

INEQUALITY OF OPPORTUNITY IN BRAZIL

BY FRANÇOIS BOURGUIGNON, FRANCISCO H. G. FERREIRA*

The World Bank

AND

MARTA MENÉNDEZ

Université Paris Dauphine and INRA, Paris-Jourdan

This paper proposes a measure of the contribution of unequal opportunities to earnings inequality. Drawing on the distinction between “circumstance” and “effort” variables in John Roemer’s work on equality of opportunity, we associate inequality of opportunities with five observed circumstances which lie beyond the control of the individual—father’s and mother’s education; father’s occupation; race; and region of birth. The paper provides a range of estimates of the importance of these opportunity-forming circumstances in accounting for earnings inequality in one of the world’s most unequal countries. We also 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 distribution of male earnings in urban Brazil, in 1996. The five observed circumstances are found to account for between 10 and 37 percent of the Theil index, depending on cohort and allowing for the possibility of biased coefficient estimates due to unobserved correlates. On average, 60 percent of this impact operates through the direct effect on earnings. Parental education is the most important circumstance affecting earnings, but the occupation of the father and race also play a role.

1. INTRODUCTION

Income inequality has many sources, not all of which are equally objectionable. One reasonable distinction is that inequality in the opportunities available to people—their basic life chances—is more objectionable than inequalities which arise because of the differential application of individual effort. In the words of Peragine (2004), “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” (p. 11).

John Roemer (1998) offered an influential formalization of the concept of unequal opportunities, suggesting that one should separate the determinants of a person’s *advantage* (i.e. desirable outcomes, such as incomes or status) into *circumstances* and *efforts*. *Circumstances* are factors which are economically exogenous to the person, such as her gender, race, family background or place of birth.

Note: 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, Ricardo Paes de Barros, Richard Blundell, Sam Bowles, Pedro Carneiro, Jishnu Das, Quy-Toan Do, Hanan Jacoby, Yoko Kijima, Stephan Klasen, Phillippe Leite, Costas Meghir, Thomas Piketty, Martin Ravallion, two anonymous referees and participants at many conferences and seminars for helpful comments. All remaining errors are ours.

*Correspondence to: Francisco H. G. Ferreira, The World Bank, 1818 H. St. NW, Washington, DC 20433, U.S. (fferreira@worldbank.org).

© 2007 The Authors

Journal compilation © 2007 International Association for Research in Income and Wealth Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main St, Malden, MA, 02148, USA.

They may affect a person's outcomes, but can not be influenced by the individual. *Efforts*, on the other hand, are outcome determinants which can be affected by individual choice. Roemer suggested a precise definition of an equal-opportunity policy: once the population is partitioned into homogeneous-circumstance *types* (i.e. groups where everyone shares exactly the same set of circumstances), then an equal-opportunity policy is a policy that equalizes "advantage" for each centile of the effort distribution, across individual types.¹

Practical applications of this approach have remained scarce, however, partly because the analysis becomes rather cumbersome as the number of "types" increases. In fact, although many authors in the fields of ethics and social choice have argued that opportunity—rather than income or some other observable outcome—should become the "currency of egalitarian justice,"² empirical usage of the concept has remained relatively 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 (observed) inequality of opportunity associated with an empirical distribution of incomes or earnings. The method derives directly from Roemer's conceptual framework, and can easily be applied to fairly standard household data, provided that the survey contains a modicum of information on a person's family background. The basic approach is to divide all observed earnings determinants into those which can be regarded as exogenous to the individual, in the sense that they cannot be influenced by her actions, and all others. Following Roemer, we refer to the first set of variables—which might include sex, race, place of birth, family wealth, parental education, or family background more generally—as "circumstance" variables. The essence of our approach is to simulate the reduction in earnings inequality which would attain if differences in these circumstance variables were eliminated. This difference between observed and counterfactual earnings inequality is then interpreted as a measure of observed inequality of opportunity. Unlike other approaches recently proposed, we take account of the fact that other earnings determinants (including "efforts")—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.

We apply this approach to the distribution of male hourly earnings in urban Brazil, exploiting the fact that the 1996 Brazilian household survey includes information on parental education and father's occupation, for a large subset of surveyed adults. We divide the sample into seven age groups, and analyze each cohort separately. Our results suggest that between 10 and 37 percent of observed earnings inequality among Brazilian males can be attributed to a set of only five exogenous circumstance variables: race; place of birth; mother's and father's education; and father's occupation. On average, some 60 percent of the effect of these circumstances operates directly through earnings, while the remaining 40 percent or so operates by affecting the level of efforts expended by individuals. We find

¹Roemer (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.

²Notably 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.

some evidence that the share of inequality attributable to observed circumstances is lower for younger cohorts. Finally, we investigate the relative importance of each individual circumstance variable separately and find that parental education is the dominant circumstance variable, accounting for a larger share of earnings inequality than race, parental occupation or region of birth. This is in contrast to some of the recent findings for the U.S., where race plays a quantitatively important role, as briefly discussed in the next section.

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 component attributable to unequal opportunities and a residual. Section 5 discusses our estimation and identification strategies. 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

In contrast to the normative literature, empirical work on inequality of opportunity is still relatively rare. A couple of papers sought 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) asked what reallocation of educational expenditures would equalize opportunities across four types of individuals in the U.S. (using the National Longitudinal Survey of Young Men). Interestingly, they found that race was a more important partitioning variable than parental education in their sample. In a related study, Page and Roemer (2001) investigated the extent to which the fiscal system could be seen as an "opportunity equalizing device" in the United States. These authors also focused on race and parental education (which they interpreted as a proxy for socioeconomic background) as the key circumstance variables, and found that the U.S. 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. A larger group of authors have extended this analysis of fiscal systems as opportunity-equalizers to other countries in Roemer *et al.* (2003).

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 seek to measure inequality of opportunity itself. Second, they are restricted to differences across a few (usually four) large groups, determined by a limited set of circumstances. As we will see, our regression-based decomposition allows for a finer treatment of circumstances.

A second and more recent set of studies is more closely related to our paper. In the specific context of measuring intergenerational mobility, van de Gaer *et al.* (2001) propose an "index of inequality of opportunity" based on the expectation of a person's income, conditional on her parents' income class and on the probabilities in the (relevant row of the) transition matrix between the two generations. While the measure is conceptually attractive in the context of measuring mobility, it is not ideal for applications in which *multiple* circumstances—rather than a

unique measure of parental income or socioeconomic status—are thought to influence opportunities. Lefranc *et al.* (2006) relate the Roemerian notion of inequality of opportunity to the concept of stochastic dominance between income distributions conditional on type. They propose two alternative “empirical definitions” of inequality of opportunity—the weaker one relying on second-order dominance, and the stronger one corresponding to first-order dominance. Types are defined by parental education and occupation, and the conditional distributions are compared across types. They apply this approach to the distributions of household income across a number of OECD countries and find that, with the exception of Sweden, the equal-opportunity hypothesis could be rejected for both definitions in all countries. Checchi and Peragine (2005) decomposed observed inequalities in earnings and in cognitive abilities (as measured by reading literacy scores for 15-year-olds) in Italy into a component due to a single exogenous circumstance—parental education—and a residual component, attributed to efforts. They use two alternative non-parametric approaches to assess inequality of opportunity: eliminating inequality (in the means) across “types,” and measuring inequality within “tranches” (groups at identical percentiles across types).

Another closely related paper is Cogneau and Gignoux’s (2005) study of unequal opportunities and earnings inequality in Brazil. Using data from four waves of Brazil’s main household survey (the PNAD), they construct two measures of inequality of opportunity for active men aged 40–49 in each survey. Whereas, in this paper, we emphasize the measurement of the component of earnings inequality that is attributable to observed differences in opportunities, Cogneau and Gignoux (2005) focus on the dynamics of inequality (in both earnings and opportunities), and their relation to changes in three factors: (i) changes in the earnings structure, or “returns”; (ii) changes in the marginal distributions of education and social origin (for a given intergenerational mobility matrix); and (iii) changes in the intergenerational mobility matrix (for given marginal distributions of education and social origin). Like Ferreira and Paes de Barros (1999), they find that earnings inequality rose from 1976 to 1988, and then declined until 1996. While the increase from 1976 to 1988 can be explained in part by the inequality-increasing nature of the educational expansion for cohorts born before the Second World War, the decline from 1988 to 1996 is accounted for in part by a reversal in the nature of the educational expansion—which became equalizing for the younger cohorts—and in part by declining returns to schooling. Although these authors report an increase in the degree of intergenerational mobility in the later years, they argue that it had little effect on the observed trajectory of earnings inequality.³

Given the importance of family background (and, more specifically, parental education) as a key circumstance variable, our paper is also related to the larger empirical literature on intergenerational mobility, which dates 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 status of interest, and y_{-1} denotes the same variable for a

³Cogneau and Gignoux (2005) and Lefranc *et al.* (2006) both cite the working paper version of this paper: Bourguignon *et al.* (2003), where we apply a variant of the regression-based approach proposed here to a broader set of income concepts, including household income per capita.

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 inequality. Its complement, $1 - \beta$, is often interpreted as a measure of intergenerational mobility.⁴ Two excellent surveys of this literature exist, in Solon (1999) and Mulligan (1999), and we do not replicate their work here.⁵

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 *et al.* (2000) for Latin American countries. Paes de Barros and Lam (1993), and Lam (1999) applied similar methods to Brazil.

Two recent papers have estimated Galtonian regressions for Brazil, using the same earnings data that 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). Dunn 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.

Ferreira and Veloso (2006) estimate Brazilian intergenerational elasticities for a broader age range (25–64), using a two-sample IV procedure with parental occupation and education as instruments. They first estimate "parent" earnings equations on earlier PNAD samples, and then use the estimated coefficients to predict earnings for the parents of the workers in the 1996 sample. In the second stage, they estimate earnings of sons as a function of the predicted earnings of their parents, and other control variables. Their preferred point estimate for the intergenerational elasticity is 0.58, but they also report evidence of non-linearities in income mobility, with lower mobility for the children of poorer fathers. They also show that mobility patterns differ across regions and races in Brazil.⁶

Although it is clearly possible to view the coefficient β in the Galtonian regression $\ln y = \alpha + \beta \ln y_{-1} + \varepsilon$ as an indicator of inequality of opportunity, the similarities between these studies and our approach should not be overstated. When properly estimated, as in Ferreira and Veloso (2006), β is an excellent (inverse) measure of *intergenerational earnings mobility*. However, it would only correspond to a measure of *inequality of opportunity* under the clearly restrictive assumption that parental earnings is a sufficient statistic for all observed circum-

⁴Behrman 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.

⁵It 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).

⁶See also the related paper on intergenerational educational mobility in Brazil, by Ferreira and Veloso (2003).

stances. Consider, for instance, the role of race, which Betts and Roemer (1999) and Page and Roemer (2001) found to be an important determinant of opportunities in the United States, even after controlling for socio-economic background. If race is included as a control variable in the second stage IV estimation of the Galtonian regression $\ln y = \alpha + \beta \ln y_{-1} + X\gamma + \varepsilon$, its effect as a circumstance is *not* captured by the coefficient on (predicted) parental earnings. The same is true for region of birth. That is exactly as it should be if one is interested in isolating the effect of socio-economic background from that of race or geography. Ferreira and Veloso's (2006) estimates are, therefore, probably the best estimates one can attain for *intergenerational mobility* using Brazil's PNAD data. They are *not*—and are not intended to be—estimates of *inequality of opportunity*.⁷ While the two concepts are closely related, they are not the same.

A second but related reason why the intergenerational mobility analysis differs from our approach is methodological. Since measurement error in y_{-1} may lead to attenuation bias in the estimate of the elasticity parameter β , a common approach is to instrument for parental earnings, typically with a vector of parental characteristics, D . One might run the two-stage least squares where $\ln y_{-1} = D\delta + \eta$, and $\ln y = \alpha + \beta(D\hat{\delta}) + \varepsilon$ where, in some cases, the two equations are run in separate samples altogether. As noted by Solon (1992), however, such a procedure will lead to an upward bias in the estimate of the elasticity, if the instruments in D are correlated with other determinants of children's earnings.⁸ Because we include the variables in D in our main regression, and address identification through an alternative approach (discussed in detail in Section 5), we avoid this particular problem altogether.

While the literature on intergenerational mobility and the more recent papers on inequality of opportunity are clearly related—because family background is a key determinant of opportunities—they are not perfect substitutes. The former seeks to measure the transmission of one specific economic indicator (generally earnings or incomes). To this end, it actually seeks to separate out the effect of other circumstances, such as race, gender or geography. The latter seeks to measure the *aggregate* effect of all observed circumstances, including but not exclusively family background, on current inequalities. Whether or not parental background is the most important circumstance determining opportunities must vary across countries and time periods, and can not be determined *ex-ante*. In this paper, in addition, we 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

⁷It turns out that parental education and occupation account for a much greater share of inequality of opportunity in Brazil than race or place of birth, but this could *not* be inferred from the results in the intergenerational mobility literature. In fact, as shown by Betts and Roemer (1999), race does play a large role in determining opportunities in the United States, and direct measurement of inequality of opportunities is needed to ascertain whether or not it is equally important in other contexts.

⁸In fact, Ferreira and Veloso (2006) note that “the degree of wage persistence is significantly smaller when only occupation is used as an instrument for father's wage (0.52) than when only education is used as instrument (0.60). This suggests that the use of father's education as an instrument may produce an upward bias in the persistence estimates, which justifies our choice of a broader set of instruments” (p. 193). Using two instruments does not, of course, fully resolve the problem, if both are correlated with other determinants of children's earnings.

insertion in the labor market. The two approaches are therefore ultimately best seen as complements, rather than as substitutes.

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 Estatística* (IBGE), Brazil's Census Bureau.⁹ The survey is nationally 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 active males aged 26–60, who report positive earnings. This sample was chosen in order to focus on individuals with the highest levels of labor market attachment. Participation rates are lower for women, as well as for men outside this age range, and participation decisions are also influenced by circumstances. Correcting for the resulting sample selection biases introduces additional complications to our method, and we choose to report only the results for the prime-age male sample here. Results for females, correcting for sample selection, are presented in Bourguignon *et al.* (2003). We study the 1996 survey because, for that year, information is available on the education of both parents and on the occupation of the fathers of all surveyed household heads and spouses.¹¹

The complete PNAD 1996 sample size is upwards of 330,000 individuals. After excluding women, individuals living in rural areas, those outside the 26–60 age range, those who are not household heads or spouses, and individuals that do not report positive earnings, we are left with a sample of 37,548. Due to missing entries for some of the relevant variables (mostly parental education or father's occupation), the sample is further reduced to 28,474 occupied men, which are representative of the universe of prime-age active urban males in Brazil.

Since non-response to questions about parental background is likely to be non-random, one might be concerned about sample selection arising from the exclusion of observations with missing entries. Table A1 in the Appendix shows that missing information on these variables was frequent, at approximately 15 percent of the sample. It was fairly evenly distributed across cohorts, except for the youngest age group (where it was higher). Table A2 compares the composition of our final sample (in column 1) with the comparable sample including those individuals for whom parental education and occupation are missing in urban areas (column 2); and the corresponding sample for both urban and rural areas (column 3). Given the sample size, the differences across columns are generally statistically significant at the usual levels. Nevertheless, it is apparent that the two urban samples are very

⁹The PNAD is an annual survey, but it is not fielded on census years.

¹⁰Ferreira *et al.* (2003) discuss a number of shortcomings with the rural income data in the PNAD, which lead us to conduct our main analysis for urban areas only. Urban areas accounted for some 80 percent of Brazil's population in 1996. We did, nevertheless, replicate our analysis for a joint urban and rural sample, as a robustness check. We return to this point below.

¹¹This information is not generally collected in the PNAD, but was also available in a special supplement in 1996.

similar in almost every respect. Even the education differential, which one might have expected to be large, is less than one tenth of a year of schooling.

Further reassurance is provided by Table A3, which reports coefficients from an earnings regression on all the variables used in the subsequent analysis, except for the parental background variables, run in both our final sample and in the complete sample. If selective non-response were likely to introduce large biases, one would expect these coefficients to differ substantively. In the event, coefficients are statistically indistinguishable (at the 95 percent confidence level) between the two urban samples in every single case. Selective non-response is a potential problem for all studies that use this kind of data (including those that focus on inter-generational mobility, as discussed in the previous section). Fortunately, the results presented in this Appendix suggest that, while non-response is statistically selective, the impact on the estimated coefficients which will be used in our analysis is negligible.

Tables A1–A3 also include a column where comparable rural workers are included. As indicated earlier, rural areas are excluded from the main analysis, because the income questions asked of agricultural producers in the PNAD questionnaire are not detailed enough to generate confidence in their estimates. Our analysis is therefore only representative of urban areas and, as the Appendix tables reveal, the rural and urban samples really are quite different. The results in this paper should not be extrapolated beyond the urban areas.¹²

Our final 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 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.” The individual circumstance variables available in the data set are: dummies for regions of birth; a race dummy variable; parental education expressed as the number of years of schooling;¹³ and the occu-

¹²Of course, inequality of opportunity in urban and rural areas may well be different. One feature of our urban data set is that, since it is representative of the active urban male population in 1996, it includes individuals who were born in rural areas but currently reside in urban areas. It excludes, however, those who were born and have remained in rural areas. As a result, our estimates of the migration coefficient should not be interpreted as representing the impact of migration for the national population. A comparison between columns 2 and 3 in Table A3 does reveal a higher coefficient for migration in the joint sample. While this has no bearing on the measurement of inequality of opportunity for today’s urban population (which does include migrants and excludes rural non-migrants), it reiterates the point that inequality of opportunity for the whole country is different than that measured in the urban areas alone. We do find, in robustness analysis, that the opportunity share of inequality is larger in rural areas, and we return to this issue in Section 7. We thank an anonymous referee for raising this point.

¹³Parental education is reported in discrete levels in the PNAD. They were converted into years of schooling (here in brackets) using the following rule: no school or incomplete 1st grade (0); complete 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 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).

pational position of the father. Our occupational categories were based on a nine-level occupational classification used by Brazilian sociological studies on occupational mobility (see Pero, 2001; Valle Silva, 1978). We have regrouped the original nine categories into three: lower status (rural workers (1); domestic servants (2); traditional sector workers (3)); medium status (service sector workers (4); modern industry workers (5); self-employed shopkeepers (6)); and higher status (technicians, artists and desk workers (7); employers (8); liberal professionals (9)).

Other earnings determinants which are included in the data set are the individual's own schooling attainment, measured in years,¹⁴ 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") or an employer; an informal employee ("sem carteira"); or self-employed ("conta própria"). Since these three variables can be affected by choices made by the individual himself, they are classified as "efforts." 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 "circumstance" variables: earnings determinants over which individuals have no control. Our main goal is to estimate the reduction in inequality which would attain if these "circumstances" had no effect on earnings or, equivalently, if there were no differences in people's circumstances. It is this reduction which we will take as a measure of the contribution of inequality in observed opportunity to 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 influenced by individual decisions—"efforts." We will keep inverted commas on that term throughout, however, to acknowledge that these variables may in general be influenced by circumstances, and may also include random shocks such as luck. Denote earnings by w , and let them be distributed in the population according to $\Phi(w)$. Denoting 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:

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

¹⁴The 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 percent of the sample.

TABLE 1
DESCRIPTIVE STATISTICS, BY 5-YEAR BIRTH COHORTS

	b1936_40	b1941_45	b1946_50	b1951_55	b1956_60	b1961_65	b1966_70	All
Mean monthly earnings (Reais, all jobs)	698.8	957.6	1,023.9	979.5	892.3	784.1	626.4	854.4
Mean number of years of schooling	4.6	5.5	6.2	6.9	7.3	7.6	7.2	6.8
Mean father's number of years of schooling	2.1	2.4	2.7	2.8	3.1	3.4	3.4	3.0
Mean mother's number of years of schooling	1.6	2.0	2.3	2.4	2.8	3.1	3.0	2.6
Race (percent)								
White & Asian	60.2%	59.5%	60.0%	60.8%	59.1%	59.7%	57.4%	59.4%
Black & MR	39.8%	40.5%	40.0%	39.2%	40.9%	40.3%	42.6%	40.6%
Regions (percent)								
North & North East	38.8%	39.4%	38.4%	34.2%	34.1%	34.4%	37.9%	36.1%
South & South East	58.0%	56.2%	56.8%	60.7%	59.9%	58.5%	53.6%	57.9%
Center-West	3.2%	4.4%	4.7%	5.1%	6.1%	7.0%	8.5%	6.0%
Migrants (percent)	70.1%	69.7%	69.1%	66.2%	63.2%	59.7%	58.9%	64.1%
Father's occupational status (percent)								
Lower status	70.8%	68.2%	63.3%	60.8%	56.3%	53.5%	53.9%	59.0%
Medium status	18.2%	19.0%	21.9%	23.9%	27.7%	29.8%	30.5%	25.8%
Higher status	11.1%	12.9%	14.8%	15.3%	16.0%	16.7%	15.6%	15.2%
Labor market status (percent)								
Formal employees & employers	43.3%	48.3%	54.1%	59.2%	61.0%	60.9%	61.5%	57.7%
Informal employees	15.4%	12.5%	12.2%	11.2%	12.2%	13.9%	16.0%	13.2%
Self-employed	41.3%	39.2%	33.7%	29.5%	26.8%	25.2%	22.4%	29.1%
Number of observations	1,730	2,457	3,726	4,877	5,488	5,643	4,553	28,474

Noting that circumstance variables are economically exogenous by definition, but that “effort” variables can be affected by circumstances, as well as by unobserved determinants, v , implies that:¹⁵

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

Equality of opportunities, in the sense of Roemer (1998), would attain if the distribution of earnings were independent of circumstances. Denoting by $F(w|X)$ the distribution of earnings conditional on variables X , equality of opportunities requires that $F(w|C) = F(w)$. Given equation (2), this independence will generally require *both* that (i) $\frac{\partial f(C, E, u)}{\partial C} = 0, \forall C$ and that (ii) $G(E|C) = G(E)$, where $G(E|C)$ is the distribution of efforts conditional on circumstances. The first condition states that circumstances have no *direct* effect on earnings, controlling for efforts, whereas the second requires that efforts be distributed independently from circumstances (which implies that circumstances have no causal effect on efforts). One useful benchmark case is *equality of circumstances*: $C_i = \bar{C}, \forall i$. In this case, condition (ii) is satisfied automatically, and condition (i) is redundant.¹⁶ Although not necessary, equality of circumstances is sufficient for equality of opportunities.

Seeking to measure *inequality* of opportunity is, therefore, seeking to measure the extent to which the observed joint distribution $\{w, C, E\}$ deviates from the conditions that define equality of opportunities. It will prove useful to rely on the benchmark case of equality of circumstances, under which equation (2) would generate a marginal distribution of earnings $\tilde{\Phi}(\tilde{w})$, where $\tilde{\Phi}$ is the cumulative distribution function of individual earnings under the counterfactual that circumstances are the same for everyone, i.e. $\tilde{w}_i = f[\bar{C}, E(\bar{C}, v_i), u_i]$. To assess the extent to which unequal opportunities affect the distribution of earnings, one is interested in comparing the observed earnings distribution $\Phi(w)$ with $\tilde{\Phi}(\tilde{w})$. Since we are interested in *inequality* of opportunities, an inequality measure I defined over the distributions $\Phi(w)$ and $\tilde{\Phi}(\tilde{w})$ would seem to be the appropriate metric for comparison. It seems natural to define the *overall* (observed) opportunity share of earnings inequality as:

$$(3) \quad \Theta_I := \frac{I(\Phi) - I(\tilde{\Phi})}{I(\Phi)}.$$

The notation Θ_I indicates that the opportunity share of inequality is contingent on the chosen inequality index.

Θ_I measures the *overall* effect of circumstances on earnings because both the direct effect of circumstances on earnings (through $\frac{\partial f(C, E, u)}{\partial C}$) and the indirect

¹⁵The distinction between economic exogeneity (that a variable is determined outside the model) and econometric exogeneity (that a variable is uncorrelated with the residual term in a regression) is important in this paper. Circumstance variables are economically exogenous by definition but, since not all relevant variables are observed, they need not be econometrically exogenous.

¹⁶ \bar{C} is any constant vector. Its elements could, for instance, take the values of the cross-sectional sample means of each element of the vector of circumstances.

effect through efforts (i.e. through $G(E|C)$) are captured. Holding circumstances constant in both places where they enter into (2) annuls both effects. It may, however, be interesting to ascertain the relative importance of the direct and indirect effects, and this can be achieved straightforwardly by an additional decomposition, namely:

$$(4) \quad \Theta_i^d := [I(\Phi) - I(\Phi^d)](I(\Phi))^{-1}$$

where $\Phi^d(w^d)$ is the distribution of simulated earnings as given by:

$$w_i^d = f[\bar{C}, E(C_i, v_i), u_i].$$

If the overall effect is given by Θ_i and the direct (or partial) effect is given by Θ_i^d , then the indirect effect is then 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* them. The indirect effect is the effect of circumstances on earnings through observed “efforts.” The next section discusses an empirical approach to carrying out these two decompositions in practice.

5. ESTIMATION STRATEGY

To implement the decompositions proposed in equations (3) and (4), an estimate is needed of equation (2). An empirically suitable first approximation can be obtained by log-linearization:

$$(5) \quad \ln(w_i) = C_i\alpha + E_i\beta + u_i$$

$$(6) \quad E_i = HC_i + v_i$$

where α and β are two vectors of coefficients, and H 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.¹⁷ u_i and v_i are white-noise disturbances. The remainder of the notation follows from the previous section.

The main empirical problem with estimating equations (5) and (6) is essentially one of omitted variables. As discussed in Section 3, five circumstance variables are observed in our PNAD 1996 dataset: race (R), parental schooling (represented here 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 $C = (R, GR, MPE, DPE, FO)$. The three observed

¹⁷A prominent example is that an individual’s own level of schooling is generally thought to be influenced by family background. This influence of parental background on education outcomes may reflect the fact that more educated parents provide more “home inputs” into an “education production function,” such as books, vocabulary and quality time spent on homework (see, e.g. Hanushek, 1986). It may also reflect the intergenerational transmission of different beliefs about the returns to effort, which may vary across families (see, e.g. Piketty, 1995).

“effort” variables are: years of schooling (S), a migration dummy (M) and labor market status (L). The model (5)–(6) could thus be written out in full as:

$$(7) \quad \ln(w_i) = \alpha_0 + R_i\alpha_R + GR_i\alpha_G + MPE_i\alpha_P + DPE_i\alpha_D + FO_i\alpha_F + S_i\beta_S + M_i\beta_M + L_i\beta_L + u_i$$

$$(8) \quad E_i = b_0 + R_i b_R + GR_i b_G + MPE_i b_P + DPE_i b_D + FO_i b_F + v_i$$

where $E = (S, M, L)$.

Although, to our knowledge, this set of circumstance and effort variables is richer than those used so far in the existing empirical literature on inequality of opportunity, it is still clearly the case that not all circumstance and “effort” variables are observed. A number of relevant circumstances (such as parental wealth, the quality of care received in early childhood, the quality of the inherited genetic endowment, etc) are not observed. Similarly, a variety of “effort” variables (such as effort in job search and in the actual work) are also unobserved. To the extent that these unobserved variables are correlated with the observed circumstance (and effort) variables, the residual terms in (7) and (8) are not orthogonal to the regressors.

Fortunately, for the purpose of conducting the decompositions defined in equations (3) and (4), it is not necessary to estimate the “structural” model (5)–(6). Substituting (6) into (5) generates the reduced form:

$$(9) \quad \ln(w_i) = C_i(\alpha + \beta H) + v_i\beta + u_i$$

which can be estimated by OLS as

$$(10) \quad \ln(w_i) = C_i\psi + \varepsilon_i$$

where $\psi = \alpha + \beta H$ and $\varepsilon_i = v_i\beta + u_i$.

If the estimates of ψ were satisfactory, then the overall (observed) opportunity share of inequality defined in equation (3) could be obtained by expressing $\tilde{w}_i = f[\bar{C}, E(\bar{C}, v_i), u_i]$ as $\tilde{w}_i = \exp[\bar{C}\hat{\psi} + \hat{\varepsilon}_i]$. The partial or direct (observed) opportunity share of earnings inequality can, in turn, be calculated by estimating (5) directly:

$$(5') \quad \ln(w_i) = C_i\alpha + E_i\beta + u_i$$

and computing the counterfactual wages: $w_i^d = \exp[\bar{C}\hat{\alpha} + E_i\hat{\beta} + \hat{u}_i]$.

That the reduced form and the single equation (5) (or (7), in full) suffice for our purposes is very helpful, in that it eliminates the need for worrying about econometric endogeneity in the full system (7)–(8). The problem, remains, of course, that if ε is not orthogonal to C , then $(\hat{\psi})$ will be a biased estimator of $\alpha + \beta H$ and, if u is not orthogonal to (C, E) , then $(\hat{\alpha}, \hat{\beta})$ will be a biased estimator of (α, β) . This remains likely, of course, because of the unobserved circumstance and effort variables contained in those residual terms. Furthermore, an instrumental variable strategy is unlikely to succeed, since it is difficult to conceive of

correlates of the circumstance variables that would not themselves have any direct influence on earnings. Most instruments generally used in the literature on returns on education (such as parental education, ability scores, education supply side variables) are unlikely to satisfy sensible exclusion restrictions in the present context.

In the absence of an adequate set of instrumental variables, the only solution is to explore the likely magnitude of the potential biases in the estimation of α and β due to the correlations between u , C and E , when estimating (5), and between ε and C when estimating (10). In what follows, we use Monte-Carlo methods to consider a wide range of estimates consistent with the condition that the variance-covariance matrix of observed variables and unobserved effects must be positive semi-definite, and with a few sign restrictions on key coefficients. We construct a counterfactual earnings distribution for each of these sets of parameter estimates, and carry out the decompositions proposed in the previous section for each of these distributions. We are thus able to present an interval of estimates of the decompositions, which is consistent with a wide range of values for the possible omitted variable biases in the estimation of equations (7) and (10).

To see how this can be done, let $X = (C, E) = (R, GR, MPE, DPE, FO; S, M, L)$ and $\gamma = (\alpha, \beta)'$, so that we can rewrite (7) as:

$$(7') \quad \ln w_i = X_i \gamma + u_i.$$

Here, u_i need not be orthogonal to the explanatory variables in X .¹⁸ Assume without loss of generality that all the variables have zero mean and define the following covariance matrix:

$$\Sigma = \begin{bmatrix} X'X & X'u \\ u'X & u'u \end{bmatrix}.$$

The bias of the OLS estimates for equation (7'), $B = E(\hat{\gamma}) - \gamma$ is given by $B = (X'X)^{-1}E(X'u) = (X'X)^{-1}(\text{cov}(X, u))$, where $\text{cov}(X, u)$ is a $k \times 1$ vector of covariances between the k -th column of X and the disturbance vector u . Each element of $\text{cov}(X, u)$ is given by $\text{cov}(x_k, u)$. Denoting the correlation coefficient between the k -th column of X and u by $\rho_{x_k u}$ we have $\rho_{x_k u} = \text{cov}(x_k, u) [\sigma_{x_k} \sigma_u]^{-1}$. The bias can thus be written as:

$$(11) \quad B = (X'X)^{-1} (\underline{\rho_{Xu} \sigma_x}) \sigma_u$$

where σ_V is the standard deviation of variable V , and the underlined term is a vector, with elements $(\rho_{x_k u} \sigma_{x_k})$.

¹⁸For ease of notation, from this point onward we drop the individual subscript i in the remainder of this section.

Evaluating the bias vector B thus requires knowing σ_u and ρ_{x_u} , $\forall k$. An unbiased estimate of σ_u would, in turn, require some knowledge of the variance of the coefficient bias, B . Since $\ln w = X\gamma + u = X\hat{\gamma} + \hat{u}$, it follows that $u = \hat{u} + X(\hat{\gamma} - \gamma)$. Therefore:

$$(12) \quad \sigma_u = E(u'u) = E(\hat{u}'\hat{u}) + E\left[(\hat{\gamma} - \gamma)' X'X(\hat{\gamma} - \gamma)\right].$$

The problem is that the last expected value term on the right hand side of (12) cannot be evaluated without knowledge of the variance of the bias or, equivalently, of the second order moment of $X'u$.

A convenient approximation consists of replacing $(\hat{\gamma} - \gamma)$ in (12) by its expected value, B , which clearly underestimates the right hand side of (12). This underestimation is likely to be small if the expected bias, B , is estimated with enough precision. We shall thus use the following proxy for σ_u :

$$(13) \quad \sigma_u^2 \cong \hat{\sigma}_u^2 + B'X'XB.$$

Substituting the value of the bias from (11) into (13) yields:

$$(14) \quad \sigma_u^2 \cong \hat{\sigma}_u^2 + (\rho_{x_u} \sigma_x)' (X'X)^{-1} (\rho_{x_u} \sigma_x) \sigma_u^2; \text{ or}$$

$$(15) \quad \sigma_u^2 \cong \hat{\sigma}_u^2 / (1 - K)$$

with K given by:

$$(16) \quad K = (\rho_{x_u} \sigma_x)' (X'X)^{-1} (\rho_{x_u} \sigma_x).$$

(11) and (14) is a two-equation system in three unknowns: B , σ_u^2 , and ρ_{x_u} . For any vector of correlation coefficients ρ_{x_u} , equations (11)–(14) 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 range 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. First, we sequentially draw random values for $\rho_{x_{ku}}$ for each k , each from a uniform distribution defined on $(-1, 1)$.¹⁹ Since the correlation coefficient vector ρ_{x_u} must satisfy the condition that the covariance matrix Σ be positive semi-definite, all drawings such that this condition is not satisfied are deleted.

A sufficient condition for Σ to be positive semi-definite is that the determinants of all its principal minors be positive. We know that this is true for the $X'X$

¹⁹By assuming a uniform distribution we are deliberately conservative, as we allow for relatively large probabilities of very high correlations between the observed variables and the residual. Had we imposed, instead, a truncated normal or any other distribution with most mass around 0, our confidence intervals would in all likelihood have been narrower.

part of Σ . The last condition needed to establish the positive semi-definiteness of Σ is therefore that its own determinant is positive. The latter is given by $Det[\Sigma] = \det[X'X] \det[u'u - u'X(X'X)^{-1}X'u] > 0$, which uses the formula for the determinant of a partitioned matrix. It can be seen that this condition is equivalent to requiring that $K \leq 1$, where K is given by (16) above. This ensures that the second determinant in the product of determinants is non-negative. We therefore discard any vector ρ_{xu} drawn which leads to $K > 1$.

Four additional assumptions are imposed on the signs of coefficient estimates $\hat{\gamma}$ as in, for instance, Manski and Pepper (2000): the coefficients on the race dummy for Afro-Brazilians (black and mixed race), on the regional dummy for the North/North-East, and on the labor market status dummy for informal employees were constrained to be non-positive. The coefficient on own schooling was constrained to be non-negative. Each of these restrictions is backed up by an extensive body of earnings regressions for Brazil; see various chapters in Henriques (2000). They are also fully consistent with our own (very significant) OLS estimates. Correlation coefficients leading to coefficient estimates that violate those restrictions are also discarded.

The first 100 randomly generated vectors of correlation coefficients that survive are substituted into (11) and (16), to generate a distribution of values for the bias vector, B . From the 100 simulations thus generated for each coefficient, the five highest and five lowest values are discarded. The next highest and lowest values are denoted the upper and lower bounds on the coefficient, generating a 90 percent confidence interval for coefficient values consistent with possible OLS biases.²⁰ A perfectly analogous procedure is employed in the estimation of the reduced-form equation (10).

We use the resulting range of coefficient estimates $\hat{\psi}^*$ to generate a series of counterfactual earnings distributions $\tilde{\Phi}(\tilde{w})$, through equation (10). Similarly, the range of estimates $\hat{\gamma}^* = (\hat{\alpha}^*, \hat{\beta}^*)'$ in the 90 percent confidence interval for the biases in equation (7) is used to generate a series of counterfactual distributions $\Phi^d(w^d)$. Theil indices of inequality are then computed over each of those distributions. Finally, in Section 7, we report for each cohort on the decompositions defined by (3) and (4) for the mean, the highest, and the lowest Theil indices generated by the coefficients in the corresponding 90 percent confidence intervals for bias. The intervals between the shares computed for the highest and lowest counterfactual Theils correspond to the range of plausible values for the share of earnings inequality that is accounted for by unequal opportunities associated with five observed circumstance variables, once possible estimation biases are taken into account.

²⁰These are “90 percent confidence intervals” in the sense that the true unbiased parameter lies in the intervals with a limiting probability of 90 percent, under the maintained model assumptions, including: (i) that $\rho_{xu} \sim U(-1, 1), \forall k$; and (ii) the four previously described restrictions on parameter signs. Note that most reasonable alternative distributional assumptions about ρ_{xu} would lead to narrower confidence intervals.

6. ESTIMATION RESULTS

Before turning to the decomposition results in Section 7, this section briefly reports on the estimation results. In order to implement the decomposition between direct and indirect effects of circumstances (equation 4), the earnings equation (7) was estimated by OLS, separately for each cohort.²¹ 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. Results are presented in Table 2, in the following manner: the OLS estimate (and its significance level) is reported in the center of each cell; above and below it are the upper and lower bounds (respectively) from the simulations, as described in Section 5. These 90 percent confidence intervals for bias are, as one would expect, reasonably wide. Their interpretation is that the true, unbiased coefficients lie in this interval with a 90 percent probability, conditional on the (conservative) assumption that the correlation between the residual and each regressor is uniformly distributed in $(-1, 1)$, and on the sign restrictions discussed in the previous section. In most cases, both the point estimates and the bulk of the intervals are firmly in economically plausible territory.

Circumstance variables have the expected effect on earnings. The coefficient of the race dummy variable is negative and significant for blacks and “pardos.”²² Point estimates (and upper-bound estimates) are higher for the three older cohorts, and decline for the two youngest cohorts. Regional differences are important, too: with the South/South-East as a reference, being born in the North/North-East has a strong and significant negative effect. The effect of the Center/West region is generally also negative, but seldom significant, and the 90 percent confidence interval for bias straddles zero.

The estimated effect of mean parental education on individual earnings is always positive and highly significant. The 90 percent confidence interval for bias lies entirely in positive territory for all but the two oldest cohorts, and even there it is predominantly positive. The coefficients decline somewhat from the older to the younger cohorts. Economically, they imply a sizable effect, with each additional year of schooling of the parents leading to a 4–6 percent increase in earnings for their children (after controlling for own schooling). We interpret the coefficient on parental schooling as capturing those 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 suggested by Behrman and Rosenzweig (2002). They may include the effect of family wealth on the quality of the school attended by the child, 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.

The coefficient on the difference between the education of the mother and the father suggests that no systematic asymmetry between the roles of the education of

²¹A model with the Heckman correction procedure for sample selection was also estimated. The Mills ratio was insignificant and most second-stage coefficients were similar to the OLS estimates, suggesting that selection for prime-age men in urban Brazil is not a significant issue.

²²Race 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. Other races account for just over 1 percent of the sample, and are grouped with whites.

TABLE 2
EARNINGS EQUATIONS BY COHORT. a), b)

	b1936_40	b1941_45	b1946_50	b1951_55	b1956_60	b1961_65	b1966_70
Race							
White & Asian (omitted)							
Black & MR	-0.048 -0.248*** -0.629	-0.072 -0.305*** -0.444	-0.104 -0.253*** -0.490	-0.084 -0.221*** -0.424	-0.043 -0.193*** -0.415	-0.020 -0.155*** -0.322	-0.057 -0.185*** -0.360
Parental schooling							
Mean parental schooling (years)	0.100 0.059*** -0.028 0.069 0	0.076 0.046*** -0.007 0.046 0.011 -0.044	0.065 0.056*** 0.022 0.022 0.004 -0.032	0.074 0.043*** 0.013 0.035 0.009 -0.020	0.057 0.038*** 0.003 0.019 0 -0.021	0.053 0.043*** 0.013 0.027 0 -0.015	0.063 0.040*** 0.020 0.036 0.007* -0.014
Region dummies							
South & South East (omitted)							
North & North East	-0.019 -0.159***	-0.006 -0.089**	-0.017 -0.070**	-0.032 -0.179***	-0.047 -0.133***	-0.097 -0.182***	-0.068 -0.161***
Center-West	-0.239 0.309 -0.152 -0.799	-0.178 0.362 -0.024 -0.435	-0.185 0.226 0.019 -0.381	-0.220 0.043 -0.188*** -0.550	-0.245 0.198 -0.014 -0.361	-0.256 0.071 -0.112** -0.335	-0.229 0.242 -0.002 -0.239
Father's occupational status							
Lower status (omitted)							
Medium status	0.360 0.099*	0.377 0.109**	0.211 0.024	0.251 0.104***	0.274 0.137***	0.189 0.057**	0.280 0.106***
Higher status	0.067 0.107 0.009 -0.184 0.130 0.120*** 0.107 0.161 0.076*	0.054 0.291 0.097 -0.082 0.114*** 0.089 0.214 0.173*** 0.121	0.021 0.165 0.051 0.005 0.124 0.118*** 0.108 0.220 0.175*** 0.151	0.077 0.187 0.076* -0.030 0.117 0.112*** 0.106 0.131 0.096*** 0.062	0.097 0.217 0.097*** -0.081 0.108 0.102*** 0.094 0.155 0.107*** 0.090	0.023 0.214 0.108*** 0.008 0.107 0.101*** 0.093 0.127 0.111*** 0.079	0.092 0.326 0.149*** 0.079 0.092 0.090*** 0.075 0.159 0.133*** 0.108
Years of schooling							
Migrant dummy							
Labor market status							
Formal employee & employer (omitted)							
Informal employee	-0.317 -0.414*** -0.431	-0.348 -0.394*** -0.488	-0.248 -0.357*** -0.368	-0.338 -0.413*** -0.417	-0.257 -0.338*** -0.334	-0.207 -0.266*** -0.280	-0.188 -0.260*** -0.254
Self employed	-0.152 -0.197*** -0.219 0.245***	-0.153 -0.191*** -0.208 0.274***	-0.085 -0.088*** -0.129 0.200***	-0.066 -0.104*** -0.129 0.296***	-0.094 -0.130*** -0.143 0.216***	-0.078 -0.090*** -0.122 0.085***	0.110 0.058** 0.069 0.001
Constant	1.730	2.457	3.726	4.877	5.488	5.643	4.553
Sample size	0.45	0.45	0.48	0.46	0.42	0.43	0.37
Adj R-squared							

a) Dependent variable is the log of hourly wage rate. b) For each variable, we present three values: the minimum and maximum coefficient estimates from our 90% confidence interval of simulations (in italics), and the OLS estimates and significance levels in between; *significant at 10%; **significant at 5%; ***significant at 1%.

the two parents seems to be present. The estimated effect of having a father with a medium- or higher-status occupation on earnings is generally positive, when compared to the reference category of lower status. For medium-status occupations, the entire 90 percent confidence interval is positive for all cohorts, but the same is only true for three cohorts in the case of higher-status occupations. Both the point estimates and the significance of these coefficients tend to increase for the younger cohorts.

Turning to the vector of “effort” variables, own education has the usual positive and significant effect on earnings. This effect is lower for younger 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.²³ The magnitude of the estimates for the returns to schooling in these equations is somewhat lower than some of the previous estimates for Brazil. For instance, Ferreira and Paes de Barros (1999) found that the average returns to a year of schooling lay in the range of 12–15 percent in 1999. In Table 2, average returns for men range from 9 to 12 percent (with surprisingly narrow 90 percent confidence intervals for bias, in the 8–13 percent range). This difference is most likely due to the inclusion of the family background variables, notably mean parental education. This resonates with earlier findings by both Lam and Schoeni (1993) and Strauss and Thomas (1996), who also found lower rates of return on own schooling when parental education is included in the regression.

Migration also has a positive and statistically significant effect on earnings, according to the OLS estimates. As with years of schooling, the 90 percent unbiasedness interval is positive everywhere. The OLS coefficients are rather large, amounting to an 8–18 percent increase in earnings for males, depending on cohort. The labor-market status coefficients confirm one’s expectations: both informal employees and the self-employed earn significantly less than formal employees (with the exception of the youngest cohort of the self-employed). Once again, the 90 percent confidence intervals for unbiasedness do not include zero for any cohort.

After estimating the “full” earnings equation (equation 7), we estimate the reduced form (10), where only circumstance variables are included. These coefficients will capture not only the direct effect of observed circumstances on earnings, controlling for efforts, but also the indirect effect *through* efforts. They will be used to calculate the overall share of (observed) opportunities in earnings inequality, defined in equation (3). The reduced-form estimation results are reported in Table 3. As expected, all coefficients have the same sign as in Table 2, and absolute values are larger. The increases in the absolute values of parental education and father’s occupation coefficients are particularly large, suggesting that the *indirect* effects (βH) of these circumstances on earnings are likely to be important.

²³The conventional Mincerian specification is such that: $\ln w = a.S + b.Exp + c.Exp^2$ where $Exp = Age - S - 6$. Expanding the Exp term leads to: $\ln w = (a - b - 12c).S - 2cAge.S + c.S^2 +$ 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.

TABLE 3
EARNINGS EQUATIONS BY COHORT. REDUCED FORM MODEL. a), b)

	b1936_40	b1941_45	b1946_50	b1951_55	b1956_60	b1961_65	b1966_70
Race							
White & Asian (omitted)							
Black & MR	-0.069 -0.390*** -0.724	-0.214 -0.451*** -0.548	-0.171 -0.392*** -0.612	-0.168 -0.363*** -0.479	-0.189 -0.331*** -0.426	-0.176 -0.305*** -0.381	-0.159 -0.269*** -0.381
Parental schooling	0.225	0.177	0.175	0.156	0.131	0.119	0.105
Mean parental schooling (years)	0.162***	0.137***	0.143***	0.119***	0.103***	0.103***	0.088***
	0.123	0.109	0.112	0.099	0.091	0.090	0.073
	0.120	0.048	0.047	0.045	0.044	0.035	0.022
Mother/father difference (years)	0.014	0.006	0.011	0.007	-0.002	0.002	0.009*
	-0.011	-0.022	-0.017	-0.022	-0.019	-0.010	-0.007
Region dummies							
South & South East (omitted)							
North & North East	-0.112 -0.267*** -0.314	-0.033 -0.190*** -0.208	-0.026 -0.122*** -0.187	-0.079 -0.279*** -0.249	-0.090 -0.236*** -0.258	-0.147 -0.242*** -0.254	-0.155 -0.227*** -0.259
Center-West	0.194 -0.171 -0.515	0.264 -0.006 -0.251	0.134 -0.021 -0.262	0.071 -0.139** -0.217	0.138 -0.045 -0.130	0.059 -0.086* -0.132	0.100 -0.021 -0.086
Father's occupational status							
Lower status (omitted)							
Medium status	0.475 0.381*** 0.246	0.494 0.370*** 0.351	0.419 0.340*** 0.296	0.440 0.379*** 0.333	0.411 0.372*** 0.321	0.266 0.243*** 0.206	0.260 0.228*** 0.203
Higher status	0.385 0.307*** 0.114	0.559 0.350*** 0.287	0.479 0.352*** 0.262	0.387 0.346*** 0.235	0.388 0.320*** 0.235	0.315 0.279*** 0.220	0.326 0.282*** 0.217
Constant	0.530***	0.699***	0.724***	0.829***	0.746***	0.656***	0.548***
Sample size	1,730	2,457	3,726	4,877	5,488	5,643	4,553
Adj R-squared	0.29	0.29	0.30	0.28	0.28	0.29	0.24

a) Dependent variable is the log of hourly wage rate. b) For each variable, we present three values: the minimum and maximum coefficient estimates from our 90th confidence interval of simulations (in italics), and the OLS estimates and significance levels in between; *significant at 10%; **significant at 5%; ***significant at 1%.

Finally, Table 4 reports the OLS coefficients for equation (8), estimated as a linear schooling regression. We do not use the coefficient estimates from regressions of “efforts” on circumstance variables in our decompositions. Both the migration and the labor status equations would have been better estimated as discrete choice models, and our estimation procedure based on confidence intervals for unbiased coefficients has not yet been extended to such non-linear models. In any case, as discussed in Section 5, using the estimates from (7) and (10) allows us to obtain exactly the same decompositions, with much greater statistical confidence.²⁴

The schooling regression is briefly mentioned here, however, because it is of intrinsic interest to the analysis. The key coefficient of interest here is that of parental education, α_p , which can be interpreted as an inverse measure of inter-generational educational mobility (conditional on other circumstances). 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 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 4 reveals rather high, but declining coefficients: educational persistence falls from 0.83 for the oldest generation to 0.52 for the youngest.²⁵ Rising average levels of schooling across generations are picked up by the large constant terms. The fact that these intercept estimates rise across cohorts indicate an acceleration across age groups in schooling differentials between generations.

Father’s occupation, race and region of birth are also significant determinants of education, even after controlling for parental schooling, and for one another. Their coefficients do not display as clear a declining trend across cohorts as parental education, but are also noticeably lower for the two youngest age groups for father’s occupation. Overall, the clearest change in the conditional distribution of educational opportunities across cohorts seems to have been a decline in the importance of parental education.²⁶

²⁴Nevertheless we have also computed Θ_j from estimates of the “structural” model (7)–(8), using the linear schooling regression reported here; a probit model for the migration decision; and a multinomial logit for the labor market status variable. We first simulated the changes in effort levels obtained by equalizing circumstances in these equations. We then simulated earnings distributions with circumstances held constant and using these counterfactual effort levels, to estimate the complete effect of opportunities. Since the estimates of the effort equations (8) did not correct for potential econometric endogeneity of the circumstances, we do not report those results here. Nevertheless, the estimates of Θ_1 obtained from that exercise were not dissimilar to those reported in Table 5. Those results are available from the authors on request.

²⁵While this pattern cannot be used to infer a decline in educational persistence over time, it does shed some light on differences in persistence across cohorts at a single point in time.

²⁶An important caveat in interpreting the results in Table 4 is that, in addition to possible omitted variable bias, there may be the usual measurement error problems associated with measuring education by years of schooling. In particular, the quality of education is not captured. It cannot be ruled out that taking changes in the quality of education into account might modify the perception that educational mobility is higher for younger cohorts in Brazil. See Behrman and Birdsall (1983) for an earlier attempt to control for quality of schooling in Brazil.

TABLE 4
LINEAR SCHOOLING REGRESSIONS a), b)

	b1936_40	b1941_45	b1946_50	b1951_55	b1956_60	b1961_65	b1966_70
Race							
White & Asian (omitted)							
Black & MR	-0.940*** [0.188]	-1.151*** [0.161]	-1.084*** [0.135]	-1.104*** [0.116]	-1.180*** [0.106]	-1.291*** [0.099]	-0.686*** [0.105]
Parental schooling							
Mean parental schooling (years)	0.828*** [0.040]	0.768*** [0.030]	0.720*** [0.025]	0.657*** [0.021]	0.612*** [0.017]	0.577*** [0.016]	0.518*** [0.018]
Mother/father difference (years)	0.097*** [0.043]	-0.02 [0.036]	0.059*** [0.029]	0.002 [0.024]	-0.009 [0.020]	0.015 [0.018]	0.003 [0.019]
Region dummies							
South & South East (omitted)							
North & North East	-0.808*** [0.190]	-0.917*** [0.160]	-0.458*** [0.136]	-0.880*** [0.120]	-0.934*** [0.111]	-0.659*** [0.104]	-0.830*** [0.109]
Center-West	0.103 [0.560]	0.343 [0.417]	-0.217 [0.334]	0.406 [0.281]	-0.106 [0.231]	0.273 [0.206]	-0.123 [0.203]
Father's occupational status							
Lower status (omitted)							
Medium status	2.239*** [0.237]	2.294*** [0.199]	2.743*** [0.159]	2.338*** [0.131]	2.267*** [0.114]	1.856*** [0.106]	1.420*** [0.115]
Higher status	2.349*** [0.308]	2.240*** [0.262]	2.569*** [0.212]	2.328*** [0.178]	2.104*** [0.157]	1.666*** [0.150]	1.425*** [0.165]
Constant	3.100*** [0.139]	3.812*** [0.124]	4.068*** [0.106]	4.913*** [0.091]	5.292*** [0.085]	5.593*** [0.083]	5.505*** [0.090]
Sample size	1,730	2,457	3,726	4,877	5,488	5,643	4,553
Adj R-squared	0.41	0.43	0.43	0.4	0.42	0.41	0.34

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

7. DECOMPOSITION RESULTS

Using the coefficient estimates from the reduced-form equation (10), reported in Table 3, we simulate the counterfactual distributions $\tilde{\Phi}(\tilde{w})$, corresponding to $\tilde{w}_i = \exp[\tilde{C}\tilde{\psi} + \hat{\epsilon}_i]$. This allows us to decompose earnings inequality for each cohort in our sample, into a component due to unequal opportunities (arising from five observed circumstance variables), and a residual component due to unobserved circumstances, “efforts,” and random elements such as transitory earnings or measurement error. The procedure described in Section 5 was used to generate a distribution of coefficient estimates, from which a 90 percent confidence interval for the unbiased coefficients was constructed. Using the range of coefficient estimates in that 90 percent confidence interval, we computed the corresponding counterfactual earnings distributions, over which inequality indices were calculated. The mean value, as well as the upper and lower bounds of these inequality indices are presented in Table 5.

Table 5 presents Theil coefficients for factual and counterfactual earnings distributions for our seven cohorts in 1996. The first row contains the observed (factual) earnings inequality. The next panel, with six rows, presents inequality in the counterfactual distribution $\tilde{\Phi}(\tilde{w})$, where the inequality of opportunities due to observed circumstances has been eliminated: $I(\tilde{\Phi})$. The mean value of all Theil indices as well as those corresponding to the extreme bounds are presented in the first three rows. The next three rows present the shares in earnings inequality accounted for by the difference between observed inequality and the mean (and extreme bounds) estimate of residual inequality, i.e.: $\Theta_l := \frac{I(\Phi) - I(\tilde{\Phi})}{I(\Phi)}$. These are our measures of inequality of opportunities in this decomposition.

For men born between 1941 and 1945, for instance, elimination of inequality due to observed circumstances reduces the Theil index from 0.997 to some value between 0.632 and 0.675, with a mean estimate of 0.656. These estimates indicate that 32–37 percent of earnings inequality in this cohort is accounted for by unequal opportunities—due only to those five observed circumstance variables. The share of inequality due to unequal opportunities varies considerably across cohorts, between 13 and 34 percent when considering mean estimates, and between 10 and 37 percent when accounting for possible biases in the estimation of the model coefficients. The intervals are narrower if the youngest cohort is excluded: 18–34 percent for the mean estimates, and 15–37 percent in the confidence intervals. The simple average of mean estimates across cohorts is 23 percent.

As indicated by the subscript l in Φ_l , these shares depend on the specific inequality measure used. An analogous decomposition was conducted using the Gini coefficient, instead of the Theil index, and the cohort average for the Gini was 13 percent. Detailed results for the Gini decomposition, as well as for an analogous decomposition for women, are available in Bourguignon *et al.* (2003). An additional set of calculations concerns extending the analysis to rural areas. As discussed in Section 3, our sample is restricted to urban areas, since there are a number of reasons for caution with rural earnings data from the PNAD. As a robustness check, however, we also calculated all these shares for a joint urban and rural sample. Excluding the youngest cohort, the overall opportunity shares of

TABLE 5
THE CONTRIBUTION OF UNEQUAL OPPORTUNITIES TO EARNINGS INEQUALITY, URBAN MEN IN BRAZIL: ACTUAL AND COUNTERFACTUAL THEIL COEFFICIENTS (AND RATIOS) FOR 5-YEAR COHORTS

	b1936_40	b1941_45	b1946_50	b1951_55	b1956_60	b1961_65	b1966_70
Total Observed Inequality (1)	0.873	0.997	0.759	0.655	0.706	0.580	0.566
PANEL 1: "Overall" (observed) opportunity share of earnings inequality							
(Upper bound estimate)	(0.692)	(0.675)	(0.644)	(0.531)	(0.572)	(0.415)	(0.507)
Mean estimate (2a)	0.654	0.656	0.619	0.519	0.562	0.407	0.494
(Lower bound estimate)	(0.642)	(0.632)	(0.609)	(0.516)	(0.555)	(0.402)	(0.485)
(Upper bound estimate)	(0.264)	(0.366)	(0.199)	(0.212)	(0.214)	(0.307)	(0.143)
Mean share ((1)-(2a))/(1)	0.251	0.343	0.184	0.208	0.205	0.298	0.128
(Lower bound estimate)	(0.207)	(0.323)	(0.151)	(0.190)	(0.190)	(0.284)	(0.104)
PANEL 2: "Partial" or "direct" (observed) opportunity share of earnings inequality							
(Upper bound estimate)	(0.834)	(0.924)	(0.710)	(0.600)	(0.661)	(0.515)	(0.533)
Mean estimate (2b)	0.749	0.821	0.659	0.561	0.620	0.474	0.508
(Lower bound estimate)	(0.719)	(0.776)	(0.634)	(0.546)	(0.601)	(0.448)	(0.491)
(Upper bound estimate)	(0.176)	(0.222)	(0.165)	(0.167)	(0.149)	(0.228)	(0.133)
Mean share ((1)-(2b))/(1)	0.142	0.177	0.132	0.144	0.122	0.184	0.102
(Lower bound estimate)	(0.044)	(0.073)	(0.065)	(0.085)	(0.065)	(0.113)	(0.058)
PANEL 3: Treating observed efforts as circumstance variables							
(Upper bound estimate)	(0.556)	(0.566)	(0.467)	(0.451)	(0.492)	(0.370)	(0.435)
Mean estimate (2c)	0.533	0.541	0.450	0.438	0.477	0.363	0.421
(Lower bound estimate)	(0.518)	(0.524)	(0.442)	(0.427)	(0.471)	(0.356)	(0.409)
(Upper bound estimate)	(0.406)	(0.475)	(0.418)	(0.348)	(0.333)	(0.387)	(0.277)
Mean share ((1)-(2c))/(1)	0.389	0.457	0.407	0.332	0.324	0.374	0.256
(Lower bound estimate)	(0.363)	(0.432)	(0.385)	(0.312)	(0.303)	(0.362)	(0.231)

inequality are slightly higher in the joint sample, by a margin of 2–19 percent (*not* percentage points), depending on the cohort.²⁷

The next panel in Table 5 presents the sub-decomposition of overall 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 $\Phi^d(w^d)$, obtained from $w_i^d = \exp[\bar{C}\hat{\alpha} + E_i\hat{\beta} + \hat{u}_i]$. 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 different sector of the labor market. These estimates are of potential interest for policy-makers, in that they separate the impact of family background, race and geography that is mediated by wage determination, from the effects operating through schooling, migration and labor market segmentation.

The fourth, fifth and sixth rows in this panel present our (respectively upper bound, mean and lower bound) summary measures of this direct effect of unequal opportunities: $\Theta_i^d := [I(\Phi) - I(\Phi^d)](I(\Phi))^{-1}$. This sub-component of inequality of opportunities averaged 14 percent across the seven cohorts, suggesting that, on average, 62 percent of the impact of opportunities on earnings takes the form of a direct effect, while the remaining 40 percent or so corresponds 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. This fraction (62 percent) is the (cross-cohort) average ratio of the direct effect to the overall effect of observed circumstances: Θ_i^d/Θ_i . Since both the numerator and the denominator are subject to both sampling error and the bias confidence interval, however, there is substantial statistical variation around it. One way to capture that uncertainty is to compute, for each cohort, a lower bound estimate of Θ_i^d/Θ_i by dividing the lower bound of the direct share by the upper bound of the overall share; and an upper bound estimate by dividing the upper bound of the direct share by the lower bound of the overall share. Excluding the youngest cohort, these intervals range from a lower bound of 16 percent and an upper bound of 76 percent for the oldest cohort; to a lower bound of 36 percent and an upper bound of 80 percent for the cohort born between 1961 and 1965. Taking averages across these six cohorts yields a mean ratio of 61 percent (suggesting that the omission of the youngest cohort matters little to the average), a lower bound of 29 percent and an upper bound of 82 percent.

It is possible, of course, that the total effect of all circumstances on our observed “effort” variables (schooling, migration and labor market status) may be greater than the term $\Theta_i^i = \Theta_i - \Theta_i^d$. If, for instance, *unobserved* circumstances

²⁷Opportunity shares are much higher in the joint urban and rural sample for the youngest cohort, born between 1966 and 1970. Since these workers are more recent entrants into the labor force, and a greater number report working part-time in urban areas, there may be greater spurious variance in their wages. This variance is part of the residual of the decomposition, and may help explain why the opportunity shares are so much lower in this particular cohort, when compared both to other urban cohorts and to the same cohort in rural areas.

account for a large share of the variance in the random terms v_i in equation (6), we would be underestimating the share of opportunities in earnings inequality. To shed some light on the magnitudes associated with this possibility, the bottom panel in Table 5 reports 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 (5)—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 $\ln(\tilde{w}_i) = \bar{C}\hat{\alpha} + \bar{E}\hat{\beta} + \hat{u}_i$, using the mean estimates and bounds reported in Table 2. Unsurprisingly, such a view of the world leads to much higher opportunity shares of inequality, as indicated in the bottom rows of Table 5. These opportunity shares average 36 percent, and the confidence intervals for the bias range (across cohorts) from a lower bound of 23 percent to an upper bound of 48 percent.

The opportunity shares in this particular thought experiment correspond to an upper bound on the share of inequality due to the five *observed* circumstance variables (because it assumes that all residual variance in the effort equations is also due to circumstances). It does not, however, correspond to an upper bound on the effects of *all* circumstances, since there are still likely to be unobserved circumstance variables in the residual term, u , of equation (5). The confidence intervals for the bias address any impact those unobservables may have on our estimates of the effect of the observable variables, but not any independent effect they might themselves have on earnings.

Figure 1 depicts the decomposition graphically: the uppermost line represents the inequality actually observed for the various cohorts: $I(\Phi)$. The solid line below it (with squares) shows the “partial” or direct effect of equalizing circumstances ($I(\Phi^d)$), whereas the bottom solid line (with triangles) shows the “overall” effect: $I(\tilde{\Phi})$. The dotted lines around these two counterfactual lines show the upper and lower bounds for the corresponding 90 percent confidence intervals. Inequality of opportunity, by each measure, corresponds to the difference between observed

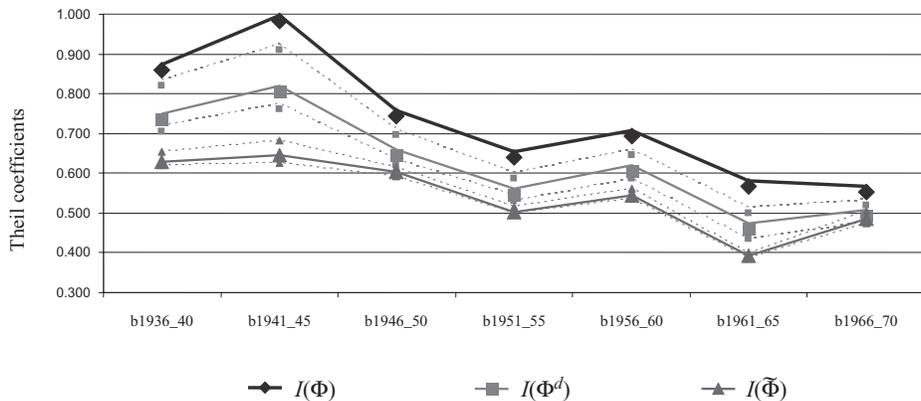


Figure 1. Effects of Equalizing Circumstances on Inequality (partial and complete effects)

inequality (at the top) and the counterfactual inequality indices: $I(\Phi) - I(\Phi^d)$ for the direct effect; and $I(\Phi) - I(\tilde{\Phi})$ for the overall effect of observed circumstances. The height of the $I(\tilde{\Phi})$ corresponds to the residual component, for each cohort.

Visual inspection of this figure reveals that male earnings inequality falls considerably from the older to the younger cohorts.²⁸ It is important to recognize that temporal patterns cannot be identified from these cohort trends. In particular, as already mentioned, returns to education in Brazil increase with age, leading to greater dispersion in earnings within older cohorts, at every time period.²⁹ It is perhaps more interesting that the share (as well as the absolute contribution) of inequality of (observed) opportunities in earnings inequality also seems to decline from the oldest to the youngest cohort. That share averages 30 percent for men born between 1936 and 1945, as opposed to 21 percent 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, more markedly, in the own schooling equations, as shown in Tables 2 and 4.³⁰ To the extent that these *shares* are unrelated to age-group-specific *levels* of earnings inequality, their decline may reflect a lower degree of inequality of opportunity for the younger cohorts in Brazil. An important caveat, however, is that this inference of a trend relies heavily on the opportunity share estimated for the youngest cohort. If, as previously mentioned, the greater incidence of part-time work in this age group, or any other reason, make it likely that this is an underestimate, then the evidence of a downward trend across cohorts becomes much more precarious.

Table 6 and Figure 2 assess the roles of *individual* circumstance variables in the preceding results. The complete effect of equalizing each individual circumstance variable in turn, while controlling for all others, is shown for the Theil coefficient separately for each cohort.³¹ It can be seen that, of all circumstance variables, parental education plays the largest role in determining inequality, 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 bound (of six school years) is imposed, as if schooling were compulsory (*de facto*, rather than merely *de jure*) 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 circumstances that shape opportunity sets for the young, father's occupation is the second most important circumstance, although the impact of race is not much smaller, particularly for younger cohorts. Father's occupation seems to have been a more important determinant of opportunity for the two oldest cohorts—where it accounted for 10 percent or more of earnings inequality—with a more muted effect for

²⁸Interestingly, this pattern is not observed for women. For women, inequality is lower for both the oldest and the youngest cohorts, and higher for those in between (see Bourguignon *et al.*, 2003).

²⁹This is particularly true for men. Evidence of this age-dependence of earnings inequality in other countries is analyzed in Deaton and Paxson (1994).

³⁰This result is consistent with the decline in intergenerational persistence of education across cohorts, described by Ferreira and Veloso (2006).

³¹This corresponds to simulating distributions such as $\tilde{w}_i^j = \exp[\bar{C}_{ij}\hat{\psi}_j + C_{i,-j}\hat{\psi}_{-j} + \hat{\epsilon}_i]$, where $\tilde{\Phi}(\tilde{w}^j)$ is the counterfactual wage distribution corresponding to holding variable C_j constant, while all other variables in the reduced-form equation (10) take their observed values.

TABLE 6
CONTRIBUTION OF INDIVIDUAL CIRCUMSTANCE VARIABLES TO EARNINGS INEQUALITY, BY COHORT

	b1936_40	b1941_45	b1946_50	b1951_55	b1956_60	b1961_65	b1966_70
Total Observed Inequality	0.873	0.997	0.759	0.655	0.706	0.580	0.566
Equalizing race	0.830	0.936	0.727	0.626	0.689	0.552	0.567
Equalizing region	0.860	0.983	0.751	0.645	0.689	0.571	0.560
Equalizing parental education	0.726	0.759	0.634	0.553	0.595	0.440	0.491
Lower-bounding parental education	0.730	0.787	0.643	0.565	0.599	0.486	0.503
Equalizing parental occupation	0.793	0.883	0.719	0.611	0.666	0.532	0.540

Note: Mean estimates from the 90 counterfactual distributions corresponding to the “90% confidence interval” of unbiasedness of the coefficients.

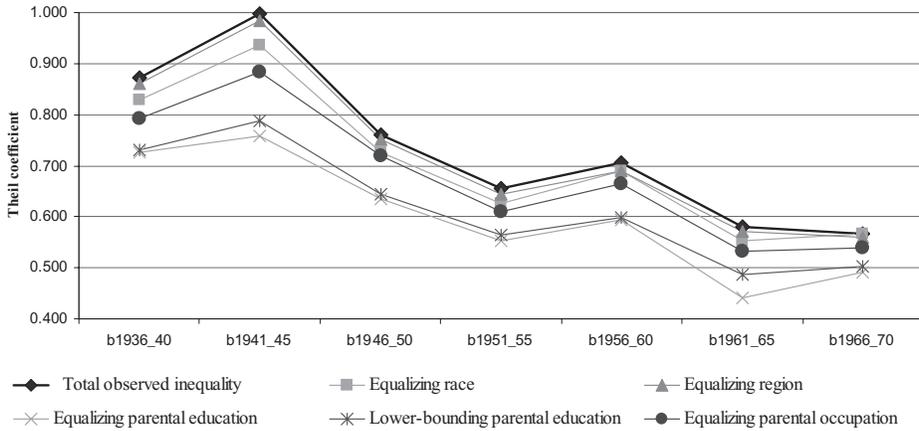


Figure 2. Overall Effect of Equalizing Individual Circumstance Variables on Inequality (Theil coefficient for 5 year cohort)

cohorts born after the Second World War. The role of spatial factors (measured by region of birth) in accounting for inequality of opportunity in Brazil is much smaller, once one controls for the racial and family background composition across regions. Controlling for the other observable circumstance variables, eliminating the impact of region of birth reduces earnings inequality by only 1–2 percent for all cohorts.

As before, the variation in these shares across cohorts cannot be interpreted as evidence of changes over time, since they are measured at the same point in time, and it is impossible to disentangle period, age and cohort effects. Nevertheless, the results are intriguing in the light of other evidence documenting, for instance, the decline in inequality between Brazil’s main regions, and between urban and rural areas, between the early 1980s and the late 1990s.³² The results can also be taken as suggesting that the most promising policies for reducing inequality of opportunities in Brazil might be those aimed at reducing the effect of parental education on the child’s schooling and earnings. Even enforcing a lower bound in years of schooling, it appears, may have a sizable impact in terms of reducing inequality in the succeeding generation. It remains to be seen whether the recent increase in education levels in Brazil’s labor force (see Table 1) will indeed have this effect on the distribution of their children’s earnings.

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 among men in urban Brazil. We estimated both the direct impact of observed opportunities (or circumstances) on earnings, and the indirect effect through three observed “effort” variables—namely own schooling,

³²See, e.g. Ferreira *et al.* (forthcoming).

the decision to migrate, and one's broad status in the labor market. We took account of the biases arising from the econometric endogeneity of each of these variables, by estimating 90 percent confidence intervals for unbiased coefficients, given the OLS estimates and a conservative assumption about the distribution of the possible correlation between the observed variables and the regression residuals. We used the ensuing distribution of admissible coefficients to generate a set of counterfactual earnings distributions, from which we constructed lower and upper bounds for the share of inequality arising from these observed circumstances.

We find that this group of five observed circumstance variables (namely father's and mother's schooling, father's occupation, race and region of birth) accounts for between 10 and 37 percent of the total earnings inequality within cohorts in Brazil in 1996, as measured by the Theil index. The simple mean estimate across cohorts is 23 percent. The impact of opportunities, defined in this way, was further decomposed into a direct effect on earnings (which accounts for some 60 percent of the total), and an indirect effect through the "effort" decisions individuals make.

Although the bounds associated with the estimates of both the overall and the direct effects of circumstances imply rather broad confidence intervals around their ratio (29–82 percent), it is nevertheless clear that the effect of family background on opportunities is not restricted to a child's schooling (or location, or broad access to formal-sector jobs), but that an additional impact occurs through the labor market, as wages are determined conditional on those observed "efforts." Indeed, at the mean across cohorts, the direct effect actually dominates.

Regardless of the channel, our analysis suggests that family background is the most important set of circumstances determining a person's opportunities. Sixty-five to 70 percent of the total effect of observed circumstances can be attributed to parental schooling alone, and this figure rises to almost 80 percent when the father's occupation is added. There is also some (weak) 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.

Given the econometric difficulties inherent in estimating the share of earnings variation associated with a set of observed variables when other, unobserved, determinants are known to be correlated with them, the unbiased confidence interval of 10–37 percent for the share of inequality accounted for by unequal opportunities (or 15–37 percent when the youngest cohort is excluded) is not too wide. This is particularly the case once one considers that a large share of this interval is due to the natural inter-cohort variation. Neither are these estimates insubstantial: they indicate that these five circumstance variables alone account for somewhere between a tenth and over a third of measured earnings inequality in Brazil, with a central estimate of just under a quarter. If the variance of the residual term u_i , which accounts for a large share of the residual component of the decomposition, also includes measurement error and transitory income components, then the true opportunity share of inequality may be even higher.³³

³³This is rather plausible. Atkinson *et al.* (1992) report that the share of transitory components in the variance of the logarithm of current earnings is around 30 percent in a number of developed countries. See also Lillard and Willis (1978) for the original U.S. study.

APPENDIX

TABLE A1
NON-RESPONSE RATES FOR PARENTAL BACKGROUND VARIABLES BY COHORT, PNAD 1996

	Urban Sample		Urban and Rural Sample	
	Obs. Missing Parental Education	Obs. Missing Parental Occupation	Obs. Missing Parental Education	Obs. Missing Parental Occupation
Birth cohorts 1936–40	14.5%	14.5%	13.7%	12.9%
Birth cohorts 1941–45	13.3%	15.6%	12.8%	13.8%
Birth cohorts 1946–50	12.6%	14.6%	11.9%	13.2%
Birth cohorts 1951–55	11.9%	14.6%	11.8%	13.4%
Birth cohorts 1956–60	13.4%	16.5%	13.0%	15.0%
Birth cohorts 1961–65	14.5%	16.3%	13.9%	15.3%
Birth cohorts 1966–70	17.7%	20.2%	16.9%	18.9%

TABLE A2
INVESTIGATING THE EFFECTS OF SELECTIVE NON-RESPONSE (MEANS)

	Sample Used in Analysis	Full Sample (urban)*	Full Sample (urban and rural)*
Mean age	40.05	39.94	40.15
Mean years of schooling	6.80	6.71	6.01
Race (percent)			
White & Asian	59.44	57.20	55.38
Black & MR	40.56	42.80	44.62
Region of origin (percent)			
South & South East	36.10	38.01	39.52
North & North East	57.91	55.92	54.27
Center-West	5.99	6.07	6.21
Migrants (percent)	64.07	63.46	60.19
Labor market status (percent)			
Formal employees & employers	57.70	57.46	52.03
Informal employees	13.19	13.71	15.75
Self employed	29.11	28.82	32.22

Note: *Full sample denotes the sample of active males, aged 26–60, with positive earnings.

TABLE A3
 INVESTIGATING THE EFFECTS OF SELECTIVE NON-RESPONSE (EARNINGS REGRESSION ON COMMON SUBSET
 OF VARIABLES)

	Sample Used in Analysis	Full Sample (urban)*	Full Sample (urban and rural)*
Race			
White & Asian (omitted)			
Black & MR	-0.237*** [0.011]	-0.245*** [0.009]	-0.219*** [0.008]
Region dummies			
South & South East (omitted)			
North & North East	-0.149*** [0.011]	-0.159*** [0.009]	-0.152*** [0.009]
Center-West	-0.097*** [0.024]	-0.085*** [0.020]	-0.070*** [0.019]
Years of schooling	0.125*** [0.001]	0.122*** [0.001]	0.131*** [0.001]
Migrant dummy	0.120*** [0.010]	0.122*** [0.008]	0.151*** [0.008]
Labor market status			
Formal employee & employer (omitted)			
Informal employee	-0.338*** [0.015]	-0.330*** [0.013]	-0.371*** [0.011]
Self employed	-0.072*** [0.011]	-0.053*** [0.009]	-0.137*** [0.009]
Constant	0.199*** [0.014]	0.197*** [0.012]	0.073*** [0.011]
Sample size	28,474	37,548	46,014
Adj R-squared	0.41	0.40	0.44

Notes: *Full sample denotes the sample of active males, aged 26–60, with positive earnings.
 ***Significant at the 1% level.

REFERENCES

- Arneson, Richard, "Equality of Opportunity for Welfare," *Philosophical Studies*, 56, 77–93, 1989.
- Atkinson, Anthony B., François Bourguignon, and Christian Morrisson, *Empirical Studies of Earnings Mobility*, Harwood Academic Publishers, Philadelphia, 1992.
- Behrman, Jere and Nancy Birdsall, "The Quality of Schooling: Quantity Alone is Misleading," *American Economic Review*, 73(5), 928–46, 1983.
- Behrman, Jere and Mark R. Rosenzweig, "Does Increasing Women's Schooling Raise the Schooling of the Next Generation?" *American Economic Review*, 92(1), 323–34, 2002.
- Behrman, Jere and Paul Taubman, "Intergenerational Transmission of Income and Wealth," *American Economic Review*, 66(2), 436–40, 1976.
- Behrman, Jere, Nancy Birdsall, and Miguel Székely, "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*, Brookings Institution, Washington DC, 135–67, 2000.
- Betts, Julian and John Roemer, "Equalizing Opportunity through Educational Finance Reform," Mimeo, Public Policy Institute of California, San Francisco, CA, 1999.
- Bourguignon, François, Francisco H. G. Ferreira, and Marta Menéndez, "Inequality of Outcomes and Inequality of Opportunities in Brazil," World Bank Policy Research Working Paper #3174, Washington DC, December 2003.
- Bowles, Samuel, "Schooling and Inequality from Generation to Generation," *Journal of Political Economy*, 80(3), S219–51, 1972.
- Bowles Samuel and Herbert Gintis, "The Inheritance of Inequality," *Journal of Economic Perspectives*, 16(3), 3–30, 2002.

- Cecchi, Daniele and Vitorocco Peragine, "Regional Disparities and Inequality of Opportunity: The Case of Italy," IZA Discussion Paper #1874, Bonn, Germany, December 2005.
- Cogneau, Denis and Jérémie Gignoux, "Earnings Inequalities and Educational Mobility in Brazil Over Two Decades," Unpublished Manuscript, DIAL and INED, Paris, 2005.
- Cohen, G. A., "On the Currency of Egalitarian Justice," *Ethics*, 99, 906–44, 1989.
- Deaton, Angus and Christina Paxson, "Intertemporal Choice and Inequality," *Journal of Political Economy*, 102(3), 437–67, 1994.
- Dunn, Christopher, "Intergenerational Earnings Mobility in Brazil and Its Determinants," University of Michigan, 2003.
- Dworkin, Ronald, "What is Equality? Part 1: Equality of Welfare; Part 2: Equality of Resources," *Philos. Public Affairs*, 10, 185–246, 283–345, 1981.
- Ferreira, Francisco H. G. and Ricardo Paes de Barros, "The Slippery Slope: Explaining the Increase in Extreme Poverty in Urban Brazil, 1976–1996," *Brazilian Review of Econometrics*, 19(2), 211–96, 1999.
- Ferreira, Sérgio G. and Fernando A. Veloso, "Mobilidade Intergeracional de Educação no Brasil," *Pesquisa e Planejamento Econômico*, 33, 481–513, 2003.
- , "Intergenerational Mobility of Wages in Brazil," *Brazilian Review of Econometrics*, 26(2), 181–211, 2006.
- Ferreira, Francisco H. G., Peter Lanjouw, and Marcelo Neri, "A Robust Poverty Profile for Brazil Using Multiple Data Sources," *Revista Brasileira de Economia*, 57(1), 59–92, 2003.
- Ferreira, Francisco H. G., Phillippe G. Leite, and Julie A. Litchfield, "The Rise and Fall of Brazilian Inequality: 1981–2004," *Macroeconomic Dynamics*, forthcoming.
- Hanushek, Eric, "The Economics of Schooling: Production and Efficiency in Public Schools," *Journal of Economic Literature*, 24, 1141–77, 1986.
- Henriques, Ricardo, *Desigualdade e Pobreza no Brasil*, IPEA, Rio de Janeiro, 2000.
- Lam, David, "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, 1999.
- Lam, David and R. F. Schoeni, "Effects of Family Background on Earnings and Returns to Schooling: Evidence from Brazil," *Journal of Political Economy*, 101(4), 710–40, 1993.
- Lefranc, Arnaud, Nicolas Pistoiesi, and Alain Trannoy, "Inequality of Opportunities vs. Inequality of Outcomes: Are Western Societies All Alike?" ECINEQ Discussion Paper # 2006-54, August 2006.
- Lillard, Lee A. and Robert J. Willis, "Dynamic Aspects of Earning Mobility," *Econometrica*, 46(5), 985–1012, 1978.
- Manski, Charles and John Pepper, "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68(4), 997–1010, 2000.
- Mazumder, Bhashkar, "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 University Press, Princeton, 2005.
- Mulligan, Casey B., "Galton Versus the Human Capital Approach to Inheritance," *The Journal of Political Economy*, 107(6), 184–224, 1999.
- Paes de Barros, Ricardo and David Lam, "Desigualdade de renda, desigualdade em educação e escolaridade das crianças no Brasil," *Pesquisa e Planejamento Econômico*, 23(2), 191–218, 1993.
- Page, Marianne and John Roemer, "The US Fiscal System as an Equal Opportunity Device," in Kevin Hassett and R. Glenn Hubbard (eds), *The Role of Inequality in Tax Policy*, The American Enterprise Institute Press, Washington DC, 2001.
- Peragine, Vito, "Ranking Income Distributions According to Equality of Opportunity," *Journal of Economic Inequality*, 2, 11–30, 2004.
- Pero, Valéria, "Et, à Rio, plus ça reste le même . . . Tendências da mobilidade social intergeracional no Rio de Janeiro," ANPEC, Salvador, 2001.
- Piketty, Thomas, "Social Mobility and Redistributive Politics," *Quarterly Journal of Economics*, CX(3), 551–84, 1995.
- Roemer, John E., *Equality of Opportunity*, Harvard University Press, Cambridge, MA, 1998.
- 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, "To What Extent do Fiscal Regimes Equalize Opportunities for Income Acquisition Among Citizens?" *Journal of Public Economics*, 87, 539–65, 2003.
- Solon, Gary, "Intergenerational Income Mobility in the United States," *American Economic Review*, 82(3), 393–408, 1992.
- , "Intergenerational Mobility in the Labor Market," in O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*, Vol. 3A, Amsterdam, North-Holland, 1761–800, 1999.

- Strauss, John and Duncan Thomas, "Wages, Schooling and Background: Investing in Men and Women in Urban Brazil," Chapter 5 in N. Birdsall and R. Sabot (eds), *Opportunity Forgone: Education in Brazil*, Inter-American Development Bank, Washington, DC, 1996.
- Valle Silva, Nelson, *Posição social das ocupações*, IBGE, Rio de Janeiro, 1978.
- van de Gaer, Dirk, Erik Schokkaert, and Michel Martinez, "Three Meanings of Intergenerational Mobility," *Economica*, 68, 519–37, 2001.