2 Earnings Inequality and Educational Mobility in Brazil over Two Decades

Denis Cogneau and Jérémie Gignoux

2.1 Introduction

Brazilian society, from a number of points of view, is one of the most inegalitarian in the world, and special attention has consequently been paid to it for a long time (Fishlow 1972). The South American and Caribbean societies are particularly inegalitarian. This characteristic has now been related to the institutions left over from the colonial period (Sokoloff and Engerman 2000). The level of inequality in Brazil is much greater even than the average on the subcontinent with, for example, a Gini index one-third higher than Argentina (UN/WIDER data), and at the same level as in South Africa (Lam 1999). The colonial legacy probably weighs heavy from this point of view, since Brazil was the region's main slave country. Correlatively, the Brazilian economy and society display an extremely high degree of dualism, visible both in the education system (private/state) and on the labor market (official/unofficial). Brazil is also among the countries with the lowest intergenerational educational mobility and equality of social and economic opportunities in the world (Dunn 2007).

A series of nationally representative annual surveys based on large samples (Pesquisa Nacional por Amostra de Domicilios, or PNAD, 1976–1996) provides a fairly accurate observation of the change in inequality in Brazil over nearly thirty years. These data show that income inequality remained remarkably stable, whether the gaps were in individual earnings or household standards of living. The 1976, 1982, 1988, and 1996 PNAD surveys also include certain information on individuals' social origin. Sociologists have used these data to produce the first quantitative analyses of intergenerational social mobility in Brazil (Pastore 1982, Pastore and Valle Silva 2000, Picanço 2004). Economists have also recently looked into the impact of family origins and inequality of educational and labor market opportunities (Lam and Schoeni 1993; Arias, Yamada, and Tejerina 2002; Andrade et al. 2003; Ferreira and Veloso 2006; Bourguignon, Ferreira, and Menendez 2007; Dunn 2007). All these studies emphasize that inequality in Brazil comes with a high degree of intergenerational transmission of education, occupational status, or income. One of the main questions put in these papers concerns the contribution of education to the reduction of economic inequality. Most argue that education is one prominent channel through which parental resources, and in particular parental education, influence the labor market position and the living standard of individuals.

This analytic question ties in with a contemporary political issue, since Brazil set up extensive means-based transfer programs in 1999 that were conditional on sending children to school (Bolsa Escola) and stopping child labor (PETI). These programs have now been combined into a single program called Bolsa Familia and are reaching cruising speed with widespread coverage. However, it is not easy to evaluate the impact of these programs since, unlike the Mexican Progresa program, no randomly allocated pilot setup has been implemented. An ex-ante evaluation of the Bolsa Escola program using a structural microeconometric model finds that the transfers have a significantly positive, albeit modest, impact on school enrollment and child labor. Hence they only have a marginal impact on income inequality and poverty (Bourguignon, Ferreira, and Leite 2003). An ex-post evaluation of the program is underway using data from the PNAD surveys, which identify the recipient households (Leite 2006).

Whatever the impact of these programs on the education of the most underprivileged children, a second question arises as to the long-run impact of a decrease in educational inequality on the distribution of income in Brazil. As regards the reduction of income inequality, the hopes raised by the huge surge in the average level of education have not yet been realized, contrary to optimistic forecasts by Lam and Levison (1991) (see Ferreira and Paes de Barros 2000 and 2004 on household income poverty). A recent paper by Bourguignon, Ferreira, and Menendez (2007) applies microsimulation techniques to the 1996 PNAD survey to analyze the contribution of inequalities of educational and income opportunity to the formation of inequality in an urban environment. It concludes that the canceling out of inequality due to social origin variables (race, region of birth, and parental education and occupation) would reduce the Theil index of individual earnings by more than one-fifth. The study's authors deem these findings disappointing since they only bring Brazil down to an average level of inequality by Latin American standards and a level way above comparable Asian countries. They however argue that this 20 percent share should be considered as a lower bound. They also find that a large part (55 to 75 percent) of the impact of factors of origin on individual earnings is associated with parental schooling. Lastly, 70 percent of this impact can be imputed to the direct effect of factors of origin on earnings while the remaining 30 percent corresponds to the indirect effect of social origins going through education—that is, the equalization of schooling opportunities.

This chapter addresses similar questions using a different set of data and other econometric methods. The remainder of the introduction provides a road map along with an overview of the main results.

The first section describes the data and the construction of the main variables. We use four PNAD surveys from 1976, 1982, 1988, and 1996 to focus on individual earnings inequalities among men aged 40 to 49 and to conduct a historical decomposition of the evolution of inequalities; in contrast, Bourguignon, Ferreira, and Menendez (2007) conduct static microsimulations by cohorts on the 1996 survey.

The second section describes the evolution of two kinds of earnings inequality over the 1976–1996 period: overall inequality in observed earnings and inequality of opportunity. Alongside traditional indicators of earnings inequality, we construct and calculate—for the first time in the case of Brazil—inequality of opportunity indicators in keeping with the axiomatics proposed by Roemer (1996 and 1998) and by Van de Gaer (1993) and Van de Gaer, Schokkaert, and Martinez (2001). Inequality of opportunity is defined as the amount of earnings inequality that can be attributed to individuals' social origins. We first show that the two kinds of earnings inequality displayed a similar historical path, including a peak in the late 1980s with the end of the dictatorship (1985) and the height of the inflationary crisis (the Cruzado, Bresser, Summer, and Collor Plans). All things considered, overall inequality rose slightly from the beginning to the end of the period, while inequality of opportunity posted a slight drop.

A third section looks at educational inequality with the same lenses as for earnings: overall inequality in education levels and inequality of opportunity. It reveals that the average number of years spent in the education system rose steadily for the generations born from 1927 to 1956, with a slight acceleration for the generations born in and after the 1940s. Intergenerational educational transmission, defined as the strength of the association between fathers' and sons' education, also recorded a very slight downturn for these generations. The rise in secondary and higher education immediately following the war, meaning the generations educated from 1945 to 1965, benefited mainly the children of the upper classes. For the generations educated from 1955 to 1975, the expansion of primary schooling had more benefit for the children of the underprivileged classes.

A fourth section then looks at the impact of these educational changes on earnings inequalities evolution. Three factors of the evolution are considered: 1) changes in the distribution of education of fathers and of sons; 2) changes in the pattern of mobility corresponding to the transition matrix between them; and 3) the structure of returns to parental education and own education. As an alternative to the parametric microsimulation techniques, we propose a semiparametric decomposition furnished by the log-linear model and nonparametric reweighting techniques inspired by Di Nardo, Fortin, and Lemieux (1996). We reveal that changes in the distribution of education levels initially had an inegalitarian effect before becoming equalizing in the late 1980s, for both kinds of earnings inequality. However, other factors, especially macroeconomic shocks, with soaring inflation and a drop in the minimum wage in real terms, provoked a sharp rise in earnings inequalities from 1982 to 1988. Yet this increase was virtually absorbed in the 40–49-year-old age bracket from 1988 to 1996 due to the expansion of primary education. Moreover, the change in the structure of earnings by education level and type of social origin had an egalitarian effect mainly at the end of the period, in particular in the form of a decrease in returns to education. Lastly, we determine that the historical growth in intergenerational educational mobility for the generations born from 1927 to 1956 was too small to play a significant part in the developments observed. This explains the persisting inequality of economic opportunity at a high level.

A fifth and final section explores the potential for a reduction in economic inequality stemming from an acceleration of intergenerational educational mobility; that is, a mitigation of what Bourguignon, Ferreira, and Menendez (2007) call the "indirect effect" of parental education on earnings. This kind of improvement indeed constitutes a long-term target for the Bolsa Escola program. We confirm that the bulk of the inequality of opportunity on the labor market can be imputed to this factor; in contrast with Bourguignon, Ferreira, and

Menendez, we attribute an even larger share to this indirect effect, in comparison with the direct impact of fathers' education on earnings. However, as found by Bourguignon, Ferreira, and Menendez, both effects only play a modest role in overall inequality. We nevertheless put forward that, in contrast with the historical decompositions or the impact on inequality of opportunity, this last evaluation is highly sensitive to earnings measurement errors.

2.2 Data

We use the data from four editions of the national survey of households (PNAD) conducted by the Brazilian Institute of Statistics (IBGE) in 1976, 1982, 1988, and 1996. The PNAD surveys cover a large sample since the data concern nearly 100,000 households every year. The sample is representative of the population of Brazil, but excludes the rural areas in the northern region (the Amazon).[1]

These four editions contain information on the adults' social origins, collected for the head of household and his spouse.[2] This concerns the father's level of education and occupation when the individual started working.[3] In addition, a question on migration provides information on the individual's place of birth (Federative Republic State)[4] and the questionnaire on demographic characteristics provides information on the individual's color. Overall, therefore, we use four data on social origins.

We restrict the sample to men aged 40 to 49 and subsequently disregard age effects on the assumption that such effects are negligible within this group. We limit the sample to men declared as the head of household or, more rarely, the spouse of the head (who combined represent 92 to 94 percent of this age bracket depending on the edition) and to employed individuals (93 to 89 percent, with this proportion decreasing over time) for whom information on social mobility, working hours, and earned income is provided. Our samples cover 2,860 observations in 1976; 18,833 in 1982; 11,304 in 1988; and 14,096 in 1996.

We construct an hourly earnings rate variable based on the information on monthly incomes in the different economic activities, wage and nonwage combined,[5] and on the weekly hours worked in these activities.[6] The incomes are discounted to September 2002 Brazilian reals using the IBGE deflators derived from the INPC national consumer price index. Ferreira, Lanjouw, and Neri (2003) posit that the PNAD underestimates agricultural earnings due to the lack of information on income in kind and production for own consumption, and overestimates the production of family businesses due to the lack of information on their expenditure on inputs. Overall, they deem that incomes are underestimated in rural areas. This would appear to be borne out by a comparison with the incomes measured by the 1996–1997 Pesquisa Sobre Padrões de Vida (PPV) living standards measurement survey containing more information on these points. Despite these potential measurement errors, we do not limit the sample to urban areas as done by Bourguignon, Ferreira, and Menendez (2007). Analyzing intergenerational mobility based on an urban subsample can result in substantial selection biases. We believe that such biases are greater than those caused by the underestimation of incomes in rural areas. Disregarding spatial variations in purchasing power constitutes another source of potential bias in the measurement of incomes. Ferreira, Lanjouw, and Neri (2003) propose a series of regional deflators based on data from the 1996 Pequisa de Orçamento Familiar (POF) household budget survey. We have tested these deflators in our empirical analyses for this year and observed that the findings changed little. We therefore do not correct these potential biases in the rest of this work.

The variable used for the education level of the individuals in the sample corresponds to the highest education level attained in numbers of years after entry into primary school, which is normally at seven years old.[7] We use a discrete decomposition of this variable into nine education levels (0, 1, 2, 3, 4, 5–7, 8, 9–11, and 12 or more years of education).

We use two characterizations of the social origins of the individuals in the sample. The first consists of four categorical variables. Variable 1 is a color variable coded into two categories, *white* for individuals declared as being white or Asian and *black* for individuals declared as being black, mixed race, or Indian. A birth region is variable 2, coded into four categories designed so as to optimize both their sample size and their discriminating power,[8] covering respectively the individuals born in (1) the Federal District and the state of Sao Paulo; (2) the states of the southern region [excepting Rio Grande do Sul], center-western region and west of the northern region; (3) the states of the southeastern region [excepting Sao Paulo], of the south of the northeastern region [Alagoas, Bahia, and Sergipe], and Rio Grande do Sul; (4) the states of the north of the northeastern region and east of the northern region [Amapa and Para]). Father's level of education is variable 3, coded into four categories covering respectively the individuals whose father (1) never went to school, (2) is literate or for whom the interviewee is unable to give an answer, (3) completed one of the first four years of primary education, and (4) completed at least the fifth year of primary education). The fourth variable concerns the father's occupation, and is coded into four categories covering respectively the individuals whose father was (1) a farmer; (2) employed in a traditional industry, a domestic employee, or whose occupation is poorly defined or for whom the interviewee is unable to give an answer; (3) employed in a modern industry, an unincorporated entrepreneur, or employed in a service sector; and (4) in a skilled profession, an employer, administrator, or manager. These four variables identify 128 groups of potential social origins.

The second characterization of social origins consists of a nine-category classification based on the father's level of education and occupation, covering the individuals whose father (1) never went to school and was a farmer, (2) never went to school and had another occupation, (3) was merely literate and was a farmer, (4) was literate and had another occupation, (5) completed one of the first four years of primary education and was a farmer, (6) completed one of the first four years of primary education and had another occupation, (7) completed one of the four years of upper primary education (5–8), (8) completed nine or more years of education, and (9) the interviewee was unable to answer.

We use resampling techniques (bootstrapping) to estimate the accuracy of the statistics calculated, including our decompositions. For this, we take into account the sample design used for the PNAD surveys; that is, the stratification of the sample into 36 natural strata corresponding to 27 Brazilian states and nine metropolitan regions (Bertail and Combris 1997).[9]

2.3 Growth in Earnings Inequalities over the 1976–1996 Period

In this section, we describe the changes in the measurements of overall inequality and inequality of opportunity regarding the hourly earnings of men aged 40 to 49.

2.3.1 Overall Earnings Inequality

Table 2.1 presents the growth in overall inequality in the distribution of hourly earnings as measured by the Gini and Theil indices.

Table 2.1
Measurements of overall inequality in hourly earnings

	1976		1982		1988		1996	
Gini index	0.570	(0.009)	0.585*	(0.004)	0.623*	(0.005)	0.599*	(0.005)
Theil index	0.625	(0.027)	0.687	(0.017)	0.772*	(0.018)	0.719	(0.028)
(Per capita GDP)	100		105.4		114.9		120.4	

Source: PNAD surveys, IBGE.
Coverage: Men aged 40 to 49.
Reading: Indices on inequality in the distribution of hourly earnings.
Notes: * indicates significance at 5 percent compared with the previous year; (in brackets): bootstrap standard deviations, 100 replications.

The Gini index remains close to 0.60 for the entire period. It increases significantly from 1976 to 1988, then falls from 1988 to 1996 before returning to a level slightly above, but not significantly different to, its 1976 level. The Theil index displays similar growth. It rises more sharply than the Gini index from 1976 to 1988, and decreases from 1988 to 1996 to a level significantly higher than in 1976. The difference between the growth in the two indices shows that overall inequality changed little over these twenty years, but that inequality rose, to the detriment of the bottom of the earnings distribution. These trends are illustrated by figure 2.1, which presents the smoothed density differences in hourly earnings from 1976 to 1996.

2.3.2 The Inequality of Labor Market Opportunity

We construct the inequality of labor market opportunity indices in keeping with the two main economic literature proposals on economic justice and equality of opportunity (Roemer 1996 and 1998; Van de Gaer 1993; Van de Gaer, Schokkaert, and Martinez 2001). For a given outcome variable (here hourly earnings), both proposals distinguish between what is due to *circumstances*, defined as an individual's characteristics that influence his outcome but over which he has no control (here social origin), and what is due to *effort*, for which the individual is held responsible. More generally, we use this latter term to cover all the outcome factors considered irrelevant to the establishment of illegitimate inequality.

The first approach proposed by Roemer considers that only the relative efforts in each group of circumstances (called *types* by this author) are comparable. The inequality between types is then measured by comparing individuals with the same relative level of effort; the in-

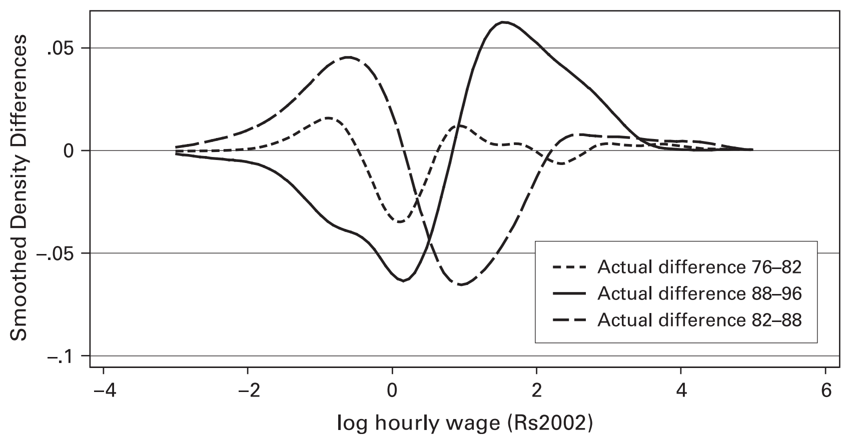

Figure 2.1
Variations in hourly earnings densities.
Source: PNAD surveys, IBGE. *Method*: Double smoothing by a Gaussian kernel function (bandwidth 0.2).

equality of opportunity is measured at different points of the distribution of relative levels of effort and these measurements are then aggregated into a single index. Roemer proposes measuring relative levels of efforts as within-types quantiles for the outcome variable. We here choose to compare deciles of hourly earnings conditional on the types of social origin.[10] We calculate the inequality indices at each decile and aggregate them, taking their average. These Roemer indices are written

$$\mathrm{ROE} = 1/10 \cdot \sum_{\pi} \mathrm{I}\{y_{o,\pi}\} \tag{2.1}$$

where o is an index for the different types of social origins, $y_{o,\pi}$ is the earning at decile π for type o, and I is an index of inequality. Instead of a traditional index of inequality like Gini or Theil, Roemer favors the minimum function (I = min), in keeping with a Rawlsian maximin principle. We also compute this original Roemer's index.

The second approach proposed by Van de Gaer (1993) considers that there is equality of opportunity when the distribution of expected earnings is independent of social origins. The extent of equality of opportunity is then measured by an indicator of the inequality of income expectations obtained by individuals of different origins. These conditional income expectations can be obtained from the distribution of

Table 2.2
Measurements of the inequality of economic opportunity (128 types of origin)

	1976		1988		1996	
VDG approach						
Gini index	0.385	(0.016)	0.409	(0.008)	0.359*	(0.007)
Theil index	0.254	(0.023)	0.280	(0.012)	0.213*	(0.009)
Roemer approach						
Minimum	1.297	(0.080)	1.048*	(0.043)	1.223*	(0.045)
Gini index	0.342	(0.013)	0.375*	(0.007)	0.343*	(0.005)
Theil index	0.211	(0.020)	0.243	(0.010)	0.197*	(0.006)

Source: PNAD surveys, IBGE.
Coverage: Men aged 40 to 49.
Reading: Inequality of opportunity indices calculated based on 128 categories of social origins constructed from four variables regarding the father's level of education (4 categories), the father's occupation (4), region of birth (4), and color (2); not available in 1982.
Notes: *indicates significance at 5 percent compared with the previous year; (in brackets): bootstrap standard deviations, 100 replications.

average earnings estimated by categories of origin; very simply, we can choose, for instance, the Gini of average earnings by category of origin.[11] In their general form, these Van de Gaer indices are written

$$\mathrm{VdG} = \mathrm{I}\{\mathrm{E}(\mathrm{y} \mid \mathrm{o})\} \tag{2.2}$$

where I is an inequality index and $\mathrm{E}(\mathrm{y} \mid \mathrm{o})$ is the earning expectation conditional on social origin o.

We therefore calculate two series of inequality of opportunity indices. We use the two social origin characterizations comprising respectively 128 and 9 categories of origins. The results are presented in tables 2.2 and 2.3.

As argued by Van de Gaer, Schokkaert, and Martinez (2001), the two Roemer and Van de Gaer measurements considered here produce the same rankings when the transition matrices between origins and outcomes are "Shorrocks monotonic" (Shorrocks 1978)—that is, when the most underprivileged types of origin in each decile are the same.

We can first of all observe that the indices based on nine types of origin (table 2.3) underestimate the inequality of opportunity by 10 percent to 20 percent compared with the indices based on 128 types (table 2.2). The Gini indices measured are situated between 0.30 and 0.40. Note that the nondecomposable nature of this index makes it impossible to use to deduce a measurement of the proportion of inequality of opportunity in overall inequality. The Theil indices measured are situ-

Table 2.3
Measurements of the inequality of economic opportunity (nine types of origin)

	1976		1982		1988		1996	
VDG approach								
Gini index	0.339	(0.015)	0.351	(0.007)	0.365	(0.009)	0.317*	(0.007)
Theil index	0.212	(0.021)	0.222	(0.009)	0.239	(0.013)	0.173*	(0.008)
Roemer approach								
Minimum	2.020	(0.056)	1.988	(0.025)	1.747*	(0.032)	2.034*	(0.046)
Gini index	0.327	(0.014)	0.339	(0.006)	0.357*	(0.007)	0.322*	(0.006)
Theil index	0.222	(0.024)	0.228	(0.009)	0.246	(0.011)	0.192*	(0.009)

Source: PNAD surveys, IBGE.
Coverage: Men aged 40 to 49.
Reading: Inequality of opportunity indices calculated based on nine categories of social origins.
Notes: * indicates significance at 5 percent compared with the previous year; (in brackets): bootstrap standard deviations, 100 replications.

ated between 0.20 and 0.30. In this case, the decomposability of the Theil index means that the contribution of social origins to overall inequality can be estimated at nearly 30 percent. These findings can be directly compared with those of Bourguignon, Ferreira, and Menendez (2007), who attribute around 26 percent of the overall inequality to social origins for men aged 40–59 in 1996.

All indices tell the same story about the evolution of the inequality of economic opportunity between 1976 and 1996, thus confirming the kind of consistency provided by Shorrocks's monotonicity. As already announced, we also present the average (over deciles) of minimum earnings for the different categories of origin at each decile of the earnings distribution.[12] This measurement corresponds to Roemer's first proposal to define the equal opportunity policies (see equation 2.1). This indicator grows in parallel with the Gini and Theil indices.

The indices also display similar growth to the overall inequality indices. All the indices find that the inequality of opportunity rises from 1976 to 1988, and that this rise is generally significant at least at 10 percent (and at 5 percent for the Gini index when using the Roemer approach). All the indices subsequently post a decrease in inequality of opportunity, and this drop is also significant (at 5 percent for all the indices). In all cases, the end-of-period indices (1996) are the lowest even though they are not significantly different from the indices at the beginning of the period (1976). Nevertheless, it is possible to say that the inequality of opportunity fell slightly from 1982 to 1996.[13]

2.4 Intergenerational Educational Mobility

In this section, we leave aside the inequality of earnings opportunity to concentrate on the inequality of educational opportunity, measured here by the number of years of education. Contrary to earnings, it would be problematic to treat the number of years of education as a suitable continuous metric for measuring the welfare procured by education. This section therefore uses another method to describe the changes in the inequality of educational opportunity: the comparison of odds ratios. We also limit our study here and in the following section to a categorization of social origins based on the father's education and occupation, in the form of the second nine-category origin variable described in section 2.2.

Table 2.4 shows growth in the average number of years of education and the distribution of years of education. It reveals that the average number of years spent in the education system rose steadily, by 2.3 years for the generations born from 1927 to 1956, with a slight acceleration for the generations born in and after the 1940s. However, it also shows that this growth was mainly in secondary and higher education for the first cohorts: the proportion of individuals having never attended school remains stable as well as the proportion of those having completed some primary education (from five to eight years of ed-

Table 2.4
Distribution of education levels by year

Year of birth	1976 1927–36	1982 1933–42	1988 1939–48	1996 1947–56
Never attended school	28.4	28.0	22.2	16.9
1 year	7.5	5.6	5.6	3.3
2 years	10.6	9.0	8.7	5.7
3 years	11.8	11.7	10.7	8.3
4 years	19.9	19.9	20.7	20.2
5–7 years	9.1	8.6	9.4	11.8
8 years	4.2	5.3	5.6	9.1
9–11 years	4.4	5.8	8.2	13.4
12 years and over	4.1	6.1	9.0	11.3
Total	100.0	100.0	100.0	100.0
Average no. of years	3.3	3.8	4.6	5.6

Source: PNAD surveys, IBGE.
Coverage: Men aged 40 to 49, employed, head of household, or spouse of head.

ucation), whereas the distribution of education levels shifts toward the top. The intermediate cohorts born during World War II post both an upturn in school attendance, which increases by 6 percentage points (the weight of the never-attended category decreases from 28 to 22 percent), and continued sharp growth in the weight of secondary and higher education sectors (more than eight years of education), from 11.9 ($= 5.8 + 6.1$) to 17.2 percent. Upper primary education still remains stable from 13.9 ($= 8.6 + 5.3$) to 15 percent. It is only in the last cohorts born after the war that primary education also shows marked growth, from 15 to 20.9 percent.

These developments were reflected at the beginning of the period by a sharp rise in the probability of access to secondary and higher education for the children of privileged families, and then at the end of the period by a rise in the probability of access to upper primary education (five to eight years) for less privileged children (see the destination matrices in the working paper version of Cogneau and Gignoux 2005). For the postwar generations, therefore, the expansion of education structurally generated ascending intergenerational growth paths among certain children of modest origin. However, it did not necessarily give rise to greater equality of educational opportunity; that is, it did not necessarily increase their access to higher levels when compared with more advantaged children. This is what we shall study now.

We use the following notations: our level of education variable S is divided into nine categories indexed by $s = 1, \dots 9$; our social origin variable O is also divided into nine categories indexed by $o = 1, \dots 9$. The analysis of educational mobility over the 1976–1996 period is based on the estimation of log-linear models using the four stacked educational mobility tables cross-tabulating S et O. The log-linear model proposes a descriptive analysis of the counts of these four mobility tables:

$$\mathrm{Ln}[n_t(s,o)] = \mu + \alpha(s) + \beta(o) + \gamma(s,o) + \mu_t + \alpha_t(s) + \beta_t(o) + \gamma_t(s,o) \tag{2.3}$$

where $n_t(s,o)$ is the number of individuals in the contingency table cross-tabulating S and O at year t. Ln denotes neperian logarithm.

This decomposition is purely descriptive and is unique under the following constraints:

$$\Sigma_s \alpha(s) = 0; \quad \Sigma_o \beta(o) = 0; \quad \Sigma_t \mu_t = 0.$$

For all s and o:

$\Sigma_s \gamma(s, o) = 0; \quad \Sigma_o \gamma(s, o) = 0,$

and for all t, s and o:

$\Sigma_s \alpha_t(s) = 0; \quad \Sigma_o \beta_t(o) = 0; \quad \Sigma_s \gamma_t(s, o) = 0; \quad \Sigma_o \gamma_t(s, o) = 0.$

Coefficients $\gamma(s, o)$ and $\gamma_t(s, o)$ are directly linked to odds ratios of the educational mobility table $n_t(s, o)$ (Bishop, Fienberg, and Holland 1975):[14]

$$\text{Odd-R}_t(s, o; s', o') = [n_t(s, o) n_t(s', o')]/[n_t(s', o) n_t(s, o')] \tag{2.4}$$

These odds ratios compare the probabilities of access to education level s versus s′ for two sons with different educational origins o and o′. For instance, let s be the 8 years of completed primary education and s′ be the level just below (5–7 years). Although symmetry is not required, let o and o′ stand for the same levels in the generation of fathers. Then, for two individuals, one whose father did not go more than 5–7 years and another whose father achieved 8 years, the odds ratio can be read as the relative probability of reproducing their father's position rather than of changing it.

Under the assumption that counts $n_t(s, o)$ follow a multinomial distribution, this model can be estimated by maximum likelihood, whether in its saturated form (equation 2.3) or in a more constrained form where, for instance, some parameters are assumed to be equal to zero. The joint test of the hypothesis [$\gamma_t(s, o) = 0$ for all (s, o) and t] can be therefore written as a likelihood ratio test following with a law of χ^2. It is used to evaluate the existence of a change in nonstructural educational mobility, in the sense of a change in the odds ratio, independently of the change in the marginal distribution of origins and education levels from one period to the next.

The global test suggests that we should reject the hypothesis of odds-ratio stability over the four years. Odds-ratio stability is also rejected for all the pairs of years from 1976 to 1996. Table 2.5 presents some odds ratios, called reproduction coefficients here since the fathers and sons categories are the same. For the categories considered, it shows that educational mobility was lower for men aged 40 to 49 in 1982 and in 1988 than for those aged 49 to 49 in 1996 and even in 1976.

The expansion of education has moreover given rise to a race for qualifications, shifting the educational hierarchy upward, and also a

Table 2.5
Educational mobility reproduction coefficients

Year of birth	1976 1927–36	1982 1933–42	1988 1939–48	1996 1947–56
Schooled/unschooled	6.29 (0.81)	8.70* (0.22)	10.80 (0.85)	7.40*° (0.23)
5 years or +/less than 5	24.66 (25.67)	28.36 (7.78)	23.72 (8.42)	22.63 (7.82)
5 years or +/1–4 years	7.66 (2.76)	10.41 (1.20)	7.82 (1.03)	11.28 (2.20)
1–4 years/unschooled	2.58 (0.18)	3.63* (0.05)	3.60 (0.12)	2.86*° (0.05)

Reading: In 1976, for an individual whose father had never been to school and for an individual whose father attended school, the probability of reproducing the paternal situations was over six times higher than the probability of interchanging them.
Notes: *: For 1996, the odds ratio is significantly different (and lower) than in 1982, at the 5 percent level; for 1982, the odds ratio is significantly different (and higher) than in 1976. °: For 1996, the odds ratio is significantly different (and lower) than in 1988, at the 5 percent level.

probable quality race (private system versus state system), both producing an apparent drop in returns (Lam and Levison 1991).

2.5 The Effects of Educational Changes on Earnings Inequalities from 1976 to 1996

2.5.1 Methodology

Our methodology is based on the nonparametric reweighting techniques introduced by Di Nardo, Fortin, and Lemieux (1996) in an application to changes in the distribution of earnings in the United States. Here, we look at the impact of the distribution of two variables on the distribution of earnings; that is, individuals' schooling S and social origin O.

Like Di Nardo, Fortin, and Lemieux (1996) and the majority of papers that analyze inequality evolutions (Bourguignon, Ferreira, and Menendez 2007 as one other example), we look at Blinder-Oaxaca decompositions (Blinder 1973, Oaxaca 1973). In other words, our decompositions reconstitute counterfactual income distributions by applying counterfactual population structures to an observed earnings structure. These decompositions consist in calculating what the overall inequality and inequality of earnings opportunity would be in 1996 if, for example, the distribution of the population between education levels and categories of social origins had remained the same as in 1976, or if the structure of earnings by education level and social origin had not changed. The changes in the distribution of the population between education levels and categories of social origin can then be broken down into two notional changes, the first altering the marginal distributions and the second the relations between social origins and levels of education,that is, educational mobility.

These decompositions assume independence between the structure of earnings (here by education level and origin) and the distribution of the population. This assumption of independence implies the absence of general equilibrium effects: the distribution of the population by education level and origin does not alter the structure of earnings. It also implies the nonendogeneity of the origin and education level variables as regards the unobserved determinants of earnings: the conditional earnings densities (vis-à-vis origin and level of education) are assumed to be invariant to the redistribution of the population by origin or education level.[15]

2.5.1.1 Construction of the Counterfactual Inequality We first of all assume that we have a counterfactual distribution $dF^*(s,o)$ of education levels and origins in the population, whose construction we present later. As variables S and O are discrete, this distribution is perfectly summed up by frequencies $p^*(s,o)$.

As regards the effect on overall inequality, the basic idea consists in reweighting the observed distribution of earnings y. The observed income density is written

$$f_t(y) = \int f(y \mid s,o,t_y = t)\, dF(s,o \mid t_{s,o} = t) \tag{2.5}$$

and the counterfactual density

$$\begin{aligned} f_t^*(y) &= \int f(y \mid s,o,t_y = t)\, dF^*(s,o) \\ &= \int f(y \mid s,o,t_y = t)\, dF(s,o \mid t_{s,o} = t)\psi(s,o), \end{aligned} \tag{2.6}$$

where $\psi(s,o) = dF^*(s,o)/dF(s,o \mid t_{s,o} = t)$ is the weighting system to be applied to the observed distribution of earnings. Let $p^*(s,o)$ be the counterfactual population frequencies and $p_t(s,o)$ those of the real population. By applying the Bayes rule, this weighting system is written simply as $\psi(s, o) = p^*(s, o)/p(s, o)$.

The Equality of Opportunity (EOp) indices are functions of the conditional distribution of y vis-à-vis o and the distribution of origins in the population.

$$\mathrm{EOp}_t = \mathrm{EOp}[f(y \mid o, t_y = t), dF(o \mid t_o = t)] \tag{2.7}$$

It is obviously hard to produce counterfactuals for the conditional densities of earnings (y) vis-à-vis origins (o), $f(y \mid o, t_y = t)$, needed to construct a Roemer index. However, the Van de Gaer index only requires the conditional expectations

$$\mathrm{VdG}_t = I[E(y \mid o, t_y = t), dF(o \mid t_o = t)], \tag{2.8}$$

where I is a usual inequality index (Gini, Theil, or other entropy indices) applied to the distribution of $E(y \mid o)$ weighted by dF(o).

From this point of view, it is relatively easy to construct a counterfactual with a fixed earnings structure since, here again, it is simply a question of constructing a counterfactual of the distribution of the population by education level and origin type: $dF^*(s, o) = dF^*(s \mid o)\, dF^*(o)$. Hence,

$$\mathrm{VdG}_t^* = I[E^*(y \mid o, t_y = t), dF^*(o)], \tag{2.9a}$$

and

$$E^*(y \mid o, t_y = t) = \int E(y \mid o, s, t_y = t)\, dF^*(s \mid o). \tag{2.9b}$$

In the case of large samples, the conditional expectations $E(y \mid o, s, t_y = t)$ can be estimated by the empirical means for each subpopulation (s, o).

2.5.1.2 Construction of Counterfactual Educational Mobility We explain here how we construct counterfactual frequencies $p^*(s, o)$ using the log-linear model.

As mentioned in section 2.4, this model, in what is known as its saturated form, provides a descriptive decomposition of the observed frequencies $p_t(s, o)$:

$$\mathrm{Ln}[p_t(s, o)] = -\mathrm{Ln}(N_t) + \mu_t + \alpha_t(s) + \beta_t(o) + \gamma_t(s, o) \tag{2.10}$$

where N_t is the total number of individuals in the sample, μ_t is a constant, $\alpha_t(s)$ the effect of the margins of s, $\beta_t(o)$ the effect of the margins of o, and $\gamma_t(s, o)$ the effect of the interactions between o and s. This decomposition is unique under the constraints

$$\Sigma_s \alpha_t(s) = 0;\ \Sigma_o \beta_t(o) = 0;\ \Sigma_s \gamma_t(s, o) = 0;\ \Sigma_o \gamma_t(s, o) = 0, \quad \text{for all s and o.}$$ [16]

If S and O are independent of one another—that is, when the observed frequencies are equal to the product of marginal frequencies $p_t(s, o) = p_t(s, .)p_t(., o)$—then all coefficients $\gamma(s, o)$ are zero. Coefficients $\gamma_t(s, o)$ are directly linked to the odds ratios of the mobility table $p_t(s, o)$:

$$\text{Ln}[\text{Odd-R}_t(s, o; s', o')] = [\gamma_t(s, o) + \gamma_t(s', o')] - [\gamma_t(s', o) + \gamma_t(s, o')].$$

For each year t, we estimate the saturated log-linear model of frequencies and retrieve the coefficients $\gamma_t(s, o)$. Then we estimate a series of constrained models where the second order interactions $\gamma_t(s, o)$ are constrained to be equal to $\gamma_{t'}(s, o)$ for $t' \neq t$:

$$\text{Ln}[p_t(s, o)] = -\text{Ln}(N_t) + \mu_t^* + \alpha_t^*(s) + \beta_t^*(o) + \gamma_{t'}(s, o). \tag{2.11}$$

We hence obtain an estimated table $p^*_{t/t'}(s, o)$ whose margins are exactly those of t and whose odds ratios are those of t′. This is the distribution of the population in t if the educational odds ratios were those of t′. For the period $[t, t']$, we can therefore break down the change in the structure of the population by education level and category of origin into two movements: an educational mobility movement from $p_t(s, o)$ to $p^*_{t/t'}(s, o)$, and a movement in the marginal distributions of education and origins from $p^*_{t/t'}(s, o)$ to $p_{t'}(s, o)$. Such a decomposition can obviously operate in the opposite direction: an educational mobility movement from $p_{t'}(s, o)$ to $p^*_{t'/t}(s, o)$, and a movement in the marginal distributions of education and origins from $p^*_{t'/t}(s, o)$ to $p_t(s, o)$.

Here is an example of a decomposition of change in the population structure. Let's assume that there are only two groups of social origins and two groups of education and that the distributions observed on dates t and t′ are given by frequency tables 2.6 and 2.7.

The change in the marginal distributions of education levels and social origins from t to t′ could represent an expansion in education, with the marginal distributions of education levels and origins changing from (0.80; 0.20) to (0.60; 0.40) for both education and origins. Two

Table 2.6
Frequencies of the assumed distribution observed in t

	Education 1	Education 2
Origins 1	0.700	0.100
Origins 2	0.100	0.100

Note: Marginal distributions (0.80; 0.20) and reproduction coefficient of 7.

Table 2.7
Frequencies of the assumed distribution observed in t′

	Education 1	Education 2
Origins 1	0.400	0.200
Origins 2	0.200	0.200

Note: Marginal distributions (0.60; 0.40) and reproduction coefficient of 2.

Table 2.8
Frequencies of the distribution simulated with the marginal distributions of t and educational mobility of t′

	Education 1	Education 2
Origins 1	0.660	0.140
Origins 2	0.140	0.060

Note: Marginal distributions (0.80; 0.20) and reproduction coefficient of 2.

individuals of different origin have respectively seven and two times less chance of changing their original situations than of reproducing them in t and in t′. This reflects an increase in educational mobility.

The change in the population structure can be broken down into two movements, representing respectively the changes in marginal distributions and educational mobility. Table 2.8 presents the structure of the simulated population obtained by applying educational mobility in t′ to the table observed in t. We first simulate the change in educational mobility from table 2.6 to table 2.8, and then the change in marginal distributions from table 2.8 to table 2.7.

A second decomposition of these developments can be made. Table 2.9 gives the structure of the simulated population obtained by applying the marginal distributions of t′ to the table observed in t. We hence first simulate the change in marginal distributions from table 2.6 to table 2.9, and then the change in educational mobility from table 2.9 to table 2.7.

Table 2.9
Frequencies of the distribution simulated with the educational mobility of t

	Education 1	Education 2
Origins 1	0.467	0.133
Origins 2	0.133	0.267

Note: Marginal distributions (0.60; 0.40) and reproduction coefficient of 7.

2.5.1.3 Semiparametric Decomposition This last methodological part sums up the construction of the semiparametric decompositions of changes in inequality using the counterfactual educational mobility tables.

As regards overall inequality between the two dates t and t′, a first counterfactual density can be constructed using the table $p^*_{t/t'}(s, o)$ (educational mobility of t′ and marginal distribution of t) to reweight the observations in t. We combine equations (2.6) and (2.11):

$$f^*_{t/t'}(y) = \int f(y \mid s, o, t_y = t)\, dF(s, o \mid t_{s,o} = t)\psi^*_{t/t'}(s, o) \tag{2.12}$$

with $\psi_{t/t'}(s, o) = p^*_{t/t'}(s, o)/p_t(s, o)$.

We can then calculate a second counterfactual density by applying the table of educational mobility observed in t′ to the earnings structures of t:

$$f^{**}_{t/t'}(y) = \int f(y \mid s, o, t_y = t)\, dF(s, o \mid t_{s,o} = t)\psi_{t/t'}(s, o) \tag{2.13}$$

with $\psi_{t/t'}(s, o) = p_{t'}(s, o)/p_t(s, o)$.

The first counterfactual describes the movement of overall inequality that can be attributed to the change in nonstructural educational mobility, and the second describes the movement of inequality that can be attributed to the change in the structure of the population by education level and origin.

The residual difference between the second counterfactual and the density observed in t′, $f_{t'}(y)$, represents not only the impact of the change in earnings structures by education level and social origin (conditional expectations, or *returns*), but also all the other factors that have contributed to the deformation of the conditional densities, like changes in the composition of the labor force by unobserved skills or changes in the remuneration of unobserved skills (Lemieux 2002). Parametric estimation of conditional density $f(y \mid s, o)$ would allow us to make the distinction between returns to education and social origins and other elements, as in Juhn, Murphy, and Pierce (1993). It would require making two major additional assumptions, the first about the function relating hourly earnings Y to observables S and O, and the second about the distribution of unobservables of Y. Many Blinder-Oaxaca decompositions rely on a log-linear relationship and log-normality of unobservables, like in Bourguignon, Ferreira, and Menendez (2007). We preferred to stick with our semiparametric methodology.

Furthermore, when we no longer consider overall inequality but inequality of opportunity—and in the case of Van de Gaer, inequality of opportunity indices—the decomposition only involves the intergenerational transition matrix between social origin and education levels $dF(S, O)$ and expected earnings conditional to schooling and origins $E(Y \mid S, O)$, as can be seen from equations (2.9a–b). The assumption of exogeneity of S and O with respect to earnings Y allows us to estimate $E(Y \mid S, O)$ from the structure of average earnings by education level and social origins. This means that neither the distribution of unobservables (selection) nor the returns to them play any role in that inequality of opportunity measurement. In contrast with overall inequality, the residual third term of the decomposition can be interpreted as the impact of changes in the returns to education levels and to social origins.

Decompositions are path-dependent. In the case of overall inequality, it is possible to start by altering the marginal distribution, given constant educational mobility, and then to alter the nonstructural educational mobility. Decompositions of overall inequality can also be made backward, starting from the final date t'. Four decompositions are hence possible: MDR (Mobility, marginal Distribution and Residual), or DMR starting from t, and RMD or RDM starting from t'. As regards the Van de Gaer inequality of opportunity indices, the decompositions of the distribution of earnings can use two earnings structures (conditional expectations), that of t $(E(y \mid o, s, t_y = t))$ or that of t' $(E(y \mid o, s, t_y = t'))$. This means that there are ultimately six possible decompositions of the change in the distribution of earnings between t and t'. Let's take the decomposition example given earlier. As shown in table 2.10, we can apply the structure of earnings observed in t or that observed in t' to each of the four tables.

In practice, we do not consider the last two counterfactual paths that introduce an earnings structure change in the middle of the population

Table 2.10
The path-dependency of decompositions: six roads

	Observed t	First-stage simulation	Second-stage simulation	Observed t′
MDS	table 2.6, $E(y \mid o, s, t_y = t)$	table 2.8, $E(y \mid o, s, t_y = t)$	table 2.7, $E(y \mid o, s, t_y = t)$	table 2.7, $E(y \mid o, s, t_y = t')$
DMS	—	table 2.9, $E(y \mid o, s, t_y = t)$	table 2.7, $E(y \mid o, s, t_y = t)$	—
SMD	—	table 2.6, $E(y \mid o, s, t_y = t')$	table 2.8, $E(y \mid o, s, t_y = t')$	—
SDM	—	table 2.6, $E(y \mid o, s, t_y = t')$	table 2.9, $E(y \mid o, s, t_y = t')$	—
MSD	—	table 2.8, $E(y \mid o, s, t_y = t)$	table 2.9, $E(y \mid o, s, t_y = t')$	—
DSM	—	table 2.9, $E(y \mid o, s, t_y = t)$	table 2.9, $E(y \mid o, s, t_y = t')$	—

structure change. Note that these paths do not have their symmetric in the decomposition of overall equalities since the conditional densities are not estimated.

2.5.2 Empty Cells and Selection Biases

Some cells in the educational mobility matrices have small and even zero values: children of illiterate fathers very rarely go to university and, conversely, it is even rarer to find children of qualified fathers not attending school. The 1976 educational mobility matrix hence contains three empty cells, and those for 1982 and 1996 contain respectively two and one empty cells. However, the 1988 matrix has none. We have solved the problem from a technical point of view by allocating a very small value (0.5) to the few empty cells for the estimation of the log-linear model. The missing occurrences are then disregarded in the calculation of the indices and counterfactual densities. A comparison of the findings obtained for 1988, whose matrix has no empty cells, with the findings for the other years shows that the problem is fairly innocuous. The values in these cells remain low regardless of the simulation considered.

However, it might still be thought that the earnings observed for the rare individuals bear a selection bias. Unfortunately, it is particularly hard to correct this type of bias, given that it concerns as much the origin variables as the levels of education. For example, the estimation of nonparametric bounds for the returns to education (Manski and Pepper 2000) by type of origin yields particularly high upper bounds.

2.5.3 Results: Historical Decomposition of Growth in Earnings Inequalities from 1976 to 1996

Table 2.11 presents the results of the decompositions of the Theil inequality of opportunity indices and the Theil overall inequality indices in the three subperiods from 1976 to 1996: 1976–1982, 1982–1988, and 1988–1996.[17] Four decompositions are presented according to the path taken for each index and subperiod. In addition, the standard deviations for each effect are calculated by fifty samples with replacement (bootstraps) based on the sample's stratified sampling plan and applied to the entire decomposition procedure (including the log-linear model estimates used to generate the counterfactual mobility tables). Our comments concern the findings that are both statistically significant and robust to the order of the decomposition; in particular, the small size of the 1976 sample means that statistically significant variations are generally not obtained.

First of all, the effects of the variations in the marginal distributions of education levels and social origins are considerable. As regards the equality of economic opportunity, changes in the distribution of the education levels of individuals and their fathers initially had an inegalitarian effect before becoming equalizing as of the late 1980s. These changes dominate the inequality of opportunity growth paths. As already pointed out (section 2.4, table 2.4), the first two subperiods analyzed (1976–1982 and 1982–1988) show an increase in secondary and higher education from 1945 to 1965, benefiting mainly children of privileged origins. The third subperiod (1988–1996) corresponds to the generations educated from 1955 to 1975. This subperiod was marked by an expansion in primary education that had more benefit for the children of the underprivileged classes.

Therefore, among the men aged 40 to 49 whose father had completed nine or more years of education, 40 percent had completed 12 years or more of education in 1976, 58 percent in 1982, and 64 percent in 1988, with this proportion peaking at 65 percent in 1996. Likewise, among the individuals whose father had completed five to eight years of education (upper primary), the probability of entering secondary education (nine or more years) rose from 50 percent to 55 percent and then 64 percent from 1976 to 1988, peaking at 68 percent in 1996.

Conversely, among the men whose fathers were uneducated farmers, 53 percent had in turn not attended school in 1976 and 1982, 46 percent in 1988, and 41 percent in 1996. Of these same men, only 3 percent had attained the upper primary level (five years and more) in 1976, 5 percent in 1982, and 7 percent in 1988—but 15 percent in 1996.

The democratization of access to school came about mainly for the generations born after World War II, educated from 1955 to 1975, who were under 50 years old in 1996. This is why the first period of education expansion was rather disadvantageous in terms of the inequality of earnings opportunities, while the 1988–1996 period was particularly equalizing.

As regards the overall earnings inequality, the expansion of education likewise initially had an inegalitarian effect on the prewar generations (1976–1982) before becoming equalizing for the postwar generations (1988–1996). Its impact is found to be negligible for the intermediate generations (1982–1988). Other factors, in particular the nosedive in the minimum wage in real terms (a 20 percent drop), generated a sharp rise in earnings inequalities at the start of hyperinflation, from 1982 to 1988. Nevertheless, this increase was virtually absorbed among the 40–49 year old age bracket from 1988 to 1996 due to the expansion of primary education. Figure 2.2 represents the im-

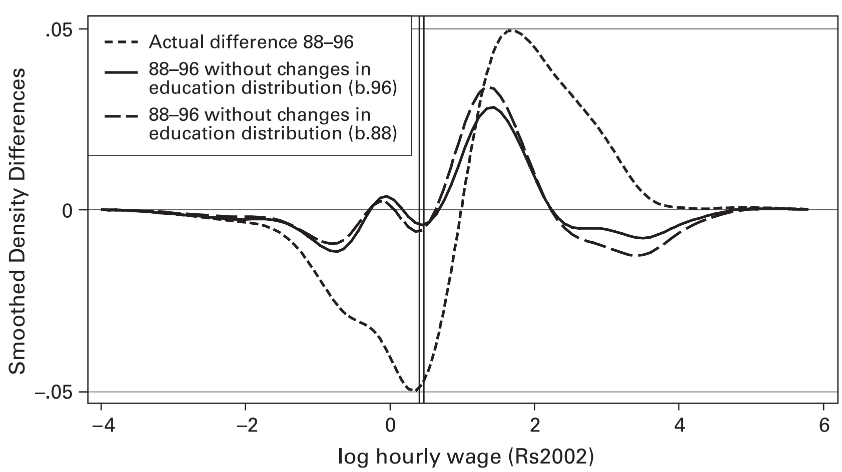

Figure 2.2
Counterfactual variation in earnings densities from 1988 to 1996.
Method: Double smoothing by a Gaussian kernel function (bandwidth 0.2). *Vertical bars*: minimum wage levels in 1988 and 1996.

pact of this expansion of primary schooling on the development of earnings densities in the last period. Vertical bars indicate the minimum wages levels in 1988 (right) and 1996 (left). It suggests that the observed reduction in poverty and inequalities would have been much lower without this change in the marginal distribution of education levels. The only notable development would have been a concentration of the distribution to the right of the minimum wage, probably due to the slow recovery in growth in the early 1990s and the end of hyperinflation in 1995.

Secondly, the change in the structure of earnings by education level and type of social origin had an egalitarian effect at the end of the period, in particular in the form of a sharp drop in returns to education from 1988 to 1996. For example, the ratio of hourly average earnings for uneducated men whose father was also uneducated to those of men with a secondary education whose father had reached the same education level was 11.4 in 1976 and 10.5 in 1982, rising to 11.3 in 1988 following a fall of over 15 percent in uneducated men's earnings, but finally descending to 8.8 in 1996. Over the 1988–1996 period, the narrowing of the earnings scale contributed almost equally with the expansion of primary education to the reduction in inequality of opportunity.

Lastly, table 2.11 shows that changes in intergenerational educational mobility for the generations born from 1927 to 1956 were too small to play a significant part in the developments observed. As the following section argues, this explains the persisting inequality of economic opportunity at a high level.

2.6 The Potential Effects of an Increase in Education Mobility on Earnings Inequalities

2.6.1 Methodology

It is also possible to consider counterfactual population structures other than the distributions observed for another year t'. We construct, for each year, the educational mobility matrices corresponding to the independence assumption

$$p_t^{(i)}(s, o) = p_t(s, .)p_t(., o) \tag{2.14}$$

where $p_t(s, .)$ (resp. $p_t(., o)$) stand for the row (resp. column) frequency of schooling level s (resp. origin o).

Table 2.11
Historical decomposition 1976–1996

	1976–1982						1982–1988						1988–1996					
	M: Ed. mobility		D: Distribution		S: Earnings		M: Ed. mobility		D: Distribution		S: Earnings		M: Ed. mobility		D: Distribution		S: Earnings	
Inequality of opportunity	effect	s.error	effect	s.error	effect	s.error	effect	s.error	effect	s.error	effect	s.error	effect	s.error	effect	s.error	effect	s.error
MDS	−0.005	0.029	0.016	0.012	−0.003	0.020	−0.002	0.004	0.011	0.006	0.011	0.011	−0.003	0.006	−0.041	0.005	−0.023	0.012
DMS	−0.001	0.011	0.013	0.019	−0.003	0.020	0.000	0.005	0.008	0.006	0.011	0.011	−0.003	0.006	−0.041	0.005	−0.023	0.012
SMD	0.011	0.014	0.024	0.013	−0.027	0.047	−0.004	0.005	0.012	0.007	0.012	0.010	−0.001	0.006	−0.035	0.005	−0.031	0.013
SDM	0.011	0.012	0.024	0.015	−0.027	0.047	−0.001	0.005	0.009	0.007	0.012	0.010	0.000	0.006	−0.036	0.005	−0.031	0.013
			Var.	s.error					Var.	s.error					Var.	s.error		
Total variation			0.008	0.026					0.020	0.016					−0.067	0.016		
percentages over the total:																		
MDS	−60%		201%		−41%		−12%		56%		57%		5%		61%		34%	
DMS	−17%		158%		−41%		2%		41%		57%		4%		62%		34%	
SMD	141%		301%		−341%		−19%		59%		60%		2%		53%		46%	
SDM	139%		302%		−341%		−4%		44%		60%		0%		54%		46%	

Overall inequality	M: Ed. mobility		D: Distribution		R: Residual		M: Ed. mobility		D: Distribution		R: Residual		M: Ed. mobility		D: Distribution		R: Residual	
	effect	s.error	effect	s.error	effect	s.error	effect	s.error	effect	s.error	effect	s.error	effect	s.error	effect	s.error	effect	s.error
MDR	−0.014	0.016	0.019	0.016	0.058	0.031	0.001	0.003	0.000	0.007	0.082	0.025	0.002	0.004	−0.052	0.006	0.001	0.035
DMR	−0.019	0.013	0.024	0.013	0.058	0.031	0.002	0.003	0.000	0.007	0.082	0.025	0.002	0.003	−0.052	0.006	0.001	0.035
SMR	−0.006	0.021	0.025	0.017	0.044	0.035	0.001	0.003	−0.001	0.006	0.084	0.026	−0.001	0.005	−0.039	0.010	−0.010	0.031
RDM	−0.004	0.011	0.022	0.012	0.044	0.035	0.002	0.003	−0.002	0.007	0.084	0.026	−0.001	0.004	−0.038	0.011	−0.010	0.031
			Var.	s.error					Var.	s.error					Var.	s.error		
Total variation			0.063	0.035					0.084	0.025					−0.049	0.035		
percentages over the total:																		
MDR	−22%		30%		93%		2%		0%		98%		−4%		106%		−3%	
DMR	−30%		38%		93%		2%		0%		98%		−3%		106%		−3%	
RMD	−10%		40%		70%		1%		−1%		100%		2%		79%		20%	
RDM	−6%		36%		70%		2%		−2%		100%		2%		78%		20%	

Reading: Semiparametric decomposition of variations in the Van de Gaer inequality of opportunity indices and overall inequality indices in terms of the respective effects of changes in educational mobility, marginal distributions of origins and education levels, and earnings from 1976 to 1996. The simulation paths are noted by the order of changes, with M denoting educational mobility, D the marginal distributions of social origins and education levels, and S the structures of earnings by education level and type of origin or R the residual (see text). Standard deviations (s.e.) obtained by bootstrapping with 50 replications.

Table 2.12
Frequencies of the assumed distribution observed in t and average earnings by level of education and type of origin

	Education 1	Education 2
Origins 1	0.40 y11	0.10 y12
Origins 2	0.10 y21	0.40 y22

Note: Marginal distributions (0.50; 0.50) and odds-ratio of 16.

We then apply these perfect mobility matrices to the earnings structures observed in the year t considered, and hence estimate the total contribution of educational mobility to the observed inequalities. This type of counterfactual simulation leaves the population distributions by type of origin and especially by education level invariant. It could therefore be thought that the general equilibrium effects count less, since the educational supply remains similar. However, there is an extensive redistribution of the population within the educational mobility matrix. So the assumption of the absence of selection effects and especially the exogeneity of social origin as regards the unobserved earning determinants has a large weight here. Our theoretical scenario consists of simulating a fictitious world far removed from reality in which the children of university-educated fathers stand as much chance of failing at primary school as the children of illiterate fathers.

To illustrate this simulation, we again assume that there are only two groups of social origins and two groups of education and that the distribution observed on date t is given by frequency table 2.12. Two individuals of different origin then are sixteen times more likely to reproduce their fathers' situations than to change them (reproduction coefficient of 16). Perfect educational mobility can be simulated by seeking the population structure that retains the marginal distributions (0.50; 0.50) of education levels and origins, but such that two individuals of different origins stand as much chance of changing their situations as of reproducing them (odds ratio of 1). Table 2.13 presents the frequencies for such a simulated distribution. The probabilities of reaching a given level of education conditional on origins are equal.

The counterfactual earnings densities are obtained by reweighting the observations by the ratios of values between tables 2.12 and 2.13, based on the formula given by equation (2.6). The counterfactual in-

Table 2.13
Frequencies of the distribution simulating perfect educational mobility

	Education 1	Education 2
Origins 1	0.25	0.25
Origins 2	0.25	0.25

Note: Marginal distributions (0.50; 0.50) and odds-ratio of 1.

dices of overall inequality are then calculated on the basis of these reweighted data.

The Van de Gaer equality of opportunity index is obtained from the conditional earnings expectations by type of social origin $E(y \mid o, t_y = t)$, estimated on the basis of the averages observed for the sample.

For the observed distribution:

$$E(y \mid o = 1, t_y = t) = (0.40/0.50)y11 + (0.10/0.50)y12$$

$$E(y \mid o = 2, t_y = t) = (0.10/0.50)y21 + (0.40/0.50)y22$$

For the simulated distribution:

$$E(y \mid o = 1, t_y = t) = (0.25/0.50)y11 + (0.25/0.50)y12$$

$$E(y \mid o = 2, t_y = t) = (0.25/0.50)y21 + (0.25/0.50)y22$$

The only source of inequality of economic opportunity remaining in the simulated distribution of earnings comes from the direct effect of social origin on earnings, which is not associated with the individual's education ($y21 \neq y11$ et $y22 \neq y12$).

2.6.2 Results: Impact of Perfect Educational Mobility on Earnings Inequalities

Table 2.14 presents the results of these perfect educational mobility simulations for 1976, 1982, 1988, and 1996.

The simulations reduce the Gini inequality of opportunity index by at least 54 percent and the Theil index by at least 78 percent.[18] Intergenerational educational mobility plays a predominant role in the inequality of opportunity on the labor market. The residual inequality is due to the earnings gaps directly associated with social origin. Given that these gaps are fairly small at the bottom of the education level distribution, the Theil index decreases considerably more than the Gini

Table 2.14
Simulations of perfect educational mobility

	1976		1982		1988		1996	
Inequality of opportunity								
Gini index								
Simulated variation	−0.184	0.015	−0.260	0.009	−0.266	0.017	−0.186	0.008
Observed	0.341	0.026	0.351	0.008	0.366	0.007	0.315	0.008
	−54%		−74%		−73%		−59%	
Theil index								
Simulated variation	−0.163	0.016	−0.207	0.009	−0.220	0.014	−0.142	0.008
Observed	0.209	0.030	0.222	0.009	0.242	0.010	0.171	0.009
	−78%		−93%		−91%		−83%	
Overall inequality								
Gini index								
Simulated variation	−0.009	0.017	−0.028	0.006	−0.032	0.009	−0.026	0.010
Observed	0.570	0.009	0.586	0.005	0.623	0.004	0.597	0.005
	−2%		−5%		−5%		−4%	
Theil index								
Simulated variation	−0.028	0.055	−0.074	0.019	−0.100	0.030	−0.069	0.043
Observed	0.626	0.030	0.687	0.021	0.771	0.017	0.712	0.024
	−4%		−11%		−13%		−10%	

Reading: Comparison of Van de Gaer inequality of opportunity indices and overall inequality indices observed and obtained by simulating independence between education levels and social origins. Standard deviations obtained by bootstrapping with 50 replications.

index. Nevertheless, this difference in variation between the two indices depends to a large extent on the sound estimation of the social origin effects in cells with low or zero values in the educational mobility matrices.

As regards overall inequality, figure 2.3, estimated by double kernel smoothing, shows that perfect educational mobility not surprisingly concentrates the distribution of earnings around the average. However, under our assumptions, the equalization of educational opportunities only generates a reduction of one to three Gini index points depending on the year, or a relative reduction of two to five percent (table 2.14). Here again, the variation in the Theil index is greater, between 4 and 13 percent (10 percent in 1996) for the aforementioned reason. For 1996, this last finding is in line with the seven percent obtained by Bourguignon, Ferreira, and Menendez (2007) for the same age bracket as regards the indirect (education-related) effect of social origin on earnings inequality.

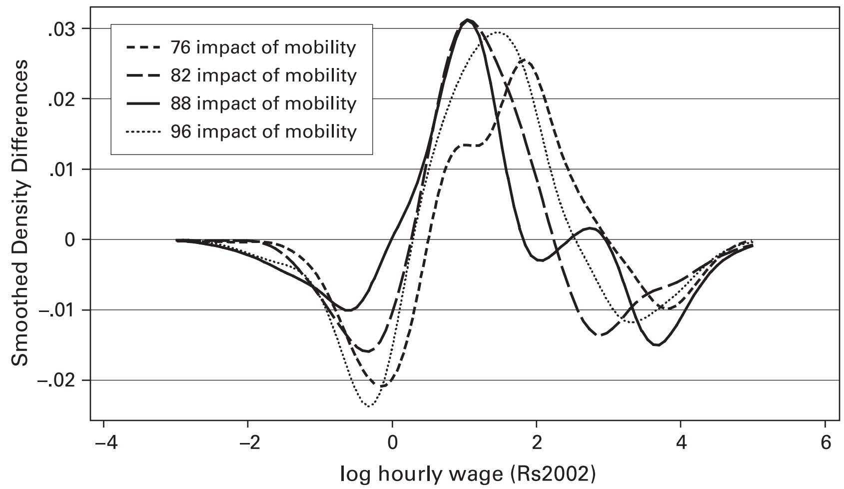

Figure 2.3
Differences between observed densities and simulated densities with perfect educational mobility.
Method: Densities simulated by reweighting using the formula given by equation (2.6) and based on educational mobility matrices, where origin and education level are independent, estimated using the formula given in equation (2.14).

However, both of our decompositions attribute a larger weight to the indirect channel going through educational mobility. When looking at the same cohorts (born between 1947 and 1956) in the same year (1996), and for overall inequality decomposition, we obtain a 42/58 indirect/direct sharing against 18/82 in Bourguignon, Ferreira, and Menendez (2007). Three main differences might explain this divergence between the two studies. A first one lies in the decomposition methology: nonparametric versus parametric. The second lies in the list of origin variables: rather restricted in our case (nine categories) due to the sample size constraints that bear on semiprametric estimations, but rather long in their case (with race, region of birth, and father's detailed occupation included, even if parental education ends up as the most important variable). A third and maybe more important difference lies in the sample selection, national versus urban, even though Bourguignon, Ferreira, and Menendez try to account for migration bias. Further research is warranted in order to understand the source of this divergence.

Coming back to the weight of educational mobility in overall inequality, we agree with Bourguignon, Ferreira, and Menendez in saying that our estimates as well as theirs only represent a lower bound. Contrary to the simulations regarding the inequality of opportunity indices, but also contrary to the historical decompositions presented in section 2.5, this last decomposition is indeed highly sensitive to measurement errors and transient components in the analyzed variable—here, hourly earnings. This is intuitively understood since this static decomposition can only concern the proportion of inequality corresponding to actual and permanent earnings gaps. In the working paper version of this chapter (Cogneau and Gignoux 2005), we use a simple case (log-normality) to show the effect of measurement errors or irrelevant transitory components in terms of their share in the variance of the analyzed variable. The review of the literature by Bound, Brown, and Mathiowetz (2001) suggests that a proportion of 20 to 30 percent is not unreasonable in the case of the measurement of hourly earnings. Yet the simulations show that a proportion of 20 percent can reduce the true effect threefold, while a proportion of 30 percent reduces it four- or fivefold. These approximations obviously only serve as notional examples, since they are based on particularly simple assumptions: the log-normality of the variables and multiplicative white noise errors. Moreover, other contradictory arguments could attenuate this underestimation of the effect of social origins on earnings (endogeneity).

In the case of the inequality of opportunity indices, the practice of considering averages or quantiles by type of social origin at least partially offsets these measurement errors. However, such a discussion calls for caution with regard to this theoretical scenario of perfect educational mobility, which has no close or even remote basis in historical fact, since intergenerational educational mobility varies little over the twenty years analyzed.[19]

2.7 Conclusion

This paper studies the impact of changes in educational opportunity on overall inequality and the inequality of opportunity on the labor market in Brazil over two decades. We use four editions of the nationally representative PNAD survey to analyze growth in earnings inequalities among 40–49-year-old men. We design and implement semiparametric decompositions of the respective effects of schooling expansion, changes in the structure of earnings, and changes in intergenerational educational mobility.

Earnings inequalities varied little over the period, with a peak in the late 1980s probably largely due to hyperinflation, which raged through to 1994 (a four-figure rate). First of all, the decompositions show that changes in the distribution of education contributed to the increase in both types of inequality among the oldest generations before sharply reducing them among the post-WWII cohorts. Second, the decrease in returns to education also contributed to equalizing labor market opportunities in the 1988–1996 period. Lastly, the changes in educational mobility were not large enough to significantly affect earnings inequalities, whereas it is shown that they should play a prominent role in equalizing opportunities in the future.

Brazil's history, at least during the macroeconomic crisis and adjustment period analyzed here, is one of steadily high income inequality. This rigidity of inequality is observed despite the expansion of education and despite the drop in returns to education, as already observed by Lam in 1991 and by Ferreira and de Barros for household income (2000 and 2004). Among the generations born before World War II, growth in education mainly concerned the spread of access to secondary and higher education for the children of the upper classes, which increased the inequality, as already noticed by Fishlow (1972). It was only with the postwar generations that the expansion of primary education and the opening of the secondary system to children of farmers and of fathers with very little education started to play a major role in the reduction of earnings inequalities. The decrease in returns to education underpinned this reduction during the period of slow growth recovery from 1988 to 1996 (marked by the Cardoso presidency and the real plan).

This last period of education-related reduction in earnings inequality could give rise to optimism as to the long-run effects of programs to educate poor children, such as conditional cash-transfer programs. The period also saw a slight upturn in intergenerational educational mobility, but this increase was too small to play a significant role in reducing the inequality of opportunity and overall inequality. The expansion of education prompted a race for qualifications and a quality race, both of which probably contributed to the decrease in returns to years of education. It will probably not be possible to attain a greater reduction in inequality *via education* in the future without a marked increase in intergenerational educational mobility. Yet it is still too soon to know whether targeted educational programs will manage to significantly stimulate this mobility.[20]

Acknowledgements

The authors would like to thank Francisco Ferreira, Michael Grimm, Marc Gurgand, Stephan Klasen, Sylvie Lambert, and Petra Todd, as well as two anonymous referees, and participants at an Education Day at INED in Paris, at the AFSE development economics meeting at CERDI in Clermont-Ferrand, at the Ibero-America Conference in Göttingen, and at the first ECINEQ conference in Palma de Mallorca. The views expressed in this paper are those of the authors alone.

Notes

1. Only the rural areas of Tocantins State were covered in this region.

2. For 1976, this information was collected solely for a subsample representing approximately 25 percent of the total sample.

3. In 1976, the question concerned the father's education when the individual was 15 years old.

4. This information was not collected by the 1982 PNAD.

5. The information on earned income is collected by a single question covering both wage and nonwage activities.

6. We thank Pierre-Emmanuel Couralet for his help in building the databases.

7. Since the 1990s, the first two levels of the Brazilian education system have been the elementary level (equivalent to primary education), lasting for eight years and normally covering children aged 7 to 14, and the intermediate level (equivalent to secondary education), lasting for three years and normally covering children aged 15 to 17. However, when the cohorts studied in this paper were educated, a basic level also existed covering the first four years of the elementary level.

8. We computed the hourly earnings means for 26 birth regions and grouped regions into four categories according to earnings differentials and geographical homogeneity. We also tried to preserve a balance in sample sizes. The highest levels of wages are observed in category 1 and the lowest in category 2.

9. A second stratification at the level of the municipalities of the metropolitan strata, the main municipalities, and grouping of municipalities of the other strata cannot be taken into account since the data do not enable these strata to be identified.

10. In the case when the number of types to be considered is too large, we implement this type of measurement by estimating decile regressions of earnings (Koenker and Bassett 1978), using dummy variables for the different types of social origin. This means we assume that the effects of the origin variables are additive. This assumption enables us to estimate a decile level for a large number of types (128) when considering the four social origin variables (see section 2.2). In this latter case, direct nonparametric estimates are in effect impossible due to sample size limitations.

11. Here again we use an intermediate regression step when considering 128 types of origin. We estimate an OLS earnings regression with the dummy variables for social origins as explanatory variables. The predictions resulting from this regression are the average earnings conditional on the different categories of origin. When the nine-category origin variable is used so that sample sizes are large enough, we estimate the means directly in a nonparametric way.

12. These minimum earnings are not normed by the average. The growth presented therefore includes an absolute component (growth in welfare) and a relative component (Rawlsian inequality index). The growth in average earnings is nevertheless virtually zero throughout the entire period.

13. Earnings inequality is slightly underestimated by the exclusion of unemployed and inactive men from the sample. This bias increases with unemployment and inactivity in 1996, but the decrease in inequality remains significant: the Theil index in 1988 reaches 0.83 when including null wages against 0.77 with strictly positive wages; in 1996 the Theil index is 0.80 against 0.70. Regarding inequality of opportunity, the sample selection seems completely innocuous. The Van de Gaer Theil index, with nine groups of social origins, is underestimated by 0.001—that is, by less than 1 percent. This very small reduction mainly comes from the higher employment rates of men whose fathers were working in agriculture.

14. For 2×2 transition matrices, there is a strict equality between the unique γ coefficient and the unique odd-ratio. For transition matrices of a higher dimension (like 9×9 here), equation (2.3) implies

$$\begin{aligned}\mathrm{Ln}[\text{Odd-Rt}(s, o; s', o')] = {} & [\gamma(s, o) + \gamma t(s, o) + \gamma(s, o) + \gamma t(s, o)] \\ & - [\gamma(s', o) + \gamma t(s', o) + \gamma(s, o') + \gamma t(s, o')].\end{aligned}$$

15. Bourguignon, Ferreira, and Menendez (2003) address the question of the endogeneity of education levels by making simulations under a number of assumptions of correlation between education level and wage unobservables. The origin variables are nevertheless assumed to be exogenous, an assumption that is also open to debate.

16. In contrast with section 2.4, we do not stack the contingency tables of different years; log-linear models are then written and estimated independently for each year.

17. The decomposition of the Gini indices can be found in the working paper version of this chapter (Cogneau and Gignoux 2005). They do not differ much from the decompositions based on the Theil index.

18. These reductions are found to be smaller in certain cases due to the existence of several empty cells, reducing the education level value taken into account in the construction of the notional matrices of perfect mobility, and hence the extent of the redistribution between education levels.

19. In the case of the historical decompositions of section 2.4, the main factor likely to confound the estimates is a variation in the variance proportion of these errors, due to a change in survey quality or methodology. Yet the effect of constant measurement errors is largely eradicated by the consideration of time differences. In addition, all of these decompositions remain influenced by the measurement errors associated with the analysis variables (level of education and social origin) and by the selection and endogeneity biases affecting the causal effect of these variables on earnings.

20. A recent paper from Ferreira, Leite, and Litchfield (2006) reveals a significant fall in household income inequality between 1993 and 2004, which they associate with five factors: declining inflation, sharp declines in the returns to education, pronounced rural-urban convergence, increases in the social transfers targeted to the poor, and a possible decline in racial inequality.

References

Andrade, E., S. Ferreira, R. Madalozzo, and F. Veloso. 2003. "Do Borrowing Constraints Decrease Intergenerational Mobility in Brazil? A Test Using Quantile Regression." Working Paper, IBMEC.

Arias, O., G. Yamada, and L. Tejerina. 2002. "Education, Family Background and Racial Earnings Inequality in Brazil." Working Paper, Inter-American Development Bank, Washington, D.C.

Bertail, P., and P. Combris. 1997. "Bootstrap Généralisé d'un Sondage." *Annales d'Economie et de Statistiques* 46: 49–83.

Bishop, Y., S. Fienberg, and P. Holland. 1975. *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MA: MIT Press.

Blinder, A. S. 1973. "Wage Discrimination: Reduced Form and Structural Estimates." *Journal of Human Resources* 8, no. 4: 436–455.

Bound, J., C. Brown, and N. Mathiowetz. 2001. "Chapter 59: Measurement Error in Survey Data." In *Handbook of Econometrics*, vol. 5, eds. J. J. Heckman and E. Leamer, 3705–3843. Amsterdam: North-Holland.

Bourguignon, F., F. H. G. Ferreira, and P. G. Leite. 2003. "Conditional Cash Transfers, Schooling, and Child Labor: Micro-Simulating Brazil's Bolsa Escola Program." *World Bank Economic Review* 17, no. 2: 229–254.

Bourguignon, F., F. H. G. Ferreira, and M. Menendez. 2007. "Inequality of Opportunity in Brazil", *Review of Income and Wealth* 53, no. 4: 585–618.

Cogneau, D., and J. Gignoux. 2005. "Earnings Inequalities and Educational Mobility in Brazil over Two Decades." Working Paper 2005/03, DIAL, Paris, and Discussion Paper No. 121, Ibero-America Institute for Economic Research, University of Göttingen. Available at http://www.dial.prd.fr/dial_publications/PDF/Doc_travail/2005–03.pdf or http://wiwi.uni-goettingen.de/vwlseminar/working_papers/ibero/DB121.pdf.

Di Nardo, J., N. Fortin, and T. Lemieux. 1996. "Labor Market Institutions and the Distribution of Wages, 1973–1992: A Semiparametric Approach." *Econometrica* 64, no. 5: 1001–1044.

Dunn, C. E. 2007. "The Intergenerational Transmission of Lifetime Earnings: Evidence from Brazil", *The B.E. Journal of Economic Analysis & Policy* 7, no. 2 (Contributions), Article 2.

Ferreira, F. H. G., P. Lanjouw, and M. Neri. 2003. "A Robust Poverty Profile for Brazil Using Multiple Data Sources." *Revista Brasileira de Economia* 57, no. 1: 60–92.

Ferreira, F. H. G., and R. Paes de Barros. 2000. "Education and Income Distribution in Urban Brazil, 1976–1996." *CEPAL Review* 71: 41–61.

Ferreira, F. H. G., and R. Paes de Barros. 2004. "The Slippery Slope: Explaining the Increase in Extreme Poverty in Urban Brazil, 1976–96." In *The Microeconomics of Income Distribution Dynamics in East Asia and Latin America*, eds. F. Bourguignon, F. H. G. Ferreira, and N. Lustig, 83–124. Washington, D.C. and New York: World Bank and Oxford University Press.

Ferreira, F. H. G., P. G. Leite, and J. A. Litchfield. 2006. "The Rise and Fall of Brazilian Inequality: 1981–2004." Policy Research Working Paper 3867, World Bank, Washington, D.C.

Ferreira, S., and F. Veloso. 2006. "Intergenerational Mobility of Wages in Brazil." *Brazilian Review of Econometrics* 26, no. 2: 181–211.

Fishlow, A. 1972. "Brazilian Size Distribution of Income." *American Economic Review* 62, no. 1–2: 391–402.

Juhn, C., K. M. Murphy, and B. Pierce. 1993. "Wage Inequality and the Rise in Returns to Skill." *Journal of Political Economy* 101, no. 3: 410–442.

Koenker, R., and G. Bassett. 1978. "Regression Quantiles." *Econometrica* 46, no. 1: 33–50.

Lam, D. 1999. "Generating Extreme Inequality: Schooling, Earnings, and Intergenerational Transmission of Human Capital in South Africa and Brazil." Research Report, Population Studies Center, University of Michigan.

Lam, D., and D. Levison. 1991. "Declining Inequality in Schooling in Brazil and its Effects on Inequality in Earnings." *Journal of Development Economics* 37, no. 1/2: 199–225.

Lam, D., and R. Schoeni. 1993. "Effects of Family Background on Earnings and Returns to Schooling: Evidence from Brazil." *Journal of Political Economy* 101, no. 4: 710–740.

Leite, P. G. 2006. "L'efficacité de Bolsa Escola par la Méthode RDD." Mimeo, EHESS and DIAL.

Lemieux, T. 2002. "Decomposing Changes in Wage Distributions: A Unified Approach." *Canadian Journal of Economics* 35, no. 4: 646–688.

Manski, C., and J. Pepper. 2000. "Monotone Instrumental Variables: With an Application to the Returns to Schooling." *Econometrica* 68, no. 4: 997–1010.

Oaxaca, R. 1973. "Male-Female Wage Differentials in Urban Labor Markets." *International Economic Review* 14, no. 3: 673–709.

Pastore, J. 1982. *Inequality and Social Mobility in Brazil*. Chicago: University of Wisconsin Press.

Pastore, J., and N. Valle Silva. 2000. "Mobilidade Social no Brasil." Sao Paulo: Makron Books.

Pesquisa National por Amostra de Domicílios (PNAD) 1976–1996. "Notas methodologicas", *Instituto Brasileiro de Geografia e Estatística*, Rio de Janeiro. For more information: http://www.ibge.gov.br/home/estatistica/populacao/trabalhoerendimento/pnad2006/default.shtm

Picanço, F. 2004. "Economic Modernization and Socio-Occupational Mobility in Brazil." Communication presented at the International Sociological Association, RC28.

Roemer, J. 1996. *Theories of Distributive Justice.* Cambridge, MA: Harvard University Press.

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

Shorrocks, A. 1978. "The Measurement of Mobility." *Econometrica* 46, no. 5: 1013–1024.

Sokoloff, K., and S. Engerman. 2000. "History Lessons: Institutions, Factor Endowments, and Paths of Development in the New World." *Journal of Economic Perspectives* 14, no. 3: 217–232.

Van de Gaer, D. 1993. "Equality of opportunity and investment in human capital", *Catholic University of Leuven*, Faculty of Economics, no. 92.

Van de Gaer, D., E. Schokkaert, and M. Martinez. 2001. "Three Meanings of Intergenerational Mobility." *Economica* 68: 519–537.