributes to his or her quality of life. The determinants of job satisfaction therefore include both objective and subjective factors.

The specific objectives of this study are as follows: (i) to evaluate the impact on job satisfaction of variables associated with objective working conditions and with perceptions regarding the organization of work and the working environment; and (ii) to measure differences in job satisfaction between the Metropolitan Region and other areas of Chile.

The information used in this study is drawn from the National Survey on Employment, Work, Health and Quality of Life of Workers in Chile (ENETS) round carried out in 2009-2010, whose results cover the various regions in the country, provide information on both objective and subjective aspects of working conditions and include the data provided by responses to a number of questions about job satisfaction.

This study is composed of five sections, including this introduction. The second section provides a review of the literature on the subject. The third describes the methodology that was used, while the fourth presents the findings regarding the study's specific objectives. The fifth and final section provides an overview of the study's main conclusions.

[1] See [online] http://www.insht.es/InshtWeb/Contenidos/Documentacion/TextosOnline/GuiasMonitor/Psicosociologia/I/Ficheros/psi23.pdf.

II. The literature

1. Approaches

Research on job satisfaction has been undertaken in a variety of fields, including psychology, human resource management and economics, each of which uses its own approaches and tools. Most of the studies that have focused on how different factors influence people's assessments of their own job satisfaction have taken a psychological approach, however; this field often makes use of a dual-factor model that distinguishes between extrinsic (or "hygiene") and intrinsic (or "motivation") factors. The former help to reduce dissatisfaction while the latter help to increase the level of satisfaction. Extrinsic factors shape the work environment, while intrinsic factors reflect how a person experiences the work environment (Bòria-Reverter, Crespi-Vallbona and Mascarilla-Miró, 2012); a number of economic studies on the subject have also used this model.

Economic research on job satisfaction is quite recent and has probably been associated with the epistemological development of work in this field. Early in the twentieth century, owing to the strong influence exerted by the industrial revolution, some researchers approached the subject of labour from the perspective of an activity that produces wealth (in the tradition of Adam Smith), while others viewed it in terms of the time required to make a product (in the tradition of Karl Marx) and still others visualized it as the activity that confers value on merchandise (in the tradition of David Ricardo) (Olarte, 2011). However, studies done in the 1960s and thereafter were the ones that focused on human capital (Mincer, 1958; Becker, 1964), and research findings in other fields sparked interest in the study of job satisfaction. This was when the idea that job satisfaction could be an important factor in determining productivity, growth and stability came to prominence. The research done by Freeman (1977) and those who followed in his footsteps marked a milestone in this area of inquiry.

Labour management thus became an increasingly fundamental consideration in the field of business administration. The evolving views regarding this factor are illustrated by three different approaches to the subject. First, Taylor focused on models for improving individual and collective physical effort as a means of boosting labour productivity (Chiavenato, 2006). Later, the focus shifted to analyses of organizational behaviour (Stephen and Timothy, 2013) that posit that greater motivation in the workplace and job satisfaction could lead to the achievement of corporate objectives (Herzberg, Mausner and Snyderman, 1959; Herzberg, 1962). A third ("new management") approach centres around human resources management aimed at shaping a management style that will help employers and employees to build a collaborative relationship whereby workers will shoulder responsibility for corporate objectives and will identify with those goals because they have come to share ownership of the company's core values (Narbona, 2012). This approach seeks to foster a subjective relationship with work that goes beyond the bounds set by employment contracts and engenders happiness. These authors also see a subtle form of control taking shape that prospers in an environment marked by labour flexibility and low levels of unionization (Narbona, 2012).

2. Measuring job satisfaction

Measuring job satisfaction is a challenging undertaking because it involves dealing with a number of elements that are difficult to evaluate. There are various possible approaches to this task, as well. The most common approach is to posit that job satisfaction can be estimated by gauging the utility of a job for the person who holds that job (Olarte, 2011) and that, therefore, a good proxy variable is the wage that the worker is willing to accept. Viewed from this standpoint, wages will be determined both by the supply of human capital and by the features of the work involved (higher qualifications, more personal skills, greater dedication, etc.) and the working environment (mobility within the company, unions, networking, job opportunities, etc.).

However, a job's utility for a given individual can also be evaluated on the basis of job quality and the quality of life that it affords. Much of the Latin American literature on this subject makes use of the decent work approach developed by the International Labour Organization (ILO) (ILO, 1999), which defines jobs as being of quality if, on the basis of objective factors, they are found to be jobs that are performed under conditions of freedom, equity, security and human dignity, if employees' rights are protected and if the jobs generate an adequate income and afford adequate social protection. Farné (2003) uses four yardsticks to measure the quality of working conditions: income, workday, social security and type of employment contract. Other researchers have used other approaches, however, with one example being the capability approach (Sen, 1999), which focuses on a person's functionalities, i.e., a set of personal characteristics (sex, age, level of education and others) and employment-related characteristics (type of contract, workday and others) that are associated with their job and that are converted into given capabilities (the ability to generate a high level of income or the ability to remain employed, for example). In line with this approach, people are seen as valuing those capabilities because they enable them to achieve a higher level of well-being (Sehnbruch, 2008). In other words, a given person may have a job (a good), but that good is going to be seen as being of quality only if it can be converted into a functionality that is valued by the individual concerned. Thus, the particular features of that job, in combination with the person's needs and characteristics, is what will determine the capabilities which that person can acquire by performing the job (Sehnbruch, 2012). Other authors have defined the concept of job satisfaction in ways that explicitly include subjective aspects. Reinecke and Valenzuela (2000), for example, describe it as the set of work-related factors that influence workers' level of well-being in economic, social, psychological and health-related terms (Farné, 2012).

Another group of research papers evaluate job satisfaction directly on the basis of the views expressed by jobholders who respond to survey questions which ask them to rate their degree of satisfaction or lack thereof on a set scale. A number of different questionnaires and rating scales have been designed for this purpose (Olarte, 2011).

Finally, some studies have used a multivariate analysis to gauge relative job satisfaction. In order to do so, they construct a synthetic indicator that incorporates various objective and subjective dimensions of the concept (Somarriba and others, 2010). These are the extrinsic and intrinsic factors used in the dual-factor model. They include individual traits that an employer cannot modify but that nonetheless influence a person's relative level of job satisfaction, such as demographic and residence variables, among others. Another type of factor is aspirational (e.g., self-esteem); these aspirations are never fully realized and therefore act as ongoing motivational forces (McGregor, 1981) that are related to expectations and life history.

In this study, job satisfaction is analysed on the basis of wage workers' perceptions as reported on a seven-point scale when they respond to the question: "How do you feel about your job?" (question D1h in ENETS module D). The question of how much certain objective and subjective factors influence those perceptions is then explored.

3. Determinants of job satisfaction

One of the aspects of job satisfaction that has been studied the most is its relationship with wages. From a utilitarian standpoint, the two variables are expected to be positively correlated, and there is even an expectation that there will be a causative relationship whereby a higher wage will translate into a higher level of job satisfaction. The results do not always bear this out, however; for example, Clark and Oswald (1996) find just the opposite effect, while others (Bòria-Reverter, Crespi-Vallbona and Mascarilla-Miró, 2012) posit that some workers may be unhappy with their wage level but are nevertheless satisfied with their jobs in general. Yet Borra and Gómez (2012) and Farné and Vergara (2007) do find a positive impact. It may be that the outcome depends on other variables, such as being higher up or lower down in the income distribution. It is thought that what usually happens is that people who accept lower wages exhibit greater satisfaction because they value the fact that they are employed at all, regardless of the wage level. Card and others (2010) note that workers' responses may also be influenced by comparisons of relative wage levels and that these ratios are not linear, even though Clark, Nicolai and Westergård-Nielsen (2007) indicate that wage comparisons may have a positive effect on job satisfaction if a worker believes that this comparison provides a basis for an expectation of higher future levels of income.

Analyses of other objective factors yield differing results. With regard to working hours and the type of employment contract, Gamero (2003) finds that, in general, workers in Spain (particularly men) do not prefer part-time work and that they have no preference between private or public contracts, even though other studies, including that of De Vries and others (2015), point to a preference for public-sector jobs. Booth and Van Ours (2007) use a panel survey of British households and find differences between men and women, with no improvement in men's job satisfaction related to working hours, whereas there is an improvement in the case of women, who prefer part-time work.

Other determinants that have been studied include job stability (Clark and Postel-Vinay, 2009; Bòria-Reverter, Crespi-Vallbona and Mascarilla-Miró, 2012) and opportunities for promotion. The literature indicates that workers feel more unhappy when they have no upward mobility even if they are given a raise (Grund and Sliwka, 2007).

As noted by Bòria-Reverter, Crespi-Vallbona and Mascarilla-Miró (2012), in this field it is not enough to look at just wages or a worker's place in the corporate hierarchy; intrinsic variables such as opportunities for participation and a sense of being useful have to be taken into account as well. Organizational factors also play a part; for example, a correlation has been found between job satisfaction and a positive organizational environment (Juárez-Adauta, 2012) and good interpersonal relationships (Juárez-Adauta, 2012).

It has also been found that perceptions of job satisfaction or dissatisfaction are a response that may be influenced by personal characteristics, including sociodemographic factors such as age, marital status, sex and level of education. Some studies have reported that, contrary to what had been expected, women have greater job satisfaction than men (Clark, 1997), although the differential disappears among more highly educated young people. However, Sloane and Williams (2000) contend that women's greater satisfaction with their jobs is not associated with the types of jobs that they hold or their pay levels but instead is attributable to innate gender-related differences. Another possible explanation is that women have fewer employment opportunities and therefore are happier when they do manage to secure a job.

People who are less educated and have less job experience are thought to have fewer expectations of being able to find a different job, and their level of dissatisfaction with a lower level of achievement is therefore less; this is also true of wage levels (Belfield and Harris, 2002; Lévy-Garboua, Montmarquette and Simonnet, 2007).

Another variable that has been studied a great deal is union membership. Interestingly —and surprisingly— membership in a union is associated with less job satisfaction (Freeman, 1977). This may be because unionized workers have higher expectations by virtue of their union membership or because the people who belong to unions have a higher threshold of dissatisfaction and have joined a union in the hope of improving that perception. It has also been argued that unionization may have an endogenous selection effect whereby unionization could account for the level of job satisfaction and, at the same time, be accounted for by it (Bryson, Cappellari and Lucifora, 2004). Others have said that union membership is endogenous and lowers job satisfaction, but being in a union is entirely exogenous and has a positive impact on job satisfaction (Rodríguez-Gutiérrez and Prieto-Rodríguez, 2004).

It is also possible that the impact of these factors on job satisfaction differs from one location to the next, probably because of differences in surrounding conditions that influence people's perceptions of their own level of satisfaction. Pouliakas and Theodossiou (2010) have studied job satisfaction in various European countries and have found that the relationship between satisfaction and wage level differs across countries: in southern Europe, lower-paid workers are less satisfied than more highly paid workers are, but this does not hold true in northern Europe – presumably because of differences in job quality. Iglesias, Llorente and Dueñas (2010) studied the relationship between job quality and job satisfaction in Spain and found that job quality in Madrid during two different time periods was high, but workers were less satisfied with their jobs.

Studies have shown that large cities offer a number of attractions for workers because of the amenities that are available in those urban centres. The literature on these urban amenities is fairly recent and has helped to account for locational differences in wage levels (Macedo and Simões, 1998) and housing prices (Rocha and Magalhães, 2013). In Chile, the amenities available in urban centres have been characterized as an explanatory variable for migration flows to certain regions, particularly the Metropolitan Region (Aroca, Geoffrey and Paredes, 2001); a classification has even been developed for regions that are attractive as places to live and those that are attractive as places to work (Aroca and Atienza, 2008), and the hypothesis has been advanced that this variable may account for some portion of regional differences in the female labour participation rate (Ferrada and Zarzosa, 2010) and in long-distance commuting rates (Jamett and Paredes, 2013). While these matters will not be analysed in any depth here, they may provide a useful perspective on some of the results of this research.

The objective here will be to discern any geographically correlated differences in job satisfaction in Chile. It is suspected that inequalities will be found between the Metropolitan Region and the rest of Chile owing to the fact that wages are much higher in the Metropolitan Region and that the production matrix is much more spatially heterogeneous in that urban area, since, because it is so much more densely populated than the rest of the country, it offers an array of amenities and financial, administrative and consumer services, while the other regions of the country rely on economic activities that are based on natural resources and that are heavily dependent on external markets.

III. Methodological aspects

As noted earlier, this study will draw on first-hand information, i.e., survey respondents' assessments of their own employment situations. Causative models were used to evaluate the extent to which the levels of job satisfaction reported by wage earners in private companies in Chile are determined by objective factors, perceptions, place of residence or demographic aspects. In addition, using a coarsened exact matching impact assessment methodology, evidence was sought of possible differences in satisfaction levels between residents of the Metropolitan Region and of other areas of the country. SPSS (Statistical Package for the Social Sciences) software was used in arriving at the descriptive analysis, and the econometric estimates were calculated using Stata.

1. Data

The data for this study were obtained from the first round of the National Survey on Employment, Work, Health and Quality of Life of Workers in Chile (ENETS), which was conducted in 2009-2010. It is the only survey of its kind in Chile; preliminary preparations for the second survey round are now under way (Ministry of Health, 2011).

This is a household survey, and the responses therefore do not necessarily correspond to the area where the respondents work. The respondents were persons over 15 years of age who had been employed during the 12-month period preceding the survey round. The survey sample (9,503 observations) was defined on the basis of the 2002 National Population and Housing Census.

The questionnaire was filled out by persons in all urban and rural regions of Chile with the exception of geographically remote areas.[2] The survey sample was designed to permit the use of the extrapolation factors to ensure that it would be representative of the various regions. Using those factors, it was determined that the sample units are representative of 7,939,170 workers in Chile's 15 regions (Ministry of Health, 2011).

Of the total sample observations, 5,802 were for private-sector wage earners (other than domestic service workers) representing 4,532,274 people. The sample distribution by region, with and without weightings, is shown in table 1. At this level of disaggregation, however, the number of observations for some areas is very small; accordingly, for the purposes of this study's geographic analysis, the regions were grouped into seven zones based on criteria having to do with geographic proximity and production-sector features, as noted below.

Table 1
Sample, by region and zone
(Number of people)

Region	Sample	Weighted sample	Zone
Arica	256	42 891	Z1
Tarapacá	216	74 926	
Antofagasta	320	164 428	Z2
Atacama	333	72 532	
Coquimbo	290	184 691	Z3
Valparaíso	594	445 357	
Metropolitan	1 199	1 973 304	Z4
O'Higgins	375	280 690	Z5
Maule	318	231 788	
Bío Bío	532	498 832	
Araucanía	275	204 647	Z6
Los Ríos	288	83 464	
Los Lagos	280	208 466	
Aysén	246	24 704	Z7
Magallanes	280	41 554	
Total	5 802	4 532 274	

Source: Prepared by the author, on the basis of data from the National Survey on Employment, Work, Health and Quality of Life of Workers in Chile.

[2] These remote areas included: Ollagüe, Easter Island and Juan Fernández Island, Lago Verde, Guaitecas, Chile Chico, Villa O'Higgins, Tortel, Torres del Paine, Río Verde, Laguna Blanca, San Gregorio, Primavera, Timaukel, Cabo de Hornos and Antarctica. The communes of Chaitén, Futaleufú and Palena were not included either.

2. Dependent variables

Job satisfaction is gauged on the basis of ENETS survey respondents' answers to the question: "How do you feel about your job?" Respondents were give seven options: very bad, bad, less than fair, fair, better than fair, good and very good. A total of 5,782 people answered this question, and their responses are calculated as being representative of 4,512,913 private-sector workers. The results for the study's target population are shown in figure 1.

Figure 1
Frequency of responses to the question "How do you feel about your job?"
(Percentages)

Source: Prepared by the author, on the basis of data from the National Survey on Employment, Work, Health and Quality of Life of Workers in Chile.

As is evident from the table, the workers' reported job satisfaction levels are quite high. This is yet another reason to try to determine what factors influence the perceptions of workers in the various geographical areas of the country, given the sharp wage differentials existing in Chile and the country's marked degree of spatial concentration.

3. Methodology

Given the distribution of the responses (see figure 1), this phenomenon is studied in two different ways, using a binary variable and a categorical variable, as follows:

- The binary SATIS2 variable takes a value of 1 when respondents report their level of job satisfaction as being very good or good and a value of 0 otherwise.
- The categorical SATIS3 variable is evaluated on the basis of three options: it takes a value of 1 when the response is "very bad", "bad", less than fair", "fair" or "better than fair"; a value of 2 if the response is "good"; and 3 if the response is "very good".

In the first case, a binary logit model is estimated; as will be seen later on, these estimates include geographical fixed effects which prove to be significant. In order to confirm the presence of such impacts, differences in job satisfaction between residents of the Metropolitan Region and residents in the rest of the country are estimated using an exact matching impact assessment methodology (Blackwell and others, 2009). The other model is a multinomial logit model (Greene, 1999).

(a) Binary logit model

In this case the dichotomous dependent variable is SATIS2. Here, we evaluate how certain variables influence the probability that workers will be satisfied or very satisfied with their jobs. In this model, the coefficients are estimated using the maximum likelihood method, which can be written as follows (Greene, 1999):

$$\hat{P}\left(SATIS2=1\right)=\frac{e^{X,\hat{\beta}}}{1+e^{X,\hat{\beta}}}=\wedge\left(\beta,X\right) \tag{1}$$

where $\hat{\beta}$ is the vector of the estimated parameters and X is the matrix of the explanatory variables.

The interpretation of the results is based on the estimated coefficients. The odds ratios for each explanatory variable are also calculated as:

$$\Omega=\frac{P\left(SATIS2=1\right)}{P\left(SATIS2=0\right)} \tag{2}$$

(b) Exact matching method

Although the preceding logit model included explanatory variables for place of residence, this method makes it possible to improve the comparison of job satisfaction for Metropolitan Region residents and persons living in other areas of Chile.

This method is primarily used to assess the impacts of public policies, but it is used here to compare groups of persons who share certain characteristics —in this case, groups of workers who gave similar answers for the X-matrix variables defined in the preceding model. These are the control variables for the matching process. Therefore, in this exercise, the difference in job satisfaction is explained solely by the condition of being a worker who resides in the Metropolitan Region or who resides elsewhere in the country.

There are various matching techniques, but here we will use the exact matching method because, as indicated in the literature, it offers a number of advantages, including the fact that it fulfils the congruence principle, restricts the data matching to areas of common empirical support and is computationally efficient (Blackwell and others, 2009).

The procedure involves first conducting a matching exercise and then using a binary logit model to arrive at an estimate. The matching is done with the variables defined in the binary logit model.

In order to check that the matching process is satisfactory, the exact matching method is used to calculate the $\mathcal{L}1$ statistic, which represents the imbalance between the treated and untreated groups, whose value can vary between 0 and 1. The idea is that, with the matching technique, the distributions should be as equal as possible, i.e., we should obtain a $\mathcal{L}1$ statistic near 0 (Iacus, King and Porro, 2012).

(c) Multinomial logit model

In order to see how the impact of the determinants changes when different levels of satisfaction are being analysed, an evaluation is conducted of the differences in the determinants for those who are very satisfied, those who are satisfied and those whose answers fall into one of the other categories;

in other words, the responses are grouped into three categories (SATIS3). In this case, the dependent variable will take the first alternative (very bad, bad, less than fair, fair or more than fair) as a reference point (base category). Hence, J = 0 is the benchmark category, which conditions the interpretations of the coefficients estimated for the other options. This will indicate the impact on the likelihood that workers' levels of job satisfaction will increase to "satisfied" (J=2) or "very satisfied" (J=3). That is, for the benchmark:

$$\hat{P}\left(SATIS3 = j\right) = \frac{e^{X,\hat{\beta}}}{1 + \sum_{j=1}^{j=2} e^{X,\hat{\beta}}} \tag{3}$$

And the likelihood for the other options is:

$$X\hat{\beta} = \alpha x + \rho x + \delta x \tag{4}$$

where β is the vector of the estimated parameters and X is the matrix of the explanatory variables.

The parameters are estimated using the maximum likelihood method, and the estimators will then be those that maximize the likelihood of the observed sample. In order to interpret the model, the odds ratio is estimated, which allows for a clearer reading of the estimated models (Long and Freese, 2014).

4. Explanatory variables

Based on the definitions and empirical evidence that have been presented, variables are then selected for four types of characteristics. Then:

$$X\hat{\beta} = \alpha x + \varphi x + \rho x + \delta x \tag{5}$$

where α, φ, ρ and δ are the estimated parameters for the explanatory variables indicated in each case (demographic characteristics, objective employment factors, area of residence and worker perceptions, respectively). Each of these elements is analysed on the basis of the variables shown in table 2.

Initially, a larger set of possible determinants was analysed, in line with the preliminary literature review; however, before proceeding to calculate the estimates, each of those determinants was examined and a descriptive analysis was undertaken. Ultimately, the variables shown in table 2 were chosen for the rest of the study. Those that were excluded were ruled out either because the previous analysis showed that they were not relevant, because, in some cases, many responses had been omitted or because the number of positive responses was too low (as happened, for example, with the question about union membership), which makes it impossible to do cross comparisons at the zone level.

This same analysis yields the categories of explanatory variables. As will be noted, they are all dichotomous, since the responses corresponding to the variables of interest in the ENETS survey are given by category. In this case, different options were explored and, in the end, those shown in table 2 were selected. The only exception was for level of education, which was retained until, finally, based on the initial analysis, it was discarded.

Table 2
Description of the variables for objective and subjective characteristics

Question	Label	Definition, variable =1	Observations	Weighted sample	Mean	Standard deviation
Sex (E2)	Sex	Male	5 802	4 532 274	0.6488	0.4774
Marital status (E5)	Casad	Married or living together	5 798	4 531 169	0.5795	0.4937
Level of education (F2t)	Educ	Technical level of education or higher	5 790	4 524 187	0.3743	0.4840
Number of workers in the company (A9)	TAM	50 or more workers	4 445	3 655 128	0.5193	0.4997
What is your net monthly income? (A48)	SALA1	Less or equal to $136,000	5 682	4 392 603	0.1389	0.3459
What is your net monthly income? (A48)	SALA2	$137.000 – $250,000	5 682	4 392 603	0.4715	0.4992
What is your net monthly income? (A48)	SALA3	$251.000 – $450,001	5 682	4 392 603	0.2707	0.4444
What is your net monthly income? (A48)	SALA4	Over $451,000	5 682	4 392 603	0.1188	0.3236
What is your net monthly income? (A48)	INGRE1	Over $651,000	5 682	4 392 603	0.0522	0.2225
How much is the salary, wage or income that you earn from your main job? (A46)	SALVAR	Other than single fixed	5 802	4 532 274	0.3372	0.4728
Do you belong to a (retirement) pension plan? (A70)	JUBILA	Any type of coverage	5 773	4 516 692	0.9154	0.2783
What type of job do you have? (A22)	TTEMP	Seasonal	5 802	4 532 274	0.1773	0.3820
What type of contract do you have? (A25)	PTIEMPO	Indefinite	4 739	3 757 711	0.2034	0.4026
At your present wage level, what type of schedule would you prefer? (A45)	CHT	The same as I have now	5 693	4 445 347	0.5695	0.4952
How satisfied are you with the opportunities for promotion or for improvements in your job? (C2a)	SPROMO4	Satisfied or very satisfied	5 728	4 395 131	0.4235	0.4942
How satisfied are you with the atmosphere where you work in terms of how people get along (fellow workers, colleagues)? (C2b)	SAMBIENT4	Satisfied or very satisfied	5 713	4 478 082	0.6739	0.4688
How satisfied are you with the physical work environment (noise, space, ventilation, temperature and lighting)? (C2c)	SCAMBIENT	Satisfied or very satisfied	5 779	4 501 101	0.5811	0.4934
How often does your direct supervisor try to make sure that workers have good development opportunities?(C1k)	OPORTDESS	Always	5 518	4 317 954	0.2707	0.4444
How often can you take sick leave without a problem? (A67b)	LICENCIA	Other than never	5 588	4 371 682	0.5373	0.4987
Do you feel motivated and committed to your job? (C1r)	MOTIVAD	Always	5 759	4 492 270	0.5834	0.4930
How do you rate your mental or emotional well-being? (D1d)	BMENTAL	Very good	5 797	4 530 543	0.1080	0.3104
How do you rate your mental or emotional well-being? (D1d)	BMENTAL2	Good or very good	5 797	4 530 543	0.7222	0.4480
Lately do you feel reasonably happy? (D14)	FELIZ	More than usual	5 802	4 532 274	0.2290	0.4202
Lately do you feel reasonably happy? (D14)	FELIZ2	Same or more than usual	5 802	4 532 274	0.9132	0.2816
How often are you afraid that you might be fired? (A66c)	MIEDODES	Almost always, always	5 763	4 499 225	0.2683	0.4431
Do you enjoy the work that you do? (D2a)	DISFTRAB	Almost always, always	5 773	4 518 873	0.7433	0.4369
Do you enjoy the work that you do? (D2a)	DISFTRAB2	Always	5 773	4 518 873	0.4891	0.4999
What changes would like to see in your job? (A69a)	DCONFSALA	Raise wages	5 802	4 532 274	0.4950	0.5000

Source: Prepared by the author, on the basis of data from the National Survey on Employment, Work, Health and Quality of Life of Workers in Chile.

IV. Results

Tables 3, 4 and 5 give the results of the logit regressions, the estimates with and without exact matching and the results obtained with the multinomial models, respectively. The analyses carried out in each case are described below.

1. Logit model

Three estimates were calculated: in the first, objective variables and respondents' characteristics were included; in the second, the zones were added; and, in the third, the subjective variables were added in.

As shown in table 3A, the likelihood that workers will be satisfied or very satisfied is significantly greater in the case of men, people who have a job that is not temporary and those who earn the most.

A majority of the workers are men (65%) and only a third of workers have a higher education or at least a technical level of education (vocational secondary schooling or completed commercial training) (see table 2). Men have a 28% greater propensity to be satisfied than women do, while the level of education turned out not to be significant (see table 3A).

The most influential variable is the income derived from the main occupation. Around half of the respondents (47%) indicated that they had a net monthly wage of between 137,000 pesos and 250,000 pesos (see table 2). The level of satisfaction of workers in this income bracket is significantly greater than it is for persons with lower wages, and the strength of the impact increases in tandem with the level of income (SALA3 and SALA4). On the other hand, the presence of temporary or seasonal contracts has a negative impact.

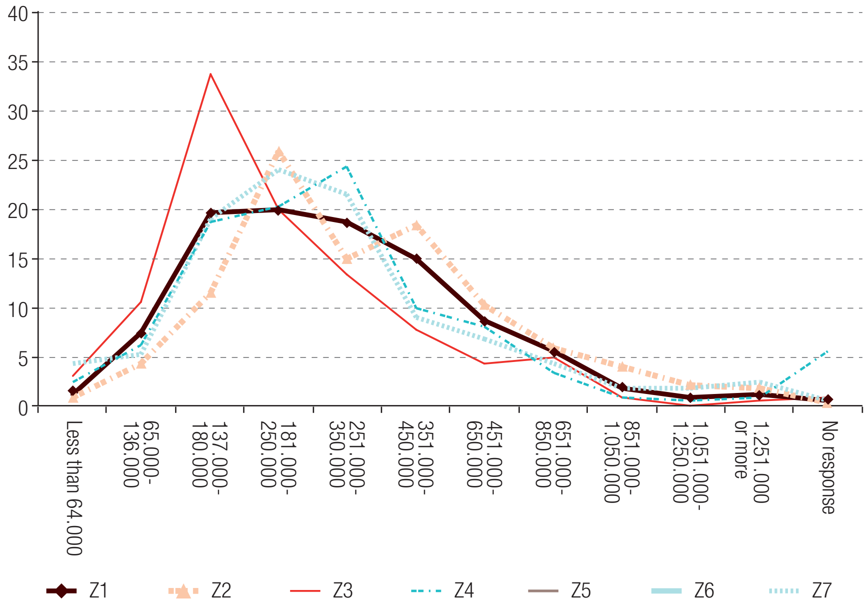
The ENETS tabulations indicate that 13% of wage earners in the private sector have a monthly income that is equal to or less than the legal minimum wage and that one fourth of them have incomes that are just slightly above the minimum wage. The Metropolitan Region is the zone in which average wages are the highest (see figure 2).[3]

The variables of marital status, company size (number of employees) pension plan coverage and fixed or variable wage were all evaluated as well, but, ultimately, they were not included because the results were not significant and did not contribute to the model.

When the variables for geographic zones (see table 3B) were incorporated, the resulting coefficient was consistently positive and significant, indicating that residence in a zone other than the Metropolitan Region (corresponding to the base zone in the estimate) has a positive influence on job satisfaction; the strength of that influence is lower in the northern zones of Chile, however. This finding is counterintuitive, inasmuch as higher wages have a positive effect on job satisfaction yet satisfaction levels are higher among people residing in zones other than the one where wages are the highest (Z4). The fact that the survey was conducted among households and that the responses came from one member of each household may have something to do with this result, since the respondents do not necessarily work in the zone where they live. This question will be explored further in the following section.

[3] The survey questions concerned net income. The minimum wage in Chile in 2009 was 165,000 pesos and in 2010 it was 172,000 pesos; these are gross monthly figures. After subtracting pension contributions and insurance, they yield net figures of approximately 140,000 pesos and 146,000 pesos, respectively.

Figure 2
Reported net wages, by zone
(Chilean pesos and percentages)



Source: Prepared by the author, on the basis of data from the National Survey on Employment, Work, Health and Quality of Life of Workers in Chile.

Table 3C shows the results obtained using the third model, which includes both objective and subjective variables. These results indicate that wage levels continue to be significant but that the variable for temporary versus more stable contracts does not.

The interesting point in this case is how important the subjective variables turn out to be. The factors having the biggest impact are whether the respondents feel good or very good mentally and emotionally (BMENTAL2) and whether they enjoy the work that they do (DISFTRAB2).

The likelihood of job satisfaction diminishes if the respondents are constantly afraid that they will be fired or laid off (MIEDODES) or if they are unhappy with their wage level (those who indicate that they would like a raise) (CONFSALA).

Other variables, such as satisfaction with promotion opportunities (SPROMO4), development opportunities (OPORTDESS), feeling reasonably happy (FELIZ2) and satisfaction with the work environment (SAMBIENT4), yield the results that were expected. It should be noted that the first two of these variables are associated with the management of the company, while the third has to do with subjective emotional factors and the fourth with the way the company is organized. The other variables did not have an impact on job satisfaction.

At the geographic level, the results indicate that the impact of the zone of residence disappears in the case of two of the zones (Z2 and Z7) but remains in the rest.

Table 3

Results obtained with the logit models

(Base option (= 1): satisfied or very satisfied with the job)

SATIS2	3A				3B				3C			
	Coefficient	Robust standard error	P>|z|	Odds ratio	Coefficient	Robust standard error	P>|z|	Odds ratio	Coefficient	Robust standard error	P>|z|	Odds ratio
Constant	0.1082	0.2166	0.617	1.1143	-0.308	0.2232	0.167	0.7348	-2.2273	0.3499	0.000	0.1078 ***
sex	0.2483	0.1311	0.058	1.2819 *	0.1945	0.1316	0.140	1.2147	-0.1187	0.1621	0.464	0.8880
educ	0.2162	0.1424	0.129	1.2414	0.2112	0.1413	0.135	1.2352	0.0977	0.1679	0.560	1.1027
TTEMP	-0.260	0.1488	0.080	0.7710 *	-0.371	0.1527	0.015	0.6898 **	-0.2510	0.1830	0.170	0.7780
SALA2	0.3932	0.1868	0.035	1.4817 **	0.4605	0.1769	0.009	1.5849 ***	0.4042	0.2051	0.049	1.4981 **
SALA3	0.6795	0.2153	0.002	1.9729 ***	0.8891	0.2127	0.000	2.4328 ***	0.5002	0.2536	0.049	1.6490 **
SALA4	1.3109	0.2863	0.000	3.7094 ***	1.4882	0.2915	0.000	4.4291 ***	0.8434	0.3127	0.007	2.3242 ***
MZ4												
MZ1					0.4040	0.2293	0.078	1.4978 *	0.4344	0.2263	0.055	1.5440 *
MZ2					0.4437	0.2466	0.072	1.5584 *	0.0623	0.3174	0.844	1.0643
MZ3					0.7258	0.1780	0.000	2.0664 ***	0.4607	0.2203	0.037	1.5851 **
MZ5					0.6003	0.1605	0.000	1.8226 ***	0.3058	0.1852	0.099	1.3578 *
MZ6					0.8434	0.1847	0.000	2.3242 ***	0.5660	0.1961	0.004	1.7612 ***
MZ7					0.9812	0.2332	0.000	2.6675 ***	0.3965	0.3013	0.188	1.4865
SPROMO4									0.7322	0.1621	0.000	2.0796 ***
SAMBIENT4									0.3913	0.1593	0.014	1.4788 **
OPORTDESS									0.6528	0.1930	0.001	1.9209 ***
BMENTAL2									1.0453	0.1648	0.000	2.8443 ***
FELIZ2									0.5183	0.2473	0.036	1.6792 **
MIEDODES									-0.4037	0.1536	0.009	0.6679 ***
DISFTRAB									1.1128	0.1942	0.000	3.0428 ***
DISFTRAB2									0.4279	0.1800	0.017	1.5340 **
CONFSALA									-0.2862	0.1429	0.045	0.7511 **
	Pseudo R2= 0.034		Number of obs= 5 653		Pseudo R2: 0.0518		Number of obs= 5 653		Pseudo R2= 0.275		Number of obs= 5 248	
	Wald chi2(6)=47		Prob > chi2 =0		Wald chi2(6):82.24		Prob > chi2 =0		Wald chi2(6)=342.89		Prob > chi2 =0	
	Log pseudo likelihood = -2 621 268,6				Log pseudo likelihood = -2 573 000,8				Log pseudo likelihood = -1 822 160,1			

Source: Prepared by the author, on the basis of data from the National Survey on Employment, Work, Health and Quality of Life of Workers in Chile.
Note: *Significant at 5%, **significant at 1%, ***significant at 10%.

2. Exact matching methodology

The next step was to test these results to see how much of an impact residence in a zone other than the Metropolitan Region has on job satisfaction. As noted in the section on the methodology, the impact analysis technique known as exact matching was used for this purpose.

This evaluation was performed using two groups of subjects having all of the same characteristics except their zone of residence. In order to use the matching method, a common support for the two groups must be found. The variables used in this case were the determinants employed in the preceding logit estimate (see table 3C), which include both objective and subjective variables (sex, educ, TTEMP, SALA2, SALA3, SALA4, SPROMO4, SAMBIENT4, OPORTDESS, BMENTAL2, FELIZ2, MIEDODES, DISFTRAB, DISFTRAB2, CONFSALA).

This yields a matched base pair in which the one difference between the subjects is the place of residence. Thus, the likelihood that these persons will feel good or very good about their work was estimated using a logit model with a single explanatory variable that identifies the zone in which they live.

Comparisons were then made between the subjects in the Metropolitan Region and those living in all other regions of the country and then between the Metropolitan Region Zone (Z4) subjects and the subjects in each of the other zones; two types of estimates were calculated: estimates with and estimates without applying the exact matching method. The results are shown in table 4.

As may be seen from the table, the ℒ1 indicator improves substantially in all cases. This indicates that the application of the exact matching methodology has produced two comparable groups, which makes it possible to improve the analysis. It has, however, also reduced the number of observations by between 54% and 61%.

The estimates confirm that living in a zone other than the Metropolitan Region has a positive impact on job satisfaction. The analysis at the zone level yields the same result for four areas (Z3, Z5, Z6, Z7); a significant impact was not found in only two of the zones: Z1 and Z2.

Table 4

Estimates calculated with and without applying the exact matching method: impact of place of residence on the likelihood of job satisfaction

		Impact of place of residence on the likelihood of job satisfaction (base zone = Z4)						
		ALL	Z1	Z2	Z3	Z5	Z6	Z7
Before matching	ℒ1 statistic	0.60113	0.713060	0.710915	0.648887	0.670845	0.687866	0.717844
	B	0.4407*** (0.1368)	0.5780** (0.2497)	0.5157** (0.2431)	0.5571*** (0.1917)	0.2506 (0.1596)	0.5807*** (0.1820)	0.9323*** (0.2513)
	Constant	0.5351*** (0.0664)	0.5351*** (0.1197)	0.5351*** (0.1197)	0.5351*** (0.1197)	0.5351*** (0.1197)	0.5351*** (0.1197)	0.5351*** (0.1197)
	Pseudo R^2	0.0083	0.0029	0.0045	0.0096	0.0025	0.0089	0.0041
	Prob > chi2	0.0013	0.0029	0.0339	0.0037	0.1164	0.0014	0.0002
	Observ. SATIS2=0	1 401	480	506	576	709	571	449
	Observ. SATIS=1	3 773	947	1 125	1 248	1 417	1 218	1 028
After matching	ℒ1 statistic	1.913e-15	2.88e-16	5.89e-16	3.29e-16	1.30e-15	7.61e-17	7.42e-16
	B	0.3902*** (0.1035)	0.3003 (0.2045)	0.2952 (.2114)	0.4196** (0.1713)	0.3171** (0.1462)	0.4751*** (0.1705)	0.8197*** (0.2695)
	Constant	1.1511*** (0.057)	1.010*** (0.1240)	1.522*** (0.1374)	1.1984*** (0.1095)	0.8739*** (0.1012)	1.040*** (0.1123)	1.582*** (0.1520)
	Pseudo R^2	0.0052	0.0035	0.0033	0.0069	0.0042	0.0091	0.0225
	Prob > chi2	0.0001	0.1391	0.1606	0.0135	0.0298	0.0050	0.0016
	Observ. SATIS2=0	737	332	360	468	470	411	306
	Observ. SATIS=1	2 143	226	322	417	502	411	265

Source: Prepared by the author, on the basis of data from the National Survey on Employment, Work, Health and Quality of Life of Workers in Chile.

Note: *Significant at 5%, **significant at 1%, ***significant at 10%. Robust standard deviations are shown in brackets.

3. Multinomial model

This model will be used to look at possible differences in impact by level of satisfaction, i.e., to calculate the probabilities that subjects will feel good or very good about their job, in different categories, as stated in equations (3) and (4).

Three models are estimated, as shown in tables 5A, 5B and 5C. In the first, it can be seen that the subjects' propensity to be very satisfied with their jobs is greater than the other options if they have a high level of education; in that latter category, the person's sex ceases to be significant. In addition, the impact of higher wages is greater than in option 3 (feel very good about their job).

In the second model, when the residence variables are incorporated (see table 5B), effects similar to those found with the binary logit model are observed, although zones 1, 2 and 5 are not significant.

Table 5C shows the model with objective and subjective variables but without incorporating place of residence. What is interesting here is the difference in the impact of the subjective variables across categories. For example, satisfaction with working hours (CHT) has a significant and positive impact on the likelihood of feeling very good about one's job, but not in the next category down (feeling good). Subjects with temporary contracts (TTEMP) are less likely to feel very good about their job.

Some subjective factors have an effect in both cases but their impact is greater for the third option (feeling very satisfied). In this case, the variables are satisfaction with the work environment (SAMBIENT4), development opportunities (OPORTDESS) and always enjoys work (DISFTRAB).

However, the variables of fear of being fired or laid off, satisfaction with the existing wage and feeling good the same usual as or more than usual only influence the propensity to be satisfied. It is possible that, in order to be very satisfied, the expectations to be met are higher, as evidenced by the DISFTRAB and DISFTRAB2 variables. In this case, always enjoying one's job has a positive impact in both cases, whereas enjoying one's job always or almost always has a positive effect only on feeling good about one's job (category 2). In turn, those who feel good or very good mentally and emotionally (BMENTAL and BMENTAL2) are 2.2 times and 10.6 times more likely, respectively, to feel very good about their job.

This analysis shows how important subjective variables associated with human resources management and the organizational environment are in determining the level of job satisfaction. The results are compelling: employees value these types of conditions even more than they value objective working conditions. This finding may have important implications for organizational management and public policy.

Table 5
Results obtained with multinomial models
(Base option (= 1): dissatisfied with the job)

SATIS3		Option 2: feels good (satisfied)				Option 3: feels very good (very satisfied)			
		Coefficient	Robust standard error	P>|z|	Odds ratio	Coefficient	Robust standard error	P>|z|	Odds ratio
5A	Sex	0.3038	0.1336	0.023	1.3551 **	-0.1161	0.2217	0.601	0.8904
	Educ	0.1583	0.1440	0.272	1.1715	0.6900	0.2497	0.006	1.9937 ***
	SALA2	0.4112	0.1792	0.022	1.5086 **	0.7928	0.3019	0.009	2.2096 ***
	SALA3	0.6833	0.2077	0.001	1.9804 ***	1.3884	0.3571	0.000	4.0086 ***
	SALA4	1.3090	0.2844	0.000	3.7026 ***	2.0669	0.4035	0.000	7.9002 ***
	Constant	-0.1107	0.1911	0.562	0.8952	-2.4200	0.3193	0.000	0.0889 ***
	Number of observations = 5 653; Wald chi2(10)= 64,4; Prob>chi2=0; Log pseudo likelihood=0,0315								

Table 5 (concluded)

	SATIS3	Option 2: feels good (satisfied)				Option 3: feels very good (very satisfied)			
		Coefficient	Robust standard error	P>lzl	Odds ratio	Coefficient	Robust standard error	P>lzl	Odds ratio
5B	Sex	0.2502	0.1335	0.061	1.2843 *	-0.1530	0.2246	0.496	0.8581
	Educ	0.1596	0.1430	0.264	1.1730	0.6792	0.2489	0.006	1.9723 ***
	SALA2	0.5027	0.1693	0.003	1.6532 ***	0.8232	0.3026	0.007	2.2778 ***
	SALA3	0.9296	0.2054	0.000	2.5335 ***	1.5170	0.3540	0.000	4.5585 ***
	SALA4	1.5276	0.2905	0.000	4.6071 ***	2.1643	0.4179	0.000	8.7089 ***
	MZ1	0.4321	0.2312	0.062	1.5405 *	0.0713	0.3638	0.845	1.0739
	MZ2	0.3930	0.2501	0.116	1.4814	0.5419	0.3467	0.118	1.7193
	MZ3	0.6967	0.1849	0.000	2.0071 ***	0.7349	0.3058	0.016	2.0853 **
	MZ5	0.5957	0.1571	0.000	1.8144 **	0.0161	0.2669	0.952	1.0162
	MZ6	0.8130	0.1853	0.000	2.2547 ***	0.8028	0.3179	0.012	2.2318 **
	MZ7	0.8567	0.2397	0.000	2.3554 ***	1.3331	0.3338	0.000	3.7929 ***
	Constant	-0.5659	0.2076	0.006	0.5678 ***	-2.7072	0.3486	0.000	0.0667 **
	Number of observations = 5 653; Wald chi2(22)=117,06; Prob>chi2=0; Log pseudo likelihood=0,0453								
5C	Sex	-0.0580	0.1500	0.699	0.9437	-0.8741	0.2438	0.000	0.4172 ***
	SALA2	0.3635	0.2109	0.085	1.4383 *	0.5673	0.4172	0.174	1.7634
	SALA3	0.4174	0.2667	0.118	1.5180	0.8117	0.4512	0.072	2.2518 *
	SALA4	0.7180	0.3036	0.018	2.0503 **	1.2273	0.5909	0.038	3.4122 **
	TTEMP	-0.1265	0.1745	0.469	0.8812	-1.6804	0.3793	0.000	0.1863 ***
	CHT	0.1126	0.1529	0.462	1.1192	0.4760	0.2523	0.059	1.6096 **
	SPROMO4	0.6989	0.1657	0.000	2.0115 ***	0.6085	0.2863	0.034	1.8376 **
	SAMBIENT4	0.3196	0.1617	0.048	1.3766 **	0.6479	0.3295	0.049	1.9115 **
	OPORTDESS	0.5315	0.1932	0.006	1.7015 **	1.5005	0.3051	0.000	4.4841 ***
	BMENTAL	-0.4088	0.3536	0.248	0.6644	2.3699	0.3483	0.000	10.696 ***
	BMENTAL2	1.0947	0.1695	0.000	2.9882 ***	0.8136	0.3386	0.016	2.2560 **
	FELIZ2	0.5045	0.2499	0.044	1.6562 **	0.5357	0.4512	0.235	1.7087
	MIEDODES	-0.4356	0.1524	0.004	0.6468 ***	-0.3295	0.3024	0.276	0.7193
	DISFTRAB	1.1149	0.1980	0.000	3.0493 ***	0.6312	0.5378	0.241	1.8799
	DISFTRAB2	0.4074	0.1834	0.026	1.5029 **	1.7481	0.3176	0.000	5.7434 ***
	CONFSALA	-0.3461	0.1436	0.016	0.7074 **	-0.3136	0.2464	0.203	0.7308
	Constant	-1.9983	0.3288	0.000	0.1356 ***	-5.2879	0.6222	0.000	0.0051 **
	Number of observations = 5 185; Wald chi2(32)= 622,62; Prob>chi2=0; Log pseudo likelihood=0,2773								

Source: Prepared by the author, on the basis of data from the National Survey on Employment, Work, Health and Quality of Life of Workers in Chile.
Note: *Significant at 5%, **significant at 1%, ***significant at 10%.

V. Conclusions

In line with the initial objective of this study, the level of job satisfaction of private-sector wage earners in Chile has been analysed by evaluating the impact of a number of different variables and by gauging differences in perceived job satisfaction on the part of persons residing in the Metropolitan Region and in other zones of Chile. This research has been undertaken with the intention of contributing to the understanding of this subject in the country, since very few studies on the topic have taken geographical factors into account.

The research was based on ENETS survey data for 2009-2010; this was the first survey that included such a wide range of subjective variables and that was regionally representative.

The international literature provides a variety of findings regarding the impact of wage levels on job satisfaction. The results of this study indicate that wages levels definitely have a positive and significant impact, and they are consistent inasmuch as, the higher the wage, the greater this effect is; when the dependent variable was analysed in the three-category model, the coefficient was the highest in category 3 (feel very good about their job) (multinomial model).

The results of this analysis regarding subjective variables are quite interesting, inasmuch as they demonstrate that these variables have a greater impact than the objective variables do. In fact, of all the factors that were initially considered, the final two models (3C and 5C) are composed primarily of subjective factors (perceptions).

The variables associated with job quality, such as social security and type of contract, were not significant; what is more, even when the wage level is the only objective variable that is left, when subjective determinants are added in, they are more significant. This finding is an important one, since the main thrust of public policy in this area in Chile is to achieve a certain minimum level in terms of job quality, whereas employees are, in all likelihood, placing greater value on perceived (subjective) conditions relating to human resources and organizational management.

In addition, when job satisfaction is analysed on the basis of three different categories, the differences in the influence exerted by subjective variables are notable.

However, the propensity to feel good about one's job is greater in areas other than the Metropolitan Region and, when the results are analysed by zone, only zones 1 and 2 in the far north cease to be significant. This finding seems to be counterintuitive, since wages are higher in the Metropolitan Region in relative terms and since that region also presumably offers greater amenities. The reasons for this result have not been analysed in this study but should, of course, be explored. In the meantime, a few possible explanations will be examined on the basis of some degree of intuition.

First of all, one hypothesis, based on an extrapolation of the material provided in the literature, is that, while wages are lower in other regions, particularly those subsumed under zones 3, 4, 5 and 6, perhaps the residents of those areas are more satisfied simply by the fact that they have a job at all; it is possible, then, that their job satisfaction threshold is lower.

Another reason might be that, since subjective variables (perceptions) are so influential, the impact of these factors, in combination with those of other location-related conditions (cultural factors, social circles, family ties, among others), outweigh the importance of wage levels. This could also account for Chile's low labour mobility.

Finally, it may be that wages exhibit endogenous behaviour as a reflection of dual causality. In other words, job satisfaction could be determining employees' wage levels; if this is the case, it would corroborate the importance of perceptions and, following on with the above line of reasoning, would explain the positive effect of living in a zone other than the Metropolitan Region, although this result would be somewhat contradictory, given the marked concentration of the population in that region.

The results to date indicate, on the one hand, that this subject area should be approached from a geographic perspective and that possible differences between regions should be explored. This, in turn, points to the importance of using decentralized public policy tools. The information provided by this study on the influence exerted by objective and subjective variables affords relevant inputs in terms of both public policymaking and business administration, since it underscores the importance of human resources management and organizational arrangements dealing with the work environment in terms of employees' job satisfaction. An interesting question —that will remain unanswered here— is whether or not the "new management" approach will have differing effects in different geographical areas.

Finally, this study points to the need for an updated survey design that will include variables that capture people's perceptions of job quality, performance, work environment and job satisfaction and that will thus make it possible to verify the temporal consistency of these results.

Bibliography

Aroca, P. and M. Atienza (2008), "La conmutación regional en Chile y su impacto en la Región de Antofagasta", *EURE*, vol. 34, No. 102, Santiago, Catholic University of Chile.

Aroca, P., J. D. Geoffrey and J. Paredes (2001), "Migracion interregional y el mercado laboral en Chile: 1977-82 y 1987-92", *Cuadernos de Economía*, vol. 38, No. 115, Santiago, Catholic University of Chile.

Becker, G. (1964), *Human Capital*, Chicago, The University of Chicago Press.

Belfield, C. and R. Harris (2002), "How well do theories of job matching explain variations in job satisfaction across education levels? Evidence for UK graduates", *Applied Economics*, vol. 34, No. 5, Taylor & Francis.

Blackwell, M. and others (2009), "CEM: coarsened exact matching in Stata", *The Stata Journal*, vol. 9, No. 4.

Booth, A. and J. Van Ours (2007), "Job satisfaction and family happiness: the part-time work puzzle", *Working Paper Series*, No. 100, Melbourne, University of Melbourne.

Bòria-Reverter, S., M. Crespi-Vallbona and O. Mascarilla-Miró (2012), "Variables determinantes de la satisfacción laboral en España", *Cuadernos de Economía*, vol. 35, No. 97, Elsevier.

Borra, C. and F. Gómez (2012), "Satisfacción laboral y salario: ¿compensa la renta laboral las condiciones no monetarias del trabajo?, *Revista de Economía Aplicada*, vol. 20, No. 3, University of Zaragoza.

Bryson, A., L. Cappellari and C. Lucifora (2004), "Does union membership really reduce job satisfaction?", *British Journal of Industrial Relations*, vol. 42, No. 3, Wiley.

Card, D. and others (2010), "Inequality at work: the effect of peer salaries on job satisfaction", *NBER Working Paper*, No. 16396, Cambridge, Massachusetts, National Bureau of Economic Research (NBER).

Chiavenato, I. (2006), *Introducción a la teoría general de la administración*, McGraw-Hill.

Clark, A. (1997), "Job satisfaction and gender: why are women so happy at work?", *Labour Economics*, vol. 4, No. 4, Amsterdam, Elsevier.

Clark, A. and A. Oswald (1996), "Satisfaction and comparison income", *Journal of Public Economics*, vol. 61, No. 3, Amsterdam, Elsevier.

Clark, A. and F. Postel-Vinay (2009), "Job security and job protection", *Oxford Economic Papers*, vol. 61, No. 2, Oxford University Press.

Clark, A., K. Nicolai and N. Westergård-Nielsen (2007), "Job satisfaction and co-worker wages: status or signal?", *IZA Discussion Paper*, No. 3073.

De Vries, W. and others (2015), "Conclusiones a contrapelo: la aportación de distintas carreras universitarias a la satisfacción en el empleo", *Revista de la Educación Superior*, vol. 37, No. 146, National Association of Universities and Higher Education Institutions.

Farné, S. (2012), *La calidad del empleo en América Latina a principios del siglo XXI*, Bogota, Universidad Externado de Colombia.

___(2003), *Estudio sobre la calidad de empleo en Colombia*, International Labour Organization.

Farné, S. and A. Vergara (2007), "Calidad del empleo: ¿qué tan satisfechos están los colombianos con su trabajo?", *Cuadernos de Trabajo*, No. 8, Bogotá, Universidad Externado de Colombia.

Ferrada, L. M. and P. Zarzosa (2010), "Diferencias regionales en la participación laboral femenina en Chile", *Cuadernos de Economía*, vol. 47, No. 136, Santiago, Catholic University of Chile.

Freeman, R. (1977), "Job satisfaction as an economic variable", *NBER Working Paper*, No. 225, Cambridge, Massachusetts, National Bureau of Economic Research (NBER).

Gamero, C. (2003), "Análisis por género de la relación entre satisfacción laboral and tipo de contrato en España", *Anales de Economía Aplicada 2007*, Spanish Association of Applied Economics.

Greene, W. (1999), *Análisis econométrico*, Madrid, Prentice Hall.

Grund, C. and D. Sliwka (2007), "Reference-dependent preferences and the impact of wage increases on job satisfaction: theory and evidence", *Journal of Institutional and Theoretical Economics (JITE)*, vol. 163, No. 2, Mohr Siebeck.

Herzberg, F. (1962), "Basic needs and satisfactions of individuals", *Industrial Relations Monograph*, No. 21.

Herzberg, F., B. Mausner and B. Sryderman (1959), *The Motivation to Work*, New York, John Wiley and Sons.

Iacus, S., G. King and G. Porro (2012), "Causal inference without balance checking: coarsened exact matching", *Political Analysis*, vol. 20, No. 1, Cambridge, Cambridge University Press.

Iglesias, C., R. Llorente and D. Dueñas (2010), "Calidad del empleo and satisfacción laboral", *Investigaciones Regionales*, No. 19, Spanish Association of Regional Science.

ILO (International Labour Organization) (1999), "Decent work", *Report of the Director-General*, 87th session of the International Labour Conference, Geneva.

INE (National Institute of Statistics) (n/d) [online] http://www.ine.cl/canales/chile_estadistico/mercado_del_trabajo/empleo/series_estadisticas/nuevas_empalmadas/series_fecha.php.

___(n/d) [online] http://www.ine.cl/canales/chile_estadistico/mercado_del_trabajo/remuneraciones/anuarios_estadisticos/anuarios_estadisticos.php.

Jamett, I. and D. Paredes (2013), "Conmutación de larga distancia en Chile: estimando el premio por trabajar muy lejos de casa", *Estudios de Economía*, vol. 40, No. 2, Santiago, University of Chile.

Juárez-Adauta, S. (2012), "Clima organizacional y satisfacción laboral", *Revista Médica del Instituto Mexicano del Seguro Social*, vol. 50, No. 3, Mexico City, Mexican Social Security Institute.

Lévy-Garboua, L., C. Montmarquette and V. Simonnet (2007), "Job satisfaction and quits", *Labour Economics*, vol. 14, No. 2, Amsterdam, Elsevier.

Long, J. and J. Freese (2014), *Regression Models for Categorical Dependent Variables Using Stata*, Stata Press.

Macedo, P. and R. Simões (1998), "Amenidades urbanas e correlação espacial: uma analise intra-urbana para BH (MG)", *Revista Brasileira de Economia*, vol. 4, No. 52, Getulio Vargas Foundation.

McGregor, D. (1981), *El aspecto humano de las empresas*, Mexico City, Diana.

Mincer, J. (1958), "Investment in human capital and personal income distribution", *Journal of Political Economy*, vol. 66, No. 4, Chicago, The University of Chicago Press.

MINSAL (Ministry of Health) (2011), "Primera encuesta nacional de empleo, trabajo, salud y calidad de vida de los trabajadores y trabajadoras en chile (ENETS 2009-2010)" [online] http://epi.minsal.cl/epi/0notransmisibles/ENETS/INTERINSTITUCIONAL_BAJA.pdf.

___(n/d), "Diseño conceptual and metodológico" [online] http://epi.minsal.cl/wp-content/uploads/2012/07/Resumen_Ejecutivo_ENETS_Final.pdf.

Narbona, K. (2012), "La producción de compromiso en la empresa flexible", Santiago, University of Chile [online] http://repositorio.uchile.cl/handle/2250/116785.

Olarte, M. (2011), *Los determinantes de la satisfacción laboral: una revisión teórica and empírica*, Bogota, Universidad Jorge Tadeo Lozano.

Pouliakas, K. and I. Theodossiou (2010), "Differences in the job satisfaction of high-paid and low-paid workers across Europe", *International Labour Review*, vol. 129, No. 1, International Labour Organization.

Rocha, R. and A. Magalhães (2013), "Valoração das amenidades urbanas: uma estimação a partir dos diferenciais salariais e do custo de habitação para as regiões metropolitanas brasileiras", *Revista de Economia Contemporanea*, vol. 17, No. 1, Rio de Janeiro, Institute of Economics of the Federal University of Rio de Janeiro.

Rodríguez-Gutiérrez, C. and J. Prieto-Rodríguez (2004), "Efecto de la afiliación sindical sobre la satisfacción laboral de los trabajadores en el caso español" [online] http://aeet-aede.ulpgc.es/Descargas/Sesion1Sala2/Rodriguez-Prieto.pdf.

Sehnbruch, K. (2012), "La calidad del empleo en Chile: teoría y medición", *La calidad del empleo en América Latina a principios del siglo XXI*, S. Farné (comp.), Bogota, Universidad Externado de Colombia.

___(2008), "From the quantity to the quality of employment: an application of the capability approach to the Chilean labour market", *The Capability Approach: Concepts, Measures and Applications*, F. Comim, M. Qizilbash and S. Alkire, New York, Cambridge University Press.

Sen, A. (1999), *Development as Freedom*, Oxford, Oxford University Press.

Sloane, P. J. and H. Williams (2000), "Job satisfaction, comparison earnings, and gender", *Labour*, vol. 14, No. 3, Wiley.

Somarriba, N. and others (2010), "La calidad del trabajo en la Unión Europea", *Estudios de Economía Aplicada*, vol. 28, No. 3.

Stephen, R. and J. Timothy (2013), *Comportamiento organizacional*, Pearson.

Currency carry trade and the cost of international reserves in Mexico

Carlos A. Rozo and Norma Maldonado[1]

Abstract

National strategies aimed at boosting economic growth following the global financial crisis have spawned monetary imbalances between industrial and emerging economies. By implementing ultra-expansionary monetary policies, the industrial economies drive down interest rates, while the emerging economies tighten their monetary policies by raising rates, thus generating a burgeoning foreign-currency carry trade. Vulnerability is caused by the sudden reversal of such capital flows or the high cost of insuring against this by accumulating reserves. This paper estimates that the cost of reserve accumulation between 2008 and 2014 averaged 1.83% of GDP, so the free capital mobility espoused by the Mexican authorities makes it very costly to play by the rules of financial globalization.

Keywords

Capital movements, foreign exchange, foreign exchange markets, monetary reserves, costs, monetary policy, Mexico

JEL classification

F31, F32, F38

Authors

Carlos Antonio Rozo Bernal is a level-C tenured professor in the Department of Economic Production of the Autonomous Metropolitan University of Mexico. Email: rozo@correo.xoc.uam.mx.

Norma Maldonado is a research assistant in the Department of Economic Production of the Autonomous Metropolitan University of Mexico. Email: normamc_27@hotmail.com.

[1] The authors are grateful for comments made by Mario Capdevielle Allevato and an anonymous referee, which helped improve the treatment of the topic.

I. Introduction

National strategies to boost economic growth following the global financial crisis have spawned monetary imbalances between industrial and emerging economies. By implementing ultra-expansionary monetary policies, industrial countries drive interest rates down, thereby reducing the value of their currencies while making their export sectors more competitive. In contrast, emerging economies tighten their monetary policies by raising interest rates to attract capital. These conflicting policies have generated large interest-rate spreads, which attract capital flows into emerging markets.

This has led to a wide-ranging and growing foreign-exchange carry trade, which energizes and strengthens what Burnside, Eichenbaum and Rebelo (2011) identify as one of the oldest and most popular currency trading strategies. While this carry trade generates large profits for footloose short-term capital, it undermines well-being in emerging markets by creating vulnerability, owing to sudden reversals of these flows and the high cost of insuring against this by accumulating reserves.

The scale of this carry trade forced the International Monetary Fund (IMF) to recognize the validity of the reactions of a number of emerging country governments to defend themselves against short-term capital pressures with cross-border control policies. It was thus recognized that emerging countries did not have to pay for the dishes that the more developed countries had broken. IMF accepted that, under certain circumstances and conditions, these countries were entitled to use such policies to address the risks caused by hot-money inflows. Thus, it was accepted that controlling cross-border capital flows is legitimate and can be a useful part of a macroeconomic strategy to prevent such flows from affecting the value of the local currency, or at least mitigate the repercussions.

Mexico is one of the emerging markets whose exchange rate has come under heavy pressure from these capital flows. Other countries that have experienced a similar fate —and in some cases with greater pressures— have included, Brazil, Chile, Colombia and Peru, along with India, the Republic of Korea and Thailand. The difference between Mexico and the other affected countries reflects the attitude of the Mexican authorities, who do not consider it prudent or necessary to intervene to control the strengthening of the domestic currency, the peso, as other countries have done. The Mexican monetary authorities continue to deny that peso appreciation is harmful for the economy, so they consider it unnecessary to impose controls or, as IMF refers to them, exchange-rate management measures. This official stance is justified by the argument that the Mexican financial system is shielded against a sudden outflow of capital due to the accumulation of high levels of international reserves and the contracting of a Flexible Credit Line (FCL) with IMF.

This paper examines the rationale by which the Mexican authorities take a positive view of what the authorities of other countries consider harmful to their economies. It argues that the free capital mobility practised by the Mexican authorities reduces welfare among the Mexican people owing to the high cost of accumulating international reserves. To demonstrate this approach, sections II, III and IV review the grounds of the debate on the need for exchange controls that led the Fund to accept that certain types of measures, under certain conditions, are positive for emerging and developing countries. The study then analyses the behaviour of the exchange rate and how it relates to the monetary policy that gave rise to the Mexican super-peso. The following three sections (V, VI and VII) review the magnitude and return on capital flows, and the prevalence of short-term capital, Then, section VIII estimates the cost of the financial shielding, and the paper ends with conclusions.

II. Currency carry trade and exchange-rate protectionism

The need to regain the growth momentum that was lost in the wake of the Great Recession has led industrial economies, particularly the United States, to implement a loose monetary policy that has forced the benchmark interest rate down to zero. The weak performance of financial investments in these economies has fuelled the transfer of huge amounts of capital to economies offering higher rates of return. In a search for yield, financial investors have flooded emerging-country money markets with private portfolio inflows since 2009.

These flows have tended to destabilize exchange rates. In industrial countries the general trend has been one of currency depreciation, while emerging country currencies have generally tended to appreciate. These exchange-rate movements have provided trade advantages to the former, while making the latter less competitive. Given these imbalances created by the monetary policy of industrial countries, the emerging countries usually respond by protecting their markets, which marks a new milestone in the development of the global crisis, as a window of exchange-rate protectionism opens. Gallagher (2011a) and Rodrik (2010) note that this confrontation means that the international economy is approaching the end of an era in global finance —the era of free mobility of speculative capital.

The wide spread between interest rates in the north and the south has created opportunities for the currency carry trade, that is, the strategy of borrowing funds denominated in a low-interest-rate currency (the funding currency) to be invested in financial securities denominated in higher-yielding currencies (the investment currency) (Burnside, Eichenbaum and Rebelo, 2011, Clarida, Davis and Pedersen, 2009). The operation seeks a fiduciary gain augmented by the possible appreciation of the investment currency, which pays the higher interest rate (Banco de México, 2010). Despite the risk that this entails, the evidence tends to show that the activity is profitable (see Brunnermeier, Nagel and Pedersen (2008), and Jordá and Taylor (2009)).

The theory of uncovered interest-rate parity states that in a frictionless and risk-neutral economy, the currency carry trade is not profitable because the gain obtained from the interest rate spread is offset by an equi-proportional depreciation in the investment currency. Although capital flows entering the country cause an immediate appreciation of the currency as the market overshoots, this is later reversed, and the currency ultimately depreciates (Fama, 1984; Dornbusch, 1976). In practice, however, by persisting over time, the foreign-currency carry trade violates these theoretical postulates, since the investment currency continues to appreciate, thereby rendering the activity profitable.

This problem, known as the "forward-premium puzzle", is associated with the risk of a crash because exchange rates take the stairs up and the elevator down, which means that this dynamic can generate an "exchange-rate bubble" in the long run.[2] The premium is the price of crash risk.

The yield on the assets in which the speculators invest has two basic drivers: a positive average rate of return and negative skewness. The gain occurs in recognition of the liquidity that the carry trade provides to the investment economy, while the negative skewness stems from the investment currency's exposure to crash risk, by movements in the opposite direction in the spot and future exchange rate, or expected exchange rate. What needs to be stressed is that positive interest-rate spreads are associated with negative skewness in exchange-rate movements, which means that the yields of the currency carry trade implicitly involve crash risk (Brunnermeier, Nagel and Pedersen, 2008, p. 4). This occurs because of the asymmetric responses of investors to movements in the expected exchange rate. When exchange-rate movements lead to losses, these are amplified by the difficulties that speculators may experience in obtaining funding and the urgency of offloading the assets. This drives their prices down and thus exacerbates the funding and volatility problems.

[2] For a literature review on the subject, see Engel (1996).

The forward-premium puzzle gives an important theoretical insight into the way international financial markets work, by suggesting that the concept of rational expectations is not valid. It also suggests that the interest-rate spread is a poor guide to efficient resource allocation, owing to the volatility it imposes on future exchange-rate movements. This is a good example of market inefficiencies, suggesting the need for government intervention in the foreign-exchange market.

III. Consequences of the carry trade

This carry trade is not neutral for the value of the investment currency, because it causes it to appreciate, thereby constituting a failure of the efficient-markets hypothesis. The key argument for this appreciation trend is that large volumes of short-term capital tend to behave pro-cyclically; in other words, too much money enters in the growth phase of the business cycle and too much flows out in the downswing (Gallagher, 2011a and 2011b). These flows generate costs and distortions in the relationship between the domestic and external balance of the recipient countries.

It is not entirely clear how these equilibria occur, given the intense debate about the most appropriate exchange-rate regime to fulfil the objectives of resource balance and allocation in a globalization context. It is not clear whether the exchange rate should be totally fixed or totally flexible, or somewhere between these extremes, as Frankel (2003) argues. No less relevant in this debate is that the national authorities officially state that they are operating under a floating regime when, de facto, they practise a fixed or controlled regime, as Calvo and Reinhart (2000) aptly express in their concept of “fear of floating”. There is also no certainty that the extreme options —total rigidity or total flexibility— will lead the market to benignly determine the appropriate level of the exchange rate for a particular country. Nonetheless, under the floating exchange-rate regime, which has been used by emerging markets since the mid-1990s, the carry trade has undoubtedly grown astronomically.

In the debate that has arisen around exchange-rate protectionism, the argument put forward by Ffrench-Davis (2010) on the “neoliberal error” seems to be a sensible proposal. The error is to believe that any exchange-rate intervention goes against the market, when in fact two types of markets need to be distinguished. In the contemporary neoliberal monetarist approach, the market being favoured is the speculative credit market of short-term operators who only seek to maximize their rents. Under this rationale, the exchange rate responds more to variations in the capital account than to changes in the trade or current-account balance, which means that the exchange rate becomes a financial asset, as Keynes (2003) postulated.

In a correct approach, the appropriate market is that of the production of tradable and non-tradable goods in a context of innovation and technological change. The equilibria being pursued, and the appropriate allocations, should foster the creation of productive wealth that satisfies social needs.

The neoliberal error is responsible for the extreme volatility of nominal exchange rates, since their level responds more to variations in capital movements than to flows of goods and services. Thus, with a flexible exchange rate, the economy is exposed to procyclical variations caused by the actions of foreign portfolio investment funds, so the exchange rate becomes volatile because of the conditions imposed by financial operators. This has adverse effects for the national economy, as noted by Ffrench-Davis, namely: (i) distortions in the evaluation of projects for resource allocation; (ii) greater promotion of speculative than productive investments; (iii) displacement of the production tradable goods that can be imported; and (iv) a reduction in export value-added.

These effects can distort a development strategy led by the production of non-traditional exports which generate externalities and interact with small and medium-sized enterprises (SMEs). For this purpose, Ffrench-Davis argues, it is better to use intermediate dirty-floating regimes or moving bands, since targeting the correct market requires coherent and selective market intervention. In this context, it is clear that capital movements that cause exchange-rate appreciation tend to make domestic economies less competitive, which constrains export growth and fuels imports.

As a result, emerging countries would be right to design and implement macroprudential policies to control the negative effects of the currency carry trade. The response has been to develop strategies to inhibit the entry of speculative capital and prevent their currencies from appreciating. The application of these schemes and the responses to such actions have given rise to a “currency war” environment (Pérez, 2010), in a struggle to obtain a more competitive place in international trade. Joseph Stiglitz (2010) has pointed out how negative this war can be, since all countries could lose out; so solutions need to be found.

Discourse and practices in the design and management of exchange-rate protection policies recently resurfaced in the face of the urgent need to mitigate or avoid the negative effects of short-term cross-border capital flows. The trigger for the confrontation is the asymmetry between the monetary policies of the developed economies and those of the emerging economies, when the latter struggle to protect the dynamism of their growth, which is undermined by the industrial countries’ policies to cheapen their currencies and gain foreign exchange competitiveness to stimulate export-led growth.[3]

IV. Exchange-rate management

In the 1970s, neoliberal market fundamentalist attitudes defended capital-account liberalization at all costs, by arguing that capital flows and their mobility make it possible for countries with limited savings to attract funding for productive investment projects, diversify investment risk, promote intertemporal trade and contribute to the development of financial markets (Ostry and others, 2010a and 2010b). Over time, the evidence has shown that capital market liberalization in developing countries has not gone hand-in-hand with economic growth, and that any such association is more feasible in countries that have attained a high level of institutional development (Gallagher, 2010a). In general, emerging countries are more vulnerable to the adverse effects of short-term capital flows, because investment inflows for less than a year only target economies with stable political-economic characteristics, simply to obtain the higher potential return available from the carry trade. When the spread that drives this trade narrows or disappears, the capital in question abruptly flows out in search of higher-yielding alternatives in other countries.

These sudden reversals in capital flows tend to complicate macroeconomic management and aggravate financial risks. From the macroeconomic perspective, the concern is that waves of capital inflows exert upward pressure on the national currency and cause both the nominal and the real exchange rate to appreciate. This in turn means that less competitive domestic producers face cheaper imports and more expensive exports on the international market. This can cause lasting damage in the economy’s export sector, even if the capital inflows subsequently decrease or reverse.

From the standpoint of financial fragility, the concern is that excessive capital inflows can undermine financial stability due to increased external indebtedness and excessive exposure to foreign exchange risks. These factors can induce vigorous domestic credit expansions and asset price bubbles, with serious adverse effects in the event of a sudden reversal of the capital inflows (Ostry and others, 2010a and 2010b; De Gregorio, 2010; Gallagher, 2010b; López-Mejía, 1999; Magud and Reinhart, 2006; Gallagher and Coelho, 2010). Consideration should also be given to the costs of a sterilization policy and the constraints that may be imposed on fiscal policy. Hence the importance of IMF acceptance of the possible use of measures to control capital inflows.

[3] Nonetheless, Bergsten argues that the Federal Reserve’s quantitative-easing strategy does not constitute a market intervention, which is a curious way of perceiving national interests (Bergsten and Gagnon, 2012).

In February 2010, the Fund recognized that regulations on cross-border capital flows can be useful and constitute a legitimate macroeconomic policy tool (Ostry and others, 2010a and 2010b; IMF, 2011a and 2011b). Subsequently, in December 2012, IMF (2012a and 2012b) officially propounded a new "institutional approach" on capital-account liberalization that nations can consider to avoid and mitigate exchange-rate volatility and avert financial crises. This position diverges dramatically from that accepted in the neoliberal era of the 1970s, and is closer to the positions adopted by John Maynard Keynes and Harry Dexter White in their debate between 1941 and 1945 that capital controls should be an essential part of the smooth functioning of the global monetary-financial system (Gallagher, 2011a).

For IMF, the priority when developing exchange-rate management policies should be based on measures that strengthen the countries' ability to absorb capital flows (IMF, 2011b). The principle is that countries need to be better prepared for the inflow of such capital and not impede its mobility, since it leads to the implementation of structural reforms that increase the capacity of domestic financial markets to process the corresponding flows.

The Fund distinguishes between different types of measure, according to whether or not they aim to prevent the free flow of capital. The capital flow management measures (CFMs) that control the flow of capital are the most relevant, because they can be used as substitutes for macroeconomic policies that are appropriate and necessary for developing economies, and may have negative externalities for other countries (IMF, 2011b, p. 6).

Capital flow management measures are also of two types: residency-based CFMs and non-residency-based CFMs. The former, commonly referred to as capital controls, affect cross-border financial activity by discriminating on the basis of residency. These measures are more circumstantial in response to capital inflows. In contrast, the second type are prudential measures designed to ensure the resilience and soundness of financial institutions, such as capitalization ratios, loan-to-value ratios, limits on net open foreign exchange positions and limits on foreign-currency mortgages. These non-residency-based measures also include some that are typically applied in the non-financial sector, such as minimum periods of stay or taxes on some types of investment. Non-residence-based measures do not have the same macroeconomic and multilateral effects as those that use a residency parameter, such as limiting currency appreciation or redirecting flows to other countries. The fundamental difference between these two types of measure is whether they directly or indirectly affect the free mobility of capital flows.

In the case of flows, it is firstly appropriate to use macroeconomic policies and, mainly, allow the exchange rate to strengthen, accumulate reserves or adjust the balance between fiscal and monetary policies. CFM measures should be applied only when the appropriate macroeconomic conditions exist, which means that the exchange rate is not undervalued, that reserves are more than sufficient, and that the economy is overheated, which makes it unwise to lower interest rates. CFM measures need to be complementary to a contractionary fiscal policy and consider the lags associated with the macroeconomic effects of fiscal consolidation. For IMF, the application of CFM measures should have a low priority to avoid affecting other countries that participate in a multilateral reference framework.

The change in the Fund's stance has been the result of a wide-ranging debate, since until recently postulating capital control policies was going against the basic neoclassical premise that emerging economies should free their capital accounts as part of a broad process of financial liberalization needed to stimulate economic growth and stability (Gallagher, 2010b). Nonetheless, IMF continues to uphold its traditional principle by insisting that the key thing is to avoid impediments to the free mobility of capital owing to the negative effect this may have on the countries that generate the funds in question, usually the industrial countries. Less relevant is the cost of maintaining this position by the recipient, generally emerging, countries. It is in this context of theoretical debate that this paper examines the support for free capital movements espoused and practised by the Mexican authorities.

V. Exchange-rate volatility and public policy in Mexico

The appreciation of the peso from mid-2009 to April 2011 did not cause the Mexican authorities to support the control of speculative flows adopted by the authorities of other countries. The Mexican monetary authorities, particularly Banco de México and the Ministry of Finance and Public Credit (SHCP), continue to insist that peso appreciation is of no concern for the Mexican economy, so there is no need for exchange-rate management measures.

It has been argued that the exchange-rate flotation regime works correctly to absorb external shocks; and the official position is also that the exchange rate should not be used as policy tool. What should be done is to protect stability and ensure responsible public-finance management and prudent decision-making (Banco de México, 2013). The official position considers that the risks for the Mexican economy come mainly from the slowdown in the United States growth rate and, to a lesser extent, from fiscal uncertainty in some European countries. Moreover, exchange-rate appreciation is assumed to reflect the soundness of the fundamentals of the Mexican economy given its low inflation and competitive advantages, such as geographic location, a solid skills base, and low transport and logistics costs (Piz, 2011a and 2011b). Nor is it accepted that the advantage of the export sector depends on the exchange rate.

Banco de México (BANXICO) and SHCP have argued that policies based on CFM measures do not have a sustainable effect in the medium term; and that controls on capital flows are not appropriate for Mexico, since the country is at an advanced stage of financial deepening. Basically, the authorities view these capital flows as unmistakable evidence that the Mexican economy is moving in the right direction thanks to sound public-policy management.[4] Nonetheless, the decisive factor in this official position is the shielding that the Mexican economy has available to absorb international turbulence, which is based on a policy of accumulation of international reserves and the credit line approved by IMF. This official position reflects the market fundamentalism with which the Mexican economy has been managed since the late 1980s, which is wholly consistent with the IMF view that countries should use exchange-rate management policies more to adapt their economies to receive the capital in question than to impede its entry.

The authorities' position is not shared by other sectors. A large part of the business community and domestic producers support diverging positions with a generalized view that the strength of the dollar is making foreign products cheaper in the domestic market and making export products more expensive. Various business organizations have urged the authorities to defend the competitiveness of industry and prevent the currency from appreciating, since, in their opinion, an over-strong currency causes loss of competitiveness with negative effects on exports from many sectors (Monroy, 2011).

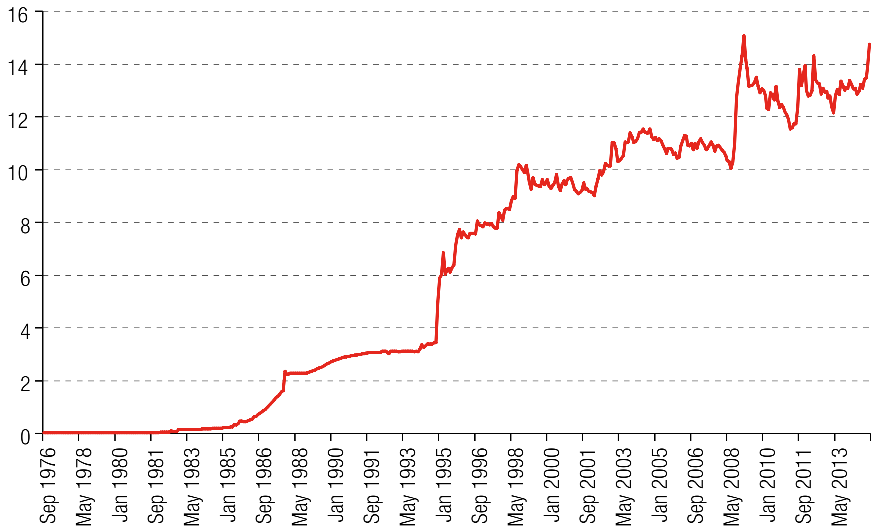
What has been clear is that the views of business leaders and the monetary authorities regarding the effects of the appreciation of the domestic currency do not coincide; but what complicates the debate is that the peso appreciation does not seem to be an obstacle to export growth. It is therefore useful to identify the factors that could give rise to these different opinions on the effect of the nominal exchange rate and the real exchange rate on the economy, but mainly on the cost for Mexico of receiving these flows of footloose capital.

[4] These positions stand in contrast to those adopted by the Central Bank of Australia, a developed country which, in the same period as Mexico, experienced a considerable currency carry trade and, consequently, a strong appreciation of its currency. The bank warns that the strength of the Australian dollar has created high risks for the economy, because manufacturing companies have lost competitiveness, with negative effects on employment (*The Economist*, 2012, p. 30).

VI. The evolution of the super-peso

Since the crisis of 1994, the Mexican peso has been extremely volatile, despite the fact that the country has been subject to various monetary and exchange-rate regimes. Figure 1 shows that, since the abandonment of the fully fixed exchange-rate regime of 12.50 pesos to the dollar in 1976, the nominal parity has lost value on a long-term basis. Between 1976 and 1994, this resulted from devaluations implemented under the logic of the fixed-but-adjustable exchange-rate regime. Starting in 1995, despite the shift towards a free-floating regime, the parity between the peso and the dollar has maintained its long-term depreciating trend in a context of greater volatility with short-term fluctuations.

Figure 1
Mexico: nominal exchange rate, 1976-2014
(Pesos per dollar)

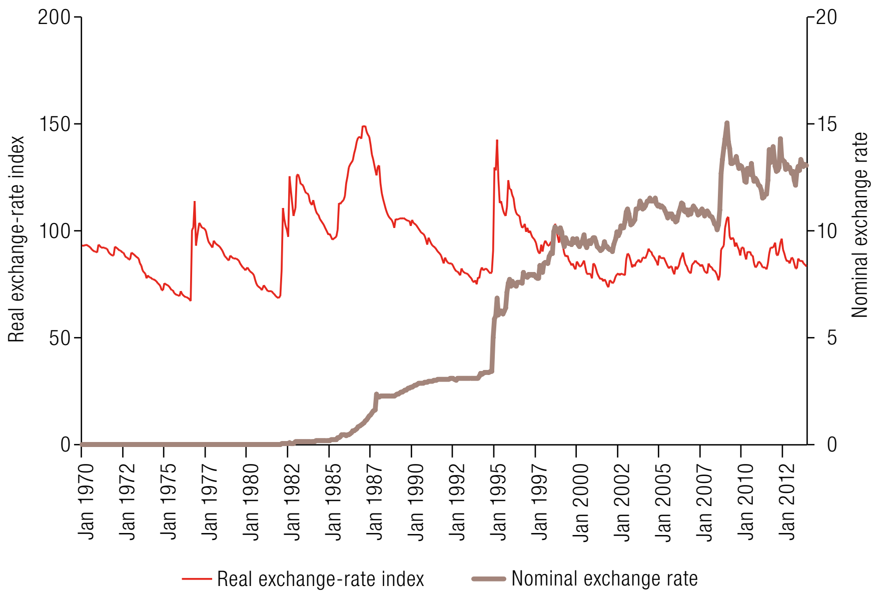


Source: Prepared by the authors, on the basis of data from Banco de México.

Figure 1 shows that, since the 2008 crisis, the nominal exchange rate has appreciated strongly: from 15.06 pesos per dollar in February 2009, the dollar price dropped to 11.52 in April 2011. This appreciation of 23.5 % places the value of the peso in its mid-2004 range, but far from the levels of under 11 pesos per dollar that it experienced in mid-2008. The most that can be said is that this appreciation has returned the nominal value of the Mexican currency to the long-term trend imposed by the free-floating regime. Analysing the problem of the exchange rate exclusively from this perspective lends support to the Mexican authorities' belief that the exchange rate does not require management.

The long-term trend of nominal and real exchange rates (see figure 2), which spans the different periods of overvaluation and undervaluation experienced by the Mexican peso against the dollar since 1970, displays two clear features. The first is the divergence of these two trends: while the nominal exchange rate trends towards persistent long-term depreciation, the real exchange rate fluctuates cyclically between appreciations and depreciations. The key factor underlying these divergences is the pattern of Mexican inflation versus that of the United States.

Figure 2
Mexico: trends in the real and nominal exchange rate, 1970-2014
(Mexican pesos)



Source: Prepared by the authors, on the basis of data from Banco de México.

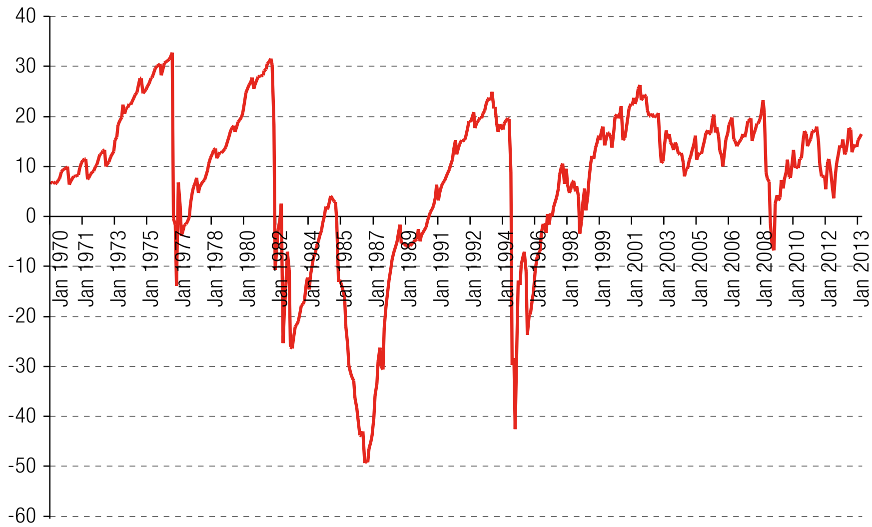
The second fact is that the Mexican economy has basically operated with an overvalued exchange rate for the last 40 years, except for brief periods of undervaluation: October 1976 to June 1977; March 1982 to October 1984; July 1985 to July 1990; January 1995 to January 1997; and in September and October 1998. From late 1998 to 2012, there was a persistent overvaluation, although with slight variations in degree.

Theoretically, having an open economy pursuing export-led growth means that Mexico should maintain a slightly undervalued exchange rate, to afford it greater competitiveness, which was first accepted as a principle of the exchange rate policy in the National Economic Development Plan of President Miguel de la Madrid Hurtado (1983-1988). This was consistent with the postulates of the Washington Consensus (Williamson, 1990; Kuczynski and Williamson, 2003). So it seems that Mexico has constantly swum against the tide in relation to the exchange rate, as explained below.

Figure 3 shows variations in the degree of overvaluation and undervaluation of the peso against the dollar between 1970 and 2012, making it possible to analyse the dynamics of the foreign-exchange market, based on the percentage difference between the nominal and real exchange rates relative to a given base year. Negative values indicate an undervaluation of the national currency against the dollar, and positive values represent an overvaluation, showing the degree of imbalance that has persisted in the exchange rate. From April 2009 to June 2012, the average monthly overvaluation was 7.2%, and over a medium-term period (from January 2000 to February 2012), the overvaluation was 14.81% on average, although it peaked at 17.3% in May 2011. These degrees of overvaluation are what determine the "super-peso" nature of Mexico's national currency.

An overvalued exchange rate makes foreign products cheaper in national currency relative to domestic products, owing to the increase in the purchasing power of foreign currency with respect to the local currency, and this fuels import growth. In contrast, an undervalued exchange rate reduces the purchasing power of the local currency relative to the foreign currency, making foreign products relatively more expensive than domestic products, which stimulates exports.

Figure 3
Mexico: undervaluation and overvaluation of the Mexican peso, 1970-2014
(Percentages)



Source: Prepared by the authors, on the basis of data from Banco de México.

The steady growth of export and import flows, however, shows that the overvaluation of the Mexican peso has not impeded the success of the export policy. In fact, Mexico has become one of the world's leading exporters, shedding the mantle of a country that exports mainly primary products to become an exporter of manufactures. The evidence shows that, in the exchange-rate appreciation stage from March 2009 to March 2012, Mexican exports grew by 75% and imports by 66.2%. From any perspective, these data suggest that the overvaluation of the peso does not undermine Mexican exports (Mold and Rozo, 2006; Rozo, 2009).

These observations show that in an economy that relies heavily on imported inputs, an overvalued exchange rate is not necessarily bad for exports. Overvaluation is convenient for the export activity on which Mexico's current development model is based, which is mainly undertaken by transnational corporations. This situation could explain the Mexican authorities' indifference to the exchange-rate overvaluation.

So, should the exchange rate appreciation not be a cause for concern? The dilemma is that the nominal exchange rate appreciation experienced by the currencies of Mexico and other emerging countries does not merely reflect a temporary trade imbalance, but is the result of a large and rapid carry trade in speculative capital fuelled by a deliberate and long-term policy by industrial countries to speed up the economic recovery process through export activity, as shown in the next section. Such behaviour occurs regardless of the adverse effects that this may have on the trade competitiveness and growth and development capacities of emerging economies. To counter these pressures to hold investment currencies, the emerging economies have been forced to accumulate international reserves in record amounts. The real dilemma caused by these speculative capital inflows, which implies maintaining an exceptionally high level of reserves, is that they entail a high cost without evident benefits.

VII.Magnitude and yield of the capital flows

The currency carry trade to Mexico is driven by the attractive return on government debt instruments. The spread between the secondary interest rates on Mexican government bonds and the rate on United States Treasury bonds is the key factor fuelling these speculative capital flows. To demonstrate this mechanism, the return (R) can be measured by the formulation of the interest parity condition, in its simplest version:

$$R = i - (i^* + (TCE - TCN) / TCN)$$

where i represents the local interest rate, i^* is the foreign interest rate, TCE is the expected exchange rate and TCN the nominal exchange rate. If R is positive, it is worth investing in the local economy; if it is negative, it is better to invest abroad. When the yield is 0, it is irrelevant which country is chosen, since the return is the same in both. To carry out this analysis of returns in the Mexican economy, the local short-term benchmark interest rate was used, represented by the yield on 28-day Treasury Certificates (CETES), in view of the growing demand for this type of government securities. The foreign interest rate is represented by the United States effective federal funds rate, and the change in parity is measured by the nominal peso per dollar exchange rate —FIX— published by the Banco de México. Figure 4 shows the average monthly yield between early 2007 and 2013, calculated applying the interest parity concept explained above.

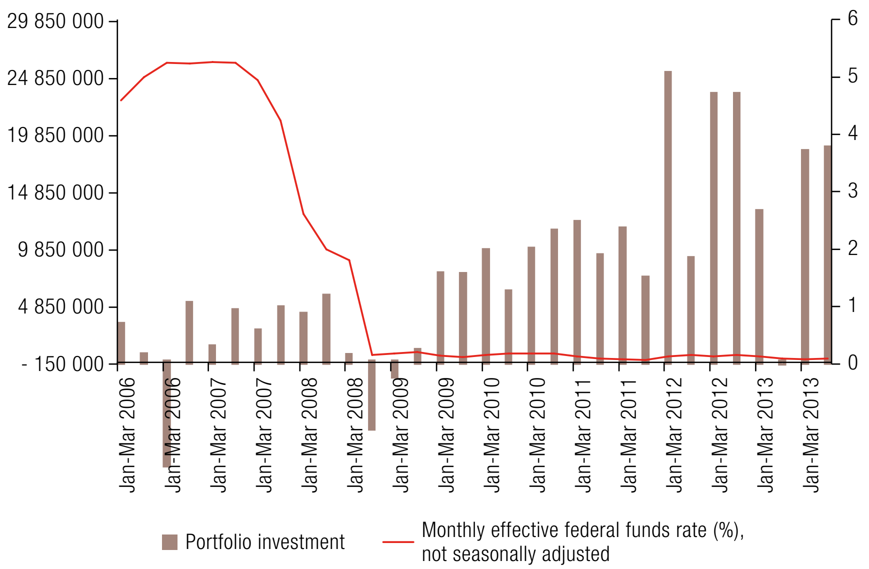
Figure 4
Mexico: trend of yields, 2007-2014
(Percentages)

Source: Prepared by the authors, on the basis of data from Banco de México.

The slowdown in the United States economy and the subsequent subprime crisis in 2008 forced the effective federal funds rate down from 5.25% in January 2007 to 0.25% in December 2008; and it was still at that level in mid-2015. Undoubtedly, this has encouraged speculative capital flight towards emerging markets; and Mexico in particular provides a haven, given its official open-door policy and the high yield on government debt. From January 2007 to December 2008, the fall in the effective federal funds rate in the United States raised the return in Mexico to a record level, reaching 15.84% on short-term debt instruments on December 18, 2008. Estimations made with short-term investment rationale show that the average rate of return of Mexican short-term debt instruments was 7.29% in 2013, a significant reduction but still about three times the yield obtained in developed economies, of barely 2.5% according to data from the National Institute of Statistics and Geography (INEGI).

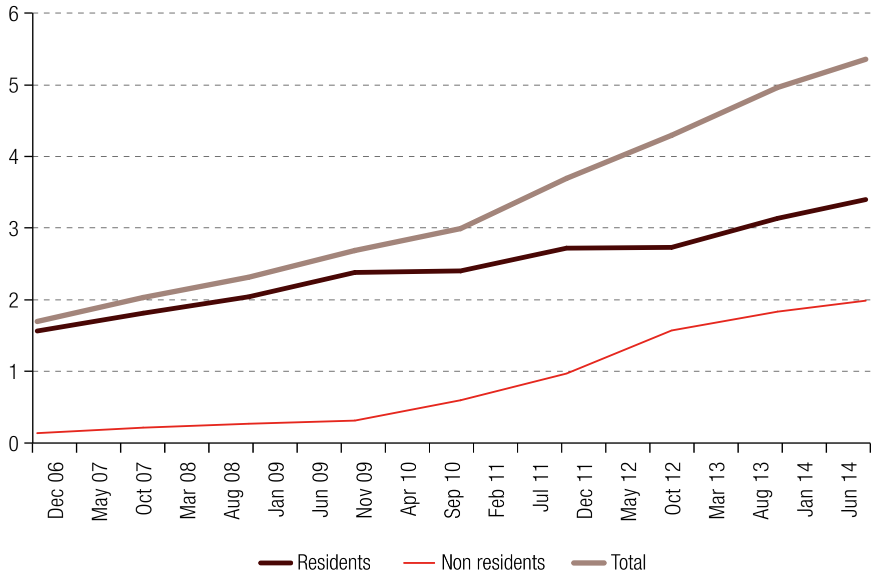
The higher yield obtainable in Mexico is what drives the large inflow of portfolio investment, as shown in figure 5. Between 2006 and 2008, these capital inflows behaved erratically with no defined trend; but from 2009 they display an unequivocal rapid growth path. The currency carry trade took off in late 2008, when the first phase of quantitative easing began (from November 2008 to March 2010), and it was consolidated with the second phase of the programme (November 2010 to June 2011). With the application of the third phase agreed on by the Federal Reserve on September 11, 2012, the high volume of speculative capital flows is maintained (see figure 6). In late 2008, non-residents held 251.114 billion pesos in Mexican government securities (11.7% of the total) —but by the end of 2014 their holdings had grown to 1.53 trillion pesos or 38% of the total. These flows also reveal a structural change in the composition of foreign investment in Mexico. Until 2009, inflows of foreign direct investment (FDI) outweighed portfolio investment flows, the former totalling US$ 17.331 billion, compared to the latter's US$ 15.261 billion. The portfolio investment share of total foreign investment grew from 29.5% in 2007 to 80.8% in 2012. Despite the approval of the structural reforms at the beginning of Enrique Peña Nieto's presidency, FDI inflows amounted to US$ 38.285 billion in 2013 (39.8% of the total foreign investment), while portfolio investment inflows amounted to US$ 50.359 billion, or 60.8%.

Figure 5
Mexico: comparison between the investment portfolio and the United States federal funds rate, 2006-2013
(Dollars and percentages)



Source: Prepared by the authors, on the basis of data from Banco de México and the Federal Reserve.

Figure 6
Mexico: trend of government securities, 2006-2014
(Billions of pesos)



Source: Prepared by the authors, on the basis of data from Banco de México.

It should not be surprising, then, that the effective United States federal funds rate and short-term investment flows into the Mexican economy were negatively correlated between 2006 and 2013. This means that, when the rate of United States interest falls, short-term foreign investment flows grow; so, as long as the United States economy maintains an ultra-expansive monetary policy with rock-bottom interest rates, yields in Mexico will continue to encourage the capital carry trade. High and secure returns are generating this flow of capital into Mexico, not the economy's sound fundamentals, as repeatedly claimed by the Mexican authorities. It also means that, when the Federal Reserve raises its interest rate, capital located in Mexico and other emerging markets will rapidly migrate to the United States, as occurred in early 2015, when a Federal Reserve interest rate hike seemed imminent towards the middle of the year. Since then, the United States economy has apparently not been able to make a robust and vigorous recovery, nor have the rest of the industrial countries; so the start of monetary policy normalization in the United States is likely to stay on hold.

Since 2010 Mexico has clearly become a highly attractive market for parking speculative capital, as a result of a public policy that encourages inflows by offering juicy and easy returns, backed by an open and explicit decision by the authorities not to impose restrictions on its earnings or stay in the country.

VIII. The cost of financial armour

Years ago, international reserves were used to manage the exchange rate that resulted from current-account supply and demand transactions; but today they manage the rate resulting from transactions on the capital account. This reflects the preponderance of financial globalization over commercial globalization; and what Peter Drucker (1986) has called the decoupling of the real economy from the financial economy is plain to see.

The starting point in strategic reserves management is to recognize that Mexico has not joined the list of economies that worry about the effects of short-term cross-border capital flows. By refusing to manage the exchange rate, the Mexican authorities have defended the policy of exchange-rate flexibility as the only valid option; and they have backed this up with a preventive policy of reserve accumulation and IMF loans. The public policy option has been to shield the economy from the turbulence that may be generated by the permanency and outflow of capital. The strategy of shielding the Mexican economy began in April 2009, when the Banco de México asked IMF to approve a Flexible Credit Line (LCF) for SDR 31.528 billion, or roughly US$ 47 billion, equivalent to 1,000% of Mexico's quota in the Fund), for a period of 12 months. In March 2010, this LCF was renewed for an additional 12 months; and, although Mexico did not need to use the available funds, it was considered that LCF met the objective of generating confidence in financial markets (Comisión de Cambios, 2010a, 2010b and 2010c). In January 2011, credit line was approved once again, but this time for two years and for US$ 72 billion (1,500% of the quota). These changes were due to the heightened uncertainty prevailing in international economic activity (Comisión de Cambios, 2011). At that time, the appreciation of the peso caused by speculative capital inflows could not be ignored. This policy of international reserves accumulation had raised official reserve assets to an historical level of US$ 195.682 billion at the end of 2014, which, in conjunction with LCF, imply financial shielding of US$ 270 billion.

The problem with this course of action is that it has high costs and is untransparent; but scant mention is made of this. The monetary authorities prefer to emphasize that the Mexican economy is well prepared to provide the liquidity needed to mitigate a shock caused by sudden capital flight, with a level of constantly rising reserves; but the cost involved in this strategy goes unmentioned. If reserves are treated simply as a portfolio asset, without considering their financing or the corresponding obligations, their accumulation is seen as a net gain. But when the issue of financing is considered, the net financial result of holding reserves can entail a loss for the country (Nugée, 2004; Rodrik, 2006).

There are at least four types of costs in reserve accumulation: the opportunity cost; the financial or sterilization cost; the balance sheet cost and the cost arising from a lack of incentives (Flores, 2010; Santaella, 2010; Cruz, 2006). Rodrik argues that there is also a social opportunity cost of public capital, which is what should be considered when analysing the cost of providing armour plating to an economy by accumulating liquid assets in the form of international reserves. Rodrik estimated that in 1995-2004 the cost of such reserve accumulation rose from 0.002% to nearly 1% of the GDP in developing countries (Rodrik, 2006).

This study estimates the financial and opportunity costs of accumulating reserves in Mexico. The financial cost is defined as the spread between the return on assets, in this case gross international reserves, and the cost of the liabilities represented by the monetary base and by monetary regulation deposits, as shown in table 1. The opportunity cost is defined as the difference between the return on assets and the cost of the external debt of the Federal Government, as presented in table 2. The cost analysis must be based on appropriate yield levels. Data from Santaella (2010) show that international reserve assets provided a return of 0.9% in 2009, while liabilities cost 5.75%. At first glance, there is a clear spread between these rates of close to 485 basis points.

The return on international reserve assets was low between 2008 and 2012 as a result of the extraordinary fall in the discount rate on first-tier instruments in which international reserves are invested, such as United States Treasury bonds. In contrast, liabilities are much more expensive, despite also decreasing over the same period. At first sight it is clear that the Mexican authorities have paid a higher cost to finance the reserves than the yield they obtain from them; and this cost has grown over time, from US$ 13.905 billion in 2009 to US$ 17.306 billion in 2013, although the amount was lower in 2014.

Table 1
Mexico: financial cost of accumulating international reserves, 2008-2014

Unit	Assets			Liabilities						Financial cost: differential between the yield of the asset and the cost of the liabilities (*US$ million*)	Percentage of GDP
	Gross international reserves (*US$ million*)	Yield		Bonds and monetary regulation deposits[b] (*US$ million*)	Yield		Monetary base (*US$ million*)	Yield			
		Rate (%)[a]	(*US$ million*)		Rate (%)[c]	(US$ million)		Rate (%)[c]	Yield (*US$ million*)		
Dec 2008	95 302	2.00	1 910	219 830	7.82	17 180	41 753	0	0	-15 270	1.72
Dec 2009	99 893	0.96	956	264 385	5.62	14 861	48 373	0	0	-13 905	1.50
Dec 2010	120 587	0.70	847	305 433	4.59	14 030	56 149	0	0	-13 182	1.23
Dec 2011	149 209	0.45	674	324 814	4.48	14 549	54 740	0	0	-13 875	1.33
Dec 2012	167 050	0.28	462	397 509	4.49	17 845	65 250	0	0	-17 383	1.44
Dec 2013	180 200	0.31	554	448 272	3.98	17 860	70 151	0	0	-17 306	1.40
Dec 2014	195 682	0.49	960	438 748	3.22	14 139	72 103	0	0	-13 178	1.14

Source: Prepared by the authors, on the basis of data from Banco de México.
[a] Rate of return on the assets in which international reserves are invested (annual average rate of two-year United States Treasury bonds).
[b] Refers to total government securities and IPAB securities in circulation.
[c] Reference rate (annual average bank funding rate): representative rate of wholesale transactions carried out by banks and brokerages in spot and one-day repo transactions with certificates of deposit, bank promissory notes and bank acceptances that have been settled in the delivery-versus-payment system of the Institute for the Deposit of Securities (INDEVAL).

Table 2
Mexico: opportunity cost of holding international reserves, 2008-2014

Unit	Assets			Liabilities			Opportunity cost: differential between its yield and the cost of gross public-sector external debt (*US$ million*)	Percentage of GDP
	Gross international reserves (*US$ million*)	Yield		Gross public-sector external debt (*US$ million*)	Cost			
		Rate (%)[a]	Yield (*US$ million*)		Rate (%)[b]	Yield (*US$ million*)		
Dec 2008	95 302	2.00	1 910	56 939.00	6.90	3 929	-2 019	-0.23
Dec 2009	99 893	0.96	956	96 354.00	5.09	4 904	-3 949	-0.43
Dec 2010	120 587	0.70	847	110 428.00	4.13	4 561	-3 714	-0.35
Dec 2011	149 209	0.45	674	116 420.00	5.59	6 508	-5 834	-0.56
Dec 2012	167 050	0.28	462	125 726.00	3.62	4 551	-4 089	-0.34
Dec 2013	180 200	0.31	554	134 436.00	2.60	3 493	-2 939	-0.24
Dec 2014	195 682	0.49	960	147 665.80	3.26	4 814	-3 854	-0.33

Source: Prepared by the authors on the basis of data from Banco de México and the Secretariat of Finance and Public Credit.
[a] Rate of return on the assets in which international reserves are invested (annual average rate of two-year United States Treasury bonds).
[b] Fixed interest rate of federal government bond issuances on the international capital markets.

The trend of the opportunity cost is more uneven, as can be seen in table 2, because in 2008 the assets were obtaining a high yield from the positive interest rates that still existed on financial markets; but rates subsequently plummeted to 0.31% in 2013, which caused these yields to decline gradually from US$ 1.919 billion in 2008 to US$ 554 million in 2013, although a higher yield was obtained in 2014.

The cost of liabilities also decreased, as the rate fell from 6.9% in 2008 to 3.26% in 2014. Nonetheless, the cost of these liabilities in absolute terms rose from US$ 3.929 billion to US$ 5.834 billion between 2008 and 2011. This constant growth in the opportunity cost of holding reserves reflects the fact that their volume practically doubled between 2008 and 2011. This cost trended down in the ensuing years to reach US$ 2.939 billion in 2013 as a result of the sharp fall in the rate of return, but it then rebounded in 2014. The relevant fact is that the opportunity cost has also grown gradually during these years, although by much less than the financial cost in absolute terms. These differences are fully captured by calculating the costs relative to gross domestic product (GDP) (see table 3). In these years the average annual financial cost was equivalent to 1.39% of GDP, while the average annual opportunity cost was 0.35% of GDP.

Table 3
Mexico: total cost of international reserves with respect to GDP, 2008-2014
(Percentages)

Year	Financial	Opportunity	LCF	Total
2008	1.72	0.23		1.95
2009	1.50	0.43	0.12	2.05
2010	1.23	0.35	0.10	1.68
2011	1.33	0.56	0.10	1.99
2012	1.44	0.34	0.09	1.87
2013	1.40	0.24	0.09	1.73
2014	1.14	0.33	0.09	1.56
Average	1.39	0.35	0.10	1.83

Source: Prepared by the authors.

To obtain a closer approximation to the cost of holding this level of reserves, the financial and opportunity cost must be augmented by the cost of contracting the IMF flexible credit line, which has an annual premium of US$ 1.08 billion. The total cost of financial shielding thus varied between 2.05% and 1.57% of GDP, and averaged 1.83% of GDP in these six years.

Like Rodrik, we consider this level of costs to be very high, regardless of the measure used. Suffice it consider that this amount far outweighs any of the anti-poverty programs that have been implemented in developing countries. For example, the *Progresa* programme in Mexico had an approximate cost of barely 0.02% of GDP in 2001, its last year of operation; and the *Oportunidades* programme, which replaced it, cost roughly 0.42% of GDP in 2010. In 2011, the financial shielding afforded by the accumulation of international reserves amounted to about 10 times the budget allocated to *Oportunidades*; and, as if that were not enough, it is equivalent to four times the budget of the Social Development Secretariat, which is tasked with combating poverty. Rodrik concludes that developing nations are paying a very high price to play by the rules of financial globalization (2006, p. 9).

The most serious aspect of this situation is the lack of evidence for the supposed benefits of this short-term borrowing. It is supposed to foster better financial intermediation, promote greater local investment and create greater risk-sharing opportunities; but this does not seem to be happening, given the low level of bank credit portfolios, which grew from 19% to 27% of GDP between 2000 and 2014 (CNBV, 2014, p. 16). As a result, a policy of accumulating reserves to deal with the accumulation of short-term liquid liabilities does not seem to make sense, since the main benefit is apparently a lengthening of debt maturities, which has allowed for a settlement horizon of longer than one year for about 80% of the debt.

A more appropriate and much less expensive policy would be to reduce exposure to short-term debt, as IMF itself has finally recognized —especially when Mexico's reserves far exceed adequacy indicators in respect of trade, debt and the money supply. Mexico's net international reserves represent 40% of annual imports, compared to the traditional 25% or three months' imports. Moreover, the Guidotti-Greenspan rule for the level of international reserves, which posits that countries should hold liquid reserves in an amount equal to their external liabilities falling due within a year (Rodrik, 2006, p. 5), would recommend reserves at just 60% of their current level.

IX. Conclusions

There is unequivocal evidence that the global crisis caused by the excesses of the United States mortgage sector spawned national economic policy practices that are inconsistent with the requirements of global stability. High unemployment and low levels of aggregate demand caused industrial countries to relax their monetary policies to an extent that would be unthinkable in normal conditions, but were made reality by the conditions of hardship experienced in these last five years. These loose monetary policies have given rise to an environment of exchange-rate volatility caused by the currency carry trade that has spread between industrial countries and developing countries —a situation leading to what is metaphorically referred to “a currency war”. What averted this war and has provided an opportunity to ease tensions is the IMF recognition that, in the current conditions of the world economy, some capital flow management practices are acceptable and positive for the stability of emerging markets.

Although Mexico has been one of the countries most affected by the carry trade, and given the latter’s effect on domestic currency appreciation, the authorities have totally ruled out the use of the exchange-rate management tools a number of emerging economies have implemented. This has partly been the result of the market fundamentalism with which the Mexican economy has been run since the late 1980s, which is entirely consistent with the traditional position of IMF that countries should use CFM policies more to adapt their economies to receive such capital, than to prevent its entry. This official attitude is also a consequence of the limited effect that Mexican peso appreciation has had on the economy’s export dynamics. What remains unclear clear is how the Mexican economy benefits from the entry and presence of short-term capital. What is clear, however, is that this reserve accumulation policy imposes a cost on the Mexican people that has averaged about 2% of GDP per year in 2009-2014. By any measurement standard, this seems a very high cost to pay to insure against an unlikely event, such as sudden capital flight, considering the amounts and the expected timeframes for the normalization of monetary policy by the Federal Reserve. To quote Rodrik, we conclude that Mexico is paying a very high price to play by the rules of financial globalization.

Bibliography

Banco de México (2013), *Reporte sobre el sistema financiero*, Mexico City, September.

___(2010), “Reporte sobre el sistema financiero. Junio de 2010”, Mexico City [online] http://www.banxico.org.mx/publicaciones-y-discursos/publicaciones/informes-periodicos/reporte-sf/%7BDC37ABCB-26F0-020D-145B-5CF397D62E68%7D.pdf.

Bergsten, C. F. and J. E. Gagnon (2012), “Currency manipulation, the US economy, and the global economic order”, *Policy Brief*, No. PB12-25, Washington, D.C., Peterson Institute of International Economics.

Brunnermeier, M. K., S. Nagel and L. H. Pedersen (2008), “Carry trade and currency crashes”, *NBER Working Paper*, No. 14473, Cambridge, Massachusetts, National Bureau of Economic Research (NBER).

Burnside, C., M. S. Eichenbaum and S. Rebelo (2011), “Carry trade and momentum in currency markets”, *NBER Working Paper*, No. 16942, Cambridge, Massachusetts, National Bureau of Economic Research (NBER).

Calvo, G. and C. Reinhart (2000), “Fear of floating”, *NBER Working Paper*, No. 7993, Cambridge, Massachusetts, National Bureau of Economic Research (NBER).

Clarida, R., J. Davis and N. Pedersen (2009), “Currency carry trade regimes: beyond the Fama regression”, *NBER Working Paper*, No. 15523, Cambridge, Massachusetts, National Bureau of Economic Research (NBER).

CNBV (National Banking and Securities Commission) (2014), *Ahorro financiero y financiamiento en México. Cifras a junio 2014*, Mexico City.

Comisión de Cambios (2011), “El director ejecutivo del FMI aprobó hoy la línea de crédito flexible de alrededor de 72 mil millones de dólares para México por dos años”, Mexico City.

___(2010a), "La Comisión de Cambios anuncia que se subastarán mensualmente opciones para vender dólares al Banco de México", *Comunicado*, Mexico City [online] http://calderon.presidencia.gob.mx/2010/02/la-comision-de-cambios-anuncia-que-se-subastaran-mensualmente-opciones-para-vender-dolares-al-banco-de-mexico/.

___(2010b), "Circular 8/2010. Asunto: subastas para la celebración de opciones de venta de dólares de los Estados Unidos de América", Mexico City, Banco de México.

___(2010c), "El Fondo Monetario Internacional renovó formalmente la línea de crédito flexible que tiene México", *Comunicado de Prensa*, Mexico City, Banco de México.

Cruz, M. (2006), "*¿Pueden las reservas internacionales contribuir al crecimiento económico?, Economía UNAM*, No. 8, vol. 3, Mexico City, National Autonomous University of Mexico.

De Gregorio, J. (2010), "Regulación macroprudencial, estabilidad financiera y flujos de capital", *Documentos de Política Económica*, No. 37, Santiago, Central Bank of Chile.

Dornbusch, R. (1976), "Expectations and exchange rate dynamics", *Journal of Political Economy*, vol. 84, No. 6.

Drucker, P. (1986), "The changed world economy", *Foreign Affairs*, vol. 64, No. 4.

Engel, C. (1996), "The forward discount anomaly and the risk premium: a survey of recent evidence", *Journal of Empirical Finance*, vol. 3, No. 2, Amsterdam, Elsevier.

Fama, E. F. (1984), "Forward and spot exchange rates", *Journal of Monetary Economics*, vol. 14, No. 3, Amsterdam, Elsevier.

Ffrench-Davis, R. (2010), "Macroeconomics for development: from 'financierism' to 'productivism', *CEPAL Review*, No. 102 (LC/G.2468-P), Santiago, Economic Commission for Latin America and the Caribbean (ECLAC).

Flores, L. A. (2010), "Costos de altas reservas internacionales", *El Economista* [online] http://eleconomista.com.mx/mercados-estadisticas/2010/06/29/costos-altas-reservas-internacionales.

Frankel, J. (2003), "Experience of and lessons from exchange rates regimes in emerging economies", *NBER Working Paper*, No. 10032, Cambridge, Massachusetts, National Bureau of Economic Research (NBER).

Gallagher, K. (2011a), "Regaining control? Capital controls and the global financial crisis", *PERI Working Paper*, No. 250, Amherst, Massachusetts, Political Economy Research Institute.

___(2011b), "Reforming United States trade and investment treaties for financial stability: the case of capital controls", *Investment Treaty News*, Art. 3, No. 3, International Institute for Sustainable Development.

___(2010a), "Control that capital", *Foreign Policy* [online] http://www.ase.tufts.edu/gdae/Pubs/rp/GallagherFPCapControl.pdf.

___(2010b), "Policy space to prevent and mitigate financial crises in trade and investment agreements", *G-24 Discussion Paper Series*, No. 58, New York, United Nations.

Gallagher, K. and B. Coelho (2010), "Capital controls and 21st century financial crises: evidence from Colombia and Thailand", *PERI Working Paper*, No. 213, Amherst, Massachusetts, Political Economy Research Institute.

IMF (International Monetary Fund) (2012a), "Liberalizing capital flows and managing outflows", *Policy Papers*, Washington, D.C., March.

___(2012b), "The liberalization and management of capital flows: an institutional view", *Policy Papers*, Washington D.C., November.

___(2011a), "Statement by the Managing Director on recent experiences in managing capital inflows. Cross-cutting themes and possible policy framework", Washington, D.C., March.

___(2011b), "Recent experiences in managing capital inflows, cross-cutting themes and possible policy framework", *Policy Paper*, Washington, D.C., February.

Jordá, Ó. and A. M. Taylor (2009), "The carry trade and fundamentals: nothing to fear but fear itself", *NBER Working Paper*, No. 15518, Cambridge, Massachusetts, National Bureau of Economic Research (NBER).

Keynes, J. M. (1936), *The General Theory of Employment, Interest and Money*, London, Macmillan.

Kuczynski, P. and J. Williamson (2003), *After the Washington Consensus. Restarting Growth and Reform in Latin America*, Washington, D.C., Institute for International Economics.

López-Mejía, A. (1999), "Large capital flows: causes consequences and policy responses", *Finance and Development*, Washington, D.C., International Monetary Fund, September.

Magud, N. and C. Reinhart (2006), "Capital controls: an evaluation", *NBER Working Paper*, No. 11973, Cambridge, Massachusetts, National Bureau of Economic Research (NBER).

Mold, A. and C. A. Rozo (2006), "Liberalization, growth and welfare: the Maquilizacion of the Mexican economy", *Trade, Growth, and Inequality in the Era of Globalization*, K. Sharma and O. Morrissey (eds.), London, Routledge.

Monroy, M. (2011), “Industriales piden frenar al súper peso”, *Expansión*, 21 January [online] http://www.cnnexpansion.com/economia/2011/01/21/Industriales-piden-frenar-al-súper-peso.

Nugée, J. (2004), “Manejo de reservas de oro y de divisas”, *Ensayos*, No. 71, Mexico City, Centre for Latin American Monetary Studies (CEMLA).

Ostry, J. and others (2010a), “Capital inflows: the role of controls”, *IMF Staff Position Note*, No. SPN/10/04, Washington, D.C., International Monetary Fund (IMF).

___(2010b), “Entradas de capital: el papel de los controles”, *Revista de Economía Institucional*, vol. 12, No. 23.

Pérez, C. (2010), “Peligro: guerra de divisas”, *El País*, 10 October.

Piz, V. F. (2011a), “Apreciación del tipo de cambio, manejable”, *El Financiero*, 7 April.

___(2011b), “México, imán de dólares”, *El Financiero*, 27 April.

Rodrik, D. (2010), “The end of an era in finance” [online] http://www.project-syndicate.org/commentary/rodrik41/English.

___(2006), “The social cost of foreign exchange reserves”, *NBER Working Paper*, No. 11952, Cambridge, Massachusetts, National Bureau of Economic Research (NBER).

Rozo, C. A. (2009), “Apertura, tratados de libre comercio y expectativas frustradas de crecimiento”, *Fin de época: de la integración tradicional al regionalismo estratégico*, A. Guerra-Borges (coord.), Mexico City, Siglo XXI.

Santaella, J. (2010), “Manejo de reservas internacionales y mercado cambiario”, Mexico City [online] http://observatorio.azc.uam.mx/pdf/CBM_8_6_2010_8.pdf.

Stiglitz, J. (2010), “No es momento para una guerra comercial”, *El País*, 2 May [online] http://elpais.com/diario/2010/05/02/negocio/1272806074_850215.html.

The Economist (2012), “Hitched to the China wagon”, 25 August.

Williamson, J. (1990), “What Washington means by policy reform”, *Latin American Adjustment: How Much has Happened?*, J. Williamson (ed.), Washington, D.C., Institute for International Economics.

The mining canon and the budget political cycle in Peru's district municipalities, 2002-2011

Carol Pebe, Norally Radas and Javier Torres

Abstract

This study seeks to determine whether access to a larger volume of funds from the mining canon affects the level of capital spending by district mayors in election years. The effect of Peru's electoral cycles on district public investment (from 2002 to 2011), and how this relates to the mining canon, is analysed in terms of the budget political cycle, using a fixed-effects panel model. The results show that the mining canon has a differential effect in the 20% of districts that receive the largest amounts; but, in general, there is no clear cyclicality between capital expenditure in these municipalities and election years.

Keywords

Mining, tax revenues, local government, municipal government, public expenditures, budget performance, political aspects, elections, Peru

JEL classification

D72, P16, H72, Q32, Q33

Authors

Carol Pebe Bernal holds a *licenciado* degree in Economics from the University of the Pacific of Peru. Email: carolpebe@gmail.com.

Norally Radas Kovalchuk holds a bachelor's degree in Economics from the University of the Pacific of Peru. Email: norallyfrk@gmail.com.

Javier Torres Gómez is a Professor in the Academic Department of Economics at the University of the Pacific of Peru. Email: j.torresgomez@up.edu.pe.

I. Introduction

The share of the extractive sectors in the Peruvian economy grew from 9.3% of gross domestic product (GDP) in 2004 to 13.2% in 2012; and this has led to a significant increase in the tax revenue received by the different governmental entities through canon transfers (mining, hydropower, gas and others).[1] Such transfers made to regional and local governments (municipalities) accounted for 23% of total transfers in 2004, and this percentage had grown to 43% by 2012.[2]

The mining canon consists of 50% of the income tax revenue received by the State for the exploitation of mineral, metallic and non-metallic resources. The revenue thus obtained is distributed as follows: 10% of the total canon is allocated to the local governments of the district municipality or municipalities in which the natural resource in question is extracted; 25% is allocated to the local governments of the district and provincial municipalities in which the natural resource is extracted; 40% is allocated to the local governments of the respective department(s); and 25% is assigned to the corresponding regional governments; and, from that percentage, the latter are required to transfer 20% to the national universities in their jurisdiction.

Among the aforementioned entities, the main beneficiaries have been local governments, since they received a total of roughly S/. 6.5 billion under the canon to carry out public investment projects in 2012. These funds can be used to finance and co-finance public investment projects, including actions to provide public services and maintain infrastructure, medical clinics and hospitals, among other things. Regional and local governments can allocate up to 20% of these transfers to current expenditure, including the maintenance of public investment projects.

As a result, from 2002 to 2011, the district municipalities' total expenditure grew by 290% (see annex A1), with the largest increase occurring in capital spending, which in 2011 was 7.3 times its 2002 level. This growth was particularly pronounced in 2006 and 2008, when capital expenditure increased by 77% and 70%, respectively. Capital spending by the 20% of municipalities receiving the largest amounts through the canon increased by more than the average of other municipalities in most years of the period 2002-2011, particularly in 2006 (109%) and 2008 (78%).

In this context, it is important to analyse the influence of the political factor on the level of capital expenditure in the district municipalities and the timing of the execution of such expenditure. For example, a report by the Office of the Comptroller General of the Republic (2011) noted that the transfers made to some local governments had not resulted in improved service coverage and attention to the population's basic needs. As a possible explanation for this, the report mentioned misalignment between the political interests of the authorities and the long-term well-being of their localities.[3] [4]

The capacity of local governments to effectively manage public investment is another problem that is present in Peruvian public management. Such capacity is limited not only by a dysfunctional budget planning and appropriation system, obstacles and constraints on the implementation of activities, and

---

1 The canon is the share in the income and rents obtained by the State from natural resource extraction, which accrue to the regional and local governments, and universities and higher public institutes, in whose jurisdictions the activities in question take place. The following canons exist: mining, hydropower, gas, fishing and forestry, along with the oil canon and its additional component (*surcanon*).

2 Regional and local governments (provinces and districts) also receive other transfers under the Municipal Compensation Fund (FONCOMUN), the Camisea Fund for Socioeconomic Development (FOCAM), regional trusts, the Glass of Milk Programme, donations and royalties, among others.

3 The report also states that some municipalities misused public resources or made improper payments. For this reason, in July 2012, the Office of the Comptroller General announced the launch of an audit, referred to as "*Operativo Canon*", to determine whether the funds obtained by local and regional governments from the extraction of natural resources had been used legally and effectively.

4 Other possible causes are: (i) inadequate definition of project priorities to be implemented within the framework of the rules governing the use of the canon; (ii) lack of strategic planning to assess their localities' unmet basic needs, such as water and sanitation, health or education; and (iii) the weak public management capacity of regional and local governments.

poor coordination between public institutions and programmes; technical capacity constraints further compound this problem. Aragón and Casas (2009) found evidence that a lack of investment management capacity, in other words project management, accounting and finance, planning and coordination with other public entities, can undermine the capacity of local governments to increase public investment, even if they have financial resources available (Aragón and Casas, 2009).[5]

This study seeks to determine whether access to larger transfers through the mining canon creates an additional incentive for opportunistic expenditure execution by the district mayors. Districts with a larger volume of transfers would be more likely to increase the amount of capital spending in pre-election years.

The study focuses on the mining canon for three reasons: firstly, because of the importance of mining canon transfers in Peru, which represented 61% of all transfers under some type of canon in 2004-2012, with a rising trend over the last decade.[6]

Secondly, the exogenous nature of the canon means that the amount of funding transferred to each subnational government is determined by pre-established rules and ultimately depends on metal prices. This means that local authorities cannot directly influence the allocation of the funds in question. Previous studies have used the exogenous effect of increasing the canon to assess the impact of mining in different sectors of Peru's economy. Aragón and Rud (2013) used the case of a large-scale mine in a region of Peru (Cajamarca) to identify the existence of positive effects of mining activities on the living conditions of the local population, the main explanation being backward linkage. Gajate-Garrido (2013) analysed the effect of public spending on the nutritional status of children in Peru, using the unanticipated regional mining canon as an instrument to account for the endogenous nature of public spending. The results revealed a positive nutritional effect only for children in urban areas.

Thirdly, transfers in respect of the mining canon do not revert to the public treasury which is very important because it could create incentives to execute expenditure in the most convenient period.

The study focuses on district municipalities because they receive a large share of mining canon transfers (see annex A2). Moreover, the districts are the smallest administrative units of the State, and their actions affect the population directly. Lastly, district-level control and oversight can be complex, which would give mayors some discretion over their use of the funds.

The rest of this article is organized as follows: section II reviews the literature on the political budget cycle and natural resource abundance. Sections III and IV describe the data sources and methodology used, while section V analyses the results obtained. Lastly, section VI sets out the main conclusions.

II. Literature review

This section relates the literature on natural resource abundance to that on the political budget cycle.

1. Resource abundance

The effect of resource abundance was initially studied by Gelb and others (1988) and by Auty (1990), who empirically analysed the relationship between natural resource extraction (oil in particular) and economic growth. Nonetheless, it was the work of Sachs and Warner (1995, 1997 and 2001) that first recognized the robustness of the negative relationship and coined the expression "resource curse".

[5] This effect is not considered in this study, however.

[6] Although the 2008 financial crisis reduced the net earnings of the mining companies, these transfers recovered as from 2010, to reach a level of S/. 5,000 million in 2012.

According to Sachs and Warner (2011), countries with abundant natural resources tend to have high price levels, which directly erodes the competitiveness of other export sectors; so countries lose the possibility of achieving export-led growth.

On the other hand, there is a vast literature demonstrating that the adverse effect of natural resource abundance can be explained by other factors, such as the level of human capital or the quality of public institutions. For example, Gylfason (2001) showed that the negative effect stems from less concern for education by the authorities in resource-rich countries, which implies a lower level of expenditure. In terms of public management, Manzano and Rigobon (2001) showed that the over-indebtedness of natural-resource-rich countries is the real problem for growth. In addition, Hausmann and Rigobon (2002) analysed the structure of trade and showed that more diversified economies are less likely to suffer adverse effects from natural resources.

Thus, countries that consider natural capital as their greatest asset can develop a false sense of security and become negligent in managing those resources and accumulating human capital. By controlling for these effects through indirect variables and using other methods to measure natural resource abundance, Brunnscheweiler (2008) found a direct positive relationship between natural resources and economic growth in 1970-2000, but no significant results confirming a negative relation between resource abundance and institutional quality. Similarly, according to Bravo-Ortega and De Gregorio (2005), higher levels of education can offset the negative effects. Thus, on average, countries with high levels of human capital are likely to benefit from natural resources. In their estimation, the authors found a positive relationship between income level and natural resources. The explanations for the underdevelopment of countries with abundant natural resources come mainly from political science and economics. According to Isham and others (2005), from the political science standpoint, the two main effects proposed are the income effect and the anti-modernization effect. The first occurs when budget growth based on income from natural resource extraction affords government greater discretion and relaxes control by the population. Weaker control allows government spending to be used for purposes other than meeting the population's needs or to improve the economy's productivity. For example, Robinson, Torvik and Verdier (2005) noted that, by enhancing the value of holding power and providing politicians with larger tax revenues which they can use to influence the outcome of elections, natural-resource boom periods aggravate the misallocation of resources in the rest of the economy. Collier and Goderis (2007) claimed that natural resource rents stimulate unproductive lobbying and rent-seeking, and encourage inefficient distribution in favour of political support. In contrast, there are fewer incentives for modernization, both for the government and for workers, because budget funding (and hence the financing of expenditure) comes from a specific extractive activity that requires specialized skills. As a result, most of the population is excluded from resource-related activity (although it maintains a passive benefit), which reduces incentives to increase productivity in other sectors and among workers generally.

In economics, the expression "Dutch disease" is used to explain the apparently contradictory result of resource wealth being associated with lower growth rates. The exploitation of natural resources alters the relative prices of other goods and services, which triggers an internal reallocation of capital and labour from tradable to extractive sectors, and fosters growth in non-tradable sectors (services).[7] The specialization of the economy is efficient in the short run; but, when extraction of the resource declines (or it is completely depleted), the economy lacks other sectors that are sufficiently developed to maintain the previous production level or absorb new job-seekers.

Nonetheless, Collier (2010) noted that while the initial explanation was purely economic, the importance of political factors is increasingly clear. Deacon (2011) argued that a number of empirical

[7] Isham and others (2005) also mention the entrenched inequality effect. This refers to the different development trajectories in former colonies (determined by the timing and nature of decolonization, property right regimes and the type of crops grown), which make it possible to explain the effects of natural resources on their economies today.

patterns had motivated the study of interactions with political institutions; and it was found, in particular, that resource abundance is more likely to be a curse when: (i) governance and the rule of law are initially weak; and (ii) the resource is highly concentrated spatially.

2. Political budget cycle

The political budget cycle is defined as a periodic fluctuation of government fiscal policy induced by the cyclicality of elections (see Shi and Svensson (2001) for further information). Opportunistic behaviour by rulers is manifested through fiscal policy management (taxes, transfers and public spending) aimed at influencing voter preferences and increasing their chances of electoral triumph (Gámez and Ibarra-Yúnez, 2007).

The literature on the political budget cycle can be divided into two groups. One of them follows the theoretical model developed by Rogoff (1990), which focuses on the question of asymmetric information on politicians' levels of competency. Voters shape their voting expectations based on observable fiscal policy outcomes in that period. The basic assumption is that only the most competent mayors can increase expenditure (or observable expenditure) in a given period. In the pre-election year, these authorities increase spending to send a signal of competency to the electorate, so that it favours them in elections.[8] Signalling thus gives rise to political budget cycles.

In contrast, the models of the second group focus on the ability of mayors to manage fiscal policy so as to bias voter preferences (Shi and Svensson, 2001). If greater provision of public goods (or increased capital expenditure) can be understood by the population as a sign of a higher level of competency, all politicians would have the same incentives to increase spending in pre-election periods. Unlike Rogoff's model, these models argue that all politicians have the same capacity to increase public spending, irrespective of their level of competency.

Apart from the models' different emphases on signalling competency or on the ability of spending to influence voter preferences, the chief conclusion is that public spending is affected by the electoral calendar.[9]

Initially, empirical studies focused on proving the existence of the political budget cycle at the country level. For example, Shi and Svensson (2006) examined the presence of this cycle in a sample of developing and developed countries, and found that the cycle existed in both types, but to a greater degree in the former.

Recently, the greater information available at the subnational level has been used to test the existence of the political budget cycle at the local level. Akhmedov, Ravichev and Zhuravskaya (2002) found evidence of short political cycles in the regions of Russia in 1996-2001. Drazen and Eslava (2005 and 2008) studied the political budget cycle in Colombian municipalities and found evidence that it influences a change in the composition of expenditure in pre-electoral periods. Gámez and Ibarra-Yúnez (2007) also found evidence of the existence of this cycle in state-level public expenditure in Mexico.

To analyse how the benefits obtained from extractive industries influence elections, Monteiro (2009) and Monteiro and Ferraz (2012) evaluated the impact of oil royalties received by municipalities in Brazil and concluded that they generate benefits for incumbent governments and reduce political competition, although the effect fades away over the medium term.

[8] A mayor is considered "competent" if he/she needs less income to provide a certain level of public services.

[9] Pre-electoral management cannot be viewed only as an increase in expenditure or generation of deficits. Drazen and Eslava (2005) argue that it occurs through changes in the composition of spending, with a larger amount of resources being allocated to the most influential voters through changes in fiscal policy or in the expenses that are most highly valued by the majority of voters.

In Peru, Carranza, Chávez, and Valderrama (2006) studied the political economy of the national budget process and demonstrated the existence of political budget cycles in nonfinancial public spending during the Fujimori administration (1995-2000). Nonetheless, at the subnational level no evaluation has been made, either of the existence of the political budget cycle or of the effect of rents obtained from natural resource exploitation on the authorities' fiscal behaviour. The study that comes closest is that of Sanguinetti (2010), who analysed the effect of the canon on fiscal practices at the regional level. This author found that canon transfers alter the composition of spending and increase the relative importance of public investment. This result, however, is probably a consequence of the restrictions placed on the use of canon resources.

The present study is one of the few to analyse the cyclicality of the execution of Peruvian public spending at the district level and how this relates to the transfers received as a result of mineral extraction.

III. Data sources

The first part of the analysis is based on data from a balanced panel of 1,131 district municipalities for the period 2002-2011.[10] The base contains budget data, such as capital expenditure executed, resources received through the mining canon, and the general budget of each district.[11] The information comes from the Economic Transparency Portal of the Ministry of Economy and Finance, and is used in both absolute value (S/. million) and in per capita terms. To properly identify the effect of the canon, it was decided to subtract the budgeted amount of the mining canon from each district's budget, so that the general budget better reflects the availability of district resources. Lastly, given that political budget cycles have to be evaluated over a period spanning at least four years (the term of office in Peru), it is clear that the database covers two political cycles (2002-2006, 2006-2010 and 2011).

In addition to the budget data, control variables (demographic, labour and income) were incorporated, which could affect the level of capital expenditure. Thus, the projected population by district, the economically active proportion of the population by district, the departmental illiteracy rate, the rural proportion of the population by department, the dependency ratio by province, and the poverty level by department are all included. These data come from calculations and estimates made by the National Institute of Statistics and Informatics (INEI).

IV. Methodology

To verify the existence of a political budget cycle and evaluate the effect of the canon on it, the study analyses following: capital expenditure as a function of economic and demographic variables, a set of dummy variables per year and the amount transferred through the mining canon. In particular, the following models are proposed:

$$G_{i,t} = \alpha_i + X'_{i,t\square}\beta + \sum_{2002}^{2011}\gamma_t D_t + \delta_1 canon_{i,t} + \sum_{t=06,10}^{\square} D_t \delta_{2t} canon_{i,t\square} + \varepsilon_{it} \quad (1)$$

$$G_{i,t} = \alpha_i + X'_{i,t}\beta + \sum_{2002}^{2011}\gamma_t D_t + \delta_1 canon_{i,t} + \sum_{t=06,10}^{\square}\left(\sum_{g=1}^{g=5}\delta_{2gt} D_t D_g\right) + \varepsilon_{it} \quad (2)$$

10 Although the universe of Peruvian district municipalities in 2011 actually numbered 1,635, this study only considers those for which annual observations are available for the whole period analysed.

11 In the "budget" variable, the modified initial budget (MIP) is used, because this is what is actually received by the municipalities.

$G_{i,t}$ is the capital expenditure per capita executed by district i in year t. The variable $X_{i,t}$ groups together sociodemographic and economic controls (for example, the per capita municipal budget at the start of the period, the illiteracy rate and the rural proportion of the population). The sum of the dichotomous variables per year (D_t) reveals the political cycle of spending.

Since a re-elected mayor has more experience in the management and execution of expenditure than one who has been elected for the first time, the number of years in office (*reelec*) is also controlled for.

The effect of the mining canon is identified with two variables. The variable $cannon_{i,t}$ reflects the direct effect of the amount of the transfer (per capita) on expenditure, while the interaction of the dummy variables in the electoral years and the amount of the canon ($D_t\delta_{2t}cannon_{i,t}$) would identify the incremental effect on election spending.

There are two specifications. The first (1) incorporates the effect of the amount of the canon in a linear way, while the second (2) separates the districts into five groups (g), according to the amount they receive under the canon, in an attempt to capture possible discontinuous effects. Both estimates use fixed effects by districts and robust standard errors. Thus, the effect of the canon would be identified from the changes occurring between the district municipalities over time.

V. Empirical Estimation

1. Budgetary budget cycle

Table 1 displays the basic results of the model in per capita terms;[12] the effect of the budget (without considering canon transfers) is significant and positive —additional funds can be expected to increase spending capacity. Moreover, the mining canon variable has a positive and significant effect on capital expenditure. That is, the municipalities that have more resources available due to the mining canon usually execute greater capital expenditure.

The dummy variables that capture the effect per year are statistically significant, which suggests that capital expenditure is influenced by the execution period. It is also noted that the coefficients for 2006 and 2010 are higher than those of 2005 and 2009 in all specifications. Nonetheless, there is no clearly marked political cycle, since the coefficients estimated for the post-electoral years (2007 and 2011) are positive and higher than those estimated for years in which the elections were held.

Although several studies state that the political budget cycle should be observable in the pre-election years (2005 and 2009), the municipal elections were held on the third Sunday of November of the year in which the municipal authorities' mandate expired. This means that the relevant year for analysing public expenditure is that in which the elections were actually held (2006 and 2010), since the incumbent authorities had nearly a whole year to organize expenditures so as to influence voter preferences.[13]

[12] In the estimated model with variables measured in levels (which incorporates the size of the district population) there are no material differences.

[13] Although there are fiscal rules in place to limit expenditure increases in election years, the fact that they are not enforced means they do not act as a restraint on spending.

Table 1
Political budget cycle

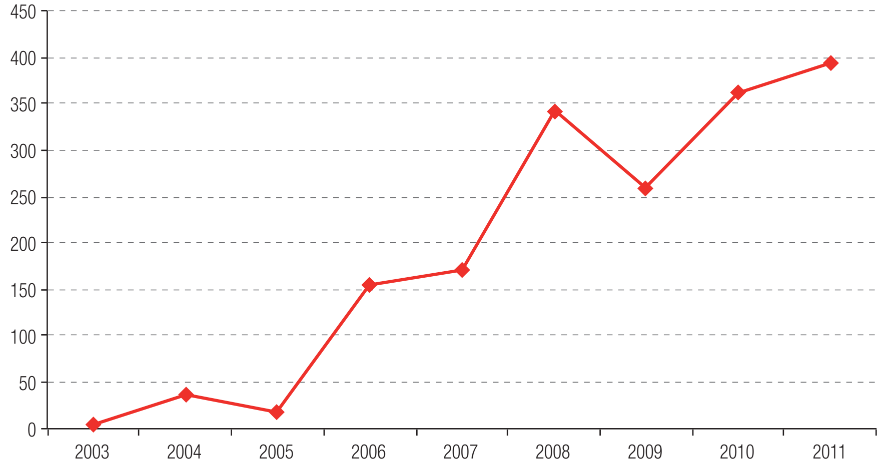
	(1)	(2)	(3)	(4)
Literate population				-0.00124
				(0.000283)***
Rural population				0.000378
				(0.000792)
Emergencies				2.13e-08
				(2.05e-07)
Re-election				-2.17e-06
				(6.89e-06)
Income (per capita)	0.154		0.130	0.129
	(0.0309)***		(0.0286)***	(0.0287)***
Mining canon (per capita)	0.171		0.170	0.170
	(0.0423)***		(0.0382)***	(0.0382)***
2003		1.81e-05	1.13e-05	4.12e-06
		(3.32e-06)***	(2.98e-06)***	(1.77e-05)
2004		6.97e-05	4.66e-05	3.55e-05
		(3.95e-06)***	(6.04e-06)***	(1.42e-05)**
2005		5.24e-05	2.78e-05	1.69e-05
		(6.83e-06)***	(6.97e-06)***	(2.02e-05)
2006		0.000213	0.000167	0.000155
		(1.52e-05)***	(1.45e-05)***	(3.51e-05)***
2007		0.000300	0.000203	0.000170
		(1.90e-05)***	(1.55e-05)***	(3.26e-05)***
2008		0.000507	0.000375	0.000341
		(2.88e-05)***	(2.40e-05)***	(4.69e-05)***
2009		0.000504	0.000296	0.000258
		(2.19e-05)***	(2.30e-05)***	(4.54e-05)***
2010		0.000590	0.000396	0.000361
		(3.90e-05)***	(3.78e-05)***	(8.12e-05)***
2011		0.000628	0.000437	0.000393
		(3.41e-05)***	(3.78e-05)***	(7.38e-05)***
Constant	0.000270	0.000125	9.17e-05	0.000113
	(2.08e-05)***	(1.43e-05)***	(9.30e-06)***	(0.000334)
No. of observations	11 310	11 310	11 310	11 310
R^2	0.377	0.161	0.448	0.449

Source: Prepared by the authors.
Note: Robust standard errors shown in parentheses. *denotes significance at 10%, **significance at 5% and ***significance at 1%.

Figure 1 shows the trend of the estimated expenditure, using the coefficients in column 4 and the average of the control variables, while assuming that the variation in the expenditure comes exclusively from the year-effect. This shows that, in 2006 and 2010, capital expenditure rose by an average of 37.1% compared to the immediately preceding years (2005 and 2009, respectively), with a much larger increase in 2006 than in 2010.[14] Nonetheless, subsequent capital expenditure in 2007 and 2011 also increased by roughly 4.6%.

[14] There was also a significant increase in 2008.

Figure 1
Predicted per capita capital expenditure deviation with respect to 2002
(Millions of soles)



Source: Prepared by the authors, on the basis of data from the Ministry of Economy and Finance (MEF).

In terms of the control variables, a relatively larger illiterate population is significantly associated with a lower level of capital expenditure. The proportion of the population who suffer from illiteracy could proxy for local management capacity, which, being smaller, would also be reflected in a lower rate of expenditure execution. On the other hand, labour participation (percentage of the population who are economically active) displays a positive and significant coefficient. The relationship between capital expenditure and the proportion of people living in rural areas is negative but not significant.[15] Similarly, neither the number of natural disasters nor the number of years in office appear to have significant effects on spending.

2. Effect of the canon on the political budget cycle

In order to determine whether access to the mining canon leads to higher capital expenditure in election years, multiplicative dummy variables were added to the foregoing model. In other words, the annual dummy variables were multiplied by the amount of the received through the canon in that period, so that the analysis focuses on the incremental effect of the mining canon in election years.

The results are reported in table 2. The first column shows that the coefficients of the annual multiplicative dummies in the pre-election years (2005 and 2009) are not significant. From the second column onwards, the analysis focuses on the election years. In the second specification, the coefficient of the multiplicative dummy variable for both years is significant and positive,[16] which would indicate that the mining canon has an incremental effect on capital expenditure in election years. The third specification also considers the years before and after the electoral period, to verify the robustness of the results. The effect is only significant for 2006 and loses all significance for 2010.

In the fourth and fifth columns the direct effect of the mining canon is eliminated and is incorporated through interactions with the annual dummy variables. The fourth column includes annual dummies for the years around the election, and the fifth column adds them for all years. The incremental effect of the canon is positive and, in this case, significant in both election years. Nonetheless, the coefficient of interaction for 2011 is even higher than that of 2010, although not significant.

[15] The “poverty” variable is highly correlated with the rural population, for which reason it has not been included. In addition, data are only available at the district level until 2010.

[16] Although for 2010 it is only significant at a 90% confidence level.

Table 2
Political budget cycle

	(1)	(2)	(3)	(4)	(5)
Literate population	-0.00116	-0.00116	-0.00116	-0.00108	-0.00110
	(0.000286)***	(0.000286)***	(0.000276)***	(0.000260)***	(0.000259)***
Rural population	7.32e-05	0.000359	2.15e-05	0.000283	0.000232
	(0.000694)	(0.000792)	(0.000656)	(0.000671)	(0.000658)
Emergencies	9.20e-08	5.81e-08	1.41e-07	9.32e-08	8.07e-08
	(1.89e-07)	(2.01e-07)	(1.82e-07)	(1.80e-07)	(1.77e-07)
Re-election	-2.80e-06	-1.87e-06	-1.18e-06	2.98e-07	-4.22e-07
	(7.09e-06)	(6.76e-06)	(5.97e-06)	(6.03e-06)	(6.25e-06)
Income (per capita)	0.140	0.134	0.139	0.200	0.193
	(0.0212)***	(0.0282)***	(0.0240)***	(0.0302)***	(0.0303)***
Mining canon (per capita)	0.206	0.145	0.182		
	(0.0488)***	(0.0375)***	(0.0592)***		
2003	-9.23e-07	5.02e-06	3.62e-06	1.04e-05	6.03e-07
	(1.87e-05)	(1.75e-05)	(1.67e-05)	(1.72e-05)	(1.85e-05)
2004	2.70e-05	3.57e-05	2.92e-05	3.18e-05	2.44e-05
	(1.25e-05)**	(1.46e-05)**	(1.36e-05)**	(1.53e-05)**	(1.94e-05)
2005	5.61e-06	1.75e-05	3.04e-06	6.32e-06	1.75e-06
	(1.78e-05)	(2.08e-05)	(1.69e-05)	(1.86e-05)	(1.92e-05)
2006	0.000140	0.000130	0.000115	0.000121	0.000115
	(2.97e-05)***	(3.07e-05)***	(2.38e-05)***	(2.45e-05)***	(2.47e-05)***
2007	0.000143	0.000179	0.000156	0.000187	0.000178
	(2.94e-05)***	(3.60e-05)***	(3.02e-05)***	(3.13e-05)***	(3.33e-05)***
2008	0.000311	0.000349	0.000319	0.000360	0.000321
	(3.86e-05)***	(5.01e-05)***	(4.31e-05)***	(4.59e-05)***	(4.29e-05)***
2009	0.000278	0.000267	0.000277	0.000288	0.000284
	(5.49e-05)***	(4.95e-05)***	(5.11e-05)***	(5.14e-05)***	(4.96e-05)***
2010	0.000323	0.000334	0.000304	0.000305	0.000302
	(6.69e-05)***	(7.08e-05)***	(5.54e-05)***	(6.18e-05)***	(6.22e-05)***
2011	0.000351	0.000400	0.000333	0.000341	0.000332
	(5.72e-05)***	(7.80e-05)***	(4.13e-05)***	(4.52e-05)***	(4.55e-05)***
Interaction: canon*2003					6.92e-05
					(3.68e-05)*
Interaction: canon*2004					4.45e-05
					(5.55e-05)
Interaction: canon*2005	2.98e-06		1.44e-05	3.23e-05	4.52e-05
	(1.80e-05)		(1.58e-05)	(1.63e-05)**	(2.23e-05)**
Interaction: canon*2006		4.48e-05	4.24e-05	5.99e-05	6.96e-05
		(1.18e-05)***	(1.46e-05)***	(1.54e-05)***	(1.91e-05)***
Interaction: canon*2007			-1.67e-06	1.62e-05	2.00e-05
			(6.83e-06)	(7.43e-06)**	(9.61e-06)**
Interaction: canon*2008					2.17e-05
					(1.25e-05)*
Interaction: canon*2009	-2.35e-05		-1.95e-05	-1.08e-05	-8.01e-06
	(1.50e-05)		(1.24e-05)	(1.09e-05)	(9.11e-06)
Interaction: canon*2010		1.51e-05	1.06e-05	2.03e-05	2.31e-05
		(8.85e-06)*	(8.67e-06)	(9.36e-06)**	(1.10e-05)**
Interaction: canon*2011			1.36e-05	2.31e-05	2.80e-05
			(1.70e-05)	(1.99e-05)	(2.13e-05)
Constant	0.000236	0.000108	0.000252	0.000105	0.000137
	(0.000289)	(0.000337)	(0.000274)	(0.000281)	(0.000277)
No. of observations	11 310	11 310	11 310	11 310	11 310
R^2	0.481	0.464	0.491	0.464	0.472

Source: Prepared by the authors.
Note: Robust standard errors shown in parentheses. *denotes significance at 10%, **significance at 5% and ***significance at 1%.

The conclusion is that the receipt of mining canon transfers had an incremental effect on capital expenditure in 2006; but the effect is less noticeable in the 2010 electoral cycle. Thus, a per capita increase in mining canon transfers results in a smaller increase in capital expenditure per person in 2010 than in 2006.

3. Categorization of municipalities according to the amount received under the mining canon

As an alternative approach (see equation (2), in the methodology section), the municipalities were classified in quintiles according to the amount received under the mining canon in 2011.[17] The transfers to the municipalities in the first group are between S/. 0 and S/. 7,000; those in the second group, between S/. 7,000 and S/. 98,000; those in the third group, between S/. 98,000 and S/. 466,000; those in the fourth group, between S/. 466,000 and S/. 1.549 million; and transfers to the fifth group exceed S/. 1.549 million. Each group consists of approximately 230 municipalities.

In addition, interactions with the annual dummy variables (since 2005) were created for all groups of municipalities: 24 interactions (6 binary variables per year for groups 2, 3, 4 and 5). This is done to evaluate whether capital expenditure is greater among municipalities that receive larger mining canon transfers. Table 3 presents the results.

In the first specification, the interactions of the pre-election years (2005 and 2009) are considered for all groups. The results are mixed: in 2005, the incremental effect is significant and negative for two groups of districts; in 2009, the effect is significant and positive for all groups except for the last one.

The second column considers the interactions in election years. In both electoral cycles (2006 and 2010) the incremental effect is positive and significant in the group 5 municipalities (which receive a larger amount under the canon). This means that the effect of canon transfers on capital expenditure is accentuated in election years by the 20% of municipalities that receive the largest transfers.

Apart from group 5, only group 4 reports positive coefficients in both electoral cycles; in the other groups the sign is negative. These results would support the hypothesis of the existence of a political budget cycle in municipalities that receive larger amounts through the mining canon.

To ensure the robustness of the results, the third specification included multiplicative variables in the years before and after the elections. This specification confirms that the incremental effect in the electoral year is positive and significant, but only for the group that receives the largest canon transfers in both elections, and for group 4 in the 2010 election.

[17] Although the total number of districts in 2011 was used in this classification, this does not differ significantly from the 1,131 districts considered for that year in subsequent analyses.

Table 3
Political budget cycle with incremental effect of the canon

	(1)	(2)	(3)	(4)
Literate population	-0.00109	-0.00116	-0.00106	-0.000959
	(0.000292)***	(0.000290)***	(0.000294)***	(0.000267)***
Rural population	0.000157	0.000375	0.000237	0.000566
	(0.000775)	(0.000800)	(0.000796)	(0.000902)
Emergencies	1.84e-08	6.63e-08	4.04e-08	2.67e-08
	(2.06e-07)	(1.94e-07)	(1.75e-07)	(1.78e-07)
Re-election	-2.24e-06	-2.30e-06	-2.95e-06	-1.61e-06
	(6.96e-06)	(6.84e-06)	(6.89e-06)	(6.59e-06)
Income (per capita)	0.130	0.128	0.130	0.197
	(0.0283)***	(0.0289)***	(0.0289)***	(0.0276)***
Mining canon (per capita)	0.171	0.169	0.167	
	(0.0383)***	(0.0378)***	(0.0381)***	
2003	2.73e-06	3.73e-06	1.17e-06	7.41e-06
	(1.79e-05)	(1.75e-05)	(1.76e-05)	(1.68e-05)
2004	3.36e-05	3.54e-05	3.33e-05	3.48e-05
	(1.41e-05)**	(1.40e-05)**	(1.37e-05)**	(1.56e-05)**
2005	4.20e-05	1.71e-05	3.32e-05	4.57e-05
	(2.73e-05)	(2.02e-05)	(2.26e-05)	(2.63e-05)*
2006	0.000151	0.000141	0.000137	0.000149
	(3.48e-05)***	(3.38e-05)***	(3.34e-05)***	(3.79e-05)***
2007	0.000166	0.000172	0.000135	0.000153
	(3.24e-05)***	(3.28e-05)***	(3.36e-05)***	(3.96e-05)***
2008	0.000337	0.000345	0.000342	0.000378
	(4.66e-05)***	(4.75e-05)***	(4.74e-05)***	(5.70e-05)***
2009	0.000206	0.000263	0.000202	0.000211
	(5.88e-05)***	(4.63e-05)***	(5.67e-05)***	(6.48e-05)***
2010	0.000355	0.000277	0.000271	0.000276
	(8.06e-05)***	(7.18e-05)***	(7.18e-05)***	(8.05e-05)***
2011	0.000386	0.000398	0.000450	0.000460
	(7.28e-05)***	(7.45e-05)***	(6.74e-05)***	(7.82e-05)***
Interaction: g2*2005	3.84e-06		-9.45e-06	-1.04e-05
	(1.57e-05)		(1.23e-05)	(1.18e-05)
Interaction: g2*2006		-1.23e-05	-1.55e-05	-1.65e-05
		(1.52e-05)	(1.70e-05)	(1.62e-05)
Interaction: g2*2007			-1.73e-05	-1.17e-05
			(1.72e-05)	(1.68e-05)
Interaction: g2*2009	9.41e-05		8.06e-05	7.90e-05
	(3.06e-05)***		(3.26e-05)**	(3.22e-05)**
Interaction: g2*2010		-5.20e-06	-5.48e-06	-1.42e-05
		(3.80e-05)	(4.01e-05)	(3.80e-05)
Interaction: g2*2011			-6.82e-05	-7.44e-05
			(4.66e-05)	(4.58e-05)
Interaction: g3*2005	-2.82e-05		-1.88e-05	-2.49e-05
	(1.54e-05)*		(1.24e-05)	(1.21e-05)**
Interaction: g3*2006		-1.43e-05	4.68e-06	1.20e-06
		(1.51e-05)	(1.64e-05)	(1.54e-05)
Interaction: g3*2007			4.39e-05	6.13e-05
			(2.10e-05)**	(2.12e-05)***
Interaction: g3*2009	0.000132		0.000142	0.000138
	(2.75e-05)***		(2.99e-05)***	(2.91e-05)***
Interaction: g3*2010		1.02e-05	3.20e-05	2.27e-05
		(3.74e-05)	(3.93e-05)	(3.74e-05)

Table 3 (concluded)

	(1)	(2)	(3)	(4)
Interaction: g3*2011			-3.10e-06	-1.15e-05
			(4.54e-05)	(4.43e-05)
Interaction: g4*2005	-1.24e-05		-9.64e-06	-1.73e-05
	(1.61e-05)		(1.38e-05)	(1.36e-05)
Interaction: g4*2006		1.49e-06	2.78e-06	5.43e-06
		(1.47e-05)	(1.52e-05)	(1.46e-05)
Interaction: g4*2007			4.25e-05	8.07e-05
			(2.31e-05)*	(2.32e-05)***
Interaction: g4*2009	8.46e-05		8.90e-05	9.31e-05
	(3.25e-05)***		(3.54e-05)**	(3.48e-05)***
Interaction: g4*2010		8.76e-05	9.04e-05	9.51e-05
		(3.61e-05)**	(3.90e-05)**	(3.72e-05)**
Interaction: g4*2011			-0.000107	-0.000104
			(4.70e-05)**	(4.56e-05)**
Interaction: g5*2005	-0.000103		-5.21e-05	-6.15e-05
	(3.13e-05)***		(2.05e-05)**	(2.19e-05)***
Interaction: g5*2006		0.000102	9.12e-05	0.000111
		(3.32e-05)***	(3.76e-05)**	(3.93e-05)***
Interaction: g5*2007			0.000105	0.000254
			(3.90e-05)***	(4.68e-05)***
Interaction: g5*2009	-7.62e-05		-1.93e-05	5.25e-05
	(0.000112)		(9.25e-05)	(7.35e-05)
Interaction: g5*2010		0.000356	0.000345	0.000405
		(9.52e-05)***	(0.000101)***	(0.000107)***
Interaction: g5*2011			-0.000117	-5.71e-05
			(0.000104)	(0.000114)
Constant	0.000190	0.000103	0.000152	-3.00e-05
	(0.000328)	(0.000338)	(0.000338)	(0.000388)
No. of observations	11 310	11 310	11 310	11 310
R*	0.450	0.454	0.456	0.423

Source: Prepared by the authors.
Note: Robust standard errors shown in parentheses. *denotes significance at 10%, **significance at 5% and ***significance at 1%.

VI. Conclusions

This study evaluates the importance of the political budget cycle and its interaction with the mining canon in the execution of capital expenditure in Peru's district municipalities. This is one of the few studies to analyse the cyclicality of the execution of public expenditure at the subnational level, and how this relates to the transfers received as a result of the extraction of mineral resources in Peru.

The first results show that, in general, there is no clear cyclicality between the capital expenditure of the municipalities and election years. Estimates of dummy variables per year show that on average, in 2006 and 2010 (capital expenditure increased by 37.1% over the previous year (2005 and 2009, respectively); and there was also a post-election rise of about 4.6% (in 2007 and 2011). Moreover, the amounts transferred through the mining canon had a positive and significant effect on capital expenditure in 2006. Yet, on average, its effect on the political budget cycle is marginal. Thus, a per capita increase in mining canon transfers results in a smaller increase in capital expenditure per person in 2010 than in 2006.

Segregating the municipalities according to the amount received through the canon shows that only the group receiving the largest amount undertakes significant additional capital expenditure in the two election years. In other words, capital expenditure increases significantly in election years only in the group that receives the largest transfers under the mining canon.

Bibliography

Akhmedov, A., A. Ravichev and E. Zhuravskaya (2002), "Opportunistic Political Cycles: Test in a Young Democracy Setting" [online] http://pdc.ceu.hu/archive/00001617/.

Aragón, F. and C. Casas (2009), "Capacidades técnicas y gasto local: el caso de las municipalidades peruanas", *Perspectivas. Análisis de temas críticos para el desarrollo sostenible*, vol. 7, No. 1, Andean Development Corporation.

Aragón, F. and J. P. Rud (2013), "Natural resources and local communities: evidence from a Peruvian gold mine", *American Economic Journal: Economic Policy*, vol. 5, No. 2, Nashville, Tennessee, American Economic Association.

___(2009), "¿Éxito en la sostenibilidad fiscal? El caso de Perú", *Who Decides the Budget? A Political Economy Analysis of the Budget Process in Latin America*, E. Stein, M. Hallerberg and C. Scartascini (eds.), Washington, D.C., Inter-American Development Bank (IDB).

Auty, R. (1990), *Resource-based Industrialization: Sowing the Oil in Eight Developing Countries*, New York, Oxford University Press.

Bravo-Ortega, C. and J. de Gregorio (2005), "The relative richness of the poor? Natural resources, human capital, and economic growth", *Policy Research Working Paper*, No. 3484, Washington, D.C., World Bank.

Brunnschweiler, C. (2008), "Cursing the blessings? Natural resource abundance, institutions and economic growth", *World Development*, vol. 36, No. 3, Amsterdam, Elsevier.

Carranza, L., J. Chávez and J. Valderrama (2006), *La economía política del proceso presupuestario: el caso peruano*, Washington, D.C., Inter-American Developmet Bank.

Collier, P. (2010), "The political economy of natural resources", *Social Research*, vol. 77, No. 4.

Collier, P. and B. Goderis (2007), "Commodity prices, growth, and the natural resource curse: reconciling a conundrum", *Working Paper*, No. 2007-15, University of Oxford.

Comptroller-General of the Republic (2011), *Distribución y utilización de los recursos del canon, sobrecanon, fondo de desarrollo socioeconómico del proyecto Camisea y regalía minera en el Perú*, Lima.

Deacon, R. T. (2011), "The political economy of the natural resource curse: a survey of theory and evidence", *Foundations and Trends in Microeconomics*, vol. 7, No. 2.

Drazen, A. and M. Eslava (2008), "Electoral manipulation via voter-friendly spending: theory and evidence", draft.

___(2005), "Electoral manipulation via expenditure composition: theory and evidence", *NBER Working Paper*, No. 11085, Cambridge, Massachusetts, National Bureau of Economic Research (NBER).

Gajate-Garrido, G. (2013), "Excluding the rural population: the impact of public expenditure on child malnutrition in Peru", *Policy Research Working Paper*, No. 6666, Washington, D.C., World Bank.

Gámez, C. and A. Ibarra-Yúnez (2007), "El ciclo político oportunista y el gasto de los estados mexicanos", *Gestión y Política Pública*, vol. 18, No. 1, Mexico City, Economic Research and Teaching Centre.

Gelb, A. and others (1988), *Oil Windfalls: Blessing or Curse*, New York, Oxford University Press.

Gylfason, T. (2001), "Natural resources, education and economic development", *European Economic Review*, vol. 45, No. 4-6, Amsterdam, Elsevier.

Hausmann, R. and R. Rigobon (2002), "An alternative interpretation of the 'resource curse': theory and policy implications", *NBER Working Paper*, No. 9424, Cambridge, Massachusetts, National Bureau of Economic Research.

Isham, J. and others (2005), "The varieties of resource experience: natural resource export structures and the political economy of economic growth", *World Bank Economic Review*, vol. 19, No. 2, Washington, D.C., World Bank.

Manzano, O. and R. Rigobon (2001), "Resource curse or debt overhang?", *NBER Working Paper*, No. 8390, Cambridge, Massachusetts, NBER National Bureau of Economic Research.

Monteiro, J. (2009), "Resource booms and politics: the effects of oil shocks on public goods and elections" [online] http://www.caf.com/attach/19/default/oil_politics_caf.pdf.

Monteiro, J. and C. Ferraz (2012), “Does Oil Make Leaders Unaccountable? Evidence from Brazil’s offshore oil boom” [online] http://www.unamur.be/en/eco/eeco/pdf/OilPaper2012-05.pdf.

Robinson, J. A., R. Torvik and T. Verdier (2005), “Political foundations of the resource curse”, Norwegian University of Science and Technology [online] http://www.svt.ntnu.no/iso/ragnar.torvik/bardhanjde.pdf.

Rogoff, K. (1990), “Equilibrium political budget cycles”, *American Economic Review*, vol. 80, No. 1, Nashville, Tennessee, American Economic Association.

Rogoff, K. and A. Sibert (1988), “Elections and macroeconomic policy cycles”, *The Review of Economic Studies*, vol. 55, No. 1, Oxford University Press.

Sachs, J. and A. Warner (2001), “The curse of natural resources”, *European Economic Review*, vol. 45, No. 4-6, Amsterdam, Elsevier.

___(1997), “Natural Resource Abundance and Economic Growth”, Cambridge, Massachusetts, Harvard University [online] http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.422.8285&rep=rep1&type=pdf.

___(1995), “Natural resource abundance and economic growth”, *NBER Working Paper*, No. 5398, Cambridge, Massachusetts, National Bureau of Economic Research (NBER).

Sanguinetti, P. (2010), “Canon minero y decisiones fiscales subnacionales en el Perú”, *Working Papers*, No. 2010/01 [online] https://www.caf.com/media/3180/201001SanguinettiFebrero2010.pdf.

Shi, M. and J. Svensson (2006), “Political budget cycles: do they differ across countries and why?”, *Journal of Public Economics*, vol. 90, No. 8-9, Amsterdam, Elsevier.

___(2001), “Conditional political budget cycles”, *CEPR Discussion Paper*, No. 3352 [online] http://people.su.se/~jsven/pbc.pdf.

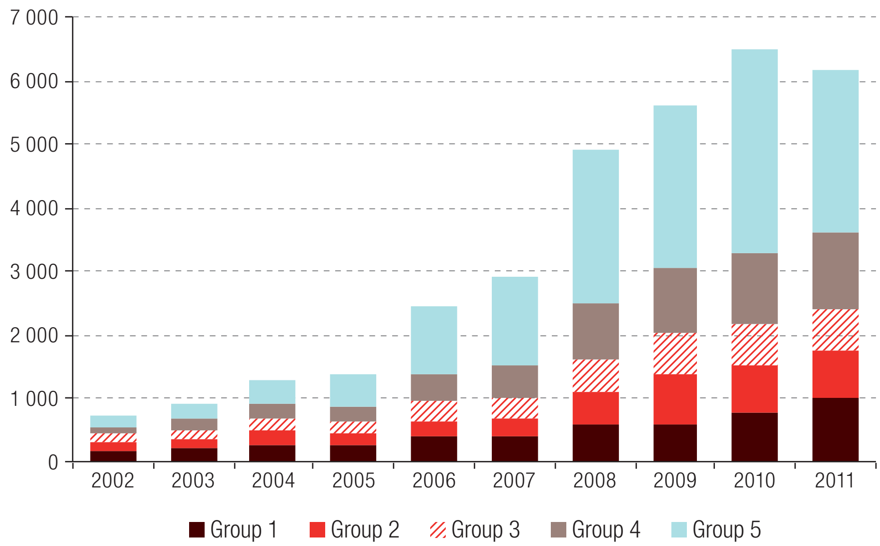
Annex A1

Table A1.1
District municipality categories according to the amount received under the mining canon in 2011
(Soles)

Category				
1	from	0	to	7 000
2	from	7 000	to	98 000
3	from	98 000	to	466 000
4	from	466 000	to	1 549 000
5	from	1 549 000	to	more

Source: Prepared by the authors.
Note: The classification by quintiles was based on the total number of districts according to the amount received under the canon in 2011.

Figure A1.1
Trend of capital expenditure executed by the district municipalities in each category (2002-2011)
(Billions of soles)



Source: Prepared by the authors, based on data from the Ministry of Economy and Finance (MEF).
Note: The classification by quintiles was based on the total number of districts according to the amount received under the canon in 2011.

Table A1.2
Trend of capital expenditure growth in the district municipalities by canon group
(Percentages)

Annual growth	2003	2004	2005	2006	2007	2008	2009	2010	2011
Group 1	34	21	-3	62	-1	55	1	28	28
Group 2	10	37	-17	34	16	73	48	-3	4
Group 3	9	52	-7	62	8	63	25	0	4
Group 4	15	56	9	71	19	60	21	9	4
Group 5	34	55	36	109	28	78	4	27	-20
Total	21	43	8	77	18	70	13	17	-5

Source: Prepared by the authors.
Note: The classification by quintiles was based on the total number of districts according to the amount received under the canon in 2011.

Annex A2

Table A2.1
Methodology used to distribute the mining canon

	Percentage	Recipient	Distribution criterion
Mining canon (50% of the income tax levied on the mining companies)	10%	District municipalities in which the resources are extracted	If there is more than one municipality, it will be distributed in equal parts.
	25%	District municipalities of the province in which the resources are extracted	According to population and unmet basic needs.
	40%	Provincial municipalities in the department where the resources are extracted	According to population and unmet basic needs.
	20%	Regional government	
	5%	National universities	

Source: Prepared by the authors, based on data from the Ministry of Economy and Finance (MEF), *Metodología de distribución*.

Annex A3

Table A3.1
Description of the political budget cycle sample variables

Variable	Districts	Years	Mean	Standard deviation	Minimum	Maximum
Departmental illiteracy rate	1 131	10	0.1211	0.0603	0.0228	0.2790
Proportion of rural departmental population	1 131	10	0.4161	0.2241	0.0000	0.8368
Number of emergencies per province	1 131	10	28.3321	45.8182	0.0000	271.0000
Number of years in office	1 131	10	2.9435	1.6953	1.0000	9.0000
Capital expenditure per capita in S/. million	1 131	10	0.0004	0.0008	0.0000	0.0254
Per capita MIP income in S/. million	1 131	10	0.0007	0.0021	0.0000	0.0944
Mining canon per capita in S/. million	1 131	10	0.0002	0.0013	0.0000	0.0470
Mining canon in S/. million	1 131	10	1.1613	8.5546	0.0000	476.9459

Source: Prepared by the authors.

Table A3.2
Description of the sample for the political organizations model, 2002, 2006 and 2010

Variable	Districts	Years	Mean	Standard deviation	Minimum	Maximum
Departmental illiteracy rate	1 131	3	0.1000	0.1000	0.0000	0.3000
Departmental poverty level	1 131	3	51.2000	19.8000	8.7000	88.7000
Number of emergencies per province	1 131	3	24.5000	43.2000	0.0000	271.0000
Income under the modified initial budget in S/. million	1 131	3	4.8000	16.3000	0.2000	431.1000
Mining canon in S/. million	1 131	3	0.9000	8.7000	0.0000	424.9000
Population	1 131	3	14 704.4000	49 909.6000	184.0000	983 095.0000

Source: Prepared by the authors.

Table A3.3
Description of the sample for the political organizations model, 2006 and 2010

Variable	Districts	Years	Mean	Standard deviation	Minimum	Maximum
Departmental illiteracy rate	1 131	2	0.1000	0.1000	0.0000	0.2000
Departmental poverty level	1 131	2	46.2000	19.9000	8.7000	88.7000
Number of emergencies per province	1 131	2	31.4000	49.5000	0.0000	271.0000
Income under the modified initial budget in S/. million	1 131	2	6.3000	19.3000	0.3000	431.1000
Mining canon in S/. million	1 131	2	1.4000	10.7000	0.0000	424.9000
Population	1 131	2	15 126.2000	51 910.5000	184.0000	983 095.0000

Source: Prepared by the authors.

A structuralist-Keynesian model for determining the optimum real exchange rate for Brazil's economic development process: 1999-2015

André Nassif, Carmen Feijó and Eliane Araújo

Abstract

The "optimum" long-run real exchange rate is the rate that will efficiently channel production resources into industries that generate and diffuse productivity gains in the economy as a whole and that will thus tend to speed up and sustain the economic development process. Rather than employing conventional models, a structuralist-Keynesian model is used to demonstrate, both theoretically and empirically, that the factors influencing the path of the long-run real exchange rate and the divergence of the observed real exchange rate from the "optimum" real exchange rate in terms of economic development are accounted for by both structural and short-term macroeconomic policy variables. Econometric estimates for 1999-2015 indicate that, following a prolonged period, beginning in late 2005, during which the Brazilian currency appreciated quite steeply, the real exchange rate in Brazil reached its "optimum" level in mid-January 2016.

Keywords

Economic development, structural adjustment, economic convergence, macroeconomics, foreign exchange rates, monetary policy, development models, econometric models, Brazil

JEL classification

F30, F32, F39

Authors

André Nassif is an Adjunct Professor with the Department of Economics of the Universidade Federal Fluminense, Brazil. Email: andrenassif27@gmail.com.

Carmen Feijó is a Professor with the Department of Economics of the Universidade Federal Fluminense, Brazil. Email: cbfeijo@gmail.com.

Eliane Araújo is an Associate Professor with the Department of Economics of the Universidade Estadual de Maringá, Brazil. Email: elianedearaujo@gmail.com.

I. Introduction

In his work in the field of structuralist development macroeconomics, Professor Luiz Carlos Bresser-Pereira has placed a great deal of emphasis on the preponderant role of the real exchange rate[1] in the development and economic convergence or catching-up process[2] of developing countries (see Bresser-Pereira, 2010).[3] He is by no means attempting to reduce the highly complex phenomenon of economic development —in which not only economic forces but historical, social, cultural and many other factors are at work— to a single variable (the real exchange rate). On the contrary, assuming that the various factors influencing the development process are helping to move it forward, in a recent book published along with his co-authors (Bresser-Pereira, Oreiro and Marconi, 2014), Professor Bresser-Pereira argues that a development process will be sustainable provided that the country in question has a stable rate of inflation, an average real interest rate below the average real rate of return on capital, average wage levels that rise in tandem with the economy's productivity and an exchange rate that can be regarded as competitive.[4] Nonetheless, these authors do take the position that the real exchange rate is the most important macroeconomic price of all these variables because it influences all the others (including the inflation rate). In their words:

> Key macroeconomic variables are composed of the current account deficit and the exchange rate […] Imports, exports, the investment rate, the savings rate, and inflation depend on the [real] exchange rate. Investments depend on it: We may think of the [real] exchange rate as the light switch that connects or disconnects the efficient business enterprises existing in a country from foreign markets and their own domestic markets (Bresser-Pereira, Oreiro and Marconi, 2014, p. 3).

It is no coincidence that the national currencies of Asian economies that are successfully catching up to developed-countries in terms of per capita income and well-being levels (especially the Republic of Korea, the Chinese province of Taiwan and Singapore) and those that continue to pursue this strategy (such as China and India), are rarely overvalued for a long period of time, as has been the case of Brazil over the past few decades.[5]

Oddly enough, the theoretical literature on economic development does not attribute due importance to the real exchange rate as a crucial strategic variable.[6] Nevertheless, there is a vast body of empirical literature that seeks to evaluate the relationship between the trend of the real exchange rate and long-term economic growth. An impressive number of econometric studies have reached the

[1] In this paper, the exchange rate is defined as the domestic price of a unit of foreign currency (e.g. the price of US$ 1.00 in reais). Accordingly, an increase in the exchange rate signals a depreciation of the national currency, while a decrease signals an appreciation.

[2] The process of catching-up is understood as the convergence of poor countries towards per capita income levels and levels of well-being similar to those found in developed countries. This concept is a central element in both the classical and neoclassical economic development literature, even though these two schools of thought are radically different from one another. In the classical development literature, see Rosenstein-Rodan (1943), Prebisch (1950), Lewis (1954), Myrdal (1957), Hirschman (1958) and Kaldor (1966), among others, while, in the neoclassical literature, see Solow (1956), Swan (1956), Romer (1986) and Lucas (1988), among others.

[3] Although some authors have acknowledged its importance (especially Kaldor (1966 and 1978)), the relationship between the real exchange rate and economic development has been analysed more thoroughly in empirical studies than in theoretical ones. As will be discussed later on, several recent empirical studies have concluded that, *ceteris paribus*, countries in the process of catching up whose currencies are marginally undervalued in real terms tend to show faster long-term growth. See Rodrik (2008), Williamson (2008) and Berg and Miao (2010). For the case of developing countries, see Gala (2008) and Araújo (2009).

[4] An exchange rate is deemed to be competitive if the national currency is slightly undervalued in relation to the United States dollar (or a basket of currencies) in real terms. As will be discussed later on in this article, the empirical evidence suggest that a "competitive" exchange rate will, *ceteris paribus*, tend to speed up the economic development process.

[5] See Amsden (1989 and 2001), Nassif, Feijó and Araújo (2011) for a discussion of the situation in Brazil, and Nassif, Feijó and Araújo (2015a) for an analysis of the cases of China and India.

[6] As noted by Gala (2008, p. 273), "while the econometric literature on this issue is relatively rich, theoretical analysis of channels through which real exchange rate levels could affect economic development are very scarce".

conclusion that, unless it is a "natural" result of the increased productivity of tradable goods compared with non-tradables —especially services— (a phenomenon that is captured by the Harrod-Balassa-Samuelson effect), an overvaluation of developing countries' exchange rates for lengthy periods of time will tend to depress their long-term economic growth rates (Razin and Collins, 1999; Dollar and Kraay, 2003; Prasad, Rajan and Subramanian, 2006; Gala, 2008).

Recently, a number of more thorough empirical studies have shown not only that an overvalued exchange rate hampers economic development, but also that a real exchange rate that is slightly higher than the long-run equilibrium exchange rate (equivalent to a slight depreciation in real terms) tends to boost the pace of development. While this empirical finding was first presented by Rodrik (2008) and later confirmed by Berg and Miao (2010), Williamson (2008, p. 14), a renowned specialist in real exchange rates, also concluded that: "Indeed, the very best policy (in terms of maximizing growth) appears to be a small undervaluation". Williamson's emphasis on the word "small" is important, since, clearly, a large undervaluation of the real exchange rate (as measured against the "neutral" equilibrium real exchange rate) that is compatible with purchasing power parity in real terms will tend to fuel more lasting inflationary processes.[7]

In a previous paper (see Nassif, Feijó and Araújo, 2011) we proposed a theoretical and econometric methodology for estimating the long-run path of the real exchange rate in developing economies. This study introduced the concept of an "optimum" (or "competitive") long-run real exchange rate —defined as the rate that will efficiently channel resources into industries that generate and diffuse productivity gains in the economy as a whole and thus, *ceteris paribus*, sustain and accelerate economic development— and found an econometric estimate of that variable (for the first time ever in Brazil, as far as we are aware). Empirical evidence suggests that such industries are located in the manufacturing sector, which is regarded as the main driver of productivity growth both in that sector and in the economy as a whole, as was noted by Kaldor (1966) in his pioneering study based on the empirical regularity originally identified by the Dutch economist Verdoorn (1949).[8]

This concept of the "optimum" real exchange rate is similar to that of the "industrial equilibrium real exchange rate" proposed by Bresser-Pereira (2010). According to that author, the industrial equilibrium exchange rate is the rate that would keep firms at the state of the art of technology in their respective industry. Despite similarities, the industrial equilibrium exchange rate is not exactly the same as the long-run "optimum" real exchange rate, because the latter —being marginally depreciated in relation to its long-term trajectory— is not necessarily an equilibrium exchange rate, but is instead comparable with the "industrial equilibrium" or the "neutral" purchasing power parity equilibrium.[9] Since the term "optimum", as used here, has nothing to do with maximization methods or equilibrium rates, it will appear in quotation marks throughout this study.

Our first study (see Nassif, Feijó and Araújo, 2011) concluded that the real exchange rate reached its "optimum" level (the estimated long-run real exchange rate rather than the observed rate) in 2004 (average for the period). In April 2011, when the "optimum" real exchange rate should have been 2.90 reais per dollar, whereas the observed rate was just 1.59 reais per dollar, there was a (mega) overvaluation of 82% in relation to the competitive rate needed to sustain Brazil's economic development process.

7 As shown by Krugman and Taylor (1978), when a country introduces a sharp, one-off correction for a large overvaluation (e.g. in Brazil in 2015), the inflationary and recessionary effects (since real wages decline) in the short run are clearly observed. However, as soon as economic agents incorporate this new equilibrium level of relative prices, ceteris paribus, the real average rate of return on capital in production activities will tend to rise and, in consequence, economic growth and productivity will increase and be sustained over the long term.

8 For recent evidence of this empirical regularity (known as the Kaldor-Verdoorn law) in Latin America, see Ros (2014).

9 In addition, although the estimation of the real industrial equilibrium exchange rate is not performed using maximization methods either, the resulting real rate of depreciation in relation to the neutral long-term equilibrium exchange rate may or may not be very high. For the proposed concept of a real "optimum" exchange rate, the level of real depreciation is only slightly above the "neutral" real exchange-rate equilibrium.

Considering that the Brazilian real depreciated sharply in nominal terms throughout 2015, this paper has two main goals: (i) to refine the theoretical discussion of the proposed structuralist-Keynesian model in order to distinguish it from conventional models, both in terms of the theoretical and empirical considerations relating to the path of the long-term real exchange rate and in terms of the misalignment of the actual real exchange rate from the optimal level; and (ii) to estimate once again the optimal real exchange rate in order to identify whether the observed level prevailing at in early 2016 was above (undervalued), below (overvalued) or close (or even equal) to the optimal level. The article is divided as follows. Section II offers a critical analysis of the conventional models for determining the real long-run exchange rate and its misalignment with respect to its long-term equilibrium level. Section III presents the proposed structuralist-Keynesian model, noting its theoretical and empirical methodological differences from the conventional models. Section IV sets forth the econometric estimation of both the long-term path of the real exchange rate in Brazil in the 1999-2015 period and the level considered "optimum" for purposes of economic development. Section V presents the study's conclusions.

II. The conventional model for the determination of the long-run real exchange rate and the degree of misalignment: a critical analysis

The mainstream theory about the behaviour of the long-run real exchange rate (RER) is based on the purchasing power parity hypothesis.[10] According to this hypothesis, in order for the purchasing power of two currencies (expressed in the same monetary unit) to remain constant over time, the nominal exchange rate in the market (expressed as the domestic price of a unit of foreign currency —for example, the price of one dollar in reais) should be adjusted by the difference between the domestic and international inflation rates.[11] The change in the real exchange rate over time can be expressed as:

$$RE\dot{R} = \dot{e} - \left(\dot{P} - \dot{P}^{*}\right) \tag{1}$$

where RER is the real exchange rate; e is the nominal exchange rate; P is the domestic price level; and P^* is the external price level (e.g. of the United States). As the dots appearing above the different variables denote instantaneous changes over time, equation (1) shows that the increase in the real exchange rate over time (i.e. a real depreciation of the domestic currency against the foreign currency) should be equal to the increase in the nominal exchange rate (in other words, to the nominal depreciation of the domestic currency) minus the differential between the domestic and external inflation rates.

That definition assumes that an increase in RER or e will result in a depreciation in the domestic currency against the foreign currency (real and nominal, respectively), while a reduction in RER or e will produce an appreciation of the domestic currency against the foreign currency (real and nominal, respectively). The main theoretical issue has to do with the key forces leading the real exchange rate to its equilibrium level in the long-term, so that it will be equal, in this case, to what is regarded as being the "neutral"—from the standpoint of competitiveness— nominal exchange rate (a "neutral" exchange rate means that exporters, importers and domestic producers who compete with foreign producers

[10] The real purchasing power parity hypothesis was developed by Cassel in 1918 and, since then, has become the principal theoretical point of reference for evaluating the behaviour of the long-run real exchange rate (see Sarno and Taylor, 2002). The empirical evidence does not bear this hypothesis out in its absolute form, however (see Sarno and Taylor, 2002). Even though the empirical evidence for the validity of the relative version of the real purchasing power hypothesis is not very robust either, Rogoff (1996, p. 647) contends that most economists instinctively believe in some variant of purchasing power parity as an anchor for long-term real exchange rates.

[11] For an excellent mathematical demonstration based on the absolute version of the purchasing power parity hypothesis, see Simonsen and Cysne (1995, pp. 99 and 100).

are equally benefited). According to the conventional theory, in the absence of nominal or real shocks, there are "fundamental" forces inherent in the capitalist economic system that will cause the nominal exchange rate to converge towards its real equilibrium level over the long term (Taylor and Taylor, 2004).[12] Any divergence of the real exchange rate from its fundamental equilibrium level would be seen as a transitory deviation caused by random shocks (Razin and Collins, 1999).

Not by chance, in conventional empirical estimates of the real exchange-rate trajectory as well as of the deviation of the real exchange rate from its long-term equilibrium, the two fundamental forces that explain such trends are: (i) the relation between change in the productivity of tradable goods compared with the change in the productivity of non-tradable goods; and (ii) the terms-of-trade (ToT) behaviour.

In the case of the first structural economic force, as productivity tends to increase faster in tradable goods than in non-tradable goods in the process of economic development, the drop in relative prices of the former implies that the domestic currency will tend to "naturally" appreciate in real terms. This is the well-known Harrod-Balassa-Samuelson effect, in which, as noted by Obstfeld and Rogoff (1996, chap. 4), price levels tend to rise (or, in other words, the real exchange rate tends to appreciate) as per capita income increases.[13]

Yet, as to the second "fundamental" force, the expected impact of ToT on the real exchange-rate trajectory is ambiguous. Baffes, Elbadawi and O'Connell (1999, p. 413) state (as do most authors) that: "An improvement in the terms of trade increases national income measured in imported goods; this exerts a pure spending effect that raises the demand for all goods and appreciates the real exchange rate". However, Edwards (1989) provides a theoretical demonstration that it can also have the opposite effect: if the increase in income generated by an improvement in the terms of trade triggers the large-scale substitution of tradables for non-tradables (notably services), then the increase in the relative prices of the former will translate into a depreciation in the currency in real terms. In other words, according to Edwards (1989), the impact of an improvement in the terms of trade on the real exchange rate can go either way: if the income effect is greater, there will be a real appreciation of the currency, but if the substitution effect is greater, there will be a real depreciation.[14]

Having identified the two main forces that influence the fundamental equilibrium path of the real exchange rate, the next step of conventional models is to use econometric methods to estimate the misalignment of the exchange rates (i.e. to estimate the percentage of overvaluation or undervaluation of the nominal exchange rate relative to its fundamental equilibrium level). Conventionally, this misalignment tends to be associated with transitory random shocks.[15] Viewed from this perspective, a real exchange rate is misaligned when "it deviates from the underlying RER that would have prevailed in the absence of price rigidities, frictions and other short-run factors", as noted by Razin and Collins (1999, pp. 59-60). While the models are of differing degrees of sophistication, the econometric equations that are intended to measure the extent of exchange rate misalignment are, essentially, estimating the deviation of the observed real exchange rate —which is estimated, for example, by the Central Bank of Brazil based on the real exchange rate indices relating to equation (1)— from a linear combination of proxy variables for

[12] According to Taylor and Taylor (2004), the empirical evidence shows that, because of the rigidity of nominal prices, nominal exchange rate variations are transmitted on a 1-to-1 basis to the real exchange rate within a very short period of time. In other words, a nominal depreciation will immediately trigger a more or less proportional real depreciation.

[13] As will be seen later on, when expressed econometrically, the sign of the estimated coefficient for per capita income (the variable used to capture the Harrod-Balassa-Samuelson effect) is negative because the increase in per capita income tends to lower (appreciate) the real exchange rate.

[14] The expected sign of the estimated coefficient for the terms of trade in the econometric equation shown in section III may be negative, if the income effect predominates, or positive, if the substitution effect prevails.

[15] The main policy implication of this approach, according to which shocks that cause the real exchange rate to diverge from its long-run equilibrium path are short-lived, is that the most suitable exchange regime would be a clean (or nearly clean) float.

the path of the long-run real equilibrium exchange rate (associated with flex-prices). In the econometric calculations, these deviations are associated, in the final analysis, with variables representing short-term shocks plus the term that represents the residual of the regression (Razin and Collins, 1999, pp. 65-67). In formal terms, the real exchange rate is expressed as follows in the conventional theoretical models (all the variables except the interest rate are expressed in logarithms):

$$RER_t = g_t\left(y_t^s, d_t, i^*\right) + f_t\left(\lambda_{mt}, \lambda_{yt}\right) \tag{2}$$

where the RER (all the t subscripts represent time t) is simultaneously determined by two sets of factors: (i) the first, represented by the function g (...), basically incorporates fundamental variables that presumably would push the real exchange rate towards its long-term equilibrium level. Therefore, all the variables represented by g are real variables: y^s is the real output; d is real aggregate demand and i^* is the real world interest rate (which is compatible with the long-term "natural" interest rate). In a world where there is perfect competition and no rigidity in nominal prices and no unforeseen random shocks, the RER would naturally converge towards g (...) over the long term and would thus be aligned with the equilibrium level determined by economic fundamentals.

Nevertheless, given the imperfections present in the real world, the conventional theory attributes the causes of real exchange rate misalignments relative to the fundamental long-term equilibrium path to the variables represented by f (...) —whether they be monetary shocks (represented by λ_m) or real shocks (represented by λ_y).

Thus, to estimate the path of the long-run real exchange rate and its misalignment relative to its equilibrium path, equation (2) can be translated into the following econometric equation (see Razin and Collins, 1999, pp. 64-65).[16]

$$RER_t = \alpha W_t + \beta Z_t + \varepsilon_t \tag{3}$$

"Fundamental" variables (determine the long-run real exchange rate)

Short-run shock variables plus the term that represents the residual of the regression (which determines the exchange-rate misalignment)

Where RER is the observed real exchange rate (expressed in logarithms), W is a set of variables that capture fundamental long-term equilibrium factors and Z is a set of variables representing short-term shocks that, together with the term that represents the residual ε, explains the exchange rates' misalignment.[17]

[16] It is useful to characterize both the theoretical and the econometric approaches to determining the long-run real exchange rate and the method used for estimating exchange-rate misalignments in conventional models, because they help to mark the radical differences of these models from the structuralist-Keynesian theoretical and empirical models proposed later and from the methodology used in this study for estimating exchange-rate misalignments.

[17] It is not by chance that the estimation of exchange rate misalignments using conventional models is expressed as the difference between the left-hand side and the first component on the right-hand side of the equation (3).

III. The determination of the long-term real exchange rate and its deviation from the "optimum" rate for development: a structuralist-Keynesian model

This section will focus on a structuralist-Keynesian model in which the path of the long-term real exchange rate and its divergence from the observed real exchange rate relative to its neutral equilibrium level and its "optimum" level (as defined earlier) are accounted for, at one and the same time, by long-term variables associated with the structure of the economy and by variables that are directly or indirectly linked to short-term macroeconomic policy (or, in other words, variables falling within the scope of action of policymakers who are either trying to deal with the ups and downs associated with short-term business cycles or to keep inflation under control). In fact, our structuralist-Keynesian model assumes that short-term macroeconomic policies (especially monetary and exchange-rate policies) have a strong and long-lasting influence on the misalignment of the observed real exchange rate relative to the "optimum" real exchange rate.

In our theoretical model, both the path of the long-run real exchange rate and the divergence of the observed real exchange rate from its "optimum" level can be represented by the following equation:

$$RER_t = g_t\left(struct_{lp_t}\right) + m_t\left(cp_t\right) \tag{4}$$

All the (structural and short-term) variables on the right side of the equation determine the path of the long-run real exchange rate as well as the degree of the misalignment of the observed real exchange rate with the "optimum" rate.

Where g is formed by a set of structural variables ($struct_{lpt}$) that influence the long-term real exchange rate (especially per capita income (Y) —which, as noted earlier, captures the Harrod-Balassa-Samuelson effect—, the terms of trade (ToT) and current account balance (CC)), while m refers to a set of variables that are directly or indirectly influenced by short-term macroeconomic policy (cp_t).[18] In the proposed econometric model, which will be set out in detail in the following section, the most important short-term variables are the spread between the domestic i and external i^* short-term nominal interest rates ($IDIFFER$), the stock of international reserves (RI) and the country risk premium (CR).

Two distinctive aspects set the proposed model apart from conventional models for determining the long-term real exchange rate. The first is that, although the variables represented by g are similar to those that influence the so-called "fundamental" variables of conventional models, the proposed model rejects the hypothesis that these variables always cause the real exchange rate to converge towards its long-term equilibrium level. This means that market forces may or may not steer the real exchange rate towards its "optimum" long-term level, but, if this does occur, it will be a random effect or be attributable to sudden, sharp adjustments in the exchange rate during crisis periods (as happened, for example, in Brazil after the speculative attack of 1999 and, more recently, as appears to have happened in 2015, as indicated by the empirical results discussed in section IV). If market forces are not always efficient in promoting exchange-rate adjustments or averting substantial misalignments (especially in

[18] In the econometric approach described in section IV, the term that represents the residual of the regression is added to the right side of the econometric equation that represents the theoretical equation (4) of the proposed model. This means that this term, together with the variables representing the structural and short-term policy component, will also influence the path of the long-run real exchange rate and the deviation of the observed real exchange rate from its "optimum" level.

the case of overvaluations), then the main policy implication would be that policymakers should work not only to avoid volatility but also to prevent the real exchange rate from deviating from its "optimum" level or appreciating in real terms. In other words, instead of a clean or dirty float, the most suitable exchange-rate regime would be a managed float, which —as shown by Aizenman, Chinn and Ito (2010)— is what is used in most of the developing countries of Asia.

The second aspect that distinguishes the two types of models from one another is that, whereas, in conventional models, the variables associated with short-term shocks are the only ones that are responsible for real exchange rate misalignments relative to the rate's long-term equilibrium level, in the proposed structuralist-Keynesian model, all the variables on the right side of the equation (4) may together account for an exchange rate misalignment, whether relative to the rate's equilibrium level or the "optimum" long-term real exchange rate, as is noted in the text table appearing below that equation.

This means that the proposed econometric procedure for estimating the deviation of the "optimum" real exchange rate from the observed real exchange rate will be different from the methodology used in conventional models. While the latter estimates the deviation by measuring the differential between the estimated observed real exchange rates and long-term exchange rates that are compatible with the "fundamental" equilibrium (and, therefore, the deviation would be influenced by the coefficients for the variables associated with short-term shocks plus the error term of the regression), the proposed procedure for calculating the deviation involves, first of all, using the Hodrick-Prescott filter technique to estimate the long-term trend of the series. The next step is to take the trend series and choose a reference period (an analysis of the selection criteria is presented in section IV) for the calculation of the observed real exchange rate's deviation from the estimated "optimum" real rate.

In short, the proposed model is structuralist because it accepts the hypothesis that the path of the long-term real exchange rate is influenced by structural variables, especially by the productivity growth of tradables relative to non-tradables (the Harrod-Balassa-Samuelson effect) and by the terms of trade. It is also a Keynesian model because, based on Keynes' argument (1936, especially chap. 12) that long-term trends are nothing more than the sum of events occurring in a succession of short periods, it rejects the conventional hypothesis that distinguishes short- and long-term real equilibrium exchange rates.[19] From an impressionistic perspective, then, we could say that both the level and the path of the short- and long-term real exchange rates are simply two sides of the same coin.[20]

Keynesian theory emphasizes that real exchange rates are heavily influenced by net short-term capital flows (Kaltenbrunner, 2008 and 2010), especially in developing economies that are very open to global financial activity. It is no coincidence that, even though he was not living in an era of financial globalization, Keynes (1923) witnessed the intensive movement of short-term capital flows that occurred in the wake of the First World War and the demise of the gold standard —triggered by defensive moves to raise short-term interest rates in the countries of the "centre", especially England— and realized that these financial flows in the international arena were the primary mechanism for the transmission of inter-country short-term interest rate differentials to the real exchange rate.

[19] See also Hahn (1984).

[20] Therefore, although structural variables such as trends in productivity and the terms of trade continue to be regarded as long-term variables, they are also heavily influenced by short-term economic policies. Thus, here again, we can see that long-term variables are nothing more than the sum of a succession of events occurring within a short-term time frame (which are, in their turn, influenced by short-term economic policies).

IV. An econometric estimate of the path and "optimum" level of the long-run real exchange rate

The econometric specification of the model for determining the path of the long-run real exchange rate can be expressed as follows:

$$\begin{gathered} ln\, RER_t = c_0 + \alpha_1\, ln\, Y_t + \alpha 2\, ln\, ToT_t + \alpha 3\, ln\, CC_t + \\ + \beta_1 \left(ln\, IDIFFER\right)_t + \beta_2 \left(ln\, IDIFFER\right)_{t-2} + \beta_3\, ln\, RI_t + \beta_4\, ln\, CR_t + \varepsilon_t \end{gathered} \tag{5}$$

Where (all the variables of the model are expressed in logarithms): RER is the observed real exchange rate, Y is real per capita GDP in dollars, ToT is the terms of trade, CC is the current account balance expressed as a percentage of GDP,[21] $IDIFFER$ is the spread between the domestic short-term interest rate (360-day swap rates) and the international short-term interest rate (the United States' target federal fund rate, which is taken as a proxy for the external short-term interest rate), $IDIFFER_{t-2}$ is the preceding variable converted into a two-period lagged variable,[22] RI is Brazil's stock of international reserves expressed as a percentage of GDP, CR is Brazil's country risk, as measured by its rating on the J.P. Morgan Emerging Market Bond Index (EMBI),[23] ε is the error term of the regression and the t subscripts are the time reference (in the proposed econometric model, the unit used is one month). While the first three variables on the right side of equation (5) are structural variables (i.e. per capita income (Y), the terms of trade (ToT) and the balance on current account (CC)), the rest of the variables are associated with short-term economic policy. The sources for the database, which covers the period from January 1999 to July 2015, are given in annex A1.

The selection of variables for inclusion in the model is not arbitrary but is instead the one used in a large number of models for determining the real exchange rate (see, for example, Helmers, 1988; Edwards, 1988; Calvo, Leiderman and Reinhart, 1993; Rodrik, 2008; Berg and Miao, 2010). Even so, the fact remains that both the structural and the short-term economic policy variables are suited to the proposed theoretical model. From a long-term perspective, trends in per capita income, the terms of trade and the balance on current account are the most important structural variables for a country such as Brazil, whose production and export structures rely heavily on natural-resource-intensive goods and whose development strategy over the past few decades has been based on attracting "external savings" (deficits on current account).[24] From a short-term perspective, the Keynesian literature (see Harvey, 2006, and Kaltenbrunner, 2008) suggests that the policy variables that we have chosen (the

[21] According to Bogdanski, Tombini and Werlang (2000), when the balance on current account (CC) has a negative sign, a positive number should be added in order to apply the logarithmic form. In those cases, the following procedure was adopted: $CC = 1 + CC$.

[22] The incorporation of the one-period lagged interest rate differential in the proposed econometric equation is justified because, given the external interest rate and assuming that everything else remains equal, an increase in the domestic short-term interest rate will tend to spur the net inflow of short-term capital, which would then (albeit with some time lag) cause the domestic currency to appreciate in both nominal and real terms.

[23] The choice of the EMBI as the most suitable measurement for evaluating Brazil's country risk is backed up by the Central Bank of Brazil (2015), which states that the most commonly used daily indicators in the market for that purpose (measuring foreign investors' credit risk in Brazil) are the EMBI+Br and the Credit Default Swap (CDS) of Brazil.

[24] Although disaggregated data for the most recent period are not yet available, Nassif (2008, table 1, p. 87) has calculated the share of Brazil's total manufacturing output that was accounted for by the value added by natural-resource-intensive manufacturers as of 2004 at 40.1%, as compared to a share of 32.7% in 1996. Bresser-Pereira, Nassif and Feijó (2016, table 2, p. 26) have calculated that exports of primary products and natural-resource-intensive manufactures (commodities) had risen to 62.1% by 2014 from 40.3% in 2000. Regarding Brazil's emphasis on financing its development with net inflows of external savings, see Bresser-Pereira and Nakano (2003) and Bresser-Pereira, Nassif and Feijó (2016).

interest rate spread, the stock of international reserves and country risk) are the best ones for capturing the direct and indirect effects on the real exchange rate in an economy that is extremely open to external capital movements, as is the case of Brazil during the period under study.

The model being proposed here differs from conventional models in its theoretical and empirical approach to the determination of the long-run real exchange rate and its degree of misalignment (in this case, its misalignment relative to the "optimum" level for purposes of economic development), as noted in earlier sections. Before moving on to the statistical tests and the application of the econometric model, it is worth analysing the expected signs of the estimated coefficients for the model's variables, which are outlined in table 1.

Table 1
Structuralist-Keynesian model for the determination of the long-run real exchange rate: expected signs of the estimated coefficients for the variables included in the model

Variables	Expected sign of the estimated coefficient
Per capita income (Y)	-
Terms of trade (ToT)	Either (+ or -)
Balance on current account (CC)	+
Domestic/external interest rate spread ($IDIFER$)	Either (+ or -, respectively, at very short, short and medium terms)
Stock of international reserves(RI)	Either (+ or -)
Country risk premium (CR)	+

Source: Prepared by the authors.

The expected signs of the estimated coefficients for per capita income (Y) and the terms of trade (ToT) were discussed in section II. The expected sign of the coefficients for the balance on current account (CC) is positive because a long-term current account surplus is associated with a depreciated currency in real terms, as discussed by the theoretical literature (Obstfeld and Rogoff, 1996). The expected sign of the coefficients for the interest rate differential could go either way: while, in the very short term, given the level of the external interest rate, an increase in the short-term domestic interest rate could lead to a depreciation of the currency (positive sign) via expectations ("fear of a float" or, in this case, "fear of a depreciation", as argued by Calvo and Reinhart, 2002), on the other hand, in the short and medium terms, an increase in the domestic interest rate (given the level of the external interest rate) would tend, *ceteris paribus*, to spur net capital inflows and, consequently, an appreciation of the currency in real terms (negative sign). The sign of the expected coefficient for the variable associated with the stock of international reserves could also go either way: an increase in external reserves over time means that the government is buying up international currency on the spot market in order to slow down a possible appreciation or even to trigger a sharper depreciation (positive sign). However, a large stock of international reserves may, *ceteris paribus*, lower the country risk premium, thereby stimulating net capital inflows and, consequently, lead to an appreciation of the domestic currency in real terms (negative sign). Clearly, the expected sign of the country risk premium is positive, since an increase in that premium may spark capital flight, leading to a depreciation of the domestic currency in real terms (positive sign).

The first step in the empirical analysis was to run the augmented Dickey-Fuller test and the Phillips-Perron test. The results indicate that all the series are integrated of order 1, or, in other words, non-stationary in level but stationary in first difference.

In addition to their stationarity, it is also important to consider the possible endogeneity of the model's variables because, since the model violates the assumption of the model of ordinary least squares (OLS) that the residual should not correlate with the explanatory variables of the regression, it may produce skewed, inconsistent and inefficient OLS estimators and thus compromise the inferential analysis.

However, as shown by Baffes, Elbadawi and O'Connell (1999, chap. 10), even the most persuasive exogeneity tests, such as the one proposed by Engle, Hendry and Richard (1983), cannot always resolve endogeneity problems when there are changes in the marginal distribution of explanatory variables. Even so, Johansen's cointegration test (1988) provides a very powerful means of dealing with the problem of endogeneity in models with more than one endogenous variable, not only because it considers all the variables in an estimation process to be endogenous, but also because it simultaneously determines the equilibrium relationship among them.

Since the variables are not stationary and have the same order of integration, the cointegration test proposed by Johansen (1988) can be used to see if a stable long-term relationship exists among them. As the test showed that there was a cointegration vector, the existence of a stable long-term relationship among the model's variables can be accepted.[25]

Once it is known that the series are non-stationary and cointegrated, equation (5) can be estimated using ordinary least squares (OLS) and the error correction model (ECM).[26] Table 2 gives the results of the econometric model. All the estimated coefficients were statistically significant and displayed the expected signs shown in table 1. It is important to note that, while the interest rate differential was included in the econometric equation (5) for economic reasons (see footnote 23), some of the model's variables were lagged one or two periods (months) simply for econometric reasons. Since the selected variables are monthly, it can be assumed that the explanatory structural and economic policy variables will not alter the real exchange rate in such a short span of time (scarcely a month), so it is reasonable to expect the model to fit the data better when the data are incorporated with some time lag.

Table 2
Brazil: estimated long-term real exchange rate models, 1999-2015
(Dependent variable: real exchange rate (RER))

Variable	Variable	OLS coefficient	ECM coefficient	
		(t statistics in brackets)	Variable	(t statistics in brackets)
C	Constant	6.650088***	C	5.9805***
		[10.41783]		
$\ln Y$-2	Log of per capita GDP	-0.33637***	$\ln Y$-3	-0.763422***
		[-7.61376]		[-7.93942]
$\ln TOT$	Log of the terms of trade	-0.26492**	$\ln TOT$-1	-0.454013*
		[-1.91535]		[-1.69178]
$\ln CC$-1	Log of the balance on current account/GDP	0.068764***	$\ln CC$-1	0.085584***
		[4.538101]		[2.34562]
$Ln(IDIFFER)$	Log of short-term interest rate differential	[2.320963]	$Ln(IDIFER)$	-
		0.296203**		-
$Ln(IDIFFER)$-2	Log of the lagged short-term interest rate differential	-0.24448**	$Ln(IDIFER)$-2	-0.26921**
		[-2.0114]		[-4.41106]
$\ln RI$-1	Log of the stock of international reserves/GDP	0.223979***	$\ln RI$-1	0.167482**
		[6.6185]		[2.37291]
$\ln CR$	Log of Brazil's risk premium	0.039893*	$\ln CR$-1	0.372263***
		[1.70786]		[5.96244]

Source: Prepared by the authors.
Notes: OLS model: R2: 0.839; adjusted R2: 0.833; Durbin-Watson statistic: 1.833; F statistic: 141.169; Prob (F test): 0.000; number of observations: 197 after adjustments. The $IDIFFER$ variable was included with two lags; the CC and RI variables were included with a lag and the Y variable with two lags. The ECM: three lags; number of observations: 193 after adjustments. The TOT, CC, RI and CR variables were included with a lag, $IDIFFER$ with two lags and Y with three. ***significant at 1%; **significant at 5%; *significant at 10%.

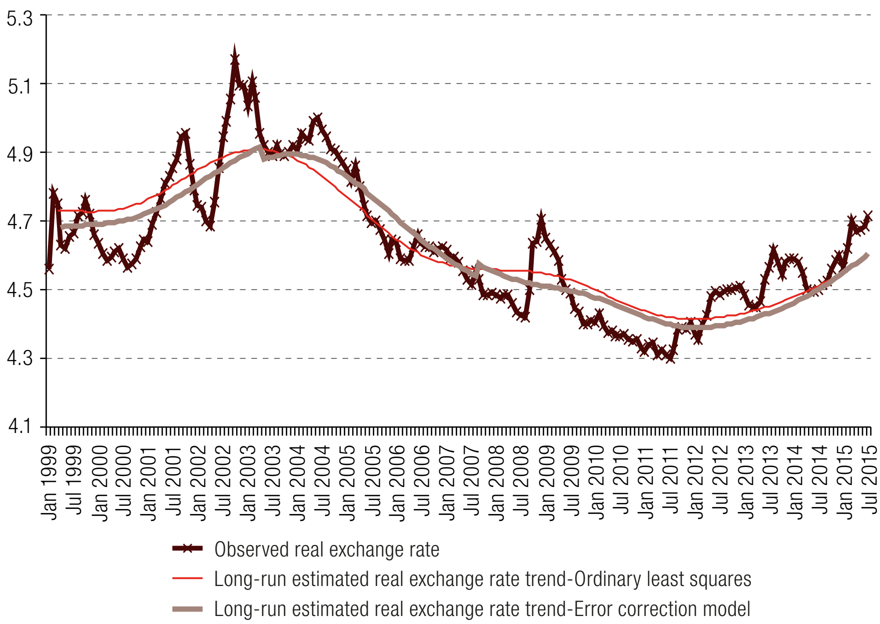
[25] The results of all the tests described in this section can be requested from the authors via email.

[26] As noted by Hamilton (1994), if the series exhibit these characteristics, then the OLS method will continue to be a highly consistent estimator. For a formal demonstration of this, see Hamilton (1994, p. 587).

The results show that per capita income, the terms of trade and short-term interest rate differential were the variables whose estimated coefficients provided the best explanations for the path of the real exchange rate in Brazil between 1999 and 2015. In other words, the proposed econometric model suggests that the path of the real exchange rate in Brazil between January 1999 and July 2015 —which reflected an appreciation trend throughout most of that period, as shown in figure 1— was influenced both by structural economic variables in Brazil (such as per capita income trends, which capture the Harrod-Balassa-Samuelson effect, and the terms of trade, which were very favourable for Brazil during much of the period under analysis) and by variables that are directly linked to short-term macroeconomic policies. [27] In fact, the wide interest rate differential observed in Brazil in recent decades led to excessive net short-term capital inflows during economic booms and high international liquidity, which spurred an appreciation of the Brazilian currency in real terms.

The coefficients of the econometric model were used to estimate the long-run trend of the real exchange rate ($R\hat{E}R$). This result was then compared with the observed real exchange rate as a basis for constructing an index that could be used to determine if the latter is overvalued, undervalued or in equilibrium relative to the estimated "optimum" level. In line with the suggestions of Edwards (1989) and Alberola (2003), this study has used the Hodrick-Prescott filter technique to estimate the $R\hat{E}R$. Figure 1 traces the observed real exchange rate (the RER that is regularly calculated by the Central Bank of Brazil, whose database is listed in annex A1) and the path of long-run real exchange rates estimated using the OLS and ECM models. Both the observed and estimated real exchange rates are expressed in logarithms in figure 1.

Figure 1
Brazil: observed real exchange rates and estimated long-term real exchange rates, January 1999 to July 2015
(In logarithms)



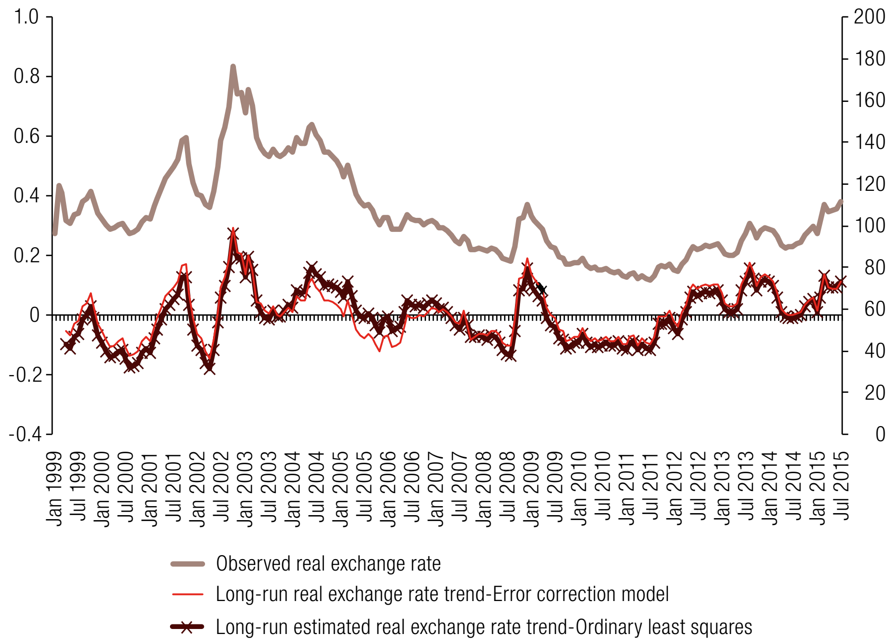
Source: Prepared by the authors.

[27] These two structural variables —per capita income and the terms of trade— were partially responsible for the long-term appreciation of the Brazilian currency in real terms, as predicted by economic theory.

The robust results of these estimates can be corroborated not only by the level of significance of the estimated coefficients (detailed in table 2), but also by the good fit between the paths of the estimated exchange rates with the two (OLS and ECM) estimation models. There is also a close correlation between the paths of the estimated real exchange rates obtained by both models and actual real exchange rates. Both models indicate that the real exchange rate in Brazil started to appreciate in late 2003 and early 2004, even though the observed real exchange rate in 2004 was considerably undervalued, as is depicted in figure 2.

Figure 2 compares the paths of the observed real exchange rates and estimated long-term real exchange rates generated by the OLS and ECM models.

Figure 2
Brazil: observed real exchange rates and degrees of undervaluation or overvaluation relative to estimated real exchange rates, January 1999 to July 2015



Source: Prepared by the authors based on the methodology described in this paper for the calculation of estimated real exchange rates and based on data from the Central Bank of Brazil for the observed real exchange rates (see annex A1).

Note: (i) The observed real exchange rates (shown by the highest line in the graph) correspond to the indices shown on the right-hand axis (the mean for 2000=100): a value of over 100 indicates a real undervaluation relative to the base year, while a value below 100 indicates an overvaluation relative to the same base year (2000).
(ii) The percentage degrees of under- and overvaluation (depicted by the lower lines on the graph) were calculated as the difference between the observed real exchange rate (RER) and the long-run trends of the real exchange rates estimated using the two models (OLS and ECM). If the result is greater than 0, then the Brazilian real is estimated to be undervalued, while if the result is less than 0, the Brazilian real is estimated to be overvalued. These results, expressed in percentages, are indicated on the left-hand side of the graph.

For the long-run estimated real exchange rates using the two models (represented by the two overlapping lines in the lower portion of figure 2), values below 0.00 indicate an overvaluation (expressed as a percentage) of the observed real exchange rates relative to the estimated real exchange rates, while values above 0.00 indicate an undervaluation. In the case of the observed real exchange rate (RER – represented by the dotted line in the upper portion of figure 2), indices below 100 indicate an overvaluation relative to the mean for 2000 (which, by hypothesis, is equal to 100 and is the year when the observed real exchange rate is said to have been at equilibrium, according to the Central Bank of Brazil), while indices over 100 denote an undervaluation relative to the mean for the year 2000.

Having estimated the path of the real exchange rate using the two models, the next step is to establish a methodology for identifying the period in which the real exchange rate has reached its "optimum" level for economic development. Once that period has been established, it becomes possible to determine whether the average nominal exchange rate during the first half of January 2016 (when this study was being finalized) was overvalued, undervalued or at its "optimum" level.

The criteria to be used in identifying the period during which the real exchange rate reached its "optimum" level must fulfill three simultaneous conditions: (i) in keeping with the empirical studies mentioned above, which indicate that a developing country's currency should be slightly undervalued (i.e. it should exhibit a small real devaluation against the United States dollar or a basket of the currencies of the country's main trading partners), the selected period should be one in which the estimated (not the observed) exchange rate exhibited a small (but not large) undervaluation (just slightly above 0.00 in figure 2);[28] (ii) the selected period should correspond to a time when macroeconomic indicators were fairly solid and, in particular, when the balance on current account was at equilibrium or showed a surplus; and (iii) the selected period should be a time when the observed real exchange rate was not overvalued (the index in figure 2 should not be below 100).

These criteria rule out all the periods in which the estimated or observed real exchange rates were overvalued (January 2010 to January 2012, to cite just one example).[29] They also rule out periods in which the estimated real exchange rate was excessively undervalued, such as, for example, the period between April 2002 and April 2003 (since a large depreciation may have longer lasting inflationary effects).[30] The period running from April 2012 to April 2013 also has to be ruled out because, although the estimated exchange rate was slightly undervalued (by an average of 7% for the period, according to the two models) —and thereby meets the first condition— the observed real exchange rate was overvalued (an average real exchange rate index of 95.70) and the country's macroeconomic indicators were far from solid.[31]

A close look at figure 2 shows that the "optimum" real exchange rate was reached at some time between June 2003 and April 2005, which is the only period in which all three conditions are met. During that period, the Brazilian economy was in the final stage of a macroeconomic adjustment that had started in mid-1999 and had higher growth rates and a surplus on its current account.[32] In addition, the average real exchange rate, as estimated using the two models, was slightly undervalued (by 5.05%)

[28] Barbosa and others (2010) state that "a structuralist theoretical model and the evidence from Brazil (1996-2009) suggest that there exists an optimal exchange rate level that maximizes growth". In that study, the authors estimate the index for a real exchange rate that would maximize Brazil's long-term economic growth at around 101.6 (or, in other words, a real depreciation (at the margin) of about 1.6% relative to the "neutral" real exchange rate). A real exchange rate that represented a very large real depreciation could, of course, hinder long-term growth, either because it could set off a longer lasting inflationary process or because it could generate a much greater and undesirable distortion in the allocation of the economy's production resources.

[29] During this period, the estimated real exchange rates were overvalued by around 7.8% (the mean for the two models), the observed real exchange rate (RER) was 81.67 (which represents an overvaluation of the Brazilian real of nearly 18.3% relative to the year 2000), and the country's macroeconomic indicators were much less sound than they were in, for example, 2007. In fact, if we look at just two indicators for the year 2011, we see that, according to the Central Bank of Brazil, real GDP growth was 3.9% (compared to 6.1% in 2007) and the deficit on current account was nearly 2.1% of GDP (compared to a surplus of 0.1% in 2007).

[30] The proposed model indicates that there was an estimated average undervaluation of nearly 11% (the mean for the two models) between April 2002 and April 2003, which is considered to be too large.

[31] In 2013, according to the database of the Central Bank of Brazil, real GDP growth amounted to 3% and there was a current account deficit equivalent to 3.6% of GDP.

[32] In 2004, for example, the database of the Central Bank of Brazil shows that real GDP growth was 5.8% and there was a current account surplus equivalent to 1.8% of GDP.

and, finally, the observed average real exchange rate was not overvalued.[33] Thus, taking the period between June 2003 and April 2005 as a point of reference for the "optimum" real exchange rate, the estimated mean index for the long-run real exchange rate was 127.82 (OLS: 125.87 and ECM: 129.87). A comparison of this last estimated index with the index for the observed real exchange rate in July 2015 —the last month for which data for all the variables of the model are available— (111.81) indicates that, in this last month, the Brazilian real exhibited a real overvaluation of about 14.36% relative to its long-term "optimum" level. In other words, in July 2015, the average nominal exchange rate should have reached approximately 3.88 Brazilian reais per dollar (compared with the observed average nominal exchange rate, of 3.39 Brazilian reais per dollar) to preserve the "optimum" level achieved between June 2003 and April 2005.

Although the available data allowed our econometric estimation to be run only until July 2015, we adjusted the results for this latter period until December 2015, based on the relative real purchasing power parity hypothesis.[34] Using data on the consumer price indices for Brazil and the United States (the extended consumer price index (IPCA) and the consumer price index (CPI), respectively) up to December 2015 and the cumulative inflation rate differential for Brazil and the United States between August and December 2015 (IPCA: 3.6%; CPI: -0.1%), it can be concluded that the "optimum" real exchange rate in December 2015 should have been around 4.02 reais per dollar.[35] This result is quite close to the average nominal exchange rate for that month (3.90 reais per dollar) and is exactly the same as the mean for the first half of January 2016 (4.02 reais per dollar), as is shown on the website of the Central Bank of Brazil. In sum, after a long period beginning in late 2005, in which the Brazilian real was appreciating considerably (and which was interrupted only for about six months immediately following the outbreak of the global financial crisis in September 2008), the real exchange rate in Brazil reached its "optimum" level in mid-January 2016.

We understand that the same adjustment procedure based on the relative purchasing power parity hypothesis could be applied in order to determine if the Brazilian real will be overvalued, undervalued or equivalent to its "optimum" level in a relatively short period of time (say, in the next two years, until the end of 2017) (i.e. using the cumulative differential between inflation in Brazil and international inflation). For longer periods, given that the long-term trajectory of the real exchange rate is strongly affected by structural variables and by short-term economic policies, the models proposed here for determining the "optimum" real exchange rate (or similar models, such as that proposed by Marconi, 2012) can be used to arrive at new estimates.

[33] The average index for the observed real exchange rate for this period was 135.52, which means that the Brazilian real was undervalued by 35.5% relative to the average for the year 2000. It might be argued that this period could not be considered as one in which the real exchange rate had reached its "optimum" level because the Brazilian currency was excessively undervalued. However, this argument is flawed, for two reasons: first, because the rate in question here is the observed real exchange rate, whose indices are based only on the differences between the domestic and external inflation rates (and, in the case of the real effective exchange rate, are weighted for the relative importance of each of Brazil's trading partners); and, second (and perhaps most importantly), as may be seen from figure 2, the Brazilian currency was shedding the effects of the excessive undervaluation seen in October 2002, when there was a clear case of overshooting in the real-dollar exchange rate. As noted by Barbosa-Filho (2015, p. 405), 2003-2005 was a period of "exchange-rate correction" because that was the time during which the appreciation of the Brazilian real was undoing the sharp depreciation of the currency that had occurred in the preceding years.

[34] As observed earlier, according to the hypothesis of relative purchasing power parity, in order to maintain a currency's real purchasing power, the nominal exchange rate should correct for the difference between the cumulative domestic and external inflation rates.

[35] The data for Brazil are taken from the Brazilian Geographical and Statistical Institute (IBGE) and the data for the United States come from that country's Bureau of Labor Statistics. For the IBGE data, see [online] http://www.ibge.gov.br/home/estatistica/indicadores/precos/inpc_ipca/defaultinpc.shtm. For the data for the United States, see [online] http://www.bls.gov/cpi/. [Consulted on 20 January 2015].

V. Conclusion

In the recent overvaluation cycle of the Brazilian real, which occurred between mid-2005 and the end of 2014 —and was interrupted only temporarily in the six months following the September 2008 global financial crisis— the Brazil's then Minister of Finance, Guido Mantega, attributed this trend to external factors. In particular, he attributed it to quantitative easing in the United States, which spurred a huge expansion of dollar-denominated liquidity in world markets, thereby forcing —in his opinion— emerging countries' currencies (and especially the Brazilian real) to appreciate. He also accused the Federal Reserve Bank of the United States of waging a "currency war".

In a paper prepared for the renowned Mundell-Fleming Lecture held each year by the International Monetary Fund (IMF), Ben Bernanke (2015), the former Chairman of the United State Federal Reserve, showed that Mantega's accusation was theoretically and empirically baseless. According to Bernanke (2015, pp. 3-4), expressions like "currency wars" in have no sense in this context because "concerns about currency wars on the part of emerging-market policymakers appear to be motivated in large part by those 4 policymakers having separate goals for their own exchange rates, over and above assuring the stability of domestic output and incomes. To the extent that they have additional exchange-rate objectives, foreign policymakers are constrained primarily by the Mundell-Fleming "trilemma" —the impossibility of combining free capital flows, independent monetary policy, and exchange rate targets— not by US policy *per se*.

However, Bernanke (2015, p. 4) went on to argue that "…monetary and exchange-rate policies should focus on macroeconomic objectives, with the problem of spillovers [from United States monetary policy to international capital flows] being tackled by regulatory and macroprudential measures, possibly including targeted capital controls, and through careful sequencing of market reforms". It is no coincidence, as recently pointed out by Hey (2015, p. 1) in a paper that has had a resounding impact in academic circles: "The global financial cycle transforms the 'trilemma' into a 'dilemma' or an 'irreconcilable duo': independent monetary policies are possible if and only if the capital account is managed".

This criticism of Bernanke is relevant here because, as has already been shown by Aizenman, Chinn and Ito (2010), most of the Asian countries have sought to avert lengthy phases of chronic real appreciation of their currencies by implementing measures designed to overcome the constraints associated with the "trilemma". According to those authors, since the disastrous consequences of the 1997 Asian crisis (whose origin can be found, to some extent, in the preceding period, when large net inflows of external capital caused their currencies to become enormously overvalued), policymakers in most of the countries of that region have sought to maintain monetary and exchange-rate stability and to avoid volatility and, above all, a cyclical and chronic appreciation of their currencies. To accomplish this, they make use of an array of the tools that are at their command, including interventions in the spot and futures markets, regulatory and macroprudential measures and ad hoc capital controls.[36] And in point of fact, the use of capital controls, which until not long ago was regarded as heresy by multilateral lending institutions, has recently been advocated by the International Monetary Fund (IMF) in official documents (see the studies of Ostry and others, 2011, and Ostry, Ghosh and Chamon, 2012). Our suggestion is that Brazilian policymakers should follow the Asian countries' example and make use —prudently and efficiently— of an array of exchange-rate policy instruments to avoid another long-lasting and chronic real appreciation of the Brazilian currency.

At the start, this paper emphasized that a slightly undervalued currency works as a powerful driver for promoting structural change, economic development and catching up in the long-term. It is important to note, however, that, even if the country manages to maintain the "optimum" real exchange

[36] See Aizenman, Chinn and Ito (2010) for a discussion of the Asian case and Subbarao (2014) for an examination of the case of India.

rate reached in January 2016 over the coming years, it will be no easy task to bridge the enormous technological lag accumulated by Brazilian industry in the past decade,[37] which is a condition *sine qua non* for reaching and maintaining higher levels of productivity, not only in this sector, but in the economy as a whole. This conclusion is backed up by robust empirical evidence that the manufacturing sector is the main driver of productivity growth in industry and in the wider economy, as first pointed out by Kaldor (1966) based on the empirical regularity originally advanced by the Dutch economist Verdoorn (1949).[38]

However, in view of the marked retreat of industry and the reprimarization of Brazil's export profile in recent decades,[39] the country's technological development —given the heavily path-dependent and locked-in nature of that process— may be hamstrung for a long time to come unless industrial and technological policies, in coordination with macroeconomic policy, can overcome the hysteresis[40] resulting from decades of high real interest rates (high cost of capital) and cyclical, chronic real appreciations of Brazil's currency. As Baldwin and Krugman (1989, p. 653) contended in their classic paper, "if it is wrong to ignore feedbacks from trade to exchange rate, it is probably also wrong to ignore feedback to the costs of capital". They also argued that "a temporary overvaluation is followed by a persistent reduction in the equilibrium exchange rate but not enough to regain lost markets" (ibid, p. 637).

As shown by Krugman (1991, p. 652) in another seminal work, in the presence of static and dynamic increasing returns to scale, an economy that has fallen far behind in relation to the international technological frontier will likely face a long-term recovery with multiple equilibrium points. Accordingly, "the choice among multiple equilibria is essentially resolved by history: past events set the preconditions that drive the economy to one or another [positive or negative] steady state".

For this reason, although it is unlikely that bringing the real exchange rate to its "optimum" level will, in and of itself, immediately revert the current regressive trend in the Brazilian economy and set it on a new path of economic convergence, the fact remains that the exchange rate is a strategic price, and its adjustment will therefore play a pivotal role in attaining that objective.[41] This is precisely why it is so important for Brazilian policymakers to regard the real exchange rate as a strategic price of crucial importance for economic development rather than simply as an eternal anchor for inflation.

[37] According to Nassif, Feijó and Araújo (2015), Brazil's technological gap, measured as the ratio between the country's labour productivity and that of the United States (a proxy for the international technological frontier), widened from 70% to 80% between 1980 and 2000, and then remained at that level until 2013.

[38] For recent evidence on this topic (known as the Kaldor-Verdoorn law) in Latin America, see Ros (2014).

[39] The loss of the share of value added by the Brazilian manufacturing sector to GDP (in real terms) has been dramatic in the last few decades, since it shrank from 21.6% in 1980 to 18.1% in 1990, 15.1% in 2000, 13.9% in 2010 and to 11.7% in the first half of 2015. The reprimarization of the country's export pattern, as noted earlier, is reflected in the expansion of the share of agricultural goods and natural-resource-intensive industrial products (commodities) in total exports, which rose from 40.3% to 62.5% between 2000 and 2014. See Bresser-Pereira, Nassif and Feijó (2016).

[40] The concept of hysteresis —a concept borrowed from physics and first incorporated into economic theory by Blanchard and Summers (1986)— refers to a situation in which a given material (in this case, the competitiveness of Brazilian industry) will not easily regain its original characteristics (rapid productivity growth) even after the main force that caused a disturbance (in this case, the greatly overvalued exchange rate) has been eliminated.

[41] For empirical evidence that the Brazilian economy is lagging farther and farther behind, see Nassif, Feijó and Araújo (2015b).

Bibliography

Aizenman, J., M. D. Chinn and H. Ito (2010), "Surfing the waves of globalization: Asia and financial globalization in the context of the trilemma", *La Follette School Working Paper*, No. 2010-009, Madison, La Follette School of Public Affairs.

Alberola, E. (2003), "Misalignment, liabilities dollarization and exchange rate adjustment in Latin America", *Working Papers*, No. 309, Bank of Spain.

Amsden, A. H. (2001), *The Rise of "the Rest": Challenges to the West from Late-Industrializing Economies*, Oxford, Oxford University Press.

___(1989), *Asia's Next Giant: South Korea and Late Industrialization*, Oxford, Oxford University Press.

Araújo, E. (2009), "Política cambial e crescimento econômico: teorias e evidências para os países em desenvolvimento e emergentes", thesis, Rio de Janeiro, Institute of Economics, Rio de Janeiro's Federal University.

Baffes, J., I. Elbadawi and S. A. O'Connell (1999), "Single-equation estimation of the equilibrium real exchange rate", *Exchange Rate Misalignment: Concepts and Measurement for Developing Countries*, L. E. Hinkle and P. J. Montiel (eds.), Oxford University Press.

Baldwin, R. and P. R. Krugman (1989), "Persistent trade effects of large exchange rate shocks", *The Quarterly Journal of Economics*, vol. 104, No. 4, Oxford, Oxford University Press.

Barbosa-Filho, N. (2015), "O desafio macroeconômico de 2015-2018", *Brazilian Journal of Political Economy*, vol. 35, No. 3, São Paulo.

Barbosa-Filho, N. and others (2010), "Real Exchange Rate, Capital Accumulation and Growth in Brazil" [online] http://eesp.fgv.br/sites/eesp.fgv.br/files/file/4c0d022668198.pdf.

Berg, A. and Y. Miao (2010), "The real exchange rate and growth revisited: the Washington Consensus strikes back?", *IMF Working Paper*, No. 10/58, Washington, D.C., International Monetary Fund (IMF).

Bernanke, B. (2015), "Federal reserve policy in an international context", paper prepared for the 16th Jacques Polak Annual Research Conference, Washington, D.C., 5-6 November [online] https://www.imf.org/external/np/res/seminars/2015/arc/pdf/Bernanke.pdf.

Blanchard, O. and L. H. Summers (1986), "Hysteresis and the European unemployment problem", *NBER Working Paper*, No. 1950, Cambridge, Massachusetts, National Bureau of Economic Research.

Bogdanski, J., A. Tombini and S. R. C. Werlang (2000), "Implementing inflation targeting in Brazil", *Working Papers Series*, No. 1, Brasilia, Central Bank of Brazil [online] https://www.bcb.gov.br/pec/wps/ingl/wps01.pdf.

Bresser-Pereira, L. C. (2010), *Globalização e competição: por que alguns países emergentes têm sucesso e outros não*, Rio de Janeiro, Ed. Campus.

Bresser-Pereira, L. C. and Y. Nakano (2003), "Crescimento econômico com poupança externa?", *Brazilian Journal of Political Economy*, vol. 23, No. 2, São Paulo.

Bresser-Pereira, L. C., A. Nassif and C. Feijó (2016), "The Case for Reindustrializing the Brazilian Economy: Connecting the Macroeconomic Regime and the Industrial Policy", paper prepared for the 13th International Conference "*Developments in Economic Theory and Policy*", Bilbao, 23-24 June.

Bresser-Pereira, L. C., J. L. Oreiro and N. Marconi (2014), *Developmental Macroeconomics: New Developmentalism as a Growth Strategy*, London, Routledge.

Calvo, G. and C. Reinhart (2002), "Fear of floating", *Quarterly Journal of Economics*, vol. 117, No. 2, Oxford, Oxford University Press.

Calvo, G., L. Leiderman and C. Reinhart (1993), "Capital inflows and real exchange rate appreciation in Latin America. The role of external factors", *IMF Staff Papers*, vol. 40, No. 1, Washington, D.C., International Monetary Fund (IMF).

Cassel, G. (1918), "Abnormal deviations in international exchanges", *Economic Journal*, vol. 28, No. 112, Wiley.

Central Bank of Brazil (2015), "Risco-país: com informações até março de 2015", Brasilia [online] http://www4.bcb.gov.br/pec/gci/port/focus/faq%209-risco%20pa%C3%ADs.pdf.

Dollar, D. and A. Kraay (2003), "Institutions, trade and growth", *Journal of Monetary Economics*, vol. 50, No. 1, Amsterdam, Elsevier.

Edwards, S. (1989), *Real Exchange Rates, Devaluation, and Adjustment: Exchange Rate Policy in Developing Countries*, Cambridge, Massachusetts, The MIT Press.

___(1988), "Real and monetary determinants of real exchange rate behavior: theory and evidence from developing countries", *UCLA Working Paper*, No. 506, Los Angeles, University of California.

Engle, R. F., D. F. Hendry and J. F. Richard (1983), "Exogeneity", *Econometrica*, vol. 51, No. 2, New York, Econometric Society.

Gala, P. (2008), "Real exchange rate levels and economic development: theoretical analysis and econometric evidence", *Cambridge Journal of Economics*, vol. 32, No. 2, Oxford University Press.

Hahn, F. (1984),"Die Allgemeine Gleichgewichtstheorie", *Die Krise der Wirtschaftstheorie*, D. Bell and I. Bristol (eds.), Berlin, Springer Verlag.

Hamilton, J. D. (1994), *Time Series analysis*, Princeton, Princeton University Press.

Harvey, J. T. (2006), "Modeling interest rate parity: a system dynamics approach", *Journal of Economic Issues*, vol. 40, No. 2, Taylor and Francis.

Helmers, F. L. C. H. (1988), "The real exchange rate", *The Open Economy: Tools for Policymakers in Developing Countries*, R. Dornbusch and F. L. C. H. Helmers (eds.), Oxford, Oxford University Press.

Hey, H. (2015), "Dilemma not trilemma: the global financial cycle and monetary policy independence", *NBER Working Paper*, No. 21162, Cambridge, Massachusetts, National Bureau of Economic Research.

Hirschman, A. (1958), *The Strategy of Economic Development*, New Haven, Yale University Press.

Johansen, S. (1988), "Statistical analysis of cointegrating vectors", *Journal of Economic Dynamics and Control*, vol. 12, No. 2-3, Amsterdam, Elsevier.

Kaldor, N. (1978), "The effect of devaluations on trade in manufactures", *Further Essays in Applied Economics*, London, Duckworth.

___(1966), *Causes of the Slow Rate of Economic Growth of the United Kingdom. An Inaugural Lecture*, Cambridge, Cambridge University Press.

Kaltenbrunner, A. (2010), "International financialization and depreciation: the Brazilian real in the international financial crisis", *Competition and Change*, vol. 14, No. 3-4, Sage.

___(2008), "A Post-Keynesian Look at the Exchange Rate Determination in Emerging Markets and its Policy Implications: the Case of Brazil", paper presented at the 12th Conference "Macroeconomic Policies on Shaky Foundations – Whither Mainstream Economics?", Berlin, 31 October-1 November.

Keynes, J. M. (1936), *The General Theory of Employment, Interest and Money*, London, Macmillan.

___(1923), *A Tract on Monetary Reform*, London, Macmillan.

Krugman, P. (1991), "History versus expectations", *The Quarterly Journal of Economics*, vol. 106, No. 2, Oxford, Oxford University Press.

Krugman, P. and L. Taylor (1978), "Contractionary effects of devaluation", *Journal of International Economics*, vol. 8, No. 3, Amsterdam, Elsevier.

Lewis, W. A. (1954), "Economic development with unlimited supplies of labor", *The Manchester School*, vol. 22, No. 2, Wiley.

Lucas Jr., R. E. (1988), "On the mechanics of economic development", *Journal of Monetary Economics*, vol. 22, No. 1, Amsterdam, Elsevier.

Marconi, N. (2012), "The industrial equilibrium exchange rate in Brazil: an estimation", *Brazilian Journal of Political Economy*, vol. 32, No. 4, São Paulo.

Myrdal, G. (1957), *Economic Theory and Underdeveloped Regions*, London, Duckworth.

Nassif, A. (2008), "Há evidências de desindustrialização no Brasil?", *Brazilian Journal of Political Economy*, vol. 28, No. 1, São Paulo.

Nassif, A., C. Feijó and E. Araújo (2015a), "The BRICS's long term economic performance" [online] https://www.boeckler.de/pdf/v_2015_10_24_nassif.pdf.

___(2015b), "Structural change and economic development: is Brazil catching up or falling behind?", *Cambridge Journal of Economics*, vol. 39, No. 5, Oxford University Press.

___(2011), "The long-term 'optimal' real exchange rate and the currency overvaluation trend in open emerging economies: the case of Brazil", *UNCTAD Discussion Papers*, No. 206, Geneva, United Nations Conference on Trade and Developement (UNCTAD).

Obstfeld, M. and K. Rogoff (1996), *Foundations of International Macroeconomics*, Cambridge, Massachusetts, The MIT Press.

Ostry, J. D., A. R. Ghosh and M. Chamon (2012), "Two targets, two instruments: monetary and exchange rate policies in emerging market economies", *Staff Discussion Notes*, No. 12/01, Washington, D.C., International Monetary Fund (IMF).

Ostry, J. and others (2011), "Managing capital inflows: what tools to use?", *Staff Discussion Notes*, No. SDN/11/06, Washington, D.C., International Monetary Fund (IMF).

Prasad, E., R. Rajan and A. Subramaniam (2006), *Foreign Capital and Economic Growth*, Washington, D.C., International Monetary Fund (IMF).

Prebisch, R. (1950), *The Economic Development of Latin America and its Principal Problems* (E/CN.12/89/Rev.1), Santiago, United Nations.

Razin, O. and S. M. Collins (1999), "Real exchange rate misalignments and growth", *The Economics of Globalization: Policy Perspectives from Public Economics*, A. Razin and E. Sadka (eds.), Cambridge, Cambridge University Press.

Rodrik, D. (2008), "The real exchange rate and economic growth", *Brookings Papers on Economic Activity*, vol. 39, No. 2, Washington, D.C., The Brookings Institution.

Rogoff, K. (1996), "The purchasing power parity puzzle", *Journal of Economic Literature*, vol. 34, No. 2, Nashville, Tennessee, American Economic Association.

Romer, P. M. (1986), "Increasing returns and long-run growth", *Journal of Political Economy*, vol. 94, No. 5, Chicago, The University of Chicago Press.

Ros, J. (2014), "Productividad and crecimiento en América Latina: ¿por qué la productividad crece más en unas economías que en otras?", *Desarrollo Económico* (LC/MEX/L.1145), Mexico City, ECLAC subregional headquarters in Mexico.

Rosenstein-Rodan, P. N. (1943), "Problems of industrialization of Eastern and South-Eastern Europe", *Economic Journal*, vol. 53.

Sarno, L. and M. P. Taylor (2002), "Purchasing power parity and the real exchange rate", *IMF Staff Papers*, vol. 49, No. 1, Washington, D.C., International Monetary Fund (IMF).

Simonsen, M. H. and R. P. Cysne (1995), *Macroeconomia*, São Paulo, Editora Atlas.

Solow, R. M. (1956), "A contribution to the theory of economic growth", *Quarterly Journal of Economics*, vol. 70, No. 1, Cambridge, Massachusetts, The MIT Press.

Subbarao, D. (2014), "Capital account management: toward a new consensus?", *What Have We Learned? Macroeconomic Policy after the Crisis*, G. Akerlof and others (eds.), Cambridge, Massachusetts, The MIT Press.

Swan, T. W. (1956), "Economic growth and capital accumulation", *Economic Record*, vol. 32, No. 1, Wiley.

Taylor, A. M. and M. P. Taylor (2004), "The purchasing power parity debate", *NBER Working Paper*, No. 10607, Cambridge, Massachusetts, National Bureau of Economic Research.

Verdoorn, P. J. (1949), "Fattori che regolano lo sviluppo della produttivitá del lavoro. *L'industria*". Published also as "Factors that determine the growth of labour productivity", *Italian Economic Papers*, L. Pasinetti (ed.), vol. II, Oxford, Oxford University Press, 1993.

Williamson, J. (2008), "Exchange rate economics", *Working Paper*, No. 08-3, Washington, D.C., Peterson Institute for International Economics.

Annex A1

Description of the variables and the databases used as sources

RER is the real effective exchange rate indicator obtained from series 11752, which is available in the Central Bank of Brazil's database [online] at:

https://www3.bcb.gov.br/sgspub/localizarseries/localizarSeries.do?method=prepararTelaLocalizarSeries.

Y is real per capita GDP in dollars. This variable was estimated by dividing monthly GDP in dollars from series 4385 by the population for the year under analysis from series 21774 (IBGE estimates), both of which are available in the Central Bank of Brazil's database (see the above entry for the Internet address). The population data for 1999 are from series 7330.

ToT is the terms of trade index of the Centre for Foreign Trade Studies (FUNCEX), which was obtained from IPEADATA, available [online] at: http://www.ipeadata.gov.br/.

CC is the balance on current account expressed as a percentage of GDP and was calculated by dividing the monthly balance of current transactions from series 2731 by monthly GDP from series 4385, both of which are available in the Central Bank of Brazil's database (see the above entry for the Internet address).

IDIFER is the spread between the monthly domestic short-term interest rate (360-day swap rate) and the monthly international short-term rate (the federal funds target rate of the United States, both of which are available [online] at:

https://www.blumberg.com/?utm_source=Microsoft&utm_medium=cpc&utm_campaign=BLUM.

RI is the stock of international reserves, expressed as a percentage of GDP and calculated by dividing the figure for the stock of international reserves taken from series 3546 by the figure for GDP taken from series 4385, both of which are available in the Central Bank of Brazil's database (see the above entry for the Internet address).

CR is the indicator of Brazil's country risk premium, which is represented by the J.P. Morgan emerging markets bond index (EMBI) rating for Brazil, which is available [online] at: www.macrodadosonline.com.br.

Impact of the Guaranteed Health Plan with a single community premium on the demand for private health insurance in Chile

Eduardo Bitran, Fabián Duarte, Dalila Fernandes
and Marcelo Villena

Abstract

In 2012, a Guaranteed Health Plan for the private health system was submitted to the Chilean National Congress, with the aim of offering a standardized flat-rate health-care plan. This paper evaluates the impact that the introduction of this plan would have on the demand for health insurance. The results suggest that the private insurance portfolio would shrink by 12.39%, which means that around 400,000 people would switch to the public system, thereby exacerbating the adverse-selection problem faced by the system and imposing an additional fiscal burden of US$ 200 million per year.

Keywords

Health, private sector, health insurance, prices, supply and demand, health services, public sector, statistical data, Chile

JEL classification

I130, C250

Authors

Eduardo Bitran is Executive Vice President of the Production Development Corporation (CORFO). Email: eduardo.bitran@uai.cl.

Fabián Duarte is Assistant Professor at the Economics Department of the Faculty of Economics and Business of the University of Chile. Email: fabduarte@fen.uchile.cl.

Dalila Fernandes is Accounts Director at App Annie. Email: dalilafernandes@gmail.com.

Marcelo Villena is Vice-Dean of the Faculty of Engineering and Sciences at Adolfo Ibáñez University (UAI), Chile. Email: marcelo.villena@uai.cl.

I. Introduction

The Chilean experience in the provision of health insurance is quite unusual. Firstly, few countries offer the option of choosing between multiple health insurance plans, as is the case in Chile. Secondly, the Chilean system displays a structural segmentation, with a low-income and high-risk population covered by public insurance through the National Health Fund (FONASA), and a high-income, low-risk population. The latter usually contracts insurance with private health insurance institutions (Isapres), thereby raising a serious problem of adverse selection. In 2004, a general regime of explicit health guarantees (GES) was introduced, which was pioneering in Latin America in seeking to establish care rights for patients and thus foster greater systemic equity (see Missoni and Solimano, 2010). Despite these reforms, the system continued to show evidence of adverse selection on the demand side.

In this context, the Health Reform Commission (CRS), set up by the Government in 2011, proposed a new Guaranteed Health Plan (PGS), designed to offer all users of the private system a flat-rate plan with standardized characteristics. This was submitted to the National Congress in 2012. Under this new plan, the premium is independent of age and gender and is charged on each member of the family group (in other words the contributor and dependants) and is financed by a 7% mandatory payroll deduction from the individual. If surpluses are generated from the mandatory contribution, the affiliate may use the difference to purchase supplementary insurance in the Isapre to which he/she contributes, subject to a table of risk factors. Should the mandatory contribution be insufficient, the affiliate must provide an additional contribution to make up the difference.[1]

According to CRS estimates, a PGS value of up to 21,000 Chilean pesos (Ch$ 21,000) is foreseen for 2011, assuming 50% coverage of the FONASA level-1 free choice modality (MLE) rate.[2] However, bargaining in the Senate has significantly reduced the co-payments, which is likely to push up the price of PGS.

The Guaranteed Health Plan combines the benefits already defined in GES, with additional coverage for catastrophic illnesses (CAEC). It also covers a group of diseases excluded from the GES plan, for which a network of providers will be specified; and it provides for a level of benefits that at least match the specifications of the Free Choice Modality of level 1 of the FONASA public insurance regime. In addition, the Guaranteed Health Plan will eliminate the waiting period and pre-existing conditions for private-system affiliates who wish to contract health insurance within the Isapre system. This measure allows individuals already affiliated to the private system to move between different insurers without restriction. This would clearly help mitigate the problem of affiliates who become captives of an Isapre when they reach a certain age or develop a chronic disease. By eliminating the waiting period and pre-existing conditions within the private system, and including a risk compensation fund that makes it possible to compensate Isapre plans that have a larger proportion of higher-risk affiliates, PGS tries to increase competition within the system and bring down entry barriers for new Isapres.

Nonetheless, the plan has fuelled debate around the consistency of its benefits. On the one hand, the introduction of a flat rate will mean a higher premium for certain groups of affiliates, who will be unable to finance the plan from their mandatory contribution alone. Using data from 2010, the authors estimate that at least 13% of affiliates, or close to 190,000 contributors, pay a premium that

[1] In this case, the option of contracting supplementary insurance is also left free.

[2] See "Informe Comisión - Octubre 2011. Evaluar la factibilidad de crear un Plan Garantizado de Salud (PGS) al interior del sub-sistema de ISAPRE" [online] https://www.politopedia.cl/wp-content/uploads/2016/06/Informe-Comisi%C3%B3n-PGS.2011.pdf.

is less than the PGS rate.[3] Moreover, PGS does not eliminate the waiting periods and pre-existing conditions for users wishing to move from the public system to an Isapre. The demand for public or private health insurance may therefore change according to the effects of the new policy on different social groups. Some groups will be more attracted to the Isapre system after the reform, while others will decide to move to FONASA. Young people facing increased premiums but with low health risks, and therefore low health costs, might naturally decide to move to FONASA.[4] Large families facing a high family premium, due to the per capita payment of the PGS rate, would also be encouraged to leave the private health system.[5] Thus, one of the key issues is how the reform will impact on demand between the two systems: Isapres and FONASA. This is the main focus of the present research. Although there are also considerations and pressures on the supply side, these are beyond the scope of this study.

The effect of the introduction of PGS on the Isapre system and FONASA is evaluated by developing a model to determine the demand for Isapres and FONASA for each affiliate, considering sociodemographic characteristics, the characteristics of the supply of services and user preferences.[6] The following variables should be taken into account: income level, risk, size of the family nucleus, age, gender and the conditions of the services offered by each system in each region of the country.

The model should therefore consider the specific features of the Chilean health system, such as the fact that dependent workers are required to purchase health insurance through a contribution consisting of a percentage of their wage; and the insurance alternatives include private plans offered by the Isapres and a single public system plan, FONASA, both of which have different pricing schemes.

This article is organized as follows. Section II discusses the relevant literature, both international and national, and section III describes the discrete-choice model, which makes several improvements on previous studies in Chile. Section IV describes the information used to calibrate the models and the descriptive statistics of the main variables used. Section V reports the study's key results, in particular the impact on demand of the implementation of PSG; and the last section sets forth the main conclusions and policy recommendations.

II. Literature review

Since the seminal study by Akerlof (1970), great attention has been paid to the effect of information asymmetry on decision-making in the health-care market. On the one hand, health insurance markets are prone to adverse selection because insurees know much more about their own health risks than the insurers, which leads to very generous coverage attracting high-risk affiliates. On the other hand, information asymmetry in the doctor-patient relationship can contribute to the overuse of health services, as the risk coverage is greater, thereby generating moral hazard.

Studies of the influence of moral hazard on decisions about choosing health insurance are based on expected utility maximization theory. Friedman (1974), for example, suggests a discrete-choice model for choosing between health insurance alternatives, which makes it possible to make inferences about perceived risk aversion with respect to the alternatives. The study, conducted for exploratory purposes in

[3] According to estimates made of the PGS price, a minimum income of about Ch$ 320,000 would be required for a single person to be insured through the mandatory contribution under PSG (Copetta, 2011). About 50,000 people would be unable to pay for PGS with the mandatory contribution; however, this estimate is biased because it does not consider the effect of dependants on the final price or on beneficiaries who pay more than the additional contribution (70% of the Isapre portfolio), many of whom would also suffer increases in the cost of their plans.

[4] The authors estimate that about 22% of young people aged under 25 years pay a lower premium than they would pay in PGS, compared to just 11% of the rest of the population.

[5] The authors estimate that about 25% of families with more than three members pay a lower premium than they would pay under PGS, compared to just 8% of families with fewer than three members.

[6] The study does not model or estimate supply, given its great complexity; so the results are only partial equilibria.

the United States, finds that family groups are less-risk averse than individuals. In subsequent research, Feldman proposes an analysis of the sensitivity of American workers' choices in relation to the health plan premiums paid directly by them (Feldman and others, 1989). The results show that individuals are highly sensitive to what they themselves pay directly. Cutler and Reber (1998) analyse the costs of adverse selection using information on how Harvard University employees choose their health plans. The authors consider a reform introduced by the institution, which in 1995 changed from a system of generous subsidies to one that only subsidizes a fixed amount of each insurance, independently of the plan and the premium. The results show that the reform reduced the costs borne by the University, since employees started to choose cheaper plans.

For the specific case of Chile, research in the field of health insurance only began in the 1990s, so it is sparse. Sapelli and Torche (1998) analyse factors underlying the choice of health insurance using the Survey of National Socioeconomic Characterization (CASEN) for 1990 and 1994. They conclude that the most important variables when deciding between public and private insurance are income, age and the individual's sector of residence. They also find that private information on the individual's health status generates adverse selection against private insurers, whereas public information on health status produces adverse selection against the public insurer. In later studies using CASEN for 1990 and 1996, Sapelli and Torche (2001) corroborate the importance of income and sector of residence in determining the individual's preference, but also add the relevance of the insurance price and expected out-of-pocket expenses in the form of co-payments. The authors stress that, at higher levels of expected health expense, public insurance becomes more attractive, so income cannot be the only variable explaining the choice of health insurance alternatives. Nonetheless, the study is limited by using a table of risk factors from a single private insurer. Although private insurers' risk factor tables are mutually correlated, price variations may reflect more than risk adjustment alone, since the different basic premiums charged in each plan depend on the type of coverage offered. The use of a single table thus fails to represent the average price of all insurers. To avoid this problem, the present study uses a vector of the average prices affiliates pay, according to age, gender, income level and number of family dependants, which has already incorporated the risk factors, the basic premium and individual preferences about whether or not to make an additional payment on top of the mandatory contribution.

Sapelli and Vial (2003) analyse adverse selection and expected expenditures in health systems. Their study confirms the problem of adverse selection in FONASA due to observable variables in the decision models; but the results do not identify adverse selection in Isapres, presumably owing to the design and variety of these plans, in which prices discriminate by risk.

Sanhueza and Ruiz-Tagle (2002) highlight the importance of the gender and age composition of the family nucleus. Yet one of the greatest shortcomings in this study is that the results of the health status variable show evidence of adverse selection against the private system, contrary to what might be expected in this system given the problem of pre-existing conditions. In another study, Tokman, Marshall and Espinosa (2007) find that the likelihood of a person being affiliated to an Isapre increases with income and decreases with the risk of members of the family group. They also find that the probability of affiliating to an Isapre increases with the availability of private beds in the region in which they live, among other factors.

Dawes Ibáñez (2010) analyses the effects of the implementation of the GES plan in the choice between public and private health systems, by comparing affiliates' preferences for the two health systems before and after implementation of the reform, using data from the 2000 and 2006 CASEN surveys. The authors use the Sapelli and Torche (2001) approach, drawing on the contributions made by Tokman, Marshall and Espinosa (2007). The results of the model before the reform concur with previous studies, highlighting the importance of income and risk in the choice of health system, and also adding the positive effects of having a spouse, education level and working in large firms on the probability of choosing private health insurance. In the post-reform analysis, the results suggest a weaker influence of individual health risk on the individual's preference, which the authors suggest is due to the better coverage of catastrophic risks in the GES plan, thereby creating a risk pool that reduces the difference in prices paid for private insurance. Nonetheless, a major shortcoming in the Dawes Ibáñez (2010) study is that it compares the pre- and post-reform situations with a six-year time interval, in which many other factors may have influenced changes in health-insurance preferences.[7]

Lastly, research by Duarte (2012) analyses the price elasticity of health expenditures, indicating possible preferences that could be invoked when analysing the choice of plans. Duarte uses data compiled from records of the Chilean private health insurance market and shows that high-income individuals are five times more price sensitive than low-income ones. The study does not cover the public system, however, since it works only with information from Isapre affiliates.

The aforementioned studies have focused on measuring the effect of prices and risk on individual choice in the Chilean case; but little has been done to analyse the influence of the quality of the systems and how this affects the choice of health insurance. The present study refines the analysis, by adding the quality component to improve the model's explanatory capacity. This should make it possible to identify individuals' preferences and their patterns of substitution between the different health systems more effectively, and thus predict the impact of PGS more accurately and obtain better results in the characterization of the groups affected by the reform.

III. The model

This study develops a discrete-choice decision model that simulates the demand for private health plans, but does not consider structural changes on the supply side. For a model to fit the discrete-choice framework, the alternatives available to choose from must be mutually exclusive; the set of alternatives must contain all the options available to the individual making the decision; and there must be a finite number of options. The individual chooses, from all the available alternatives, the one that gives him/her the greatest utility.

The individual's utility function is modelled using the random-utility approach, the application of which to economics was initially proposed by McFadden and later developed by Manski. Under this approach, utility functions are treated as random, to reflect not a lack of rationality but rather the observer's lack of information on the characteristics of the alternatives or individuals. This model assumes that the utility of the individual (U_i) can be separated additively into an observable deterministic component (ε_i) and an unobservable random component (V_i) for each alternative. Thus, considering individuals who choose different insurance systems:

$$U_{iI} = V_{iI} + \varepsilon_{iI} \tag{1}$$

$$U_{iF} = V_{iF} + \varepsilon_{iF} \tag{2}$$

where U_{iI} represents the utility that individual i obtains from the Isapre health insurance (subindex I), and U_{iF} is the utility individual i obtains by choosing the FONASA insurance (subindex F). The component ε_i captures all factors that affect utility but are not included in V_i. The probability that individual i selects the private health system is therefore:

[7] During these years, for example, inequality indicators in Chile fell sharply, which may influence the reduction in the income effect. Moreover, the number of Isapre plans shrank from 16,696 in December 2003 to 7,454 in January 2006, indicating lower price discrimination in private plans as a result of the restrictions implemented by the Isapres Law. Since price discrimination in this case involves the creation of more affordable plans for lower-income affiliates, this could be a relevant factor in making the private health system less attractive. Another limitation is that the implementation of GES cannot be analysed using control groups, since it applies to both Isapre and FONASA affiliates. Since GES is a new attribute of both the public and the private insurance, it cannot be claimed a preference for one or the other is a result of its implementation.

$$\begin{aligned} P_{if} &= Prob(U_{iI} > U_{iF}) \\ &= Prob(V_{iI} + \varepsilon_{iI} > V_{iF} + \varepsilon_{iF}) \\ &= Prob(\varepsilon_{iF} + \varepsilon_{iI} < V_{iI} + V_{iF}) \end{aligned} \tag{3}$$

This consists of the cumulative probability that each random term $\varepsilon_{iF\_}\varepsilon_{iI}$ is below the value observed for $V_{iI\_}V_{iF}$. For example, for logit models, the error term is distributed according to the extreme value density function. So:

$$P_{iI} = \int I\left(\varepsilon_{iF} - \varepsilon_{iI} < V_{iI} \text{ - } V_{IF}\right) f\left(\varepsilon_i\right) d\varepsilon_i \tag{4}$$

where $I(\cdot)$ is the indicator function, equal to 1 when the expression in parentheses is true and 0 otherwise. In this theoretical framework, we assume the following utility functions:

$$U_{iI} = \alpha(Y_i - P_{iI}) + \beta X_{iI} + \gamma X_{iI} Z_i + \delta Z_i + \varepsilon_{iI} \tag{5}$$

$$U_{iF} = \alpha(Y_i - P_{iF}) + \beta X_{iF} + \gamma X_{iF} Z_i + \delta Z_i + \varepsilon_{iF} \tag{6}$$

where

$Y_i =$ the individual's income;

$P_{iI} =$ price of private insurance charged by the Isapre for the individual i;

$P_{iF} =$ price of FONASA public insurance for individual i, corresponding to 7% of individual income, or 0.07 Y_i;

$X_{iI} =$ vector of attributes that Isapre health insurance offers the individual i;

$X_{iF} =$ vector of attributes that the FONASA health insurance offers to individual i;

$Z_i =$ vector of sociodemographic characteristics of individual i;

Equations (5) and (6) represent the utility function of Isapres and FONASA for individual i, respectively. Preferences are assumed stable over time.

The expected utility for each alternative is considered to depend positively on net income after the insurance premium has been paid. So, as health insurance becomes more expensive, its expected utility declines. Expected utility also depends on the attributes of the service delivered by the health plan: the higher the quality, the greater the expected utility. This is empirically supported by an opinion survey conducted by the Chilean Health Superintendency, which shows that the quality of medical care, the speed of care and the quality of the doctors are among the characteristics most highly valued by users of the private system. In the public system, however, one of the most prized characteristics is the low cost of care, in other words low co-payment levels (Superintendency of Health, 2012). The individual's sociodemographic characteristics also influence the level of utility perceived from a health insurance policy. Someone who is more risk averse, for example, tends to value greater coverage and be less sensitive to price; while an individual who resides in rural areas where there are few private hospitals tends to value public health insurance more highly.

When deciding on a health system, each individual evaluates the difference between the utility of being in the Isapre system and the utility of being in FONASA, and chooses the option that maximizes his/her utility. Thus, if the difference is positive, private insurance is chosen; if the difference is negative, public insurance is preferred.

Subtracting equation (6) from (5), gives:

$$U_{iI} - U_{iF} = \alpha(0.07\,Y_i - P_{iI}) + \beta\Delta X_i + \gamma\Delta X_i Z_i + \Delta\varepsilon_i \quad (7)$$

Unlike previous studies, the model being proposed here specifies that the vector of health insurance attributes consists of: (i) a vector of observable quality and service characteristics, ΔX_{i1}; and (ii) a vector of unobservable characteristics ΔX_{i2}. Lastly, $\Delta X_{i=} \Delta X_{i1} + \Delta X_{i2}$, so:

$$U_{iI} - U_{iF} = \alpha(0.07\,Y_i - P_{iI}) + \beta(\Delta X_{i1} + \Delta X_{i2}) + \gamma(\Delta X_{i1} + \Delta X_{i2})Z_i + \Delta\varepsilon_i \quad (8)$$

In the estimation, an unobservable is assumed for each option (FONASA or Isapre), so, the difference for this unobservable is normalized to 1. As the sociodemographic characteristics of the individual are the same regardless of their choice,

$$I = U_{iI} - U_{iF} = \alpha(0.07\,Y_i - P_{iI}) + \beta_1\Delta X_{i1} + \beta_2\Delta X_{i2} + \gamma_1\Delta X_{i1}Z_i + \gamma_2 Z_i + \Delta\varepsilon_i \quad (9)$$

Unlike previous studies, this model does not use the individual risk factor as a proxy variable for the price of private insurance, but instead a vector of estimated prices according to the contributor's age, number of dependants and income level.[8] Use of the price vector rather than the individual-risk proxy variable allows the new estimated price for PGS to be used as a counterfactual in the decision model, which makes it possible to simulate individuals' choices between public and private insurance following implementation of the reform. In particular, the study evaluates a single community premium. This is an average price and not the price that the average individual observes, so it biases the results; but it is a better approximation than those previously reported in the literature.

Once the effect of the reform on the relative size of the two systems and their characteristics in terms of income, age, family size and health status of individuals that move from one system to the other, this analysis will make it possible to estimate the eventual fiscal cost of the reform, to the extent that affiliates who spend more than they contribute move to FONASA.

As this research only models the demand side, it implicitly assumes a scenario where the supply is adjusted to maintain the economic rent perceived by the constant sector. This requires a significant increase in the price of PGS, which explains the fall in the demand for the Isapre service. This can then be considered an extreme scenario. If the Isapres have large economic rents, they could for example reduce the impact of the reform, by absorbing the higher costs entailed by this more solidarity-based system, generating a smaller price increase, reducing the final price to their affiliates and thus mitigating the drop in demand.

IV. Data

The data used in this study were obtained from CASEN 2013, which provides information on individuals' housing conditions, employment status, educational level, health conditions and demographic characteristics. The survey, which questioned 66,724 households across Chile between November 2013 and January 2014, is representative nationally, by urban and rural area, for the 15 regions of the country and for a total of 324 boroughs (*comunas*).

8 Authors' estimation from data extracted from Duarte (2012).

One of the necessary requirements for using the discrete-choice model is that the alternatives facing an individual must be mutually exclusive. For this reason, the decision model only applies to employees, manual labourers or retirees, because these individuals are subject to the mandatory health contribution, so they have to choose between contracting a plan in an Isapre or joining FONASA, paying at least 7% of their wage in both cases.

The developed model is applied to the head of the family nucleus, since it is he/she who chooses which health insurance alternative to affiliate to, according to his/her individual characteristics and those of the family nucleus, and the attributes offered by the available alternatives. The CASEN 2013 survey contains 218,490 observations, distributed among 78,938 households. The expanded sample has a population of 17,273,085 people, distributed over 6,256,912 homes.

As noted above, the sample for analysis includes only employees, labourers and retirees with a non-zero taxable income and who are affiliated to the Isapre system or FONASA groups B, C and D. Table 1 reports the distribution of total health system affiliates in 2013.

Table 1
Chile: distribution of affiliates by income quintile and type of health insurance, 2013
(Percentages)

	Quintile I	Quintile II	Quintile III	Quintile IV	Quintile V	Total
FONASA	11.0	18.4	17.6	15.6	7.0	69.6
Isapres	0.7	0.5	1.6	4.6	15.2	22.6
Other	1.2	1.2	1.2	2.1	2.0	7.7
Total	12.9	20.1	20.4	22.4	24.2	100.0

Source: Prepared by the authors, on the basis of data from the National Socioeconomic Survey 2013.

Table 2 shows the 2013 distribution of total affiliates between the public and private systems in the final sample. Table 3 displays a number of demographic characteristics of the final sample.

Table 2
Chile: distribution of affiliates by income quintile and type of health insurance after purging the sample, 2013
(Percentages)

	Quintile I	Quintile II	Quintile III	Quintile IV	Quintile V	Total
FONASA	9	21	20	17	7	74
Isapres	1	1	2	5	17	26
Total	9.3	21.8	21.9	22.7	24.2	100.0

Source: Prepared by the authors, on the basis of data from the National Socioeconomic Survey 2013.

A major disadvantage in using surveys is that they contain data declared by the affiliates themselves, which may be subject to serious inaccuracies. In this case, for example, information on individual income corresponds to the value declared as net rather than gross income. To be used in the model, this value is transformed into taxable income, adjusting the corresponding values of income tax and health and pension fund contributions. Moreover CASEN 2013 does not contain effective information on the number of dependants per affiliate, which is why members of the family nucleus under 18, or over 18 but under 25 if attending an educational establishment are considered as dependants. It is also not possible to distinguish how the corresponding expenses are distributed in families in which both spouses work and contribute to the health system. For this reason, it is assumed that the head of the family nucleus makes the decision and bears all the expenses.

The public insurance premium, 7% of the wage, was estimated using a variable that represents the wage of the head of the family nucleus. If the spouse is an employee, labourer or retiree, his/her wage is attributed to this variable, which defines a "single contributor", who is assumed to bear all family expenses. This simplification reduces the problem of the lack of information on the distribution of expenses between the spouses. Nonetheless, it assumes that spouses who work and pay contributions choose the same health system.

For the private insurance premium, as noted above, an average estimate will be used according to the affiliate's age, number of dependants and income level, using data from the Superintendency of Health Insurance Institutions.

The other variables used in this study represent the individual characteristics of the survey respondents, frequently used in the studies reviewed for the Chilean case. These variables are used on a standalone basis in the model, and also to interact with the variables that contain the attributes related to the quality and service of public and private health insurance providers. These include: age, an indicator of gender, an indicator of whether the head of the family has a spouse, the number of dependants and two indicators of geographical location (one corresponding to Santiago and the other representing whether the individual lives in a small city). Tables 3 and 4 show selected statistics on these variables.

Table 3

Chile: descriptive statistics, by quintile and system, 2013

(Percentages, years of age, Chilean pesos and number)

	System	Quintile I	Quintile II	Quintile III	Quintile IV	Quintile V
Affiliates	FONASA	94.3	97.3	91.0	76.5	28.7
	Isapres	5.7	2.7	9.0	23.5	71.4
Female affiliates	FONASA	97.3	97.7	95.6	79.7	33.6
	Isapres	2.7	2.3	4.4	20.3	66.4
Male affiliates	FONASA	92.3	97.0	89.5	75.4	27.3
	Isapres	7.7	3.0	10.5	24.6	72.7
Average age (years)	FONASA	51.3	49.1	46.7	46.7	47.1
	Isapres	51.4	42.3	41.9	44.4	43.6
Average taxable income (Chilean pesos)	FONASA	198 771	271 453	364 389	549 420	1 181 443
	Isapres	207 345	290 311	415 292	646 902	2 232 254
Average risk	FONASA	3.5	3.4	3.2	3.4	3.5
	Isapres	3.6	3.0	2.6	2.8	2.8
Dependants	FONASA	2.1	2.3	2.4	2.4	2.6
	Isapres	2.2	2.1	2.3	2.1	2.2

Source: Prepared by the authors, on the basis of data from the National Socioeconomic Survey 2013.

Table 4

Chile: descriptive statistics of independent sociodemographic variables, 2013

Variable	Mean	Standard deviation
Age	43.25	12.36
Education	5.88	3.25
Spouse	0.25	0.43
Santiago or Metropolitan Region	0.51	0.50
Small cities	0.41	0.49
Health status	5.75	1.14
Risk aversion	0.16	0.37
Pre-existing conditions	0.23	0.42
Days of hospitalization	0.01	0.11
Risk	3.08	1.30
Number of people	2.36	1.16

Source: Prepared by the authors, on the basis of data from the National Socioeconomic Survey 2013.

Health status variables are also used, including: two indicators of individuals' (self-reported) health status. The first takes the value 1 if the individual self-evaluates with grades 1 or 2, ("very bad" or "bad"). The second indicator takes the value 1 if the respondent self-evaluates 3 or 4 ("less than reasonable" or "reasonable"). The remaining scores of 5, 6 and 7 correspond to "more than reasonable", "good" and "very good", respectively. These variables attempt to capture trends in adverse selection and moral hazard after the implementation of the reform. A variable is constructed that captures the sum of the risk factors among the dependants and the head of the family nucleus. This is based on a table of risk factors updated to 2009, calculated from a weighted average of the tables of Isapres functioning in 2005 and the proportion of beneficiaries of each Isapre operating in 2009.

Lastly, the model incorporates quality and service variables to represent another dimension of the attributes of health insurance. Firstly, using data from the Department of Health Statistics and Information (DEIS) of the Ministry of Health,[9] the number of health facilities available per 100,000 inhabitants in each region for each system is constructed, as shown in table 5. Secondly, a variable is constructed to represent affiliates' average expenses in the two systems, by region.[10]

Table 5
Chile: number of health facilities, by region, 2013
(Per 100,000 inhabitants)

Region	Isapres	FONASA
I	0.65	0.33
II	1.58	0.35
III	1.08	0.36
IV	0.28	0.42
V	0.63	0.46
VI	0.34	0.23
VII	0.30	0.30
VIII	0.59	0.40
IX	0.31	0.31
X	0.24	0.36
XI	0.96	0.96
XII	1.27	0.63
RM	1.03	0.34
XIV	0.53	0.26
XV	3.23	0.54

Source: Prepared by the authors, on the basis of data from the Department of Health Statistics and Information of the Ministry of Health.

Table 6 shows the differences in the averages of the quality and service variables, by income quintile.

[9] The most recent DEIS database of hospital discharges is used for 2011. To construct the index, only hospitals classified as more complex establishments are selected. Rural health posts, and urban, family, family-community or rural health centres are not considered; nor are vaccination centres, facilities of medium or lower complexity, field hospitals or military campaign hospitals, primary emergency care services (SAPU); mobile dental clinics, sanitary bureaus, clinics under contract, general rural consulting rooms, closed care centres; health referral centres, therapeutic diagnostic centres and delegated hospitals. In the case of private establishments, only hospitals and clinics are included.

[10] The average is calculated including emergency and specialty medical consultations, x-rays or ultrasound procedures and preventive check-ups. The average does not include expenses in respect of hospitalization or surgery. The averages are based on data from the Social Protection Survey (EPS) of 2009; it was decided not to work with the latest available EPS (2012), since the Undersecretariat of Social Security considers it unreliable.

Table 6
Chile: quality variables, by income quintile, 2009
(Chilean pesos)

	Quintile I	Quintile II	Quintile III	Quintile IV	Quintile V
Average co-payment difference	20 178	17 502	14 862	12 809	9 254
Average difference in hospitals per inhabitant	0.32	0.43	0.53	0.56	0.64

Source: Prepared by the authors, on the basis of data from the 2009 Social Protection Survey.

Lastly, other variables are used for the purpose of replicating previous results, such as: the logarithm of the income of the head of the family nucleus; a price variable (risk factor) in two versions —one with the same data used by Sapelli and Torche (2001) and another updated to 2009— calculated from a weighted-average of the tables of Isapres open in 2005 and the proportion of beneficiaries of each Isapre open in 2009; and the number of members of the family nucleus who have been hospitalized for more than eight days in the three months prior to the survey.

V. Results

The demand for private health insurance is a function of the difference in prices between public and private insurance: the more negative this difference, the less likely it is that the individual will join an Isapre. The equation also includes insurance quality and service attributes, described in the vector of observable attributes: the better the quality and service of Isapre plans compared to FONASA, the more likely it is that the individual will choose an Isapre. Given the normalization of the vector of unobservable attributes, the individual's sociodemographic characteristics in themselves represent an effect in terms of preference for a health plan. Similarly, individual characteristics, such as age, income, gender, sector of residence and others, influence how people perceive the specific attributes of each plan. For this reason, the model includes a series of interactions between individual sociodemographic variables and health insurance quality and service variables, to more effectively characterize the population's heterogeneous preferences. The main purpose of the analysis is to study the individual's sensitivity to changes in the health insurance premium, while controlling for sociodemographic and quality variables. This approach makes it possible to analyse how the reform may affect adverse selection, which could affect the public insurance because the public system does not impose restrictions on accepting affiliates with any type of risk or expected cost to the system.

1. Comparison of the results with those reported in the literature

The variable of interest is a binary variable that takes the value 1 if the individual chooses the Isapre system and 0 if he/she chooses FONASA. Table 7 shows the results of the regression for the model proposed by Sapelli and Torche. Regression A replicates the model using the risk factor table used in the original model; B uses the updated risk factor table and, lastly, C uses the price vector constructed on the basis of the real prices paid by individuals.

The results reported in the table show that price sensitivity has ceased to be significant for 2013 (regression A), while income sensitivity has remained. The use of the risk factor as a proxy for the private premium (regression B) is not significant, unlike the price created from real premiums (regression C). This implies that the risk factor would not be the most efficient proxy variable for the average price, because it does not incorporate the effect of the basic premium, which is different for each plan.

Table 7
Chile: coefficients and standard deviations of the regressions

Variables[a]	Sapelli and Torche CASEN 1996	Regression A	CASEN2013	Regression B	CASEN2013	Regression C	CASEN2013
	Coefficients	Coefficients	Standard deviation	Coefficients	Standard deviation	Coefficients	Standard deviation
L_ingreso	2.10**	2.647***	0.199	2.650***	0.198	2.906***	0.183
Precio	-0.20**	-0.105	0.079	-0.081	0.077	-0.0142**	0.005
Cónyuge	0.47**	0.332***	0.079	0.301***	0.081	0.439***	0.075
Edad	-0.04**	-0.004	0.003	-0.002	0.003	-0.0101**	0.003
Días_Hosp	-0.49	-0.002	0.102	0.002	0.094	-0.044	0.095
Santiago_RM	5.11**	3.824	2.905	4.140	2.898	5.875*	2.750
Ciud_Pequeñas	3.32**	4.724	2.912	4.961	2.908	7.048*	2.755
Lin_Stgo	-0.29*	-0.168	0.218	-0.174	0.217	-0.348	0.208
Lin_Pequeñas	-0.18	-0.290	0.218	-0.301	0.218	-0.481*	0.209
Pre_Stgo	-0.05	-0.127	0.092	-0.170	0.087	0.000	0.000
Pre_Pequeñas	-0.03	-0.054	0.088	-0.065	0.084	0.000	0.000
Constant	-24.84***	-36.91***	2.644	-37.01***	2.63	-40.21***	2.407
Pseudo-R^2	0.26	0.377		0.378		0.373	

Source: Prepared by the authors.
Note: ***Significant at 1%; **significant at 5%; *significant at 10%.
[a] These variables are as defined in Sapelli and Torche (1998): L_ingreso (logarithm of income); Precio (price – the constructed index); Cónyuge (spouse dummy variable); Edad (age); Días_Hosp (number of family members who have been hospitalized for more than eight days in the last three months); Santiago_RM (dummy variable identifying Santiago (Chile)); Ciud__Pequeñas (dummy variable identifying small cities); Lin_Stgo and Lin_Pequeñas (interactions between the dummy variables of cities and the logarithm of income); Pre_Stgo and Pre__Pequeñas (interactions between city dummy variables and price).

2. Results of the proposed model

Table 8 shows the regression results for the proposed model. The variable Dif_Precio represents the difference between the public and private premiums. As expected, the more negative the price difference between public insurance and private insurance, the smaller the chance of affiliation to the Isapre system. The other variables related to health insurance attributes of were constructed as their value in the case of private insurance minus their value in public insurance.

A greater difference in co-payments between Isapres and FONASA (the *Dif_Copago_sis* variable) implies a higher probability of affiliating to FONASA. On the other hand, a larger difference between the number of health-care facilities in the individual's region increases the likelihood of affiliation to the Isapre system.

The positive sign found for the age variable (*Edad*) indicates an increase in the risk profile of Isapre affiliates. One possible reason for this effect is the introduction of the additional coverage for catastrophic illnesses (CAEC) in 2006, which may have affected those with a higher health-risk profile, making private health insurance more attractive to them. Even so, the health status variable, *Estado_Salud*, which captures information on the individual's health risk that is not available to insurers, indicates a preference for public insurance if there is evidence of a worse health status.

On the other hand, the existence of a working spouse who pays into a health system increases the likelihood of affiliation to Isapres, due to the interaction between the *Dif_Copago_sis* and *Cónyuge* (spouse) variables and despite the larger co-payments in the private system. This is because a working spouse increases family income, which explains a greater preference for the Isapre system.

On average, the *Santiago_RM* and *Ciud_Pequeñas* variables show that people living in metropolitan Santiago or small cities have a stronger preference for Isapres. Moreover, as expected, an increase in income increases the likelihood of affiliation to the private system.

Table 8
Chile: coefficients and standard deviations of the regression of the proposed model

	Coefficients	Standard deviation
Dif_Precio	0.076	0.047
Dif_Copago_Sis	-0.647	0.740
Dif_Hosp_Hab	0.763	0.505
Ysueldo	0.000***	2.61E-07
Edad	0.002	0.006
Sexo	-0.185	0.119
Cónyuge	0.323	0.177
Estado_Salud1	0.083	0.616
Estado_Salud2	-0.511	0.272
Santiago_RM	1.495***	0.297
Ciud_Pequeñas	0.797**	0.265
Num_Pers	0.003	0.0818
Dif_Cop x Ysueldo	0.000	4.85E-07
Dif_Cop x Edad	-0.000	0.011
Dif_Cop x Sexo	0.152	0.234
Dif_Cop x Cónyuge	0.325	0.221
Dif_Cop x Salud1	0.739	0.613
Dif_Cop x Salud2	0.555	0.313
Dif_Cop x Numpers	-0.069	0.144
DifHosp x Ysueldo	0.000	5.08E-07
DifHosp x Edad	-0.012*	0.005
DifHosp x Sexo	-0.203	0.145
DifHosp x Cónyuge	-0.593***	0.154
DifHosp x Salud1	-0.312	0.386
DifHosp x Salud2	0.109	0.163
DifHosp x Santiago_RM	0.889*	0.383
DifHosp x Ciud_Peq	0.357	0.246
DifHosp x Numpers	-0.103	0.0587
DifPrecio x Cónyuge	-0.034	0.0334
DifPrecio x Salud1	0.118	0.104
DifPrecio x Salud2	-0.025	0.0403
DifPrecio x Santiago_RM	0.043	0.0528
DifPrecio x Ciud_Peq	-0.026	0.047
Constant	-3.836***	0.394

Source: Prepared by the authors.
Note: ***significant at 1%; **significant at 5%; *significant at 10%.

3. Effects of the implementation of the Guaranteed Health Plan with a single community premium

Using the model described in the foregoing sections, the transfer of affiliates between the Isapre and FONASA systems is estimated after the implementation of PGS with a single community premium. For the purposes of the simulation, the data provided on the PGS coverage level by the Health Reform Commission were corrected. An average coverage level of 75% was assumed, at 2013 prices. This generates a flat rate of Ch$ 26,989, above the range estimated by CRS, which considered lower coverage (50% in ambulatory FONASA free-choice and 60% for hospitalization in the José Joaquín Aguirre Clinical Hospital of the University of Chile). This increase in coverage was considered to approach the proposal made in the Senate Health Committee,[11] which, at current prices, implies a cost estimated

[11] The Ministry's proposal considers 80% hospitalization coverage (J. J. Aguirre hospital fee) and 60% ambulatory coverage, thereby raising the coverage defined in the Health Reform Council study.

by the Ministry of Health of a flat rate Ch$ 28,432 to pesos per person at 2013 prices.[12] The value of Ch$ 26,989 per beneficiary was obtained by imposing the 75% coverage and calculating the premium generated by the same total profit margin of the Isapre system as existed pre-reform. This value could overestimate the community premium if the total profit margin falls as a result of the increased competition that could be generated by eliminating the pre-existing conditions and waiting periods within the Isapre system. Lastly, it should be noted that the value includes a work disability subsidy (SIL) of 1.5% of the affiliate's average wage.

The results obtained and displayed in table 9 indicate that 12.39% of the Isapre system's total portfolio, or roughly 1 million people, transferred to FONASA. In all the quintiles the net effect is a transfer from Isapres to FONASA, but it is the lower-income groups that move on a massive scale.

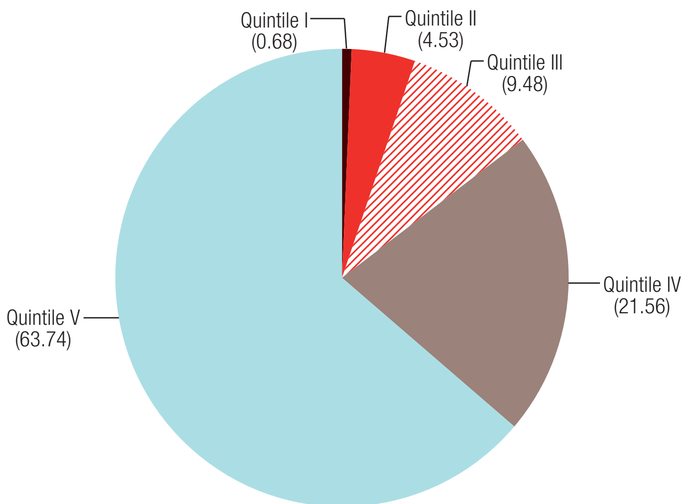
Table 9
Chile: changes in the Isapres portfolio, by quintile
(Percentages)

Income quintile	Variation in Isapre portfolio	Group-specific variation
I	-0.83	-71.64
II	-2.41	-53.10
III	-3.46	-36.45
IV	-3.94	-18.27
V	-1.76	-2.78
Total	-12.39	-12.39

Source: Prepared by the authors, on the basis of regression results and simulated scenario with the Guaranteed Health Plan.

A comparison of the proportion of affiliates that changed by income quintile, relative to each group's initial participation in the Isapres, shows that the effect of the reform is to accentuate adverse selection against FONASA. Only 1.76% of the families that change belong to the fifth quintile, yet that group represents 63.74% of Isapre affiliates (see figure 1).[13]

Figure 1
Chile: participation in the Isapres, by income quintile, 2013
(Percentages)



Source: Prepared by the authors, on the basis of data from the National Socioeconomic Survey 2013.

[12] Report of the Minister of Health to the Senate Health Committee, June 2013.

[13] The lowest-income families are over-represented among those who migrate to FONASA, relative to the average Isapre population.

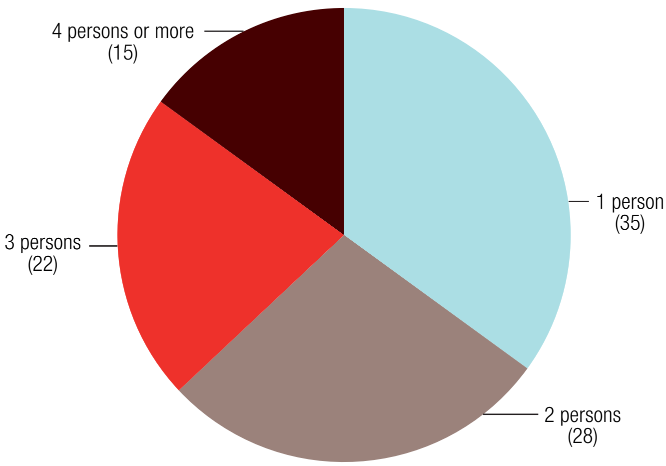
A higher proportion of large families change than single individuals or small families. As shown in table 10, 20.26% of families with four or more members switch to FONASA; this group represents 15% of the total system (see figure 2).

Table 10
Chile: changes in Isapre portfolio, by family group
(Percentages)

Family nucleus	Variation in Isapre portfolio	Group-specific variation
1 person	-1.96	-5.69
2 persons	-3.44	-12.34
3 persons	-3.87	-17.42
4 persons or more	-3.12	-20.26
Total	-12.39	-12.39

Source: Prepared by the authors, on the basis of regression results and simulated scenario with the Guaranteed Health Plan.

Figure 2
Chile: participation in the Isapres, by family group, 2013
(Percentages)



Source: Prepared by the authors, on the basis of data from the National Socioeconomic Survey 2013.

This also represents an adverse selection effect that generates an increase in the FONASA deficit, since FONASA families pay a contribution that is proportional to their income (7%) but the expense depends on the number of dependants.

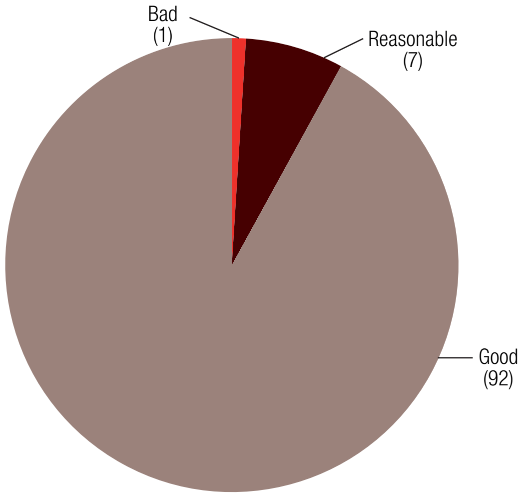
Table 11 shows that the health status of persons who migrate to FONASA is not a significant factor generating adverse selection (see also figure 3).

Table 11
Chile: changes in the Isapre portfolio, by health status
(Percentages)

Health status	Variation in Isapre portfolio	Group-specific variation
Good	-9.69	-10.52
Reasonable	-3.09	-46.83
Bad	0.40	33.04
Total	-12.39	-12.39

Source: Prepared by the authors, on the basis of regression results and simulated scenario with the Guaranteed Health Plan.

Figure 3
Chile: participation in Isapres, by health status, 2013
(Percentages)



Source: Prepared by the authors, on the basis of data from the National Socioeconomic Survey 2013.

Lastly, in order to simulate the cost to FONASA of the 400,000 beneficiaries transferring from Isapres, we consider the data available up to 2011 on costs for FONASA group D as an estimate of the cost of care, adjusted by the health CPI up to 2013 and considering a cost for FONASA of Ch$ 26,190.[14]

As the FONASA premium is 7% of income, from which the Subsidy for Work Disability due to Illness must be financed (1.5% of income), the income of those that switch system averages Ch$ 251,003 and the average number of family members is 2.9, so the average contribution per beneficiary is Ch$ 5,522 per month. The monthly FONASA deficit for each person who changes would be Ch$ 20,668. Considering the price situation of 2013, the annual deficit per beneficiary transferred would be Ch$ 248,015, similar to FONASA's total expenditure per average beneficiary. In conclusion, the reform implies a 12% increase in FONASA expenditure to serve beneficiaries who move to the public system.

VI. Conclusions

This study developed a discrete-choice model, using data from the CASEN 2013 survey, to model how people in Chile choose their health insurance system. Unlike previous studies, the model does not use the individual risk factor as a proxy for the price of private insurance, but uses a vector of estimated prices according to age, number of dependants and the affiliate's income level. The model also explicitly considers the influence of system quality of when selecting health insurance, which, as shown in the analysis, improves the model's explanatory capacity.

The model is used to simulate how the affiliates' choice would change when introducing a guaranteed health plan with a flat-rate premium, in other words to simulate changes in the demand for health plans, so that the results represent a partial analysis of the effect of PGS. This does not consider the possible supply-side reaction in terms of adding complementary or supplementary plans to PGS. Any such new plans would change the "total premium" and the "total coverage" that individuals would face, which would thus change the demand for health plans. For example, if an individual has strong

[14] See Bitran, Arpón and Debrott (2013).

preferences for coverage, but also for price, when facing two plans that are similar in those dimensions it is unclear which would be chosen. On the other hand, if the same individual is offered a plan that complements the basic plan with an acceptable premium, the individual in question would choose this plan. This partial effect on the demand side is large enough not to be considered in a more exhaustive analysis of the supply of and demand for private insurance.

The results of the study indicate that the Isapre portfolio would shrink by 12.39%, which means that around 400,000 people would switch to FONASA.[15] The changers would mainly consist of large families and individuals belonging to lower income quintiles. This result exacerbates the adverse selection problems already existing in the FONASA system. In other words, the introduction of PGS would have an adverse financial impact on FONASA, thereby reducing its ability to meet its obligations.

As noted above, the reform implies a 12% increase in FONASA expenditure to serve beneficiaries who move to the public system. Considering the 2013 budget, the cost to the Treasury of implementing PGS with a flat rate is estimated at around US$ 200 million per year. In addition to the financial stress implied by an immediate increase in spending of this magnitude, FONASA would have to incorporate some 400,000 people who are accustomed to a different service modality, with all the difficulties that this entails. These families will probably spend more than the average of FONASA group D affiliates, since they are accustomed to the free-choice modality. Moreover, this massive change would to some extent displace the current FONASA population in the use of resources, since the latter has less training in the use of insurance policies and fewer resources.

In conclusion, it is not advisable to implement a Guaranteed Health Plan with a single community premium in a health insurance system as segmented as Chile's, since it tends to exacerbate the problem of adverse selection already present in the system, increasing segmentation and requiring a huge amount of additional State funding. Accordingly, the recommendation is to analyse a more inclusive pricing policy for PGS, addressing demand on a differentiated and comprehensive basis (between the public and private system), with premiums that differ according to factors such as the age and socioeconomic level of the affiliates and the interaction with supply.

[15] The simulation is based on 2 million affiliates, who represent a large percentage of the total portfolio. This is necessary owing to the data restrictions that apply when using surveys. In particular, the database excludes self-employed affiliates, which could introduce a bias.

Bibliography

Akerlof, G. A. (1970), "The market for '*lemons*': quality uncertainty and the market mechanism", *The Quarterly Journal of Economics*, vol. 84, No. 3.

Bitran, R., P. Arpón and D. Debrott (2013), "Opciones de reforma a la seguridad social en salud en Chile", *Documento de Referencia*, No. 13, Santiago, Espacio Público.

Copetta, C. (2011), "Minuta Plan Garantizado de Salud Para el Sector Isapres", *Working Paper*, Santiago, Superintendence of Health.

Cutler, D. M. and S. J. Reber (1998), "Paying for health insurance: the trade-off between competition and adverse selection", *The Quarterly Journal of Economics*, vol. 113, No. 2, Oxford, Oxford University Press.

Dawes Ibáñez, A. (2010), "Health care reform and its effect on the choice between public and private health insurance: evidence from Chile", thesis, Santiago, Institute of Economics, Catholic University of Chile.

Duarte, F. (2012), "Price elasticity of expenditure across health care services", *Journal of Health Economics*, vol. 31, No. 6, Amsterdam, Elsevier.

Feldman, R. and others (1989), "The demand for employment-based health insurance plans", *Journal of Human Resources*, vol. 24, No. 1, Wisconsin, University of Wisconsin Press.

Friedman, B. (1974), "Risk aversion and the consumer choice of health insurance option", *The Review of Economics and Statistics*, vol. 56, No. 2, Cambridge, Massachusetts, The MIT Press.

Missoni, E. and G. Solimano (2010), "Towards universal health coverage: the Chilean experience", *World Health Report (2010) Background Paper*, No. 4, World Health Organization (WHO) [online] https://pdfs.semanticscholar.org/226c/4621cd397092dfd556f7b19f02048d6f5620.pdf?_ga=2.163282104.523300194.1501966237-1570332244.1501185329.

Sanhueza, R. and J. R. Tagle (2002), "Choosing health insurance in a dual health care system: the Chilean case", *Journal of Applied Economics*, No. 001.

Sapelli, C. and A. Torche (2001), "The mandatory health insurance system in Chile: explaining the choice between public and private insurance", *International Journal of Health Care Finance and Economics*, vol. 1, No. 2, Springer.

___(1998), "El seguro previsional de salud: determinantes de la elección entre seguro público y privado, 1990-1994", *Cuadernos de Economía*, vol. 35, No. 106, Santiago, Institute of Economics, Catholic University of Chile.

Sapelli, C. and B. Vial (2003), "Self-selection and moral hazard in Chilean health insurance", *Journal of Health Economics*, vol. 22, No. 3, Amsterdam, Elsevier.

Superintendency of Health (2012), "Estudio de opinión y satisfacción de usuarios", Santiago [online] http://www.supersalud.gob.cl/documentacion/666/articles-7632_recurso_1.pdf.

Tokman, M., C. Marshall and C. Espinosa (2007), "Un matrimonio difícil: la convivencia entre un seguro público solidario y seguros de salud competitivos", *Financiamiento del Desarrollo series*, No. 199 (LC/L.2851-P), Santiago, Economic Commission for Latin America and the Caribbean (ECLAC).

Guidelines for contributors to the *CEPAL Review*

In order to facilitate the submission, consideration and publication of articles, the editorial board of the *cepal Review* has prepared the following information and suggestions to serve as a guide for future contributors.

The submission of an article implies an undertaking by the author not to submit it simultaneously to other publications. The copyright to all articles published in the *Review* shall be owned by the United Nations.

Each article will be reviewed by the editorial board, which may decide to submit it to external referees.

Papers should be submitted in the original language (English, French, Portuguese or Spanish). They will be translated into the appropriate language by the relevant eclac services.

Each article must be accompanied by a summary, no more than 150 words in length, giving a brief description of its subject matter and main conclusions.

Each article must also carry three jel (Journal of Economic Literature) classification codes. The JEL Classification Codes Guide is available online at www.aeaweb.org/jel/jel_class_system.php.

Papers should be no longer than 10,000 words, including the summary, notes and bibliography. Shorter papers will also be considered.

Articles should be sent by e-mail to *revista@cepal.org* in Word format. Papers in pdf format will not be accepted.

Style guide:

Titles should not be excessively long.

Footnotes

— It is recommended that footnotes be kept to a minimum.
— It is recommended that footnotes not be used to cite bibliographical references; such references should preferably be incorporated into the text.
— Footnotes should be numbered consecutively using superscript Arabic numerals.

Tables, figures and equations

— It is recommended that tables and figures be kept to a minimum, avoiding any redundancy with the text.
— Equations should be written using the Office equation editor “MathType”; they should not be inserted as “pictures”.
— Tables, figures and other elements should be inserted at the end of the text in the format in which they were designed; they should not be inserted as “pictures”. Figures in Excel should include the corresponding worksheets.
— The location of tables and figures in the body of the article should be indicated in the appropriate place as follows:

Insert figure 1

Insert table 1

— Tables and figures should include an explicit and complete reference to their sources.
— Tables should indicate the period covered at the end of the title, and should indicate the units in which the data are expressed in a subtitle (in italics and between brackets).
— The symbols referred to in the “Explanatory notes” which appear on the page preceding the table of contents should be taken into account in the preparation of tables and figures.
— Footnotes to tables and figures should be ordered consecutively using superscript lower-case letters.
— Figures should be prepared bearing in mind that they will be printed in black and white.

Acronyms and abbreviations

— Acronyms and abbreviations should not be used unless absolutely necessary, in which case the full name should be written out the first time it occurs in the article.

Bibliography

— Bibliographical references should be directly related to the content of the article and should not be excessively long.
— At the end of the article, under the title “Bibliography”, all the necessary information should be included accurately and in alphabetical order by author: name of author(s), year of publication, full name of article (if any) and publication (including any subtitle), city of publication, publisher and, in the case of a periodical, month of publication.

The editorial board of the *Review* reserves the right to make any necessary editorial changes in the articles, including their titles.

Authors will receive a one-year courtesy subscription to the *Review*, plus 30 offprints of their article in Spanish and 30 in English, at the time of publication in each language.

Publicaciones recientes de la CEPAL
ECLAC recent publications

www.cepal.org/publicaciones

Informes periódicos / *Annual reports*

También disponibles para años anteriores / *Issues for previous years also available*

- Estudio Económico de América Latina y el Caribe 2017, 218 p.
  Economic Survey of Latin America and the Caribbean 2017, 214 p.
- La Inversión Extranjera Directa en América Latina y el Caribe 2017, 206 p.
  Foreign Direct Investment in Latin America and the Caribbean 2016, 196 p.
- Anuario Estadístico de América Latina y el Caribe 2016 / *Statistical Yearbook for Latin America and the Caribbean 2016, 132 p.*
- Balance Preliminar de las Economías de América Latina y el Caribe 2016, 132 p.
  Preliminary Overview of the Economies of Latin America and the Caribbean 2016, 124 p.
- Panorama Social de América Latina 2016, 290 p.
  Social Panorama of Latin America 2016, 272 p.
- Perspectivas del Comercio Internacional de América Latina y el Caribe 2017. Documento informativo, 176 p.
  International Trade Outlook for Latin America and the Caribbean 2017. Briefing paper, 170 p.

Libros y documentos institucionales / *Institutional books and documents*

- Plan de acción regional para la implementación de la nueva agenda urbana en América Latina y el Caribe, 2016-2036, 2017, 64 p.
  Regional Action Plan for the implementation of the New Urban Agenda in Latin America and the Caribbean 2016-2036, 2017, 62 p.
- Brechas, ejes y desafíos en el vínculo entre lo social y lo productivo, 2017, 182 p.
  Linkages between the social and production spheres: Gaps, pillars and challenges, 2017, 170 p.
- Derechos de las personas mayores: retos para la interdependencia y autonomía, 2017, 176 p.
  Challenges to the autonomy and interdependent rights of older persons, 2017, 162 p.
- Panorama fiscal de América Latina y el Caribe 2017: la movilización de recursos para el financiamiento del desarrollo sostenible, 2017, 115 p.
  Fiscal Panorama of Latin America and the Caribbean 2017: Mobilizing resources to finance sustainable development, 2017, 108 p.
- La matriz de la desigualdad en América Latina, 2016, 96 p.
  The social inequality matrix in Latin America, 2016, 94 p.
- Autonomía de las mujeres e igualdad en la agenda de desarrollo sostenible, 2016, 184 p.
  Equality and women's autonomy in the sustainable development agenda, 2016, 168 p.
  Autonomia das mulheres e igualdade na agenda de desenvolvimento sustentável. Síntese, 2016, 106 p.
- La Unión Europea y América Latina y el Caribe ante la Agenda 2030 para el Desarrollo Sostenible: el gran impulso ambiental, 2016, 112 p.
  The European Union and Latin America and the Caribbean vis-à-vis the 2030 Agenda for Sustainable Development: The environmental big push, 2016, 112 p.
- Horizontes 2030: la igualdad en el centro del desarrollo sostenible, 2016, 176 p.
  Horizons 2030: Equality at the centre of sustainable development, 2016, 174 p.
  Horizontes 2030: a igualdade no centro do desenvolvimento sustentável, 2016, 176 p.
- 40 años de agenda regional de género, 2016, 130 p.
  40 years of the regional gender agenda, 2016,128 p.

Libros de la CEPAL / *ECLAC books*

- 150 ¿Quién cuida en la ciudad?: aportes para políticas urbanas de igualdad, María Nieves Rico, Olga Segovia (eds.), 2017, 476 p.
- 149 Manufactura y cambio estructural: aportes para pensar la política industrial en la Argentina, Martín Abeles, Mario Cimoli, Pablo José Lavarello (eds.), 2017, 336 p.
- 148 Planificación para el desarrollo en América Latina y el Caribe: enfoques, experiencias y perspectivas, Jorge Máttar, Luis Mauricio Cuervo (eds.), 2017, 336 p.
- 147 Inclusión financiera de pequeños productores rurales, Francisco G. Villarreal (ed.), 2017, 218 p.
  147 Financial inclusion of small rural producers, Francisco G. Villarreal (ed.), 201, 196 p.
- 146 Institucionalidad social en América Latina y el Caribe, Jorge Martínez (ed.), 2017, 334 p.
- 145 Política industrial rural y fortalecimiento de cadenas de valor, Ramón Padilla (ed.), 2017, 242 p.
  145 Rural industrial policy and strengthening value chains, Ramón Padilla (ed.), 2017, 236 p.

Páginas Selectas de la CEPAL / *ECLAC Select Pages*

- Empleo en América Latina y el Caribe. Textos seleccionados 2006-2017, Jürgen Weller (comp.), 2017, 446 p.
- Planificación y prospectiva para la construcción de futuro en América Latina y el Caribe. Textos seleccionados 2013-2016, Jorge Máttar y Mauricio Cuervo (comps.), 2016, 222 p.
- Desarrollo inclusivo en América Latina. Textos seleccionados 2009-2016, Ricardo Infante (comp.), 2016, 294 p.
- Globalización, integración y comercio inclusivo en América Latina. Textos seleccionados 2010-2014, Osvaldo Rosales (comp.), 2015, 326 p.

Copublicaciones / *Co-publications*

- El imperativo de la igualdad, Alicia Bárcena, Antonio Prado, CEPAL/Siglo Veintiuno, Argentina, 2016, 244 p.
- Gobernanza global y desarrollo: nuevos desafíos y prioridades de la cooperación internacional, José Antonio Ocampo (ed.), CEPAL/Siglo Veintiuno, Argentina, 2015, 286 p.
- Decentralization and Reform in Latin America: Improving Intergovernmental Relations, Giorgio Brosio and Juan Pablo Jiménez (eds.), ECLAC/Edward Elgar Publishing, United Kingdom, 2012, 450 p.
- Sentido de pertenencia en sociedades fragmentadas: América Latina desde una perspectiva global, Martín Hopenhayn y Ana Sojo (comps.), CEPAL/Siglo Veintiuno, Argentina, 2011, 350 p.

Coediciones / *Co-editions*

- Perspectivas económicas de América Latina 2017: Juventud, Competencias y Emprendimiento, 2016, 338 p.
  Latin American Economic Outlook 2017: Youth, Skills and Entrepreneurship, 2016, 314 p.
- Desarrollo e integración en América Latina, 2016, 314 p.
- Hacia un desarrollo inclusivo: el caso del Uruguay, 2016, 174 p.
- Perspectivas de la agricultura y del desarrollo rural en las Américas: una mirada hacia América Latina y el Caribe 2015-2016, CEPAL/FAO/IICA, 2015, 212 p.

Documentos de Proyectos / *Project Documents*

- El gran impulso ambiental en el sector de la energía, Andrés Arroyo Peláez, 2017, 62 p.
- La transversalización del enfoque de género en las políticas públicas frente al cambio climático en América Latina, Marina Casas Varez, 2017, 101 p.
- Financiamiento para el cambio climático en América Latina y el Caribe en 2015, Joseluis Samaniego y Heloísa Schneider, 2017, 76 p.
- El cambio tecnológico y el nuevo contexto del empleo: tendencias generales y en América Latina, Sebastian Krull, 2016, 48 p.
- Cambio climático, políticas públicas y demanda de energía y gasolinas en América Latina: un meta-análisis, Luis Miguel Galindo, Joseluis Samaniego, Jimy Ferrer, José Eduardo Alatorre, Orlando Reyes, 2016, 68 p.

Cuadernos estadísticos de la CEPAL

- 44 Las cuentas de los hogares y el bienestar en América Latina. Más allá del PIB, 2016
- 43 Estadísticas económicas de América Latina y el Caribe: Aspectos metodológicos y resultados del cambio de año base de 2005 a 2010

Series de la CEPAL / *ECLAC Series*

Asuntos de Género / Comercio Internacional / Desarrollo Productivo / Desarrollo Territorial / Estudios Estadísticos / Estudios y Perspectivas (Bogotá, Brasilia, Buenos Aires, México, Montevideo) / *Studies and Perspectives* (The Caribbean, Washington) / Financiamiento del Desarrollo / Gestión Pública / Informes y Estudios Especiales / Macroeconomía del Desarrollo / Medio Ambiente y Desarrollo / Población y Desarrollo / Política Fiscal / Políticas Sociales / Recursos Naturales e Infraestructura / Seminarios y Conferencias

Manuales de la CEPAL

- 5 Estimación de las erogaciones sociales a partir del sistema de cuentas nacionales: una propuesta para las funciones de educación, salud y protección social, María Paz Colinao, Federico Dorin, Rodrigo Martínez y Varinia Tromben, 2016, 63 p.
- 4 Territorio e igualdad: planificación del desarrollo con perspectiva de género, 2016, 84 p.
- 3 Manual de formación regional para la implementación de la resolución 1325 (2000) del Consejo de Seguridad de las Naciones Unidas relativa a las mujeres, la paz y la seguridad, María Cristina Benavente R., Marcela Donadio, Pamela Villalobos, 2016, 126 p.

Revista CEPAL / *CEPAL Review*

La Revista se inició en 1976, con el propósito de contribuir al examen de los problemas del desarrollo socioeconómico de la región. La *Revista CEPAL* se publica en español e inglés tres veces por año

CEPAL Review first appeared in 1976, its aim being to make a contribution to the study of the economic and social development problems of the region. CEPAL Review is published in Spanish and English versions three times a year

Observatorio demográfico / *Demographic Observatory*

Edición bilingüe (español e inglés) que proporciona información estadística actualizada, referente a estimaciones y proyecciones de población de los países de América Latina y el Caribe. Desde 2013 el *Observatorio* aparece una vez al año

Bilingual publication (Spanish and English) proving up-to-date estimates and projections of the populations of the Latin American and Caribbean countries. Since 2013, the Observatory appears once a year

Notas de población

Revista especializada que publica artículos e informes acerca de las investigaciones más recientes sobre la dinámica demográfica en la región. También incluye información sobre actividades científicas y profesionales en el campo de población. La revista se publica desde 1973 y aparece dos veces al año, en junio y diciembre

Specialized journal which publishes articles and reports on recent studies of demographic dynamics in the region. Also includes information on scientific and professional activities in the field of population. Published since 1973, the journal appears twice a year in June and December

Las publicaciones de la CEPAL están disponibles en:
ECLAC publications are available at:

www.cepal.org/publicaciones

También se pueden adquirir a través de:
They can also be ordered through:

www.un.org/publications

United Nations Publications
PO Box 960
Herndon, VA 20172
USA

Tel. (1-888)254-4286
Fax (1-800)338-4550
Contacto / *Contact*: publications@un.org
Pedidos / *Orders*: order@un.org

REVIEW

www.eclac.org/revista

United Nations

ECONOMIC COMMISSION FOR LATIN AMERICA AND THE CARIBBEAN

ECLAC

United Nations Publication • S.17-00682 • December 2017 • ISSN 0251-2920

ISBN 978-92-1-121972-2