Ibero-Amerika Institut für Wirtschaftsforschung
Instituto Ibero-Americano de Investigaciones Económicas
Ibero-America Institute for Economic Research
(IAI)

Georg-August-Universität Göttingen
(founded in 1737)

Diskussionsbeiträge · Documentos de Trabajo · Discussion Papers

Nr. 133

Inequality of Opportunity in Brazil

François Bourguignon, Francisco H.G. Ferreira,
Marta Menéndez

November 2005

Platz der Göttinger Sieben 3 · 37073 Goettingen · Germany · Phone: +49-(0)551-398172 · Fax: +49-(0)551-398173

e-mail: uwia@gwdg.de · http://www.iai.wiwi.uni-goettingen.de

Inequality of Opportunity in Brazil

François Bourguignon, Francisco H.G. Ferreira and Marta Menéndez[*]

Keywords: Inequality of opportunity, earnings inequality, intergenerational mobility.

JEL Codes: D31, D63, J62

Abstract: This paper proposes a method to decompose earnings inequality into a component due to unequal opportunities and a residual term. Drawing on the distinction between 'circumstance' and 'effort' variables in John Roemer's work on equality of opportunity, we associate inequality of opportunities with the inequality attributable to circumstances which lie beyond the control of the individual – such as her family background, her race and the region where she was born. We interpret the decomposition as establishing a lower bound on the contribution of opportunities to earnings inequality. We further decompose the effect of opportunities into a direct effect on earnings and an indirect component which works through the "effort" variables. The decomposition is applied to the distributions of male and female earnings in Brazil, in 1996. While the residual term is large, observed circumstances nevertheless account for around a quarter of the value of the Theil index. Parental education is by far the most important circumstance affecting earnings, dwarfing the effects of race and place of birth.

[*] Bourguignon and Ferreira are with The World Bank; Menendez is at the Université Paris Dauphine. Correspondence to Francisco H.G. Ferreira, The World Bank, 1818 H. St. NW, Washington, DC, 20433, USA. Fax: (202) 522-1151. E-mail: fferreira@worldbank.org. The views expressed in this paper are solely those of the authors, and should not be attributed to The World Bank; its Board of Directors; or the countries they represent. We would like to thank Tony Atkinson, Richard Blundell, Sam Bowles, Pedro Carneiro, Jishnu Das, Quy-Toan Do, Hanan Jacoby, Yoko Kijima, Phillippe Leite, Costas Meghir, Thomas Piketty, Martin Ravallion, participants at various seminars, and two anonymous referees for helpful comments. All remaining errors are ours.

1. Introduction

Income inequality has many sources, not all of which are equally objectionable. A reasonable distinction, which is often made informally, is that inequality in the opportunities available to people – their basic life chances, determined by factors such as their race or gender, the families they are born to or the countries they are born in, and perhaps even their endowments of natural talent – is more objectionable than inequalities which arise because of the differential application of individual effort.[1] John Roemer (1998) offered one formalization of the concept of unequal opportunities across different types of individuals, suggesting that one should separate the determinants of a person's earnings into circumstance variables (those which are exogenous to the person) and effort variables (which she can influence). He suggested a precise definition of an equal-opportunity policy, which would be to equalize "advantage" (which one might somewhat coarsely proxy for by income) for each centile of the effort distribution, across individual types.[2]

Practical applications of this approach have remained scarce, however, and this is in part because the analysis becomes rather cumbersome as the number of "types" increases. In fact, although various other authors in the fields of ethics and social choice have recently argued that opportunity, rather than income or other observable outcomes, should become the "currency of egalitarian justice",[3] empirical usage of the concept has remained rare. Basically, this is because economists do not know how to measure inequality of opportunity.

This paper proposes a simple method to quantify the degree of inequality of opportunity associated with an empirical distribution of incomes or earnings. The basic approach is to divide all observed earnings determinants (available from a household or labor force survey, say) into those which can be regarded as exogenous to the individual, in the sense that they can not be influenced by her actions, and all others. Following Roemer, we refer to the first set of variables

[1] One recent paper argued that "according to the opportunity egalitarian ethics, economic inequalities due to factors beyond the individual responsibility are inequitable and to be compensated by society, whereas inequalities due to personal responsibility are equitable and not to be compensated." (Peragine, 2004, p.11).
[2] Roemer (1998, p.27) recognizes that there may be no single policy which does this for every centile, and proposes an averaging of "indirect advantage functions" across centiles to generate a well-defined maximand.
[3] Notably Arneson (1989), Cohen (1989) and Dworkin (1981). Cohen argues that Amartya Sen's capability approach is not too far removed from the concept of opportunities either.

– which could include sex, race, place of birth, family wealth, parental education, or family background more generally - as “circumstance” variables. Our approach is simply to simulate the reduction in earnings inequality which would attain if differences in circumstance variables were eliminated. This difference between observed and counterfactual inequality is interpreted as a lower bound for inequality of opportunities, for reasons which we discuss in below. Our main challenge is to take account of the fact that other earnings determinants – such as one’s own level of education or position in the labor market – are endogenous, since they are also influenced by those same circumstances, and to address the identification problems which this entails for the estimation of the relevant empirical model.

We apply this approach to the distribution of hourly earnings in urban Brazil, separately for males and females, exploiting the fact that the 1996 Brazilian household survey includes information on parental education and occupation. Our results suggest that over a fifth, and for some cohorts almost a third, of observed earnings inequality can be attributed to a set of only four exogenous circumstance variables, namely race, place of birth, parental education and father’s occupation. To the extent that other circumstances – such as family wealth, primary school quality, etc – are uncorrelated with those determinants, these estimates should be seen as a lower bound on inequality of opportunity. We also divide the sample into seven age groups, and investigate the evolution of both the opportunity and the residual components of inequality across cohorts. We find some evidence of a moderate decline in the opportunity share of inequality. Finally, we investigate the relative importance of each individual circumstance variable separately, and find that parental education dwarfs all other circumstance variables, including race, in importance.

The paper is organized as follows. The next section briefly reviews the related empirical literature. Section 3 describes our data set. Section 4 lays out our empirical approach and formally defines the decomposition of earnings inequality into a (lower bound) component attributable to unequal opportunities and a residual. Section 5 discusses our estimation strategy and the bounds analysis used for identification. Section 6 presents the estimation results, and Section 7 summarizes the results of the final decompositions. Section 8 briefly concludes.

2. Empirical Approaches to Inequality of Opportunity: the literature.

While we are not aware of other papers which seek to measure the share of earnings inequality attributable to unequal opportunities, there is a (small) empirical literature which aims to quantify the costs and effects of implementing Roemer's "equal opportunity policy" in different contexts. Focusing on race and parental education as determinants of opportunities in the United States, Betts and Roemer (1999) ask what reallocation of educational expenditures would equalize opportunities across four types of individuals in the US (using the National Longitudinal Survey of Young Men). Interestingly, they find that race is a more important partitioning variable than parental education in their sample. In a related study, Page and Roemer (2001) ask to what extent the United States fiscal system can be seen as an "opportunity equalizing device". These authors also focus on race and parental education (which they interpret as a proxy for socioeconomic background) as the key circumstance variables, and find that the US tax system does contribute to an equalization of opportunities (as compared to the pre-tax earnings distribution) across socio-economic groups, but much less so across racial groups.[4]

These studies differ from ours in two ways: first, they seek to assess specific (actual or counterfactual) policies with respect to their opportunity-equalizing impact, whereas we ask what share of observed overall inequality is due to unequal opportunities. Second, by virtue of Roemer's definition of an equal-opportunity policy, they are restricted to differences across a few (usually four) large groups, determined by a very limited set of circumstances. As we will see, our decomposition allows for a broader treatment of circumstances.

Given the importance of family background (and, more specifically, parental education) as a key circumstance variable, this paper is also related to the much larger empirical literature on intergenerational mobility. This literature has a long and distinguished tradition, dating back at least to Bowles (1972), and Behrman and Taubman (1976). Much of it focuses on the intergenerational elasticity of some measure of economic status, estimated as the coefficient β in the Galtonian regression: $\ln y = \alpha + \beta \ln y_{-1} + \varepsilon$, where y denotes the measure of economic

[4] A larger group of authors have extended this analysis of fiscal systems as opportunity-equalizers to other countries in Roemer et. al. (forthcoming).

status of interest, and y_{-1} measures the same variable for a person's parent(s). This elasticity measures the degree of transmission of economic status across generations, and is thus interpreted as a measure of persistence of its inequality. Its complement, 1- β, is often interpreted as a measure of intergenerational mobility.[5]

A variety of economic status variables have been used, including wages (from the PSID and the NLS data sets for the US, as well as for a number of European countries); total earnings; family income; family wealth; family consumption; and years of schooling. Results obviously vary by indicator and by country. Two excellent recent surveys of this literature exist, in Solon (1999) and Mulligan (1999), and we do not replicate their work here.[6]

Two recent papers have estimated such Galtonian regressions for Brazil, using exactly the same data set as we use in this paper. Dunn (2003) considers only males aged 25-34, and instruments for father's earnings using father's education. He finds an elasticity of 0.69, "higher than in any country previously studied" (Dunn, 2003, p.1).[7] Sérgio Ferreira and Veloso (2004) also estimate Brazilian intergenerational elasticities for various cohorts, using parental occupation and education as instruments, but their procedure involves first estimating "parent" earnings equations on earlier PNAD samples, and then using the estimated coefficients to predict earnings for the parents of the workers in the 1996 sample.[8] For the middle cohorts (i.e. excluding the oldest and the youngest cohorts) they find intergenerational elasticities in the 0.50-0.68 range.

---

[5] Behrman and Taubman (1976) took a different route. They used information on white male twins to estimate the contribution of a genetic component (and of two separate environment components) to the variance in four different measures of adult individual achievement.

[6] It is worth noting, though, that since those surveys were published, some recent findings for the United States have challenged the 1990s consensus that the intergenerational elasticity of earnings in that country was of the order of 0.4 (see, e.g. Solon, 1992). Although this figure was already seen as indicating lower intergenerational mobility than previously thought, more recent analysis suggests an even higher elasticity – and hence lower mobility. Much of the variation hinges on how one averages out transitory components and measurement errors associated with earnings observations at a single point in time. While Solon (1992) already used an average of earnings across periods, Mazumder (2005) uses longer earnings histories, and estimates an elasticity of closer to 0.6. See also Bowles and Gintis (2002).

[7] Although he acknowledges that "If fathers' educations are independently positively correlated with sons' earnings, then the IV elasticity estimate will be upwards-inconsistent."(p.5). He correctly treats his estimate as an upper-bound.

[8] Their approach is essentially a two-sample instrumental variable procedure. See Angrist and Krueger (1992).

While these papers are related and complementary to our study, they have different objectives. They seek to understand intergenerational mobility in one specific economic indicator, in this case earnings. We are interested in quantifying the share of current earnings inequality which is attributable to inequality of opportunities, using Roemer's concept of circumstance variables. The point of contact between the two approaches is that parental education (and occupation) are important circumstances, and account for a substantial share of current inequality. But we also consider other circumstances observed in the data set, chiefly race and region of birth, and quantify their importance. We further identify two separate channels for the impact of parental education (and other circumstances) on current earnings: a direct impact and an indirect effect through the child's own schooling, migration decisions, and insertion in the labor market. This is clearly not possible when, as in Dunn (2003), parental education is used as an instrument for parental earnings. The two approaches are ultimately best seen as complements, rather than as substitutes.

Another branch of the intergenerational mobility literature has focused directly on educational transmission. Some papers estimate the part of schooling inequality which is explained by the characteristics of parents, which they take to quantify the inequality of opportunities, whereas the remainder is attributed to heterogeneous individual efforts. This approach has been followed by Behrman, Birdsall and Székely (2000) for Latin American countries. Barros and Lam (1993), and Lam (1999) applied similar methods to Brazil. Given the nature of their data sets, however, those studies focused on expected *future* mobility or, in other words, the relation between the *potential* earnings of children when they become adults, and the earnings of their parents. The approach was not suitable for disentangling the *actual* share of the inequality of opportunity in today's level of overall (outcome) inequality.

3. The Data

Our data comes from the 1996 wave of the Pesquisa Nacional por Amostra de Domicílios (PNAD – National Household Survey), which is conducted annually by the Instituto Brasileiro de Geografia e Estatítica (IBGE), Brazil's Census Bureau.[9] The survey is nationally

[9] The PNAD is an annual survey, but is not fielded on Census years.

representative, except for the rural areas of the Northern Region. This exception does not affect our analysis, which is restricted to urban areas because of the general imprecision of earnings and income measurement in rural areas.[10] The sample is also restricted to individuals 26 to 60 years old, in an effort to concentrate on individuals who had finished schooling and were active in the labor market. We study the 1996 survey because, on that year, information is available on the education and occupation of the parents of all surveyed household heads and spouses.[11] The complete PNAD 1996 sample size is upwards of 330,000 individuals. After excluding individuals living in rural areas; those outside the 26-60 age range; those who are not household heads or their spouses; and those with missing entries for the relevant variables, we are left with a sample of 48,392 active adults. We further excluded 179 adults who reported being non-remunerated workers in family enterprises.[12]

This sample was then divided into seven 5-year birth cohorts - from individuals born between 1936-40 up to those born between 1966-70. This allows us not only to measure the role of the inequality of opportunities in shaping the inequality of observed earnings at a point in time, but also to study how this role may vary across cohorts. An important question is indeed whether the increase in the educational level of successive cohorts was accompanied by more or less inequality of opportunities, or whether it corresponded to a uniform upward shift in schooling achievements, with constant inequality of opportunities. Comparing various cohorts observed at a single point in time allows us to shed some light on this issue.

Our key dependent variable is current individual earnings, measured as "real hourly earnings from all occupations". Other variables used in the analysis include a race dummy variable, parental education expressed as the number of years of schooling [13], the occupational position of

[10] See Ferreira, Lanjouw and Neri (2003) for a discussion of the shortcomings of the rural income data in the PNAD.

[11] This information is not generally collected in the PNAD, but was also available in both the 1982 and the 1988 surveys. Earnings in the 1982 survey were collected with respect to a different reference period (of three months) and are therefore not comparable with other years. The 1988 survey was used to check the robustness of some of the results reported in this paper.

[12] To control for the effect of the productivity of these workers on the reported earnings of their household heads, we included a dummy variable for the latter in the earnings equation discussed in the next section. The coefficient on this dummy variable was statistically insignificant, and it was subsequently dropped.

[13] Parental education is given in discrete levels. They were converted into years of schooling (here in brackets) using the following rule. No school or incomplete 1st grade (0); incomplete elementary (2); complete elementary, or complete 4th grade (4); incomplete 1st cycle of secondary or 5th to 7th grade (6); complete 1st cycle of secondary or

the father[14] - a 9-level categorical variable- and dummies for the regions of origin. We also use the individual's own schooling attainment, measured in years[15]; years of schooling squared, to capture possible non-linearities, and a migration dummy, defined as whether the observed municipality of residence is different from the one where the individual was born. Finally, a categorical variable for labor market status is included, which indicates whether the worker is a formal employee ("com carteira"); an informal employee ("sem carteira"); self-employed ("conta própria"); or an employer. Descriptive statistics of the main variables are shown in Table 1.

4. The Decomposition

We are interested in estimating the share of observed inequality in current earnings which can be attributed to inequality of opportunity. Loosely following John Roemer, we associate opportunity with the impact on earnings of variables which are completely independent from individual effort, and which we call "circumstances". Our main goal is to estimate the reduction in inequality which would attain if there were no differences in people's circumstances. It is this reduction which we will take to measure the contribution of inequality in opportunity to observed earnings inequality.

We will also follow Roemer in calling the other variables which help determine earnings – i.e. those which affect earnings but can be affected by individual decisions – as "efforts". We will keep inverted commas on that term throughout, however, to acknowledge that – besides being explicitly influenced by observed circumstances - they may also capture the effects of unobserved circumstances with which they are correlated. Denoting earnings by w, circumstance variables by the vector C; "effort" variables by the vector E; and other, unobserved determinants by u, one can write the earnings function most generally as:

---

complete 8th grade (8); incomplete 2nd cycle (9.5); complete 2nd cycle of secondary (11); incomplete superior (13); complete superior (15); master or doctorate (17).

[14] The 9-level occupational categories are: rural workers (1); domestic servants (2); traditional sector workers (3); service sector workers (4); modern industry workers (5); self-employed shopkeepers (6); technicians, artists and desk workers (7); employers (8); liberal professionals (9). This classification is borrowed from Brazilian sociological studies on occupational mobility (see Pero, 2001; Valle e Silva, 1978)

[15] The number of years of schooling directly provided in the PNAD is bounded at 15. For consistency with the scale used for parents' schooling, this variable was changed to 17 for individuals reporting a "master or doctorate" degree. The distinction between these two levels is not made in the data and, although a doctorate is likely to take at least 20 years to complete, a Masters degree is likely to be more common. In any case, this affects less than 1% of the sample.

$$w_i = f(C_i, E_i, u_i) \quad (1)$$

Noting that circumstance variables are exogenous by definition, but that "effort" variables can be affected by circumstances, as well as by unobserved determinants, implies that :

$$w_i = f(C_i, E_i\,(C_i, v_i), u_i) \quad (2)$$

For the purposes of empirical estimation, we could log-linearize (2) to obtain a system :

$$Ln(w_i) = C_i.\alpha + E_i.\beta + u_i \quad , \quad u_i \perp C_i \quad (3)$$

$$E_i = BC_i + v_i \quad , \quad v_i \perp C_i \quad (4)$$

where w_i denotes current hourly earnings, α and β are two vectors of coefficients and u_i is an i.i.d. random variable with zero mean, that accounts for unobserved circumstance and effort variables; sheer luck; and measurement error. If one wished to interpret current wages, w_i, as a proxy for permanent income or 'economic status', then u_i would also include transitory income shocks. B is a matrix of coefficients linking the circumstance variables to the "effort" variables. This matrix explicitly allows for the fact that some of these "effort" variables are clearly affected by circumstances. Formal schooling, for example, is determined at least in part by family background. This effect of parental background on the educational outcomes of the next generation may occur because more educated parents provide more "home inputs" into an "education production function", such as books, vocabulary and quality time spent on homework, but it may also reflect individual learning about the returns to effort, which may themselves depend on the circumstances – and indeed on the previous mobility history – of the family.[16] v_i is another white-noise disturbance term, orthogonal to the vector C, as indicated.[17]

---

[16] Hanushek (1986) reviews the literature on the first, 'production function' view of the impact of family characteristics on individual schooling outcomes. Piketty (1995) models the second type of channel, where individual beliefs about opportunities and mobility are formed on the basis of the experience of their own lineages, and in turn rationally determine their own effort levels. Our empirical model allows for both types of indirect effects of circumstances (and past history) on earnings *through* efforts.

[17] One could substitute (4) into (3) and obtain an even more reduced-form equation, $Ln(w_i) = C_i.(\alpha + \beta.B) + v_i\,\beta + u_i$. While running our main decomposition on the basis of this specification is both possible and econometrically simpler, it would not then be possible to distinguish the direct (α) from the indirect (βB) effect of circumstances on earnings inequality, as we do.

If the system (3)-(4) could be sensibly estimated (which will be discussed in the next section), one could define the opportunity share of earnings inequality as:

$$\Theta_I := \frac{I(w) - I\left(w \middle| C_i = \overline{C}, \forall i\right)}{I(w)} = \frac{I(w) - I(\tilde{w})}{I(w)} \tag{5}$$

Where I(w) is a particular inequality index defined over earnings, and $\overline{C}$ is a constant vector, the elements of which could, for instance, take the value of the cross-sectional sample mean of each element of the vector of circumstances. In practical terms, Θ_I can be estimated by simulating a vector $\tilde{w}$ given by:

$$\begin{aligned} Ln(\tilde{w}_i) &= \overline{C}\hat{\alpha} + \tilde{E}_i\hat{\beta} + \hat{u}_i \\ \tilde{E}_i &= \hat{B}\overline{C} + \hat{v}_i \end{aligned} \tag{6}$$

where $\hat{B}$, $\hat{\alpha}$ and $\hat{\beta}$ are the estimates obtained for the parameters in (3) and (4) – in a manner which will soon be described – and $\hat{u}_i$ and $\hat{v}_i$ are the estimation residuals for each individual.

One can further decompose the contribution of opportunities to observed inequality into a direct and an indirect effect. Define the direct effect as:

$$\Theta_I^d := \left[I(w) - I\left(w^p\right)\right]\left(I(w)\right)^{-1}, \quad \text{where } Ln\left(w_i^p\right) = \overline{C}\hat{\alpha} + E_i\hat{\beta} + \hat{u}_i \tag{7}$$

The indirect effect then is simply given by $\Theta_I^i = \Theta_I - \Theta_I^d$. According to this definition, the direct effect of opportunities on earnings is the impact of circumstance variables controlling for "effort" variables, but ignoring any effect *through* the latter. The indirect effect is the effect of circumstances on earnings through observed "efforts".

Before discussing the estimation of (3)-(4), we should acknowledge that this system of equations, like any parametric model, is somewhat restrictive. It relies on certain functional form assumptions, which will be described in more detail below. Additionally, to simplify the micro-simulation procedures used to generate the counterfactual earnings distributions {$\tilde{w}$} and {w^p}, we have chosen to be parsimonious in our specification, and have omitted various possible interaction terms between circumstance and effort variables in (3). This is not entirely innocuous, in that it restricts returns to “effort” variables, such as schooling, not to vary with circumstances. This may be an unrealistic assumption which it would be interesting to test, but that remains for further research.

5. Estimation Strategy

To carry out the decompositions proposed in (5) and (7), one needs to obtain meaningful estimates for the parameters in the system (3) – (4). Based on the information available in the PNAD 1996 data set, we observe four basic circumstance variables: race (R), parental schooling (represented as mean parental schooling, MPE, and the difference between the mother’s and the father’s schooling, DPE), region of birth (GR) and father’s occupational status (FO). So C = (R, GR, MPE, DPE, FO). Our “effort” variables are years of schooling (S), years of schooling squared (S^2), a migration dummy (M) and labor market status (L). These variables were described in Section 3. The model (3)-(4) can thus be rewritten as:

$$Ln(w_i) = \alpha_0 + R_i\alpha_R + GR_i.\alpha_G + MPE_i.\alpha_P + DPE_i.\alpha_D + FO_i.\alpha_F + S_i.\beta_S + S_i^2.\beta_{S2} + M_i.\beta_M + L_i\beta_L + u_i \tag{8}$$

$$S_i = b_0 + R_i b_R + GR_i.b_G + MPE_i.b_P + DPE_i.b_D + FO_i.b_F + v_i \tag{9}$$

$$\Pr\left(M = 1 \middle| R, GR, MPE, DPE, FO\right) = \Pr(\xi > -c'\left(R, GR, MPE, DPE, FO\right) \tag{10}$$

$$\Pr\{L = s\} = \Pr\{U^s = Cd_s + \varepsilon^s \geq U^k = Cd_k + \varepsilon^k, \forall k \neq s\} = P^s(C_i, d) = \frac{e^{C_i d_s}}{e^{C_i d_s} + \sum_{k \neq s} e^{C_i d_k}} \tag{11}$$

Equation (8) is simply the linear earnings equation (3) written out in full. Equations (9), (10) and (11) now replace equation (4). (9) is a linear schooling equation which, since it includes parental education on the right-hand side, could be seen as an extended Galtonian equation in years of schooling. Although (4) was written as a linear system, our two remaining “effort” variables[18] are discrete, and are thus estimated by a probit model (for migration, through equation 10, where

[18] No separate equation is estimated for S^2, given the similarity with the linear schooling variable.

ξ is assumed to be normally distributed with mean zero) and by a multinomial logit model for selection of labor market status (11).[19]

As before, we assume that u is orthogonal to C. In a sense, we are therefore effectively treating as exogenous circumstances not only parental education, say, but also the components of unobserved determinants which are correlated with it, such as parental wealth or intergenerationally correlated ability.[20] To the extent that these unobserved correlates of the exogenous variables in C are themselves also exogenous circumstances which affect earnings, we do in fact want to capture them in our decomposition, and they 'belong' in C. The overall impact of circumstances, thus redefined, will be truthfully described by the estimates of α. In other words, if parental wealth or income is not observed, the estimate of α will account for *both* the effect on children's earnings of the variables in *C and* the effect of that part of parental wealth or income which is correlated with elements of C.

The real problem arises because unobservables (u) in the earnings equation cannot be assumed to be independent of the "effort" variables (E). Again, imagine that parental wealth has a direct impact on either the schooling or the current earnings of their children (or both), independently of the child's own education. This correlation between u and E, or u and v, introduces bias in the estimation of the (α,β) coefficients and therefore in the decomposition of the total inequality into circumstance and effort components. Similar arguments apply if u is not independent from ξ or ε, in (10) or (11).

One way out of this difficulty would be to observe instrumental variables, Z that would influence efforts but not earnings. Equation (4) would then be replaced by:

[19] The assumptions underlying the use of the multinomial logit are that occupational choice across the different employment alternatives can be modeled using a linear utility function of the circumstance variables, with unobserved determinants (ε) which follow a Weibull distribution. See McFadden (1974).

[20] There is a recent literature that has tried to disentangle the "true" effect of some of the observed circumstance variables included here, such as family background. Behrman and Rosenzweig (2002), for instance, have pointed out that the impact of women's schooling on children's schooling may be overstated because of intergenerationally correlated genetic ability and motivation endowments. For our purposes, however, we want to include these latter influences along with the narrowly defined effect of mother's schooling, as a broader measure of family circumstances.

$$E_i = BC_i + DZ_i + v_i \tag{12}$$

with the vector Z_i being orthogonal to both u_i and v_i. Then instrumenting the effort variables in (3) through (12) would yield an unbiased estimator of (α,β) and then an unbiased decomposition of total inequality into inequality of observed opportunities, or circumstances, and inequality of efforts. Instruments have, of course, been extensively used in the literature on returns to education. In the standard Mincerian equation, for instance, instrumenting education with family background has been often used to correct for the endogeneity of education. But if family background is an independent determinant of earnings in its own right, and indeed one is interested in separating out the impacts of the instrument and of the variable it is instrumenting for, as is presently the case, then some other instrument is required. Ability tests taken while attending school have also been used. But ability may also belong in the earnings equation and, besides, that information is seldom available in developing countries. Other popular instruments include distance to school or, equivalently, the density of the school network in the respondent's vicinity. This would have been obtainable for Brazil, but it is likely to be correlated with other omitted earnings determinants, such as city size. Also, school distance might not have been an appropriate instrument for other variables in E, such as migration or labor market status.[21]

In the absence of an adequate set of instrumental variables Z, the only solution is to explore the likely sign and magnitude of the potential bias in the estimation of α and β due to the correlation between u and v, and then to decide, on that basis, what is the most reasonable range of estimates. This is the approach we adopt in this paper, relying on the following parametric bounds analysis.

Let X = (C, E) = (R, GR, MPE, DPE, FO; S, S^2, M, L) and $\gamma = (\alpha, \beta)'$, so that we can rewrite (3) or (8) as:

$$Lnw_i = X_i\gamma + u_i \tag{3'}$$

[21] Early contributions to the literature on estimating returns to schooling include Bowles (1972) and Griliches and Mason (1972). For an excellent survey of more recent empirical estimates, using instruments based on institutional features of the supply side of the education system, see Card (2001).

In (3'), as in (3), u_i is not necessarily orthogonal to all explanatory variables in *X*. Assume without loss of generality that all the variables have zero mean and define the following covariance matrices:

$$\Sigma = \begin{bmatrix} X'X & X'u \\ u'X & u'u \end{bmatrix} \quad \text{and} \quad S = X'X$$

The bias of OLS estimates of equation (3') is given by B in (13):

$$E(\hat{\gamma}) = \gamma + B \quad \text{with} \quad B = S^{-1}X'u = S^{-1}(\rho_{Xu} \otimes \sigma_X)\sigma_u \tag{13}$$

where ρ_{Xu} stands for the correlation coefficients between the components of *X* and the residual term, *u*, and σ_V is the standard error of variable *V*. Evaluating the bias vector *B* thus requires knowing σ_u and ρ_{Xu} . An unbiased estimator of σ_u is readily obtained for any set of correlation coefficients ρ_{Xu} . Indeed, it can be shown that $\sigma_u^2 = \hat{\sigma}_u^2 + B'SB$ where $\hat{\sigma}_u^2$ is the variance of the OLS residuals. Substituting the value of the bias given in (13) in that expression yields:

$$\sigma_u^2 = \hat{\sigma}_u^2 / (1 - K) \tag{14}$$

with K given by:

$$K = (\rho_{Xu} \otimes \sigma_X)' S^{-1} (\rho_{Xu} \otimes \sigma_X). \tag{15}$$

For any set of correlation coefficients ρ_{Xu} , equations (13)-(15) would permit computing the bias vector B and thus obtaining unbiased estimates of the variance of the error term and of the coefficients of the model, γ. Since these correlation coefficients are not known, we explore the set of possible values for the bias by randomly generating a large number of correlation coefficients, and checking for consistency with a set of conditions which must hold for them to be valid. We first randomly draw 1000 values for ρ_{Su} , ρ_{S2u} , ρ_{Mu} and ρ_{Lu} , each from a uniform distribution defined on (-1, 1). Since the correlation coefficient vector ρ_{Xu} must satisfy the condition that the covariance matrix Σ be positive semi-definite, all drawings such that this condition is not satisfied are deleted. Four additional assumptions are imposed on the signs of coefficient estimates $\hat{\gamma}$, as in modern bounds analysis (see, for instance, Manski and Pepper, 2000): the coefficients on the race dummy for Afro-Brazilians, on the regional dummy for the Northeast and on the labor market status dummy for informal employees were constrained to be non-positive; the coefficients on own schooling and on the labor market status dummy for employers were constrained to be non-negative. Each of these assumptions is backed up by an extensive body of earnings regressions for Brazil; see various chapters in Henriques (2000). Correlation coefficients leading to coefficient estimates that violate those restrictions are also discarded.[22]

The surviving values of the randomly generated correlation coefficients are substituted into (15) and (13), to generate a distribution of values for each element of the bias vector, B. We then use the corresponding distribution of estimates of coefficients $\hat{\gamma} = (\hat{\alpha}, \hat{\beta})$ to generate a series of counterfactual earnings distributions $\{\tilde{w}\}$, through equation (6), and $\{w^p\}$, through equation (7).[23] Inequality indices – Theil indices and Gini coefficients - are then computed over each of those distributions. Finally, in Section 7, we report the decompositions described in (5) and (7) for the mean, the highest, and the lowest values in the set of these inequality measures, for each cohort/gender combination.

6. Estimation Results

Earnings equations (8) were estimated separately for men and women, and by cohort, using OLS for men and a two-stage selection bias correction procedure for women. Table 2a reports the equation for males; Table 2b reports the second stage equation for women, and Table 2c reports

[22] Bounds analysis of this kind is not used as commonly to address identification issues as one might expect, because it usually leads to rather large bounds around coefficient estimates, considerably reducing their usefulness for hypothesis testing. For the purposes of our decomposition, however, the bounds that matter are those around the inequality measure defined over a distribution of counterfactual earnings predicted using the estimated coefficients (rather than around the coefficients themselves). For this application, the bounds are acceptably narrow, and this kind of analysis is probably preferable to the use of instruments of questionable validity.

[23] The second line of equation (6) implies that circumstance variables are equalized in (9), (10) and (11), to predict counterfactual schooling levels, migration decisions and labor status categories. This is straight forward for the linear schooling equation (9). For simulating with the probit model (10) and the multinomial logit (11), one does not observe actual residuals, so that pseudo-residuals must be simulated in a way which is consistent with the observed choices in the actual distribution. An easy way to do this is to draw the residuals randomly in the law of extreme values, discarding drawings which are inconsistent with observed choices. The operation is repeated until a single set of values is drawn such that the migration and occupational choices are in accordance with the ones actually observed. See Bourguignon et al. (2004) for a detailed explanation of the methodology.

the first-stage equation for women. For each entry in Tables 2a and 2b, we first report our estimate of the unbiased coefficient, obtained by correcting the OLS estimate with our mean estimate for its bias, as described in Section 5. Below this number, we report the unadjusted OLS coefficient and its standard error, for comparison. Apart from the oldest female age-group, biases turn out to have been quite small, so that the general discussion below largely refers to these coefficients interchangeably. In the equation for females, family composition variables, the number of children and household log income per capita – excluding own earnings – were used as instrumental variables in the first stage. A similar Heckman correction procedure was initially applied to men as well, but available instruments proved unsatisfactory. Unlike in the standard Mincer specification, age or imputed experience do not appear among the regressors because we treat cohorts as age-homogeneous by definition.

Circumstance variables have the expected effect on earnings. The coefficients of racial dummy variables are negative and significant for both blacks and 'pardos'.[24] They are generally positive, but not always significant for people with an Asian origin. The racial gap against blacks and 'pardos' is much less pronounced for women and in fact it is generally insignificant. Regional differences are important, too. With the South-East as a reference, being born in the North-East has a strong and significantly negative effect for both men and women. The effect of the other regions is generally also negative, but seldom significant.[25] The estimated effect of mean parental education on individual earnings is always positive and highly significant. It tends to be somewhat higher for women than for men, and in both cases tends to decline from the older to the younger cohorts. It is also sizable, since it amounts to a 3 to 6 per cent increase in earnings, for each additional year of schooling of the parents.

The difference between the education of the mother and the father is meant to detect a possible asymmetry in the role of the two parents. But no such asymmetry seems to be systematically present. The estimated effect of the father's occupation on earnings, which we do not report on the tables due to space constraints, is generally positive when compared to the reference category

[24] Race is self-reported in the PNAD: the respondent, rather than the interviewer, chooses his or her race. 'Pardo' is meant to refer to people of mixed-race, generally involving some Afro-Brazilian component.
[25] Except for the South region, for men.

of rural workers, but seldom significant. Significance tends to increase for the younger cohorts, but this seems to be driven by sample size differences, rather than changes in the coefficients.[26]

Turning now to the vector of "effort" variables, own education has the usual positive and significant effect on earnings for men. This effect is decreasing as one considers younger male cohorts. This is consistent with the negative coefficient generally found for the squared imputed experience term –i.e. age minus number of years of schooling minus first schooling age - in the standard Mincerian specification. This implies that returns to schooling increase with age, which is exactly what is found here.[27] The overall effect of education for women is insignificant at very low levels of schooling, but picks up at higher level of education. This is captured by the coefficient on the quadratic term, which is positive and significant for both genders, indicating that the marginal returns to education in Brazil increase with the number of years of schooling.

The magnitude of the estimates for the returns to schooling in these equations is somewhat lower than previous estimates for Brazil. For instance, Ferreira and Paes de Barros (1999) found that the marginal returns to a year of schooling lay in the range of 12 to 15 percent, for both men and women in 1999. In Table 2, marginal returns at 5 years of schooling range from 7 to 11 percent for men and from 11 to 13 percent at 10 years of schooling. This difference is most likely due to the inclusion of the family background variables, notably mean parental education. This resonates with earlier findings by Lam and Schoeni (1993), who also found that returns on own schooling decline when parental schooling is introduced in the regression. Lam and Schoeni (1993) interpret this finding as indicating a correlation between parental schooling and unobserved characteristics of workers, which are also correlated with their own schooling. We interpret the coefficient on parental schooling as capturing the effects of family background on current earnings which do not go through an individual's own education, migration decisions, or labor market status. They may include intergenerationally correlated ability, as proposed by Behrman and Rosenzweig (2002), which we also treat as a circumstance that can contribute to

[26] Interestingly, we observe that all father's occupational categories are highly significant determinants of individuals' years of schooling in our effort equations described further on.
[27] The conventional Mincerian specification is such that: $Lnw = a.S + b.Exp – c.Exp^2$ where $Exp = Age – S – 6$. Expanding the Exp term leads to: $Lnw = (a-b-12c).S + 2cAge.S - c.S^2$ + terms in Age or Age squared. If this equation is estimated within groups with constant age, one should indeed observe that the coefficient of S is higher in older cohorts. Note that the present specification also includes an independent S^2 term.

unequal opportunities. It may include the effect of family wealth on the quality of the school attended by the child, an effect which will be present both on the school quantity equation and on the earnings equation, controlling for number of years. It may capture the effect of the parents' social network in finding their child a high-paying job, and so on. A moment's reflection will suggest that our interpretation is not inconsistent with Lam and Schoeni's, and is also adequate to the decomposition we propose.

Migration has a positive effect on earnings for both men and women, although it is more often significant for men. This sign could be consistent with a human capital interpretation of migration. The coefficient is rather large, amounting to a 10-20 per cent increase in earnings for males. While migration decisions taken as an adult are consistent with the "effort" classification, the definition of the variable in the data does not allow us to separate adult migrants from those who migrated as children, reflecting a parental decision. In the latter case, this variable might be better classified as a circumstance variable. This ambiguity in the (circumstance versus "effort") status of the migration variable – which is of course also applicable to years to schooling, since early enrollment decisions are probably taken by one's parents – is part of the reason why we must interpret the unequal opportunities component of our decomposition as a lower bound on the contribution of inequality of opportunities to earnings inequality in urban Brazil.

The labor market status variable confirms one's expectations: employers earn significantly more than formal employees, given these controls, and everyone else earns less (with the exception of self-employed women, for whom there is no significant difference with respect to formal employees).

"Effort" equations: schooling, migration and position in the labor market

We now briefly report on estimates for equations (9)-(11), and consider the impact of circumstances on our "effort" variables, namely own schooling; migration status and position in the labor market. The results for the linear schooling regressions are shown in Tables 3a and 3b separately for men and women. They call for several remarks.

The first set of remarks has to do with intergenerational educational mobility as measured, inversely, by the coefficient of parental education, α_P. This is an education analogue to the intergenerational elasticity of earnings discussed in Section 2. It would be natural to expect this educational persistence coefficient to lie in the (0, 1) interval, with zero suggesting complete independence between the schooling levels across generations, and one indicating that parental education fully determines schooling in the next generation, up to an average increase given by the constant term in the regression, and subject to a random term. Table 3 reveals rather high, but declining coefficients: educational persistence falls from 0.74 for the oldest generation to 0.5 for the youngest among men, and from 0.82 to 0.48 among women. Rising average levels of schooling across generations are picked up by the large constant terms. The fact that these estimates also rise across cohorts indicate an acceleration in schooling differentials across generations, for these age groups.

Father's occupation, race and region of birth are also significant determinants of education, even after controlling for parental schooling, and one another. Unlike for parental education, coefficients for these other circumstance variables do not display a clear declining trend across cohorts. Afro-Brazilians seem to have roughly the same absolute disadvantage in years of schooling in the cohort born in the 1960s - 1 to 1.5 years of schooling - as they have in the cohorts born in the 1930s or 1940s. Likewise, the disadvantage of being born in the North-East for men displays no clear trend. In effect, only the coefficient reflecting the disadvantage of being born from parents with a low level of schooling seems to have been clearly and steadily falling across cohorts. In other words, the conditional distribution of educational opportunities seems to have remained approximately constant over time, except with respect to family background.[28]

An interesting feature of intergenerational educational mobility is that intra-household decision mechanisms seem to matter more for the education of daughters than for sons. The transmission of education from parents to girls is higher, the greater the mother's schooling relative to the

[28] There are some interesting changes in the coefficients for specific father's occupational dummies for women: the advantage of women whose fathers were employers or domestic servants, with respect to those whose fathers were rural workers has increased substantially. Conversely, those women whose fathers were "liberal professionals" (white-collar workers), have seen their educational advantage over the daughters of rural workers decline across cohorts.

father's. This effect is significant and persistent across cohorts. For boys, however, although the mother's education still seems to weigh more than the father's, the difference is only significant at the 5% level for a single cohort. In both cases, it is difficult to find a trend in the evolution across cohorts.

This discussion of inequality of educational opportunities has been based on measuring education in terms of the number of years of schooling. One might well prefer a more sophisticated proxy for 'human capital' - such as the cost of education, including forgone earnings, or possibly the earnings that a given schooling level actually commands - rather than unadjusted years of schooling. One problem with simple years of schooling is that the quality of the education received is not captured. It cannot be ruled out that taking into account the quality of education might modify the preceding conclusion of an increasing educational mobility in Brazil. Unfortunately, we know of no data source in Brazil which would allow us to combine information on plausible measures of educational quality with information on the schooling of parents, for individuals currently active in the labor market.[29]

We now turn to the effect of circumstance variables on the probability of migration, as estimated by the probit model in (10). Marginal effects calculated from the probits are shown for men and for women in Tables 4a and 4b. As is well-known, and as the constant terms confirm, migration was much less common for the later cohorts than it had been for those born in the late 30s and 40s. Those born to fathers working in rural areas were likelier to migrate, and all the more so if they were born in the Northeast. For younger cohorts, migrants from the South and Center West also became more common. Afro-Brazilians and those with more educated parents are somewhat less likely to migrate, but neither race nor parental education were often significant.

Tables 5a and 5b present the estimated marginal effects derived from the coefficients from the labor market status multinomial logit model in (11). The results suggest that both men and women with more educated parents are likelier to be employers or formal sector employees, and

[29] Behrman and Birdsall (1983) attempted to control for quality of schooling in earnings functions in Brazil, by linking data on individuals aged 15-35 years old with observations on the average schooling of teachers in the area where they had obtained their schooling. In our case, a similar analysis would not be plausible, since we cover a much larger time interval (seven cohorts with their corresponding parents), and do not know where most of the sampled individuals actually went to school.

less likely to be informal sector employees ('sem carteira'). Parental education was not a significant influence on the probability of being self-employed, vis-à-vis being a formal sector employee, which may be a reflection of the greater heterogeneity of the self-employed sector. Having parents who worked as domestic servants seems to increase the probability that one is self-employed (particularly for men). Afro-Brazilians seem to be generally more likely to work in the informal sector, and less likely to employ others. Self-employment is relatively more common in the North-East, relative to formal employment. All in all, it seems that this basic set of circumstances can indeed have a significant impact on the nature of one's insertion in the labor market in urban Brazil.

7. Decomposition Results.

Using the coefficient estimates for equations (8)-(11), we simulate the counterfactual distributions $\{\tilde{w}\}$, from equation (6), and $\{w^p\}$ from equation (7). This allows us to decompose earnings inequality for each gender/cohort combination in our sample, into a component due to unequal opportunities, and a residual component due to unobserved circumstances, "efforts", and random elements such as transitory earnings or measurement error. The bounds analysis described in Section 5 was used to estimate a distribution of coefficient estimates that would give us a sense of the upper and lower limits for the biases. Using this distribution of coefficient estimates resulting from the bounds analysis, we computed counterfactual distributions and decompositions, over which inequality indices were calculated. The mean value, as well as the upper and lower bounds of inequality indices are presented in Table 6, for men, and in Table 7, for women.

Each of these tables presents Theil coefficients for factual and counterfactual earnings distributions for our seven cohorts in 1996. The first row contains the observed inequality. The next panel, with four rows, presents inequality in the counterfactual distribution $\{\tilde{w}\}$, where the inequality of opportunities due to observed circumstances has been eliminated: $I(\tilde{w}) = I\left(w \middle| C_i = \overline{C}, \forall i\right)$. The mean value of all Theil indices as well as those corresponding to the extreme bounds are presented in the first three rows. The fourth row presents the share in earnings inequality accounted for by the difference between observed inequality and the mean estimate of residual inequality, i.e.: $\Theta_I := \frac{I(w) - I(\tilde{w})}{I(w)}$. This is our measure of inequality of opportunities in this decomposition. Because the coefficients should be unbiased, but we can not be sure that there are no unobserved circumstance variables in the residual term u_i, we interpret this share as a lower bound.

For men born between 1951 and 1955, for instance, elimination of inequality due to observed circumstances reduces the Theil index from 0.68 to some value between 0.527 and 0.554, with a mean estimate of 0.536. This mean estimate indicates that at least 21% of earnings inequality in this cohort is accounted for by unequal opportunities. Similarly, for women born between 1956 and 1960, replacing individual values of circumstance variables with their means reduces the Theil from 0.71 to somewhere between 0.490 and 0.532, with a mean estimate of 0.503. This value yields an opportunity share of inequality of around 0.29.

The share of inequality due to unequal opportunities varies across cohorts, between 16% and 30% for men, and 14% and 32% for women. The simple average across cohorts is 22% for men, and 23% for women. As indicated by the subscript I in Θ_I, these shares depend on the specific inequality measure used. A perfectly analogous decomposition was conducted using the Gini coefficient, instead of the Theil index, and cohort averages for the Gini were 13% for men and 12% for women. Detailed results for the Gini decomposition are available from the authors on request.

The next panels (with four rows each) in Tables 6 and 7 present the sub-decomposition of inequality of opportunity into its direct and indirect components. The central and bounds estimates for counterfactual Theil coefficients in this panel correspond to the simulated distribution $\{w^p\}$, obtained from $Ln(w_i^p) = \overline{C}_i \hat{\alpha} + E_i \hat{\beta} + \hat{u}_i$. The difference between observed inequality and this inequality level can be accounted for by holding observed circumstances constant only in the earnings regressions, whilst not taking account of the impact of unequal circumstances on the levels of “efforts”, such as own schooling levels, a decision to migrate to an area with greater income earning opportunities, or efforts to find a job in a better sector of the labor market.

The fourth row of this panel presents our summary measure of this direct effect of unequal opportunities: $\Theta_I^d := \left[I(w) - I\left(w^p\right)\right]\left(I(w)\right)^{-1}$. This sub-component of inequality of opportunities averaged 15% for men across all seven cohorts, and 16% for women. This suggests that, on average, approximately 70% of the impact of opportunities on earnings takes the form of a direct effect, while the remaining 30% correspond to the indirect effect of circumstance on earnings through their impact on "efforts" – such as the effect of parental education on earnings through own schooling. A similar breakdown is found for the decomposition using the Gini coefficient.

It is possible, of course, that the effect of circumstances on our "effort" variables (schooling, migration and labor market status) is greater than the effects captured through equations (9) – (11). If unobserved circumstances accounted for a large share of the variance in the random terms *v*, ξ, and ε, we could be underestimating the share of opportunities in earnings inequality. To shed some light on the magnitudes associated with this possibility, the bottom panels in both Table 6 and Table 7 report results from a thought experiment. Suppose schooling and migration decisions were taken entirely by a person's parents, with no room for individual decision-making, and that labor market status were also somehow exogenous. Those (rather extreme) assumptions would correspond to treating all of our observed "effort" variables as circumstances. Equalizing all observed variables in (8) – and treating all unobserved variance in *u* as the only true source of "effort" – yields the decomposition in this panel, derived from a simulation of $Ln\left(\tilde{\tilde{w}}_i\right) = \overline{C}_i\hat{\alpha} + \overline{E}_i\hat{\beta} + \hat{u}_i$ Unsurprisingly, such a view of the world leads to much higher opportunity shares, as indicated in the bottom rows of Tables 6 and 7. These opportunity shares average of 40% for men and 45% for women, across cohorts.[30]

Figures 1 and 2 depict the decomposition graphically, for men and women. In both figures, the top line represents the inequality actually observed for the various cohorts. The line below it

[30] It is important to note that the opportunity shares in this thought experiment correspond neither to a lower bound on inequality of opportunity (because it seems unreasonable to interpret schooling, migration and labor market status as fully exogenous), nor to an upper bound (because there are still likely to be unobserved circumstance variables in the residual term, *u*, of equation 8).

(with squares) shows the "partial" or direct effect of equalizing circumstances, whereas the bottom line (with circles) shows the "complete" effect. The dotted lines around the two bottom lines show upper and lower bounds for the preceding estimates. Inequality of opportunity, by each measure, corresponds to the difference between observed inequality (at the top) and the simulated inequality indices. The height of the counterfactual inequality measures corresponds to the residual component.

Visual inspection of these figures reveals that male earnings inequality falls considerably from the older to the younger cohorts, while for women, inequality is lower for both the oldest and the youngest cohorts, and higher for those in between. It is important to recognize that temporal patterns can not be identified from these cohort trends. In particular, as already alluded to, returns to education in Brazil increase with age, leading to greater dispersion in earnings within older cohorts, at every time period.[31] It is also interesting that the share (as well as the absolute contribution) of inequality of opportunities in earnings inequality also seems to decline, albeit moderately, from the oldest to the youngest cohort. That share averages 25% (31%) for men (women) born between 1936 and 1945, as opposed to 21% (18%) for those born between 1961 and 1970. This decline is in all likelihood driven by the falling coefficients on parental education in both the earnings and own schooling equations for both genders, as shown in Tables 2 and 3. To the extent that these *shares* are unrelated to age-group-specific *levels* of earnings inequality, the reductions in the shares of both parental education and father's occupation (particularly for men) may signify a lower degree of inequality of opportunity for the younger cohorts in Brazil.

Figures 3 and 4 evaluate the role of individual circumstance variables in the preceding results. The complete effect of equalizing each individual circumstance variable, while controlling for all others, but allowing them to take their observed values, is shown for the Theil coefficient separately for men and women. It can be seen that, of all circumstance variables, parental education plays the largest role in determining inequality, for both genders and across all cohorts. Interestingly, the contribution of parental education to reducing earnings inequality is not much smaller when parental schooling is not equalized across the board, but instead a lower

[31] This is particularly true for men. Evidence of this age-dependence of earnings inequality in other countries is analyzed in Deaton and Paxson (1994).

bound (of six school years) is imposed, as if schooling were compulsory until a certain age. This suggests that it is the inequality of education at the bottom of the distribution that matters most to explaining the contribution of opportunities to earnings inequality.

Reinforcing the importance of family background as the key set of circumstances which shape opportunity sets for the young, father's occupation is the second most important circumstance, although this is more pronounced for men than for women. Even for men, there is a clear downward trend in the contribution of this variable, as one moves from the older to the younger cohorts. When parental occupation and education are already controlled for, race alone seems to account for very little inequality in opportunities. These results suggest that the most effective policies for reducing inequality of opportunities in Brazil would be those that reduced the effect of parental education on their child's schooling and earnings. This conclusion is even stronger for women than for men.

8. Conclusions.

This paper sought to quantify the role of inequality of opportunity – associated with family background, race, and region of origin – in generating inequality in current earnings in Brazil. We estimated the impact of opportunities (or circumstances) both directly on earnings, and indirectly on a number of individual "effort" variables – such as schooling, the decision to migrate, or one's broad status in the labor market. We took account of the biases arising from the endogeneity of these "efforts", by calculating upper and lower bounds for the unbiased estimates, drawn from Monte-Carlo simulations based on the universe of admissible values for the bias.

Altogether, we find that a group of four observed circumstance variables (namely parental schooling, father's occupation, race and region of birth) account for more than a fifth of the total earnings inequality within gender/cohort groups in Brazil in 1996, as measured by the Theil index. For some cohorts, particularly the older ones, this share is nearer 30%. The impact of opportunities, defined in this way, was further decomposed into a direct effect on earnings (which accounts for some 70% of the total), and an indirect effect through the effort decisions individuals make. Regardless of the channel, our analysis suggests that family background is the most important set of circumstances determining a person’s opportunities. Fifty-five to 75 percent of the total effect of circumstances can be attributed to parental schooling alone, and this figure rises to 70 to 80 percent when the father's occupation is added. There is also some evidence that inequality of opportunity may account for a lower share of earnings inequality in the younger cohorts, which may be consistent with an actual decline in that component of inequality over time.

Considering the fact that our analysis was based on only four observed circumstance variables, and that components of unobserved circumstances which are not correlated with them[32] may be present in the residual term, we interpret these shares as lower-bound estimates of the inequality of opportunity in Brazil. Furthermore, the variance of the residual term u_i, which accounts for a large share of the residual component of the decomposition, includes not only genuine rewards to differences in unobserved effort, but also measurement error and transitory income components.[33] In this light, a lower bound share of inequality of opportunity of over 20% is not insubstantial. While the remaining residual inequality is still very large – and in fact larger than total inequality in most OECD countries – we can certainly conclude that unequal opportunities account for a substantial share of observed inequality in Brazil. It would be interesting for future research to explore this question further, with better measures of life-time earnings or permanent consumption, for which transitory components might be lower. As the literature on intergenerational mobility in the United States has shown, the extent to which one can average out earnings volatility and measure “permanent” economic status has a considerable impact on estimates of intergenerational persistence and, hence, on the share of inequality accounted for by opportunities.

To the extent that Brazilian policy-makers are averse to unequal opportunities, our analysis may have some policy implications. Figures 1 and 2 showed that both the direct and the indirect effects of circumstances on earnings are important, each in its own right. This suggests that

[32] In the absence of wealth data, it is impossible to determine how well family wealth correlates with parental education and occupation. As a rough gauge, we ran a regression of current income on current occupation, education and region of birth for the PNAD respondents with children. R-squares are in the range 0.31 - 0.42, depending on the cohort.

[33] Atkinson et. al (1992) report that the share of transitory components in the variance of the logarithm of current earnings is around 30% in a number of developed countries. See also Lillard and Willis (1978) for the original US study.

policies aimed at equalizing opportunity may be warranted *both* inside and beyond the classroom. Parental education – and, to a lesser extent, occupation – do affect the length of children's school careers. Efforts to reduce this dependence, through conditional cash-based assistance to poor students and their families, or through measures to enhance the quality of schooling supplied across the country's schools – such as after-school programs for students who may be falling behind, or through curriculum changes in teacher-training programs which aim to strengthen teacher capacity to support the weakest students - might well deserve consideration. But family background clearly also impacts on earnings directly, even after conditioning on own schooling. In fact, this direct effect of family background seems to be quantitatively more important than the indirect effects. This is consistent with hypotheses that both employment and career advancement opportunities may be allocated in part through socially-based networks.[34] In this paper, we have not presented any evidence on the existence of such networks, or on their welfare properties, but further investigation of their operation in the Brazilian case is warranted, and may be relevant for the ongoing debate on affirmative action in Brazil, and in other countries.

[34] Evidence from elsewhere suggests that socially-based networks can be effective in matching workers from certain families to coveted jobs. One example is the traditional segmentation of blue-collar occupations in Bombay by *jati* (or sub-caste) groups. While such labor-market effects are likely to impact on educational decisions (as explored by Munshi and Rosenzweig, 2003), they also constitute a direct impact of family background on earning opportunities, conditional on the child's schooling attainment.

References :

Angrist, Josh and Alan Krueger (1992): "The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples", *Journal of the American Statistical Association*, **87**, pp.328-336.

Arneson, R. (1989): "Equality of Opportunity for Welfare", *Philosophical Studies*, **56**, pp.77-93.

Atkinson, A. B., F. Bourguignon and C. Morrisson (1992): *Empirical Studies of Earnings Mobility*, Philadelphia: Harwood Academic Publishers.

Barros, R. P. and D. Lam (1993): "Desigualdade de renda, desigualdade em educação e escolaridade das crianças no Brasil", *Pesquisa e Planejamento Econômico*, **23** (2), pp.191-218.

Behrman, J. and N. Birdsall (1983), "The Quality of Schooling: Quantity alone is misleading", *American Economic Review*, **73** (5), pp. 928-946.

Behrman, J., N. Birdsall and M. Szekely (2000), "Intergenerational mobility in Latin America: deeper markets and better schools make a difference", in N. Birdsall and C. Graham (eds), *New markets, new opportunities? Economic and social mobility in a changing world*, Washington, D. C. : Brookings Institution, p. 135-67

Behrman, J. and M.R. Rosenzweig (2002), "Does Increasing Women's Schooling Raise the Schooling of the Next Generation?", *American Economic Review,* **92** (1), pp.323-334

Behrman, J. and P. Taubman (1976): "Intergenerational Transmission of Income and Wealth", *American Economic Review*, **66** (2), pp.436-440.

Betts, Julian and John Roemer (1999): "Equalizing Opportunity through Educational Finance Reform", mimeo, Public Policy Institute of California, San Francisco, CA.

Bourguignon, F., F.H.G. Ferreira and N. Lustig (2004): *The Microeconomics of Income Distribution Dynamics in East Asia and Latin America*, Washington, DC: The World Bank and Oxford University Press.

Bowles, S. (1972), "Schooling and inequality from generation to generation", *Journal of Political Economy*, **80** (3), pp.S219-S51

Bowles S. and Gintis , H. (2002), "The inheritance of inequality", *Journal of Economic Perspectives*, **16** (3), pp. 3-30

Card, D. (2001), "Estimating the returns to schooling: progress on some persistent econometric problems", *Econometrica*, **69** (5), pp. 1127-60

Cohen, G.A. (1989), "On the Currency of Egalitarian Justice", *Ethics*, **99**, pp.906-944.

Deaton, A. and C. Paxson (1994), "Intertemporal choice and inequality", *Journal of Political Economy*, **102** (3), pp. 437-67

Dunn, Christopher (2003): "Intergenerational Earnings Mobility in Brazil and Its Determinants", University of Michigan, unpublished.

Dworkin, R. (1981): "What is Equality? Part 1: Equality of Welfare; Part 2: Equality of Resources", *Philos. Public Affairs*, **10**, pp.185-246; 283-345.

Ferreira, F.H.G.; P. Lanjouw and M. Neri (2003): "A Robust Poverty Profile for Brazil Using Multiple Data Sources", *Revista Brasileira de Economia*, **57** (1), pp.59-92.

Ferreira, F.H.G. and R. Paes de Barros (1999): "The Slippery Slope: Explaining the Increase in Extreme Poverty in Urban Brazil, 1976-1996", *Brazilian Review of Econometrics*, **19** (2), pp.211-296.

Ferreira, Sérgio G. and Fernando A. Veloso (2004): "Intergenerational Mobility of Wages in Brazil", Banco Nacional de Desenvolvimento Econômico e Social (BNDES), Rio de Janeiro, unpublished.

Griliches, Z. and W. Mason (1972), "Education, Income and Ability", *Journal of Political Economy*, **80** (3), pp.S74-S103

Hanushek, E. (1986): "The Economics of Schooling: Production and Efficiency in Public Schools", *Journal of Economic Literature*, **24**, pp.1141-1177.

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

Lam, D. and R.F. Schoeni (1993), "Effects of Family Background on Earnings and Returns to Schooling: Evidence from Brazil", *Journal of Political Economy*, **101** (4), pp. 710-740.

Lam, D. (1999), "Generating extreme inequality: schooling, earnings, and intergenerational transmission of human capital in South Africa and Brazil", Population Studies Center, University of Michigan, Report 99-439

Lillard, L.A. and R. J. Willis (1978): "Dynamic Aspects of Earning Mobility", *Econometrica*, **46** (5), pp.985-1012.

Manski, C. and J. Pepper (2000), "Monotone instrumental variables: with an application to the returns to schooling", *Econometrica*, **68** (4), pp. 997-1010

Mazumder, Bhashkar (2005): "The Apple Falls Even Closer to the Tree Than We Thought", Chapter 2 in S. Bowles, H. Gintis and M. Groves, eds., *Unequal Chances: Family Background and Economic Success*, Princeton: Princeton University Press.

McFadden, Daniel (1974): “Conditional Logit Analysis of Qualitative Choice Behavior”, in Paul Zarembka, ed., *Frontiers in Econometrics*, New York: Academic Press.

Mulligan, C. (1999), “Galton versus the human capital approach to inheritance”, *The Journal of Political Economy*, **107** (6), pp. 184-224.

Munshi, K. and M. Rosenzweig (2003), “Traditional Institutions Meet the Modern World: Caste, Gender and Schooling Choice in a Globalizing Economy”, Harvard University, unpublished.

Page, Marianne and John Roemer (2001): “The US Fiscal System as an Equal Opportunity Device”, in Kevin Hasset and R. Glenn Hubbard, eds. *The Role of Inequality in Tax Policy*, Washington DC: The American Enterprise Institute Press.

Peragine, Vito (2004): “Ranking Income Distributions According to Equality of Opportunity”, *Journal of Economic Inequality*, **2**, pp.11-30.

Pero, V. (2001) "Et, à Rio, plus ça reste le même... Tendências da mobilidade social intergeracional no Rio de Janeiro", ANPEC, Salvador.

Piketty, T. (1995): “Social Mobility and Redistributive Politics”, *Quarterly Journal of Economics*, **CX** (3), pp.551-584.

Roemer, J. E. (1998): *Equality of Opportunity*, (Cambridge, MA: Harvard University Press)

Roemer, J.E., R. Aaberge, U. Colombino, J. Fritzell, S. P. Jenkins, I. Marx, M. Page, E. Pommer J. Ruiz-Castillo, M.J.S. Segundo, T. Traanes, G. Wagner and I Zubiri (forthcoming): “To What Extent do Fiscal Regimes Equalize Opportunities for Income Acquisition Among Citizens?”, *Journal of Public Economics*.

Solon, Gary (1992): “Intergenerational Income Mobility in the United States”, *American Economic Review*, **82** (3), pp. 393-408.

Solon, Gary (1999), “Intergenerational mobility in the labor market”, in O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*, Amsterdam, North-Holland, Vol. 3A, pp. 1761-1800

Valle Silva, N. (1978), *Posição social das ocupações*. Rio de Janeiro: IBGE.

Table 1. Descriptive statistics, by 5-year birth cohorts.

	b1936_40	b1941_45	b1946_50	b1951_55	b1956_60	b1961_65	b1966_70
Mean monthly earnings (Reais, all jobs)	561.3	753.6	812.2	808.0	736.8	658.5	543.7
Mean number of years of schooling	4.5	5.6	6.4	7.1	7.5	7.8	7.7
Mean father's number of years of schooling	2.2	2.4	2.7	2.9	3.2	3.5	3.5
Mean mother's number of years of schooling	1.7	2.0	2.3	2.5	2.8	3.2	3.2
Race (Percents)							
Branca (whites)	60.0%	59.2%	59.4%	59.9%	58.7%	59.6%	57.9%
Preta (blacks)	6.5%	6.7%	6.3%	6.2%	5.7%	5.2%	5.7%
Amarela (Asians)	0.4%	0.8%	0.5%	0.5%	0.4%	0.3%	0.3%
Parda (MR)	33.0%	33.3%	33.8%	33.4%	35.2%	34.8%	36.1%
Regions (Percents)							
North	6.2%	6.8%	6.9%	7.5%	7.7%	7.5%	7.8%
North East	23.8%	25.6%	24.2%	22.2%	23.1%	23.3%	24.3%
South East	39.5%	37.8%	38.6%	38.7%	37.1%	36.5%	34.0%
South	21.3%	20.4%	20.4%	21.7%	21.2%	21.3%	20.9%
Center-West	9.1%	9.4%	9.9%	10.0%	10.9%	11.4%	13.0%
Migrants (Percents)	70.2%	69.4%	68.5%	66.3%	63.0%	59.4%	57.5%
Father's occupational status							
Rural workers	28.65%	24.47%	22.89%	21.02%	18.76%	16.28%	14.97%
Domestic servants	35.06%	36.05%	32.30%	30.32%	27.88%	25.86%	25.07%
Traditional sector workers	6.72%	7.37%	8.22%	8.84%	10.08%	11.09%	12.84%
Service sector workers	5.95%	6.01%	6.27%	7.00%	7.85%	8.61%	9.42%
Modern industry workers	5.95%	6.56%	8.00%	9.59%	11.68%	12.94%	14.43%
Self-employed shopkeepers	6.18%	6.92%	7.67%	7.55%	7.77%	8.15%	7.30%
Technitians, artists and desk workers	4.75%	4.63%	6.12%	7.18%	7.68%	9.02%	8.55%
Employers	5.25%	5.02%	4.96%	4.85%	4.60%	3.73%	3.89%
Liberal professionals	1.51%	2.97%	3.57%	3.65%	3.71%	4.33%	3.54%
Labor market status							
Formal employees	32.6%	37.6%	45.9%	50.3%	51.5%	51.3%	53.3%
Informal employees	16.3%	15.2%	15.0%	14.5%	16.1%	17.9%	20.0%
Self-employed	44.3%	39.3%	32.8%	28.8%	25.9%	25.0%	22.3%
Employers	6.8%	7.8%	6.3%	6.5%	6.5%	5.8%	4.3%
Number of individuals	2590	3842	6025	8167	9327	9498	7463

Table 2.a: Earnings equations by cohort, men. a), b).

	b1936_40	b1941_45	b1946_50	b1951_55	b1956_60	b1961_65	b1966_70
Race							
Branca (whites, omitted)							
Preta (Black)	*-0.380*	*-0.260*	*-0.318*	*-0.271*	*-0.247*	*-0.165*	*-0.194*
	-0.385***	-0.248***	-0.321***	-0.270***	-0.249***	-0.166***	-0.195***
	[0.086]	[0.070]	[0.057]	[0.050]	[0.049]	[0.044]	[0.047]
Amarela (Asians)	*-0.097*	*0.300*	*0.510*	*0.037*	*0.322*	*0.198*	*-0.363*
	-0.115	0.318**	0.506***	0.035	0.315**	0.194	-0.363*
	[0.316]	[0.158]	[0.157]	[0.157]	[0.150]	[0.152]	[0.215]
Parda (MR)	*-0.192*	*-0.323*	*-0.252*	*-0.210*	*-0.169*	*-0.156*	*-0.194*
	-0.197***	-0.317***	-0.254***	-0.209***	-0.171***	-0.155***	-0.194***
	[0.049]	[0.040]	[0.032]	[0.027]	[0.025]	[0.023]	[0.025]
Parental schooling							
Mean parental schooling (years)	*0.051*	*0.033*	*0.046*	*0.033*	*0.031*	*0.032*	*0.032*
	0.048***	0.037***	0.045***	0.033***	0.030***	0.031***	0.031***
	[0.011]	[0.009]	[0.007]	[0.005]	[0.005]	[0.004]	[0.004]
Mother/father difference (years)	*0.000*	*0.010*	*0.005*	*0.010*	*0.000*	*0.001*	*0.009*
	0.000	0.011	0.005	0.009*	0.000	0.001	0.009**
	[0.011]	[0.009]	[0.006]	[0.005]	[0.004]	[0.004]	[0.004]
Region dummies							
South East (omitted)							
North	*-0.222*	*-0.067*	*-0.144*	*-0.082*	*-0.190*	*-0.107*	*-0.146*
	-0.227*	-0.053	-0.147*	-0.082	-0.190***	-0.109**	-0.151***
	[0.123]	[0.111]	[0.078]	[0.065]	[0.057]	[0.055]	[0.058]
North East	*-0.204*	*-0.167*	*-0.119*	*-0.224*	*-0.167*	*-0.256*	*-0.208*
	-0.210***	-0.162***	-0.120***	-0.223***	-0.170***	-0.257***	-0.210***
	[0.050]	[0.041]	[0.033]	[0.029]	[0.028]	[0.025]	[0.027]
South	*-0.131*	*-0.166*	*-0.106*	*-0.070*	*-0.045*	*-0.089*	*-0.085*
	-0.138**	-0.157***	-0.110***	-0.068**	-0.048*	-0.091***	-0.087***
	[0.060]	[0.051]	[0.039]	[0.031]	[0.028]	[0.026]	[0.029]
Center-West	*-0.171*	*-0.079*	*-0.008*	*-0.219*	*-0.028*	*-0.159*	*-0.028*
	-0.175	-0.061	-0.010	-0.219***	-0.032	-0.163***	-0.032
	[0.137]	[0.097]	[0.075]	[0.062]	[0.053]	[0.045]	[0.046]
Years of schooling	*0.089*	*0.069*	*0.079*	*0.066*	*0.056*	*0.028*	*0.029*
	0.094***	0.068***	0.080***	0.068***	0.055***	0.029***	0.030***
	[0.015]	[0.013]	[0.010]	[0.009]	[0.009]	[0.008]	[0.009]
Years of schooling-squared	*0.002*	*0.003*	*0.002*	*0.003*	*0.003*	*0.005*	*0.004*
	0.002*	0.003***	0.002***	0.003***	0.003***	0.005***	0.004***
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
Migrant dummy	*0.099*	*0.152*	*0.180*	*0.110*	*0.121*	*0.131*	*0.162*
	0.101**	0.199***	0.181***	0.112***	0.120***	0.133***	0.159***
	[0.044]	[0.036]	[0.029]	[0.024]	[0.022]	[0.020]	[0.022]
Father's occupational status c)	yes	yes	yes	yes	yes	yes	yes
Labor Market status							
Formal employee (omitted)							
Informal employee	*-0.327*	*-0.318*	*-0.309*	*-0.399*	*-0.298*	*-0.245*	*-0.237*
	-0.325***	-0.299***	-0.309***	-0.395***	-0.302***	-0.245***	-0.233***
	[0.061]	[0.055]	[0.043]	[0.038]	[0.034]	[0.029]	[0.030]
Self employed	*-0.101*	*-0.101*	*-0.041*	*-0.061*	*-0.080*	*-0.047*	*0.082*
	-0.085*	-0.071*	-0.035	-0.055**	-0.081***	-0.042*	0.096***
	[0.047]	[0.038]	[0.030]	[0.026]	[0.025]	[0.023]	[0.027]
Employer	*0.590*	*0.655*	*0.377*	*0.375*	*0.408*	*0.444*	*0.490*
	0.573***	0.613***	0.376***	0.370***	0.398***	0.449***	0.482***
	[0.076]	[0.061]	[0.051]	[0.041]	[0.040]	[0.039]	[0.049]
Constant	*0.205*	*0.409*	*0.299*	*0.413*	*0.331*	*0.301*	*0.185*
	0.203***	0.329***	0.299***	0.400***	0.348***	0.300***	0.184***
	[0.067]	[0.059]	[0.049]	[0.043]	[0.042]	[0.041]	[0.045]
Sample size	1684	2397	3637	4797	5375	5544	4473
Adj R-squared	0.470	0.480	0.490	0.470	0.440	0.460	0.400

a) Dependent variable is the log of hourly wage rate. b) For each variable, we present three values: the unbiased mean coefficient estimates (in italics), OLS estimates and OLS standard errors (in brackets); * significant at 10%; ** significant at 5%; *** significant at 1% c) Coefficients for the father's occupational status dummy variables are here omitted to save space, though available from the authors on request.

Table 2.b: Earnings equations by cohort,women. a), b)

	b1936_40	b1941_45	b1946_50	b1951_55	b1956_60	b1961_65	b1966_70
Race							
Branca (whites, omitted)							
Preta (Black)	*-0.082*	*-0.025*	*-0.090*	*-0.111*	*-0.232*	*-0.126*	*-0.172*
	0.043	0.008	-0.081	-0.114	-0.206	-0.121*	-0.162
	[0.158]	[0.134]	[0.074]	[0.259]	[0.154]	[0.066]	[0.269]
Amarela (Asians)	*0.347*	*-0.044*	*-0.187*	*-0.095*	*0.262*	*0.562*	*-0.101*
	-0.091	-0.091	-0.108**	-0.062	-0.103	-0.079**	-0.148
	[0.092]	[0.083]	[0.043]	[0.163]	[0.083]	[0.033]	[0.137]
Parda (MR)	*-0.150*	*-0.126*	*-0.110*	*-0.052*	*-0.097*	*-0.084*	*-0.117*
	0.467	0.038	-0.215	-0.168	0.202	0.553**	-0.122
	[0.475]	[0.415]	[0.229]	[0.905]	[0.523]	[0.237]	[0.927]
Parental schooling							
Mean parental schooling (years	*0.043*	*0.043*	*0.048*	*0.025*	*0.037*	*0.039*	*0.041*
	0.061***	0.044***	0.048***	0.024	0.035**	0.038***	0.033
	[0.019]	[0.014]	[0.008]	[0.025]	[0.014]	[0.005]	[0.021]
Mother/father difference (years	*0.009*	*-0.004*	*0.000*	*-0.004*	*0.004*	*-0.003*	*0.001*
	0.016	-0.004	-0.001	-0.004	0.004	-0.003	-0.001
	[0.016]	[0.014]	[0.007]	[0.023]	[0.013]	[0.005]	[0.019]
Region dummies							
South East (omitted)							
North	*0.429*	*-0.076*	*-0.014*	*-0.080*	*-0.091*	*-0.024*	*-0.093*
	0.451*	-0.068	-0.036	-0.101	-0.111	-0.014	-0.109
	[0.267]	[0.142]	[0.092]	[0.265]	[0.158]	[0.062]	[0.264]
North East	*-0.239*	*-0.314*	*-0.289*	*-0.246*	*-0.266*	*-0.264*	*-0.241*
	-0.240**	-0.312***	-0.286***	-0.250*	-0.249***	-0.267***	-0.239
	[0.102]	[0.079]	[0.043]	[0.151]	[0.093]	[0.034]	[0.149]
South	*-0.118*	*-0.083*	*-0.008*	*0.005*	*0.022*	*0.047*	*-0.024*
	-0.052	-0.047	-0.017	0.004	0.069	0.052	-0.028
	[0.110]	[0.092]	[0.051]	[0.169]	[0.106]	[0.039]	[0.153]
Center-West	*-0.430*	*-0.382*	*-0.168*	*-0.137*	*-0.088*	*-0.014*	*-0.134*
	-0.361*	-0.422**	-0.162*	-0.155	-0.129	-0.004	-0.107
	[0.188]	[0.174]	[0.087]	[0.279]	[0.153]	[0.057]	[0.205]
Years of schooling	*0.003*	*0.008*	*-0.003*	*-0.029*	*-0.017*	*-0.041*	*-0.038*
	-0.037	-0.004	0.002	-0.08	-0.018	-0.044***	-(0.06)
	[0.040]	[0.032]	[0.016]	[0.069]	[0.031]	[0.013]	[0.056]
Years of schooling-squared	*0.004*	*0.006*	*0.007*	*0.008*	*0.007*	*0.009*	*0.006*
	0.006***	0.007***	0.007***	0.010***	0.007***	0.009***	0.008**
	[0.002]	[0.002]	[0.001]	[0.003]	[0.002]	[0.001]	[0.003]
Migrant dummy	*-0.041*	*0.101*	*0.102*	*0.031*	*0.052*	*0.075*	*0.098*
	0.119	0.103	0.107***	0.089	0.052	0.080***	0.127
	[0.075]	[0.063]	[0.037]	[0.121]	[0.070]	[0.026]	[0.108]
Father's occupational status	yes	yes	yes	yes	yes	yes	yes
Labor Market status							
Formal employee (omitted)							
Informal employee	*-0.325*	*-0.104*	*-0.210*	*-0.081*	*-0.195*	*-0.170*	*-0.044*
	-0.13	-0.084	-0.203***	-0.116	-0.194**	-0.172***	-0.132
	[0.094]	[0.079]	[0.047]	[0.156]	[0.088]	[0.034]	[0.135]
Self employed	*-0.357*	*0.037*	*-0.013*	*0.162*	*0.005*	*0.059*	*0.308*
	-0.072	0.061	0.003	0.135	0.009	0.057*	0.126
	[0.081]	[0.070]	[0.042]	[0.140]	[0.085]	[0.033]	[0.140]
Employer	*0.329*	*0.367*	*0.717*	*0.618*	*0.619*	*0.448*	*0.859*
	0.571***	0.365***	0.730***	0.657**	0.624***	0.455***	0.483
	[0.186]	[0.140]	[0.090]	[0.303]	[0.168]	[0.063]	[0.296]
Constant	*0.662*	*0.175*	*0.131*	*0.407*	*0.127*	*0.059*	*0.383*
	0.315	0.215	0.079	0.613	0.276	0.097	0.503
	[0.219]	[0.175]	[0.097]	[0.462]	[0.198]	[0.085]	[0.377]

a) Dependent variable is the log of hourly wage rate. b) For each variable, we present three values: the unbiased mean coefficient estimates (in italics), two-stage Heckman estimates and related standard errors (in brackets); * significant at 10%; ** significant at 5%; *** significant at 1% c) Coefficients for the father's occupational status dummy variables are here omitted to save space, though available from the authors on request.

Table 2.c: First Stage Selection Equations by cohort, women (cont'd).

	b1936_40	b1941_45	b1946_50	b1951_55	b1956_60	b1961_65	b1966_70
SELECTION EQUATION ***(Dep.var.=1 if hourly wages>0)***							
Race dummies							
Branca (whites, omitted)							
Preta (Black)	0.169	-0.307	0.134	0.092	-0.046	-0.183	0.363
	[0.268]	[0.213]	[0.213]	[0.204]	[0.175]	[0.202]	[0.248]
Amarela (Asians)	4.848	-1.267**	-0.786	4.575	4.595	4.059	4.576
	[0.000]	[0.539]	[0.537]	[0.000]	[0.000]	[0.000]	[0.000]
Parda (MR)	-0.058	-0.275**	-0.116	-0.207**	-0.021	-0.162	0.146
	[0.152]	[0.137]	[0.116]	[0.104]	[0.109]	[0.108]	[0.123]
Years of schooling	0.195***	0.142***	0.091***	0.086**	-0.008	0.03	-0.097*
	[0.046]	[0.043]	[0.035]	[0.034]	[0.040]	[0.040]	[0.058]
Years of schooling-squared	-0.009***	-0.003	-0.001	-0.001	0.006*	0.004	0.014***
	[0.004]	[0.003]	[0.002]	[0.002]	[0.003]	[0.003]	[0.005]
Family type dummies							
Couple with children (omitted)							
Couple, no children	0.053	-0.246	-0.334*	0.016	0.038	0.376	-0.081
	[0.210]	[0.182]	[0.171]	[0.213]	[0.235]	[0.300]	[0.200]
Mother with children	0.388**	0.372**	0.453***	0.321**	0.525***	0.647***	0.822***
	[0.172]	[0.154]	[0.141]	[0.127]	[0.165]	[0.239]	[0.302]
Others	0.306	-0.159	0.029	-0.142	4.681	4.756	0.392
	[0.256]	[0.246]	[0.281]	[0.299]	[0.000]	[0.000]	[0.452]
Number of children	0.117	-0.064	-0.069*	-0.03	-0.043	-0.069*	-0.001
	[0.077]	[0.042]	[0.036]	[0.031]	[0.032]	[0.039]	[0.054]
Region dummies							
South East (omitted)							
North	-0.772***	-0.009	-0.515***	0.25	0.055	-0.034	0.006
	[0.277]	[0.250]	[0.192]	[0.214]	[0.225]	[0.232]	[0.241]
North East	0.272	0.052	0.058	0.071	-0.114	-0.067	-0.01
	[0.169]	[0.145]	[0.131]	[0.116]	[0.121]	[0.127]	[0.137]
South	-0.22	-0.309**	-0.280**	-0.143	-0.371***	-0.338***	0.034
	[0.163]	[0.155]	[0.133]	[0.124]	[0.121]	[0.126]	[0.146]
Center-West	0.072	5.535	-0.223	-0.025	0.119	-0.371**	-0.009
	[0.304]	[0.000]	[0.214]	[0.209]	[0.234]	[0.168]	[0.202]
Household income (excluding own)	0.083	0.078	-0.022	0.085	0.142**	0.033	0.288***
	[0.075]	[0.067]	[0.057]	[0.054]	[0.057]	[0.063]	[0.071]
Constant	0.109	0.801**	1.532***	0.997***	1.113***	1.567***	0.221
	[0.433]	[0.374]	[0.331]	[0.308]	[0.315]	[0.352]	[0.403]
Self-selection correction term	-1.035**	-1.086**	-0.191	-3.195	-2.005**	-0.554	-2.754**
	[0.507]	[0.494]	[0.372]	[2.170]	[0.994]	[0.385]	[1.246]
Standard error of residual	1.03	1.09	0.76	3.2	2.01	0.75	2.75
Number of obs	812	1276	2129	3006	3425	3354	2554
Censored obs	92	105	115	116	112	111	94

a) Dependent variable is participation. b) Two-stage Heckman estimates, standard errors in brackets; * significant at 10%; ** significant at 5%; *** significant at 1%

Table 3.a: Schooling determinants, men. a), b)

	b1936_40	b1941_45	b1946_50	b1951_55	b1956_60	b1961_65	b1966_70
Race							
Branca (whites, omitted)							
Preta (Black)	-1.505***	-1.247***	-0.906***	-1.205***	-1.638***	-1.480***	-1.032***
	[0.344]	[0.298]	[0.257]	[0.227]	[0.214]	[0.203]	[0.213]
Amarela (Asians)	5.967***	2.788***	1.999***	2.761***	1.694**	2.416***	-1.329
	[1.281]	[0.659]	[0.707]	[0.709]	[0.670]	[0.695]	[0.982]
Parda (MR)	-0.826***	-1.192***	-1.156***	-1.117***	-1.145***	-1.238***	-0.729***
	[0.193]	[0.170]	[0.144]	[0.124]	[0.112]	[0.105]	[0.111]
Parental schooling							
Mean parental schooling (years)	0.742***	0.745***	0.684***	0.622***	0.585***	0.552***	0.494***
	[0.042]	[0.034]	[0.027]	[0.023]	[0.019]	[0.017]	[0.018]
Mother/father difference (years)	0.075*	-0.014	0.058**	0.004	-0.012	0.015	0.005
	[0.043]	[0.037]	[0.029]	[0.025]	[0.020]	[0.018]	[0.019]
Region dummies							
South East (omitted)							
North	-1.374***	-0.733	-0.688**	-0.391	-0.298	-0.418*	-0.493*
	[0.487]	[0.474]	[0.344]	[0.297]	[0.254]	[0.252]	[0.265]
North East	-0.853***	-0.953***	-0.586***	-1.037***	-1.085***	-0.768***	-0.987***
	[0.197]	[0.169]	[0.145]	[0.130]	[0.120]	[0.114]	[0.120]
South	0.008	-0.305	-0.596***	-0.486***	-0.288**	-0.135	-0.305**
	[0.240]	[0.215]	[0.173]	[0.142]	[0.126]	[0.118]	[0.132]
Center-West	0.287	0.39	-0.332	0.373	-0.172	0.239	-0.19
	[0.541]	[0.418]	[0.338]	[0.283]	[0.234]	[0.209]	[0.207]
Father's occupational status							
Rural workers (omitted)							
Domestic servants	0.441**	0.451**	0.400**	0.708***	0.760***	0.972***	0.522***
	[0.200]	[0.180]	[0.157]	[0.143]	[0.136]	[0.136]	[0.149]
Traditional sector workers	1.890***	1.503***	1.775***	1.636***	1.547***	1.531***	1.360***
	[0.335]	[0.307]	[0.237]	[0.196]	[0.176]	[0.166]	[0.175]
Service sector workers	1.927***	2.201***	2.797***	2.538***	2.278***	2.132***	2.045***
	[0.347]	[0.329]	[0.266]	[0.221]	[0.190]	[0.182]	[0.193]
Modern industry workers	2.521***	2.599***	2.707***	2.945***	3.024***	2.678***	1.701***
	[0.387]	[0.328]	[0.249]	[0.208]	[0.175]	[0.165]	[0.178]
Self-employed shopkeepers	4.190***	3.342***	4.029***	3.331***	3.439***	3.228***	2.651***
	[0.391]	[0.332]	[0.262]	[0.230]	[0.211]	[0.202]	[0.227]
Technitians, artists and desk workers	3.920***	2.883***	3.817***	3.188***	2.961***	2.702***	2.032***
	[0.449]	[0.416]	[0.303]	[0.252]	[0.213]	[0.200]	[0.224]
Employers	2.026***	2.570***	2.506***	2.747***	2.670***	2.219***	1.988***
	[0.405]	[0.361]	[0.305]	[0.271]	[0.255]	[0.264]	[0.281]
Liberal professionals	3.291***	2.485***	2.951***	3.098***	2.584***	2.446***	2.288***
	[0.842]	[0.553]	[0.430]	[0.373]	[0.323]	[0.301]	[0.338]
Constant	2.756***	3.491***	3.857***	4.499***	4.804***	4.909***	5.138***
	[0.174]	[0.167]	[0.145]	[0.129]	[0.123]	[0.126]	[0.138]
Sample size	1760	2462	3702	4844	5427	5594	4509
Adj R-squared	0.44	0.44	0.44	0.41	0.44	0.42	0.35

a) Dependent variable is years of schooling. b) OLS estimates standard errors in brackets;* significant at 10%; ** significant at 5%; *** significant at 1%

Table 3.b: Schooling determinants, women. a), b)

	b1936_40	b1941_45	b1946_50	b1951_55	b1956_60	b1961_65	b1966_70
Race							
Branca (whites, omitted)							
Preta (Black)	0.433	-1.083**	-1.878***	-1.059***	-1.824***	-1.610***	-1.411***
	[0.425]	[0.422]	[0.325]	[0.281]	[0.256]	[0.275]	[0.278]
Amarela (Asians)	2.145**	1.601	3.053***	0.857	2.641***	3.159***	1.625
	[1.082]	[1.123]	[0.908]	[0.887]	[0.810]	[0.903]	[1.016]
Parda (MR)	-0.455*	-0.843***	-1.392***	-1.359***	-0.654***	-1.152***	-1.167***
	[0.251]	[0.257]	[0.188]	[0.159]	[0.140]	[0.135]	[0.149]
Parental schooling							
Mean parental schooling (years)	0.820***	0.719***	0.730***	0.620***	0.607***	0.517***	0.476***
	[0.050]	[0.049]	[0.034]	[0.028]	[0.023]	[0.021]	[0.023]
Mother/father difference (years)	0.125**	-0.032	0.069**	0.080***	0.095***	0.037*	0.077***
	[0.049]	[0.052]	[0.034]	[0.027]	[0.023]	[0.021]	[0.022]
Region dummies							
South East (omitted)							
North	-0.007	-0.162	0.268	0.389	0.051	-0.047	-0.087
	[0.662]	[0.581]	[0.434]	[0.343]	[0.313]	[0.307]	[0.343]
North East	-0.529**	-0.308	-0.06	-0.117	-0.733***	-0.305**	-0.525***
	[0.251]	[0.258]	[0.189]	[0.168]	[0.149]	[0.146]	[0.160]
South	-0.092	-0.315	-0.590***	-0.432**	-0.511***	-0.484***	-0.722***
	[0.301]	[0.301]	[0.218]	[0.187]	[0.161]	[0.152]	[0.167]
Center-West	0.101	-0.092	-0.563	0.731**	0.064	0.326	0.009
	[0.562]	[0.595]	[0.415]	[0.349]	[0.288]	[0.255]	[0.252]
Father's occupational status							
Rural workers (omitted)							
Domestic servants	0.662**	0.585**	0.744***	0.641***	0.989***	1.231***	1.085***
	[0.262]	[0.270]	[0.207]	[0.185]	[0.171]	[0.174]	[0.198]
Traditional sector workers	1.689***	2.495***	2.177***	1.815***	1.723***	2.141***	1.762***
	[0.453]	[0.430]	[0.306]	[0.278]	[0.227]	[0.218]	[0.232]
Service sector workers	1.738***	2.306***	3.169***	2.235***	2.300***	2.501***	2.111***
	[0.462]	[0.465]	[0.347]	[0.286]	[0.254]	[0.235]	[0.254]
Modern industry workers	3.462***	2.770***	3.246***	2.734***	2.883***	3.070***	2.668***
	[0.455]	[0.491]	[0.319]	[0.268]	[0.229]	[0.217]	[0.231]
Self-employed shopkeepers	4.109***	3.354***	4.482***	3.954***	3.927***	4.010***	3.331***
	[0.500]	[0.454]	[0.345]	[0.287]	[0.249]	[0.244]	[0.279]
Technitians, artists and desk workers	2.006***	2.356***	3.149***	3.144***	3.146***	3.509***	2.592***
	[0.574]	[0.612]	[0.394]	[0.320]	[0.279]	[0.261]	[0.283]
Employers	1.547***	1.571***	3.324***	2.847***	2.619***	2.743***	3.031***
	[0.502]	[0.596]	[0.427]	[0.355]	[0.321]	[0.329]	[0.370]
Liberal professionals	3.229***	3.758***	3.344***	2.770***	2.815***	2.870***	2.133***
	[0.893]	[0.868]	[0.529]	[0.456]	[0.368]	[0.345]	[0.415]
Constant	1.842***	3.301***	3.704***	4.698***	4.924***	5.087***	5.719***
	[0.240]	[0.251]	[0.190]	[0.168]	[0.156]	[0.159]	[0.185]
Sample size	899	1439	2373	3360	3938	3939	2976
Adj R-squared	0.48	0.34	0.44	0.38	0.39	0.4	0.36

a) Dependent variable is years of schooling. b) OLS estimates standard errors in brackets;* significant at 10%; ** significant at 5%; *** significant at 1%

Table 4.a: Probit estimates for migration, men. a), b)

	b1936_40	b1941_45	b1946_50	b1951_55	b1956_60	b1961_65	b1966_70
Race							
Branca (whites, omitted)							
Preta (Black)	0.007	-0.02	-0.024	-0.046	-0.066**	-0.047	-0.102***
	[0.049]	[0.041]	[0.035]	[0.032]	[0.032]	[0.032]	[0.034]
Amarela (Asians)	.	0.158**	0.190***	.	0.032	0.219***	0.131
	.	[0.068]	[0.061]	.	[0.094]	[0.084]	[0.143]
Parda (MR)	-0.055*	-0.029	-0.015	-0.007	-0.008	-0.023	-0.032*
	[0.029]	[0.024]	[0.019]	[0.017]	[0.016]	[0.016]	[0.018]
Parental schooling							
Mean parental schooling (yea	-0.006	-0.011**	-0.006	-0.002	-0.003	-0.005*	0
	[0.006]	[0.004]	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Mother/father difference (yea	-0.003	-0.008*	0	-0.006*	-0.004	-0.001	-0.002
	[0.006]	[0.005]	[0.004]	[0.003]	[0.003]	[0.003]	[0.003]
Region dummies							
South East (omitted)							
North	-0.130*	-0.111*	-0.043	-0.106**	-0.011	0.06	0.025
	[0.075]	[0.067]	[0.047]	[0.042]	[0.037]	[0.037]	[0.041]
North East	0.114***	0.118***	0.104***	0.120***	0.117***	0.168***	0.162***
	[0.027]	[0.022]	[0.018]	[0.017]	[0.017]	[0.016]	[0.018]
South	0.022	0.067**	0.056***	0.048**	0.133***	0.149***	0.185***
	[0.033]	[0.027]	[0.021]	[0.019]	[0.017]	[0.017]	[0.019]
Center-West	0.009	0.019	0.03	0.137***	0.139***	0.135***	0.125***
	[0.076]	[0.054]	[0.042]	[0.033]	[0.030]	[0.028]	[0.029]
Father's occupational status							
Rural workers (omitted)							
Domestic servants	-0.038	-0.058**	-0.102***	-0.046**	-0.044**	-0.021	-0.084***
	[0.030]	[0.025]	[0.022]	[0.021]	[0.020]	[0.021]	[0.024]
Traditional sector workers	-0.057	-0.07	-0.079**	-0.139***	-0.127***	-0.083***	-0.137***
	[0.050]	[0.044]	[0.034]	[0.028]	[0.027]	[0.026]	[0.028]
Service sector workers	-0.083	-0.134***	-0.132***	-0.169***	-0.081***	-0.125***	-0.139***
	[0.054]	[0.048]	[0.038]	[0.032]	[0.029]	[0.028]	[0.031]
Modern industry workers	-0.149**	-0.133***	-0.168***	-0.178***	-0.158***	-0.129***	-0.186***
	[0.059]	[0.048]	[0.036]	[0.030]	[0.026]	[0.026]	[0.028]
Self-employed shopkeepers	-0.069	-0.107**	-0.167***	-0.142***	-0.151***	-0.141***	-0.178***
	[0.058]	[0.048]	[0.038]	[0.034]	[0.032]	[0.031]	[0.036]
Technitians, artists and desk w	-0.180***	-0.063	-0.221***	-0.230***	-0.126***	-0.102***	-0.183***
	[0.069]	[0.059]	[0.044]	[0.036]	[0.032]	[0.031]	[0.035]
Employers	-0.001	-0.013	-0.122***	-0.122***	-0.046	-0.068	-0.105**
	[0.059]	[0.050]	[0.044]	[0.039]	[0.038]	[0.042]	[0.045]
Liberal professionals	-0.226*	0.011	-0.077	-0.150***	-0.185***	-0.094**	-0.128**
	[0.133]	[0.074]	[0.061]	[0.054]	[0.048]	[0.047]	[0.054]
Sample size	1685	2402	3651	4783	5389	5557	4486
Pseudo R-squared	0.03	0.02	0.03	0.03	0.03	0.03	0.04

a) Dependent variable is migrant status. b) Probit marginal effect estimates reported (standard errors in brackets);* significant at 10%; ** significant at 5%; *** significant at 1%

Table 4.b: Probit estimates for migration, women. a), b)

	b1936_40	b1941_45	b1946_50	b1951_55	b1956_60	b1961_65	b1966_70
Race							
Branca (whites, omitted)							
Preta (Black)	-0.024	-0.117**	-0.032	-0.065*	-0.087**	-0.113***	-0.064
	[0.068]	[0.054]	[0.043]	[0.037]	[0.036]	[0.040]	[0.043]
Amarela (Asians)		0.094	0.261***	0.131	0.174*	0.09	-0.398***
		[0.122]	[0.059]	[0.094]	[0.092]	[0.124]	[0.116]
Parda (MR)	-0.092**	-0.107***	0.014	-0.004	-0.004	0.007	-0.001
	[0.040]	[0.032]	[0.024]	[0.020]	[0.019]	[0.019]	[0.023]
Parental schooling							
Mean parental schooling	-0.014*	-0.005	-0.014***	-0.006*	-0.005*	-0.008**	-0.003
	[0.008]	[0.006]	[0.004]	[0.003]	[0.003]	[0.003]	[0.003]
Mother/father difference	-0.012*	0	-0.003	-0.003	-0.006**	-0.005	0.002
	[0.008]	[0.006]	[0.004]	[0.003]	[0.003]	[0.003]	[0.003]
Region dummies							
South East (omitted)							
North	0.003	-0.1	-0.096*	-0.061	-0.045	-0.018	-0.026
	[0.097]	[0.073]	[0.058]	[0.044]	[0.043]	[0.044]	[0.052]
North East	0.175***	0.072**	0.082***	0.108***	0.132***	0.092***	0.122***
	[0.035]	[0.030]	[0.023]	[0.020]	[0.019]	[0.020]	[0.023]
South	-0.024	-0.016	0.084***	0.080***	0.082***	0.125***	0.181***
	[0.046]	[0.036]	[0.026]	[0.022]	[0.021]	[0.021]	[0.024]
Center-West	0.029	0.03	0.092*	0.133***	0.080**	0.081**	0.146***
	[0.082]	[0.068]	[0.048]	[0.037]	[0.037]	[0.035]	[0.035]
Father's occupational status							
Rural workers (omitted)							
Domestic servants	-0.064	0.024	-0.057**	-0.070***	-0.048**	-0.021	-0.085***
	[0.042]	[0.032]	[0.028]	[0.025]	[0.024]	[0.025]	[0.031]
Traditional sector workers	0.001	-0.063	-0.184***	-0.206***	-0.177***	-0.088***	-0.110***
	[0.072]	[0.053]	[0.042]	[0.037]	[0.032]	[0.032]	[0.036]
Service sector workers	-0.160**	-0.015	-0.07	-0.181***	-0.161***	-0.095***	-0.157***
	[0.078]	[0.056]	[0.047]	[0.038]	[0.035]	[0.034]	[0.038]
Modern industry workers	-0.123	0.001	-0.179***	-0.221***	-0.130***	-0.110***	-0.172***
	[0.076]	[0.058]	[0.044]	[0.036]	[0.032]	[0.031]	[0.035]
Self-employed shopkeepe	-0.166**	-0.171***	-0.063	-0.170***	-0.105***	-0.156***	-0.148***
	[0.084]	[0.058]	[0.047]	[0.039]	[0.035]	[0.035]	[0.042]
Technitians, artists and de	-0.126	-0.102	-0.048	-0.191***	-0.208***	-0.121***	-0.223***
	[0.096]	[0.077]	[0.053]	[0.043]	[0.038]	[0.038]	[0.040]
Employers	0.067	0.029	-0.004	-0.078	-0.044	-0.055	-0.160***
	[0.073]	[0.070]	[0.057]	[0.048]	[0.045]	[0.048]	[0.055]
Liberal professionals	-0.116	-0.154	-0.105	-0.218***	-0.181***	-0.121**	-0.130**
	[0.147]	[0.112]	[0.072]	[0.061]	[0.051]	[0.050]	[0.062]
Sample size	894	1439	2373	3360	3938	3939	2976
Pseudo R-squared	0.05	0.02	0.03	0.04	0.03	0.03	0.04

a) Dependent variable is migrant status. b) Probit marginal effect estimates (standard errors in brackets);* significant at 10%; ** significant at 5%; *** significant at 1%

Table 5.a: Multinomial logit estimates for labor market status, men.

	b1936_40	b1941_45	b1946_50	b1951_55	b1956_60	b1961_65	b1966_70
INFORMAL EMPLOYEE							
Race							
Branca (whites, omitted)							
Preta (Black)	0.034 *	0.082 ***	0.006	0.062 ***	0.064 ***	0.060 ***	0.104 ***
Amarela (Asians)	-0.165 *	0.055	-0.028	-0.043	-0.004	0.149	0.184
Parda (MR)	0.048 **	0.039 **	0.020	0.036 ***	0.037 ***	0.056 ***	0.043 ***
Parental schooling							
Mean parental schooling (years)	-0.005	-0.012 ***	-0.011 ***	-0.004	-0.009 ***	-0.006 ***	-0.005 **
Mother/father difference (years)	-0.004	0.006	0.002	0.003	0.006	0.000	-0.005 *
Region dummies							
South East (omitted)							
North	0.045	-0.016	0.001	0.029	0.041 **	0.042 *	0.024
North East	-0.021	0.000	0.009	0.004	0.017	0.007	0.036 **
South	-0.020	0.012	-0.001	-0.014	-0.003	0.006	-0.005
Center-West	0.100 *	0.078 **	0.029	0.003	0.067 ***	0.040 **	0.076 ***
Father's occupational status							
Rural workers (omitted)							
Domestic servants	-0.070 ***	-0.025	-0.013	-0.010	-0.010	-0.026 *	-0.031 *
Traditional sector workers	-0.108 ***	-0.049 *	-0.057 **	-0.019	-0.026	-0.030 *	-0.053 ***
Service sector workers	-0.075 **	-0.056 *	-0.038	-0.058 ***	-0.022	-0.039 **	-0.044 *
Modern industry workers	-0.069 *	-0.026	-0.014	-0.042 **	-0.028	-0.054 ***	-0.069 ***
Self-employed shopkeepers	-0.048	-0.055 *	-0.031	-0.049 **	-0.049 **	-0.041 *	-0.052 *
Technitians, artists and desk workers	-0.021	0.065	-0.034	-0.041 *	-0.033	-0.043 *	-0.079 ***
Employers	-0.075 *	0.003	-0.075 **	-0.055 **	-0.048	-0.058 *	-0.075 **
Liberal professionals	-0.088	-0.029	0.022	-0.024	0.034	-0.045	-0.075 *
SELF-EMPLOYED							
Race							
Branca (whites, omitted)							
Preta (Black)	-0.020 *	-0.169 ***	-0.041	-0.066 *	-0.056 *	-0.004	-0.010
Amarela (Asians)	0.139 ***	-0.184	0.000	-0.105	-0.072	-0.060	0.111
Parda (MR)	-0.027	-0.016	-0.007	0.001	0.006	-0.003	0.006
Parental schooling							
Mean parental schooling (years)	-0.005	-0.008	-0.005	-0.011 ***	-0.005 *	-0.003	-0.002
Mother/father difference (years)	-0.003	-0.005	-0.003	0.006	0.001	0.001	0.002
Region dummies							
South East (omitted)							
North	0.102 *	0.087	0.187 ***	0.065 *	0.084 ***	0.088 ***	0.159 ***
North East	0.130 ***	0.075 ***	0.089 ***	0.058 ***	0.059 ***	0.038 **	0.077 ***
South	0.056	0.013	0.085 ***	0.010	0.007	0.017	0.019
Center-West	-0.037	-0.038	0.064	0.055	0.045	0.023	0.068 ***
Father's occupational status							
Rural workers (omitted)							
Domestic servants	0.092 ***	0.057 **	0.062 ***	0.070 ***	0.024	0.009	0.064 ***
Traditional sector workers	0.126 **	0.113 **	0.029	0.039	0.020	-0.019	0.044 *
Service sector workers	0.100 *	0.089 *	0.025	-0.005	0.006	-0.039	0.051 *
Modern industry workers	0.014	-0.027	-0.013	-0.001	-0.037	-0.017	0.002
Self-employed shopkeepers	0.059	0.038	0.006	0.021	0.017	0.006	0.052 **
Technitians, artists and desk workers	-0.063	-0.001	-0.020	-0.010	-0.058 *	-0.048	0.007
Employers	0.091	-0.096	-0.035	0.030	-0.033	0.023	0.044
Liberal professionals	0.102	-0.070	0.000	0.014	0.029	-0.040	0.077

Table 5.a: Multinomial logit estimates for labor market status, men. Cont'd.

	b1936_40	b1941_45	b1946_50	b1951_55	b1956_60	b1961_65	b1966_70
EMPLOYER							
Race							
Branca (whites, omitted)							
Preta (Black)	-0.077 ***	-0.034	-0.038 *	-0.042 **	-0.024	-0.042 ***	-0.029 **
Amarela (Asians)	-0.006 ***	0.053	0.083	0.021	-0.040	0.001	0.010
Parda (MR)	-0.012 ***	-0.040 ***	-0.033 ***	-0.041 ***	-0.029 ***	-0.031 ***	-0.013 **
Parental schooling							
Mean parental schooling (years)	0.001 *	0.010 ***	0.007 ***	0.010 ***	0.009 ***	0.004 ***	0.003 ***
Mother/father difference (years)	0.001 *	0.004	0.001	0.000	0.001	0.000	0.000
Region dummies							
South East (omitted)							
North	-0.002	0.020	-0.019	-0.007	-0.022	-0.023	-0.012
North East	-0.006 **	-0.013	-0.009	-0.026	-0.006	-0.005	-0.016 **
South	-0.005 **	0.009	-0.010	0.012	0.010	0.003	-0.011
Center-West	-0.003	0.020	0.008	0.007	-0.004	0.028 **	-0.009
Father's occupational status							
Rural workers (omitted)							
Domestic servants	0.009 **	0.021 *	0.024 *	0.025 *	0.028 **	0.034 **	0.013
Traditional sector workers	0.001	-0.026	0.025	0.027	0.005	0.015	-0.019
Service sector workers	0.008	-0.001	0.018	0.026	0.020	0.030 *	0.006
Modern industry workers	0.013 *	0.025	0.028	0.022	0.023	0.015	0.035 **
Self-employed shopkeepers	0.025 ***	0.104 ***	0.134 ***	0.097 ***	0.113 ***	0.116 ***	0.116 ***
Technitians, artists and desk workers	0.003	-0.002	0.037	-0.033 *	0.007	0.023	0.016
Employers	0.049 ***	0.109 ***	0.234 ***	0.101 ***	0.188 ***	0.125 ***	0.101 ***
Liberal professionals	-0.004	0.083	0.022	0.017	0.003	0.072 **	0.065 **
Observations	1691	2402	3651	4805	5389	5557	4486
PseudoR2	0.05	0.04	0.04	0.04	0.04	0.03	0.04

Formal employee is the base category. Marginal effect estimates. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 5.b: Multinomial logit estimates for labor market status, women.

	b1936_40	b1941_45	b1946_50	b1951_55	b1956_60	b1961_65	b1966_70
INFORMAL EMPLOYEE							
Race							
Branca (whites, omitted)							
Preta (Black)	-0.023	0.063	0.149 ***	0.086 ***	0.050 *	0.125 ***	0.171 ***
Amarela (Asians)	0.154	-0.025 ***	-0.200 ***	-0.028	-0.008	-0.223 ***	0.006
Parda (MR)	0.043	0.038	0.044 ***	0.071 ***	0.038 **	0.067 ***	0.104 ***
Parental schooling							
Mean parental schooling (years)	-0.022 ***	-0.025 ***	-0.014 ***	-0.014 ***	-0.021 ***	-0.020 ***	-0.016 ***
Mother/father difference (years)	-0.001	0.000	-0.003	-0.008 **	0.000	0.000	-0.006
Region dummies							
South East (omitted)							
North	-0.066	-0.089	-0.078 **	-0.069 **	-0.061 **	0.020	-0.038
North East	-0.063 *	-0.060 *	-0.039 **	-0.064 ***	-0.018	0.020	-0.009
South	-0.025	0.035	-0.012	-0.003	0.010	0.024	0.028
Center-West	0.007	0.026	0.031	-0.023	0.017	0.020	0.042
Father's occupational status							
Rural workers (omitted)							
Domestic servants	-0.022	-0.043	0.004	-0.011	-0.042 **	-0.054 **	-0.036
Traditional sector workers	-0.076	-0.009	-0.016	-0.038	-0.059 **	-0.036	-0.068 **
Service sector workers	-0.124 *	-0.022	-0.042	-0.063 **	-0.074 **	-0.116 ***	-0.050
Modern industry workers	0.047	-0.083	-0.052	-0.101 ***	-0.105 ***	-0.086 ***	-0.112 ***
Self-employed shopkeepers	-0.110 *	-0.054	-0.057	-0.119 ***	-0.126 ***	-0.129 ***	-0.104 ***
Technitians, artists and desk workers	-0.019	-0.030	-0.007	-0.088 ***	-0.082 **	-0.082 **	-0.153 ***
Employers	-0.075	0.046	-0.066	-0.061 *	-0.087 **	-0.114 ***	-0.109 **
Liberal professionals	-0.054	0.141	-0.057	-0.121 **	-0.091 *	-0.133 ***	-0.081
SELF-EMPLOYED							
Race							
Branca (whites, omitted)							
Preta (Black)	-0.195 **	-0.079	-0.125 **	-0.098 **	-0.039	-0.019	-0.092 **
Amarela (Asians)	-0.311	0.410 ***	-0.267 ***	0.064	-0.137	0.056 ***	0.060
Parda (MR)	-0.041	0.010	-0.001	-0.010	0.015	0.022	-0.008
Parental schooling							
Mean parental schooling (years)	0.006	-0.003	-0.013 **	-0.004	-0.005	-0.005 *	-0.006 *
Mother/father difference (years)	-0.017 *	-0.004	-0.002	0.006 *	-0.002	0.003	-0.005 *
Region dummies							
South East (omitted)							
North	0.269 ***	0.058	0.077 *	0.073 **	0.098 ***	-0.033	0.100 **
North East	0.088 *	0.101 ***	0.080 ***	0.066 ***	0.084 ***	0.069 ***	0.076 ***
South	0.101 *	0.036	0.016	-0.018	0.008	0.003	-0.016
Center-West	-0.010	-0.007	-0.013	-0.066	0.004	0.045	-0.006
Father's occupational status							
Rural workers (omitted)							
Domestic servants	0.051	0.054	0.034	0.000	0.049 **	-0.004	0.072 **
Traditional sector workers	0.071	-0.071	0.028	0.019	0.029	-0.017	0.009
Service sector workers	0.183 **	0.006	-0.002	0.055	0.044	0.016	0.075 *
Modern industry workers	-0.132	-0.006	0.011	-0.015	-0.006	-0.020	0.024
Self-employed shopkeepers	0.011	0.029	-0.016	0.045 *	0.009	0.027 **	0.006
Technitians, artists and desk workers	-0.011	0.028	-0.023	-0.032	0.010	-0.078 *	0.042
Employers	0.115	0.031	0.031	-0.026	0.042	-0.060	0.025
Liberal professionals	-0.038	-0.200 *	0.027	-0.009	-0.012	0.085 **	0.015

Table 5.b: Multinomial logit estimates for labor market status, women. Cont'd.

	b1936_40	b1941_45	b1946_50	b1951_55	b1956_60	b1961_65	b1966_70
EMPLOYER							
Race							
Branca (whites, omitted)							
Preta (Black)	-0.009 ***	-0.035 ***	-0.019	-0.030 **	-0.020	-0.029 ***	-0.017 **
Amarela (Asians)	-0.001 ***	0.008	0.194 ***	0.098 **	0.127 **	0.045 ***	-0.020 ***
Parda (MR)	-0.001 ***	-0.002	-0.024 ***	-0.020 ***	-0.025 *	-0.012 *	-0.008
Parental schooling							
Mean parental schooling (years)	0.000 *	0.001 ***	0.004 ***	0.003 ***	0.002 ***	0.004 ***	0.003 ***
Mother/father difference (years)	0.000	0.000	0.002	0.001	0.001	-0.001	0.000
Region dummies							
South East (omitted)							
North	-0.002 ***	-0.002	-0.026 **	0.010	-0.011	-0.019 *	-0.010
North East	0.001	0.000	-0.001	0.014 *	0.001	-0.007	-0.001
South	-0.001	0.001	-0.006	0.004	-0.002	0.001	-0.002
Center-West	0.000	0.000	-0.004	0.019	0.001	-0.009	0.002
Father's occupational status							
Rural workers (omitted)							
Domestic servants	0.001	0.003	0.010	0.014	0.008	0.074 ***	0.012
Traditional sector workers	0.000	0.000	-0.009	0.027	0.005	0.073 **	0.003
Service sector workers	0.002	0.010 **	-0.005	-0.018	0.027	0.132 ***	-0.004
Modern industry workers	0.002	-0.001	0.014	0.009	0.021	0.088 ***	0.032 *
Self-employed shopkeepers	0.002	0.013 ***	-0.002	0.041 **	0.061 ***	0.170 ***	0.054 ***
Technitians, artists and desk workers	0.003	-0.001	-0.003	-0.007	0.043 **	0.073 ***	0.031 *
Employers	0.010 ***	0.011 **	0.013	0.019	0.084 ***	0.208 **	0.125 ***
Liberal professionals	0.000	0.001	0.017	0.003	0.067 **	0.080 ***	0.027
Observations	899	1439	2373	3360	3938	3939	2976
PseudoR2	0.06	0.05	0.04	0.04	0.04	0.05	0.05

Formal employee is the base category. Marginal effect estimates. * significant at 10%; ** significant at 5%; *** significant at 1%

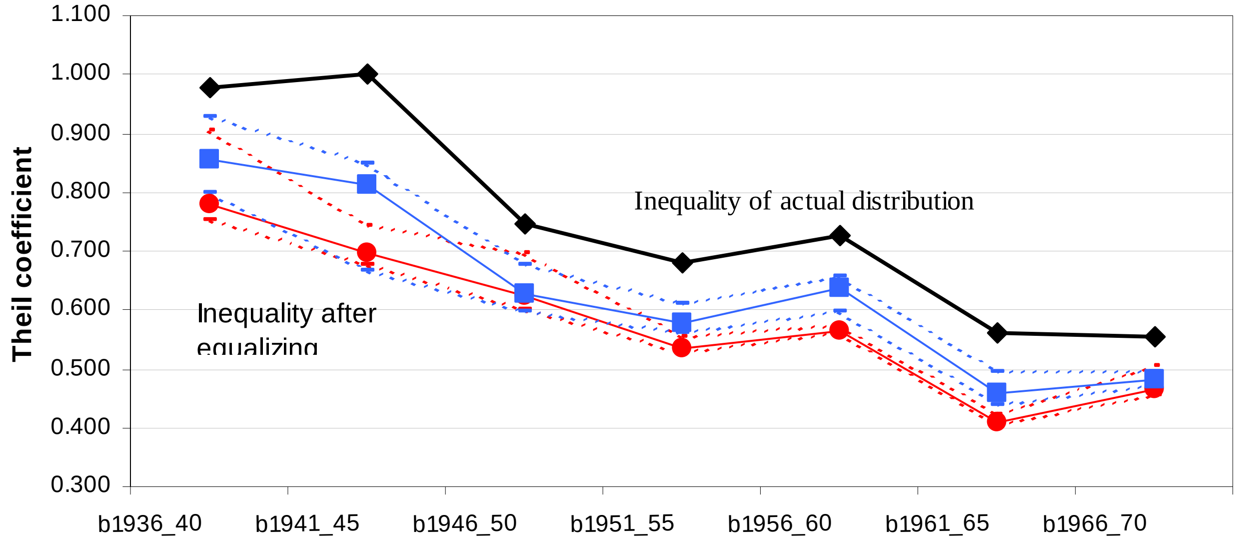
Table 6. Effects of equalizing circumstances on inequality. Simulation results.
Theil coefficient for 5-year cohorts of men.

	b1936_40	b1941_45	b1946_50	b1951_55	b1956_60	b1961_65	b1966_70
Total Observed Inequality (1)	**0.977**	**1.001**	**0.746**	**0.680**	**0.728**	**0.561**	**0.553**
"Complete" effect of equalizing circumstances							
(Upper bound estimate)	(0.904)	(0.742)	(0.696)	(0.554)	(0.575)	(0.424)	(0.505)
Mean estimate (2a)	0.780	0.698	0.626	0.536	0.565	0.410	0.466
(Lower bound estimate)	(0.752)	(0.678)	(0.602)	(0.527)	(0.561)	(0.405)	(0.457)
Share =((1)-(2a))/(1)	**0.202**	**0.302**	**0.161**	**0.212**	**0.223**	**0.269**	**0.157**
"Partial" effect of equalizing circumstances							
(Upper bound estimate)	(0.927)	(0.848)	(0.675)	(0.611)	(0.657)	(0.494)	(0.494)
Mean estimate (2b)	0.854	0.811	0.629	0.579	0.636	0.459	0.480
(Lower bound estimate)	(0.800)	(0.667)	(0.597)	(0.560)	(0.598)	(0.440)	(0.471)
Share =((1)-(2b))/(1)	**0.126**	**0.190**	**0.157**	**0.148**	**0.126**	**0.183**	**0.132**
Treating observed efforts as circumstance variables							
(Upper bound estimate)	(0.701)	(0.749)	(0.670)	(0.443)	(0.452)	(0.337)	(0.415)
Mean estimate (2c)	0.638	0.483	0.422	0.432	0.436	0.317	0.370
(Lower bound estimate)	(0.618)	(0.454)	(0.410)	(0.424)	(0.423)	(0.314)	(0.356)
Share =((1)-(2c))/(1)	**0.347**	**0.518**	**0.434**	**0.365**	**0.401**	**0.435**	**0.332**

Table 7. Effects of equalizing circumstances on inequality. Simulation results.
Theil coefficient for 5-year cohorts of women.

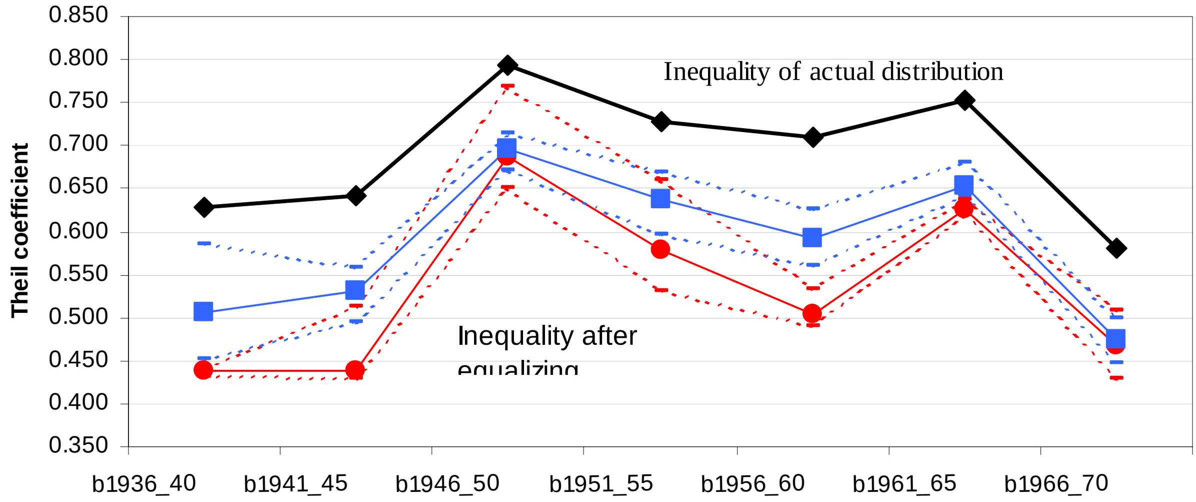
	b1936_40	b1941_45	b1946_50	b1951_55	b1956_60	b1961_65	b1966_70
Total Observed Inequality (1)	**0.629**	**0.643**	**0.795**	**0.728**	**0.710**	**0.754**	**0.581**
"Complete" effect of equalizing circumstances							
(Upper bound estimate)	(0.441)	(0.513)	(0.768)	(0.660)	(0.532)	(0.637)	(0.507)
Mean estimate (2a)	0.437	0.438	0.686	0.579	0.503	0.626	0.468
(Lower bound estimate)	(0.432)	(0.430)	(0.651)	(0.531)	(0.490)	(0.620)	(0.429)
Share =((1)-(2a))/(1)	**0.305**	**0.319**	**0.136**	**0.205**	**0.291**	**0.169**	**0.194**
"Partial" effect of equalizing circumstances							
(Upper bound estimate)	(0.585)	(0.557)	(0.714)	(0.670)	(0.625)	(0.681)	(0.500)
Mean estimate (2b)	0.507	0.531	0.697	0.637	0.593	0.652	0.474
(Lower bound estimate)	(0.451)	(0.494)	(0.672)	(0.597)	(0.560)	(0.643)	(0.448)
Share =((1)-(2b))/(1)	**0.195**	**0.174**	**0.122**	**0.125**	**0.165**	**0.135**	**0.184**
Treating observed efforts as circumstance variables							
(Upper bound estimate)	(0.419)	(0.322)	(0.444)	(0.426)	(0.363)	(0.443)	(0.365)
Mean estimate (2c)	0.406	0.312	0.430	0.387	0.346	0.421	0.353
(Lower bound estimate)	(0.397)	(0.305)	(0.420)	(0.370)	(0.341)	(0.399)	(0.341)
Share =((1)-(2c))/(1)	**0.354**	**0.514**	**0.459**	**0.468**	**0.513**	**0.442**	**0.392**

**Figure 1. Inequality Decomposition into Opportunity and Residual Components.
Theil coefficient for 5-year cohorts of men.**



a) Partial and complete effect shown respectively on intermediate and bottom curves. Dotted lines correspond to upper and lower bounds.

**Figure 2. Inequality Decomposition into Opportunity and Residual Components.
Theil coefficient for 5-year cohorts of women.**



a) Partial and complete effect shown respectively on intermediate and bottom curves. Dotted lines correspond to upper and lower bounds.

Figure 3: Complete effect of equalizing individual circumstance variables on inequality. Theil coefficient for 5-year cohorts of men.

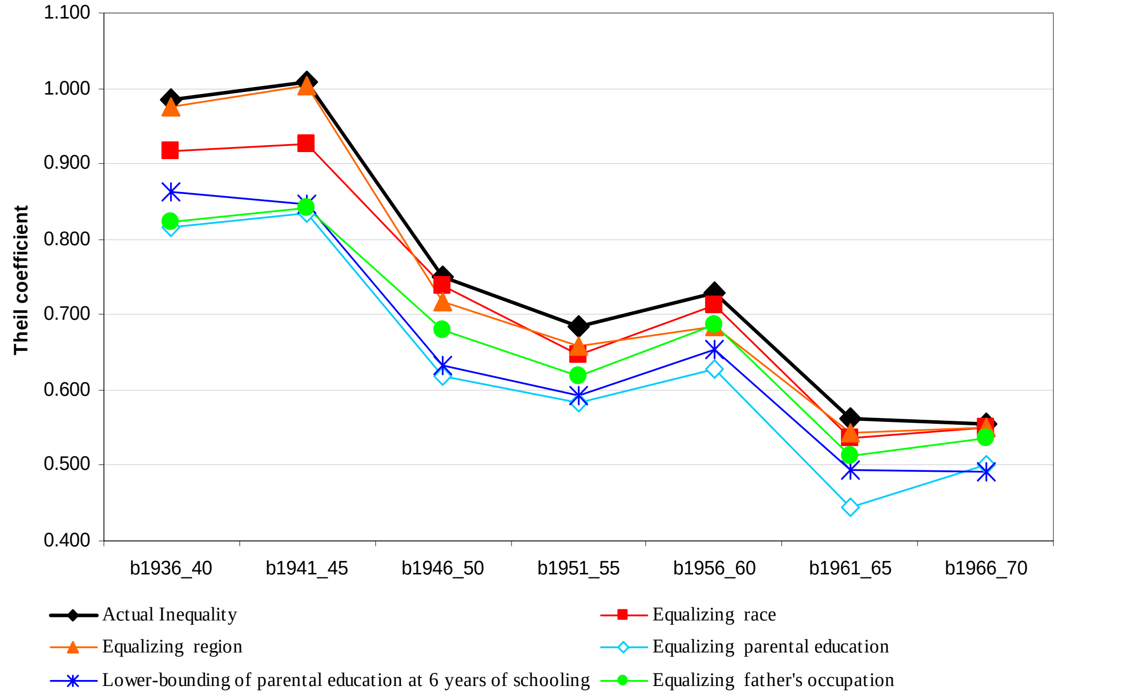


Figure 4: Complete effect of equalizing individual circumstance variables on inequality. Theil coefficient for 5-year cohorts of women.