DOCUMENT DE TRAVAIL DT/2005-03

Earnings inequalities and educational mobility in Brazil over two decades

Denis COGNEAU
Jérémie GIGNOUX

DIAL • 4, rue d'Enghien • 75010 Paris • Téléphone (33) 01 53 24 14 50 • Fax (33) 01 53 24 14 51
E-mail : dial@dial.prd.fr • Site : www.dial.prd.fr

EARNINGS INEQUALITIES AND EDUCATIONAL MOBILITY IN BRAZIL OVER TWO DECADES[1]

Denis Cogneau
IRD-Paris, DIAL
cogneau@dial.prd.fr

Jérémie Gignoux
INED, IEP-Paris, Université Paris-IX Dauphine, LEA/INRA
jeremie.gignoux@ined.fr

Document de travail DIAL
Janvier 2005

ABSTRACT

This paper studies the impact of changes in educational opportunities on various definitions of labour market inequalities in Brazil over two decades (1976-96). Using four editions of the nationally representative PNAD survey, we analyze the evolution of overall inequalities and inequalities of opportunity in 40-49 year old males' earnings. We design and implement semiparametric decompositions of the respective effects of (i) schooling expansion, (ii) changes in the structure of earnings, and (iii) changes in intergenerational educational mobility. Earnings inequalities varied little over the period, with a peak in the late 1980s that can be imputed to hyperinflation. First of all, the decompositions show that changes in the distribution of education contributed to the increase in both overall earnings inequalities and inequalities of opportunity among the oldest generations, before sharply reducing them among the post-WWII cohorts. Secondly, the decrease in returns to education also contributed to equalizing labour market opportunities in the 1988-96 period. Thirdly and 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.

Key words: Equality of opportunities, Labour market, Inequality decomposition, Brazil

JEL Code: D32, D63, J62, O15

RESUME

Ce papier étudie les conséquences des changements dans les opportunités scolaires sur plusieurs définitions des inégalités sur le marché du travail au Brésil sur deux décennies. En utilisant quatre éditions de l'enquête nationale représentative PNAD, nous analysons l'évolution des inégalités globales et de l'inégalité des chances de rémunérations des hommes de 40 à 49 ans. Nous construisons et mettons en œuvre des décompositions semi-paramétriques des effets respectifs de (i) l'expansion de la scolarisation (ii) les changements dans la structure des rémunérations, et (iii) les changements dans la mobilité scolaire intergénérationnelles. Les inégalités de rémunération ont peu varié sur l'ensemble de la période, avec un pic à la fin des années 1980 attribuable à l'hyperinflation. Premièrement, les décompositions montrent que les changements dans la distribution de l'éducation ont contribué à l'accroissement des inégalités de rémunération globales et de l'inégalité des chances dans les générations les plus anciennes, avant de les réduire fortement parmi les cohortes nées dans l'après-guerre. Deuxièmement, la baisse des rendements de l'éducation a aussi contribué à égaliser les opportunités sur le marché du travail pendant la période 1988-96. Troisièmement et enfin, les évolutions de la mobilité scolaire n'ont pas été suffisamment importantes pour affecter significativement les inégalités de rémunération, alors qu'il est montré qu'elles devraient jouer un rôle primordial pour l'égalisation des opportunités dans le futur.

Mots clés : Egalité des chances, Marché du travail, Décomposition des inégalités, Brésil

[1] The authors would like to thank Marc Gurgand and participants at an Education Day at the French National Institute for Demographic Studies (INED). The views expressed in this paper are those of the authors alone.

Contents

List of tables

List of figures

List of appendices

1. INTRODUCTION

Brazilian society can be regarded as one of the most inegalitarian in the world from a number of points of view and special attention is consequently paid to it. The South American and Caribbean societies are particularly inegalitarian. This characteristic has now been related to the institutions left over from the colonial period (Engerman and Sokoloff, 2000). The level of inequalities in Brazil is much greater even than the average on the sub-continent with, for example, a Gini index one-third higher than Argentina (UN/WIDER data), and is 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 labour market (official/unofficial). Brazil also has among the lowest intergenerational educational mobility and equality of social and economic opportunities in the world (Dunn, 2003).

A series of nationally representative annual surveys based on large samples (PNAD, *Pesquisa Nacional por Amostra de Domicilios*) provides a fairly accurate observation of the change in inequalities in Brazil over nearly thirty years. These data show that income inequalities remained remarkably stable, whether gaps 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, 2003). Economists have also recently looked into the impact of family origins and inequalities of educational and labour market opportunities (Lam and Schoeni, 1993; Arias, Yamada and Tejerina, 2002; Bourguignon, Ferreira and Menendez, 2003; Andrade, Ferreira, Madalozzo and Veloso, 2003; Dunn, 2003; Ferreira and Veloso, 2003). The main question put in these papers is the contribution of education to the reduction of economic inequalities.

This analytic question ties in with a contemporary political issue, since Brazil set up extensive means-based transfer programmes in 1999 conditional on sending children to school (*Bolsa Escola*) and stopping child labour (PETI). These programmes have now been combined into a single programme called Bolsa Familia and are reaching cruising speed with widespread coverage. However, it is not easy to evaluate the impact of these programmes since, unlike the Mexican *Progresa* programme, no randomly allocated pilot set-up has been implemented. An *ex-ante* evaluation of the *Bolsa Escola* programme using a structural micro-econometric model finds that the transfers have a significantly positive, albeit modest, impact on school enrolment and child labour. The sums of transfers distributed are moreover lower than in the *Progresa* programme. Hence they only have a marginal impact on income inequalities and poverty (Bourguignon, Ferreira and Leite, 2004). An ex-post evaluation of the programme is underway using data from the PNAD surveys, which identify the recipient households (Leite, 2005).

Whatever the impact of these programmes on the education of the most underprivileged children, a second question arises as to the long-run impact of a decrease in educational inequalities on the distribution of income in Brazil. As regards the reduction of income inequalities, 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 (2003) applies microsimulation techniques to the 1996 PNAD survey to analyze the contribution of inequalities of educational and income opportunities to the formation of inequalities in an urban environment. It concludes that the cancelling out of inequalities due to social origin variables (race, region of birth, and parental education and occupation) would reduce the Gini index of individual incomes by just over 15% (10 points) and the Theil index by approximately 30%. The impact on household incomes would be slightly greater (15 Gini points), given the impacts on female labour force participation and fertility. A large part of the impact of factors of origin on individual earnings is associated with paternal education. The microsimulations also show that a minimal six years of education would produce more or less similar effects on the following generation. Lastly, half of this impact can be imputed to the direct effect of factors of origin on earnings while the other half corresponds to the indirect effect of education, i.e. the equalization of schooling opportunities. 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.

This paper addresses the same type of question using a different set of data and other econometric methods (semiparametric decompositions). We use four PNAD surveys in 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 these inequalities (Bourguignon *et al.* conduct static microsimulations by cohorts on the 1996 survey). For the first time in the case of Brazil, we construct and calculate inequality of opportunity indicators in keeping with the axiomatics proposed by Roemer (1996 and 1998) and Van De Gaer *et al.* (2000). We also look at the impact of educational mobility on the income distribution, i.e. the Bourguignon *et al.* 'indirect' effect, which is the variable for *Bolsa Escola* programme action. Lastly, as an alternative to the microsimulation techniques, we propose a semiparametric estimation of the respective effects of schooling expansion, educational mobility and the structure of earnings on changes in individual earnings inequalities, covering as much overall inequalities as inequalities of opportunity. We use decompositions of educational mobility furnished by the log-linear model and nonparametric reweighting techniques inspired by Di Nardo, Fortin and Lemieux (1996).

The two types of earnings inequalities for men aged 40 to 49 (1927-1956 cohorts) displayed a similar growth path including a peak in the late 1980s with the end of the dictatorship (1985) and the height of the inflationary crisis (Cruzado, Bresser, Summer and Collor plans; see also Appendix 4). All things considered, overall inequalities rose slightly from the beginning to the end of the period, while inequalities of opportunity posted a slight drop. Intergenerational educational mobility also recorded a very slight upturn.

As regards the equality of economic opportunities, changes in the distribution of education levels initially had an inegalitarian effect before becoming equalizing as of the late 1980s. The first sub-period analyzed shows a rise in secondary and higher education immediately following the war, i.e. for the generations educated from 1945 to 1965, which benefited mainly the children of the upper classes. The second sub-period corresponds to the generations educated from 1955 to 1975. This period was marked by the expansion of primary schooling, which benefited more the children of the underprivileged classes. 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.

As regards the overall earnings inequalities, the expansion of education likewise initially had an inegalitarian effect before becoming equalizing at the end of the period. 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 (see also Bittencourt, 2005). 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.

Lastly, 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 inequalities of economic opportunity at a high level. We confirm that the majority of the inequalities of opportunity on the labour market can be imputed to this factor. However, as found by Bourguignon *et al.* (2003), this factor only plays a modest role in overall inequalities. We nevertheless put forward that this last evaluation is highly sensitive to earnings measurement errors.

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)[2].

These four editions contain information on the adults' social origins, collected for the head of household and his spouse[3]. This concerns the father's level of education and occupation when the individual started working[4]. In addition, a question on migration provides information on the individual's place of birth (Federative Republic State)[5] and the questionnaire on demographic characteristics provides information on the individual's colour. 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% of this age bracket depending on the edition) and to employed individuals (93% to 89%, 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 non-wage combined[6], and on the weekly hours worked in these activities[7]. The incomes are discounted to September 2002 Brazilian Reals using the IBGE deflators derived from the INPC national consumer price index. Ferreira *et al.* (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 the incomes are underestimated in the 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 *et al.* (2004). Analyzing intergenerational mobility based on an urban sub-sample 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 *et al.* (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 numbers of years as of entry into primary school, which is normally at seven years old[8]. 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: a colour 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 variable coded into four categories (covering respectively the individuals born in a) the Federal District and the State of Sao Paulo, b) the states of the south, centre-west and west of the Northern region, c) the Rio Grande do Sul, the states of the south-east and south of the

[2] Only the rural areas of Tocantins State were covered in this region.
[3] For 1976, this information was collected solely for a sub-sample representing approximately 25% of the total sample.
[4] In 1976, the question concerned the father's education when the individual was 15 years old.
[5] This information was not collected by the 1982 PNAD.
[6] The information on earned income is collected by a single question covering both wage and non-wage activities.
[7] We thank Pierre-Emmanuel Couralet for his help in building the databases.
[8] 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.

Northeastern region, and d) the states of the north of the Northeastern region and east of the Northern region); a father's level of education variable coded into four categories (covering respectively the individuals whose father a) never went to school, b) is literate or for whom the interviewee is unable to give an answer, c) completed one of the first four years of primary education, and d) completed at least the fifth year of primary education); and a father's occupation variable coded into four categories (covering respectively the individuals whose father was a) a farmer, b) employed in a traditional industry, a domestic employee, whose occupation is poorly defined or for whom the interviewee is unable to give an answer, c) employed in a modern industry, an unincorporated entrepreneur or employed in a service sector, and d) 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 respectively the individuals whose father a) never went to school and was a farmer, b) never went to school and had another occupation, c) was merely literate and was a farmer, d) was literate and had another occupation, e) completed one of the first four years of primary education and was a farmer, f) completed one of the first four years of primary education and had another occupation, g) completed one of the four years of upper primary education (5-8), h) completed nine or more years of education, and i) 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 (Nascimento Silva *et al.*, 2002), i.e. the stratification of the sample into 36 "natural" strata corresponding to 27 Brazilian states and nine metropolitan regions (Bertail and Combris, 1997)[9].

3. GROWTH IN WAGE INEQUALITIES OVER THE 1976-1996 PERIOD

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

3.1. Overall earnings inequalities

Table 1 presents the growth in overall inequalities in the distribution of hourly earnings as measured by the Gini and Theil indices.

Table 1: Measurements of overall inequalities 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.
(*): Variation significant at 5% compared with the previous year
(In brackets): bootstrap standard deviations.

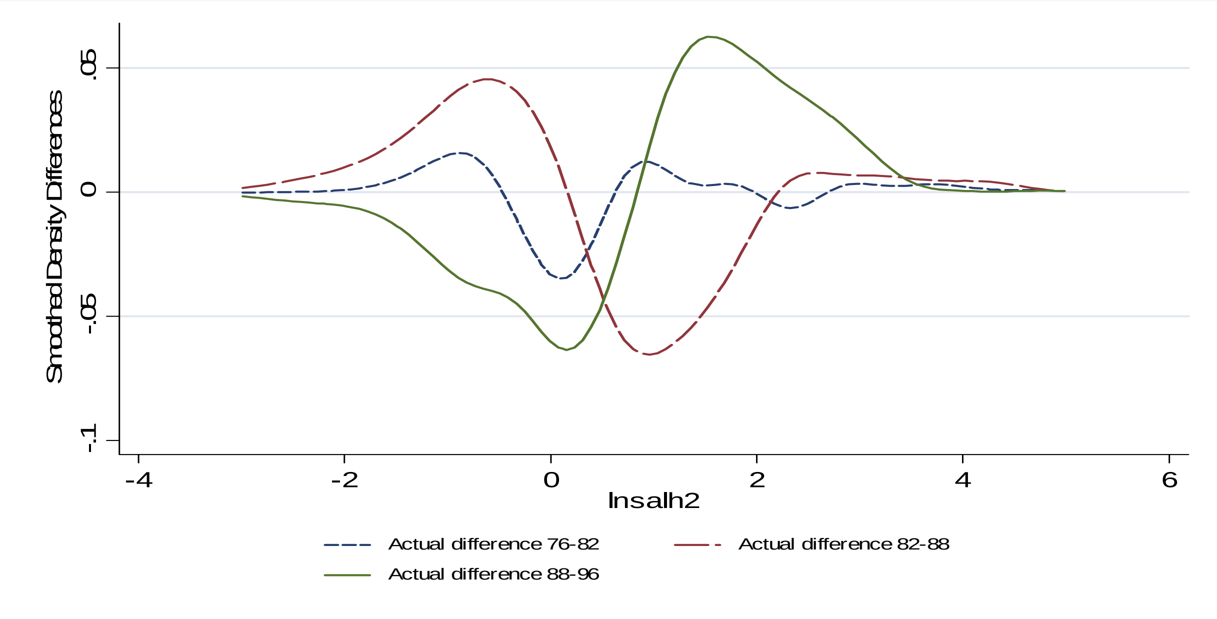
The Gini index remains close to 0.60 for the entire period. It increases significantly from 1976 to 1988, then falls from 1988 to 1996 before returning to a level slightly above, but not significantly different to, its 1976 level. The Theil index displays similar growth. It rises more sharply than the 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 inequalities changed little over these twenty years, but that inequalities rose to the detriment of the bottom of the earnings distribution. These trends are illustrated by Figure 1, which presents the smoothed density differences in hourly earnings from 1976 to 1996.

[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.

3.2. The inequality of labour market opportunities

We construct the inequality of labour market opportunity indices in keeping with the two main economic literature proposals on economic justice and equality of opportunities (Roemer 1996 and 1998; Van de Gaer, 2000). For a given outcome variable (here hourly earnings), both proposals distinguish between what is due to *"circumstances"*, defined as an individual 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 inequalities.

Figure 1: Variations in hourly earnings densities



Method: Double smoothing by a Gaussian kernel function (bandwidth 0.2)

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 inequalities between types are then measured by comparing individuals with the same relative level of effort. The inequality 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. We implement this type of measurement by estimating decile regressions of earnings using dummy variables for the different types of social origin. For this, we assume that the effects of the origin variables are additive[10]. The predictions made based on these estimates at the different deciles measure the hourly earnings deciles conditional on the types of origin. We hence reconstitute the earnings distributions predicted by the origin variables at each relative level of effort. We calculate the inequality indices at each decile and aggregate them taking their average[11]. These "Roemer" indices are written:

$$\mathrm{ROE} = 1/10 \, . \sum_{\pi} \mathrm{I} \, \{ \, w_{o,\pi} \, \} \tag{1}$$

where o is an index for the different types of social origins, $w_{o,\pi}$ is the predicted wage at decile π for type o, and I is an index of inequality (here Gini or Theil) between the different types o.

[10] The parametric assumptions about the additivity of the decile regressions enable us to estimate a decile level for a large number of types when considering the four social origin variables (see Section 2). In this latter case, direct nonparametric estimates are in effect impossible.

[11] This average can be weighted to place more weight on the inequality of opportunity at the bottom or top of the distribution of relative efforts.

The second approach proposed by Van de Gaer *et al.* 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. Very simply, we can choose the Gini of average earnings by category of origin. We implement this measurement by estimating an earnings regression based on the dummy variables for social origins. The predictions resulting from this regression are the average earnings conditional on the different categories of origin[12]. The "Van de Gaer" inequality of opportunity indices are obtained from this distribution of average earnings estimated by categories of origin. These indices are written:

$$VdG = I \{ E(w|o) \} \quad (2)$$

where I is an inequality index and E(w|o) is the wage expectation conditional on social origin o.

We therefore calculate two series of inequality of opportunity indices and present the Gini and Theil indices. We use the two social origin characterizations comprising respectively 128 and nine categories of origins. The results are presented in tables 2 and 3.

Table 2: Measurements of the inequality of economic opportunities (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 colour (2); not available in 1982; (*) indicates significance at 5% compared with the previous year; (in brackets): bootstrap standard deviations, 100 replications.

Table 3: Measurements of the inequality of economic opportunities (9 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; (*) indicates significance at 5% compared with the previous year; (in brackets): bootstrap standard deviations, 100 replications.

We can first of all observe that the indices based on nine types of origin (Table 3) underestimate the inequality of opportunities by 10% to 20% compared with the indices based on 128 types. The Gini

12 See note 9. Here again, we only use this intermediate regression step when considering 128 types of origin. When the nine-category origin variable is used, we estimate the means directly in a nonparametric way.

indices measured are situated between 0.30 and 0.40. Note that the non-decomposable nature of this index makes it impossible to use to deduce a measurement of the proportion of inequalities of opportunity in overall inequalities. The Theil indices measured are situated between 0.20 and 0.30. In this case, the decomposability of the Theil index means that the contribution of social origins to overall inequalities can be estimated at nearly 30%. These findings can be directly compared with those of Bourguignon *et al.* (2003), who attribute 15% of the overall inequalities to social origins in the case of the Gini index and 30%, as here, in the case of the Theil index. Moreover, as argued by Van de Gaer *et al.* (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, i.e. when the most underprivileged types of origin in each decile are the same. The findings confirm in general this property of monotonicity.

The indices display similar growth to the overall inequality indices. All the indices find that the inequalities of opportunity rise from 1976 to 1988, and that this rise is generally significant at least at 10% (and at 5% for the Gini index when using the Roemer approach). All the indices subsequently post a decrease in inequalities of opportunity, and this drop is also significant (at 5% for all the indices). We also present the growth in minimum earnings for the different categories of origin at each decile of the earnings distribution[13]. This measurement corresponds to Roemer's first proposal to define the equal opportunities policies. This indicator grows in parallel with the Gini and Theil 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 opportunities fell slightly from 1982 to 1996.

4. INTERGENERATIONAL EDUCATIONAL MOBILITY

In this section, we leave aside the inequality of earnings opportunities to concentrate on the inequality of educational opportunities, 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 opportunities: 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 the second section.

Table 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 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 intermediate cohorts born during World War II post both an upturn in school attendance and continued sharp growth in the secondary and higher education sectors (more than eight years of education). It is only in the last cohorts born after the war that primary education shows marked growth (from five to eight years of education).

[13] 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.

Table 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
	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.

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 probabilities of access to upper primary education (five to eight years) for less privileged children (see the destination matrices in the appendix). For the post-war 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 opportunities. 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,…9; our social origin variable O is also divided into nine categories indexed by o=1,…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 saturated log-linear model is written:

$$Ln[n_t(s,o)]= \mu + \alpha(s) + \beta(o) + \gamma(s,o) + \mu_t + \alpha_t(s) + \beta_t(o) + \gamma_t(s,o) \qquad (3)$$

Where $n_t(s,o)$ is the value of the table cross-tabulating S and O at year t.

This decomposition is unique under the constraints:

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

For all s and o:

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

And for all t, s and o:

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

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

$$\text{Odd-R}_t(s,o;\ s',o') = [n_t(s,o)\ n'_t(s',o')] / [n_t(s',o)\ n'_t(s,o')] \qquad (4)$$

This model is estimated by maximum likelihood. The joint test of the assumption [$\gamma_t(s,o)=0$ for all (s,o) and t] is therefore written as a likelihood ratio test in keeping with a law of χ^2. It is used to evaluate the existence of a change in "non-structural" 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 5 presents some odds ratios, called reproduction coefficients here since the fathers and sons' categories are the same, calculated for each of the four years. 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.

Table 5: Educational mobility reproduction coefficients

	1976	1982	1988	1996
Year of birth	1927-36	1933-42	1939-48	1947-56
Schooled/unschooled	6.29	8.70*	10.80	7.40*°
	(0.81)	(0.22)	(0.85)	(0.23)
5 years or +/less than 5	24.66	28.36	23.72	22.63
	(25.67)	(7.78)	(8.42)	(7.82)
5 years or +/1-4 years	7.66	10.41	7.82	11.28
	(2.76)	(1.20)	(1.03)	(2.20)
1-4 years/unschooled	2.58	3.63*	3.60	2.86*°
	(0.18)	(0.05)	(0.12)	(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 changing them.
*: For 1996, the odds ratio is significantly different (and lower) than in 1982, at the 5% 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% level.

The expansion of education has moreover given rise to a race for qualifications shifting the educational hierarchy upwards, and also probably a quality race (private system vs. state system), both producing an apparent drop in returns (Lam, 1991).

5. THE EFFECTS OF EDUCATIONAL CHANGES ON EARNINGS INEQUALITIES FROM 1976 TO 1996

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, i.e. individuals' education S and social origin O.

In common with Di Nardo *et al.* and the majority of papers (Bourguignon *et al.* being one other example), we look at Oaxaca decompositions. In other words, our decompositions reconstitute counterfactual income distributions by applying counterfactual population structures to an observed earnings structure. These decompositions consist of calculating what the overall inequalities and inequalities of earnings opportunities 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 wages by education level and social origin had not changed. The changes in the distribution of the population between education levels and categories of social origins 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, i.e. 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 non-endogeneity of the origin and education level variables as regards the unobserved determinants of earnings: the conditional earnings densities[14]

[14] And especially the conditional expectations or deciles used for the inequality of opportunity indices.

(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].

5.1.1 Construction of the counterfactual inequalities

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 inequalities, the basic idea consists of reweighting the observed distribution of earnings y. The observed income density is written:

$$f_t(y) = \int f(y|s,o,t_y=t)\, dF(s,o|t_{s,o}=t) \quad (5)$$

And the counterfactual density:

$$f^*_t(y) = \int f(y|s,o,t_y=t)\, dF^*(s,o) = \int f(y|s,o,t_y=t)\, dF(s,o|t_{s,o}=t)\, \psi(s,o) \quad (6)$$

where $\psi(s,o) = dF^*(s,o) / dF(s,o|t_{s,o}=t)$ 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 EOp indices are functions of the conditional distribution of y vis-à-vis o and the distribution of origins in the population.

$$EOP_t = EOP\,[\, f(y|o,t_y=t)\,,\, dF(o|t_o=t)] \quad (7)$$

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

$$VdG_t = I\,[\, E(y|o,t_y=t)\,,\, dF(o|t_o=t)], \quad (8)$$

where I is a usual inequality index (Gini, Theil or other entropy indices) applied to the distribution of E(y|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|o) dF*(o). Hence:

$$VdG^*_t = I\,[\, E^*(y|o,t_y=t)\,,\, dF^*(o)], \quad (9)$$

and:

$$E^*(y|o,t_y=t) = \int E(y|o,s,t_y=t)\, dF^*(s|o)$$

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

[15] Bourguignon et al. (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.

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

$$\mathrm{Ln}[p_t(s,o)] = -\mathrm{Ln}(N_t) + \mu_t + \alpha_t(s) + \beta_t(o) + \gamma_t(s,o) \tag{10}$$

where N_t is the total number of individuals in the sample, μ is a constant, $\alpha(s)$ the effect of the margins of s, $\beta(o)$ the effect of the margins of o and $\gamma(s,o)$ the effect of the interactions between o and s. This decomposition is unique under the constraints: $\Sigma_s\ \alpha(s)=0$; $\Sigma_o\ \beta(o)=0$; $\Sigma_s\ \gamma(s,o)=0$; $\Sigma_o\ \gamma(s,o)=0$, for all s and o. If S and O are independent of one another, then coefficients $\gamma(s,o)$ are zero ($p_t(s,o)= p_t(s,.)p_t(.,o)$). Coefficients $\gamma_t(s,o)$ are directly linked to the odds ratios of the mobility table $p_t(s,o)$.

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 $t' \neq t$:

$$\mathrm{Ln}[p_t(s,o)] = -\mathrm{Ln}(N_t) + \mu_t^* + \alpha_t^*(s) + \beta_t^*(o) + \gamma_{t'}(s,o) \tag{11}$$

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 $p^*_{t/t'}(s,o)$;
- 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 $p^*_{t'/t}(s,o)$;
- A movement in the marginal distributions of education and origins from $p^*_{t'/t}(s,o)$ to $p_t(s,o)$.

Below 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 A and B.

Table A: Frequencies of the assumed distribution observed in t
[Marginal distributions (0.80; 0.20) and reproduction coefficient of 7]

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

Table B: Frequencies of the assumed distribution observed in t'
[Marginal distributions (0.60; 0.40) and reproduction coefficient of 2]

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

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 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 = C 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 A to Table C, and then the change in marginal distributions from Table C to Table B.

Table C: Frequencies of the distribution simulated with the marginal distributions of t and educational mobility of t'
[Marginal distributions (0.80; 0.20) and reproduction coefficient of 2]

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

A second decomposition of these developments can be made. Table D 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 A to Table D, and then the change in educational mobility from Table D to Table B.

Table D: Frequencies of the distribution simulated with the educational mobility of t
[Marginal distributions (0.60; 0.40) and reproduction coefficient of 7]

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

5.1.3 Semiparametric decomposition

This last methodological part sums up the construction of the semiparametric decompositions of changes in inequalities using the counterfactual educational mobility tables.

As regards overall inequalities 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 (6) and (11):

$$f^*_{t/t'}(y) = \int f(y|s,o,t_y=t)\, dF(s,o|t_{s,o}=t)\, \psi^*_{t/t'}(s,o) \quad (12)$$

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:

$$f^{**}_{t/t'}(y) = \int f(y|s,o,t_y=t)\, dF(s,o|t_{s,o}=t)\, \psi_{t/t'}(s,o) \quad (13)$$

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

The first counterfactual describes the movement of overall inequalities that can be attributed to the change in non-structural educational mobility, and the second the movement of inequalities 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.

It is also possible to start by altering the marginal distribution, given constant educational mobility, and then to alter the non-structural educational mobility. Decompositions of overall inequalities can also be made “backwards” 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 ($E(y|o,s,t_y=t)$) or that of t’ ($E(y|o,s,t_y=t')$). 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 above. We can apply the structure of earnings observed in t or that observed in t’ to each of the four tables:

	Observed t	First-stage simulation	Second-stage simulation	Observed t’
MDS	Table A, $E(y\|o,s,t_y=t)$	Table C, $E(y\|o,s,t_y=t)$	Table B, $E(y\|o,s,t_y=t)$	Table B, $E(y\|o,s,t_y=t')$
DMS	_	Table D, $E(y\|o,s,t_y=t)$	Table B, $E(y\|o,s,t_y=t)$	_
SMD	_	Table A, $E(y\|o,s,t_y=t')$	Table C, $E(y\|o,s,t_y=t')$	_
SDM	_	Table A, $E(y\|o,s,t_y=t')$	Table D, $E(y\|o,s,t_y=t')$	_
MSD	_	Table C, $E(y\|o,s,t_y=t)$	Table C, $E(y\|o,s,t_y=t')$	_
DSM	_	Table D, $E(y\|o,s,t_y=t)$	Table D, $E(y\|o,s,t_y=t')$	_

In practice, we do not consider the last two counterfactual paths that introduce an earnings structure change “in the middle of” the population structure change. Note that these paths do not have their symmetric in the decomposition of overall equalities since the conditional densities are not estimated.

5.1.4 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 (see the tables in Appendix A1). We have solved the problem from a technical point of view by allocating a very small value (0.5) to the few empty cells in the matrices 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 (see below). 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[16]. 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.

5.2. Findings: historical decomposition of growth in earnings inequalities from 1976 to 1996

Tables 6 to 8 present the results of the decompositions of the Van de Gaer inequality of opportunity indices and the overall inequality indices for the three sub-periods from 1976 to 1996: 1976-1982, 1982-1988 and 1988-1996. Four decompositions are presented according to the path taken (see below) for each index and sub-period. 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[17].

[16] These low values for the extreme categories also partially explain why the wage regressions rarely reveal significant interactions between level of education and category of origin.

[17] 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 opportunities, 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 4, Table 4), the first two sub-periods analyzed (1976-1982 and 1982-1988) show an increase in secondary and higher education from 1945 to 1965, benefiting mainly the children of privileged origins. The third sub-period (1988-1996) corresponds to the generations educated from 1955 to 1975. This sub-period was marked by an expansion in primary education benefiting more 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% had completed 12 years or more of education in 1976, 58% in 1982 and 64% in 1988, with this proportion peaking at 65% in 1996. Likewise, among the individuals whose father had completed five to eight years of education (upper primary), the probabilities of entering secondary education (nine or more years) rose from 50% to 55% and then 64% from 1976 to 1988, peaking at 68% in 1996. Conversely, among the men whose father was an uneducated farmer, 53% had in turn not attended school in 1976 and 1982, 46% in 1988 and 41% in 1996. Of these same men, only 3% had attained upper primary level (five years and more) in 1976, 5% in 1982, 7% in 1988, but 15% in 1996[18].

The "democratization" of access to school came about mainly for the generations born after the war, educated from 1955 to 1975 and 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 inequalities, the expansion of education likewise initially had an inegalitarian effect among the pre-war generations (1976-1982) before becoming equalizing for the post-war generations (1988-1996). Its impact is found to be negligible for the intermediate generations (1982-1988). Other factors, in particular the nose-dive in the minimum wage in real terms (-20%), 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 represents the impact of this expansion of primary schooling on the development of earnings densities in the last period. 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 wages for uneducated men whose father was uneducated to those of men with a secondary education whose father had reached the same level was 11.4 in 1976 and 10.5 in 1982, rising to 11.3 in 1988 following a fall of over 15% 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 equally with the expansion of primary education to the reduction in inequalities of opportunity.

[18] See also the tables in appendix A1.

Figure 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

Table 6: Historical decomposition 1976-1982

	Gini						Theil					
	M mobility		D distrib		S wages		M mobility		D distrib		S wages	
		s.e.		s.e.		s.e.		s.e.		s.e.		s.e.
<u>Inequalities of opportunity</u>												
MDS	-0.008	0.021	0.018	0.009	-0.001	0.013	-0.005	0.029	0.016	0.012	-0.003	0.020
DMS	-0.006	0.008	0.016	0.015	-0.001	0.013	-0.001	0.011	0.013	0.019	-0.003	0.020
SMD	0.007	0.021	0.025	0.009	-0.023	0.047	0.011	0.014	0.024	0.013	-0.027	0.047
SDM	0.005	0.012	0.027	0.018	-0.023	0.047	0.011	0.012	0.024	0.015	-0.027	0.047
Total	0.010	0.018					0.008	0.026				
Percentages:												
MDS	-81%		187%		-6%		-60%		201%		-41%	
DMS	-57%		164%		-6%		-17%		158%		-41%	
SMD	75%		258%		-234%		141%		301%		-341%	
SDM	55%		279%		-234%		139%		302%		-341%	
	M mobility		D distrib		R residual		M mobility		D distrib		R residual	
<u>Overall inequalities</u>												
MDR	-0.004	0.005	0.013	0.006	0.007	0.009	-0.014	0.016	0.019	0.016	0.058	0.031
DMR	-0.006	0.004	0.015	0.005	0.007	0.009	-0.019	0.013	0.024	0.013	0.058	0.031
SMR	-0.001	0.005	0.016	0.005	0.001	0.010	-0.006	0.021	0.025	0.017	0.044	0.035
RDM	-0.001	0.003	0.016	0.005	0.001	0.010	-0.004	0.011	0.022	0.012	0.044	0.035
Total	0.016	0.010					0.063	0.035				
Percentages:												
MDR	-22%		79%		43%		-22%		30%		93%	
DMR	-35%		91%		43%		-30%		38%		93%	
RMD	-6%		102%		4%		-10%		40%		70%	
RDM	-4%		100%		4%		-6%		36%		70%	

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 wages from 1976 to 1982; the simulation paths are noted by the order of changes, with M denoting educational mobility, D the marginal distributions 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 7: Historical decomposition 1982-1988

	Gini						Theil					
	M mobility		D distrib		S wages		M mobility		D distrib		S wages	
		s.e.		s.e.		s.e.		s.e.		s.e.		s.e.
Inequalities of opportunity												
MDS	-0.003	0.003	0.013	0.005	0.006	0.008	-0.002	0.004	0.011	0.006	0.011	0.011
DMS	-0.001	0.004	0.011	0.005	0.006	0.008	0.000	0.005	0.008	0.006	0.011	0.011
SMD	-0.004	0.003	0.014	0.005	0.006	0.007	-0.004	0.005	0.012	0.007	0.012	0.010
SDM	-0.002	0.004	0.012	0.005	0.006	0.007	-0.001	0.005	0.009	0.007	0.012	0.010
Total	0.016	0.011					0.020	0.016				
Percentages:												
MDS	-20%		83%		37%		-12%		56%		57%	
DMS	-5%		69%		37%		2%		41%		57%	
SMD	-26%		89%		37%		-19%		59%		60%	
SDM	-12%		75%		37%		-4%		44%		60%	
	M mobility		D distrib		R residual		M mobility		D distrib		R residual	
Overall inequalities												
MDR	0.000	0.001	0.008	0.003	0.031	0.006	0.001	0.003	0.000	0.007	0.082	0.025
DMR	0.000	0.001	0.008	0.003	0.031	0.006	0.002	0.003	0.000	0.007	0.082	0.025
SMR	0.000	0.001	0.007	0.002	0.032	0.006	0.001	0.003	-0.001	0.006	0.084	0.026
RDM	0.000	0.001	0.007	0.002	0.032	0.006	0.002	0.003	-0.002	0.007	0.084	0.026
Total	0.038	0.007					0.084	0.025				
Percentages:												
MDR	0%		21%		80%		2%		0%		98%	
DMR	0%		20%		80%		2%		0%		98%	
RMD	0%		18%		82%		1%		-1%		100%	
RDM	1%		17%		82%		2%		-2%		100%	

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 wages from 1982 to 1988; the simulation paths are noted by the order of changes, with M denoting educational mobility, D the marginal distributions 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 8: Historical decomposition1988-1996

	Gini						Theil					
	M mobility		D distrib		S wages		M mobility		D distrib		S wages	
		s.e.		s.e.		s.e.		s.e.		s.e.		s.e.
Inequalities of opportunity												
MDS	-0.006	0.004	-0.021	0.004	-0.021	0.010	-0.003	0.006	-0.041	0.005	-0.023	0.012
DMS	-0.004	0.004	-0.022	0.004	-0.021	0.010	-0.003	0.006	-0.041	0.005	-0.023	0.012
SMD	-0.004	0.004	-0.019	0.004	-0.025	0.010	-0.001	0.006	-0.035	0.005	-0.031	0.013
SDM	-0.002	0.004	-0.021	0.005	-0.025	0.010	0.000	0.006	-0.036	0.005	-0.031	0.013
Total	-0.048	0.012					-0.067	0.016				
Percentages:												
MDS			43%		45%		5%		61%		34%	
DMS			46%		45%		4%		62%		34%	
SMD			40%		52%		2%		53%		46%	
SDM			44%		52%		0%		54%		46%	
	M mobility		D distrib		R residual		M mobility		D distrib		R residual	
Overall inequalities												
MDS	0.001	0.001	-0.011	0.002	-0.013	0.008	0.002	0.004	-0.052	0.006	0.001	0.035
DMS	0.001	0.001	-0.011	0.002	-0.013	0.008	0.002	0.003	-0.052	0.006	0.001	0.035
SMD	-0.001	0.001	-0.009	0.003	-0.014	0.007	-0.001	0.005	-0.039	0.010	-0.010	0.031
SDM	-0.001	0.001	-0.009	0.003	-0.014	0.007	-0.001	0.004	-0.038	0.011	-0.010	0.031
Total	-0.024	0.008					-0.049	0.035				
Percentages:												
MDS	-2%		47%		55%		-4%		106%		-3%	
DMS	-2%		47%		55%		-3%		106%		-3%	
SMD	2%		38%		59%		2%		79%		20%	
SDM	3%		38%		59%		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 wages from 1988 to 1996; the simulation paths are noted by the order of changes, with M denoting educational mobility, D the marginal distributions 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.

6. THE POTENTIAL EFFECTS OF AN INCREASE IN EDUCATION MOBILITY ON EARNINGS INEQUALITIES

6.1. Methodology

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

$$p^{(i)}_t(s,o) = p_t(s,.)p_t(.,o) \tag{14}$$

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 wage determinants have 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 E. Two individuals of different origins 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 F presents the frequencies for such a simulated distribution. The probabilities of reaching a given level of education conditional on origins are equal.

Table E: Frequencies of the assumed distribution observed in t and average earnings by level of education and type of origin
[Marginal distributions (0.50; 0.50) and odds ratio of 16]

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

Table F: Frequencies of the distribution simulating perfect educational mobility
[Marginal distributions (0.50; 0.50) and odds ratio of 1]

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

The counterfactual earnings densities are obtained by reweighting the observations by the ratios of values between tables E and F, based on the formula given by equation (6). The counterfactual indices of overall inequalities 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, $E(y|o, t_y=t)$, estimated on the basis of the averages observed for the sample.

For the observed distribution:

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

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

For the simulated distribution:

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

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

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

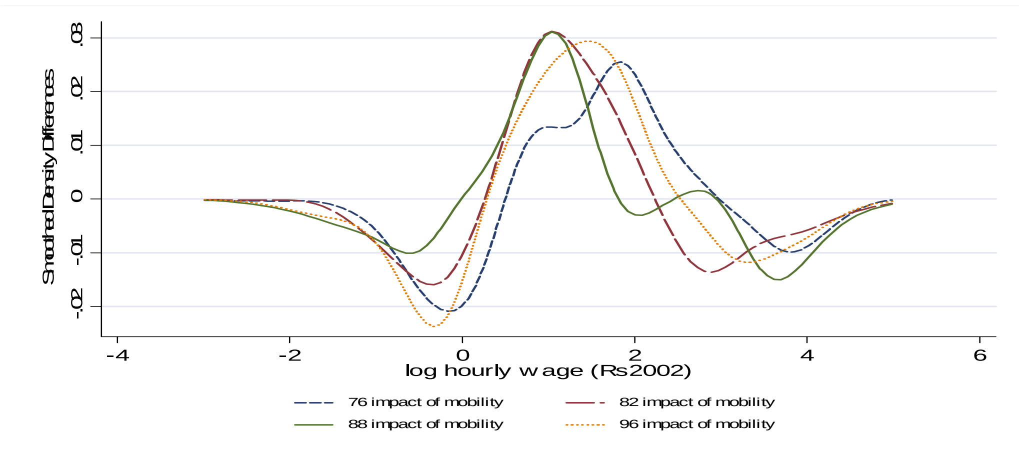
6.2. Findings: impact of perfect educational mobility on earnings inequalities

Table 9 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% and the Theil index by at least 78% (and by 73% and 91% respectively for 1982 and 1988)[19]. Intergenerational educational mobility plays a predominant role in the inequalities of opportunity on the labour market. The residual inequalities are due to the earnings gaps directly associated with social origin. Given that these gaps are fairly small at the bottom of the distribution of education levels (see tables in Appendix A2), the Theil index decreases considerably more than the Gini 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 (see above).

As regards overall inequalities, Figure 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 1 to 3 Gini index points depending on the year, or a relative reduction of 2% to 5%. This finding is in line with those obtained by Bourguignon *et al.* (2003) for several age brackets in 1996 as regards the 'indirect' (education-related) effect of social origin on earnings inequalities. Here again, the variation in the Theil index is greater for the abovementioned reason.

Figure 3: Differences between observed densities and simulated densities with perfect educational mobility



Method: Densities simulated by reweighting using the formula given by equation (6) and based on educational mobility matrices, where origin and education level are independent, estimated using the formula given in equation (14).

---

19 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.

Appendix A3 nevertheless shows that this last decomposition could represent a considerable underestimation of the impact of intergenerational educational mobility on overall earnings inequalities. Contrary to the simulations regarding the inequality of opportunity indices, but also contrary to the historical decompositions presented above (Section 5), this last decomposition is highly sensitive to measurement errors in the analyzed variable, here hourly earnings. This is intuitively understood since this static decomposition can only concern the proportion of inequalities corresponding to actual earnings gaps. In the case of the inequality of opportunity indices, the fact of considering averages or quantiles by type of social origin at least partially offsets these measurement errors. In the case of the historical decompositions, the main factor likely to confound the estimates is a variation in the 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.

Coming back to the static decomposition of overall inequalities, Appendix A3 uses a simple case (log-normality) to show the effect of measurement errors in terms of their share in the variance of the analyzed variable. The review of the literature made by Bound, Brown and Mathiowetz (2001) suggests that a proportion of 20% to 30% is not unreasonable in the case of the measurement of hourly wages. Yet the simulations presented in Appendix A3 show that a proportion of 20% can reduce the true effect threefold, while a proportion of 30% 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). However, such a discussion calls for caution with regard to this theoretical scenario, which has no close or even remote basis in historical fact, since intergenerational educational mobility varies little over the twenty years analyzed.

7. CONCLUSION

This paper studies the impact of changes in educational opportunities on overall inequalities and the inequality of opportunities on the labour 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 (i) schooling expansion, (ii) changes in the structure of earnings, and (iii) 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 1995 (a four-figure rate, see Appendix A4). First of all, the decompositions show that changes in the distribution of education contributed to the increase in both types of inequalities among the oldest generations before sharply reducing them among the post-WWII cohorts. Secondly, the decrease in returns to education also contributed to equalizing labour market opportunities in the 1988-1996 period. Thirdly and 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 inequalities. This rigidity of inequalities 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 & de Barros for household income (2000 and 2004). Among the generations born before World War II, growth in education concerned mainly the spread of access to secondary and higher education for the children of the upper classes, which increased the inequalities. It was only with the post-war generations that the expansion of primary education and the opening of the secondary system to children of farmers and 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 (Cardoso presidency and Real plan).

This last period of education-related reduction in earnings inequalities could give rise to optimism as to the long-run effects of programmes to educate poor children such as *Bolsa Escola*. 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 opportunities 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 inequalities *via education* in the future without a marked increase in intergenerational educational mobility. Yet it is still too soon to know whether programmes such as *Bolsa Escola* will manage to significantly stimulate this mobility.

Table 9: Simulations of perfect educational mobility

	1976		1982		1988		1996	
Inequalities of opportunity								
Gini								
Simulated	-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								
Simulated	-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 inequalities								
Gini								
Simulated	-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								
Simulated	-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.

BIBLIOGRAPHY

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

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

Bertail P. and Combris P. (1997), « Bootstrap généralisé d'un sondage », *Annales d'Economie et de Statistiques*, 46, 49-83.

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

Bittencourt M. (2005), "Macroeconomic Performance and Inequality: Brazil 1983-94", Working Paper, CMPO, University of Bristol.

Bound J., Brown C. and Mathiowetz N. (2001), "Chapter 59: Measurement Error in Survey Data", *in* J. J. Heckman, E. Leamer, *Handbook of Econometrics*, Vol. 5, North-Holland.

Bourguignon F., Ferreira F. and Leite Ph. (2004), "Conditional Cash Transfers, Schooling, and Child Labor: Micro-Simulating Brazil's *Bolsa Escola* Program", *World Bank Economic Review*, 17 (2).

Bourguignon F., Ferreira F. and Menendez M. (2003), "Inequality of Outcomes and Inequality of Opportunities in Brazil", Working Paper, World Bank.

DiNardo J., Fortin N. and Lemieux T. (1996), "Labor Market Institutions and the Distribution of Wages, 1973-92: A Semiparametric Approach", *Econometrica*, 64 (5).

Dunn Ch. (2003), "Intergenerational Earnings Mobility in Brazil and its Determinants", Working Paper, University of Michigan.

Engerman S. and Sokoloff K. (2000), "History Lessons: Institutions, Factor Endowments, and Paths of Development in the New World", *Journal of Economic Perspectives*, 14 (3).

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

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* F. Bourguignon, F.H.G. Ferreira and N. Lustig (eds.), *The Microeconomics of Income Distribution Dynamics in East Asia and Latin America*, World Bank and Oxford University Press, 83-124.

Ferreira S. and Veloso F. (2004), "Intergenerational Mobility of Wages in Brazil", Working Paper, IBMEC.

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 Levison D. (1991), “Declining Inequality in Schooling in Brazil and its Effects on Inequality in Earnings”, *Journal of Development Economics*, 37 (1).

Lam D. and Schoeni R. (1993), “Effects of Family Background on Earnings and Returns to Schooling: Evidence from Brazil”, *Journal of Political Economy*, 101 (4).

Leite Ph. (2004), « Méthodes d’évaluation Ex Ante et Ex Post : le Cas du Programme *Bolsa Escola* au Brésil », DEA APE dissertation, EHESS.

Manski C. and Pepper J. (2000), “Monotone Instrumental Variables: With an Application to the Returns to Schooling”, *Econometrica*, Notes and Comments, 68(4), 997-1010.

Moreno-Ternero J. (2004), “On Roemer’s Equality of Opportunity”, Working Paper, Yale University.

Pastore J. (1982), *Inequality and Social Mobility in Brazil*, University Wisconsin Press.

Pastore J. and Valle Silva N. (2000), “Mobilidade Social no Brasil”, Makron, Sao Paulo.

Picanço F. (2004), “Economic Modernization and Socio-Occupational Mobility in Brazil”, Communication at the International Sociological Association, RC28.

Roemer J. (1996), *Theories of Distributive Justice*, Harvard Univ. Press, Cambridge, Ma.

Roemer J. (1998), *Equality of Opportunity*, Harvard Univ. Press, Cambridge, Ma.

Van de Gaer D., Schokkaert E. and Martinez M. (2001), “Three Meanings of Intergenerational Mobility”, *Economica*, 68.

Appendix A 1: Educational mobility tables

Rows: social origin; columns: number of years of education

1976

	margin	0	1	2	3	4	5-7	8	9-11	12+	Total
Father farmer, without education	29.6	52.6	10.1	12.0	11.0	11.5	2.4	0.1	0.1	0.2	100.0
Father other, without education	6.3	33.4	7.5	10.5	8.5	22.1	7.4	7.0	2.1	1.4	100.0
Father farmer, literate	18.2	25.9	9.3	14.3	16.1	23.1	8.3	1.8	1.0	0.3	100.0
Father other, literate	10.0	7.7	3.7	13.6	12.2	28.1	16.5	9.0	6.8	2.3	100.0
Father farmer, primary education (1-4)	8.9	17.7	8.1	11.1	18.1	28.9	10.4	1.9	2.5	1.4	100.0
Father other, primary education (1-4)	12.5	4.9	4.0	3.0	8.7	27.0	18.9	9.4	10.4	13.6	100.0
Father upper primary education (5-8)	2.9	0.0	1.4	2.2	5.7	9.1	12.4	19.6	25.6	24.0	100.0
Father secondary education or more (9+)	2.4	0.0	0.0	3.3	5.0	3.7	10.6	8.8	29.0	39.7	100.0
No answer	9.3	32.9	7.7	9.6	8.9	19.1	9.6	4.2	5.2	2.8	100.0
Margin		28.4	7.5	10.6	11.8	19.9	9.1	4.2	4.4	4.1	100.0
Total sample number	2,860										

Source: PNAD 1976, social mobility questionnaire sub-sample, men aged 40 to 49, employed and head or spouse of head of household.
Reading: Sample numbers by social origins (father's education level and occupation) and level of education.

1982

	margin	0	1	2	3	4	5-7	8	9-11	12+	Total
Father farmer, without education	28.9	53.6	7.9	10.6	11.0	11.9	3.2	0.8	0.8	0.3	100.0
Father other, without education	7.7	37.7	6.0	10.8	11.7	19.6	7.1	2.9	2.7	1.5	100.0
Father farmer, literate	11.9	30.3	8.1	13.3	14.3	22.0	6.9	2.4	1.6	1.0	100.0
Father other, literate	7.8	11.2	3.6	7.0	12.3	29.6	15.2	9.3	6.6	5.3	100.0
Father farmer, primary education (1-4)	13.2	14.3	5.0	10.9	17.8	30.1	10.5	4.2	3.8	3.5	100.0
Father other, primary education (1-4)	15.5	3.4	1.9	2.9	7.8	23.6	15.9	13.0	16.3	15.2	100.0
Father upper primary education (5-8)	2.1	1.7	0.5	0.9	3.0	10.2	12.3	16.4	23.7	31.4	100.0
Father secondary education or more (9+)	2.9	0.6	0.0	0.0	1.4	4.8	4.6	9.5	21.2	57.8	100.0
No answer	9.9	26.8	7.1	10.4	12.5	20.6	9.0	6.0	5.1	2.6	100.0
Margin		28.0	5.6	9.0	11.7	19.9	8.6	5.3	5.8	6.1	100.0
Total sample number	18,832										

Source: PNAD 1982, men aged 40 to 49, employed and head or spouse of head of household.
Reading: Sample numbers by social origins (father's education level and occupation) and level of education.

Appendix A 1: Educational mobility tables (cont.)
Rows: social origin; columns: number of years of education

1988

	margin	0	1	2	3	4	5-7	8	9-11	12+	Total
Father farmer, without education	26.2	46.4	9.1	11.0	11.2	15.1	3.9	1.7	1.3	0.4	100.0
Father other, without education	10.0	28.6	5.4	9.4	10.8	21.5	10.7	6.3	5.0	2.4	100.0
Father farmer, literate	12.5	20.3	7.1	13.5	16.1	26.1	8.1	2.8	3.8	2.1	100.0
Father other, literate	9.0	9.8	4.3	6.5	12.0	25.4	14.4	8.0	12.4	7.2	100.0
Father farmer, primary education (1-4)	9.8	10.5	5.4	10.1	15.3	29.7	11.6	6.7	5.6	5.1	100.0
Father other, primary education (1-4)	16.1	1.7	1.4	3.0	4.4	21.8	14.4	11.3	20.2	21.7	100.0
Father upper primary education (5-8)	2.1	0.8	1.1	0.4	3.1	8.0	11.8	11.5	22.5	40.8	100.0
Father secondary education or more (9+)	3.9	0.1	0.2	0.5	1.0	5.7	2.7	5.0	20.9	64.0	100.0
No answer	10.4	24.0	6.0	9.2	12.2	21.4	12.0	5.2	6.0	4.0	100.0
Margin		22.2	5.6	8.6	10.7	20.7	9.4	5.6	8.2	9.0	100.0
Total sample number	11,304										

Source: PNAD 1982, men aged 40 to 49, employed and head or spouse of head of household.
Reading: Sample numbers by social origins (father's education level and occupation) and level of education.

1996

	margin	0	1	2	3	4	5-7	8	9-11	12+	Total
Father farmer, without education	22.3	41.0	6.4	8.5	11.3	18.1	7.5	3.5	2.9	0.8	100.0
Father other, without education	7.4	24.4	4.8	8.0	9.4	20.9	13.2	8.6	8.0	2.7	100.0
Father farmer, literate	7.3	17.8	4.8	6.5	11.9	28.5	13.0	6.7	7.6	3.2	100.0
Father other, literate	8.2	6.0	1.0	3.7	5.3	21.2	18.2	13.5	20.1	11.0	100.0
Father farmer, primary education (1-4)	16.5	9.6	3.3	6.9	10.8	29.8	12.8	9.8	10.7	6.3	100.0
Father other, primary education (1-4)	21.9	3.2	1.0	2.1	4.5	17.5	14.1	14.2	24.2	19.2	100.0
Father upper primary education (5-8)	3.5	1.1	0.6	0.2	2.6	4.3	9.4	14.0	28.1	39.7	100.0
Father secondary education or more (9+)	4.4	0.0	0.3	0.2	0.5	2.2	4.0	4.7	23.2	64.8	100.0
No answer	8.6	21.3	3.6	8.9	10.4	21.6	12.4	8.5	10.1	3.2	100.0
Margin		16.9	3.3	5.7	8.3	20.2	11.8	9.2	13.4	11.3	100.0
Total sample number	14,096										

Source: PNAD 1982, men aged 40 to 49, employed and head or spouse of head of household.
Reading: Sample numbers by social origins (father's education level and occupation) and level of education.

Appendix A 2: Average hourly earnings by level of education and type of origin

1976

	0	*1*	*2*	*3*	*4*	*5-7*	*8*	*9-11*	*12+*	*Total*
Father farmer, without education	1.61	2.16	2.19	3.07	4.28	5.24	14.54	6.08	18.92	2.34
Father other, without education	2.47	2.29	3.05	3.77	4.32	3.63	4.01	8.67	18.78	3.60
Father farmer, literate	2.19	3.42	2.74	3.54	4.29	4.11	6.49	13.30	10.06	3.45
Father other, literate	3.45	4.82	3.70	5.03	5.71	4.25	8.20	9.10	17.03	5.62
Father farmer, primary education (1-4)	3.03	3.18	3.25	3.52	4.59	4.66	7.62	10.49	14.02	4.20
Father other, primary education (1-4)	2.66	4.69	7.66	6.38	7.31	7.27	8.90	11.68	23.06	9.64
Father upper primary education (5-8)		2.70	2.35	3.73	1.98	12.57	7.35	17.91	26.96	14.53
Father secondary education or more (9+)			3.55	8.28	2.13	8.37	19.08	24.58	29.16	21.87
No answer	1.73	2.23	3.79	2.55	3.79	6.14	10.97	15.01	37.51	4.92
Total	1.94	2.86	3.01	3.81	4.96	5.71	8.53	14.55	25.09	5.07

Source: PNAD 1976, social mobility questionnaire sub-sample, men aged 40 to 49, employed and head or spouse of head of household.
Reading: Average hourly earnings by social origins (father's education level and occupation) and level of education.

1982

	0	*1*	*2*	*3*	*4*	*5-7*	*8*	*9-11*	*12+*	*Total*
Father farmer, without education	1.61	2.29	2.29	2.86	3.91	3.61	6.30	9.81	19.11	2.36
Father other, without education	2.01	2.10	2.83	3.01	4.47	4.26	5.90	8.28	13.68	3.32
Father farmer, literate	1.71	2.25	3.18	3.51	4.25	3.99	6.11	9.16	14.33	3.28
Father other, literate	2.52	2.51	2.83	4.05	5.32	5.13	8.91	10.52	19.94	5.99
Father farmer, primary education (1-4)	2.08	2.37	2.70	3.32	4.68	5.68	6.92	11.23	17.12	4.62
Father other, primary education (1-4)	2.42	3.40	3.16	4.22	6.38	6.50	9.29	12.67	23.04	9.89
Father upper primary education (5-8)	1.44	4.51	2.19	3.28	4.81	5.30	9.31	15.29	25.50	14.45
Father secondary education or more (9+)	1.38		2.19	5.24	5.24	8.88	9.45	15.21	29.14	21.72
No answer	2.35	2.02	2.64	2.84	4.20	4.96	7.74	10.43	22.92	4.29
Total	1.81	2.32	2.68	3.31	4.82	5.33	8.37	12.27	23.93	5.31

Source: PNAD 1982, men aged 40 to 49, employed and head or spouse of head of household.
Reading: Average hourly earnings by social origins (father's education level and occupation) and level of education.

Appendix A 2: Average hourly earnings by level of education and type of origin (cont.)

1988

	0	*1*	*2*	*3*	*4*	*5-7*	*8*	*9-11*	*12+*	*Total*
Father farmer, without education	1.38	1.96	2.09	2.57	3.32	3.51	5.58	6.72	15.44	2.21
Father other, without education	1.62	2.76	2.55	2.84	4.26	4.24	6.86	7.56	15.54	3.71
Father farmer, literate	1.39	1.98	2.75	3.01	3.97	3.53	5.10	9.27	13.44	3.39
Father other, literate	1.84	2.51	2.41	3.20	4.18	3.89	6.59	11.06	18.35	5.67
Father farmer, primary education (1-4)	1.31	2.32	2.34	3.37	4.04	4.40	7.13	8.28	17.60	4.57
Father other, primary education (1-4)	2.17	2.77	1.88	3.58	4.17	6.26	9.48	10.81	20.58	9.83
Father upper primary education (5-8)	1.03	0.93	0.82	2.67	5.77	4.13	7.10	11.12	26.97	15.38
Father secondary education or more (9+)	2.86	2.03	1.17	8.10	3.82	4.41	8.49	14.51	27.11	21.23
No answer	1.64	1.87	2.45	2.96	3.88	5.39	6.91	8.85	13.65	4.00
Total	1.47	2.13	2.35	2.99	3.95	4.70	7.57	10.45	21.95	5.48

Source: PNAD 1988, men aged 40 to 49, employed and head or spouse of head of household.
Reading: Average hourly earnings by social origins (father's education level and occupation) and level of education.

1996

	0	*1*	*2*	*3*	*4*	*5-7*	*8*	*9-11*	*12+*	*Total*
Father farmer, without education	1.66	1.92	2.10	2.39	3.62	3.56	3.87	7.11	9.79	2.60
Father other, without education	1.57	2.75	2.50	3.43	3.41	5.06	5.28	5.69	15.21	3.75
Father farmer, literate	1.55	1.81	1.93	2.78	4.52	4.34	8.83	7.48	11.84	4.20
Father other, literate	3.15	2.86	3.88	3.94	4.44	4.01	6.11	10.18	15.44	6.81
Father farmer, primary education (1-4)	1.99	2.32	2.60	2.80	3.99	3.95	6.56	7.98	19.15	5.14
Father other, primary education (1-4)	3.51	2.89	2.51	3.27	5.04	6.47	6.51	10.59	17.70	9.02
Father upper primary education (5-8)	6.88	4.11	11.12	4.06	5.39	2.90	5.75	11.87	20.19	12.89
Father secondary education or more (9+)		0.31	2.35	3.95	5.41	5.71	9.40	11.38	25.93	20.27
No answer	2.06	1.63	1.93	2.66	3.91	3.82	5.47	7.09	19.59	4.08
Total	1.85	2.13	2.35	2.84	4.17	4.67	6.23	9.61	19.79	6.23

Source: PNAD 1996, men aged 40 to 49, employed and head or spouse of head of household.
Reading: Average hourly earnings by social origins (father's education level and occupation) and level of education.

Appendix A 3: Impact of measurement errors on the decomposition of the inequality indices

We endeavour to assess the impact of the distribution of a continuous variable ln*x* with standard deviation σ_x (here social origin or level of education) on the distribution of a continuous variable ln*w* (here log of earnings) with standard deviation σ_w. We assume that (ln*w*, ln*x*) obey a bivariate normal law. We also assume that the two variables are affected by measurement errors in the form of white noise. The errors on ln*w* (resp. ln*x*) have standard deviation θ_w (resp. θ_x).

The naive regression of ln*w* on ln*x* lessens the ln*x* coefficient:

$$\beta^2 \sigma^2_x = \alpha^2 (\sigma^2_x - \theta^2_x)(1- \theta^2_x/\sigma^2_x)$$

where α is the real value of this coefficient, and β the naive estimator.

Let $\mu \equiv \alpha^2 (\sigma^2_x - \theta^2_x) / (\sigma^2_w - \theta^2_w)$ be the real proportion of the variance of ln*w* explained by ln*x*.
And $1-\lambda_w \equiv \theta^2_w/\sigma^2_w$ be the proportion of measurement errors in the variance of ln*w*.
And $1-\lambda_x \equiv \theta^2_x/\sigma^2_x$ be the proportion of measurement errors in the variance of ln*x*.

The R^2 of the naive regression of ln*w* on ln*x* is given by:

$$R^2 = \beta^2 (\sigma^2_x / \sigma^2_w) = \mu\lambda_w\lambda_x$$

The relative impact of ln*x* on the variance of ln*w* is thus given by this R^2.
Now the real decomposition would assume the calculation of μ. We therefore see that even if $\lambda_x = 1$ (no measurement error on ln*x*), the presence of measurement errors on y lessens the impact of ln*x* on the variance of ln*w*.
Given that (*w*, *x*) is log-normal, decompositions of the Gini index can be calculated.
The Gini index of *w* hence takes the value: $G = 2\ \Phi [\sigma_w /\sqrt{2}]-1$, where Φ is the cdf of the standard normal law.
The Gini index of *w* without measurement errors takes the value $G^* = 2\ \Phi [\sigma_w\sqrt{\lambda_w} /\sqrt{2}]-1$
The Gini index obtained by cancelling out the impact of *x* in the naive decomposition is:

$$G_\Delta = 2\ \Phi [\sigma_w \sqrt{(1-\mu\lambda_w\lambda_x)} /\sqrt{2}]-1$$

The counterfactual Gini index in the real decomposition is:

$$G^*_\Delta = 2\ \Phi [\sigma_w \sqrt{((1-\mu)\lambda_w)} /\sqrt{2}]-1$$

The following two tables present the relative impact of *x* on the Gini index of *w* in the case of the naive decomposition $((G_\Delta - G)/G)$ and in the case of the real distribution $((G^*_\Delta - G^*)/G^*)$. This impact is broken down according to the proportion of measurement errors and for two values of the observed Gini index G (the row in bold corresponds to the magnitude observed in the Brazilian case). The first table makes the assumption that *x* has a small influence (the variance of *x* only explains 25% of the true variance of *w*, but also the absence of measurement errors on *x*). The second table makes the assumption that *x* has a strong influence (50% of the variance of *w*), but introduces a measurement error on x ($1-\lambda_x=0.2$).

$\mu = 0.25$ (small influence of *x*) and $1-\lambda_x=0$ (absence of measurement errors on x)

Measurement errors on w	$1-\lambda_w = 0$		$1-\lambda_w = 0.2$		$1-\lambda_w = 0.3$	
Observed Gini (G)	Naive	True	Naive	True	Naive	True
0.30	-12.9%	-12.9%	-10.1%	-27.0%	-8.8%	-37.7%
0.60	**-11.0%**	**-11.0%**	**-8.6%**	**-22.9%**	**-8.5%**	**-36.3%**

$\mu = 0.5$ (strong influence of *x*) and $1-\lambda_x = 0.2$ (measurement errors on x)

Measurement errors on w	$1-\lambda_w = 0$		$1-\lambda_w = 0.2$		$1-\lambda_w = 0.3$	
Observed Gini (G)	Naive	True	Naive	True	Naive	True
0.30	-21.8%	-28.4%	-16.9%	-44.4%	-14.6%	-56.3%
0.60	**-19.1%**	**-25.3%**	**-14.6%**	**-39.0%**	**-14.1%**	**-54.6%**

As shown, the impact estimated in Gini points can come out well below the true impact depending on the proportion of measurement errors in the variance of the analyzed variable. The naive decomposition bias is even more sensitive to errors regarding the analyzed variable than to errors made regarding the influence of the decomposition variable.

In the case where we observe historical growth, this underestimation problem associated with measurement errors on the analyzed variable is largely diminished. Assume that, during this growth, the role of variable *x* becomes zero and that the measurement errors on *w* retain the same distribution.

We observe the initial inequalities: $G_t = 2\ \Phi\ [\sigma_w /\sqrt{2}\ ]\text{-}1$

Then the end inequalities: $G_{t'} = 2\ \Phi\ [\sigma_w \sqrt{(1\text{-}\mu\lambda_w)}\ /\sqrt{2}\ ]\text{-}1$

The counterfactual is always: $G_\Delta = 2\ \Phi\ [\sigma_w \sqrt{(1\text{-}\mu\lambda_w\lambda_x)}\ /\sqrt{2}\ ]\text{-}1$

We see that, in the absence of measurement errors on x, the counterfactual measures historical growth perfectly. Obviously, if the proportion of measurement errors in w also changes, then the historical decomposition is confounded: $G_{t'} = 2\ \Phi\ [\sigma_w \sqrt{(1\text{-}\mu\lambda_{w,t'})}\ /\sqrt{2}\ ] - 1$ is no longer equal to G_Δ when $\lambda_x = 1$.

The case of inequality of opportunity indices calls for the inclusion of two decomposition variables x_1 (here social origin) and x_2 (here level of education). We again assume that (lnw, ln x_1, ln x_2) obeys a normal three-dimensional law. In the case of the Van der Gaer index, we consider the conditional expectations at x_1. These are only influenced by the measurement errors on $x=(x_1, x_2)$.

The naive estimation yields:

$E(\ln w \mid \ln x_1) = \beta_1 \ln x_1 + \beta_2 E(\ln x_2 \mid \ln x_1) = \beta' \ln x_1$

When the true model is:

$E(\ln w \mid \ln x_1) = \alpha_1 \ln x_1 + \alpha_2 E(\ln x_2 \mid \ln x_1) = \alpha' \ln x_1$

The counterfactual that cancels out the effect of x_1 on x_2 is calculated by $\beta_1 \ln x_1$, whereas it theoretically takes the value $\alpha_1 \ln x_1$.

With the Gini index, we compare $VdG=2\ \Phi\ [\sigma_{x1}\ \beta' / \sqrt{2}\ ]\text{-}1$ with $VdG_\Delta=2\ \Phi\ [\sigma_{x1}\ \beta_1 / \sqrt{2}\ ]\text{-}1$

The true decomposition is $VdG^*=2\ \Phi\ [\sigma_{x1}\ \alpha'\ \sqrt{\lambda_{x1}}\ / \sqrt{2}\ ]\text{-}1$ compared with $VdG_\Delta=2\ \Phi\ [\sigma_{x1}\ \alpha_1 \sqrt{\lambda_{x1}} / \sqrt{2}\ ]\text{-}1$

Only the measurement errors on $x=(x_1, x_2)$ come into play.

This paper's simulations, which isolate the impact of intergenerational educational mobility, come within this two-variable framework even if no parametric assumption (log-normality) is made. We can hence check whether a static decomposition of the overall inequality index raises the same problems as in the univariate framework addressed above, i.e. the strong influence of measurement errors on the analyzed variable.

Appendix A 4: Per capita GDP and inflation rate (World Development Indicators 2003)