DEPARTAMENTO DE ECONOMIA
PUC-RIO

TEXTO PARA DISCUSSÃO
N°. 452

BEYOND OAXACA-BLINDER: ACCOUNTING FOR DIFFERENCES IN HOUSEHOLD INCOME DISTRIBUTIONS ACROSS COUNTRIES

FRANÇOIS BOURGUIGNON
FRANCISCO H. G. FERREIRA
PHILLIPPE G. LEITE

MARÇO 2002

Beyond Oaxaca-Blinder: Accounting for Differences in Household Income Distributions Across Countries

François Bourguignon, Francisco H. G. Ferreira and Phillippe G. Leite[1]

Keywords: Inequality, Distribution, Micro-simulations

JEL Classification Codes: C15, D31, I31, J13, J22

Abstract: This paper develops a micro-econometric method to account for differences across distributions of household income. Going beyond the determination of earnings in labor markets, we also estimate statistical models for occupational choice and for the conditional distributions of education, fertility and non-labor incomes. We import combinations of estimated parameters from these models to simulate counterfactual income distributions. This allows us to decompose differences between functionals of two income distributions (such as inequality or poverty measures) into shares due to differences in the structure of labor market returns (price effects); differences in the occupational structure; and differences in the underlying distribution of assets (endowment effects). We apply the method to the differences between the Brazilian income distribution and those of the United States and Mexico, and find that most of Brazil's excess income inequality is due to underlying inequalities in the distribution of two key endowments: access to education and to sources of non-labor income, mainly pensions.

---

[1] Bourguignon is with DELTA, Paris, and the World Bank. Ferreira and Leite are at the Department of Economics of the Pontifícia Universidade Católica do Rio de Janeiro. We thank David Lam, Dean Jolliffe and seminar participants at PUC-Rio, IBMEC-Rio, the University of Michigan, the World Bank and DELTA for helpful comments; and Nora Lustig and Cesar Bouillon at the IDB for making the Mexican data available to us, ready to use. The opinions expressed here are those of the authors and do not necessarily reflect those of the World Bank, its Executive Directors or the countries they represent.

1. *Introduction*

The distribution of personal welfare varies enormously across countries. The Gini coefficient for the distribution of household per capita incomes, for instance, ranges from 0.20 in the Slovak Republic to 0.63 in Sierra Leone (World Bank, 2002) and similar (or greater) international variation can be found for any alternative measure of inequality. Given that inequality levels *within* countries are generally rather stable, one would think that there ought to be considerable interest in understanding why income distributions vary so much *across* countries. Is it because the underlying distributions of wealth differ greatly, perhaps due to historical reasons? Or is it because returns to education are higher in one country than in the other? What is the role of differences in labor market institutions? Do different fertility rates and family structures play a role? And if, as is likely, differences in income distributions reflect all of these (and possibly other) factors, in what manner and to what extent does each one contribute?

Yet, applied research on differences across income distribution has not been as abundant as one might expect.[2] Increasingly, this seems to have less to do with lack of data and more to do with inadequate methodological tools. Through initiatives like the Luxembourg Income Study, the WIDER International Income Distribution Dataset and others, the availability of high-quality household-level data is growing. Methodologically, however, those seeking an understanding of why distributions are so different - and reluctant to rely exclusively on cross-country regressions with inequality measures as dependent variables - have often resorted to comparing Theil decompositions across countries.[3] We will argue below that, while these can be informative, their ability to shed light on determinants of differences across distributions is inherently limited.

[2] Theoretical models of why income distributions might differ across countries have been more abundant. Banerjee and Newman (1993) and Bénabou (2000) are two well-known examples. See Aghion et. al. (1999) for a survey.
[3] Theil decompositions are known more formally as decompositions of Generalized Entropy inequality measures by population subgroups. They were developed independently by Bourguignon (1979), Cowell (1980) and Shorrocks (1980).

Meanwhile, substantial progress has been made in our ability to understand differences in wage (or earnings) distributions. Some of this work, such as Almeida dos Reis and Paes de Barros (1991), Juhn, Murphy and Pierce (1993), Blau and Khan (1996) and Machado and Mata (2001), draws on variants of a decomposition technique based on simulating counterfactual distributions by combining data on individual characteristics (X) from one distribution, with estimated parameters (β) from another, which is due originally to Oaxaca (1973) and Blinder (1973).[4] Another strand, which includes DiNardo, Fortin and Lemieux (1996) and Donald, Green and Paarsch (2000), is based on alternative semi-parametric approaches. DiNardo et.al. (1996) use weighted kernel density estimators - instead of regression coefficients - to generate counterfactual density functions that combine population attributes (or labor market institutions) from one period, with the structure of returns from another. Donald et. al. (2000) adapt hazard-function estimators from the spell-duration literature to develop density-function estimators, and use these to construct counterfactual density and distribution functions (comparing the US and Canada).[5]

These approaches have been very fruitful, but they have not yet been generalized from wage distributions to those of household incomes, largely because the latter involve some additional complexities. The distribution of wages is defined over those currently employed. Taking the characteristics of these workers as given, earnings determination can be reasonably well understood by estimating returns to those characteristics in the labor market, through a Mincerian earnings equation: $y_i = X_i \boldsymbol{b} + \boldsymbol{e}_i$. Most of the aforementioned recent literature on differences in wage inequality is based on simulating counterfactual distributions on the basis of equations such as this, and many further restrict their samples

[4] Some of these studies, like Juhn, Murphy and Pierce (1993) and Machado and Mata (2001) decompose changes in the wage distribution of a single country, over time. Others, like Almeida dos Reis and Paes de Barros (for metropolitan areas within Brazil) and Blau and Khan (for ten industrialized countries) decompose differences across wage distributions for different spatial units. For a less well known but also pioneering work, see Langoni (1973).

[5] The distinction between "parametric" and "semi-parametric" methods is not terribly sharp. DiNardo et. al. (1996) use a probit model to estimate one of their conditional reweighing functions. Donald et. al. (2000) rely entirely on maximum likelihood estimates of parameters in a proportional-hazards model, and what is non-parametric about their method is a fine double-partitioning of the income space, allowing for considerable flexibility in both the estimation of the baseline hazard function, and in the manner in which it is shifted by the proportional-hazards estimates. Conversely, in the current paper, which follows a predominantly parametric route, some non-parametric reweighing of joint distribution functions is also used (see below). These techniques are often more complementary than substitutable.

to include prime-age, full-time male workers only. In addition, some authors are quite clear that they are interested in wages primarily as indicators of the price of labor, rather than as measures of welfare.

Naturally, the distribution of household incomes also depends on the returns and characteristics of its employed members, and will thus draw on earnings models too. But it also depends on their participation and occupational choices and on decisions concerning the size and composition of the family. In addition, changes in some personal characteristics, such as education, affect household incomes through more than one channel. Suppose we ask what the effect of "importing" the US distribution of education to Mexico is on the Mexican distributions of earnings and incomes. Whereas for earnings it might very well suffice to replace the relevant vector of X with US values, the distribution of household incomes will also be affected through changes in participation and fertility behavior. This greater complexity of the determinants of household income distributions seems to have prevented counterfactual simulation techniques from being applied to them, thus depriving those interested in understanding cross-country differences in the distribution of welfare from the powerful insights they can deliver.

Nevertheless, a more general version of the Oaxaca-Blinder idea – of simulating counterfactual distributions on the basis of combining models estimated for different real distributions - can fruitfully be applied to household incomes. What is required is an expansion of the set of models to be estimated, to include labor market participation, fertility behavior and educational choices. In this paper, we first propose a general statement of statistical decompositions applied to household income distributions; and then suggest a specific model of household income determination that enables us to implement the decomposition empirically. In particular, we investigate the comparative roles of three factors: the distribution of population characteristics (or endowments); the structure of returns to these endowments, and the occupational structure of the population. We apply the method to an understanding of the differences between the income distributions in Brazil, Mexico and the US. [6]

The paper is organized as follows. Section 2 summarizes what can be learned from conventional comparisons of income distributions across these three countries, and presents an empirical motivation. Section 3 contains a general statement of statistical decomposition analysis, which encompasses all variants currently in use as special cases. Section 4 proposes a specific model of household income determination and describes the estimation and simulation procedures needed for the decomposition. The results obtained in the case of the Brazil-US comparison are discussed in some detail in Section 5. Section 6 discusses the Brazil-Mexico comparison and Section 7 concludes.

2. Income Distribution in Brazil, Mexico and the United States.

This section compares the distributions of household income in the three most populous countries in the Western Hemisphere.[7] The comparisons are based on an analysis of the original household-level data sets: the Pesquisa Nacional por Amostra de Domicílios (PNAD) 1999 is used for Brazil; the Encuesta Nacional de Ingresos y Gastos de Hogares (ENIGH) 1994 for Mexico; and the Annual Demographic Survey in the March Supplement to the Current Population Survey (CPS) 2000, for the United States. As always with the March Supplement of the CPS, total personal income data refers to the preceding calendar year:1999. Sample sizes for each data set (actually used) are as follows: the CPS 2000 contained 50,982 households (133,649 individuals); the ENIGH 1994 contained 6,614 households (29,149 individuals); and the PNAD 1999 contained 80,972 households (294,244 individuals).

[6] This approach is a cross-country extension of a methodology previously developed to analyze the dynamics of the distribution of income within a single country. See Bourguignon, Ferreira and Lustig (1998).
[7] Our emphasis here is purely comparative. We make no attempt to present a detailed analysis of inequality or poverty in each of these countries. There is a large literature on these topics for each of our three countries, but see Henriques (2000) for a recent compilation of work on Brazil, and Székely (1998) on Mexico. For earlier studies comparing the Brazilian and US *earnings* distributions, see Lam and Levison (1992) and Sacconato and Menezes-Filho (2001).

We use income, rather than consumption, data because the decompositions described in the remainder of the paper rely in part on the determination of earnings.[8] In Brazil and Mexico, the income variable used was monthly total household income per capita, available in the surveys as a constructed variable from the disaggregated income questionnaire. In the US, the variable used was the sum (across individuals in the household) of annual total personal income and other incomes, excluding disability benefits, educational assistance and child support, divided by 12.[9] All three income definitions are before tax, but include transfers. While total annual incomes are not top-coded in the CPS, some of their components might be. The US Census Bureau warns that weekly earnings, in particular, are "subject to top-coding at U$1923", so as to censor the distribution of annual earnings from the main job at U$100,000. Inspection of our sample revealed, however, that 2.1% (2.5%) of observations had reported weekly (annual) earnings above those value. The maximum reported weekly value was U$2884. We therefore did not correct for top-coding in the US. Incomes are not top-coded in Brazil or Mexico either.

As usual, there are reasons to suspect that incomes may be measured with some error. In the case of Brazil, the problem is particularly severe in rural areas, to the extent that the usefulness of any estimate based on rural income data is thrown into doubt.[10] For this reason, we prefer to confine our attention to urban areas only, in Brazil and Mexico.[11] Care is taken to ensure that the distributions used are as comparable as possible, and this requires that we work with data unadjusted for misreporting, imputed rents, or for regional price level differences within countries. [12]

---

[8] And also because consumption data for Brazil is either very old (ENDEF, 1975) or incomplete in geographical coverage (POF, 1996; PPV, 1996).

[9] These income sources were excluded from the analysis because non-retirement public transfers are proportionately much more important in the US than in Brazil or Mexico, and their allocation follows rules which are not modelled in our approach. When they were included, the residual term of the decomposition was slightly larger, but all of our conclusions remained qualitatively valid.

[10] For evidence on the weaknesses of income data for rural Brazil, see Ferreira, Lanjouw and Neri (2000) and Elbers, Lanjouw, Lanjouw and Leite (2001).

[11] For the US, since the CPS does not disaggregate non-metropolitan areas into urban and rural, and the former dominate, we included both metropolitan and non-metropolitan areas.

[12] All three datasets are well-known in their respective countries. For more detailed information about the CPS, go to www.census.gov. Information on the PNAD is available from www.ibge.gov.br. Information on the ENIGH is available from http://www.inegi.gob.mx/.

Table 1 below reports some key summary statistics of the income distributions for our three countries. In addition to population, GDP per capita and mean income from the household survey, three inequality measures are computed: the Gini Coefficient, the Theil T and L indices – in what follows, the last two are sometimes labeled E(1) and E(0), respectively, as members of the class of generalized entropy inequality measures. Each of these statistics is presented for the distribution of household income per capita, as well as for a distribution of equivalised incomes, where the Buhmann et. al. (θ = 0.5) equivalence scale is used. [13] All households are weighted by the number of individuals they comprise.

Table 1: Descriptive Statistics

Country	Population (millions, 1999)	GDP per capita (monthly, USD)	Mean equivalised income (monthly, USD)	Gini Coefficient	Theil-T	Theil-L
$\theta = 1.0$						
Brazil	168	526.42	290.34	0.587	0.693	0.646
Mexico	97	643.25	280.90	0.536	0.580	0.511
USA	273	2550.00	1691.64	0.445	0.349	0.391
$\theta = 0.5$						
Brazil	168	526.42	551.08	0.560	0.613	0.572
Mexico	97	643.25	587.91	0.493	0.478	0.423
USA	273	2550.00	2791.78	0.415	0.298	0.344

Notes: Population and GDP per capita figures are from World Bank (2001). The other figures are from calculations by the authors from the household surveys. GDP per capita and mean equivalised income (MEY) are monthly and measured in 1999 US dollars at PPP exchange rates. Mexican survey data is for 1994; Brazilian survey data is for 1999, and US survey data is for 2000. Values of θ are for the economy of scale parameter in the Buhmann et.al. (1988) equivalence scale - θ = 1 corresponds to income per capita.

Similarities between Brazil (in 1999) and Mexico (in 1994) are immediately apparent. Across those different years, the two countries had broadly similar levels of GDP per capita. Mexico's was 22% higher than Brazil's , which pales in comparison to the difference between the two countries and the US: 384% higher than Brazil's. Brazil's inequality is ranked highest by all three measures reported, followed by Mexico and the United States. The difference between Brazil's and Mexico's Ginis, at approximately five points, is not too large, while there are a full fourteen points between Brazil and the US. It is interesting to note that the effect of allowing for (a good deal) of scale economies in household consumption differs across both countries and measures. Focusing on the Gini coefficient,

[13] According to that method, the equivalised income of a household with income y and size N is taken to be y/N^{θ}. This definition coincides with income per capita when θ=1.

the reduction in inequality in Mexico from reducing θ from 1.0 to 0.5 is larger than either in the US or Brazil.

The considerable differences in both mean incomes and inequality across these three countries must translate into different poverty levels as well. Table 2 below presents the three standard FGT[14] poverty measures for each country, based on the distribution of per capita household incomes. The first panel shows poverty rates for the entire countries, whereas the second panel shows them for urban areas only, which is the universe for the analysis carried out in the next sections of the paper. In both cases, we use two alternative poverty thresholds. The first block in each panel employs an absolute poverty line, originally calculated as a strict indigence line for Brazil by Ferreira, Lanjouw and Neri (2000). Translated to 1999 values, it was set at R$74.48, or US$83.69 at PPP exchange rates. Having the lowest mean and the highest inequality of the three countries, Brazil has the most poverty by all three measures, in urban areas and overall. The United States has, by this ungenerous developing country standards, only traces of poverty. As for Mexico, it is striking how much of its poverty is rural: poverty incidence falls from 23% nationally, to less than 7% in urban areas. While being mindful that urban-rural definitions vary across countries, it would seem that poverty has an even more predominantly rural profile in Mexico than in Brazil.

But when one considers welfare across countries at such different levels of development and per capita income as these three countries, a strong argument can be made that a relative poverty concept might be more appropriate. For this reason we also present the same poverty measures, in the same distributions, calculated with respect to a line set at half the median income in each distribution, in the second block of each panel. By these more relative standards, poverty in the US reaches a full quarter of the population, which happens to be quite similar to Brazil's urban incidence. Mexico's P(0) also rises to 15% in urban areas.

FGT(α) measures for Urban and Rural areas

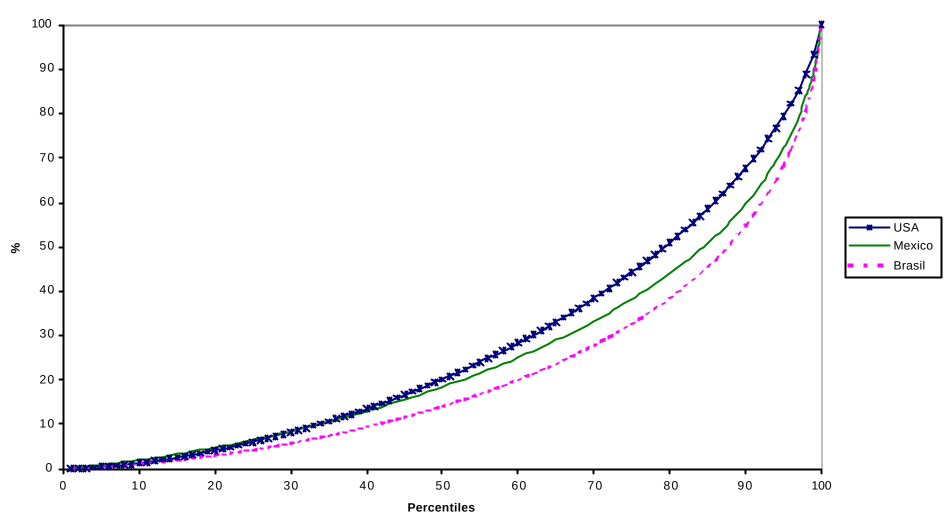
	P(0)	P(1)	P(2)	***Poverty line***[1]
Brazil	29,18	12,10	6,74	83,69
Mexico	23,29	8,02	3,84	83,69
USA	1,41	0,75	0,54	83,69
Brazil	30,02	12,22	6,82	84,27
Mexico	17,86	5,59	2,57	70,11
USA	25,02	10,19	5,92	687,70

FGT(α) measures for Urban areas

	P(0)	P(1)	P(2)	***Poverty line***[1]
Brazil	22,33	8,40	4,37	83,69
Mexico	6,66	1,52	0,51	83,69
Brazil	26,74	10,42	5,55	95,51
Mexico	14,98	3,73	1,39	110,46

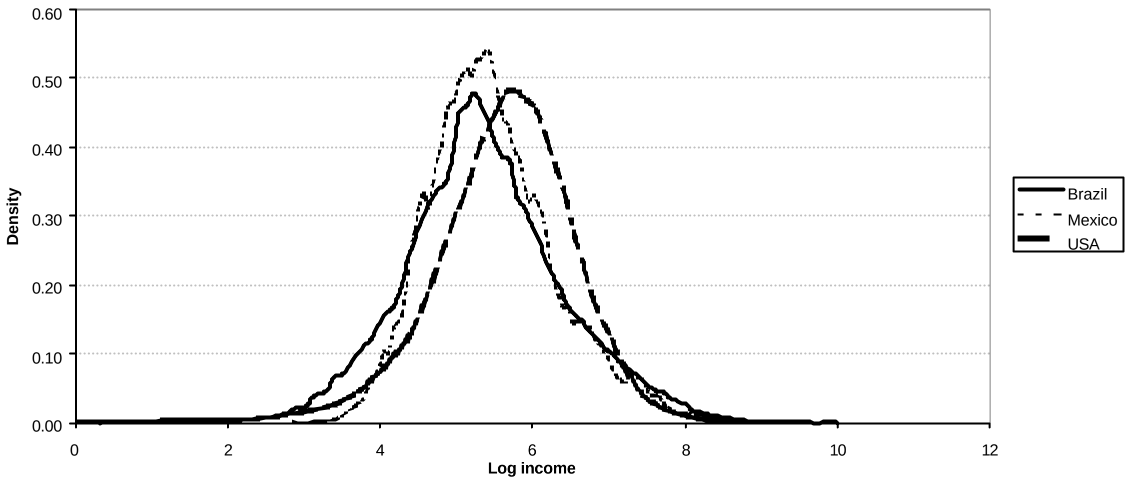
Figure 1, which contains the Lorenz curves for the urban household income distributions for Brazil, Mexico and the US, is a useful complement to the indices presented so far. Brazil is Lorenz dominated by both Mexico and the United States, whereas those two countries, at least with only urban Mexico being considered, can not be Lorenz ranked. The Atkinson Theorem (1970) – which establishes the link between normalized second-order stochastic dominance and unambiguous inequality ranking - makes Lorenz Curves very useful diagrammatic tools to compare income distributions. Nevertheless, because they are two levels of integration above a density function, we can do even better in terms of picturing the distribution. Figure 2 below plots kernel estimates of the (mean normalized) density functions for the distribution of (the logarithm of) household per capita income in our three countries. The greater dispersion of the Brazilian distribution is noticeable with respect to the Mexican, as is the greater skewness of the Brazilian and Mexican distributions, vis-à-vis that of the United States.

Figure 1: Urban Lorenz Curve For Brazil, Mexico and the US.



[14] Foster, Greer and Thorbecke (1984). In what follows, we use the three common measures of that family of

Figure 2: Income Distributions for Brazil, Mexico and The United States



Sources: PNAD/IBGE 1999, CPS/ADS 2000
Note: Gaussian Kernel Estimates (with optimal window width) of the density functions for the distributions of the logarithms of household per capita incomes. The distribution were scaled so as to have the brazilian mean. Brazil and Mexico are urban areas only. Incomes were converted to US dollar at PPP exchange rates (see Appendix).

Finally, Table 3 reports on standard decompositions of E(0), E(1) and E(2) by population subgroups[15], computing the R_B statistic developed by Cowell and Jenkins (1995). This statistic is an indicator of the relative importance of each attribute used to partition the population, in the process of "accounting for" the inequality. The idea is that the larger the share of dispersion which is between groups defined by some attribute - rather than within those groups - the more likely it is that something about the distribution of or returns to that attribute are causally related to the observed inequality. The attributes to be used include education of the household head (or main earner for the distribution of household incomes); his or her age; his or her race or ethnic group; his or her gender; as well as the location of the household (both regional and rural/urban) and its size or type.

The results are suggestive. In Brazil, education of the head is clearly the most important partitioning characteristic, followed by race and family type. In the US, family type dominates, with education a surprisingly low second, and age of head third. In Mexico, education and urban/rural vie for first place, with family type third. It is clear that education accounts for more inequality in Brazil (and Mexico) than in the US, although this technique can not tell us whether this is due predominantly to different returns or different

poverty indices : P(0), the headcount, P(1), the poverty gap and P(2), the cumulated squared gap.

endowments of education – i.e. a different distribution of the population across educational levels. The greater role of the urban/rural partition in Mexico is in line with our findings regarding total and urban poverty rates there. Strikingly little of overall US inequality is between different regions of the country, reinforcing the widespread perception of a well-integrated economy. This is in contrast to the two Latin American countries, where some 10% of the Theil-L is accounted for by the regional partition.[16] Finally, it is interesting to note that inequality between households headed by people of different races - which one would expect to be prominent in the US - is five to six times as large in Brazil.

Table 3: Theil Decompositions of Inequality by Population Characteristics

	Brasil			USA			Mexico		
	RB(0)	RB(1)	RB(2)	RB(0)	RB(1)	RB(2)	RB(0)	RB(1)	RB(2)
Region	0,092	0,076	0,031	0,003	0,004	0,003	0,113	0,103	0,050
Household Type	0,126	0,121	0,060	0,192	0,210	0,155	0,194	0,180	0,092
Urban / Rural	0,101	0,073	0,026	-	-	-	0,253	0,194	0,079
Gender of the Head	0,000	0,000	0,000	0,002	0,002	0,002	0,000	0,000	0,000
Race of the Head	0,137	0,119	0,051	0,024	0,024	0,016	-	-	-
Education Level	0,266	0,316	0,213	0,129	0,133	0,093	0,247	0,255	0,150
Age Group	0,051	0,047	0,021	0,082	0,091	0,066	0,042	0,037	0,017

But although this is a useful preliminary exercise, there are at least three reasons why one would wish to go further. First, none of these decompositions control for any of the others: some of the inequality between regions in Mexico is also between individuals with different races, and there is no way of telling how much. Second, the decompositions are of scalar measures, and therefore "waste" information on how the entire distributions differ (along their support). Although some information can be recovered from knowledge of the

[15] See Bourguignon (1979), Cowell (1980) and Shorrocks (1980).
[16] The regional breakdowns used in this decomposition were standard for each country. Brazil was divided up into five regions: North, Northeast, Centre-West, Southeast and South. Mexico was divided up into nine regions: "Noroeste", "Noreste", "Norte", "Centro Occidente", "Centro", "Sur", "Sureste", "Suroeste" and "Distrito Federal". The US was broken down into four regions: Northeast, Midwest, South and West. For a

different sensitivities of each measure, this is at best a hazardous and imprecise route. Finally, even to the extent that one is prepared to treat inequality between subgroups defined by age or education, say, as being driven by those attributes – rather than by correlates – the share of total inequality attributed to that partition tells us nothing of whether it is the distribution of the characteristic (or asset), or the structure of its returns that matters. In the next section, we propose an alternative approach, which suffers from none of these shortcomings.

3. A General Statement of Statistical Decomposition Analysis.

In order to understand the differences between two distributions of household incomes, $f^A(y)$ and $f^B(y)$, it seems natural to depart from the joint distributions $\varphi^C(y, T)$, where T is a vector of observed household characteristics, such as family size, the age, gender, race, education and occupation of each individual member of the household, etc.. The superscript C (= A, B) denotes the country. Because a number (but not all) of the characteristics in T clearly depend on others (e.g. family size, via the number of children, will vary with the age and education of the parents), it will prove helpful to partition T = [V, W] where, for any given household h in C, each element of V_h may be thought of as logically depending on W_h, and possibly on some other elements of V_h, but W_h is to be considered as fully exogenous to the household.

The distribution of household incomes, $f^C(y)$, is of course the marginal distribution of the joint distribution $\varphi^C(y, T)$: $f^C(y) = \iiint \varphi^C(y,T)dT$. It can therefore be rewritten as $f^C(y) = \iiint g^C(y|V,W)\phi^C(V,W)dVdW$, where $g^C(y \mid V,W)$ denotes the distribution of y conditional on V and W, and $\phi^C(V, W)$ is the joint distribution on all elements of T in country C. Given the distinction made above between the "semi-exogenous"[17] household characteristics V and the "truly exogenous" characteristics W, this can be further rewritten as:

much more detailed analysis of the importance of regional effects in Mexican inequality, see Legovini, Bouillon and Lustig (2000).

(1) $$f^{C}(y)=\iiint g^{C}(y|V,W)h_1^{C}(v_1|V_{-1},W)h_2^{C}(v_2|V_{-1,2},W)..h_u^{C}(v_u|W)\psi^{C}(W)dW$$

In (1), the joint distribution of all elements of T = [V,W] has been replaced by the product of υ conditional distributions and the joint distribution of all elements in W, $\psi^C(W)$. Each conditional distribution h_n is for an element of V, conditioning on the υ-n elements of V not yet conditioned on, and on W. The order n = {1,…υ) obviously does not matter for the product of the conditional distributions. (1) is an identity, invariant in that ordering. However, the order does matter for the definition of each individual conditional distribution $h_n(v_n|V_{-1,\ldots,n}, W)$, and therefore for the interpretation of each decomposition defined below.[18]

Once we have written the distributions of household incomes for countries C = A, B as in (1), one could investigate how $f^B(y)$ differs from $f^A(y)$ by replacing some of the observed conditional distributions in the ordered set $k^A = \{g^A, \mathbf{h^A}\}$ by the corresponding conditional distributions in the ordered set $k^B = \{g^B, \mathbf{h^B}\}$. Each such replacement generates a counterfactual (ordered) set of conditional distributions k^s, the dimension of which is $\upsilon+1$, (like k^A and k^B) whose elements are drawn either from k^A or k^B. It is now possible to define a counterfactual distribution $f^s_{A\to B}(y; k^s, \psi^A)$ as the marginal distribution that arises from the integration of the product of the conditional distributions in k^s and the joint distribution function $\psi^A(W)$, with respect to all elements of W. As an example, the counterfactual distribution $f^s_{A\to B}(y; g^A, h_1^B, \mathbf{h_{-1}^A}, \psi^A)$ is given by:

$f^{s}_{A\to B}(y)=\iiint g^{A}(y|V,W)h_1^{B}(v_1|V_{-1},W)h_2^{A}(v_2|V_{-1,2},W)..h_u^{A}(v_u|W)\psi^{A}(W)dW$. The number of possible such counterfactual distributions is the number of possible combinations of elements of the set k, i.e. the dimension of its sigma-algebra.[19]

---

[17] This terminology is motivated by the fact that we do not pretend that our models of V should be interpreted causally, and make no claims to be endogenizing these variables in a behavioural sense.

[18] Shorrocks (1999) proposes an algorithm based on the Shapley Value in order to calculate the correct "average" contribution of a particular $h_n(\ )$ or of g( ), over the set of possible orderings, to the overall difference across the distributions. Rather than constructing these values in this paper, we present our results by showing a number of different orderings explicitly in Sections 5 and 6 below.

[19] When we turn to the empirical implementation of these counterfactual distributions, we will see that is also possible, of course, to simulate replacing the joint distribution $\psi^A(y)$ by a non-parametric approximation of $\psi^B(y)$. Depending on how each specific conditional distribution is modelled, it is also possible to have more

For each counterfactual distribution, it is possible to decompose the observed difference in the income distributions for countries A and B as follows:

$$(2) \qquad f^B(y) - f^A(y) = \left[f^s(y) - f^A(y)\right] + \left[f^B(y) - f^s(y)\right]$$

where the first term on the right-hand side measures the "explanatory power" of decomposition s, and the second term measures the "residual" of decomposition s.[20] Since these are differences in densities, they can be evaluated for all values of y. Furthermore, any functional of a density function can be evaluated for f^A, f^B or f^s, and similarly decomposed, according to its own metric.

So, we have the same decomposition relationship as (2) for the cumulative distribution $F^C(y) = \int_0^y f^C(x)dx$. Likewise, for the mean income of quantile q: $m_q^C(y) = \frac{1}{Q} \int_{F_C^{-1}(q)}^{F_C^{-1}(q+1)} y f^C(y)dy$, we have:

$$(3) \qquad m_q^B(y) - m_q^A(y) = \left[m_q^s(y) - m_q^A(y)\right] + \left[m_q^B(y) - m_q^s(y)\right]$$

And we have analogous decompositions for any inequality measure I(f(y)) or poverty measure P(f(y); z). In the applications discussed in Sections 5 and 6, the results are presented exactly in this form: Tables 5 and 7 contain inequality and poverty measures, evaluated for $\hat{f}^A(y)$, $\hat{f}^B(y)$ and for a set of counterfactual distributions $\hat{f}^s(y)$, so that the reader can make his own subtractions. Figures 4-8 and 10-14 plot the differences in the (log) mean income of "hundredths" q $\in$ [1, 100], in a graphical representation of Equation (3). In recognition of their parentage, we call these the Generalized Oaxaca-Blinder decompositions.

---

than one counterfactual distribution per element of k. These matters pertain more properly to a discussion of the empirical application of the approach, however, and we return to them later.

[20] A decomposition is defined (by (2)) with respect to a unique counterfactual distribution s, and is thus also indexed by s.

4. *The Decompositions in Practice: A Specific Model*

The essence of the approach outlined above is to compare two actual income distributions, by means of a sequence of "intermediate" counterfactual distributions. These are constructed by replacing one or more of the underlying conditional distributions of A by those imported from B. In practice, this requires generating statistical approximations to the true conditional distributions. This may be done either through parametric models - following the tradition of Oaxaca (1973), Blinder (1973) and Almeida dos Reis and Paes de Barros (1991) - or through non-parametric techniques – as in DiNardo, Fortin and Lemieux (1996).[21] Because of the direct economic interpretations of the parameter estimates in our approximated distributions, we find it convenient in this paper to follow (mainly) the parametric route, by approximating each of the true conditional distributions through a set of standard econometric models, with pre-imposed functional forms.[22]

In particular, we will find it convenient to propose two (sets of) models:

(4) $y = G(V, W, \mathbf{e}; \Omega)$ and

(5) $V = H(W, \mathbf{h}; \Phi)$,

where Ω and Φ are sets of parameters and $\mathbf{e}$ and $\mathbf{h}$ stand for vectors of random variables, with $\varepsilon \perp \{V, W\}$, and $\eta \perp W$, by construction. G and H have pre-imposed functional forms. We can then write an approximation $f^*(y)$ to the true marginal distribution $f^C(y)$ in Equation (1) as:

$$(1') \qquad f^{*C}(y) = \int_{G(V,W,e;\Omega)=y} \boldsymbol{p}^{y}(\boldsymbol{e})d\boldsymbol{e}\left[\int_{H(W,h,\Phi)=V} \boldsymbol{p}^{v}(\boldsymbol{h})d\boldsymbol{h}\right]\Psi^{C}(W)dW$$

where $\pi^y(\varepsilon)$ is the joint probability distribution function of $\mathbf{e}$ and $\pi^v(\eta)$ is the joint probability distribution function of $\mathbf{h}$.

[21] Although, as noted earlier, these authors too rely on parametric approximations to some conditional distributions, such as the probit for the conditional distribution of union status on individual characteristics.

[22] This is an advantage of our approach vis-à-vis, for instance, the hazard-function estimators of Donald et. al. (2000), who "note that the estimates of the hazard function for wages, earnings or incomes are difficult to interpret" (p.616)

Just as an exact decomposition was defined by (2) for each true counterfactual distribution, we can now define the (actually operational) decomposition s in terms of the approximated distributions $f^*(y)$, as follows:

(2') $$f^B(y) - f^A(y) = \left[f^{*s}(y) - f^A(y)\right] + \left[f^B(y) - f^s(y)\right] + \left[f^s(y) - f^{*s}(y)\right].$$

Recall that a counterfactual distribution s is conceptually given by $f^s_{A\to B}(y;\ k^s,\ \psi^A)$, and is thus defined by (ψ^A and) the simulated sequence of conditional distributions k^s, which consists of some original distributions from A, and some imported from B. Analogously, an approximated distribution $f^{*s}_{A\to B}\left(y;\Omega^s,\Phi^s,\Psi^A\right)$ is defined with respect to (ψ^A and) the two sets of simulated parameters Ω^s and Φ^s, which consist of some original parameters from the models estimated for country A, and some imported from the models estimated for country B.

The last term in (2') gives the difference between the approximated and the true counterfactual distribution We therefore call it the approximation error and denote it by R_A. Clearly, how useful this decomposition methodology is in gauging differences between income distributions depends to some extent on the relative size of the approximation error. The applications in the next two sections illustrate that it can be surprisingly small.

Following from (1'), our statistical model of household incomes has three levels. The first corresponds to model G (V, W, $\mathbf{e}$; Ω), which seeks to approximate the conditional distribution of household incomes on observed characteristics: g(y| V,W). This level generates estimates for the parameter set Ω, which we associate with the structure of returns in the labor markets and with the determination of the occupational structure in the economy. The second level corresponds to model H (W, $\mathbf{h}$; Φ) which seeks to approximate the conditional distributions $h_n(v_n|V_{-1,\ldots,n}, W)$, for V ={number of children in the household (n_{ch}); years of schooling of individual i (E_{ih}); and total household non-labor income (y_{0h})} In the third level, we investigate the effects of replacing $\psi^A(W)$ with a (non-parametric) estimate of $\psi^B(W)$. This largely corresponds to the racial and demographic make-up of the population.

First-level model G (V, W, $\mathbf{e}$; Ω) is given by equations (6-8) below. Household incomes are an aggregation of individual earnings y_{hi}, and of additional, unearned income such as transfers or capital income, y_0. Per capita household income for household h is given by:

(6) $$y_h = \frac{1}{n_h}\left[\sum_{i=1}^{n_h}\sum_{j=1}^{J} \mathrm{I}_{hi}^{j} y_{hi}^{j} + y_0\right]$$

where I_{hi}^{j} is an indicator variable that takes the value 1 if individual i in household h participates in earning activity j, and 0 otherwise. The allocation of individuals across activities (i.e. labor force participation and the occupational structure of the economy) is modeled through a multinomial logit of the form:

(7) $$\Pr\{j=s\} = P^s(Z_{hi}, \boldsymbol{l}) = \frac{e^{Z_{hi} l_s}}{e^{Z_{hi} l_s} + \sum_{j \neq s} e^{Z_{hi} l_j}}$$

where $P^s(\ )$ is the probability of individual i in household h being in occupational category s, which could be: inactivity, formal employment in industry, informal employment in industry, formal employment in services or informal employment in services. Separate but identically specified models are estimated for males and females. The vector of characteristics $Z \subset T$ is given by Z = {1, age, age squared, education dummies, age interacted with education, race, and region for the individual in question; average endowments of age and education among adults in his or her household; numbers of adults and children in the household; whether the individual is the head or not; and if not whether the head is active}.

As is well known, the multinomial logit model may be interpreted as a utility-maximizing discrete choice model where the utility associated with choice j is given by $U_{hi}^{j} = Z_{hi}.\boldsymbol{l}_j + \boldsymbol{e}_{hi}^{Uj}$. The last term stands for unobserved choice determinants of individual i, and it is assumed to be distributed according to a double exponential law in the population. We prefer, however, not to insist on this utility-maximizing interpretation of the multi-logit and to treat it merely as a building block of the statistical model G, defined in equation (4).

Turning to the labor market determination of earnings, y_{hi}^{j} in (6) is assumed to be log-linear in α_j and $\mathbf{b}_j$, and the individual earnings equation is estimated separately for males and females, as follows:

(8) $$\log y_{hi}^{j} = \alpha_j + \mathbf{x}_{hi}\beta_j + \varepsilon_i$$

where $\mathbf{x} \subset T$ is given by $\mathbf{x}$ = {education dummies, age, age squared, age * education, and intercept dummies for region, race}. In the absence of specific information on experience, the education and age variables are the standard Becker - Mincer human capital terms. The racial and regional intercept dummies allow for a simple level effect of possible spatial segmentation of the labor markets, as well as for the possibility of racial discrimination. Earning activities are defined by sector and formality status. To simplify, it is assumed that earnings functions across activities also differ only through the intercepts, so that the sets of coefficients $\mathbf{b}_j$ are the same across activities ($\mathbf{b}_j = \mathbf{b}$). We interpret these β coefficients in the usual manner: as estimates of the labor market rates of return on the corresponding individual characteristics.

This first level of the methodology generates estimates for the set Ω, comprising occupational choice parameters λ, and (random) estimates of the residual terms ε_{hi}^{Us} [23], as well as for $\mathbf{a}_j$ and $\mathbf{b}$ and for the variance of the residual terms, $\sigma_{em}^2, \sigma_{ef}^2$.

In the second level of the model, H (W, η; Φ), we estimate the conditional distributions of V ={number of children in the household (n_{ch}); years of schooling of individual i (E_{ih}); and total household non-labor income (y_{0h})} on W = {number of adults in the household (n_{ah}), its regional location (r_h), individual age (A_{ih}), race (R_{ih}) and gender (g_{ih})}. This is done by imposing the functional form associated with the multinomial logit (such as the one in Equation 7) on both the conditional distribution of E_{ih} on W: ML_E ($E \mid A, R, r, g, n_{ah}$) and on

[23] For details on how the latter may be determined, see Bourguignon, Ferreira and Lustig (1998).

the conditional distribution of the number of children in the household on {E, W}: ML_C (n_{ch} | E, A, R, r, g, n_{ah}).

Unlike Equation (7), these models are estimated jointly for men and women. The educational choice multilogit ML_E has as choice categories 1-4; 5-6; 7-8; 9-12; and 13 and more years of schooling, with 0 as the omitted category. Estimation of this model generates estimates for the educational endowment parameters, $\mathbf{g}$. The demographic multilogit ML_C has as choice categories the number of children in the household: 1, 2, 3, 4 and 5 and more, with 0 as the omitted category. Estimation of this model generates estimates for the demographic endowment parameters, $\mathbf{y}$. Finally, the conditional distribution of total household non-labor incomes on {E, W} is modelled as a Tobit: T (y | E, A, R, r, g, n_{ah}).[24] Estimation of this model generates estimates for the non-human asset endowment parameters, $\mathbf{x}$. These three vectors constitute the set of parameters $\Phi=\{\mathbf{g}, \mathbf{y}, \mathbf{x}\}$.

After each of these reduced-form models has been estimated for two countries (Brazil and a comparator nation), the approximate decompositions in (2') can be carried out. Each decomposition is based on the construction of one approximated counterfactual distribution $f^{*s}_{A\to B}\left(y;\Omega^s,\Phi^s,\Psi^A\right)$, defined largely by which set of parameters in Ω^A and Φ^A is replaced by their counterparts in Ω^B and Φ^B. All of our results in the next two sections are presented in this manner. Tables 5 and 7, for example, list mean incomes, four inequality measures and three poverty measures for a set of approximated counterfactual distributions, denoted by the vectors of parameters which were replaced with their counterparts from B. Similarly, Figures 4-8 and 10-14 draw differences in log mean quantile incomes between actual and

[24] We also experimented with an alternative approximation for the conditional distribution of non-labor incomes. This was a (non-parametric) rank-preserving transformation of the observed distribution of y_0, conditional on earned incomes in each country. In practical terms, we ranked the two distributions by per capita household earned income $y_e = y_h - \frac{y_0}{n_h}$. If $p = F_B(y_e)$ was the rank of household with income y_e in country B, then we replaced y^B_{op} with the unearned income of the household with the same rank (by earned income) in country A, after normalizing by mean unearned incomes: $y^A_{op}\frac{m_B(y_0)}{m_A(y_0)}$. The results, which are available from the authors on request, were similar in direction and magnitude to those of the parametric exercise reported in the text.

approximated counterfactual distributions, where these are denoted by the vectors of parameters which were replaced with their counterparts from B to generate them.

As an example, consider line 4 of Table 5 (denoted "α, β, and σ^2"). It lists the mean income and the inequality and poverty measures calculated for the distribution obtained by replacing the Brazilian α and β in equation (8), with those estimated for the US; scaling up the variance of the residual terms ε_i by the ratio of the estimated variance in the US to that of Brazil; and then predicting values of y_{ih} for all individuals in the Brazilian income distribution, given their original characteristics (ψ^A). The density function defined over this vector of predicted incomes is $f^{*s}_{A\to B}\left(y;\Omega^s,\Phi^s,\Psi^A\right)$ for $\Omega^s=\left\{\boldsymbol{a}^B,\boldsymbol{b}^B,\boldsymbol{s}^{2B},\boldsymbol{l}^A,\boldsymbol{h}^A\right\}$ and $\Phi^s = \Phi^A$.

Whenever $\boldsymbol{l}^B\in\Omega^s$, individuals may be reallocated across occupations. This involves drawing counterfactual ε^U's from censored double exponential distributions with the relevant empirically observed variances.[25] The labor income ascribed to the individuals who change occupation (to a remunerated one) is the predicted value by equation (8), with the relevant vector of parameters, and with ε's drawn from a Normal distribution with mean zero and the relevant variance. And when $\Phi^s \neq \Phi^A$, so that the values of the years of schooling variable and/or the number of children in households may change, these changes are incorporated into the vector V, and counterfactual distributions are recomputed for the new (counterfactual) household characteristics. As the discussion in the next two sections will show, the interactions between these various simulations are often qualitatively and quantitatively important. The ability to shed light on them directly and the ease with which they can be interpreted are two of the main advantages of this methodology.

The third and final level of the model consists of altering the joint distribution of the truly exogenous household characteristics, $\psi^C(W)$. The set W is given by the age (A), race (R), gender (g) of each adult individual in the household, as well as by adult household size (n_{ah})

[25] The censoring of the distribution from which the unobserved choice determinants are drawn is designed to ensure that they are consistent with observed behaviour under the alternative vector λ. See Bourguignon, Ferreira and Lustig (1998) for details.

and the region where the household is located (r). Since these variables do not depend on other exogenous variables in the model, this estimation is carried out simply by re-calibrating the population by the weights corresponding to the joint distribution of these attributes in the target country.[26]

In practice, this is done by partitioning the two populations by the numbers of adults in the household. To remain manageable, the partition is in three groups: households with a single adult; households with two adults; and households with more than two adults. Each of these groups is then further partitioned by the race (whites and non-whites) and age category (six groups) of each adult.[27] The number of household in each of these subgroups can be denoted $M_{a,r}^{n,C}$, where a stands for the age category of the group, r for the race of the group, n for the number of adults in the household, and C for the country. If we are importing the structure from country A (population of households P^A) to country B (population of households P^B), we then simply re-scale the household weights in the sample for country B by the factor:

$$f_{a,r}^{n} = \left(\frac{M_{a,r}^{n,A}}{M_{a,r}^{n,B}}\right)\frac{P^B}{P^A} \tag{9}$$

Results for this final level of simulations are reported in Tables 5 and 7 under the letter ϕ.

5. The Brazil-US Comparison.

The decompositions described in the previous section were conducted for differences in distributions between Brazil in 1999 and the United States in 2000. The estimated coefficients for equations (7) and (8), as well as those for the multinomial logit models for the demographic and educational structures and the tobit model of the conditional distribution of non-labor incomes are included in Tables A1 – A5, in the Appendix. Table 4 – at the end of the paper - presents the results for importing the parameters from the US into Brazil, in terms of means and inequality measures for the individual earnings distributions,

---

[26] The spirit of this procedure is very much the same as in DiNardo et. al. (1996).

separately for men and women. Table 5 displays analogous results for household per capita incomes, and includes also three poverty measures.[28] Figures 4 to 8 present the full picture, by plotting differences in log incomes between the distributions simulated in various steps and the original distribution, for each percentile of the new distribution.[29]

Looking first at individual earnings, the observed differences between the Gini coefficients in Brazil and the US are nine points for men, and ten for women. Brazil's gender-specific earnings distributions have a Gini of 0.5, whereas those of the US are around 0.4. Roughly speaking, price effects (identified by simulating Brazilian earnings with the US á and â parameters) account for half of this difference. As we shall see, this is a much greater share than that which will hold for the distribution of household incomes per capita. Among the different price effects, the coefficient on the interaction of age and education stands out as making the largest difference.

Differences in participation behavior are unimportant in isolation. Importing the US participation parameters only contributes to reducing Brazilian earnings inequality when combined with importing US prices, as may be seen by comparing the rows α,β (viii) and the row λ,α,β. Educational and fertility choices are more important effects. The former raises educational endowments and hence both increases and upgrades the sectoral profile of labor supply. The latter leads to increased participation rates by women. This effect accounts for nearly all of the remaining four to five Gini points. As one would expect, demographic effects are particularly important for the female distribution, where, in combination with the effect of education, it reduces the Brazilian Gini by a full five points

[27] In the case of households with more than two adults, this is done for two adults only: the head and a randomly drawn other adult. In this manner, the group of single adult households is partitioned into 12 sub-groups, and the other two groups into 144 sub-groups each.

[28] In order for the poverty comparisons to make sense across two countries as different as the US and Brazil, the US **earnings** distributions were scaled down so as to have the Brazilian mean. This was done by appropriately adjusting the estimate for α^{US}, as can be seen from the means reported in Tables 4 and 5. Accordingly, counterfactual poverty measures are not reported for simulations which do not include an α estimate. The same procedure was used in Section 6, to rescale the Mexican earnings distributions to have the Brazilian means.

[29] Analogous figures for differences in log incomes by percentiles ranked by the original distribution – which show the re-rankings induced by each simulation - are available from the authors on request.

even before any changes are made to prices. Reweighing the purely exogenous endowments - including race - has no effect.

Table 5, which reports on the simulations for the distribution of household incomes per capita, can be read in an analogous way. The first two lines present inequality and poverty measures for the actual distributions of household per capita income by individuals in Brazil (in 1999) and the US (in 2000). In terms of the Gini coefficient, the gap we are trying to "explain" is substantial: it is twelve and a half points higher in Brazil than in the US. The difference is even larger when the entropy inequality measures E() are used.

Table 5 : Simulated Poverty and Inequality for Brazil in 1999, Using 2000 USA coefficients.

		Mean p/c Income	Inequality				Poverty Z =median/2 per month		
			Gini	E(0)	E(1)	E(2)	P(0)	P(1)	P(2)
1	Brasil	294,8	0,569	0,597	0,644	1,395	26,23	10,10	5,36
2	USA	294,8	0,445	0,391	0,349	0,485	25,02	10,19	5,92
3	α, β	294,9	0,516	0,486	0,515	1,049	20,32	7,53	3,92
4	α, β, σ^2	294,9	0,530	0,517	0,545	1,119	21,92	8,39	4,46
5	λ	277,9	0,579	0,632	0,653	1,313			
6	λ, α, β	255,4	0,535	0,536	0,542	1,022	28,06	11,58	6,46
7	$\lambda, \alpha, \beta, \sigma^2$	255,5	0,548	0,565	0,572	1,093	29,59	12,50	7,06
8	γ	454,0	0,505	0,489	0,460	0,719			
8a	γ, α, β	283,9	0,480	0,425	0,425	0,732	18,81	7,12	3,75
8b	$\gamma, \alpha, \beta, \sigma^2$	283,9	0,494	0,453	0,452	0,786	20,33	7,84	4,18
9	λ, γ	469,0	0,511	0,514	0,467	0,711			
10	$\lambda, \gamma, \alpha, \beta$	274,2	0,490	0,450	0,445	0,780	21,15	8,36	4,54
11	$\lambda, \gamma, \alpha, \beta, \sigma^2$	274,2	0,505	0,480	0,474	0,837	22,73	9,19	5,07
12	ψ	295,2	0,576	0,613	0,663	1,449			
13	ψ, γ	464,6	0,505	0,493	0,454	0,686			
14	$\psi, \gamma, \alpha, \beta$	287,1	0,486	0,437	0,434	0,746	19,31	7,31	3,85
15	$\psi, \gamma, \alpha, \beta, \sigma^2$	287,1	0,499	0,464	0,459	0,794	20,85	8,09	4,35
16	ψ, λ e γ	507,2	0,500	0,492	0,441	0,641			
17	$\psi, \lambda, \gamma, \alpha, \beta$	299,2	0,481	0,433	0,423	0,709	18,14	7,00	3,75
18	$\psi, \lambda, \gamma, \alpha, \beta, \sigma^2$	299,2	0,495	0,462	0,448	0,755	19,59	7,77	4,24
19	ξ	317,5	0,534	0,531	0,551	1,144	20,58	7,97	4,32
20	$\psi, \lambda, \gamma, \alpha, \beta, \sigma^2; \xi$	356,3	0,428	0,353	0,315	0,416	11,17	4,33	2,38
21	ϕ	404,7	0,585	0,637	0,683	1,496			
22	$\phi, \psi, \lambda, \gamma, \alpha, \beta, \sigma^2$	387,7	0,511	0,490	0,489	0,874	14,35	5,43	2,88
23	$\phi,\psi, \lambda, \gamma, \alpha, \beta, \sigma^2; \xi$	436,4	0,432	0,359	0,325	0,448	8,14	3,11	1,71

Source: PNAD 1999 and CPS March 2000

The first block of simulations suggests that differences in the structure of returns to observed personal characteristics in the labor market can account for some five of these thirteen points.[30] When one disaggregates by individual âs, it turns out that returns to education, conditionally on experience – as for individual earnings - play the crucial role.

Overall, it can thus be said that difference in returns to schooling and experience together explain approximately 40 per cent of the difference in inequality between Brazil and the US. The order of magnitude is practically the same with E(1) and E(2) but it is higher with E(0), suggesting that the problem is not only that returns to schooling are relatively higher at the top of the Brazilian schooling scale but also that they are relatively lower at the bottom. This is confirmed by the fact that importing US prices lowers poverty in Brazil, even though (relative) poverty is initially comparable in the two countries.

Importing the US variance of residuals goes in the opposite direction, contributing to an increase of almost 1.5 Gini points in Brazilian inequality.[31] Two candidate explanations suggest themselves: either there is greater heterogeneity amongst US workers along unobserved dimensions (such as ability) than among their Brazilian counterparts, or the US labor market is more efficient at observing and pricing these characteristics. This is an interesting question, which deserves further investigation. In the absence of additional information on, say, the variance of IQ test results or other measures of innate ability, orthogonal to education, we are inclined to favor the second interpretation. It may be that the lower labor market turnover and longer tenures that characterize the US labor market translate into a lessened degree of asymmetric information between workers and managers in that country, with a more accurate remuneration of endowments which are unobserved to researchers. We thus consider the $ó^2$ effect as a price effect, which dampens the overall contribution of price effects to some 3.5 to 4 points of the Gini.

[30] The relative importance of each effect varies across the four inequality measures presented, but the orders of magnitude are broadly the same, and the main story could be told from any of them. All are presented in Table 5, but we use the Gini for the discussion in the text.
[31] This result is in line with the earlier findings of Lam and Levinson (1992), who noted that the variance of residuals from earnings regressions such as these was considerably higher in the US than in Brazil.

The next block shows that importing the US occupational structure (l) by itself, has almost no impact on Brazilian inequality, but lowers average incomes and raises poverty. This is a consequence of the great differences in the distribution of education across the two countries, as revealed by Figure 3 below. Since education is negatively correlated with inactivity, and positively with employment in industry and with formality in the US, when we simulate participation behavior with US parameters but Brazilian levels of education, we withdraw a non-negligible number of people from the labor force, and 'downgrade' many others. Figure 5 shows the impoverishing effect of imposing US occupational choice behavior, combined with its price effect, on Brazil's original distribution of endowments.

Turning to the second-level model, $H(W, \eta, \Phi)$, we see further support for the aforementioned role of education in determining occupational choice. When US educational parameters are imported by themselves, this raises education levels in Brazil substantially, thus significantly increasing incomes and reducing poverty. Education endowments increase more for the poor (as expected by the upper-bounded nature of the education distribution), and inequality also falls dramatically. The ã simulation alone takes six points of the Gini off the Brazilian coefficient and, crucially, takes the impoverishing effect away from the occupational structure simulation. The latter result suggests that the most important difference in the distribution of educational endowments between Brazil and the US might actually be in the lack of minimum compulsory level in Brazil – see figure 3.

At this stage, it might seem that almost all of the difference in inequality between the US and Brazil is explained by education-related factors. Six points of the Gini are explained by the differences in the distribution of education and five points by the difference in the structure of earnings by educational level (that is, the coefficients of the earning functions). Yet, when these changes - i.e. α, β and γ - are simulated together, as in row 8a in table 5, it turns out that their overall effect is not the sum of the two effects (eleven points), but only eight points. The two education-related effects, distribution and earnings structure, are therefore far from being additive. The same is true of the decomposition of earnings inequality in Table 6.

The explanation for this non-additivity property is straight-forward. As can be seen in figure 3, only a tiny minority of US citizens have fewer than 9 years of education, whereas practically 60% of the Brazilian population do. At the same time, the structure of US earnings for the few people below that minimum level of schooling is approximately flat, possibly because of minimum wage laws. In Brazil, on the contrary, earnings are strongly differentiated over that range. People with less than full primary education earn on average 70% of the mean earnings of people with some secondary education.[32] This proportion is 95 per cent for the few people with such a low level of schooling in the US. Thus, importing the earnings structure from the US to Brazil contributes to a drastic equalization of the distribution when the demographic structure of education of Brazil remains unchanged. Many people with less than secondary education are then paid at practically the same rate as people with completed secondary.

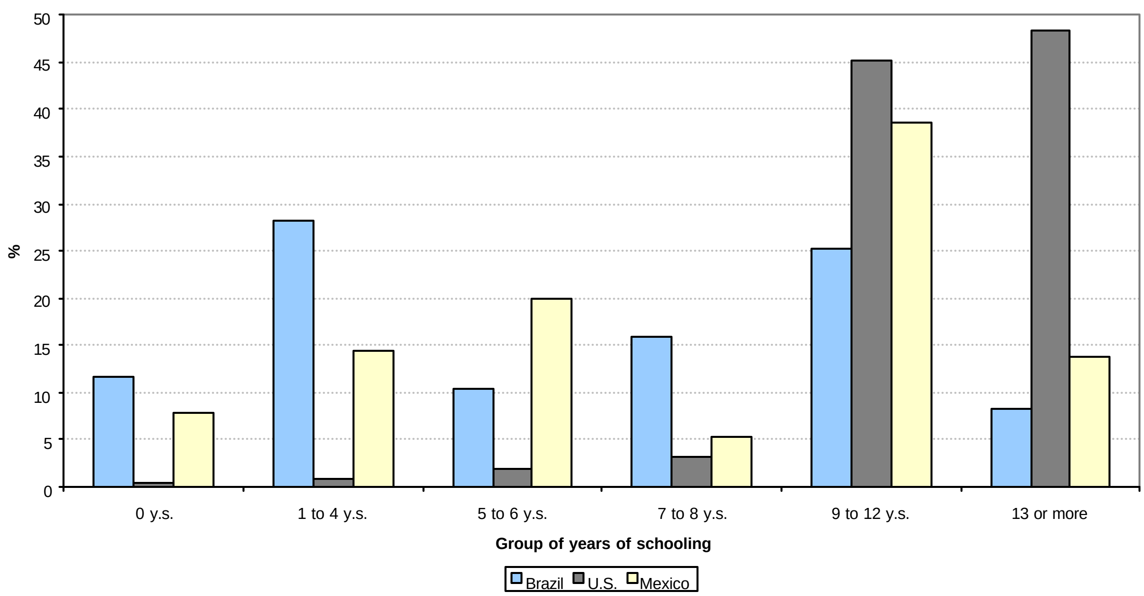
Doing the same exercise with the US demographic structure of education has much less effect, because there are very few people in that country with less than secondary. This appears clearly in table 5 when comparing rows 1 and 3 on the one hand, and rows 8 and 8a on the other. The basic effect of switching to US earnings when the US demographic educational structure is used comes from the fact that the relative earnings of college versus high school graduates is substantially higher in Brazil.

The question which remains is: how much of the excess inequality in Brazil with respect to the US is due to the distribution of education, and how much is due to the structure of schooling returns.[33] The foregoing argument makes it tempting to place greater weight on the distribution of education effect. This is because the structure of educational returns at low schooling levels is relevant to very few people in the US, and yet it has such an important effect when imported to Brazil. One may also hold that the structure of returns actually reflects the educational profile of both populations. There are positive returns at the

[32] These figures refer to mean earnings by educational level and differ from what may be inferred from the regression coefficients for schooling in table A2.
[33] This is not a new question. In fact, it was at the heart of the public debate about the causes of increasing inequality in Brazil during the 1960s. See Fishlow (1972) and Langoni (1973) for different views on the matter at that time.

bottom in Brazil because many people in the labor force have zero or a very low level of schooling, whereas this is exceptional in the US. There are also larger returns in Brazil at the top of the schooling range because there are relatively fewer people with a college education.

Figure 3: Distribution of education across the countries



Sources: PNAD/IBGE 1999, CPS/ADS 2000, ENIGH 1994

Moving on to demographic behavior, we observe a similar role for education. As with occupational structure, importing ø alone hardly changes inequality – it would even increase it slightly. However, fertility is negatively correlated with educational attainment, particularly of women. If the change in fertility were taking place in the Brazilian population with US levels of schooling and participation behavior, inequality would drop by 1 percentage point of the Gini coefficient and poverty would fall. This seems to mean that fertility behavior differs between the two countries mostly for lowest educated households.

When the effects of some of the “semi-exogenous” endowments (embodied in the approximations to the educational and demographic counterfactual conditional distributions) are combined with occupational structure and price effects (as in the row for ą, ë, ã á, â, ó^2), we see an overall reduction of seven points in the Gini. Most of this (around five points) seems to be associated with adopting the US endowments of education, either directly or indirectly, through knock-on effects on participation and fertility. The remainder is due to the price effects.[34] This still leaves, however, some additional five Gini points - a rather substantial amount - in the difference in inequality between the two countries unexplained. Figure 7 illustrates the results of the combined simulations for the entire distribution: while the simulated line has moved much closer to the actual (log income percentile) differences, it is not yet a very good fit.

Of the various candidate factors we are considering, two remain: the differences in the joint distributions of exogenous observed personal endowments: $\psi^A(W)$ and $\psi^B(W)$; and non-labor incomes. The two final blocks of simulations show that it is the latter, rather than the former, that accounts for the remaining inequality differences. While reweighing the households in accordance with Equation (9) actually has an increasing effect on Brazilian inequality (see line 21) - thus weakening the explanatory power of the overall simulation by about one and half Gini points - importing the conditional distribution of non-labor incomes has a surprisingly large explanatory power. As may be seen from line 20 of Table 5, it actually moves the simulated Gini coefficient for Brazil to within 1.7 Gini point of the true US Gini.

When reweighing the joint distributions of exogenous observed personal endowments is combined with all the previous steps, in line 23, the difference is further reduced to 1.3 Gini points It also does remarkably well by all other inequality measures in Table 5. Figure 8 shows the simulated income differences for two different counterfactual distributions with non-labor incomes - one with and the other without reweighing. The fit with regard to the actual differences is clearly much improved with respect to the preceding simulations, and it is evident that reweighing the exogenous endowments has a limited effect. The fact that the curve for simulated income differences now lies much nearer the actual differences

[34] This allocation of the various effects is made difficult by the fact that their size depends on the other effects already being accounted for. The figures mentioned here are obtained as averages over the various possible configurations appearing in table 5.

curve graphically illustrates the success of the simulated decomposition. This suggests that the approximation error R_A is very small, at least in this application.

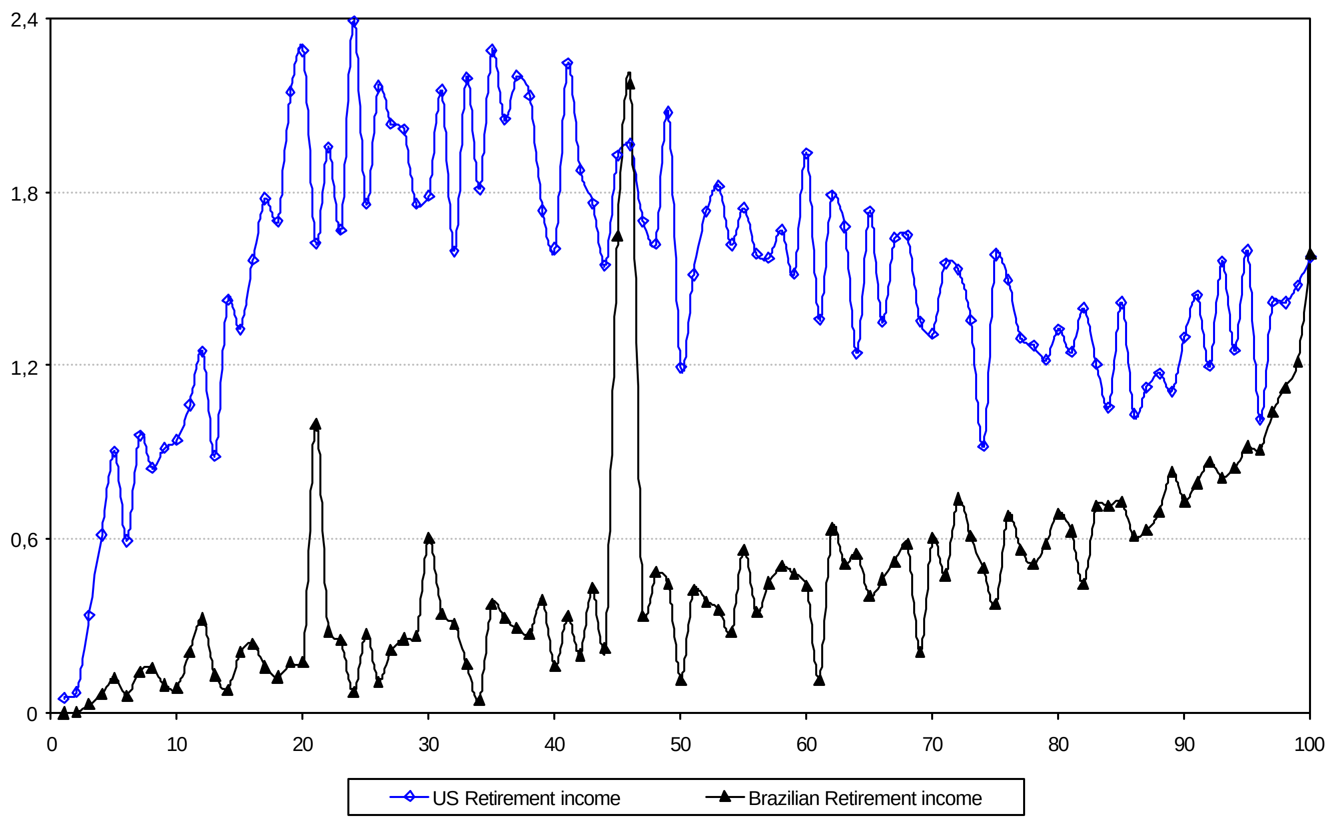
In order to identify the relative importance of the various components of non-labor income, we considered the effect of each source separately.[35] Private transfers are responsible for a drop in the Gini coefficient equal to 0.7 percentage points, certainly not a negligible effect. But most of the effect of unearned income is in effect due to retirement income. Retirement income is strongly inequality-increasing in Brazil, whereas it would be (mildly) equalizing in the US. This can be seen in Figure 9, which shows the mean retirement pension income for each hundredth of the distribution of household income. Apart for some outliers in the middle of the distribution, retirement income clearly concentrates among the richest households in Brazil, whereas it is the largest in the deciles just below the median in the US. The explanation of that difference is simple. Retirement income in Brazil concentrates among retirees of the formal sector who tend to be better off than the rest of the population.[36] In the US, on the contrary, retirees are more evenly distributed in the population. When summing up all income sources, they tend to be around the median of the distribution. Hence the switch from Brazilian to US retirement income is very strongly equalizing, reflecting first of all the universality of retirement in the US and the privilege that it may represent in Brazil.

Overall, the bottom line seems to be that differences in income inequality between Brazil and the United States are predominantly due to differences in the underlying distributions of endowments in the two countries, including among endowments the right to retirement income. Of the almost thirteen Gini points difference, almost ten can be ascribed to endowment effects. Among these, the data suggest almost equally important roles to inequalities in the Brazilian distribution of human capital (as proxied by years of schooling), and other claims on resources, measured by flows of non-labor income.

[35] This analysis is available from the authors on request.
[36] See Hoffman (2001) for an interesting analysis of the contribution of retirement pensions to Brazilian inequality. His findings confirm the importance of this income source to the country's high levels of inequality, but he shows that this effect is particularly pronounced in the metropolitan areas of the poorer Northeastern region, as well as in the states of Rio de Janeiro, Minas Gerais and Espírito Santo. The effect appears to be much weaker in rural areas.

The remaining three points of the Gini are due to price effects and, in particular, steeper returns to education in Brazil than in the US. Combined to the more unequal distribution of educational endowments themselves, this confirms the importance of education (prices and quantities) in driving Brazilian inequality, as previewed by the Theil decompositions reported in Section 2. While human capital remains firmly at the center-stage, our results suggest that it is joined there by the distribution of non-labor incomes and, in particular, of post-retirement incomes.

Figure 9: Incidence of Retirement Pensions in Brazil and the US



6. The Brazil - Mexico Comparison

The differences between the distributions of household income per capita in Brazil and Mexico are much smaller than those between either country and the US. The two Latin

American countries are at roughly the same level of development, and both are high inequality countries in international terms. Nevertheless, urban Brazil is much poorer than urban Mexico, and more unequal by any of the four measures reported in Table 7 below. The Lorenz curve for urban Brazil, in Figure 1, lies everywhere below Mexico's. The estimated coefficients for equations (7) and (8), run now so as to be strictly comparable between Brazil and Mexico, as well as those for the multinomial logit models for the demographic and educational structures, are included in Tables A6 – A9, in the Appendix.

In terms of the Gini coefficient, Brazil's excess inequality amounts to some seven points. Price effects account for 1.2 of these, with the variance of the residuals making no contribution at all to differences between Mexico and Brazil. Participation behavior and occupational structure also account for about a Gini point, but its interaction with the price effects is more-than-additive. The combined impact of all price and participation effects is of more than three points of the Gini.

Education alone also accounts for some three Gini points, but its interaction with occupational choice and price effects is less-than-additive. Joint simulation of Mexican ë, ã, á, â and ó^2 account for some four and a half of the seven-point difference. Interacting demographic effects takes away another Gini point from Brazil's measure, but again only once the Mexican approximated conditional distribution of education has been imported too. As in the case of the US, the educational structure of the population seems to be, either directly or indirectly, a powerful explanatory factor of the difference in household income distribution between Brazil and Mexico.

Replacing $\psi^A(W)$ by $\psi^B(W)$ - i.e. reweighing the Brazilian population so that its make-up in terms of exogenous characteristics such as age, race and household type is the same as Mexico's - has a small inequality-reducing effect: the Gini coefficient falls by 0.7 percentage point. This effect is slightly bigger when these new exogenous endowments are interacted with Mexican ("semi-exogenous" endowments of) education and fertility, as well as its price and occupational choice effects. They also help subtract a Gini point.

Altogether, the preceding effects account for almost all the difference observed between Brazil and Mexico, in terms of the Gini coefficient. This is not true, however, of the other inequality measures or of poverty, as shown in table 7. In particular, it can be seen that very little of the excessive relative poverty in Brazil is explained by the decomposition methodology, when it is limited to price, occupational structure and endowment effects, a feature that also appears quite clearly in Figure 12. As in the comparison with the US, it may thus be expected that what is left unexplained actually corresponds to the factors behind unearned income.

Table7 : Simulated Poverty and Inequality for Brazil in 1999, Using 1994 Mexico coefficients.

		Mean p/c Income	Inequality					Poverty Z =median/2 per month		
			Gini	E(0)	E(1)	E(2)	V(log)	P(0)	P(1)	P(2)
1	Brasil	294.8	0.569	0.597	0.644	1.395	1.101	26.23	10.10	5.36
2	Mexico	294.8	0.498	0.420	0.495	1.028	0.703	14.98	3.73	1.39
3	α , β	294.8	0.556	0.567	0.610	1.303	1.059	24.50	9.33	4.90
4	α , β , $\sigma 2$	294.8	0.557	0.570	0.613	1.314	1.063	24.62	9.39	4.94
5	λ	289.5	0.557	0.567	0.608	1.229	1.053	25.47	9.56	5.00
6	λ, α , β , $\sigma 2$	281.3	0.535	0.518	0.552	1.079	0.977	23.64	8.68	4.46
7	γ	375.3	0.537	0.544	0.532	0.908	1.112	18.04	6.87	3.62
8	λ , γ	399.2	0.535	0.540	0.525	0.889	1.108	16.47	6.12	3.18
9	λ, γ, α , β	285.1	0.522	0.500	0.513	0.950	0.981	22.95	8.56	4.44
10	λ, γ, α , β , $\sigma 2$	285.1	0.524	0.502	0.516	0.957	0.985	23.09	8.61	4.46
11	ψ	275.5	0.579	0.619	0.671	1.496	1.133	29.94	11.90	6.44
12	ψ , γ	348.0	0.537	0.550	0.529	0.891	1.144	20.48	8.13	4.41
13	ψ, λ , γ	389.7	0.532	0.538	0.514	0.844	1.125	17.41	6.61	3.50
14	ψ, λ, γ, α , β	282.6	0.514	0.490	0.493	0.887	0.991	22.85	8.82	4.70
15	ψ, λ, γ, α , β , $\sigma 2$	282.6	0.515	0.491	0.494	0.888	0.992	22.88	8.81	4.69
16	ξ_0	291.9	0.529	0.488	0.554	1.216	0.848	20.6	6.3	2.8
16	ψ, λ, γ, α , β , $\sigma 2$; $\xi 0$	279.9	0.447	0.348	0.356	0.539	0.678	14.75	4.40	1.87
17	ϕ	284.5	0.562	0.579	0.625	1.330	1.074	26.51	10.17	5.39
18	ϕ, ψ, λ, γ, α , β , $\sigma 2$	269.2	0.506	0.471	0.473	0.834	0.955	23.39	8.92	4.72
19	ϕ ; ξ_0	283.7	0.522	0.475	0.535	1.138	0.832	20.9	6.5	2.8
20	ϕ ; ψ, λ, γ, α , β , $\sigma 2$; $\xi 0$	268.6	0.437	0.331	0.337	0.496	0.650	14.94	4.43	1.88

Source: PNAD 1999 and ENIGH 1994

Unlike in Section 5, the conditional distribution of non-labor incomes in Mexico was approximated by a non-parametric method, described in footnote 18. As Figure 13 illustrates, the impact of this approximation is powerfully equalizing. By itself, it subtracts four points from the Brazilian Gini, and six points from the headcount index (see row 16: ξ_0, in Table 7). Tellingly, it almost halves the distribution-sensitive poverty measure FGT(2). At the same time, it may also be seen that, when combined with all the preceding changes, importing the structure of Mexican unearned incomes overshoots the observed difference between the two countries – see also figure 13. This means that the approximation error R_A for this decomposition is negative – and larger in module than in the previous section.[37]

In any case, however, the results obtained so far suggest that the Brazilian urban poor are at a disadvantage in terms of access to non-human assets and to public or private transfers when compared not only to their US counterparts - which might not be so surprising - but also when compared to the Mexican urban poor. This is an issue of clear relevance for the design of poverty-reduction policy in Brazil. Identifying more precisely the reasons of the difference with Mexico deserves further investigation.

7. Conclusions

This paper proposed a micro-econometric approach to investigating the nature of the differences between income distributions across countries. Since a distribution of household incomes is the marginal of the joint distribution of income and a number of other observed household attributes, simple statistical theory allows us to express it as an integral of the product of a sequence of conditional distributions and a (reduced order) joint distribution of exogenous characteristics. Our method is then to approximate these conditional distributions by pre-specified parametric models, which can be econometrically estimated in each country. We then construct counterfactual approximated income distributions, by importing sets of parameter estimates from the models of country B into

[37] In addition, the Brazil - Mexico decompositions appear, on the whole, to be less additively separable than the Brazil - US ones. The sum of individual effects in Table 7 is further away from the corresponding combined effects than in Table 5.

country A. This allows us to decompose the difference between the density functions (evaluated at any point) of the two distributions - or any of their functionals, such as inequality or poverty indices – into a term corresponding to the effect of the imported parameters, a residual term, and an approximation error. The decomposition residual can be reduced arbitrarily by combining the sets of parameters to be imported into a given simulation. The approximation error is shown to be small for the applications considered.

The sets of counterfactual income distributions constructed in this paper were designed to decompose differences across income distributions into effects due to three broad sources: differences in the returns or pricing structure prevailing in the countries' labor markets; differences in the parameters of the occupational structure of the economy; and differences in the endowments of age, race, gender, education, fertility and non-labor assets, broadly defined. By comparing the counterfactual distributions corresponding to each of these effects and to various combinations of them, we shed light on the nature of the inter-relationships between returns, occupations, and the underlying distributions of endowments. These can lead to interesting findings, such as a quantification of the impact of educational expansion on inequality through a specific channel: its effect on women's fertility behavior and labor force participation.

We applied this approach to the question of what makes the Brazilian distribution of income so unequal. In particular, we considered the determinants of the differences between it and the distributions of two other large American nations: Mexico and the United States. We found that differences in the structure of occupations account for little in both cases. Prices were not insubstantial in explaining difference between the US and Brazil, with this being due largely to steeper returns to education in Brazil. But the most important source of Brazil's uniquely large income inequality is the underlying inequality in the distribution of its human and non-human endowments. In particular, the main causes of Brazil's inequality - and indeed of its urban poverty - seem to be poor access to education and claims on assets and transfers that potentially generate non-labor incomes.

The importance of these non-labor incomes was one of our chief findings. Income distribution in Brazil would be much improved if only the distribution of this income component was more similar to those of the US or Mexico - themselves hardly paragons of the Welfare State. If this is due to public transfers, which needs to be investigated further, it is possible that our findings would vindicate those who have argued for a speedier public approach to the reduction in inequality than that which would be available from educational policies alone.

References.

Aghion, Philippe, Eve Caroli and C. Garcia-Peñalosa (1999): "Inequality and Economic Growth: The Perspective of the New Growth Theories", *Journal of Economic Literature*, **XXXVII** (4), pp.1615-1660.

Almeida dos Reis, José G. and Ricardo Paes de Barros (1991): "Wage Inequality and the Distribution of Education: A Study of the Evolution of Regional Differences in Inequality in Metropolitan Brazil", *Journal of Development Economics*, **36**, pp. 117-143.

Atkinson, Anthony B. (1970): "On the Measurement of Inequality", *Journal of Economic Theory*, **2**, pp.244-263.

Banerjee, Abhijit and Andrew Newman (1993): "Occupational Choice and the Process of Development", *Journal of Political Economy*, **101** (2), pp.274-298.

Bénabou, Roland (2000): "Unequal Societies: Income Distribution and the Social Contract", *American Economic Review*, **90** (1), pp.96-129.

Blau, Francine and Lawrence Khan (1996): "International Differences in Male Wage Inequality: Institutions versus Market Forces", *Journal of Political Economy*, **104** (4), pp.791-837.

Blinder, Alan S. (1973): "Wage Discrimination: Reduced Form and Structural Estimates", *Journal of Human Resources*, **8**, pp.436-455.

Bourguignon, François (1979): "Decomposable Income Inequality Measures", *Econometrica*, **47**, pp.901-20.

Bourguignon, François, Francisco H.G. Ferreira and Nora Lustig (1998): "The Microeconomics of Income Distribution Dynamics in East Asia and Latin America", World Bank DECRA mimeo.

Buhmann, B., L. Rainwater, G. Schmaus e T. Smeeding (1988): "Equivalence Scales, Well-being, Inequality and Poverty: Sensitivity Estimates Across Ten Countries using the Luxembourg Income Study database", *Review of Income and Wealth*, **34,** pp.115-42.

Cowell, Frank A. (1980): "On the Structure of Additive Inequality Measures", *Review of Economic Studies*, **47**, pp.521-31.

Cowell, Frank A. and Stephen P. Jenkins (1995): "How much inequality can we explain? A methodology and an application to the USA", *Economic Journal*, **105,** pp.421-430**.**

DiNardo, John, Nicole Fortin and Thomas Lemieux (1996): "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semi-Parametric Approach", *Econometrica*, **64** (5), pp.1001-1044.

Donald, Stephen, David Green and Harry Paarsch (2000): "Differences in Wage Distributions between Canada and the United States: An Application of a Flexible Estimator of Distribution Functions in the Presence of Covariates", *Review of Economic Studies*, **67**, pp.609-633.

Elbers, Chris, Jean O. Lanjouw, Peter Lanjouw and Phillippe G. Leite (2001): "Poverty and Inequality in Brazil: New Estimates from Combined PPV-PNAD Data", World Bank, DECRG Mimeo.

Ferreira, Francisco H.G., Peter Lanjouw and Marcelo Neri (2000): "A New Poverty Profile for Brazil using PPV, PNAD and census data", PUC-Rio, Department of Economics, TD#418.

Fishlow, Albert (1972): "Brazilian Size Distribution of Income", *American Economic Review*, **62** (2), pp.391-410.

Foster, J., J. Greer, and E. Thorbecke, 1984, “A class of decomposable poverty measures.” *Econometrica,* **52**, pp.761-65.

Henriques, Ricardo (2000): *Desigualdade e Pobreza no Brasil*, (Rio de Janeiro: IPEA)

Hoffman, Rodolfo (2001): "Desigualdade no Brasil: A Contribuição das Aposentadorias", UNICAMP, Instituto de Economia, mimeo.

Juhn, Chinhui, Kevin Murphy and Brooks Pierce (1993): “Wage Inequality and the Rise in Returns to Skill”, *Journal of Political Economy*, **101** (3), pp. 410-442.

Lam, David and Deborah Levinson (1992): "Age, Experience, and Schooling: Decomposing Earnings Inequality in the United States and Brazil", *Sociological Inquiry*, **62** (2), pp.218-145.

Langoni, Carlos G. (1973): *Distribuição da Renda e Desenvolvimento Econômico do Brasil*, (Rio de Janeiro: Expressão e Cultura).

Legovini, Arianna, César Bouillon and Nora Lustig (2001): "Can Education Explain Income Inequality Changes in Mexico", IADB Poverty and Inequality Unit, mimeo.

Machado, José A.F. and José Mata (2001): “Earning Functions in Portugal 1982-1994: Evidence from Quantile Regressions”, *Empirical Economics*, **26** (1), pp.115-134.

Oaxaca, Ronald (1973): "Male-Female Wage Differentials in Urban Labor Markets", *International Economic Review*, **14**, pp.673-709.

Sacconato, André L. and Naércio Menezes Filho (2001): "A Diferença Salarial entre os Trabalhadores Americanos e Brasileiros: Uma Análise com Micro Dados", Universidade de São Paulo, Instituto de Pesquisas Econômicas, TD No. 25/2001.

Shorrocks, Anthony F. (1980): "The Class of Additively Decomposable Inequality Measures", *Econometrica*, **48**, pp.613-25.

Shorrocks, Anthony F. (1999): "Decomposition Procedures for Distributional Analysis: A Unified Framework Based on the Shapley Value", University of Essex, Department of Economics, mimeo.

Székely, Miguel (1998): *The Economics of Poverty, Inequality and Wealth Accumulation in Mexico,* (New York: St. Martin's Press).

World Bank (2002): *Building Institutions for a Market Economy: World Development Report, 2002*, (New York: Oxford University Press).

Table 4: Simulated Poverty and Inequality for Brazilian earnings in 1999, Using 2000 USA coefficients.

	MEN						WOMEN					
	Mean p/c Income	Inequality					Mean p/c Income	Inequality				
		Gini	E(0)	E(1)	E(2)	V(log)		Gini	E(0)	E(1)	E(2)	V(log)
Brazil	636,3	0,517	0,467	0,510	0,902	0,837	411,1	0,507	0,450	0,488	0,838	0,819
USA	636,3	0,427	0,355	0,325	0,441	0,820	411,1	0,409	0,336	0,288	0,362	0,814
α, β												
i. Intercept	636,3	0,517	0,467	0,510	0,902	0,837	411,1	0,507	0,450	0,488	0,838	0,819
ii. Education	636,3	0,513	0,479	0,485	0,783	0,948	411,1	0,479	0,401	0,423	0,674	0,761
iii. Experience	636,3	0,575	0,609	0,644	1,244	1,120	411,1	0,535	0,506	0,549	0,986	0,914
iv. Race	636,3	0,515	0,463	0,507	0,893	0,830	411,1	0,497	0,430	0,467	0,791	0,783
v. Interaction: Age/Education	636,3	0,439	0,332	0,344	0,504	0,642	411,1	0,461	0,374	0,386	0,586	0,731
vi. Sector of Activity	636,3	0,513	0,457	0,502	0,884	0,817	411,1	0,508	0,451	0,489	0,839	0,823
vii. Formal/Informal	636,3	0,517	0,476	0,509	0,900	0,887	411,1	0,517	0,484	0,506	0,876	0,929
viii. All Betas	636,3	0,460	0,379	0,376	0,545	0,767	411,1	0,453	0,371	0,368	0,544	0,761
α, β, σ^2												
i. Intercept	636,3	0,540	0,516	0,562	1,039	0,927	411,1	0,545	0,533	0,578	1,084	0,971
ii. Education	636,3	0,536	0,528	0,536	0,910	1,038	411,1	0,519	0,483	0,510	0,888	0,913
iii. Experience	636,3	0,594	0,659	0,697	1,415	1,210	411,1	0,570	0,590	0,640	1,260	1,066
iv. Race	636,3	0,538	0,512	0,559	1,030	0,920	411,1	0,535	0,512	0,556	1,028	0,935
v. Interaction Age/Education	636,3	0,465	0,379	0,392	0,600	0,733	411,1	0,503	0,454	0,470	0,779	0,883
vi. Sector of Activity	636,3	0,536	0,506	0,554	1,020	0,907	411,1	0,545	0,534	0,578	1,085	0,975
vii. Formal/Informal	636,3	0,538	0,523	0,557	1,028	0,977	411,1	0,551	0,561	0,589	1,116	1,080
viii. All Betas	636,3	0,484	0,424	0,421	0,638	0,857	411,1	0,492	0,446	0,446	0,720	0,913
λ	722,9	0,502	0,434	0,475	0,803	0,772	465,4	0,503	0,439	0,471	0,781	0,800
λ, α, β	636,3	0,442	0,336	0,345	0,492	0,649	411,1	0,432	0,321	0,332	0,479	0,624
$\lambda, \alpha, \beta, \sigma^2$	636,3	0,468	0,382	0,392	0,584	0,735	411,1	0,476	0,400	0,415	0,651	0,773
γ	1210,0	0,477	0,408	0,400	0,572	0,825	705,9	0,468	0,391	0,384	0,545	0,789
λ e γ	1306,8	0,464	0,382	0,375	0,526	0,769	809,2	0,456	0,369	0,363	0,506	0,742
$\lambda, \gamma, \alpha, \beta$	636,3	0,428	0,322	0,315	0,421	0,654	411,1	0,415	0,300	0,297	0,396	0,608
$\lambda, \gamma, \alpha, \beta, \sigma^2$	636,3	0,455	0,367	0,361	0,505	0,741	411,1	0,460	0,378	0,376	0,547	0,761
ψ e γ	1235,3	0,469	0,397	0,381	0,529	0,818	732,2	0,457	0,373	0,361	0,500	0,762
$\psi, \gamma, \alpha, \beta$	636,4	0,441	0,346	0,333	0,447	0,717	411,1	0,431	0,328	0,319	0,425	0,674
$\psi, \gamma, \alpha, \beta, \sigma^2$	636,4	0,465	0,391	0,378	0,532	0,808	411,1	0,474	0,405	0,395	0,573	0,828
ψ, λ, γ	1281,8	0,463	0,385	0,369	0,506	0,796	797,2	0,449	0,361	0,348	0,477	0,743
$\psi, \lambda, \gamma, \alpha, \beta$	636,3	0,430	0,328	0,315	0,413	0,681	411,1	0,412	0,297	0,289	0,378	0,611
$\psi, \lambda, \gamma, \alpha, \beta, \sigma^2$	636,3	0,455	0,373	0,359	0,496	0,772	411,1	0,457	0,374	0,365	0,523	0,764
ϕ	818,7	0,528	0,492	0,518	0,865	0,907	508,7	0,524	0,485	0,510	0,834	0,896
$\phi, \psi, \lambda, \gamma, \alpha, \beta, \sigma^2$	704,3	0,448	0,362	0,349	0,484	0,751	435,3	0,454	0,369	0,362	0,520	0,752

Source: PNAD, 1999 and CPS, March 2000

Table 6: Simulated Poverty and Inequality for Brazilian earnings in 1999, Using 1994 Mexico coefficients.

	MEN						WOMEN					
	Mean p/c Income	Inequality					Mean p/c Income	Inequality				
		Gini	E(0)	E(1)	E(2)	V(log)		Gini	E(0)	E(1)	E(2)	V(log)
Brazil	636,2	0,517	0,467	0,511	0,906	0,837	410,3	0,507	0,449	0,486	0,831	0,818
Mexico	636,3	0,498	0,432	0,492	0,925	0,765	411,1	0,466	0,416	0,387	0,565	0,944
α, β												
i. Intercept	636,2	0,517	0,467	0,511	0,906	0,837	410,3	0,507	0,449	0,486	0,831	0,818
ii. Education	636,2	0,500	0,435	0,470	0,804	0,795	410,3	0,459	0,368	0,384	0,585	0,709
iii. Experience	636,2	0,516	0,463	0,509	0,904	0,827	410,3	0,516	0,466	0,508	0,891	0,840
iv. Interaction Age/Education	636,2	0,504	0,445	0,467	0,756	0,831	410,3	0,511	0,457	0,495	0,848	0,833
v. Sector of Activity	636,2	0,519	0,471	0,514	0,911	0,847	410,3	0,513	0,469	0,497	0,847	0,886
vi. Formal/Informal	636,2	0,539	0,509	0,563	1,052	0,890	410,3	0,520	0,470	0,520	0,934	0,831
vii. All Betas	636,2	0,500	0,431	0,469	0,794	0,776	410,3	0,490	0,421	0,449	0,745	0,793
α, β, σ^2												
i. Intercept	636,2	0,511	0,453	0,497	0,869	0,812	410,3	0,532	0,504	0,546	0,989	0,921
ii. Education	636,2	0,493	0,421	0,456	0,769	0,769	410,3	0,488	0,423	0,442	0,713	0,812
iii. Experience	636,2	0,509	0,449	0,494	0,867	0,802	410,3	0,541	0,521	0,568	1,057	0,942
iv. Interaction Age/Education	636,2	0,497	0,431	0,453	0,723	0,806	410,3	0,536	0,512	0,554	1,008	0,935
v. Sector of Activity	636,2	0,512	0,457	0,499	0,873	0,822	410,3	0,538	0,524	0,556	1,006	0,988
vi. Formal/Informal	636,2	0,533	0,494	0,547	1,009	0,864	410,3	0,546	0,528	0,584	1,115	0,933
vii. All Betas	636,2	0,493	0,417	0,454	0,758	0,751	410,3	0,518	0,479	0,512	0,903	0,895
λ	657,0	0,508	0,449	0,491	0,854	0,805	439,2	0,519	0,477	0,506	0,857	0,882
λ, α, β	636,2	0,478	0,392	0,421	0,675	0,718	410,3	0,481	0,399	0,425	0,673	0,738
$\lambda, \alpha, \beta, \sigma^2$	636,2	0,471	0,378	0,406	0,643	0,692	410,3	0,510	0,456	0,486	0,814	0,842
γ	912,5	0,523	0,486	0,499	0,803	0,916	615,4	0,514	0,479	0,471	0,703	0,950
λ, γ	926,6	0,525	0,493	0,501	0,794	0,940	736,0	0,516	0,495	0,467	0,673	1,022
$\lambda, \gamma, \alpha, \beta$	636,2	0,493	0,426	0,439	0,686	0,809	410,3	0,487	0,421	0,421	0,623	0,830
$\lambda, \gamma, \alpha, \beta, \sigma^2$	636,2	0,486	0,412	0,425	0,654	0,784	410,3	0,514	0,477	0,480	0,759	0,934
ψ, γ	922,4	0,517	0,479	0,484	0,771	0,924	628,2	0,504	0,465	0,444	0,637	0,949
$\psi, \gamma, \alpha, \beta$	636,2	0,495	0,435	0,443	0,701	0,842	410,3	0,474	0,402	0,391	0,553	0,813
$\psi, \gamma, \alpha, \beta, \sigma^2$	636,2	0,489	0,422	0,430	0,669	0,817	410,3	0,499	0,453	0,441	0,661	0,916
ψ, λ, γ	909,3	0,522	0,491	0,494	0,787	0,950	721,9	0,505	0,479	0,441	0,605	1,015
$\psi, \lambda, \gamma, \alpha, \beta$	636,2	0,483	0,414	0,416	0,629	0,811	410,3	0,469	0,398	0,380	0,520	0,823
$\psi, \lambda, \gamma, \alpha, \beta, \sigma^2$	636,2	0,477	0,401	0,404	0,600	0,786	410,3	0,494	0,449	0,429	0,624	0,926
ϕ	621,3	0,511	0,455	0,500	0,887	0,814	401,3	0,500	0,437	0,474	0,809	0,798
$\phi, \psi, \lambda, \gamma, \alpha, \beta, \sigma^2$	615,8	0,476	0,398	0,403	0,602	0,777	400,0	0,495	0,448	0,431	0,630	0,921

Source: PNAD, 1999 and ENIGH, 1994

Figure 4: Brazil-US Differences, Actual and Simulated, Steps 1 and 2

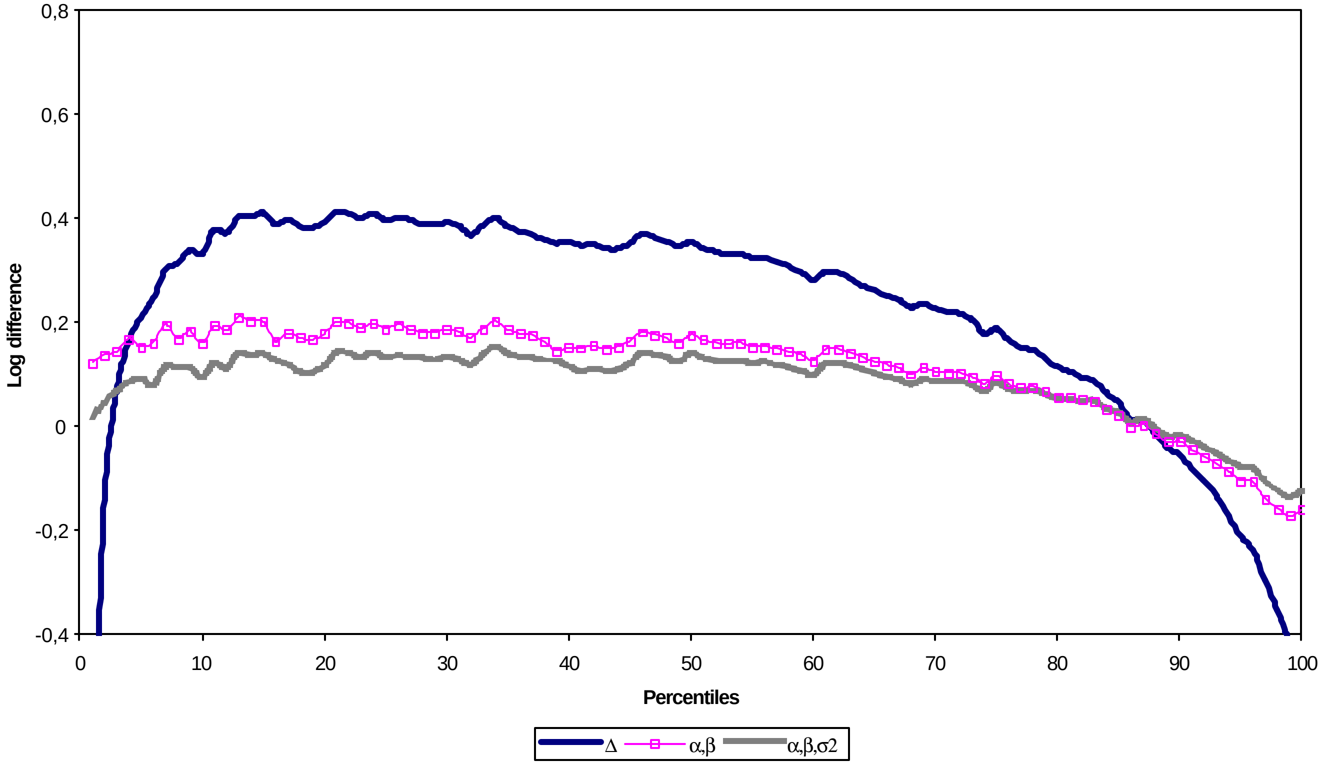


Figure 5: Brazil - US Differences, Actual and Simulated, Step 4

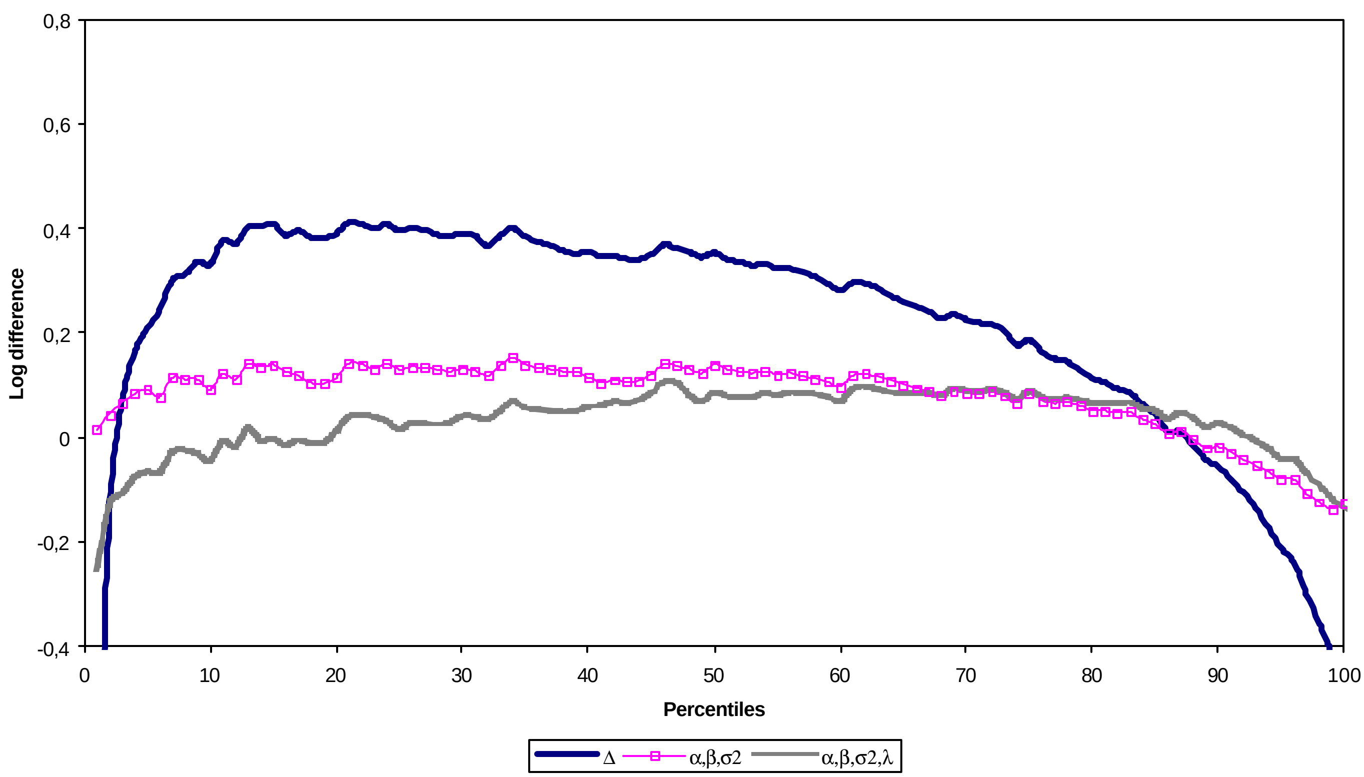


Figure 6: Brazil - US Differences, Actual and Simulated, Step 6.

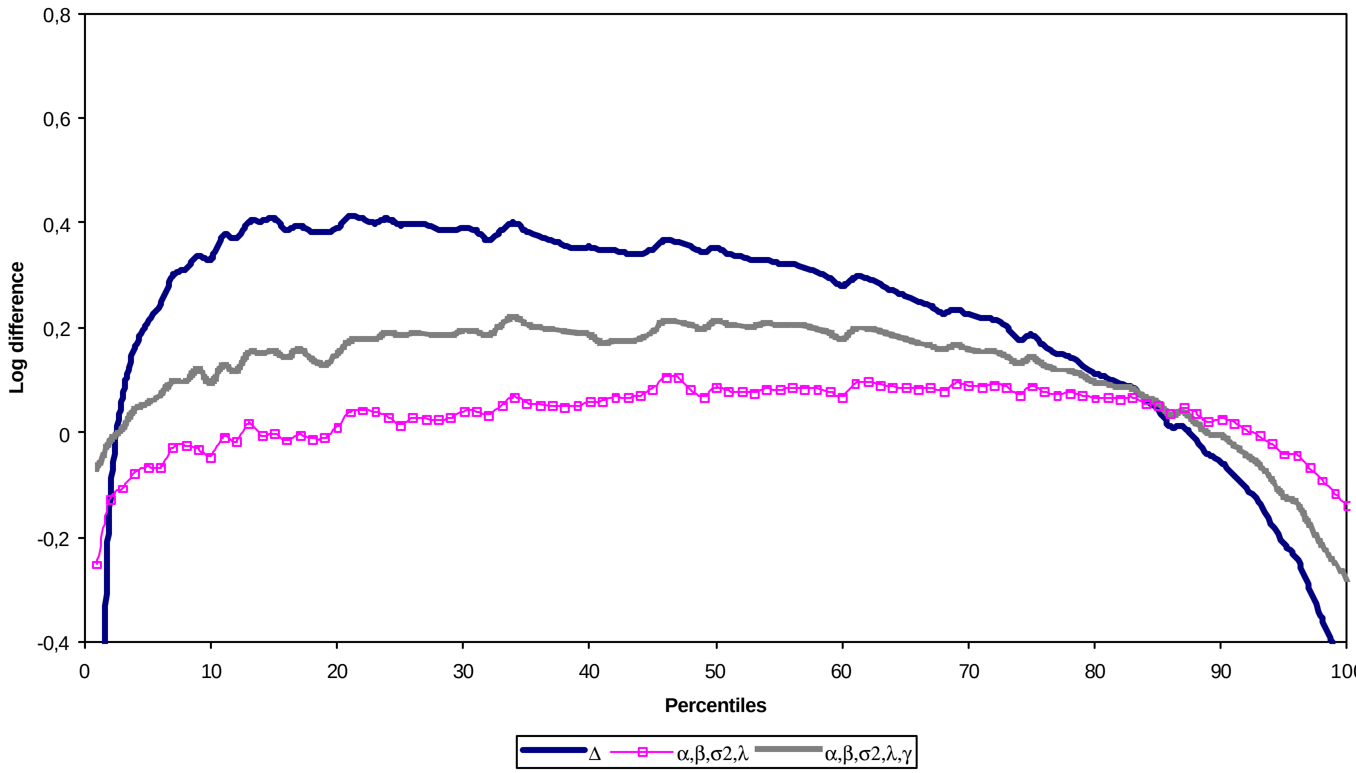


Figure 7: Brazil - US Differences, Actual and Simulated, Step 8

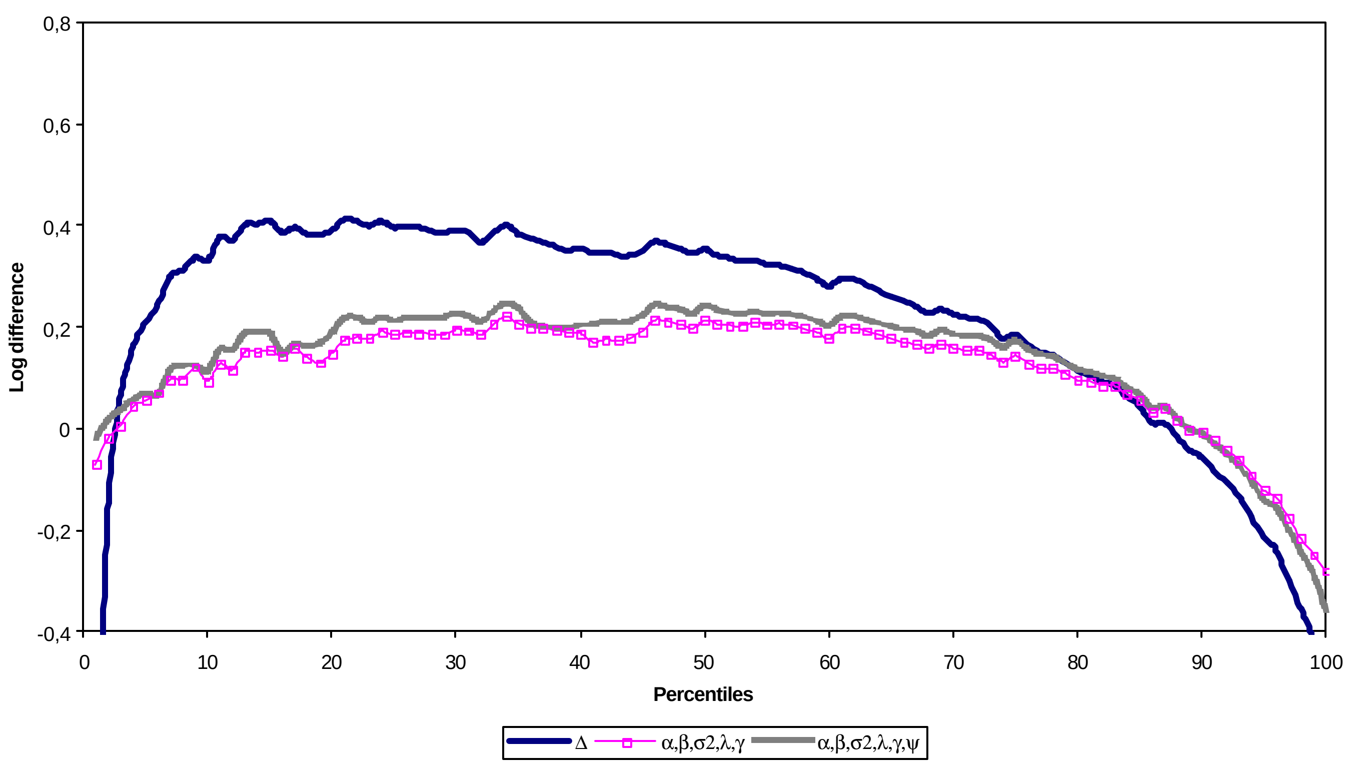


Figure 8: Brazil - US Differences, Actual and Simulated

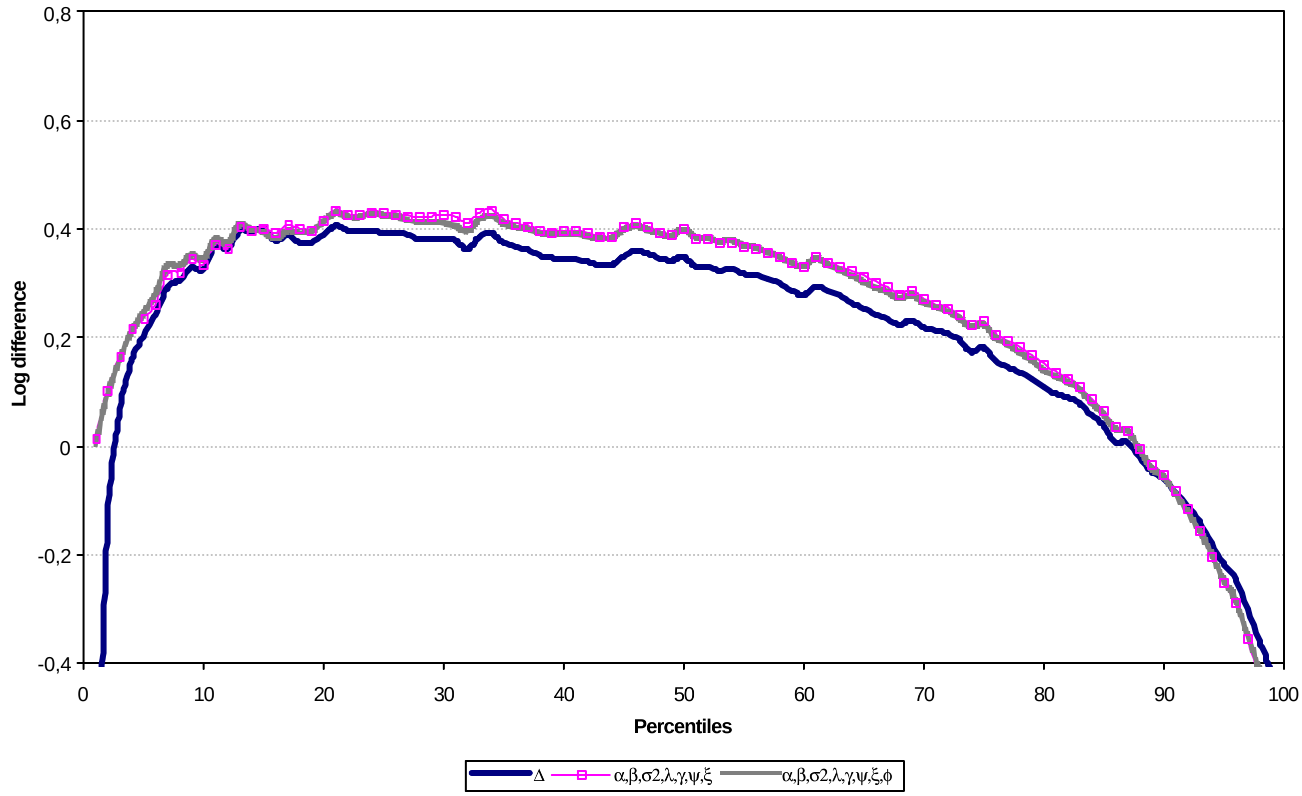


Figure 10: Brazil - Mexico Differences, Actual and Simulated, Steps 1 and 2

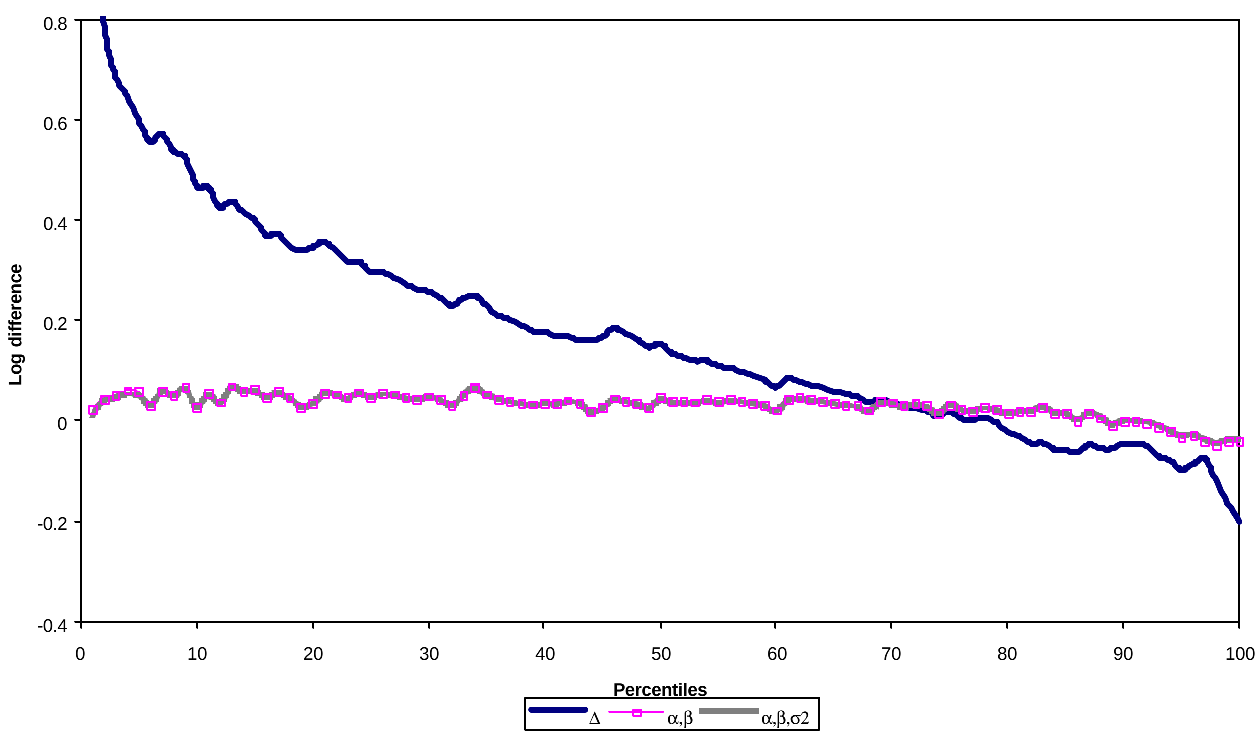


Figure 11: Brazil - Mexico Differences, Actual and Simulated, Step 4

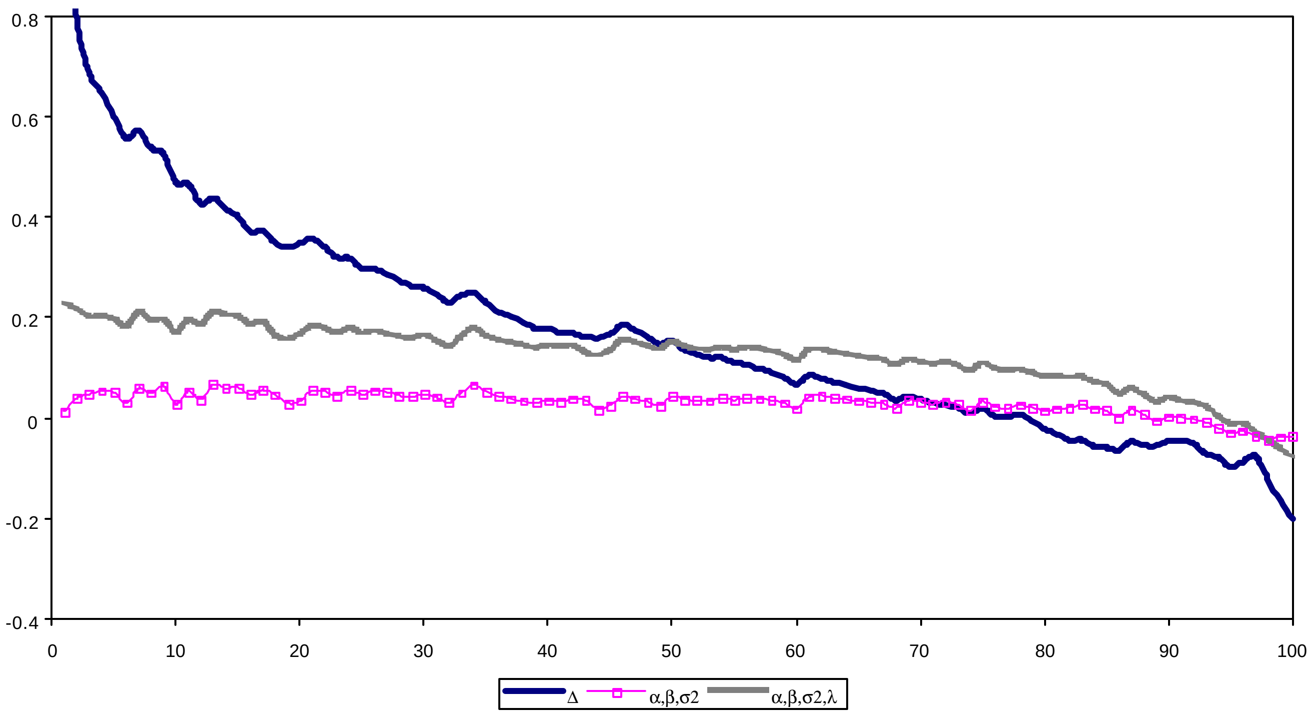


Figure 12: Brazil - Mexico Differences, Actual and Simulated, Step 6

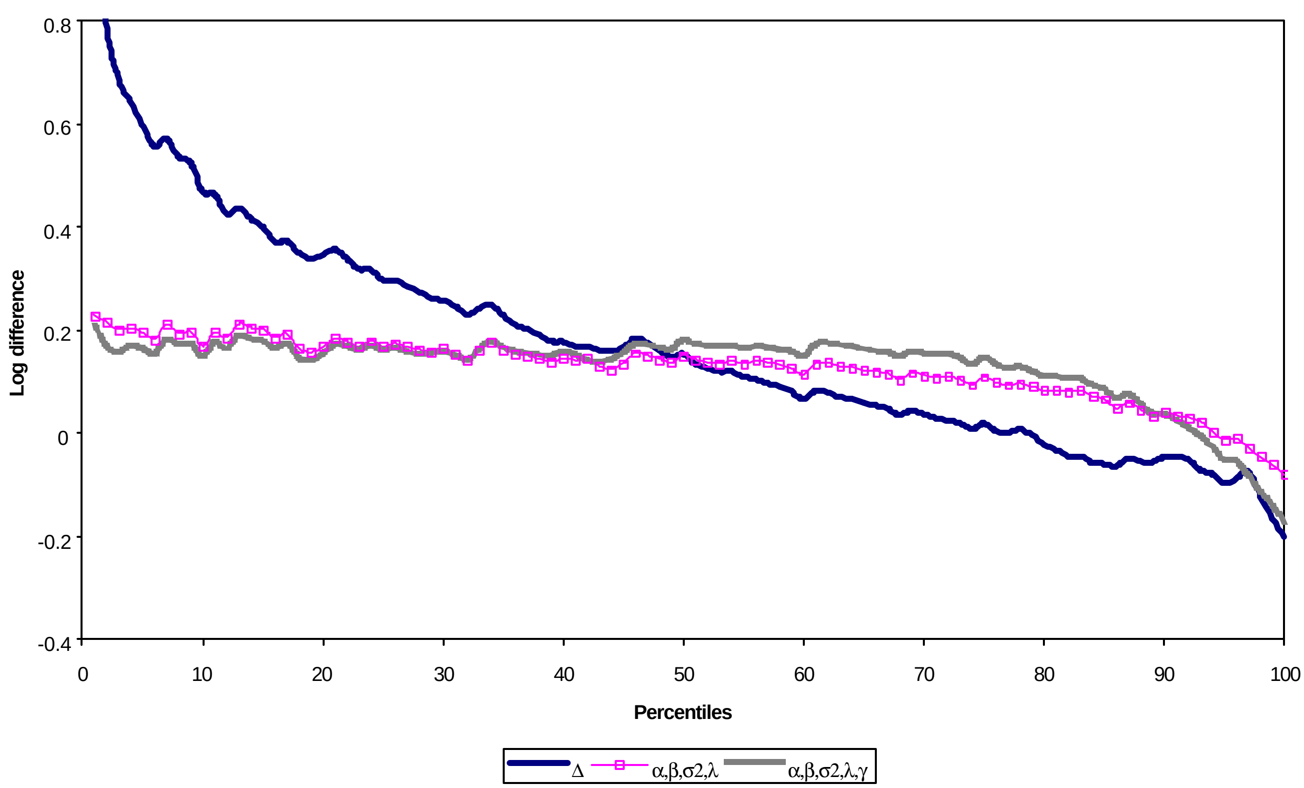


Figure 13: Brazil - Mexico Differences, Actual and Simulated, Step 8

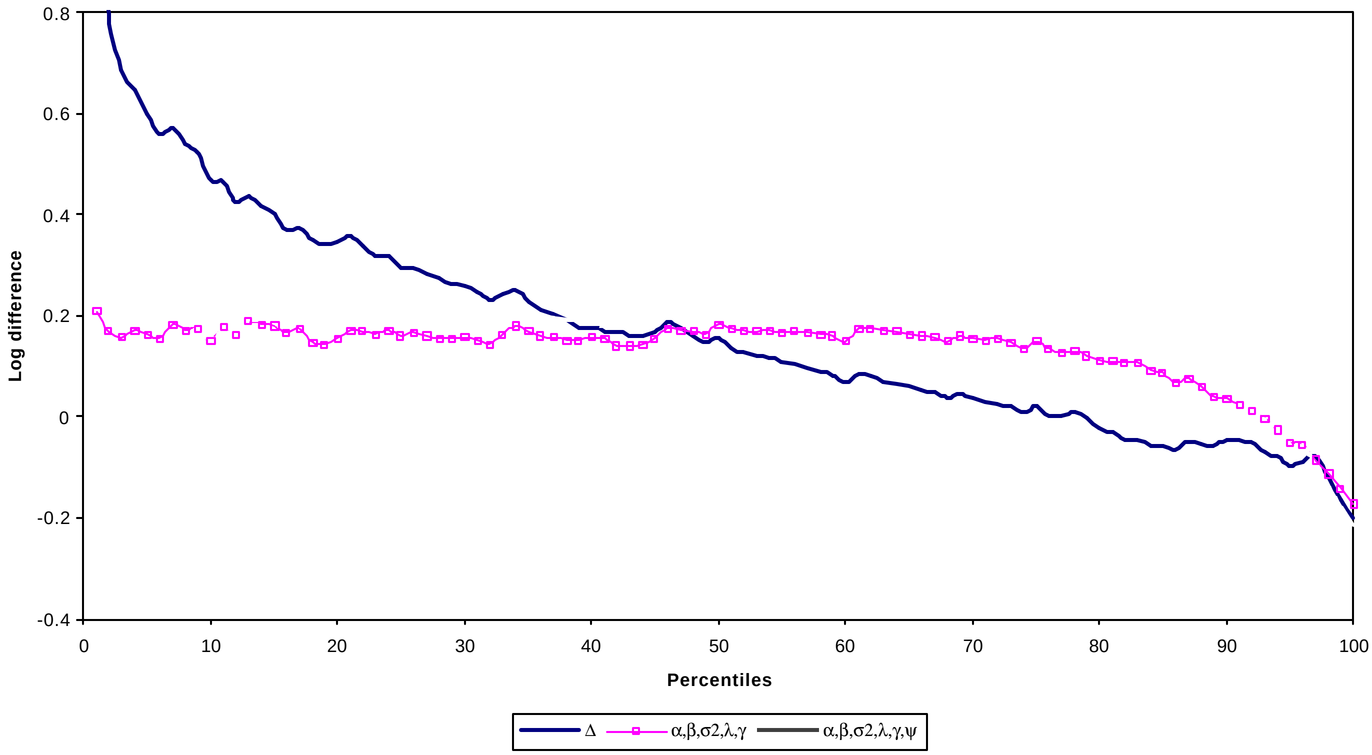


Figure 14: Brazil - Mexico Differences, Actual and Simulated, Step 12

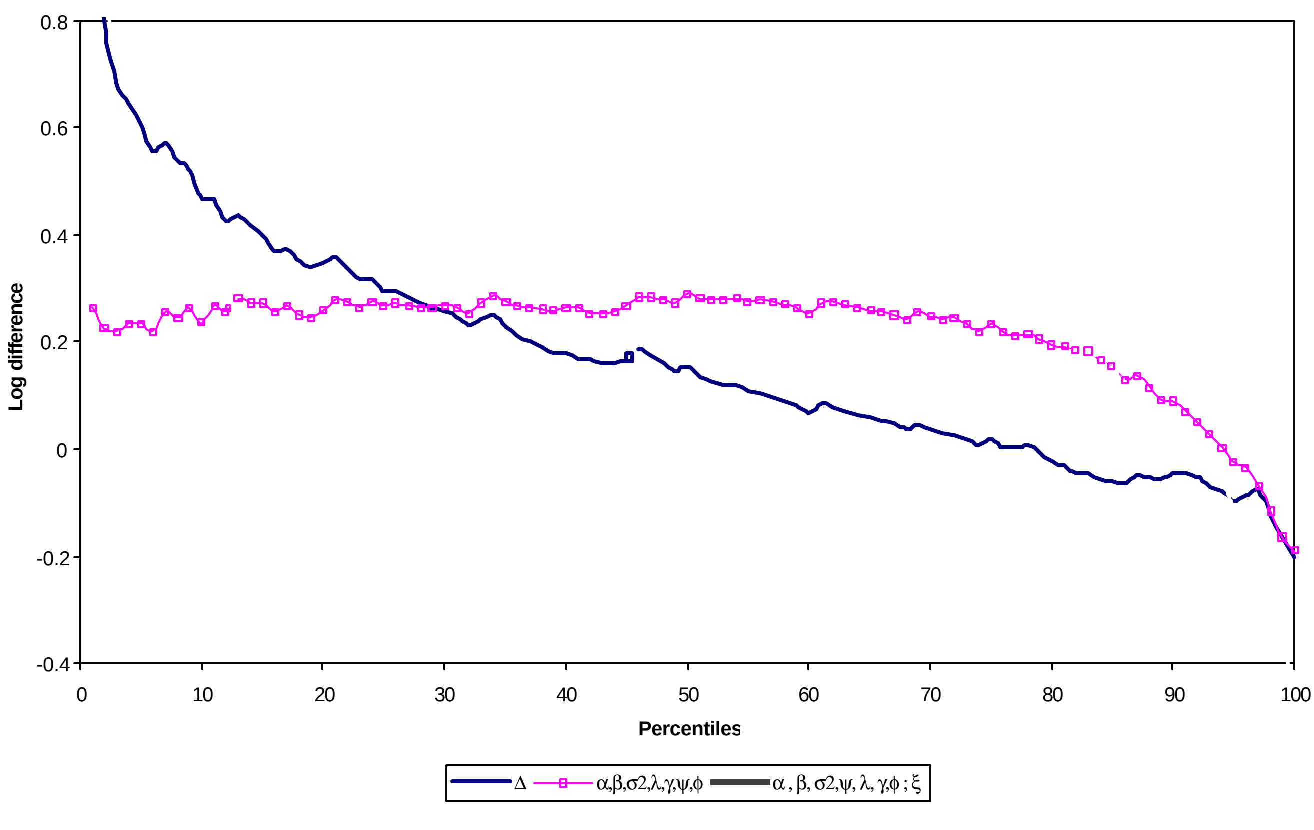


Table A1: The Multinomial Logit estimates for participation behavior and occupational choice: Brazil and the united States

	Brazil (1999)										
	Men				Women				Men		
	Formal employment in industry	Informal employment in industry	Formal employment in services	Informal employment in services	Formal employment in industry	Informal employment in industry	Formal employment in services	Informal employment in services	Formal employment in industry	Informal employment in industry	Formal employment in services
Age	0.281	0.352	0.288	0.326	0.389	0.263	0.306	0.316	0.442	0.389	0.320
Age2	-0.004	-0.004	-0.003	-0.004	-0.005	-0.003	-0.004	-0.003	-0.005	-0.005	-0.004
Education											
1 to 4	1.207	1.381	1.556	1.284	1.355	0.837	1.034	1.145	4.359	0.345	1.830
5 to 6	1.082	1.107	1.735	1.017	1.957	1.661	1.318	1.463	5.335	-2.486	2.156
7 to 8	0.472	0.682	1.310	0.905	2.125	1.064	0.813	1.525	4.370	-2.548	0.902
9 to 12	0.020	-0.725	1.464	0.424	2.076	1.293	1.364	1.644	4.265	-0.430	1.433
13 or more	-1.339	-2.139	0.627	0.085	1.773	1.225	1.228	1.973	3.064	-2.828	1.129
Age * education											
1 to 4	-0.020	-0.022	-0.022	-0.018	-0.016	-0.007	-0.012	-0.011	-0.064	0.028	-0.020
5 to 6	-0.016	-0.016	-0.020	-0.006	-0.029	-0.030	-0.020	-0.015	-0.076	0.090	-0.019
7 to 8	-0.005	-0.016	-0.011	-0.008	-0.031	-0.006	-0.004	-0.013	-0.058	0.086	0.005
9 to 12	0.005	0.006	-0.011	0.004	-0.028	-0.008	-0.006	-0.013	-0.050	0.056	0.005
13 or more	0.035	0.036	0.014	0.013	-0.018	-0.009	0.013	-0.013	-0.027	0.096	0.022
Race - White	0.040	0.059	0.076	0.409	0.105	0.237	-0.113	0.067	0.703	1.258	0.309
Average endowments of age	-0.019	-0.018	-0.014	-0.009	-0.011	0.003	-0.008	0.002	-0.028	-0.030	-0.022
Education among adults in his or her household											
0	1.456	0.865	0.859	0.466	0.617	-0.033	0.020	-0.331	2.462	0.879	1.772
1 to 4	1.443	1.021	0.942	0.550	0.429	-0.057	-0.025	-0.232	2.010	1.462	1.457
5 to 6	1.329	0.887	0.998	0.520	0.388	-0.214	-0.058	-0.330	1.973	1.131	1.468
7 to 8	1.153	0.706	0.969	0.648	0.297	-0.172	-0.155	-0.284	1.821	1.096	1.070
9 to 12	0.969	0.720	0.888	0.715	0.033	0.135	-0.338	-0.230	1.777	1.806	1.222
13 or more	0.443	0.410	0.494	0.623	-0.271	0.115	-0.603	-0.230	1.569	1.789	1.340
Numbers of children in the household	-0.042	-0.089	-0.089	-0.137	-0.003	0.018	-0.028	-0.047	-0.176	-0.309	-0.183
Numbers of children in the household	0.066	0.083	0.021	0.047	-0.129	0.085	-0.080	-0.024	0.036	0.130	-0.006
The individual is the head in the household	0.778	0.846	0.714	1.078	0.432	1.437	0.420	1.297	0.654	-0.208	0.219
The individual is not the head in the household	-0.106	-0.125	-0.025	0.115	0.179	0.442	0.133	0.510	0.470	-0.385	0.095
The individual is the spouse in the household					-0.429	0.750	-0.358	0.418			
If is not then head, is the head active?	-0.157	-0.223	-0.152	0.084	0.224	0.302	0.237	0.472	0.237	-0.678	-0.176
Intercept	-6.242	-8.622	-6.586	-8.523	-9.976	-12.134	-6.338	-10.038	-12.695	-11.357	-7.402

Source: PNAD 1999 and CPS March 2000

Table A2: Estimates for the Mincerian Equation: Brazil (1999) and USA (2000)

	Brazil								USA							
	MEN				WOMEN				MEN				WOMEN			
	R^2	coef	std	p-value	R^2	coef	std	p-value	R^2	coef	std	p-value	R^2	coef	std	p-value
	0.499				0.485				0.368				0.286			
Intercept		3.947	0.038	0.000		4.055	0.058	0.000		2.983	0.297	0.000		3.826	0.454	0.000
Education																
1 to 4		-0.073	0.031	0.019		-0.166	0.049	0.001		0.778	0.330	0.018		0.347	0.523	0.506
5 to 6		0.009	0.038	0.813		0.023	0.057	0.686		0.878	0.306	0.004		0.287	0.477	0.547
7 to 8		0.063	0.034	0.067		-0.008	0.052	0.885		0.638	0.306	0.037		0.029	0.472	0.951
9 to 12		0.067	0.033	0.040		0.202	0.049	0.000		0.925	0.295	0.002		0.453	0.452	0.317
13 or more		0.680	0.041	0.000		0.891	0.055	0.000		1.243	0.295	0.000		0.928	0.452	0.040
Age		0.079	0.001	0.000		0.046	0.002	0.000		0.150	0.007	0.000		0.100	0.010	0.000
Age^2		-0.001	0.000	0.000		-0.001	0.000	0.000		-0.001	0.000	0.000		-0.001	0.000	0.000
Age * education																
1 to 4		0.008	0.001	0.000		0.007	0.001	0.000		-0.015	0.007	0.031		-0.012	0.011	0.308
5 to 6		0.009	0.001	0.000		0.005	0.001	0.000		-0.018	0.007	0.008		-0.005	0.010	0.649
7 to 8		0.012	0.001	0.000		0.012	0.001	0.000		-0.009	0.007	0.188		0.002	0.010	0.811
9 to 12		0.022	0.001	0.000		0.018	0.001	0.000		-0.009	0.006	0.150		0.000	0.010	0.978
13 or more		0.026	0.001	0.000		0.022	0.001	0.000		-0.006	0.006	0.338		0.000	0.010	0.970
Race - White		0.188	0.006	0.000		0.159	0.007	0.000		0.164	0.012	0.000		-0.008	0.013	0.504
Sector of activity																
Agriculture		-0.352	0.010	0.000		-0.213	0.028	0.000		-0.180	0.024	0.000		-0.253	0.044	0.000
Industry		0.018	0.006	0.002		0.103	0.010	0.000		0.099	0.009	0.000		0.221	0.014	0.000
Employees		-0.035	0.005	0.000		0.098	0.007	0.000		0.454	0.013	0.000		0.666	0.016	0.000

Source: PNAD 1999 and CPS March 2000

Table A3: The Multinomial Logit Estimates for Demographic choices, Brazil and the United States

	Brazil (1999)					USA (2000)				
	Number of children					Number of children				
	0	1	2	3	4	0	1	2	3	4
Race - White	-0.574	-0.262	-0.324	-0.280	-0.155	-0.784	-0.160	-0.202	-0.212	-0.121
Numbers of adults in the household	0.865	0.779	0.730	0.501	0.283	0.979	0.644	0.802	0.776	0.407
Age	0.094	0.036	0.022	0.013	0.005	0.091	0.027	0.014	0.009	0.005
Education										
1 to 4	0.627	0.403	0.468	0.386	0.103	-0.286	-0.198	-0.382	-0.338	-0.461
5 to 6	1.023	0.926	1.110	0.890	0.383	-0.017	0.072	-0.268	-0.115	-0.450
7 to 8	1.825	1.690	1.766	1.362	0.686	0.449	0.347	-0.004	0.008	0.034
9 to 12	2.688	2.467	2.432	1.798	0.813	1.719	1.474	1.224	0.796	0.269
13 or more	4.125	3.645	3.679	2.950	0.844	2.357	2.022	1.865	1.217	0.580
Intercept	-0.723	0.874	1.307	1.026	0.440	0.310	1.043	1.472	1.198	0.668

Source: PNAD 1999 and CPS March 2000

Table A4: The Multinomial Logit Estimates fo Educational Structure: Brazil and the United States

	Brazil (1999)					USA (2000)				
	Years of schooling					Years of schooling				
	0	1 to 4	5 to 6	7 to 8	9 to 12	0	1 to 4	5 to 6	7 to 8	9 to 12
Gender - male	-0.041	0.070	0.112	0.071	-0.134	0.051	0.183	0.106	0.161	-0.041
Age	0.047	0.009	-0.040	-0.037	-0.038	0.047	0.047	0.034	0.045	0.007
Race - White	-2.221	-1.723	-1.622	-1.324	-0.889	-0.781	-0.417	-0.184	-0.223	-0.185
Cohort										
1931 to 1940	-4.224	-3.706	-4.715	-4.282	-3.969	-1.969	-0.685	-1.484	-3.969	-1.985
1941 to 1950	-5.020	-4.462	-5.442	-5.177	-4.719	-2.308	-0.946	-1.927	-4.837	-2.453
1951 to 1960	-5.535	-4.957	-5.493	-5.399	-4.902	-1.735	-0.962	-1.920	-4.902	-2.473
1961 to 1970	-5.486	-5.280	-5.292	-5.508	-4.913	-1.965	-0.585	-1.507	-4.653	-2.354
1971 to 1980	-5.173	-5.216	-5.271	-5.489	-4.630	-1.927	-0.574	-1.266	-4.126	-2.293
Intercept	4.287	6.675	7.745	8.133	7.794	-4.363	-5.101	-3.037	-0.387	2.058

Source: PNAD 1999 and CPS March 2000

TableA5: Tobit Model Estimates for Non-Labor Incomes in Brazil and the United States

	Brazil (1999)			USA (2000)		
	coef	std	p-value	coef	std	p-value
Gender - male	-239.38	7.77	0.000	173.88	7.87	0.000
Race - White	88.19	7.12	0.000	225.24	11.37	0.000
Age	28.65	1.11	0.000	10.78	1.17	0.000
Age^2	0.12	0.01	0.000	0.29	0.01	0.000
Education						
1 to 4	116.39	10.63	0.000	-99.07	69.60	0.155
5 to 6	236.22	16.47	0.000	-68.14	65.17	0.296
7 to 8	277.86	13.81	0.000	91.34	62.61	0.145
9 to 12	456.68	12.45	0.000	435.98	59.33	0.000
13 or more	902.82	13.96	0.000	863.07	59.35	0.000
The individual is the head in the household	557.85	8.24	0.000	184.05	8.11	0.000
Intercept	-2925.83	29.39	0.000	-2103.24	65.68	0.000

Source: PNAD 1999 and CPS March 2000

Standart desviations of residual	**953.36**			**1088.00**		
Left-censored observation (<=0)	**153,143**	**79%**		**35,300**	**36%**	
Uncensored observations	**39,972**	**21%**		**61,894**	**64%**	
Total	**193,115**			**97,194**		
R^2	**0.07**			**0.03**		

Table A6: TheMultinomial Logit Estimates for participation behavior and occupational choice: Brazil and Maxico

	Brazil (1999)										
	Men				Women				Men		
	Formal employment in industry	Informal employment in industry	Formal employment in services	Informal employment in services	Formal employment in industry	Informal employment in industry	Formal employment in services	Informal employment in services	Formal employment in industry	Informal employment in industry	Formal employment in services
Age	0.281	0.352	0.287	0.324	-0.388	-0.262	-0.307	-0.316	-0.212	-0.223	-0.227
Age2	-0.004	-0.004	-0.003	-0.004	0.005	0.003	0.004	0.003	0.003	0.003	0.003
Education											
1 to 4	1.205	1.376	1.551	1.263	-1.345	-0.830	-1.043	-1.142	-2.609	-0.436	-1.253
5 to 6	1.080	1.104	1.731	1.002	-1.943	-1.647	-1.331	-1.458	-2.479	-0.835	-1.823
7 to 8	0.470	0.679	1.308	0.909	-2.120	-1.064	-0.816	-1.524	-2.133	-0.337	-1.236
9 to 12	0.021	-0.722	1.468	0.455	-2.076	-1.304	-1.361	-1.646	-0.709	0.490	-0.614
13 or more	-1.329	-2.125	0.644	0.184	-1.786	-1.261	-1.210	-1.983	1.101	2.550	0.158
Age * education											
1 to 4	-0.020	-0.022	-0.022	-0.017	0.016	0.007	0.012	0.011	0.043	-0.010	0.009
5 to 6	-0.016	-0.015	-0.020	-0.005	0.028	0.029	0.020	0.014	0.041	0.009	0.023
7 to 8	-0.005	-0.016	-0.011	-0.007	0.031	0.005	0.004	0.013	0.035	0.016	0.011
9 to 12	0.005	0.006	-0.011	0.004	0.028	0.007	0.006	0.013	0.000	-0.003	-0.014
13 or more	0.035	0.036	0.014	0.013	0.018	0.009	-0.013	0.013	-0.048	-0.038	-0.047
Average endowments of age	-0.019	-0.018	-0.014	-0.008	0.011	-0.003	0.008	-0.002	0.027	0.031	0.019
Education among adults in his or her household											
0.000	1.445	0.850	0.844	0.420	-0.593	0.076	-0.043	0.344	-1.508	-1.110	-0.733
1 to 4	1.437	1.013	0.935	0.539	-0.415	0.088	0.010	0.241	-1.340	-0.820	-0.870
5 to 6	1.324	0.879	0.992	0.508	-0.376	0.243	0.044	0.339	-0.925	-0.542	-0.671
7 to 8	1.149	0.701	0.967	0.652	-0.288	0.191	0.146	0.290	-1.097	-0.294	-1.209
9 to 12	0.968	0.719	0.890	0.738	-0.031	-0.131	0.337	0.232	-0.660	-0.185	-0.587
13 or more	0.449	0.416	0.506	0.680	0.266	-0.128	0.611	0.227	0.296	1.748	0.174
Numbers of children in the household	-0.042	-0.089	-0.090	-0.141	0.004	-0.016	0.027	0.047	0.060	0.121	0.094
Numbers of children in the household	0.065	0.082	0.019	0.039	0.132	-0.079	0.077	0.026	-0.123	-0.134	-0.061
The individual is the head in the household	0.779	0.848	0.716	1.089	-0.436	-1.437	-0.419	-1.297	-0.969	-1.447	-0.848
The individual is not the head in the household	-0.106	-0.124	-0.023	0.122	-0.183	-0.447	-0.130	-0.511	0.854	-0.256	0.938
The individual is the spouse in the household					0.425	-0.757	0.362	-0.421			
If is not then head, is the head active?	-0.157	-0.223	-0.153	0.084	-0.225	-0.304	-0.236	-0.473	0.196	0.332	0.461
Intercept	-6.213	-8.581	-6.538	-8.263	9.900	11.961	6.417	9.989	3.858	5.123	3.233

Source: PNAD 1999 and ENIGH 1994

Table A7: Estimates for the Mincerian Equation: Brazil (1999) and Mexico (1994)

	Brazil								**Mexico (1994)**							
	MEN				**WOMEN**				**MEN**				**WOMEN**			
	R^2	coef	std	p-value	R^2	coef	std	p-value	R^2	coef	std	p-value	R^2	coef	std	p-value
	0.491				0.480				0.430				0.432			
Intercept		4.052	0.038	0.000		4.141	0.058	0.000		5.110	0.140	0.000		4.668	0.243	0.000
Education																
1 to 4		-0.089	0.032	0.005		-0.169	0.050	0.001		0.256	0.127	0.044		0.157	0.205	*0.444*
5 to 6		0.004	0.038	*0.913*		0.022	0.058	0.707		0.044	0.125	*0.727*		0.359	0.203	*0.077*
7 to 8		0.064	0.035	*0.065*		-0.002	0.052	0.967		-0.104	0.150	*0.488*		0.290	0.266	*0.276*
9 to 12		0.083	0.033	0.012		0.223	0.049	0.000		-0.038	0.123	*0.758*		0.413	0.198	0.037
13 or more		0.736	0.041	0.000		0.934	0.055	0.000		0.657	0.132	0.000		0.778	0.231	0.001
Age		0.079	0.001	0.000		0.046	0.002	0.000		0.070	0.004	0.000		0.054	0.008	0.000
Age^2		-0.001	0.000	0.000		-0.001	0.000	0.000		-0.001	0.000	0.000		-0.001	0.000	0.000
Age * education																
1 to 4		0.008	0.001	0.000		0.007	0.001	0.000		-0.001	0.003	*0.585*		-0.001	0.004	*0.875*
5 to 6		0.010	0.001	0.000		0.006	0.001	0.000		0.008	0.003	0.004		-0.001	0.004	*0.747*
7 to 8		0.013	0.001	0.000		0.012	0.001	0.000		0.017	0.004	0.000		0.008	0.007	*0.262*
9 to 12		0.023	0.001	0.000		0.019	0.001	0.000		0.020	0.003	0.000		0.013	0.004	0.003
13 or more		0.026	0.001	0.000		0.022	0.001	0.000		0.021	0.003	0.000		0.019	0.006	0.001
Sector of activity																
Agriculture		-0.355	0.010	0.000		-0.219	0.028	0.000		-0.406	0.045	0.000		-1.746	0.109	0.000
Industry		0.011	0.006	*0.054*		0.108	0.010	0.000		0.012	0.017	*0.478*		-0.107	0.033	0.001
Employees		-0.042	0.005	0.000		0.090	0.007	0.000		0.115	0.020	0.000		0.512	0.031	0.000

Source: PNAD 1999 and ENIGH 1994

Table A8: The Multinomial Logit Estimates for Demographic Choice: Brazil (1999) and Mexico (1994)

	Brazil (1999)					Mexico (1994)				
	Number of children					Number of children				
	0	1	2	3	4	0	1	2	3	4
Numbers of adults in the household	-0.578	-0.265	-0.328	-0.284	-0.158	-0.459	-0.135	-0.193	-0.176	-0.097
Age	0.096	0.037	0.024	0.014	0.006	0.115	0.046	0.034	0.020	0.015
Education										
1 to 4	0.703	0.466	0.525	0.422	0.122	0.467	0.321	0.325	0.383	0.465
5 to 6	1.114	1.003	1.180	0.935	0.407	1.552	1.278	1.567	1.306	1.053
7 to 8	1.965	1.810	1.878	1.435	0.724	2.618	2.380	2.178	1.946	1.537
9 to 12	2.892	2.645	2.598	1.906	0.871	3.032	2.743	2.766	2.077	1.480
13 or more	4.459	3.944	3.959	3.141	0.950	5.013	4.284	4.395	3.512	1.086
Intercept	-0.419	1.151	1.565	1.197	0.533	-2.373	-0.433	0.081	0.396	-0.374

Source: PNAD 1999 and ENIGH 1994

Table A9: The Multinomial Logit Estimates for Educational Structure: Brazil (1999) and Mexico (1994)

	Brazil (1999)					Mexico (1994)				
	Years of schooling					Years of schooling				
	0	1 to 4	5 to 6	7 to 8	9 to 12	0	1 to 4	5 to 6	7 to 8	9 to 12
Gender - male	-0.014	0.092	0.133	0.087	-0.126	-1.011	-0.573	-0.627	-0.214	-0.626
Age	0.041	0.005	-0.044	-0.040	-0.040	0.041	0.011	-0.006	-0.109	-0.046
Cohort										
1931 to 1940	-4.216	-3.720	-4.734	-4.299	-3.979	0.000	0.000	0.000	0.000	0.000
1941 to 1950	-4.973	-4.447	-5.432	-5.172	-4.717	-1.041	-0.928	-0.838	-1.783	-0.852
1951 to 1960	-5.462	-4.919	-5.462	-5.377	-4.889	-1.758	-1.671	-1.418	-2.551	-1.504
1961 to 1970	-5.410	-5.232	-5.249	-5.477	-4.895	-2.491	-2.216	-1.764	-3.606	-1.604
1971 to 1980	-5.110	-5.171	-5.230	-5.459	-4.611	-1.758	-1.730	-1.155	-3.020	-0.900
Intercept	2.912	5.504	6.621	7.174	7.112	-1.776	0.566	1.564	5.021	3.767

Source: PNAD 1999 and ENIGH 1994