Latin American and Caribbean Economic Association
Asociación de Economía de América Latina y el Caribe

CLIMBING A MOVING MOUNTAIN:
Explaining the Decline in Income Inequality in Brazil from 1976 to 1996

Francisco H.G. Ferreira and Ricardo Paes de Barros [1]

October 1998
Preliminary and Incomplete.
Please do not cite.

First Workshop of the
LACEA/IDB/World Bank Inequality and Poverty Network

1 Respectively the World Bank and IPEA (Rio). We are very grateful to Francois Bourguignon for his guidance and support, and to Carlos Henrique Corseuil, Phillippe George Leite e Daniele Reis for superb research assistance.

October 1998
Preliminary and Incomplete.
Please do not cite.

CLIMBING A MOVING MOUNTAIN:
Explaining the Decline in Income Inequality in Brazil from 1976 to 1996

Francisco H.G. Ferreira and Ricardo Paes de Barros [1]

Keywords: Brazil; Income Distribution; Education; Demographics; Micro-Simulations.

JEL Classification: C15, D31, J22, I21, I32

Abstract: Brazilian income inequality fell between 1976 and 1996. The Theil index declined from 0.86 to 0.69, and the Gini coefficient also fell, although by less. Over the same period, however, the earnings-education profile became more convex, and our simulations indicate that applying the 1996 structure of returns to education to the (otherwise unchanged) 1976 population would have *increased*, rather than lowered, inequality. Changes in other returns (to experience, or gender) as well as changes in labor force participation behavior also contribute to an increase in dispersion. Residual changes (in the joint distribution of observable and unobservable individual characteristics across the population) must account for the observed declines in inequality. We hypothesize that one important component of these equalizing changes might have been a change in the educational composition of the population. By moving a sufficient number of workers up along the (shifting) earnings-education profile, this effect would have overwhelmed the others, and reduced inequality. It would also have offset a tendency for poverty to increase over the period.

---

[1] Respectively the World Bank and IPEA (Rio). We are very grateful to Francois Bourguignon for his guidance and support, and to Carlos Henrique Corseuil, Phillippe George Leite e Daniele Reis for superb research assistance.

1. Introduction.

Between 1976 and 1996, the Brazilian economy underwent significant structural transformations. The country's population grew from 110 million to 160 million; real GDP more than tripled to U$ 750 billion; and its sectoral make up also changed considerably. Against that background, a change in the Gini coefficient, based on an individual distribution of household incomes per capita from the Pesquisa Nacional por Amostra de Domicilios (PNAD) from 0.6049 to 0.5864 may seem rather small. One is tempted to speak about stability, rather than of a decline in inequality. However, other measures of inequality, such as the Theil index, record a more substantial decline, from 0.8574 to 0.6856.[2]

Over the same period, however, the evidence which we will examine below suggests that the earnings-education profile became more convex, controlling for age and gender. This implies an increase in the rate of return to education for a number of pairwise comparisons of education levels. Furthermore, as we will show below, returns to other factors, notably experience and gender, changed very little. Their impact, if any, seems to have been to incompletely mitigate the inequality-augmenting effect of the change in the structure of returns to education. The combined effect of changes in these returns – the 'price effects' – was still an increase in simulated inequality. And this remained the case once the determinants of labor force participation decisions were also taken into account. The puzzle which motivates this study, then, is what explains the observed decline in inequality.

We address these issues by means of a micro-simulation based decomposition of distributional changes, developed by Bourguignon et. al. (1998), which itself builds upon the work of Juhn, Murphy and Pierce (1993). This approach has two distinguishing features. First, unlike other dynamic inequality decompositions such as that proposed by

[2] This paper, as the reader will rapidly realize, is an intermediate input. It is work in progress, and a number of results are still preliminary. On a number of occasions, we will refer to planned future improvements. This is the first such instance: the final version of this paper will test for Lorenz dominance between the 1976 and 1996, and if it exists, for whether it is statistically significant.

Mookherjee and Shorrocks (1982), it decomposes the effects of changes in an entire distribution, rather than on a scalar summary statistic (such as the mean log deviation). This allows for a greater wealth of simulations, investigating the effect of the changes in specific parameters on any number of inequality or poverty measures (and then for any number of poverty lines or assumptions about equivalence scales. Second, the distribution which it decomposes changes in is a distribution of household incomes per capita (with the recipient unit generally being the individual). Therefore, moving beyond labor market studies, the effect of household composition on living standards and participation decisions is explicitly taken into account.

The remainder of the paper is organized as follows. Section 2 outlines the basic model and the describes the empirical methodology. Section 3 presents the results of the estimation stage, and discusses some of its implications. Section 4 presents the (preliminary) results of the simulation stage, and draws inferences from them. Section 5 concludes and suggests the next steps in this ongoing research project.

2. The Model and the Decomposition Methodology

The model described below is the Brazilian version of the general semi-reduced form model for household income and labor supply in Bourguignon et. al. (1998). At this stage, we examine only two years, 1976 and 1996, but we hope to extend the analysis to another year in the middle of the period, in order to gain a better understanding of the dynamics between the end-points of the interval. Also, this paper covers only Brazil's urban areas (which account for some three quarters of its population). The general model therefore collapses to two occupational sectors: wage earners and self-employed in urban areas.[3] We experiment with two different formulations for the income of the self employed. In the first version, their income is assumed to be produced through a household production function, resembling a backyard shoe repair shop where all self-employed family members

[3] We will eventually extend the model to cover rural areas too, by incorporating two additional sectors: wage earners and self-employed in the rural areas. In Brazil, wage earners include employees with or without formal documentation ('com ou sem carteira'). The self-employed are own-account workers ('conta propria').

lend a hand. In the second, in equations denoted by primes, the incomes of the self-employed are estimated individually, according to the same specification as applied to wage earners. See below.

Therefore, the household income equations are given by:

(1) $$Y = \sum_{i=1}^{n} w_i L_i^w + \Pi + Y_0$$ or

(1') $$Y = \sum_{i=1}^{n} w_i L_i^w + \sum_{i=1}^{n} p_i L_i^{se} + Y_0$$

Where Y is total household income; w_i are the total wage earnings of individual i, L^w is a dummy variable that takes the value one if individual i is a wage earner (and zero otherwise); Π is total self employment income; π_i is the self-employment profit of individual i: L^{se} is a dummy that takes the value one if individual i is self-employed (and zero otherwise); and Y_0 is income from any other sources, such as transfer or capital incomes. Equations 1 or 1' are not estimated econometrically. They aggregate information on right-hand-side terms 1 (from equations 2 and 4), 2 (from equation 3 or 3') and 3 from the household data set.

The wage-earnings equation is given by:

(2) $$Logw_i = X_i^P b^w + e_i^w$$

where X_i^P = (ed, ed^2, age, age^2 , D_g). Ed denotes completed years of schooling (net of repetition). Age is used instead of experience, because the latter, more desirable variable is not available for 1976. D_g is a gender dummy, which takes the value 1 for females (and zero for males). w are monthly earnings. [4] This extremely simple specification was chosen so as to make the simulation stage of the decomposition feasible, as described below. It

[4] Hours may be introduced in a separate paper, when the entire model changes to take them into account. Similarly, a more satisfactory specification for experience, such as (age – education – 6) will be experimented with.

embodies the assumption that the Brazilian labor market was not segmented by region, firm size, race, or any attribute other than gender. [5]

The self-employed earnings equations are given by:

(3) $$Log\Pi = \left[Log\sum_i (L_i^{se} + L_i^{nr}), .. g(X_i^P) \right] b^{se} + e_i^{se}$$

where

$$g(X_h^P) = \left(\frac{\sum (L_i^{se} + L_i^{nr}) ed_i}{\sum (L_i^{se} + L_i^{nr})}, \left[\frac{\sum (L_i^{se} + L_i^{nr}) ed_i}{\sum (L_i^{se} + L_i^{nr})} \right]^2, \frac{\sum (L_i^{se} + L_i^{nr}) \exp_i}{\sum (L_i^{se} + L_i^{nr})}, \left[\frac{\sum (L_i^{se} + L_i^{nr}) \exp_i}{\sum (L_i^{se} + L_i^{nr})} \right]^2, \frac{\sum (L_i^{se} + L_i^{nr}) D_{gi}}{\sum (L_i^{se} + L_i^{nr})} \right)$$

and the coefficient on the number of workers (β_1) can be transformed into a coefficient on self-employment earnings per worker by the simple transformation $b_1 = 1 - \beta_1$. Or:

(3') $$Logp_i = X_i^P b_i^{'se} + e_i^{'se}$$

Equations 2, 3 and 3' are estimated by simple OLS. Equation (2) is estimated for all employees (whether or not heads of household), and whether com or sem carteira. Equation (3) is estimated as a "household production function", where labor is supplied by household members reporting self-employment status (conta-propria) (L^{se}) and those who report themselves as active but unpaid (L^{nr}). Equation 3' is estimated for all self-employed individuals (whether or not heads of households). Because the errors ε are unlikely to be independent from the exogenous variables, a sample selection bias correction procedure should be used. Rather than using the standard Heckman procedure, we intend to use an amended version, as in Lee (1984), in the final version. In this version, we assume all errors are independently distributed, and do not correct for sample selection bias in the earnings regressions.

[5] In a later version, the gender dummy will be interacted with the other variables, to allow for a more comprehensive understanding of how they interact.

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

(4) $$P_i^s = \frac{e^{Z_i^o g_s}}{e^{Z_i^o g_s} + \sum_{j \# s} e^{Z_i^o g_j}} \quad \text{where s = (0, w, se)}$$

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

For household heads:

$$Z_1^j = \left(\begin{array}{l} \underline{X_1^P}; n_{0-13}, n_{14-65}, n_{>65}, \frac{1}{n_{14-65}} \sum_{-1} D_{14-65} ed, \left[\frac{1}{n_{14-65}} \sum_{-1} D_{14-65} ed \right]^2, \frac{1}{n_{14-65}} \sum_{-1} D_{14-65} age, \\ \left[\frac{1}{n_{14-65}} \sum_{-1} D_{14-65} age \right]^2, \frac{1}{n_{14-65}} \sum_{-1} D_{14-65} Gd, D \end{array} \right)$$

For other members of the household:

$$Z_i^j = \left(\begin{array}{l} \underline{X_i^P}; n_{0-13}, n_{14-65}, n_{>65}, \frac{1}{n_{14-65}} \sum_{-i} D_{14-65} ed, \left[\frac{1}{n_{14-65}} \sum_{-i} D_{14-65} ed \right]^2, \frac{1}{n_{14-65}} \sum_{-i} D_{14-65} age, \\ \left[\frac{1}{n_{14-65}} \sum_{-i} D_{14-65} age \right]^2, \frac{1}{n_{14-65}} \sum_{-1} D_{14-65} Gd, D_1^{se}, L_1^w w_1, D \end{array} \right)$$

Where n_{k-m} is the number of persons in the household whose age falls between k and m; D_{14-65} is a dummy that takes the value one for individuals whose age is between 14 and 65; D^{se} is a dummy for a self-employed head, the penultimate term is the earnings of a wage-earning head; and D is a dummy variable that takes the value one if there are no individuals aged 14-65 in the household. The sums defined over {-j} are sums over { $\forall i \in h / j$ }

The multinomial logit model in (4) corresponds to the following discrete choice process:

(5) $$s = Arg \max_j \{U_j = Z^j . g_j + x^j ,\ j = (0, w, se)\}$$

where Z^j are given above, separately for household heads and other members; the ξ_j are random variables with a double exponential density function and U_j may be interpreted as the utility of alternative j. Once the vector γ is estimated by (4), and a random term ξ is drawn, each individual chooses an occupation j so as to maximize the above utility function.

A Decomposition of Changes in the Distribution of Household Income[6]

Once equations 2, 3 (or 3') and 4 have been estimated, we have two vectors of parameters for each of the two years in our sample ($t \in \{1976, 1996\}$): β_t from the earnings equations, for both wage earners and the self-employed, and γ_t from the participation equation. In addition, from equation 1, we have Y_{0ht} and Y_{ht} . Let $X_{ht} := \{X_i^P, Z_i \mid i \in h\} \cup \{X_p^h\}$ and $\Omega_{ht} := \{\varepsilon^w_i, \varepsilon^{se}_i, \xi^j_i \mid i \in h\}$. We can then write the total income of household h at time t as follows:

(6) $$Y_{ht} = H\left(X_{ht}, Y_{oht}, \Omega_{ht}; b_t, g_t\right) \qquad h=1,\dots,m$$

Based on this representation, the distribution of household incomes:

(7) $$D_t = \{Y_{1t}, Y_{2t}, ..., Y_{mt}\}$$

6 This section draws heavily on Bourguignon et.al. (1998), adapting it to our specifications.

can be rewritten as:

(8) $$D_t = D\left[\{X_{ht}, Y_{oht}, \Omega_{ht}\}, b_t, g_t\right]$$

Where *{.}* refers to the joint distribution of the corresponding variables over the whole population. We are interested in understanding the evolution of D_t over time, or possibly that of a set of alternative summary poverty or inequality measures based on it.

The basic decomposition proposed in this project consists of estimating the effects of changing one or more of the arguments of *D[.]* on D_t. The simplest decomposition applies to those arguments which are exogenous to the household: that is, the *bs*, *gs*, and the variance of the residual terms. Changing the *bs* amounts to assuming a change in the rate of return on human capital variables in equation (2) and (3) or (3'). We refer to this as the "price effect". In algebraic terms, it can be expressed as:

(9) $$B_{tt'} = D\left[\{X_{ht}, Y_{oht}, \Omega_{ht}\}, b_{t'}, g_t\right] - D\left[\{X_{ht}, Y_{oht}, \Omega_{ht}\}, b_t, g_t\right]$$

This expression measures the contribution to the overall change in the distribution $D_{t'} - D_t$ of a change in *b* between *t* and *t'*, holding all else constant. Likewise, the "labor supply effect" may be defined by:

(10) $$L_{tt'} = D\left[\{X_{ht}, Y_{oht}, \Omega_{ht}\}, b_t, g_{t'}\right] - D\left[\{X_{ht}, Y_{oht}, \Omega_{ht}\}, b_t, g_t\right]$$

The price effect $B_{tt'}$ is obtained by comparing the distribution at date *t* with the hypothetical distribution obtained by simulating on the population observed at date *t* the remuneration structure of period *t'*. A price effect can be computed individually – that is, for one element of the vector β, or collectively – that is, for all elements of the vector β. Both types of simulations are reported below.

Likewise, the labor supply effect, $L_{tt'}$, is obtained by comparing the initial distribution with the hypothetical distribution obtained by simulating on the population observed at date *t* the occupational preferences observed at date t'. Again, a labor-supply effect can be computed individually – that is, for one element of the vector γ, or collectively – that is, for all elements of the vector γ. We only report collective labor-supply decompositions in this paper.

Considering only the collective price and labor supply effects, one can then write the change in the distribution of household income as the sum of a price effect, a labor supply effect and a residual:

(11) $$D_t - D_{t'} = B_{tt'} + L_{tt'} + R_{tt'}$$

The residual $R_{tt'}$ measures the contribution to the change in the distribution of income of changes in the distributions of observable and unobservable characteristics, respectively all the X_s and Y_0, and all the e_s and h_s. (11) is an exact decomposition.

3. Estimating the Model

Equation (2) was estimated for wage earners (formal and informal) only, using OLS and based on the earnings, education, age and gender information available in the PNAD surveys for both 1976 and 1996,. The results are shown in Table 1 below:

Table 1: Dependent variable: Monthly earnings of wage earners

	Year					
	1976			1996		
	Coefficient	St. Error	P-value	Coefficient	St. Error	P-value
Education	0.0942	0.0018	0.0001	0.0603	0.0020	0.0001
Education2	0.0018	0.0001	0.0001	0.0039	0.0001	0.0001
Age	0.0949	0.0011	0.0001	0.0942	0.0010	0.0001
Age2	-0.0010	0.0000	0.0001	-0.0010	0.0000	0.0001
Gender	-0.6535	0.0058	0.0001	-0.5047	0.0051	0.0001

Source: Based on "Pesquisa Nacional por Amostra de Domicílios" (PNAD) of the 1976 and 1996.

The static results are not surprising. All variables are significant and have the expected signs. The coefficients on education and its square are positive and significant. The effect of experience, as proxied by age, is positive but concave. The gender dummy (female =1) is negative, significant and large.

The dynamics are more interesting. Between 1976 and 1996, the earnings-education profile changed shape. The linear component fell substantially, but the coefficient of squared years of schooling doubled, implying that the relationship became more convex. This would suggest a fall in the conditional mean incomes of people with low levels of education, and a steepening of the rate with which these increase from low levels.

Remarkably, returns to experience were basically constant over the period. The coefficients (and standard errors) of both age and its square are almost identical, implying that the concave shape of the partial earnings-experience profile among wage earners was unchanged. Female earnings, controlling for both age and experience, were substantially lower in both periods, with only a marginal decline in the this effect between 1976 and 1996.

Table 2 below reports the results of the OLS estimation of Equation (3), for the self-employed. As discussed in Section 2, this equation assumes that the labor income of the self-employed comes from their pooling their labor with other self-employed members of their households, as well as those in the household who report themselves as active but unpaid. The stylized picture it seeks to capture is that of the family farm or, in an urban setting, a family shop or workshop.

On the main variables, there is remarkable consistency with the trends detected for wage workers. Returns to education among the self-employed (here measured as household averages) changed from concave to convex over the period. Returns to experience were not very different from the wage sector, and also broadly unchanged over the period. The negative effect of females in self-employed production units was less pronounced, but still substantial and significant in 1996.

Table 2: Dependent variable: Labor Income of the Self-Employed
Version 1: Household Production Function

	Year					
	1976			1996		
	Coefficient	Standard	P-value	Coefficient	Standard	P-value
Number of self employed and unpaid worker	0.5963	0.0233	0.0001	-0.2696	0.0180	0.0001

Mean education of self employed and unpaid worker	0.1929	0.0053	0.0001	0.1170	0.0049	0.0001
Mean education of self employed and unpaid worker ^{2}	-0.0033	0.0004	0.0001	0.0012	0.0003	0.0003
Mean age of self employed and unpaid worker	0.0939	0.0031	0.0001	0.0913	0.0027	0.0001
Mean age of self employed and unpaid worker ^{2}	-0.0010	0.0000	0.0001	-0.0010	0.0000	0.0001
Women proportion between self employees and unpaid worker	-1.0621	0.0176	0.0001	-0.7405	0.0150	0.0001
Source: Based on "Pesquisa Nacional por Amostra de Domicílios" (PNAD) of the 1976 and 1996.						

Notice, however, that raw labor involved in the presumed self-employed production function changed sign from significantly positive in 1976 to significantly negative in 1996. The interpretation of the negative coefficient in the latter year is not straight forward, as it suggests a negative marginal product of raw labor in these units. One possibility is that the ‘household production function’ specification is inappropriate for the Brazilian urban own-account sector. Whereas it may be well-suited to the family farm sector, in Brazil or elsewhere, it is possible that most of the urban self-employed do not actually work jointly with other household members. It was to allow for that possibility that we ran equation (3’), where reported individual self-employment incomes are regressed on the same explanatory variables used for wage-earners (in equation 2). The results are reported in Table 3 below.

This table reveals that education is also an important correlate of incomes in the self-employment sector. The coefficient on the linear term has a higher value in both years than for wage-earners, but the quadratic term is lower. This implies that, ceteris paribus, the return to low levels of education might be higher in self-employment than in wage work, but would eventually become lower as years of schooling increase. This will clearly have an impact on occupational choice, estimated through equation (4). Dynamically, the same trend was observed as for wage-earners: the coefficient on the linear term fell over time, but the relationship became more convex. [7] The coefficients on age and age squared are very similar in value to those observed for wage earners, and also fail to change much over the period, although there is a slightly more pronounced decline in the coefficient of the linear term. The effect of being female, ceteris paribus, is even more markedly negative in this sector than in the wage sector, although it also fell somewhat between the two years.

Table 3: Dependent variable: Labor Income of the Self-Employed Version 2: Individual Estimation						
	Year					
	1976			1996		
	Coefficient	Standard	P-value	Coefficient	Standard	P-value
Education	0.1836	0.0051	0.0001	0.1098	0.0045	0.0001
Education2	-0.0028	0.0004	0.0001	0.0012	0.0003	0.0001
Age	0.1047	0.0027	0.0001	0.0974	0.0023	0.0001
Age2	-0.0012	0.0000	0.0001	-0.0010	0.0000	0.0001
Gender	-1.1106	0.0160	0.0001	-0.7297	0.0128	0.0001
Source: Based on "Pesquisa Nacional por Amostra de Domicílios" (PNAD) of the 1976 and 1996.						

Let us now turn to the estimation of the multinomial logit in equation (4). This was estimated separately for household heads and for others, since the set of explanatory variables was slightly different in each case (see the description of vectors Z_1 and Z_i in Section 2 above). Table 4 below presents the results for household heads, and Table 5 for other household members. In table 4, the results are presented as the effects of other choices versus that of being a wage-earner. In table 5, they are presented as the effects of other choices versus that of remaining outside the labor force ('unoccupied'). [8]

For household heads, education was not significantly related to the likelihood of choosing to work in the wage sector vis-à-vis staying out of the labor force, in either 1976 or 1996. It did, however, significantly enhance the probability of being in wage work vis-à-vis in self-employment in1976. This latter effect had disappeared by 1996.

[7] In this case, it in fact switched from concave to convex.
[8] The next version of the paper will present both tables with 'unoccupied' as the default category.

Table 4: Dependent variable: Participation of the Household Head

	Year					
	1976			1996		
	Coefficient	Standard	P-value	Coefficient	Standard	P-value
Unoccupied versus occupied as employee						
Education	-0.0123	0.0096	0.1970	-0.0012	0.0078	0.8810
Education^2	-0.0023	0.0006	0.0000	-0.0018	0.0005	0.0000
Age	-0.0658	0.0065	0.0000	-0.1527	0.0049	0.0000
Age^2	0.0016	0.0001	0.0000	0.0024	0.0001	0.0000
Gender	1.8979	0.0389	0.0000	0.9224	0.0255	0.0000
Number of members from 0 to 14*	0.0020	0.0087	0.8180	0.0149	0.0097	0.1240
Number of members from 14 to 65*	0.0689	0.0108	0.0000	0.0790	0.0115	0.0000
Number of members older than 65*	-0.0904	0.0495	0.0680	-0.0061	0.0450	0.8920
Presence of other members from 14 to 65 (dummy)	-0.3284	0.1298	0.0110	-0.4939	0.1065	0.0000
Mean education*	0.0880	0.0122	0.0000	-0.0091	0.0111	0.4100
Mean education^2*	-0.0008	0.0008	0.3370	0.0039	0.0007	0.0000
Mean age*	-0.0125	0.0077	0.1040	-0.0040	0.0060	0.5100
Mean age^2*	0.0004	0.0001	0.0010	0.0002	0.0001	0.0140
Women proportion*	-0.0106	0.0054	0.0490	-0.0003	0.0039	0.9400
Constant	-1.7775	0.1408	0.0000	1.0521	0.1048	0.0000
Occupied as self-employed versus occupied as employee						
Education	-0.0725	0.0093	0.0000	0.0075	0.0083	0.3680
Education^2	-0.0013	0.0006	0.0270	-0.0040	0.0005	0.0000
Age	0.0380	0.0066	0.0000	0.0222	0.0057	0.0000
Age^2	-0.0001	0.0001	0.3370	0.0002	0.0001	0.0140
Gender	0.0181	0.0437	0.6780	-0.5567	0.0318	0.0000
Number of members from 0 to 14*	0.0309	0.0071	0.0000	0.0697	0.0092	0.0000
Number of members from 14 to 65*	-0.0079	0.0109	0.4700	-0.0106	0.0124	0.3910
Number of members older than 65*	-0.0336	0.0495	0.4980	-0.0873	0.0504	0.0830
Presence of other members from 14 to 65 (dummy)	0.1166	0.1334	0.3820	-0.0248	0.1193	0.8350
Mean education*	0.0019	0.0118	0.8710	0.0297	0.0114	0.0090
Mean education^2*	0.0019	0.0008	0.0140	0.0012	0.0007	0.0980
Mean age*	-0.0127	0.0078	0.1010	-0.0063	0.0066	0.3450
Mean age^2*	0.0001	0.0001	0.2800	0.0001	0.0001	0.2970
Women proportion*	-0.0039	0.0062	0.5310	-0.0143	0.0045	0.0020
Constant	-2.0339	0.1385	0.0000	-1.8078	0.1233	0.0000

Source: Based on "Pesquisa Nacional por Amostra de Domicílios" (PNAD) of the 1976 and 1996.
Note: * excluding the person whose participation is estimated.

Table 5: Dependent variable: Participation of Other Members						
	Year					
	1976			1996		
	Coefficient	Standard	P-value	Coefficient	Standard	P-value
Occupied as employee versus unoccupied						
Education	0.2060	0.0059	0.0000	0.0733	0.0065	0.0000
Education^2	-0.0055	0.0004	0.0000	0.0030	0.0004	0.0000
Age	0.3340	0.0038	0.0000	0.3027	0.0031	0.0000
Age^2	-0.0051	0.0001	0.0000	-0.0042	0.0000	0.0000
Gender	-1.3968	0.0185	0.0000	-0.8605	0.0166	0.0000
Number of members from 0 to 14*	-0.1061	0.0046	0.0000	-0.1230	0.0065	0.0000
Number of members from 14 to 65*	0.2390	0.0052	0.0000	0.1581	0.0061	0.0000
Number of members older than 65*	0.2123	0.0218	0.0000	-0.0682	0.0200	0.0010
Presence of other members from 14 to 65 (dummy)	-1.2206	0.1307	0.0000	0.2948	0.1106	0.0080
Mean education*	-0.3143	0.0074	0.0000	-0.1630	0.0073	0.0000
Mean education^2*	0.0063	0.0005	0.0000	-0.0027	0.0005	0.0000
Mean age*	0.0449	0.0059	0.0000	-0.0109	0.0050	0.0290
Mean age^2*	-0.0004	0.0001	0.0000	0.0001	0.0001	0.2460
Women proportion*	0.1089	0.0046	0.0000	0.0906	0.0035	0.0000
Self-employed head (dummy)	-0.5784	0.0197	0.0000	-0.2812	0.0172	0.0000
Labor income of the head (if employee)	-0.0001	0.0000	0.0000	-0.0001	0.0000	0.0000
Constant	-4.8529	0.0858	0.0000	-5.1909	0.0728	0.0000
Occupied as self-employed versus unoccupied						
Education	0.1955	0.0132	0.0000	0.0854	0.0105	0.0000
Education^2	-0.0113	0.0010	0.0000	-0.0024	0.0007	0.0000
Age	0.3709	0.0068	0.0000	0.3475	0.0051	0.0000
Age^2	-0.0046	0.0001	0.0000	-0.0040	0.0001	0.0000
Gender	-1.8048	0.0424	0.0000	-1.3430	0.0298	0.0000
Number of members from 0 to 14*	-0.0434	0.0094	0.0000	-0.0280	0.0109	0.0100
Number of members from 14 to 65*	0.0598	0.0119	0.0000	0.0212	0.0114	0.0630
Number of members older than 65*	0.2445	0.0393	0.0000	-0.0341	0.0315	0.2790
Presence of other members from 14 to 65 (dummy)	0.1675	0.2281	0.4630	0.8982	0.1727	0.0000
Mean education*	-0.2615	0.0168	0.0000	-0.1139	0.0125	0.0000
Mean education^2*	0.0078	0.0011	0.0000	0.0008	0.0008	0.2940
Mean age*	0.0055	0.0109	0.6100	-0.0356	0.0079	0.0000
Mean age^2*	-0.0001	0.0001	0.6320	0.0004	0.0001	0.0000
Women proportion*	0.0580	0.0111	0.0000	0.0608	0.0060	0.0000
Self-employed head (dummy)	0.1720	0.0362	0.0000	0.5107	0.0264	0.0000

Labor income of the head (if employee)	-0.0001	0.0000	0.0000	-0.0001	0.0000	0.0000
Constant	-7.9520	0.1467	0.0000	-7.9046	0.1195	0.0000
Source: Based on "Pesquisa Nacional por Amostra de Domicílios" (PNAD) of the 1976 and 1996. Note: * excluding the person whose participation is estimated.						

For other members of the household, education did seem to raise the probability of choosing wage work vis-à-vis staying out of the labor force in both periods, with the relationship changing from concave to convex (and weak) over the period. It also significantly enhanced the probability of being in self employment vis-à-vis outside the labor force in both periods, and this relationship was concave throughout. The number of children in the household significantly discouraged participation in both sectors, although more so in the wage–earning one. This effect became stronger in that sector, and weaker for the self-employed.

The proportion of women in the household significantly increased the probability of participating in both sectors for other members, and in wage work for the head. This was true in both years. For other members, having a self-employed head decreased the probability of joining the wage sector, and increased that of entering self-employment. Wage income of the head was significantly and negatively associated with labor force participation by other members of the household, in both sectors and years, as expected. However, the quantitative effect seems surprisingly small. [9]

4. The Simulation Results.

Having estimated earnings and participation equations for both sectors of the model – wage-earners and the self-employed, we are now in a position to carry out the decompositions described in equations (9) and (10), which generate the 'joint price effect' and the 'joint labor supply effect', respectively. The simulations, as was discussed above, are carried out for the entire distribution (as in equations 7 and 8). However, the results are summarized below in Table 6, which reports mean household per capita income $\mu(y)$,

[9] Further discussion of participation decisions and occupational choices is postponed until the next version of this paper, in which Tables 4 and 5 will be more easily comparable.

two inequality indices (the Gini coefficient and the Theil T index), and the standard three members of the Foster-Greer-Throbecke class of poverty measures, P(α), α = 0, 1, 2, computed with respect to a monthly poverty line of R$60 (in 1996 prices). The results presented in Table 6 are based on version (3') and (1') of the model, using the individual earnings specification for the self-employed.

Table 6: Joint Price Effects and Joint Labor Supply Effects: Applying the 1996 coefficients on the 1976 population						
Simulation	μ(y)	Theil	Gini	P(0)	P(1)	P(2)
1976 original	267.23	0.85	0.60	0.23	0.09	0.05
Joint Price Effects: Applying 1996 Betas to 1976						
1996 wage-earners Betas	222.81	0.87	0.61	0.30	0.12	0.07
1996 self-employed Betas	252.87	0.87	0.61	0.25	0.10	0.05
1996: all Betas	208.48	0.88	0.62	0.32	0.13	0.07
Joint Labor Supply Effects: Applying 1996 Gamas to 1976						
1996: all gamas	265.31	0.87	0.61	0.24	0.11	0.07
Combined Effects: Applying all 1996 Betas and Gamas to 1976						
1996: all betas and gamas	226.66	0.89	0.62	0.30	0.13	0.08

Table 6 contains the three key results of this paper, all of which follow from the intuition derived from the regression estimations in the previous section. First, the joint price effect from 1976 to 1996 ($B_{tt'}$: see equation 9) was inequality augmenting. It raised both the Theil index and the Gini coefficient, both for wage-workers and for the self employed. When the combined effect of applying all 1996 betas to the 1976 distribution was tested, this had the 'worst' impact of all: mean household per capita income was at its lowest across simulations, the Theil and Gini were at their highest, as were all three poverty measures. Although this comes from changing all coefficients in equations (2) and (3'), we investigate their separate effects below, in tables 7 –12. As one would expect, from the fact that the age coefficients hardly changed from 1976 to 1996, the dominant effect is that of a steepening earnings education profile. We return to this below.

The joint labor-supply effect ($L_{tt'}$: see equation 10) was also inequality augmenting, although not as much as the joint price effect. In particular, it seemed to affect the incomes of the poor by less, since all three poverty measures rose by substantially less in this simulation. The overall combined effect ($B_{tt'} + L_{tt'}$), reported in the last row of table 6, is also inequality and poverty augmenting. In fact, any hope that the two effects might somehow 'cancel out' in the distributional dynamics underlying these summary statistics is dispelled by this last row, where both inequality and all three poverty measures are at their highest for the table.

Nevertheless, a comparison of that overall combined row, with the 'all betas' row suggests that the price effect was first order, with smaller inequality augmenting contributions from the labor supply effects. Once again, as suggested above, since the 'returns' to age and gender, as revealed in Tables 1 and 3, appeared to change much less than the shape of the returns to education, it would appear that the convexification of the earnings-education profile might be the strongest driving force behind the inequality augmenting effect of both price and labor-supply effects described so far. To investigate that, we carried out simulations on the individual betas, which are presented in tables 7-12.

Table 7: Inequality registered for simulated incomes - 1976

	Theil Coefficient	Gini Coefficient
1976 Observed	0.8574	0.6049
1996 Education coefficient	0.8920	0.6117
1996 Age coefficient	0.8619	0.6074
1996 Gender coefficient	0.8554	0.6057
1976 Fitted income	0.8619	0.6066

Source: Based on "Pesquisa Nacional por Amostra de Domicílios" (PNAD) of 1976.

Table 8: Inequality registered for simulated incomes - 1996

	Theil Coefficient	Gini Coefficient
1996 Observed	0.6856	0.5864
1976 Education coefficient	0.6610	0.5807
1976 Age coefficient	0.6874	0.5869
1976 Gender coefficient	0.6937	0.5887
1996 Fitted income	0.6885	0.5875

Source: Based on "Pesquisa Nacional por Amostra de Domicílios" (PNAD) of 1996.

Table 9: Poverty registered for simulated incomes - 1976 Poverty line of R$30,00			
	MP0	MP1	MP2
1976 Observed	0.0713	0.0243	0.0136
1996 Education coefficient	0.0833	0.0297	0.0174
1996 Age coefficient	0.0710	0.0257	0.0157
1996 Gender coefficient	0.0712	0.0256	0.0155
1976 Fitted income	0.0747	0.0270	0.0162
Source: Based on "Pesquisa Nacional por Amostra de Domicílios" (PNAD) of 1976.			

Table 10: Poverty registered for simulated incomes - 1996 Poverty line of R$30,00			
	MP0	MP1	MP2
1996 Observed	0.0889	0.0465	0.0371
1976 Education coefficient	0.0790	0.0449	0.0368
1976 Age coefficient	0.0903	0.0483	0.0384
1976 Gender coefficient	0.0938	0.0495	0.0391
1996 Fitted income	0.0895	0.0474	0.0379
Source: Based on "Pesquisa Nacional por Amostra de Domicílios" (PNAD) of 1996.			

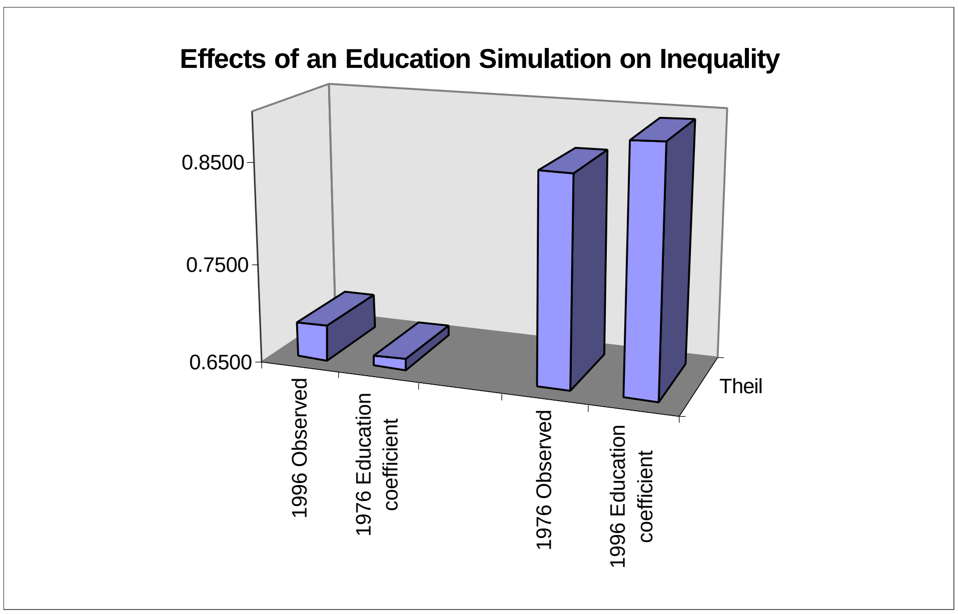
Table 11: Poverty registered for simulated incomes - 1976 Poverty line of R$60,00			
	MP0	MP1	MP2
1976 Observed	0.2290	0.0880	0.0468
1996 Education coefficient	0.2610	0.1013	0.0549
1996 Age coefficient	0.2269	0.0877	0.0476
1996 Gender coefficient	0.2248	0.0875	0.0474
1976 Fitted income	0.2331	0.0913	0.0497
Source: Based on "Pesquisa Nacional por Amostra de Domicílios" (PNAD) of 1976.			

Table 12: Poverty registered for simulated incomes - 1996 Poverty line of R$60,00			
	MP0	MP1	MP2
1996 Observed	0.2223	0.0973	0.0641
1976 Education coefficient	0.1941	0.0895	0.0604
1976 Age coefficient	0.2253	0.1009	0.0665
1976 Gender coefficient	0.2315	0.1038	0.0684

1996 Fitted income	0.2221	0.0982	0.0651
Source: Based on "Pesquisa Nacional por Amostra de Domicílios" (PNAD) of 1996.			

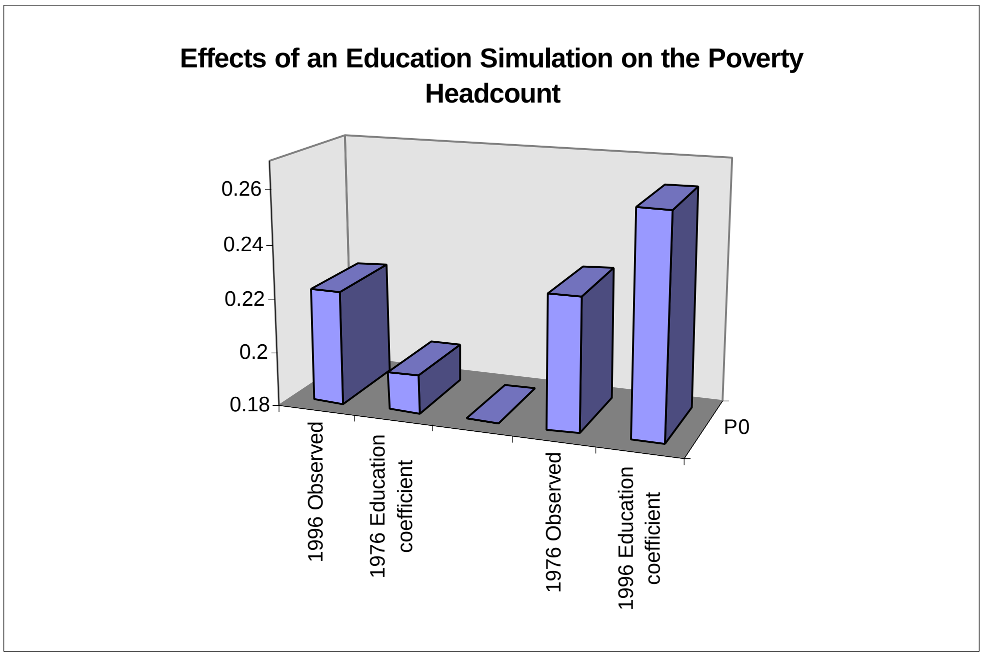
Table 7 fails to falsify our hypothesis: it is indeed the 1996 coefficient of education which is responsible for the largest increase in the 1976 Theil and Gini indices, with age and gender having a much smaller impact. This is confirmed by the converse simulation, reported in Table 8: applying the 1976 education coefficients on the 1996 model. As expected, this leads to a decline in inequality from the actual level observed in 1996. And once again, it is the returns on education which have the largest impact. These impacts of the education simulation on the Theil index are highlighted in Figure 1 below.

Figure 1: Observed Inequality Fell between 1976 and 1996. But he 1996 Education-Earnings Profile would have led to More Inequality…



Tables 9 and 10 consider the impact of the same individual price effect simulations on poverty measures, with respect to a R$30 per month poverty line. Poverty does increase with this exercise, suggesting that the lower returns to very low levels of education ruling in 1996 (see Tables 1 and 3') would have caused an increase in poverty in the 1976 population, all else constant. This is confirmed for the higher poverty line of R$60 per month. Again, all three measures of poverty rise when the 1996 education coefficient is applied to the 1976 model, and the change is far more significant than for any of the other individual β simulations. The converse simulations (applying 1976 individual coefficients to the 1996 model) result in poverty declines, as expected: the higher linear component of the earnings-education profile in the mid-70s would have left fewer people in poverty, given the educational composition (and that of other observable and unobservable characteristics) of the 1996 population. These effects are highlighted in Figure 2 below:

Figure 2: Observed Poverty also Fell Slightly between 1976 and 1996. But the effects of the change in the Earnings-Education Profile would have led to an increase…



Now, we know from the exact decomposition in equation (11) : $D_t - D_{t'} = B_{tt'} + L_{tt'} + R_{tt'}$ that the actual change in the distribution between t and t' (e.g. 1976 and 1996) must be the sum of the joint price effect, the joint labor supply effect, and the residual effect, which encompasses changes in the joint distribution of observed (e.g. education) and unobserved (e.g. ability) characteristics. The last row in Table 6 contains summary statistics for both poverty and inequality, from the distribution arising from adding $B_{tt'}$ and $L_{tt'}$ to $D_{t'}$. The fact that the combined application of these effects to the 1976 distribution would have caused inequality (and poverty) to rise whereas, in fact, we observed a decline in inequality (and also, albeit less pronounced, in poverty) points by necessity to the fact that the residual term $R_{tt'}$ must be inequality-reducing, and to an extent capable of more than offsetting the price and labor supply effects. That is to say, changes in the joint distribution of population characteristics, observable and unobservable, between 1976 and 1996, must have more than compensated for the inequality augmenting tendencies which arose, as we have seen, largely from a convexification of the earnings-education profile, with lower returns accruing to the least educated, and steeper increases associated with higher levels of education.

At this stage, it must be left only as a hypothesis that the change in that joint distribution of characteristics which will account for the bulk of the equalizing residual term is a change in the educational composition of the population, with population mass moving up the steepening mountain… To conduct the simulation which will be able to test that hypothesis, estimated changes in each element of the joint distribution must control for changes in other exogenous variables. Regressions which prepare the path towards that next stage of the investigation have already been run, and are reported in the Appendix. [10] The simulations of changes in specific elements of the vector of personal characteristics X_I^P will be reported in the next version of this paper.

5. Conclusions

This paper presented a simple two-sector model of household income determination and occupational choice for urban Brazil, and relied on a micro-simulation-based decomposition of changes in the distribution of household per capita incomes in order to understand what were the factors driving the observed decline in inequality (and in poverty, to a lesser extent) over the period. The main apparent puzzle uncovered by the estimation of earnings regressions for both wage-earners and self-employed workers was that the change in the returns to education seemed to be inequality and poverty augmenting. This was confirmed at the simulation stage, when imposing 1996 education coefficients on the 1976 population led to an increase in simulated inequality.

The effects of other price and labor supply effects was also investigated, and seemed to reinforce the inequality-augmenting effect of the education price effect, as revealed by Tables 7-12 and, in particular, by the joint simulations in Table 6. Given the exact decomposition in equation 11, the observed decline in inequality (and to a lesser extent in poverty) between 1976 and 1996 must be due to changes in the joint distribution of observed and unobserved characteristics of the population. The hypothesis with which we close this stage of the investigation, and which sets the scene for the next version of this study, is that the main such equalizing change is the change in the educational composition of the population, with population mass moving along the education scale, just the returns to it become steeper.

[10] The four regressions A1-A4 are included merely for the sake of completeness. They are not particularly informative in themselves, and are mainly preparatory steps for the simulations of changes in elements of X_I^P. These will be reported in the next version of this paper.

Appendix.

Table 1: Dependent variable: Education*						
	Year					
	1976			1996		
	Coefficient	Standard	P-value	Coefficient	Standard	P-value
Age	0.3058	0.0021	0.0001	0.2263	0.0019	0.0001
Age 2	-0.0038	0.0000	0.0001	-0.0032	0.0000	0.0001
Gender	-0.2062	0.0178	0.0001	0.1954	0.0169	0.0001
North region	-0.7175	0.0530	0.0001	-1.0917	0.0380	0.0001
Northeast region	-1.1573	0.0228	0.0001	-1.3717	0.0213	0.0001
West-center region	-0.7372	0.0427	0.0001	-0.5689	0.0344	0.0001
South region	0.0879	0.0257	0.0006	-0.1519	0.0249	0.0001

Source: Based on "Pesquisa Nacional por Amostra de Domicílios" (PNAD) of the 1976 and 1996.
Note: * People older than 10 years

Table 2: Dependent variable: Total members of households younger than 14 years						
	Year					
	1976			1996		
	Coefficient	Standard	P-value	Coefficient	Standard	P-value
schooling of the head	-0.0806	0.0045	0.0001	-0.0413	0.0028	0.0001
schooling of the head 2	0.0012	0.0003	0.0001	0.0007	0.0002	0.0001
age of the head	0.1116	0.0024	0.0001	0.0086	0.0013	0.0001
age of the head 2	-0.0015	0.0000	0.0001	-0.0004	0.0000	0.0001
North region	0.6456	0.0376	0.0001	0.3680	0.0175	0.0001
Northeast region	0.4695	0.0163	0.0001	0.2304	0.0097	0.0001
West-center region	0.3616	0.0304	0.0001	0.0467	0.0152	0.0021
South region	0.0565	0.0178	0.0015	-0.0045	0.0108	0.6773

Source: Based on "Pesquisa Nacional por Amostra de Domicílios" (PNAD) of the 1976 and 1996.

Table 3: Dependent variable: Total members of households with age between 14 to 65 years						
	Year					
	1976			1996		
	Coefficient	Standard	P-value	Coefficient	Standard	P-value
schooling of the head	0.0212	0.0040	0.0001	0.0049	0.0031	0.1114
schooling of the head 2	-0.0027	0.0003	0.0001	-0.0018	0.0002	0.0001
age of the head	0.2522	0.0021	0.0001	0.2046	0.0015	0.0001
age of the head 2	-0.0026	0.0000	0.0001	-0.0021	0.0000	0.0001
North region	0.1230	0.0336	0.0003	0.1963	0.0194	0.0001
Northeast region	0.0045	0.0145	0.7555	0.1167	0.0108	0.0001
West-center region	0.0869	0.0271	0.0014	0.0329	0.0169	0.0513
South region	-0.0124	0.0159	0.4334	-0.1055	0.0120	0.0001

Source: Based on "Pesquisa Nacional por Amostra de Domicílios" (PNAD) of the 1976 and 1996.

Table 4: Dependent variable: Total members of households older than 65 years						
	Year					
	1976			1996		
	Coefficient	Standard	P-value	Coefficient	Standard	P-value
schooling of the head	0.0053	0.0009	0.0001	0.0053	0.0009	0.0001
schooling of the head 2	-0.0002	0.0001	0.0018	-0.0001	0.0001	0.0325
age of the head	-0.0568	0.0005	0.0001	-0.0525	0.0004	0.0001
age of the head 2	0.0008	0.0000	0.0001	0.0008	0.0000	0.0001
North region	0.0025	0.0079	0.7529	0.0016	0.0055	0.7761
Northeast region	0.0046	0.0034	0.1781	-0.0020	0.0030	0.5128
West-center region	-0.0148	0.0064	0.0211	-0.0143	0.0048	0.0026
South region	-0.0085	0.0038	0.0239	-0.0039	0.0034	0.2454
Source: Based on "Pesquisa Nacional por Amostra de Domicílios" (PNAD) of the 1976 and 1996.						

References.

Bourguignon, F., F. Ferreira and N. Lustig (1998): "The Microeconomics of Income Distribution Dynamics in East Asia and Latin America", World Bank Research Proposal, (Washington, DC: World Bank, April).

Juhn, C., K. Murphy and B. Pierce (1993): "Wage Inequality and the Rise in Returns to Skill", *Journal of Political Economy*, **101** (3), pp. 410-442.

Mookherjee, D. and A.F. Shorrocks (1982): "A Decomposition Analysis of the Trend in UK Income Inequality", *Economic Journal*, **92**, pp. 886-902.