

Determinants of unfair inequality in Brazil, 1995 and 2009

Ana Claudia Annegues, Erik Alencar de Figueiredo and Wallace Patrick Santos de Farias Souza

ABSTRACT

This article analyses the trend of unfair inequality in Brazil (1995-2009) using a non-parametric approach to estimate the income function. The entropy metrics introduced by Li, Maasoumi and Racine (2009) are used to quantify income differences separately for each effort variable. A Gini coefficient of unfair inequality is calculated, based on the fitted values of the non-parametric estimation; and the robustness of the estimations, including circumstantial variables, is analysed. The trend of the entropies demonstrated a reduction in the income differential caused by education. The variables “hours worked” and “labour-market status” contribute significantly to explaining wage differences imputed to individual effort; but the migratory variable had little explanatory power. Lastly, the robustness analysis demonstrated the plausibility of the results obtained at each stage of the empirical work.

KEYWORDS

Economic conditions, income distribution, income, education, employment, statistical data, statistical methodology, Brazil

JEL CLASSIFICATION

C14, D63

AUTHORS

Ana Claudia Annegues is a doctoral student in the Graduate Programme in Economics at the Federal University of Rio Grande do Sul (UFRGS), Brazil. annegues.ana@gmail.com

Erik Alencar de Figueiredo is a professor with the Department of Economics of the Postgraduate Programme in Economics at the Federal University of Paraiba, Brazil. efigueiredo@gmail.com

Wallace Patrick Santos de Farias Souza is a doctoral student in the Graduate Programme in Economics at the Federal University of Rio Grande do Sul, Brazil. wpsfarias@gmail.com

I

Introduction

The Brazilian economy, which endured a long period of rising inequality from the 1960s onwards, has seen an improvement in its social indicators since the mid-1990s (Azevedo, 2007). This has resulted from a macroeconomic environment favouring the transformations caused by the stabilization of inflation and Brazil's integration into the international market. Government income transfer programs also increased substantially in that period, which led to a rise in income and a consequent reduction in poverty levels (Figueiredo and Netto Júnior, 2014).

Nonetheless, although those indicators might lead one to believe that Brazil became a fairer country, the bibliography has questioned the way income inequality is usually treated, in which perfect equality is considered the ideal of social justice. Authors supporting this line of thought are highly diverse, ranging from the followers of the Rawlsian tradition, such as Dworkin (1981) and Arneson (1989), who support income differences related to individual preferences, to authors such as Roemer (1998), who propose the concept of inequality of opportunities.

Following the latter approach, individual economic outcomes depend partly on responsibility variables (defined as effort) and partly on variables that are beyond the scope of individual responsibility (defined as circumstantial). Following that line of reasoning, several researchers argue that only unfair inequality, or inequality caused by variables beyond the scope of responsibility, is socially undesirable.

Some studies have attempted to measure inequality of opportunities and its contribution to total inequality. For that purpose, direct and indirect parametric techniques have been used based on a defined functional form — as in Bourguignon, Ferreira and Menéndez (2007) — where the determinants of the counterfactual income distribution are considered by assuming that all individuals are in the same circumstances; and in Salvi (2007), which uses panel data to distinguish circumstances and effort in time-variant and time-invariant variables.

Applying the concept of sensitivity to responsibility, Devooght (2008) measures income inequality and a

fairness criterion calculated on the basis of an income function g , approximated through log-linear regression. The author only uses responsibility variables, because it is not always possible to obtain a well-defined set of variables, particularly with respect to family background (the educational level and profession of the parents); so the component of inequality that is explained by the non-responsibility variables are included in the stochastic term.

Nonetheless, that methodology raises a number of problems; in particular the endogeneity generated by the degree of relation between the effort and circumstantial variables. Moreover, in many studies, assuming a functional form for individual incomes can result in specification bias, given the random nature of the relations between the variables.

This article attempts to bridge that gap by adopting a non-parametric approach (without a functional form defined for g), with the aim of answering the following questions, based on the hypothesis that incomes essentially reflect responsibility variables: which of variable has the greatest explanatory power? Have effort variables been losing their influence in the determination of income differentials over the period analysed? Does the suppression of variables beyond the individual's control expose the results to some degree of bias?

For that purpose, the following four-step empirical strategy is used. Firstly, a specification analysis is undertaken to justify the use of the non-parametric method for estimating the income equation. The second step consists of dividing the sample into two groups of individuals, classified on an ad hoc basis as “high-effort” and “low-effort”, according to each responsibility variable, and then estimating incomes in relation to each of the determinants of effort (while holding the others constant) and observing their capacity to explain income inequality in the period analysed. The curves are compared on the basis of the entropy metrics discussed in Racine (2006). In the third step, Gini coefficients of unfair inequality are produced to measure the repercussion of the non-parametric inferences on the calculation of indices of fairness, and to verify whether the traditional parametric approach tends to overestimate those measures or not. To make inferences on the results obtained, the fourth and last procedure makes an analysis of robustness using data from the 1996

□ The authors gratefully acknowledge financial support from the National Scientific and Technological Development Council (CNPq) (Project 470202/2012-2).

National Household Survey (PNAD). That year was chosen because the PNAD published a social mobility supplement that summarizes data on the education, schooling and occupation of the parents of the selected individual (family background).

The period between 1995 and 2009 is analysed using data from the annual PNADS. The effort variables used are education level, the decision to emigrate, hours worked, and labour-market status.¹ The logarithm of nominal

wages is used as the dependent variable (proxying for individual income).

Following this introduction, the rest of the article is organized as follows: section II includes the methodological procedures, and presents the empirical strategy which is divided into the analysis of the relation between inequality and effort, estimation of the variable without a defined functional form, and the non-parametric model. Section III is devoted to the presentation of results and discussion, along with a comparison of the distributions for each responsibility variable used and for determining the entropy metric. Section IV presents debates on the topic, and section V sets out final thoughts.

¹ Those variables are presented in greater detail in subsection 1.c of section II, which describes the databank.

II

Methodological procedures

The objective of this section is to present the key methodological procedures of the study. It firstly highlights the theoretical concepts related to the bibliography on unequal opportunities, or unfair inequality; then, the steps in the empirical analysis are described, to clarify the study's analytical procedures (see subsection 1.a of this section). The methods involved in each step are described in subsection 1.b of this section. Lastly, the databank is presented.

1. Inequality of opportunities: theoretical and technical procedures

The traditional approach to unequal opportunities considers that an individual's economic outcome stems from circumstantial factors, such as family background, race or gender, among others, combined with effort, which is associated with variables that can be controlled by the person in question, including years of schooling or hours worked, for example (Roemer, 1998).² In short, unlike the modern egalitarian vision,³ a society is considered fair if it guarantees the same access to social benefits to all of its members, irrespective of their circumstances, such that effort turns access to those benefits into real

gains for individuals. Put another way, only inequality the results from circumstantial variables is socially undesirable. Accordingly, equalizing opportunities means correcting unequal circumstances while keeping effort differences unchanged.⁴

Nonetheless, although the concept of equal opportunities is quite simple, implementing it raises a number of significant challenges, not least the definition of the effort variable. Fleurbaey (1998) considers that effort is not observable, and uses a non-parametric approach to identify it, based on the Roemer identification assumption that multi-dimensional effort variables can be distributed independently of the individuals' circumstances. In contrast, Bourguignon, Ferreira and Menéndez (2007) believe effort is observable; and they develop a parametric model in which effort endogenously depends on circumstances. In short, effort can be considered either a latent (non-observable) variable, or not.⁵ When a structure is adopted in which this is an observable factor, the traditional parametric representations are based on the fact that an economic outcome, such as the wage,

² Alternatively, inequality of opportunities can be studied by comparing economic outcomes in relation to the sets of opportunity variables. See Pattanaik and Xu (1990) and Kranich (1996) for more detailed information. Ooghe, Schokkaert and Van de Gaer (2007) provide a summary of the different approaches.

³ See, for example, Dworkin (1981) and Arneson (1989).

⁴ Discussion on the social norm of fairness is complex and involves a series of mutually conflicting perspectives. For a synthesis, see Thomson (2011) and Fleurbaey and Maniquet (2011).

⁵ Bourguignon, Ferreira and Menéndez (2007), for example, approach effort variables on the basis of years of education, the decision to migrate and the worker's labour-market status. Checchi and Peragine (2009) consider effort as a non-observable factor, thereby adopting the Roemer axiom (Roemer, 1996); and they assume that individuals with different opportunity sets, but who are in the same percentile of the corresponding distribution, make the same level of effort.

is a function of the variables of circumstances (C_i) and effort (E_i) following an additively separable structure:

$$Y_i = \alpha C_i + \lambda E_i + u_i \quad (1)$$

Equation (1), however, requires the two sets of variables (effort and circumstances) to be available, which is generally not possible in most international surveys. In view of this, a variety of studies assume that the economic outcome can be determined from a series of effort variables, together with assumptions concerning the random term.

Devooght (2008), for example, applies a normative criterion under which the error term is included in the set of circumstantial variables (C_i). Following the same line of reasoning, Almås (2008), and Almås and others (2011), devise a criterion of fairness that is sensitive to responsibility; and they propose a Gini coefficient of unfair inequality that is applied to data from Norway. Figueiredo and Netto Júnior (2014) use a similar strategy to calculate unfair inequalities in Brazil between 1995 and 2009.

The results for the Brazilian economy are striking, because the country shows a significant reduction in total income inequality (Gini coefficient), without any alteration in unfair inequality. As the coefficients of unfair inequality are calculated on the basis of effort variables (education, hours of work, decision to migrate, and labour-market status), the maintenance of their values means that the importance of those co-variant variables in explaining the income equations have remained constant over the years. In fact, the equations contained in Figueiredo and Netto Júnior (2014) showed that the goodness of fit of the regressions (R^2) has deteriorated. In view of this, the authors postulate the following:

- (i) given that the construction of the fairness rule considers that the unobservable factors (error term) are the non-responsibility variables, a decrease in R^2 implies that these factors have a heavier weight in 2009 than in 1995. In other words, earnings density in 2009 depends much more on variables related to origin, color, and family background than on factors related to effort; or
- (ii) such a pattern would for instance occur if, with the quantitative expansion in the access to education, differences in the quality of education tend to matter more over time (Figueiredo and Netto Júnior, 2014).

Nonetheless, at least one caveat needs to be raised: the authors use a linear parametric structure similar to that of equation (1). If, for example, that specification

is not appropriate and the relation between income and the effort variables is non-linear, the values predicted by the estimated equations (used to calculate the unfair inequality coefficients) will not be valid; nor, consequently, will the degrees of fit of the regressions (R^2).⁶

In view of that limitation, the study proposes an alternative model which represents the economic outcome as follows:

$$\ln Y_i = m(E_i) + g(\xi_i) \quad (2)$$

where $g(\xi_i) = s(C_i, u_i)$. Unlike equation (1), equation (2) does not impose a linear form to describe the link between the economic outcome and its covariant variables. Instead, it assumes that this relation arises from unknown functions $m(\cdot)$, $g(\cdot)$ and $n(\cdot)$. It also considers that $E_i \perp \xi_i$, in other words, it postulates that the two sets of variables are mutually independent.⁷ This approach is inspired in the normative assumption of Devooght (2008) combined with the non-parametric structure widely adopted in economic studies.⁸ The application of those methods follows the following empirical sequence.

(a) *Sequence of the empirical work*

The empirical strategy includes the following procedures:

- (1) Specification analysis to compare the parametric and non-parametric alternatives for the income equations. This stage is necessary to avoid adopting the non-parametric structure on an ad hoc basis.
- (2) Creation of high- and low-effort groups. Four divisions are considered, each linked to one dimension of effort, namely:
 - Education: low-effort in the case of persons with less than four years' schooling and high-effort in other cases.
 - Migration: low-effort for non migrants and high-effort for migrants.
 - Hours of work: low-effort for persons who work less than 20 hours a week and high-effort for others.
 - Labour-market status: low-effort for own-account and informal workers and high-effort for workers in the formal sector.

The idea is to measure the distance between the high- and low-effort groups in a period covering several

⁶ See Maasoumi, Racine and Stengos (2007) for a discussion on the degree of fit in non-linear models and a presentation of alternatives to that parameter.

⁷ This hypothesis is relaxed in the section on robustness analyses.

⁸ For a synthesis see Li and Racine (2007).

years. The identification is based on the fitted regression values. This makes it possible to gauge whether effort is losing its influence in determining incomes among the Brazilian population.

- (3) Using the fitted values of the non-parametric regressions, fairness criteria and Gini coefficients of unfair inequality are calculated.
- (4) This procedure makes it possible to establish whether the gains obtained in applying a non-parametric method for the wage regressions is significantly reflected in the unfairness indices.

Lastly, a robustness analysis is performed based on a sample of effort and circumstantial variables.

The main lines of the analysis are the hypothesis of independence between the two sets of variables and the omission of circumstantial variables in procedures (1) to (3). The data used come from the social supplement of the 1996 PNAD.

The implementation of this sequence (procedures 1 to 4) is based on a non-parametric instrumental and the data set described in the next two subsections.

(b) *Inference*

Three non-parametric methods are used. The first is an analysis of specifications and relates to procedure (1) of the empirical strategy; in other words, the robustness of the non-parametric specification is tested against the parametric specification. The second method refers to the non-parametric estimations for the wage equations. The third, lastly, entails calculating the distances (entropies) between the low- and high-effort groups.

Specification analysis. This study chooses the kernel-based analysis developed by Racine (2006). To understand the structure of the Racine (2006) specification analysis, the following parametric representation should be considered:

$$y_i = q(x_i, \beta) + \eta_i, \quad i = 1, \dots, n$$

If the specification is correct, the expected error term η_i conditional on x_i will be equal to zero. Thus, the β estimator will be consistent. In contrast, if $q(x_i, \beta)$ was incorrectly specified, following the normal linear regression structure, for example, the β estimator would be inconsistent. Consequently, the correct specification would need to be considered as the null hypothesis:⁹

$$H_0: P\left[F\left(q(x_i, \beta) \mid x_i\right)\right] = 1, \text{ for some } \beta \in \aleph \forall i \geq 1$$

where $q(x_i, \beta)$ is a known function, β is a factor of unknown parameters, and \aleph is a compact subset of \mathbb{R}^d . The alternative hypothesis is as follows:

$$H_1: P\left[F\left(q(x_i, \beta) \mid x_i\right)\right] < 1, \text{ for some } \beta \in \aleph \forall i \geq 1$$

The null hypothesis is true if $E(\varepsilon_i \mid x_i) = 0$, and $I = E\left\{\left[E(\varepsilon_i \mid x_i)\right]^2 f(x_i)\right\} \geq 0$, where $f(x_i)$ is the marginal density for X . In this case, $I = 0$ if and only if H_0 is true. Thus, I is a candidate for analysing H_0 .

An observation is in order at this point. In general, kernel-based methods are appropriate for continuous data. Nonetheless, using a frequencies method (Li and Racine, 2007, chapter 3), it is possible to deal with discrete and continuous variables in the same dataset. Thus x_i^d can be considered as a vector of discrete variables of dimension $r \times 1$ and $x_i^c \in \mathbb{R}^d$ while the other variables are continuous. Accordingly, $x_i = (x_i^c, x_i^d)$.

The I sample is defined as:

$$I_n = n^{-2} \sum_i \sum_{j \neq i} \hat{\varepsilon}_i \hat{\varepsilon}_j K_{\gamma, ij}$$

where $K_{\gamma, ij} = W_{h, ij} L_{\lambda, ij}$, $\gamma = h, \varphi$ are the bandwidths, $W_{h, ij}$ and $L_{\lambda, ij}$ are the multivariate kernel functions for the discrete and continuous data, respectively; and $\hat{\varepsilon}_i$ are the model's sample errors.

Racine (2006) recommends using the cross-validation method for selecting the bandwidths,¹⁰ replacing $(h_1, \dots, h_q, \varphi_1, \dots, \varphi_r)$ by their estimators $(\hat{h}_1, \dots, \hat{h}_q, \hat{\varphi}_1, \dots, \hat{\varphi}_r)$. In this way, \hat{I}_n is defined. Under certain assumptions the author shows that, with H_0 ,

$$\hat{J}_n = n(\hat{h}_1, \dots, \hat{h}_q)^{1/2} \hat{I}_n / \sqrt{\hat{\Omega}} \rightarrow N(0, 1) \quad (3)$$

and

$$\hat{\Omega} = \frac{2(\hat{h}_1, \dots, \hat{h}_q)}{n^2} \sum_i \sum_{j \neq i} \hat{\varepsilon}_i^2 \hat{\varepsilon}_j^2 W_{h, ij}^2 L_{\lambda, ij}^2$$

⁹ The null hypothesis is based on the consistent analysis definition. For further information, see Li and Racine (2007).

¹⁰ See Härdle (1990) for further information on the cross-validation method.

The critical values are tabulated according to the *wild bootstrap* method. Racine (2006) presents and discusses details of the inference process, as well as the performance of the analysis in finite samples. Broadly speaking, the analysis \hat{J}_n consists of several stages, including the estimation of regressions with combined or mixed data (Racine and Li, 2004), obtaining estimates for the bandwidths of the discrete and continuous variables.

Non-parametric estimations. Many empirical studies admit a specific functional form, generally log-linear, for measuring the inequality of opportunities. Examples include Bourguignon, Ferreira and Menéndez (2007); Devooght (2008), and Figueiredo and Netto Júnior (2014). Nonetheless, parametric modelling might not be appropriate, given the uncertainty in relations between the explained and the explanatory variables. Alternatively, a non-parametric estimation does not impose functional forms for the equation, so it is capable of revealing aspects of the data structure that are not captured by the traditional parametric approach.

This study uses the constant local kernel estimator for mixed data, suggested by Racine and Li (2004), bearing in mind that the regressions include categorical variables. The authors use a kernel function to smooth the discrete variables, which adopts the following forms:

$$l(X_{t,i}^d, x_t^d) = \begin{cases} 1 & \text{se } X_{t,i}^d = x_t^d \\ \lambda & \text{se } X_{t,i}^d \neq x_t^d \end{cases}$$

If $d_{xi,x} = \sum_{t=1}^k 1(X_{t,i}^d \neq x_t^d)$ is defined as the number of unequal components between $X_{t,i}^d$ and x_t^d , then the product kernel for discrete variables will be

$$L(X_i^d, x^d, \lambda) = \prod_{t=1}^k l(X_{t,i}^d, x_t^d) = 1^{k-d_{xi,x}} \lambda^{d_{xi,x}} = \lambda^{d_{xi,x}}$$

Let $W(\cdot)$ be the kernel function associated with the continuous variables, and h its smoothed parameters. Using the notation $K_{h,ix} = W_{h,ix} L_{\lambda,ix}$, where $W_{h,ix} = h^{-p} W\left(\frac{X_i^c - x^c}{h}\right)$ and $L_{\lambda,ix} = L(X_i^d, x^d, \lambda)$, the kernel estimator for the joint density function of (X_i^c, X_i^d) is given by:

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n K_{h,ix} \tag{4}$$

In short, the true income function corresponds to the conditional average of the dependent variable, and its estimator is defined as:

$$\hat{g}(x) = \frac{n^{-1} \sum_{i=1}^n Y_i K_{h,ix}}{\hat{f}(x)} \tag{5}$$

where Y_i is the dependent variable, and $K_{h,ix}$ satisfies the aforementioned definition. The results of this function are summarized in the form of partial regression graphs. The explained part of the multivariate regression is plotted in relation to each effort variable, holding everything else constant, in a series of bivariate graphs. In other words, \hat{g} is plotted against the expected value of the regressors, keeping two of these at their respective averages; all of this for two different groups of individuals divided between low-effort and high-effort, according to each responsibility factor analysed. This procedure, used in Li, Maasoumi and Racine (2009), makes it possible to compare the curves of the two groups based on the distance between them, and thus observe the capacity of those regressors to explain income inequality over time. The distances are quantified through an entropy metric, which is explained in further detail in the next section.

Distance entropy metric. The entropy metrics represent a formal quantification of the distances between the distributions. An ideal metric of the distance between two random variables has a number of desirable properties, namely: (i) it is normalized on zero, if X and Y are identical; (ii) the modulus of the measure is equal to one, if there is an exact measurable (non-linear) relation between the variables $Y = g(X)$; (iii) it has a value close to the coefficient of linear correlation in the case of a bivariate normal distribution; (iv) it is a true metric of distance and not just of divergence between the distributions; (v) the ideal measure is well-defined for both continuous and discrete variables, and (vi) it is invariant to continuous and strictly increasing transformations such as $h(\cdot)$. This is useful provided X and Y are independent, and if and only if $h(X)$ and $h(Y)$ are also independent.

Li, Maasoumi and Racine (2009) use entropy metrics to make inferences on the degree of fit of the models, measuring the distance between observed values and fitted values, because the traditional R^2 and other measures of correlation are inadequate if the estimated function is non-linear. The entropy used by the authors was suggested by Granger, Maasoumi and Racine (2004)

and consists of a normalization with the aforementioned properties, as well as an appropriate distance metric:

$$S_{\rho} = \frac{1}{2} \int (f_1^{1/2} - f_2^{1/2})^2 dx \quad (6)$$

where f_1 and f_2 are the marginal densities of the random variables. To improve the reliability of the results, the bootstrap method is applied to a new sample, refitting the joint distribution of X and Y .

As noted above, this study will use entropy metrics as an indicator of the inequality of effort between the groups (procedure 3). The fitted values of the dependent variable for low- and high-effort individuals will be plotted against each explanatory variable, holding everything else constant. The entropy metrics are used to calculate the distances between the curves, and an analysis of equality of univariate densities is immediately performed, based on the null hypothesis $S_{\rho} = 0$.

(c) *Data*

The data used in the study come from the PNAD for the period between 1995 and 2009. They cover male heads of family, of between 25 and 60 years of age, living in urban areas. Individuals with these characteristics were chosen so as to homogenize the sample used as far as possible, restricting the analysis to individuals of active age, while also avoiding the influence of gender factors, both in the selection and in the individual economic outcomes. Heads of families were chosen because they are the main source of family income. Income is represented by the hourly wage, expressed in reais at September 2009 prices, as in Bourguignon, Ferreira and Menéndez (2007).

The vector of effort variables is represented by: (i) the individual's years of schooling; (ii) a dummy variable for migration; (iii) the individual's labour-market status (formal work, without a work contract, or working on own account), and (iv) weekly hours of work. High-effort and low-effort categories were defined for each of the variables, as indicated in procedure 2 of the empirical strategy sequence (see subsection 1.a of section II).

The variables were selected following Bourguignon, Ferreira and Menéndez (2007) and Figueiredo and Netto Júnior (2014); and they include individual responsibility factors, since, despite being influenced by circumstances, they are ultimately determined by individual choices. In the case of years of schooling, it is reasonable to assume that individuals with higher levels of schooling are those that have tried to obtain higher qualifications, they have invested in human capital so have made a greater effort.

The variables "labour-market status" and "weekly hours worked" are included because they represent the pursuit of higher wages, since individuals who work more hours and in formal jobs tend to obtain higher income. In addition, formal jobs are generally held by individuals with a higher level of schooling.

In relation to the migratory factor, individuals who decide to migrate are selected by the labour market and, according to Assis, Costa and Mariano (2012), they display a higher level of human capital, in addition to being more productive than non-migrants, except in the southeast region. Bearing those characteristics in mind, migrants would have capacity to achieve a higher level of income, which makes it plausible to include this variable as a measurement of effort.

III Results

This section sets out the results of each of the steps in the empirical strategy. Firstly, the income equation specifications are analysed; then, a graphical analysis is made of the non-parametric densities, comparing income differentials between the high- and low-effort groups over the period analysed, varying each responsibility factor separately. The formal quantification of those distances is obtained from the entropy metrics for the distances between the densities of the two groups.

Lastly, robustness analyses are performed using data on the circumstances of the individuals, followed by measurement and comparison of the results related to non-parametric inequality of opportunities in Brazil.

Step 1. The analyses of the income equation specifications are summarized in table 1. As noted above, this involves comparing the parametric and non-parametric estimations of the income equations. This

procedure avoids making an ad hoc decision to adopt an estimation structure. In short, the analysis undertaken by Racine (2006) demonstrates the superiority of the non-parametric models in all years investigated.

Step 2. Once the superiority of the non-parametric models has been demonstrated, the study makes inferences based on the constant kernel estimator.¹¹ After obtaining the fitted values for each equation, the high- and low-effort groups are identified according to the categories indicated in section II. Figure 1 displays the estimations of the kernel densities for the fitted values by effort group in 1995 and 2009.

A visual analysis shows that the education curves moved closer together in terms of inequality, and there was less dispersion in the distribution of individuals in the high-effort category. In the case of migration, the difference between the groups, both in 1995 and in 2009, is relatively insignificant and hardly varies from one year to another.

In the case of labour-market status, the distance between the density functions increases, while the dispersion of the curve corresponding to the low-effort group decreases slightly. Analogously, the difference between the distributions attributed to the hours-worked factor grows in the period, without major alterations to the shape of the curves.

Nonetheless, the distances between the curves synthesized in figure 1 need to be formally quantified, and their trend during the period analysed needs to be identified. Figure 2 shows the evolution of the distances between the fitted values of a high and low effort groups, for each effort dimension, over the period considered.¹² Any measure of entropy is useful as an indicator of divergence between distributions and thus serves as a metric of their inequality or concentration.

When considering education as an effort factor, the values of the entropies decline during the period as a whole. From 1995 to 2009, the distance between the groups, which had been slightly more than 0.50 in 1997, dropped to a level close to 0.36 in 2009. This shows that effort, measured in terms of a higher educational level, has become less important in individual income differentials. In other words, years of schooling in 1995 (entropy = 0.41) involved a higher level of income than in 2009 (entropy = 0.35) relative to individuals who had a lower level of education.

In terms of migration, the values of the entropies are lower than those of the other effort variables, which suggests that this variable has little explanatory power. Thus, the fact that an individual is a migrant does not imply a significant improvement in his or her income.

An explanation for the weak effect of migration can be found in the study by Assis, Costa and Mariano (2012), which uses data on non-native migrants, returning migrants, and non-migrants in the states of São Paulo and Bahia, representing the south east and north-east regions, respectively. Their results show that migrants from the north-east region, particularly returnees, have higher income levels than non-migrants. The opposite is the case in São Paulo, however, because the migrants of this state are individuals who have failed to find a place in a more dynamic local market. Accordingly, as this study used data for Brazil as a whole, those effects may have cancelled each other out; and, migration has not been a factor associated with a significant wage difference between the groups at the national level.

Moreover, during the period, the distances (S_p) increased for the variables of labour-market status and weekly hours worked. The values of the entropies for hours worked were higher than those associated with labour-market status in all years, which shows that hours worked explains more of the income inequality between the groups than labour-market status. The trend of the entropies through time reveals a considerable variation in the distances for both variables with respect to the initial year.

¹¹ The results of the estimations were excluded from the text owing to space constraints. They can be requested from the corresponding author by e-mail.

¹² The values of the entropies for each year of the series are shown in table A.1 of the annex to this article.

TABLE 1

Analysis of parametric specifications

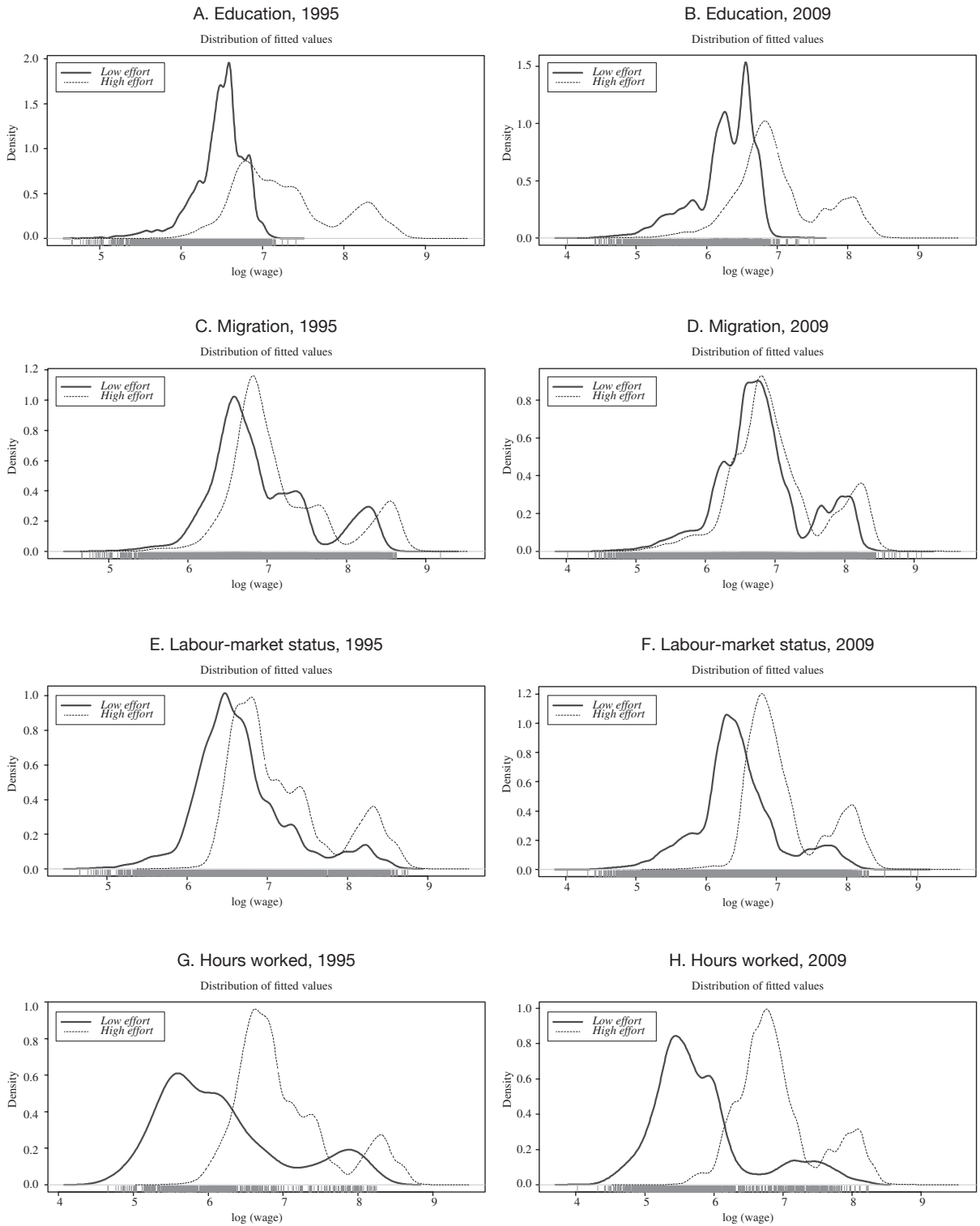
	H ₀ : perfect parametric specification														
Years	1995	1996	1997	1998	1999	2001	2002	2003	2004	2005	2006	2007	2008	2009	
Statistic	34.45*	33.21*	47.62*	64.03*	66.43*	52.54*	66.76*	16.01**	32.43*	45.09*	45.06*	65.92*	15.43**	54.10*	

Source: prepared by the authors.

Note: * and ** represent rejection of the null hypothesis at the 1% and 5% significance level, respectively.

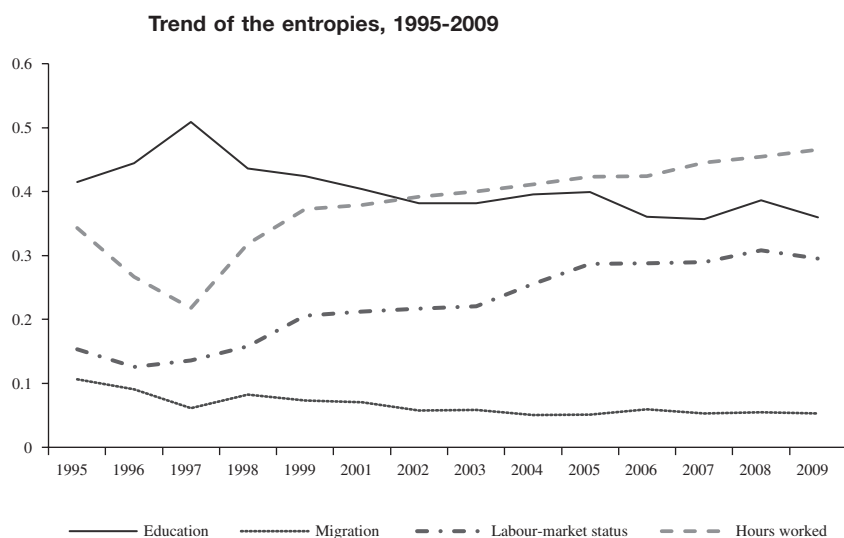
FIGURE 1

Brazil: non-parametric densities for the fitted values by effort factors, 1995 and 2009



Source: prepared by the authors.

FIGURE 2



Source: prepared by the authors.

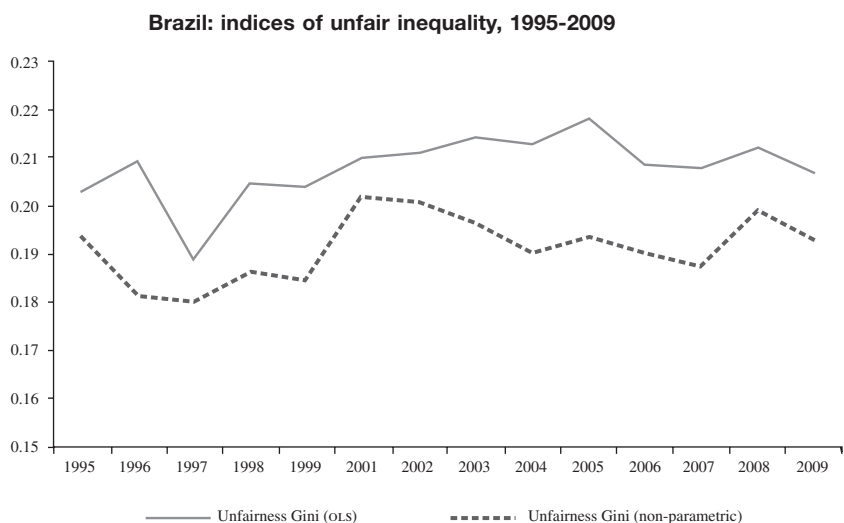
Step 3. The main question in this stage is what impact does the use of non-parametric inferences have on the fairness indicators. Figueiredo and Netto Júnior (2014) propose a method for estimating the indices of unfair inequality in the period between 1995 and 2009. The authors basically use parametric models, although they address issues related to the endogeneity of the equations and the identification of the parameters. It is beyond the scope of this study to propose more robust estimations for the measures of unfair inequality, since topical inferences, parametric or otherwise, can have biases.

Accordingly, only the comparison of the unfair inequality indices is presented here (see figure 3), calculated on the basis of a parametric specification with these same indicators calculated using the non-parametric method.¹³

It should be noted that the unfair inequality index calculated from specification (2) is always lower than the

¹³ See Almás (2008) and Almás and others (2011) for information on the development of the fairness criterion and the unfair inequality index.

FIGURE 3



Source: prepared by the authors.

value calculated by ordinary least squares (OLS), although its behaviour is similar. As there is an average difference of 10% between these inferences, the parametric method tends to overestimate the unfair inequality index.

Step 4. The results shown thus far depend on a fundamental assumption: independence between the responsibility and circumstantial variables. In other words, it is postulated that hours of work, education, the decision to migrate, and labour-market status are not affected by the circumstantial variables, such as family background, gender or race.

That assumption is often made in the bibliography on unequal opportunities (see O'Neill, Sweetman and Van de Gaer, 2001, and Checchi and Peragine, 2009). Nonetheless, recent developments in certain theories of justice alert to the possibility that responsibility factors and non-responsibility factors are not independent (Betts and Roemer, 2005). The study by Bourguignon, Ferreira and Menéndez (2007) is a key empirical reference for that hypothesis. With data on Brazil, these authors show that circumstances have direct and indirect effects on total inequality. The indirect effect reflects the influence of circumstances on effort. In short, circumstances are responsible for about 23% of total inequality: 13 percentage points through the direct effect and 10 percentage points through the indirect effect.

It is therefore possible that the results obtained up to that point in time, were affected by the hypothesis of independence between the two sets of variables. Accordingly, a robustness analysis needs to be performed, based on a simple procedure: the effect of circumstances on effort needs to be isolated, particularly in the years of the study; and whether the omission of circumstantial variables affects the entropies also needs to be analysed.¹⁴

For the robustness analyses, an equation is estimated in which years of schooling are explained by a set of variables that are outside the scope of individual responsibility. This provides the total number of years of schooling without the effect of circumstances (years of schooling minus years of study predicted by the variables); the entropy distances between the low- and high-effort groups are calculated, considering this new

variable; and, lastly, a test is conducted to see whether the distance between the two groups is significant.

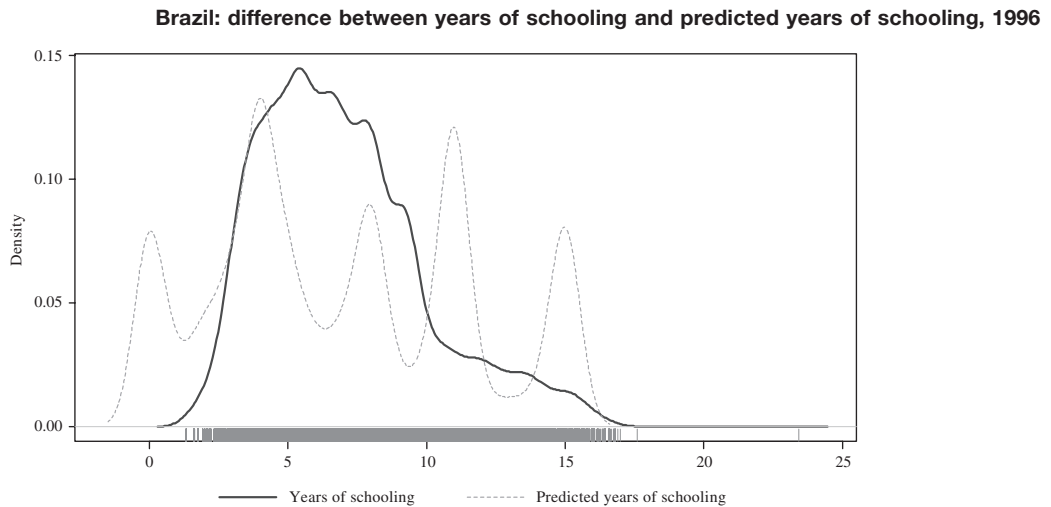
The data used in the estimations come from the social supplement of the 1996 PNAD, which summarizes information relating to the circumstances of the individuals, such as education, schooling, and the occupation of the parents of the selected persons (family background). For these years, the effort variables described above are used, along with a vector of circumstantial variables, defined by: (i) a dummy race variable, which takes the value 1 in the case of non-white persons (black, mixed race and indigenous) and zero for whites (including Asiatics); (ii) a dummy variable for the regions, which takes the value 1 if the individual lives in the most dynamic regions of the country (south, south-east and centre-west) and zero otherwise (north and north-east regions); (iii) the education of the father and mother, expressed in years of schooling, and (iv) the father's type of occupation, according to the six categories proposed by Pastore and Silva (2000).¹⁵

Figure 4 shows the difference between actual and predicted years of schooling (the latter defined as years of schooling controlled for the effect of circumstances). The entropy is calculated on the basis of the creation of the low- and high-effort groups in the variable "predicted years of schooling". Low-effort is assigned to individuals with less than five years of schooling predicted (including the two first modes of the discontinuous plot curve) and high-effort for other individuals. The result of the entropy (0.4023) shows that there has been no significant reduction, compared to the entropy for 1996, without considering the circumstantial variables (entropy = 0.4447). In that regard, while bearing in mind that the two sets of variables, effort and circumstances, are not independent, there is no clear change in the distance between the two effort groups, thereby demonstrating the plausibility of the results described in steps 1 to 3.

¹⁴ This procedure was also used for the other effort variables. The results were similar to those of the "years of schooling" variable, and were eliminated owing to space constraints.

¹⁵ Categories: (1) Lower low level: farm labourer, fisherman, or woodcutter, among others; (2) Upper low level: manual labourer, domestic employee, or security guard, among others; (3) Lower middle level: bricklayer, electrician or carpenter, among others; (4) Middle level: street trader, travelling salesperson or first level manager, among others; (5) Upper middle level; consultant, senior manager or director, among others, and (6) High level: large-scale proprietor, judge or other higher-level occupation.

FIGURE 4



Source: prepared by the authors.

IV

Discussion of the results

As noted above, the entropy metrics represent a measure of the inequality of effort between individuals. By isolating the effect of each variable, the aim is to test its power to explain the income differentials between the groups, as quantified by the distance between the distributions.

Some of the results obtained should be highlighted for their theoretical and practical implications. One of these is the reduction in inequality when education is considered as an effort factor. The role of education in the income distribution can help expand or narrow gaps between individuals, depending on its rate of return and the composition of the population by education level. When describing the structure of income inequality in Brazil, Ferreira and others (2006) suggest that the reduction in total income inequality attributable to education differences reflects the rising education level of the labour force. The increase in the relative supply of individuals with primary and secondary education probably caused a reduction in the average rate of return to education, thereby reducing the wage differential between such persons and those with less education (Menezes-Filho, 2001).

Another factor that could explain that result is the expansion of social income-transfer programmes and the fact that these have targeted the poorest individuals. As many individuals in that segment of the population have

few years of schooling, incomes in the low-effort group were raised during the period, so the distance from the high-effort group (higher level of schooling) narrowed. Another possible explanation for the narrowing of the distance between the curves could relate to a fall in the quality of education. Although educational policies have made it possible to increase the population's average number of years schooling, this does not necessarily guarantee higher skill or, therefore, significant wage increases.

The trend of labour-market status reveals an increase in income differentials between formal and informal workers. It is generally believed that the formal sector offers higher wages because of its intrinsic characteristics. Nonetheless, studies such as those of Menezes-Filho, Mendes and De Almeida (2004) provide clear evidence of a self-selection bias, suggesting that the reason why the formal sector offers better wages relates primarily to unobservable individual attributes of the workers themselves.

The results obtained can also be analysed in the light of a number of government policies, including those that have promoted social transfer programmes, such as the *Bolsa Família* family subsidy program, and mechanisms for raising the minimum wage, used as tools to reduce income disparities and poverty. Nonetheless,

despite the reduction in inequality in Brazil in recent years, some studies question the effective influence of the succession of minimum wage hikes that have been implemented by the government (Barros, Carvalho and Franco, 2006) and the contribution made by the *Bolsa Família* program (Soares, 2006).

Since the implementation of the Real Plan, the minimum wage has been increased nearly every year. Nonetheless, these real increases have proven ineffective in terms of redistribution, since most individuals who earn the minimum wage are not considered poor under the Brazilian the poverty criterion (Saboia, 2006). An example of this situation is provided by the north-east region, where a large proportion of workers earn incomes below the minimum wage and, therefore, do not benefit when it is raised (Institute of Applied Economic Research-IPEA). According to Giambiagi and Franco (2007), the average real income of workers fell in that period.

Between 1995 and 2009, the minimum wage rose by more than productivity growth (increase in per capita GDP) nearly every year. In 1997, it was raised by 7%, whereas productivity grew by around 3%; and the entropy value with respect to hours worked, as an effort factor, was lowest in that year. That confirms the belief that, given the circumstances, there are few incentives to increase skills and improve labour-market status.

Nonetheless, the minimum wage policy does have a significant effect on the labour market, but most of all in the case of workers in the informal sector of the economy (Menezes-Filho, Rodrigues and De Souza, 2009). In short, while minimum-wage hikes can help reduce poverty, it is also possible to discern negative effects in the labour market when the increases outpace wage-earners' productivity. Maloney (2000) states that the minimum wage is also a significant determinant of employment and wages in the informal sector in most Latin American economies.

V

Final thoughts

This article has analysed the trend of the inequality of opportunities in Brazil between 1995 and 2009, in relation to a set of effort variables. For that purpose, data for the period from the National Household Surveys (PNADS) have been analysed, following a non-parametric approach presented in Racine and Li (2004), both to estimate the income function and to determine the distribution of income between individuals. Two groups were created, one of low effort and another of high effort; and inequality between the two groups was evaluated by using metrics of the entropy distance between the two income levels.

The results reveal a reduction in the explanatory power of years of schooling in the income differential

between the groups, and a very weak participation by migration. The robustness analysis for 1996, based on information from that year's PNAD social supplement, revealed that the omission of circumstantial variables for other years, such as family background, did not significantly change the results, thereby corroborating the plausibility of the strategy used. Future improvements to this study could include: an analysis of the true reasons why the education has lost capacity to explain income inequality; an exploration of optimal policies for reducing inequality of opportunities and stimulating differences through effort; and more in-depth discussion of fairness criteria as applied to Brazil.

ANNEX

TABLE A.1

Entropies

Years	Effort factors			
	Education	Migration	Labour-market status	Hours worked
1995	0.4151*	0.1061*	0.1535*	0.3428*
1996	0.4447*	0.0905*	0.1253*	0.2664*
1997	0.5094*	0.0616*	0.1361*	0.2175*
1998	0.4358*	0.0825*	0.1580*	0.3179*
1999	0.4240*	0.0736*	0.2055*	0.3726*
2001	0.4039*	0.0701*	0.2122*	0.3795*
2002	0.3816*	0.0580*	0.2171*	0.3916*
2003	0.3814*	0.0588*	0.2210*	0.3999*
2004	0.3954*	0.0505*	0.2554*	0.4109*
2005	0.3994*	0.0510*	0.2874*	0.4234*
2006	0.3602*	0.0596*	0.2880*	0.4246*
2007	0.3567*	0.0534*	0.2898*	0.4456*
2008	0.3865*	0.0548*	0.3085*	0.4546*
2009	0.3594*	0.0530*	0.2955*	0.4654*

Source: prepared by the authors.

Note: * indicates values that are significant at 1%.

Bibliography

- Almås, I. (2008), "Equalizing income versus equalizing opportunity: a comparison of the United States and Germany", *Research on Economic Inequality*, vol. 16, Emerald.
- Almås, I. and others (2011), "Measuring unfair (in)equality", *Journal of Public Economics*, vol. 95, No. 7-8, Amsterdam, Elsevier.
- Arneson, R. (1989), "Equality and equal opportunity for welfare", *Philosophical Studies*, vol. 56, No. 1, Springer.
- Assis, R.S. de, E.M. Costa and J.L. Mariano (2012), "Impacto da migração de não naturais e da migração de retorno sobre a distribuição de renda dos estados da Bahia e de São Paulo: um olhar sobre a inserção desses indivíduos no mercado de trabalho local", *40º Encontro Nacional de Economia* [online] <http://www.anpec.org.br/novosite/br/encontro-2012>.
- Azevedo, J. (2007), "Avaliando a significância estatística da queda na desigualdade no Brasil", *Desigualdade de renda no Brasil: uma análise da queda recente*, R. Barros, M. Foguel and G. Ullysea (orgs.), Brasília, Institute of Applied Economic Research (IPEA).
- Barros, R.P. de, M. Carvalho and S. Franco (2006), "A efetividade do salário mínimo como um instrumento para reduzir a pobreza no Brasil", *Boletim de Conjuntura*, No. 74, Institute of Applied Economic Research (IPEA).
- Betts, J. and J. Roemer (2005), "Equalizing Opportunity for Racial and Socioeconomic Groups in the United States through Educational Finance Reform" [online] <http://www.escholarship.org/uc/item/0gq4z4m9#page-3>.
- Bourguignon, F., F. Ferreira and M. Menéndez (2007), "Inequality of opportunity in Brazil", *Review of Income and Wealth*, vol. 53, No. 4, Wiley.
- Checchi, D. and V. Peragine (2009), "Regional disparities and inequality of opportunity: the case of Italy", *Journal of Economic Inequality*.
- Devooght, K. (2008), "To each the same and to each his own: a proposal to measure responsibility-sensitive income inequality", *Economica*, vol. 75, No. 298.
- Dworkin, R. (1981), "What is equality? Part 2: equality of resources", *Philosophy and Public Affairs*, vol. 10, No. 4, Princeton, Princeton University Press.
- Ferreira, F. and others (2006), "Ascensão e queda da desigualdade de renda no Brasil", *Econômica*, vol. 8, No. 1.
- Figueiredo, E. and J.L. Netto Júnior (2014), "More equal but not so fair: an analysis of Brazilian income distribution from 1995 to 2009", *Empirical Economics*, vol. 46, No. 4, Springer.
- Fleurbaey, M. (1998), "Equality among responsible individuals", *Freedom in Economics: New Perspectives in Normative Analysis*, J.-F. Laslier and others (eds.), London, Routledge.
- Fleurbaey, M. and F. Maniquet (2011), "Compensation and responsibility", *Handbook of Social Choice and Welfare*, vol. II, K. Arrow, S. Sen and K. Suzumura (eds.), Amsterdam, Elsevier.
- Giambiagi, F. and S. Franco (2007), "O esgotamento do papel do salário mínimo como mecanismo de combate à pobreza extrema", *Texto para Discussão*, No. 1290, Rio de Janeiro, Institute of Applied Economic Research (IPEA).
- Granger, C., E. Maasoumi and J. Racine (2004), "A dependence metric for possibly nonlinear processes", *Journal of Time Series Analysis*, vol. 25, No. 5, Wiley.
- Härdle, W. (1990), *Applied Nonparametric Regression*, Cambridge, Cambridge University Press.
- Kranich, L. (1996), "Equitable opportunities: an axiomatic approach", *Journal of Economic Theory*, vol. 71, No. 1, Amsterdam, Elsevier.
- Li, Q. and J. Racine (2007), *Nonparametric Econometrics: Theory and Practice*, Princeton, Princeton University Press.
- Li, Q., E. Maasoumi and J.S. Racine (2009), "A nonparametric test for equality of distributions with mixed categorical and continuous data", *Journal of Econometrics*, vol. 148, No. 2, Amsterdam, Elsevier.
- Maasoumi, E., J. Racine and T. Stengos (2007), "Growth and convergence: a profile of distribution dynamics and mobility", *Journal of Econometrics*, vol. 136, No. 2, Amsterdam, Elsevier.
- Machado de Assis, J. (1891), *Quincas Borba*, Rio de Janeiro, B.L. Garnier, Livrero-Editor.
- Maloney, W.F. (2000), "A note on minimum wages in Latin America", Washington, D.C., World Bank.
- Menezes-Filho, N. (2001), "A evolução da educação no Brasil e seu impacto no mercado de trabalho", Instituto Futuro Brasil, São

- Paulo, March [online] <http://www.anj.org.br/pje/biblioteca/publicacoes/A%20Evolucao%20da%20educacao%20no%20Brasil%20e%20seu%20impacto%20no%20Mercado%20de%20trabalho.pdf>.
- Menezes-Filho, N., R. Fernandes and P. Picchetti (2006), "Rising human capital but constant inequality: the education composition effect in Brazil", *Revista Brasileira de Economia*, vol. 60, No. 4, Rio de Janeiro.
- Menezes-Filho, N., M. Mendes and S.E. de Almeida (2004), "O diferencial de salários formal-informal no Brasil: segmentação ou viés de seleção?", *Revista Brasileira de Economia*, vol. 58, No. 2, Rio de Janeiro.
- Menezes-Filho, N., E. Rodrigues and A. de Souza (2009), "Salário mínimo e desigualdade no Brasil entre 1981-1999: uma abordagem semiparamétrica", *Revista Brasileira de Economia*, vol. 63, No. 3, Rio de Janeiro.
- O'Neill, D., O. Sweetman and D. Van de Gaer (2001), "Equality of opportunity and kernel density estimation: an application to intergenerational mobility", *Applying Kernel and Nonparametric Estimation to Economic Topics*, vol. 14, Stanford, JAI Press.
- Ooghe, E., E. Schokkaert and D. Van de Gaer (2007), "Equality of opportunity versus equality of opportunity sets", *Social Choice and Welfare*, vol. 28, No. 2, Springer.
- Pastore, J. and N. Silva (2000), *Mobilidade social no Brasil*, São Paulo, Macron Books.
- Pattanaik, P. and Y. Xu (1990), "On ranking opportunity sets in terms of freedom of choice", *Recherches Économiques de Louvain*, vol. 56, No. 3-4, Catholic University of Louvain.
- Racine, J. (2006), "Consistent specification testing of heteroskedastic parametric regression quantile models with mixed data", unpublished.
- Racine, J.S. and Q. Li (2004), "Nonparametric estimation of regression functions with both categorical and continuous data", *Journal of Econometrics*, vol. 119, No. 1, Amsterdam, Elsevier.
- Roemer, J. (1998), *Equality of Opportunity*, New York, Harvard University Press.
- _____ (1996), *Theories of Distributive Justice*, Cambridge, Massachusetts, Harvard University Press.
- Saboia, J. (2006), "Salário mínimo e combate à pobreza", *Valor Econômico*.
- Salvi, A. (2007), "An empirical approach to the measurement of equality of opportunity", Milan, Università degli Studi di Milano, unpublished.
- Soares, S. (2006), "Distribuição de renda no Brasil de 1976 a 2004 com ênfase no período entre 2001 e 2004", *Texto para discussão*, No. 1166, Rio de Janeiro, Institute of Applied Economic Research (IPEA).
- Thomson, W. (2011), "Fair allocation rules", *Handbook of Social Choice and Welfare*, vol. 2, K. Arrow, A. Sen and K. Suzumura (eds.), Amsterdam, Elsevier.