

de Albuquerque e Arraes, Ronaldo; Souza Menezes, Francisca Livia;
Simonassi, Andrei Gomes

Article

Earning differentials by occupational categories: Gender, race and regions

EconomiA

Provided in Cooperation with:

The Brazilian Association of Postgraduate Programs in Economics (ANPEC), Rio de Janeiro

Suggested Citation: de Albuquerque e Arraes, Ronaldo; Souza Menezes, Francisca Livia;
Simonassi, Andrei Gomes (2014) : Earning differentials by occupational categories: Gender,
race and regions, EconomiA, ISSN 1517-7580, Elsevier, Amsterdam, Vol. 15, Iss. 3, pp.
363-386,
<https://doi.org/10.1016/j.econ.2014.07.001>

This Version is available at:

<https://hdl.handle.net/10419/179582>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by-nc-nd/3.0/>

Earning differentials by occupational categories: Gender, race and regions[☆]

Ronaldo de Albuquerque e Arraes^{a,*}, Francisca Lívia Souza Menezes^b,
Andrei Gomes Simonassi^a

^a Graduate Program in Economics, Federal University of Ceará, Brazil

^b Department of Economics, University of Leicester, United Kingdom

Available online 21 July 2014

Abstract

Analyzing the income differentials amongst the Brazilian workers' occupations is the focus of this paper. Due to the wide diversity of occupations cataloged by the IBGE (around 800), a theoretical procedure is applied to reduce them to only seven in order to allow statistical treatment of the data. The methodological approach is based on Mincerian quantile equations to be estimated in various strata of the workers' income distribution, on which a breakdown is made to check the gap among the individuals' income with distinct inherent attributes and between those living in both more and less developed regions. The estimation results ensure the importance of breakdown analysis for occupational strata as well as for quantiles, since the effects of explanatory variables are distinct along the income distribution and vary among occupations. Regarding the gaps in this distribution, there is a glass ceiling effect in some occupations, that is, the gap is greater at the top of distribution, although in most cases the wider gap occurs at the bottom of the distribution – sticky floor effect, which makes this a distinguished result from other studies. Moreover, contrary to the gap in gender, which is due to the compensation characteristics of individuals only (discrimination), income differentials between regions and races also occur from the heterogeneous characteristics of workers.

© 2014 National Association of Postgraduate Centers in Economics, ANPEC. Production and hosting by Elsevier B.V. All rights reserved.

JEL classification: J31; J44; J71; C21

Keywords: Earning differentials in occupations; Counterfactual decomposition; Mincerian quantile equations; Discrimination

Resumo

Analisar os diferenciais de rendimentos nas ocupações dos trabalhadores brasileiros é o foco central deste trabalho. Em virtude da vasta diversidade de ocupações catalogadas pelo IBGE (cerca de 800), aplica-se um procedimento teórico para reduzi-las em apenas sete a fim de permitir tratamento estatístico dos dados. São estimadas equações mincerianas quantílicas, sobre as quais são feitas decomposições para verificar o hiato entre os rendimentos dos indivíduos com atributos inerentes distintos, bem como entre aqueles residentes em regiões mais e menos desenvolvidas do país. Os resultados das estimações ratificam a importância da análise desagregada em estratos ocupacionais e por quantis, uma vez que os efeitos das variáveis explicativas são distintos ao longo das distribuições de rendimentos e entre as ocupações. Em relação às lacunas existentes entre as distribuições de rendimentos, verifica-se

[☆] We thank CNPq for the financial support.

* Corresponding author.

E-mail address: ronaldo@caen.ufc.br (R.A. Arraes).

Peer review under responsibility of National Association of Postgraduate Centers in Economics, ANPEC.

um *glass ceiling effect* em algumas ocupações, ou seja, a defasagem é maior no topo da distribuição, embora na maioria dos casos, o hiato seja mais acentuado na base da distribuição – *sticky floor effect*, resultado este que se distingue de outros trabalhos. Além disso, diferente das lacunas entre gêneros, que são decorrentes apenas da remuneração às características dos indivíduos (discriminação), os diferenciais de rendimentos entre raças e regiões também decorrem das diferentes características dos trabalhadores.

© 2014 National Association of Postgraduate Centers in Economics, ANPEC. Production and hosting by Elsevier B.V. All rights reserved.

Palavras-chave: Diferencial de Renda em Ocupações; Decomposição Contrafactual; Equações Mincerianas Quantílicas; Discriminação

1. Introduction

Economic growth and the increase in demand for jobs have generated significant changes in the Brazilian job market dynamics. Besides that, the recent phenomenon of income inequality reduction has a direct implication for the transformations that have occurred in the demand and supply sides of the market occupational distribution (Barros et al., 2007).

However, age long problems such as salary inequalities between individuals who work at similar activities and also have the same qualifications are actually still observable. For instance, the salary gap between men and women, whites and non-whites, and residents of more and less developed regions. In some cases, acquired attributes such as education level and work experience seem to have less relevance than a worker's gender or race. In order to illustrate, data from PNAD (Brazil's National Household Survey) of 2011 reveal that for individuals 10 years and older, the averages of education years of women and men are respectively 7.5 and 7.1, which means a 5.6% superiority in favor of women. On the other hand, men's average monthly earnings are 73.7% higher than that for women, that is, R\$1129.00 and R\$650.00, respectively.

So, it becomes relevant to investigate how education and other personal attributes have affected the Brazilian worker. It is also believed to be indispensable for such analysis to take into account not only distinctions between inherent attributes and regionalization, but also differences among the diverse occupational categories due to requirements on each of them regarding human capital level.

It is usual to get theoretical support based on Mincerian equations for empirical application of such nature that spot the effects of explanatory factors on the average variation of worker's income. Worth reminding is that the impact of every factor can significantly vary among the income distribution strata. It is expected, for instance, that effects of inherent characteristics such as ethnicity and gender, or acquired characteristics such as human capital are differentiated according to occupation and income stratification. Therefore, the results obtained by applying this methodological technique on non-stratified samples, as is often applied in literature, produce inconsistent and distorted effects.

Based on this, quantile Mincer equations are estimated to verify the impact of explanatory variables on every income distribution quantile, allowing for the occurrence of differentiated effects when stratifying workers' income distribution by occupation. In proceeding so, an investigation is conducted within each quantile about the lag between workers' income caused by gender, and region differences, according to the methodology by Machado and Mata (2005), which expands the formulation of the Blinder–Oaxaca decomposition (Blinder, 1973; Oaxaca, 1973). For this purpose, the conditional wage distributions of these categories are estimated via quantile equations, marginal distributions are obtained, when comparisons are made between the above mentioned categories, they make it possible to decompose the income distributions into two components; one that comes from observable characteristics of workers and another from the returns to these characteristics, which may be interpreted as discrimination.

The main methodological hurdle for the actualization of the estimations is the considerable number of more than eight hundred occupational categories designated by The Brazilian Institute of Geography and Statistics Foundation (IBGE). In order to overcome such hurdles and make estimations possible, it is vital to implement a special handling on the data so as to create occupation groups and then have a statistically practicable numbers that lead to inferences of greater degree of freedom and accuracy, despite the possibility of information loss.

The seminal contribution to the investigation of occupational choices is believed to have been made by Roy (1951), who identified in such choices, and the distribution of the return of these skills for every occupation. In order to achieve continuity of this analysis through the application of adequate inference methods with discrete choices, some researchers suggested the grouping of those numerous occupational categories in smaller strata. Since we have applied

the introduction of working categories in the methodology of earnings decomposition in this paper, further comments on this are noteworthy.

One of the earliest proposals was made by Erickson et al. (1979) which was later honed by Goldthorpe et al. (1987), who designed a scheme known as EGP (initials of the previous authors), as well as Erickson and Goldthorpe (1987, 1992) who proposed the CASMIN project (Comparative Analysis of Social Mobility in Industrial Nations). The CASMIN and EGP models comprise all working categories into seven broad occupational classes, whose purpose is centered on the analysis of labor mobility in the industrial realities of various European Union countries. Although Hout and Hauser (1992) recognize that the CASMIN model treats the main aspects of labor mobility related to hierarchy, inheritance and labor sectors parameters on solid theoretical rationale, they make some criticisms for not detailing as much information as possible about the pattern of social mobility, such as the status and prestige of the working category from the origin to the destination as well as the distance between origin and destination on the rate of mobility. They end up breaking each of the before five categories into two others which accounted for a total of twelve working classes. Since the reasoning for this is solely to explain mobility rate, which is not undertaken here, we kept the EGP/CASMIN classification in this paper.

Many other grouping formulations came up, but all of them are like to EGP proposal. Silva and Roditi (1988) formulation stands out for incorporating significant insights for the identification of similarities and differences between categories, such as the dichotomous rural/urban sectors and manual/non-manual labor, so he grouped the numerous categories into eighteen sets. Nevertheless, such scheme has not been widely used due to its analytical difficulty, so researchers have been opting for the implementation of a grouping proposal which contains a lower number of categories.

It is expected that this study sheds some light on the debate about this subject as we proceed an analysis of the earning distributions in the quantiles among the established varieties of occupations by individual characteristics, and identify the factors that affect them. To implement the EGP scheme, firstly a recoding of the PNAD (Brazil's National Household Sampling Survey) data is done in order to group those various occupational categories into just seven sets, and then an analysis of the Brazilian job market by occupational categories is performed. Such analysis is approached by the estimation of Mincerian quantile equations and by identifying and decomposing the existing hiatus between men's income and women's incomes, whites and non-whites, individuals who reside in the Northeast region or in another more-developed region (namely the South, the Southeast or the Midwest). Besides verifying the strata in which the applied explanatory variables affect the income distribution, this procedure makes possible the detection of distortions that may occur between personal attribute effects and in addition it also detects the existence of likely effects such as *glass ceiling* or *sticky floor*, as previously seen in economic literature. Once the distribution strata that are affected by each factor are identified, it becomes less difficult to implement policies that aim to reduce income inequalities, which leads to the improvement of workers' well-being.

This article comprises four additional sections. In the next section, one can see the literature on the topic, and subsequently, both methodology and results are shown. Then finally, the main conclusions are stated.

2. A selective review of literature: Mincerian equations, decomposition of income distribution and occupations

In the 20th century, Schultz (1961) and Ben-Porath (1967) formulated, despite the lack of theoretical strictness, the relation between education and income. However, the debates and the empirical applications were intensified after the publishing of the seminal article by Mincer (1974), which formalized the theory of the derivation of such relation whose equation specifies the worker's income as a function of his or her formal education and experience acquired on the job. A further comment on this is worthy to note.

Mincer developed a theoretical equation from the simple optimization of future income of the individual taking into account the stock of human capital acquired (ed) and experience (xp), resulting in the world-famous Mincerian wage equation. Nevertheless, Mincer himself, based on empirical observations, tried several alternative specifications of this equation involving ed and xp and concluded that it is impossible to establish which one would be the best.¹ In

¹ For instance, he estimated an alternative equation by introducing the variables ed^2 and $ed * xp$ to test the linearity of the return to education and the trend of income with educational levels, whose results are: $\ln w = 4.87 + 0.255ed - 0.0029ed^2 - 0.0043ed * xp + 0.148xp - 0.0018xp^2$. Unexpectedly, the returns to schooling decreases both with the level of education and with experience in the labor market.

this regard, [Murphy and Welch \(1990\)](#), [Heckman et al. \(2006\)](#) and [Lemieux \(2006\)](#), launch critiques to the quadratic specification in the Mincerian equation, and test adjustments for polynomials of higher degrees based on scattering data. They conclude that the quartic specification fits better than the quadratic. On the other hand, [Polachek \(2008, p. 188\)](#) is incisive with respect to education: The positive correlation between earnings and schooling is so evident in the literature that one cannot do justice to the rate of return. For him, this is based on surveys carried out by too many authors, and states that the main stream of results was carried out for the original Mincer equation (*ed* linear and *xp* quadratic) based upon 70 countries and estimates for over 25 years. All of them confirm positive coefficients for *ed* and *xp*, but negative for xp^2 . Since our work relies on measuring and qualifying discretionary effects in the labor market, there is no reason to deviate from the mainstream of the original specification to perform such tests. Thus, testing the degree of the polynomial for increasing/decreasing returns to education is a specific goal pursued by others, such as [Dias et al. \(2013\)](#) who specify a cubic effect of education on earnings.

It is worth mentioning that most applications use a simplified version of the Mincer model in which the rigidity hypothesis is assumed; that is, returns to education and experience are invariable among individuals. However, the most general of his model, and the most realistic one, acknowledges the possibility of such returns to vary among individuals. And that leads to a random-coefficient model ([Harmon et al., 2003](#)). So, when one stratifies earning distribution as it is done herein, this is the model to be applied.

In this sense, this paper relies on equations in that methodological structure, despite having a focus that differs from the focus usually seen in literature. For each occupational stratum, analyses of the Brazilian workers' income distribution and of the factors that cause changes in such distribution are conducted. In order to achieve that, initially, quantile Mincer equations are estimated so as to analyze the effects of variables like education and experience on the workers' income distribution. From these results, conditional distributions are obtained from which in turn, the marginals will be extracted. This way, it becomes possible to decompose the workers' income distributions into various components.

This sort of analysis was initially due to [Oaxaca \(1973\)](#) and [Blinder \(1973\)](#), who decomposed the income distributions of males and females into a component that is attributed to distinct individual characteristics, and into another component derived from the differences between the returns to these characteristics. Such decomposition, however, was conducted via the Least Squares (LS) method, only at the mean of the distribution. Subsequent studies proposed to apply quantile regressions in order to perform such analysis, thus providing information pertaining not only at the income distribution center, but also to substrates of it.

A similar procedure was applied by [Blau and Kahn \(1996\)](#) to investigate the income inequalities in the U.S. in comparison to nine OECD countries, with data from the 1980 decade. Through the estimation of the income densities of male workers, it is concluded that salary inequality in the U.S. is larger than in all other OECD countries. However, after disaggregating the income distributions into several parts, they verified that inequality in the United States is considerably larger than in the other countries in the lower quantiles, although smaller in the upper quantiles.

[Dinardo et al. \(1996\)](#) applied nonparametric methods to measure the effects of institutions and labor market factors on the changes of income distribution in the U.S. in the period 1979–1988. They concluded that labor unions and supply/demand shocks are vital to explain the income inequality rise in that period. Besides, there was also evidence that the decline of the minimum wage contributed mostly to such inequality increase, especially due to the gender factor.

Based upon the Blinder–Oaxaca decomposition with the quantile regressions procedure, [Albrecht et al. \(2003\)](#) found in the Swedish workers' income distribution what is conventionally named in literature as the *glass ceiling effect*. This means that the difference between men's and women's income is positive, it grows along the distribution and intensifies at the top.

The technique proposed by [Machado and Mata \(2005\)](#) has been applied in some recent articles. It utilizes quantile regressions in order to identify and decompose income gaps among distinct categories of works at different points in time. Their aim was to verify the changes in income distribution in Portugal regarding the period 1986–1995, and concluded that the changes of the individuals' attributes as well as the returns to such attributes do contribute to the increase of income inequality, besides emphasizing that education plays a decisive role to this increase.

This technique has been applied worldwide with distinguishing approaches. [Rica et al. \(2005\)](#) found a steep pattern of gender wage gap in Denmark, Sweden, United Kingdom, Italy and Greece in contrast with the flatter evolution of the gap in Spain, where for higher education range there is evidence for glass ceiling effect, while for lower education group

the gap is much higher at the bottom than at the top of the distribution. Albrecht et al. (2009) provided greater theoretical support to the methodology of Machado and Mata (2005) by evidencing the asymptotic consistency and normality regarding such technique, besides taking into consideration sample selection adjustments. The empirical application of such technique aimed to test the hiatus between genders among Dutch workers' income, and concluded for the presence of a *glass ceiling effect* on that distribution. Using these authors' methodology, Paredes (2012) also analyzes the gender discrimination in the Chilean economy by grouping up occupations in sets of nine,² and identifies that the income hiatus between men and women has been narrowed in categories such as managers, upscale self-employed professionals, and especially in occupations dealing with finances and insurance. Similarly for Brazil and using data from PNAD/1999, Santos and Ribeiro (2006) found in the aggregated data of PNAD/1999 a glass ceiling effect in the female earnings (top of the distribution), but was uncertain about a glass floor effect (bottom of the distribution). Coelho et al. (2010) have also identified such an effect through the earning distribution gap of black and white Brazilian women by performing quantile regression with PNAD/2007 data. Their main finding is that the returns to education are higher at the top of the distribution.

This work also follows the Methodology of Machado and Mata (2005) to Brazilian workers' income distributions, although it distinguishes from others by considering occupational groups which shall provide valuable information has more accurate results in the labor market. So it will benefit policies that aim to reduce salary inequalities as they might be driven to specific working categories.

The surveyed studies analyze the variations of salary distributions either in an aggregate way or considering gender differences. This research proposes an analysis that takes into account individuals' different occupations. The greatest difficulty present in works that focus on this approach is occupation definition, due to the large number of working classes. So it becomes necessary to aggregate them in order to accomplish a practicable analysis, even if that may eventually lead to the loss of some information. Thus, various researchers have proposed ways to aggregate such categories into smaller groups while simultaneously trying to avoid information loss.

The first versions of occupational distribution were proposed by Erickson et al. (1979) and improved by Goldthorpe et al. (1987), a scheme internationally known as EGP (first authors' initials), which is comprised of seven classes. Based on this aggregation, occupations were then grouped in several countries studied by the CASMIN Project (*Comparative Analysis of Social Mobility in Industrial Nations*), which was coordinated by Goldthorpe in the early 1980s.

This methodological scheme takes into account the work relations between employers and employees in both rural and urban activity sectors. According to Ribeiro and Scalón (2001) the resulting classes at the end of the aggregation are classes of comparable individuals in terms of their income level, financial safety degrees, career-advancement probabilities and these individuals' autonomy to perform work.

There are other class-aggregation issues which share some degree of similarity with the Goldthorpe's proposal. Although Pastore (1979) suggests the aggregation of the numerous categories to only six groups, he points out the heterogeneity inherent within each of them due to the reduced number of strata. Aiming to reduce such in-between group differences, Silva and Roditi (1988) proposed the use of eighteen categories, which was rejected by most researchers as they considered too many categories. The opposite happened to Scalón (1999) who rearranged Silva's and Roditi's (1988) eighteen categories into nine but managed to minimize the loss of information. To achieve this, Scalón applied cluster analysis techniques and also a criterion for combining categories based on log-linear models.

With the goal of verifying which are the determinants of the distinctions in Brazilian workers' earning distributions, we rather use the EGP scheme's theoretical principle, despite the critiques made by Hout and Hauser (1992) as already commented. To do so, a recoding of the PNAD data about occupations has to be done, so as to group more than eight hundred occupational category classifications into just seven, which are: professional managers, routine non-manual laborers, small business owners, rural owners/employers, skilled manual laborers such as technicians and supervisors, unskilled manual laborers, and rural-sector unskilled manual laborers. Contrast to most analyses found in literature, where job market approaches are done in an aggregate fashion or split by education groups, this research stands out in this respect for homogenizing the workers by peers in each occupational category strata, consequently, robustness is enhanced. This shall lead to inferences that produce more realistic results for implementing policies.

² The author makes no reference on the methodology for grouping the occupations.

3. Methodology

The methodological framework to be carried out comprises two procedures. In the first procedure the return to workers' education is measured through quantile Mincer equations which were estimated for each occupational stratum. The second procedure is a counterfactual decomposition of workers' earning, taking into account differences of gender, ethnicity and region. The data used and the variables that make up the models are to be described in the next sections.

3.1. Earning equations

Initially, the empirical exercise is the estimation of earning equations in the traditional way, although differentiated by the application of quantile regression model which is due to [Koenker and Bassett \(1978\)](#). This way, a detailed analysis of workers' income distribution can be conducted and it differs from the usual application of LS method in this kind of investigation which provides partial information of the response variable distribution. However, for developing countries like Brazil, where there is a marked heterogeneity among workers in terms of income and also inherent and acquired attributes, and the impact of individual characteristics on earnings is utterly significant, the analysis by quantile regressions is the most suitable one. This type of modeling is indeed suitable for deriving the conditioned income distributions to comply with the decomposition proposed by [Machado and Mata \(2005\)](#), which comes up next.

3.2. Decomposition of earning distribution

Counterfactual decomposition can be understood as a generalization of the [Blinder \(1973\)](#) and [Oaxaca \(1973\)](#) procedure, which the LS method is applied to obtain the estimations of the median of the distribution, but with the advantage of using quantile regressions that are modeled on each quantile of the distribution. This technique allows for the decomposition of the differences of income distributions of various categories. For instance, men and women become two components which is derived from differences between observable characteristic distributions (education, experience, etc.) for genders, and also derived from the returns of these characteristics. Therefore, decomposition makes possible to identify the existing gap between the incomes of distinct classes and also to measure how much of this gap is distributed as a consequence of characteristics and coefficients. The effect of this may be considered as discrimination between categories. This decomposition has some stages to be described as follows.

3.2.1. Conditioned income distributions

The first phase of decomposition is the obtaining of conditioned income distributions which are obtained through the estimation of quantile Mincer equations. This kind of equation models the relation between regressors type (z) – education level, experience, square experience, gender, ethnicity, regionalization – in the corresponding quantiles of the dependent variable ($w = \ln \text{earnings}$):

$$Q_q(w|z) = z' \beta_q \quad (1)$$

where $Q_q(w|z)$, for $q \in [0, 1]$, is the conditional q th quantile of the distribution of the dependent variable and β_q is the q th coefficient vector of the Mincerian equation.

As the sample average is the solution to minimizing a residual sum of squares and the median is the solution to a sum of absolute values of the residuals, the quantile regression can be defined as a solution to an asymmetric minimization of a weighted sum of the absolute values of the residuals. This way, β_q can be estimated by solving this following problem ([Koenker and Bassett, 1978](#)):

$$\min_{\beta \in R^k} \sum_{i:w_i \geq z'_i \beta} q |w_i - z'_i \beta_q| + \sum_{i:w_i < z'_i \beta} (1 - q) |w_i - z'_i \beta_q| \quad (2)$$

Despite not having an explicit form, this problem can be efficiently solved by the application of linear programming methods and the standard errors can be known by *bootstrapping* ([Martins and Pereira, 2004](#)). Besides, [Cameron and Trivedi \(2009\)](#) emphasize that quantile regressions become robust in the presence of *outliers* and the absence of normality of the dependent variable, and they do not need the existence of the conditional mean to be consistent. These characteristics favor and justify its application in numerous economic investigations just like this one.

3.2.2. Marginal density and decomposition of changes in earnings

After specifying the conditional quantile function, the estimated parameters are used in order to simulate the conditional distribution of w given z , by applying the Probability Integral Transform Theorem: If Y is a continuous random variable with *fc* $F_Y(y)$, then, the random variable $F_Y^{-1}(X)$, where X is $U(0, 1)$, has a distribution $F_Y(x)$, if U is a uniformly distributed variable in $[0,1]$, then $F^{-1}(U)$ has density function F . If q_1, q_1, \dots, q_M derived from a uniform distribution $(0,1)$, the corresponding M estimations of the quantiles of the w conditionals to z_i , $\hat{w}_i \equiv \{z_i' \hat{\beta}_q\}_{i=1}^M$, constitute a random sample of the (estimated) conditional distribution of the income given z_i .

This simulation procedure characterizes the conditional quantiles for all z , but it does not provide the w marginal density, since this marginal density depends not only on the conditional function $\hat{\beta}_q$, but also on the distribution of covariates $g(z)$. In order to generate a random sample of the marginal density w , one can simulate $g(z)$ lines and, for every para z_i line, starting from a uniform distribution $(0,1)$, simulate a random q_i and then form $\hat{w}_i \equiv z_i \hat{\beta}'_{q_i}$, which is a simulation of the w density implied by the model. By replicating this technique several times, it is possible to simulate a random sample of the desired distribution. This procedure is equivalent to numerically integrating the estimated quantile conditional function $\hat{Q}_q(w|z)$ on the distributions of z and q in order to form the function $f(\hat{w}) = \int \int \hat{Q}_q(w|z)g(z)dzdq$.

In order to infer from the gap between the incomes of two categories (e.g. between men and women) it is necessary to estimate two marginal densities of incomes: the function of density estimated according to the distribution of women's salaries, based on a generated sample, can be defined as the marginal distribution from the model; and the counterfactual density function, which in this case corresponds to the men's density function corresponding to the women's characteristic distribution (Machado and Mata, 2005).

The procedure to calculate the marginal distribution derived from the model is summarized in these steps:

1. A size- M random sample is generated corresponding to the M estimates of the conditional quantiles, from a uniform distribution $[0,1]$: q_1, q_1, \dots, q_M .
2. Using data about female individuals, it is then estimated that: $Q_{q_i}(w|z) = z' \beta_q$. As a result, M estimates for the quantile regression coefficients, $\hat{\beta}_{fq_i}$, where f stands for females.
3. A size- M random sample is generated, with replacement of the vector of covariates, z_f ;
4. Using this sample, the estimation of the density function is done through the multiplication of the estimated vector of coefficients times the generated size- M vector of covariates: $\{\hat{w}_i \equiv z_{if} \hat{\beta}'_{fq_i}\}_{i=1}^M$.

The counterfactual distribution is calculated according to the procedure above but on step 2, the male workers data is computed to estimate the vector of coefficients and then obtain: $\{\hat{w}_i \equiv z_{im} \hat{\beta}'_{mq_i}\}_{i=1}^M$, where m stands for males. The difference of these two density functions is described as gender discrimination.

In this work, the above-mentioned procedure is used in order to investigate the hiatus between income distributions caused by differences of gender, ethnicity and region. This way, for every occupational stratum the procedure was implemented three times, once for every duo of categories, namely: men and women, whites and non-whites, northeast-residing workers and workers residing in other regions.

3.3. Database

The database for the estimation of the proposed models is the National Household Survey (PNAD) of 2011, which is conducted yearly by The Brazilian Institute of Geography and Statistics (IBGE). A further comment on this database is worthy.

The PNAD is a complex sample, in which a sampling design is stratified with more than one stage of selection. Given these characteristics, point estimates may be affected by sampling plans, whereas the variance is influenced by both the sampling weights and the stratification and clustering. Aware of this prior to estimating the models, we try to proceed to the recommended procedure that basically consists in combining data from individual and household and use a command of STATA to state the sample as complex. Unfortunately an estimation error states that the command does not support the quantile regressions, which are essential to the contribution of the article, and disregarding them cause a severe damage to it.

An eventual recurring problem is that sample selection biases arise when the data for estimation are not random. We believe that researchers accept as true that the process of sample selection is done randomly on PNAD, and then the individuals treated in this study are those that are included in this sample and are in the labor market. The universe of the study is delimited by all individuals in that sample who fall within the selected occupational categories. The exclusion of individuals who do not participate in this universe cannot produce selection bias, as is the case of the military class, ruled out from this study or any other for not belonging to any working category and rural workers as well, since the study focuses only on workers in the urban sector. In this regard, Polachek (2008, p. 206) points out that if the sample of women in the labor force is random then the fact that a greater proportion of women are out of the labor force would not bias the estimate of the male-female wage differential.

3.3.1. Occupation definitions

Starting from the theoretical principle of the EGP/CASMIN scheme, a recoding of the PNAD data was done so as to aggregate more than 800 occupational categories into just seven which are described as follows:³

1. Professional general directors and managers (P.G.D.M.)

Legislators, general managers of government departments, operation and production managers, bioengineering professionals, biotechnology and genetic engineering professionals, mechatronics engineers, mathematicians, physicists, chemists, architects, doctors, surgeon dentists, veterinarians, pharmacists, teachers, lawyers, judges, managers and accountants.

2. Routine non-manual laborers (R.N.M.L.)

Registrars in general, agents, assistants and managerial assistants, secretaries, office machinery operators, office juniors, storekeepers and warehouse keepers, postmen, checkout operators and box office staff, receptionists, telephone operators, telemarketers, clerks, pollsters/interviewers.

3. Small business owners (S.B.O.)

Employers of more than five employees, providing they have declared themselves as employers or self-employed individuals: wholesale agents and sales technicians, dry-cleaners, laundry workers and similar kinds of professionals, replenishers and retail associates responsible for markups and markdowns, sales supervisors.

4. Skilled manual laborers such as technicians and supervisors (S.M.L.T.S.)

Technical designers and fashion designers, civil engineering/construction technicians, petrochemical technicians, topographic/land survey/hydrographic technicians, production technicians and food technologists, industrial lab staff, medical and dental instrument technicians.

5. Unskilled manual laborers (U.M.L.)

Workers in extraction of solid minerals, earth works and foundations, masonry structures, weaving preparation, textile craft, feeders in production lines.

6. Rural owners/employers (R.O.E.)

7. Rural-sector unskilled manual laborers (R.S.U.M.L.)

Due to the heterogeneity among rural and urban workers, as well as the different work environments in which these two groups operate, the categories (6) and (7) were excluded from sampling. These categories are the ones including rural workers. The purpose of such exclusion was to give the analysis the necessary robustness. Also, in order to homogenize the sample, the north-region workers were left aside, due to certain socio-geographic circumstances. Even though workers from the north region may be categorized as urban workers, they are very likely to have rural-worker characteristics. Needless to say that militaries are excluded from the sample since they are out of the labor force.

³ As already commented, Hout and Hauser (1992) launch critiques on this classification for not taking into account, among others, the spatial mobility of workers, which is not a concern of this work. They expand to twelve categories by just dividing each of five categories into two others. For instance, category (1) is broken up into lower grade and upper grade professionals. Besides the difficulties to handle their classification with PNAD data, sample in each category would be reduced. So, we prefer to keep the classification as is.

Table 1
Descriptive statistics by occupational categories.

Variables	1. <i>P.G.D.M.</i>		2. <i>R.N.M.L.</i>		3. <i>S.B.O.</i>		4. <i>S.M.L.T.S.</i>		5. <i>U.M.L.</i>	
	Average	Standard error	Average	Standard error	Average	Standard error	Average	Standard error	Average	Standard error
<i>w</i>	3421.27	4263.71	1177.91	1442.35	5917.30	7409.11	1094.28	1077.51	844.72	816.86
<i>ed</i>	14.14	2.82	11.78	3.22	12.71	3.33	9.37	3.75	8.28	3.84
<i>xp</i>	16.93	10.02	15.24	11.12	22.72	9.61	19.85	11.83	21.83	11.89
<i>GEN</i>	0.47	0.50	0.59	0.49	0.31	0.46	0.33	0.47	0.43	0.49
<i>ETH</i>	0.69	0.46	0.54	0.50	0.78	0.42	0.47	0.50	0.41	0.49
<i>MW</i>	0.16	0.36	0.15	0.35	0.15	0.36	0.13	0.34	0.14	0.34
<i>NE</i>	0.23	0.42	0.29	0.45	0.20	0.40	0.27	0.45	0.33	0.47
<i>SE</i>	0.39	0.49	0.36	0.48	0.35	0.48	0.38	0.48	0.34	0.47
<i>S</i>	0.23	0.42	0.21	0.41	0.30	0.46	0.22	0.41	0.19	0.39
Number of observations	11,211		36,150		743		27,904		28,050	

Source: Own calculations using PNAD/2011 data.

3.3.2. Variable definition

For each occupational group mentioned in the previous subsection, quantile Mincer equations were estimated, and then from these equations, a counterfactual decomposition of wage distribution was conceived. The equations were estimated according to the following specification:

$$w_i = \alpha + \beta_1 ed_i + \beta_2 xp_i + \beta_2 (xp_i)^2 + \beta_3 ETH_i + \beta_4 GEN_i + \beta_5 MW_i + \beta_6 NE_i + \beta_7 SE_i + \beta_8 S_i + u_i \quad (3)$$

where

- *w* is the logarithm of worker's main earning.
- *ed* are the education years – corresponds to a grade or a level. For the purpose of equivalence, each accomplished passed grade is considered as one education year.
- *xp* is the calculated experience, it is conceived as age minus six, minus the education years.⁴ When adopting this procedure, it was assumed that all the analyzed individuals started out their school years at the age of six and they do not spend any time outside school or outside the job market. Berndt (1991 in Kassouf, 1998), and also it is assumed that school-attending individuals do not work. Besides, only individuals in the age range 15–55 were analyzed in order to homogenize the sample and avoid any eventual selection bias.
- *ETH* is for ethnicity – 1 for whites and 0 for non-whites;
- *GEN* is for gender – 1 for female and 0 for male;
- *MW* represents Midwest – 1 for the midwest region residents and 0 for non-midwest residents.
- *NE* represents Northeast – 1 for the northeast region residents and 0 for non-northeast residents;
- *SE* represents Southeast – 1 for the southeast region residents and 0 for non-southeast residents.
- *S* represents South – 1 for the south region residents and 0 for non-south residents.

The descriptive statistics of the variables applied in the model of each occupational category can be seen in Table 1. There is heterogeneity between the occupational categories; this fact can be corroborated by the marked variation amplitude of the continuous variables. This justifies the suitability for this type of disaggregated analysis. It is worth noticing that the low relative representativeness of the small business owners category, may lead to results that diverge from the other categories' results.

⁴ This procedure was adopted because the PNAD did not have information about individuals' experience.

4. Results

In this section, the results of both proposed empirical applications are presented. In the first subsection shows the quantile Mincer equations, followed by the income decomposition.

4.1. Quantile Mincerian equations

The estimates of quantile Mincerian equations by occupational stratum as well as by LS are shown in [Table 2](#). It reveals that all the coefficients are significant up to a maximum of 5%. Besides differing from the LS estimates, the effects of the explanatory variables are distinct in most cases along the distribution of the dependent variable. Such characteristic is statistically proved by the Wald Test, whose results are shown in [Table A.1](#) of [Appendix A](#). This effect also varies between the occupational strata and can be easily seen in [Figs. B1–B5](#) of [Appendix B](#). These figures, besides showing the results by LS estimation, also evidence the influence of each explanatory variable on the dependent quantiles by occupational category.

The coefficients of the education (*ed*) and experience (*xp*) variables are positive in all cases, with stronger effect on the general directors' and managers' category (*P.G.D.M.*) implying that, according to literature, human capital significantly contributes to the increase of workers' earnings. It is worth emphasizing, that except for the unskilled manual laborers' category (*U.M.L.*) the effect of these variables is greater at the top of the earnings distribution. For the total number of workers, a similar effect was noticed in the USA ([Chamberlain, 1994](#); [Buchinsky, 1994](#)), Germany ([Fitzenberger and Kurz, 1997](#)), Uruguay ([González and Miles, 2001](#)), Zambia ([Nielsen and Rosholm, 2001](#)), Portugal ([Machado and Mata, 2005](#)) and Brazil ([Coelho et al., 2010](#)) for a sample of women in the range of 20–60 years of age. The xp^2 variable coefficients measure to what extent additional experience years translate into salary increases. This, in most cases, shows the expected negative sign, evidencing that at some point the effect of experience on salaries goes null or even negative. In certain points of the distribution of some occupational strata, these coefficients differ from what was expected, although there is absence of statistical significance.

Gender and race present negative- and positive sign coefficients, respectively, in all quantiles and occupations, evidencing that if a worker is a female or a non-white person, she or he will get smaller earnings. Regarding gender, this result is similar to the result found by [Machado and Mata \(2005\)](#) in their study for Portuguese workers. The effects of these variables between the quantiles are distinct in the magnitudes of the coefficients amongst occupations, since the effect is stronger on some of higher quantiles while in some others the effect is reverted.

The dummy variables inserted to detect variation in differences in earnings between residents of more-developed and less-developed regions proved to be relevant in most of the studied cases. As expected, the negative effects are predominant on the earnings of individuals who reside in the least-developed region (the northeast) in almost all quantiles and occupations. Except for small business owners, on which the null effect prevails in all regions, but principally in the southeast, and also income strata whose results might have been derived from the homogeneity degree and low sampling representativeness relative to other occupations.

4.2. Decomposition of earning distribution

In order to decompose the workers' income distribution into two components, one component due to individual distinct characteristics, and another component that comes from the coefficients (that is, remuneration to these characteristics), the procedure described in [Section 3](#) was conducted, with a replication number (*M*) equal to 1000. The characteristics included were: years of education, experience, ethnicity (when conducting decomposition by gender) and gender (when decomposition by ethnicity and region are carried out).

The results of gap decomposition in-between men's and women's income, as well as results of the whites and non-whites and workers residing in the northeast region residents in comparison to other regions, by occupational stratum, are displayed in [Table 3](#).

Regarding the gap between the men's and women's income, it is clear that in all occupational strata, the gap is negative while it is significantly wider in the unskilled manual laborer category, where in smaller quantiles, it surpasses 100% mark, and that the proportion originated by the characteristics is positive. Therefore, the wage differential has to do with the coefficients, that is, it has to do with the remuneration given to the individual due to his or her characteristics, and this means the presence of discrimination by gender in the Brazilian job market. [Santos and Ribeiro \(2006\)](#) found

Table 2
Quantile Mincer equations by occupational categories.

Variable	1. P.G.D.M.					2. R.N.M.L.					3. S.B.O.				
	5%	25%	50%	75%	LS	5%	25%	50%	75%	LS	5%	25%	50%	75%	LS
<i>ed</i>	0.095 (0.005)	0.121 (0.003)	0.136 (0.004)	0.145 (0.004)	0.128 (0.003)	0.088 (0.003)	0.067 (0.002)	0.084 (0.001)	0.104 (0.002)	0.097 (0.001)	0.102 (0.016)	0.093 (0.017)	0.108 (0.016)	0.109 (0.014)	0.099 (0.010)
<i>xp</i>	0.054 (0.004)	0.054 (0.003)	0.059 (0.003)	0.058 (0.004)	0.056 (0.003)	0.050 (0.002)	0.031 (0.001)	0.033 (0.001)	0.043 (0.001)	0.040 (0.001)	0.045 (0.020)	0.056 (0.017)	0.013* (0.020)	0.026* (0.020)	0.032 (0.012)
<i>xp</i> ²	-0.001 (0.000)	-0.001 (0.000)	-0.001 (0.000)	-0.001 (0.000)	-0.001 (0.000)	-0.001 (0.000)	-0.001 (0.000)	0.000 (0.000)	0.000 (0.000)	-0.001 (0.000)	0.000* (0.000)	-0.001* (0.000)	0.000* (0.000)	0.000* (0.000)	0.000* (0.000)
<i>GEN</i>	-0.323 (0.025)	-0.367 (0.016)	-0.388 (0.016)	-0.388 (0.020)	-0.370 (0.014)	-0.262 (0.013)	-0.216 (0.006)	-0.250 (0.006)	-0.316 (0.009)	-0.308 (0.006)	-0.510 (0.120)	-0.450 (0.080)	-0.465 (0.081)	-0.410 (0.084)	-0.490 (0.063)
<i>ETH</i>	0.088 (0.027)	0.129 (0.018)	0.172 (0.021)	0.211 (0.022)	0.170 (0.016)	0.076 (0.014)	0.064 (0.006)	0.084 (0.007)	0.128 (0.009)	0.107 (0.006)	0.297 (0.150)	0.168* (0.089)	0.163* (0.123)	0.082* (0.083)	0.157 (0.072)
<i>NE</i>	-0.313 (0.036)	-0.221 (0.025)	-0.148 (0.026)	-0.07* (0.029)	-0.140 (0.022)	-0.501 (0.024)	-0.327 (0.010)	-0.314 (0.009)	-0.280 (0.013)	-0.320 (0.009)	-0.15* (0.175)	0.002* (0.125)	0.267 (0.133)	0.465 (0.117)	0.222 (0.086)
<i>MW</i>	0.006* (0.035)	0.093 (0.031)	0.182 (0.028)	0.310 (0.035)	0.190 (0.023)	0.016* (0.020)	-0.044 (0.009)	-0.022 (0.011)	0.055 (0.017)	0.038 (0.010)	-0.10* (0.213)	-0.04* (0.108)	0.263 (0.107)	0.391 (0.107)	0.211 (0.093)
<i>SE</i>	-0.067 (0.028)	-0.03* (0.020)	0.001* (0.020)	0.060 (0.025)	0.012* (0.018)	0.001* (0.016)	-0.046 (0.008)	-0.043 (0.008)	-0.02* (0.011)	-0.027 (0.008)	-0.14* (0.144)	-0.07* (0.078)	0.016* (0.090)	0.109* (0.088)	0.032* (0.072)
<i>c</i>	4.952 (0.073)	5.085 (0.052)	5.197 (0.052)	5.444 (0.064)	5.356 (0.045)	4.845 (0.051)	5.617 (0.019)	5.640 (0.016)	5.595 (0.021)	5.463 (0.016)	5.045 (0.333)	5.774 (0.294)	6.302 (0.300)	6.627 (0.272)	6.304 (0.193)
Variable	[0,2-6]4. S.M.L.T.S.					[0,7-11]5. U.M.L.									
	5%	25%	50%	75%	LS	5%	25%	50%	75%	LS					
<i>ed</i>	0.041 (0.004)	0.039 (0.001)	0.051 (0.001)	0.073 (0.001)	0.065 (0.001)	0.061 (0.004)	0.055 (0.001)	0.050 (0.001)	0.061 (0.001)	0.063 (0.001)					
<i>xp</i>	0.034 (0.004)	0.024 (0.001)	0.030 (0.001)	0.037 (0.001)	0.034 (0.001)	0.042 (0.003)	0.037 (0.001)	0.032 (0.001)	0.034 (0.001)	0.039 (0.001)					
<i>xp</i> ²	-0.001 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	-0.001 (0.000)	-0.001 (0.000)	0.000 (0.000)	0.000 (0.000)	-0.001 (0.000)					
<i>GEN</i>	-0.621 (0.030)	-0.339 (0.009)	-0.390 (0.007)	-0.458 (0.009)	-0.452 (0.007)	-1.182 (0.025)	-0.746 (0.013)	-0.549 (0.009)	-0.558 (0.008)	-0.712 (0.007)					
<i>ETH</i>	0.067 (0.016)	0.047 (0.007)	0.073 (0.008)	0.106 (0.009)	0.097 (0.007)	0.036* (0.021)	0.058 (0.010)	0.063 (0.008)	0.091 (0.008)	0.079 (0.008)					
<i>NE</i>	-0.673 (0.033)	-0.370 (0.010)	-0.353 (0.010)	-0.324 (0.012)	-0.373 (0.011)	-0.651 (0.034)	-0.471 (0.014)	-0.395 (0.011)	-0.343 (0.011)	-0.429 (0.011)					
<i>MW</i>	-0.070 (0.021)	-0.108 (0.011)	-0.077 (0.013)	-0.003* (0.017)	-0.02* (0.012)	-0.02* (0.033)	-0.038 (0.015)	-0.032 (0.012)	-0.006* (0.012)	0.002* (0.013)					
<i>SE</i>	-0.090 (0.019)	-0.080 (0.009)	-0.050 (0.008)	-0.024 (0.012)	-0.044 (0.009)	-0.03* (0.026)	-0.047 (0.013)	-0.039 (0.010)	-0.029 (0.011)	-0.027 (0.011)					
<i>c</i>	5.583 (0.056)	6.052 (0.020)	6.111 (0.018)	6.104 (0.021)	5.950 (0.017)	5.125 (0.052)	5.698 (0.023)	6.022 (0.022)	6.180 (0.021)	5.854 (0.019)					

Source: Own estimations using PNAD/2011 data.
Non-significant at 5%.

Notes: Standard errors – in parentheses – calculated by Bootstrapping with 1000 repetitions. The South variable was excluded in order to prevent from perfect collinearity.

Table 3
Decomposition of earning differentials by gender, ethnicity and regions by selected quantiles.

	Gender				Ethnicity				Region			
	5%	25%	50%	75%	5%	25%	50%	75%	5%	25%	50%	75%
<i>P.G.D.M.</i>												
Total	−0.1913 (0.0171)	−0.2715 (0.0126)	−0.3228 (0.0129)	−0.3499 (0.0149)	−0.1997 (0.0185)	−0.2777 (0.0149)	−0.3017 (0.0158)	−0.3067 (0.0180)	−0.3220 (0.0287)	−0.2893 (0.0192)	−0.2694 (0.0200)	−0.2426 (0.0232)
Coefficients	−0.2909 (0.0160)	−0.3582 (0.0124)	−0.3949 (0.0127)	−0.4051 (0.0140)	−0.0886 (0.0144)	−0.1712 (0.0108)	−0.2076 (0.0113)	−0.2218 (0.0131)	−0.2640 (0.0129)	−0.2482 (0.0100)	−0.2407 (0.0105)	−0.2233 (0.0121)
Characteristics	0.0996 (0.0172)	0.0867 (0.0132)	0.0721 (0.0133)	0.0552 (0.0142)	−0.1111 (0.0195)	−0.1065 (0.0134)	−0.0941 (0.0128)	−0.0850 (0.0140)	−0.0579 (0.0185)	−0.0411 (0.0132)	−0.0288 (0.0129)	−0.0193 (0.0144)
<i>R.N.M.L.</i>												
Total	−0.2143 (0.0131)	−0.1383 (0.0035)	−0.2037 (0.0043)	−0.2754 (0.0060)	−0.2595 (0.0151)	−0.1588 (0.0039)	−0.2169 (0.0048)	−0.2827 (0.0067)	−0.5945 (0.0206)	−0.2970 (0.0066)	−0.3160 (0.0059)	−0.3599 (0.0094)
Coefficients	−0.3016 (0.0127)	−0.2088 (0.0049)	−0.2731 (0.0060)	−0.3535 (0.0081)	−0.1904 (0.0118)	−0.1049 (0.0040)	−0.1551 (0.0051)	−0.2073 (0.0070)	−0.5691 (0.0093)	−0.2820 (0.0032)	−0.3008 (0.0042)	−0.3400 (0.0059)
Characteristics	0.0873 (0.0113)	0.0705 (0.0051)	0.0694 (0.0064)	0.0781 (0.0084)	−0.0690 (0.0141)	−0.0539 (0.0047)	−0.0618 (0.0049)	−0.0754 (0.0065)	−0.0254 (0.0103)	−0.0149 (0.0040)	−0.0153 (0.0046)	−0.0199 (0.0064)
<i>S.B.O.</i>												
Total	−0.3655 (0.0942)	−0.3668 (0.0657)	−0.3795 (0.0602)	−0.3656 (0.0720)	−0.1980 (0.1274)	−0.1887 (0.0795)	−0.0726 (0.0924)	0.0365 (0.0871)	−0.1750 (0.1693)	−0.0462 (0.0975)	0.1048 (0.0984)	0.2817 (0.1061)
Coefficients	−0.4684 (0.0676)	−0.4556 (0.0442)	−0.4540 (0.0425)	−0.4277 (0.0495)	−0.2316 (0.0596)	−0.2017 (0.0403)	−0.0786 (0.0378)	0.0383 (0.0442)	−0.1446 (0.0543)	−0.0220 (0.0411)	0.1243 (0.0366)	0.3069 (0.0426)
Characteristics	0.1029 (0.0638)	0.0888 (0.0460)	0.0745 (0.0476)	0.0621 (0.0506)	0.0336 (0.0666)	0.0130 (0.0468)	0.0060 (0.0429)	−0.0018 (0.0480)	−0.0304 (0.0651)	−0.0242 (0.0510)	−0.0194 (0.0445)	−0.0252 (0.0474)
<i>S.M.L.T.S.</i>												
Total	−0.6427 (0.0283)	−0.2201 (0.0044)	−0.3944 (0.0058)	−0.4620 (0.0089)	−0.2735 (0.0200)	−0.1362 (0.0034)	−0.1953 (0.0052)	−0.2395 (0.0064)	−0.7232 (0.0271)	−0.2796 (0.0085)	−0.3770 (0.0066)	−0.4280 (0.0094)
Coefficients	−0.6944 (0.0121)	−0.2679 (0.0044)	−0.4479 (0.0047)	0.0557 (0.0059)	−0.2527 (0.0160)	−0.1111 (0.0052)	−0.1595 (0.0058)	−0.1834 (0.0077)	−0.7048 (0.0118)	−0.2575 (0.0040)	−0.3418 (0.0042)	−0.3782 (0.0057)
Characteristics	0.0517 (0.0109)	0.0479 (0.0051)	0.0536 (0.0053)	−0.5177 (0.0064)	−0.0207 (0.0171)	−0.0251 (0.0053)	−0.0358 (0.0057)	−0.0561 (0.0070)	−0.0184 (0.0129)	−0.0222 (0.0044)	−0.0352 (0.0047)	−0.0498 (0.0062)
<i>U.M.L.</i>												
Total	−1.1780 (0.0203)	−0.7111 (0.0124)	−0.4656 (0.0059)	−0.5994 (0.0061)	−0.3330 (0.0175)	−0.2371 (0.0089)	−0.2089 (0.0047)	−0.2255 (0.0062)	−0.7512 (0.0268)	−0.5371 (0.0144)	−0.3873 (0.0065)	−0.3914 (0.0084)
Coefficients	−1.1978 (0.0142)	−0.7279 (0.0043)	−0.4804 (0.0049)	−0.6134 (0.0065)	−0.2666 (0.0218)	−0.1793 (0.0075)	−0.1490 (0.0060)	−0.1622 (0.0079)	−0.7257 (0.0168)	−0.4985 (0.0052)	−0.3521 (0.0044)	−0.3509 (0.0059)
Characteristics	0.0197 (0.0137)	0.0168 (0.0046)	0.0148 (0.0052)	0.0139 (0.0067)	−0.0663 (0.0231)	−0.0578 (0.0086)	−0.0600 (0.0058)	−0.0632 (0.0070)	−0.0255 (0.0173)	−0.0386 (0.0065)	−0.0352 (0.0053)	−0.0405 (0.0061)

Source: Own estimations using PNAD/2011 data.

Note: Standard errors in parentheses – sampling by bootstrapping with 1000 repetitions.

that the critical discrimination occurs at the top of the earning distribution. A similar result was verified by [Albrecht et al. \(2009\)](#) for the Netherlands, [Arulampalam et al. \(2005\)](#) for several European countries and [Rica et al. \(2005\)](#) for Spain.

The gaps between whites' and non-whites' incomes who reside in the northeast or in another region are also negative gaps. Except for the small business owners category, there are positive gaps in some points of the distribution. Just like in the decomposition by gender, the largest parts of these lags come from the coefficients themselves, which evidences the existence of discrimination. However, the proportion that comes from the workers' characteristics is also negative. Such fact means that non-white individuals who reside in the northeast region have characteristics (years of education and experience) that make them get smaller incomes than those of white individuals or individuals who reside in other regions. The implementation of policies that promote social integration of these categories becomes highly advisable in the light of the evidences shown herein.

[Figs. C1–C3](#) in [Appendix C](#) generated by the Stata software after the decomposition procedure, illustrate, for every occupational stratum, the gaps between the men's and women's income distributions, as well as between whites and non-whites and northeast residents and other regions, coming from the coefficients (discrimination) with a 95% confidence interval. Some occupations show a glass ceiling effect, that is, the lag is higher at the top of the distribution. That can be observed, for instance, in the professional general directors and managers category (*P.G.D.M.*), likewise in decompositions by gender and by race, where the income hiatus grows from 19% at the bottom of the distribution to over 30% at the top. However, the incidence of a sticky floor effect (situations when the hiatus is higher at the bottom of the distribution) is more likely to occur. This effect was observed by [Arulampalam et al. \(2005\)](#). In the decomposition by regions, such effect is observed in all occupational strata, indicating that discrimination decreases as distribution evolves.

The results of the decomposition over all of the distribution can be better seen in [Figs. D1–D3](#) depicted in [Appendix D](#). It is clear the hiatus varies between the incomes along the quantiles and between occupations. In the decomposition by ethnicity, for example, in the category of professional general directors and managers, the income differentials vary significantly along the quantiles and do a strikingly distinct trajectory, different from the path seen in the small business owners (*S.B.O.*) category. Facts like these ratify the importance of the overall results here found by disaggregating the analysis into distinct strata and by quantiles, without which a solid conclusion could be withdrawn as pointed out earlier from the literature surveying.

5. Concluding remarks

By grouping over eight hundred occupational categories (as classified by the PNAD survey in 2011) into just five based upon the schematic EGP theoretical principle, an analysis of the Brazilian urban job market by occupational and individual earnings strata has been conducted herein.

Quantile Mincerian equations were estimated in order to verify the effects of inherent and acquired attributes on various quantiles along workers' income distribution. Abiding by the [Machado and Mata \(2005\)](#) methodology, a decomposition of such distribution allowed measuring the individuals' income differentials according to gender, race and region of residence. This procedure made it possible to evidence how much of the difference between these categories' incomes is caused by distinctions in characteristics (experience and education) and how much of the difference is caused by dissimilarities in remuneration to such characteristics, which conclusively means discrimination.

The results from the Mincerian equations ratify the importance of the disaggregated analysis of occupational strata by quantiles and shed merit to the findings, once the effects of the explanatory variables are distinct along income distributions and do vary amongst occupations. This suggests an alternative and realistic view of such equations with variable coefficients, especially regarding human capital return. Variables such as education and experience bring on a growth effect on workers' earnings, mainly in the professional general directors and managers category. The regional effects show that the earnings from those who reside in the northeast tend to be smaller than the others in all categories, except in the small business owners category.

Regarding the income gaps in gender, race and region, some occupations display clearly the glass ceiling effect, that is, the lag is larger at the top of the distribution. However, contrary to other studies that did not disaggregate data by occupations, it is found that there is sharp evidence toward a sticky floor effect, meaning that the hiatus is larger at the bottom of the distribution. In the decomposition by region, such effect is observable in all occupational strata, indicating that discrimination decreases as distribution progresses.

The income differentials between men and women are negative in all occupational strata, expressively higher in the Unskilled Manual Laborers category and the proportion derived from the characteristics is positive. This means that such differential is originated mainly from the earning equation coefficients, which implies that the remuneration is due to the individual characteristics; therefore, discrimination by gender in the Brazilian job market is evident in most types of jobs.

Similarly, the individuals' income gaps between whites and non-whites as well as northeast residents and other-region residents are negative, except on quite a few points of the distribution in the small business owners category. These lags are mostly consequential to the equation coefficients, which in turn dictate that there exists discrimination. However, unlike the gender results, these differentials are also due to workers' distinct acquired characteristics. This indicates that non-white individuals and northeast residents have characteristic endowments, such as education and experience, which do not allow them to earn more than the other individuals. These solid inferences confirm the imperativeness of policies that foster equity by competence, promote social integration and ban discrimination in the labor market.

Appendix A. Coefficient equality between quantiles

Table A.1.

Table A.1

Wald test.

Occupations	Coefficients	<i>p</i> value						
		q05 = q25	q05 = q50	q05 = q75	q25 = q50	q25 = q75	q50 = q75	q05 = q25 = q50 = q75
1. <i>P.G.D.M.</i>	<i>ed</i>	0.0000	0.0000	0.0000	0.0000	0.0000	0.0101	0.0000
	<i>xp</i>	0.9742	0.2505	0.4888	0.0845	0.3761	0.6821	0.3861
	<i>xp</i> ²	0.1437	0.2976	0.0590	0.7151	0.3357	0.1394	0.2076
	<i>GEN</i>	0.0679	0.0163	0.0293	0.1685	0.3160	0.9852	0.1090
	<i>ETH</i>	0.1235	0.0079	0.0003	0.0245	0.0008	0.0488	0.0017
	<i>NE</i>	0.0104	0.0000	0.0000	0.0022	0.0000	0.0004	0.0000
	<i>SE</i>	0.2242	0.0302	0.0002	0.0712	0.0003	0.0044	0.0007
	<i>co</i>	0.0210	0.0000	0.0000	0.0010	0.0000	0.0000	0.0000
2. <i>R.N.M.L.</i>	<i>ed</i>	0.0000	0.2286	0.0000	0.0000	0.0000	0.0000	0.0000
	<i>xp</i>	0.0000	0.0000	0.0066	0.0060	0.0000	0.0000	0.0000
	<i>xp</i> ²	0.0000	0.0000	0.0000	0.0003	0.3393	0.0378	0.0000
	<i>GEN</i>	0.0001	0.3710	0.0004	0.0000	0.0000	0.0000	0.0000
	<i>ETH</i>	0.3658	0.5852	0.0007	0.0005	0.0000	0.0000	0.0000
	<i>NE</i>	0.0000	0.0000	0.0000	0.1161	0.0002	0.0010	0.0000
	<i>SE</i>	0.0018	0.0089	0.2953	0.6428	0.0147	0.0089	0.0007
	<i>co</i>	0.0011	0.0665	0.1189	0.0238	0.0000	0.0000	0.0000
3. <i>S.B.O.</i>	<i>ed</i>	0.6458	0.7464	0.7421	0.2900	0.3946	0.9879	0.7533
	<i>xp</i>	0.6296	0.2076	0.4830	0.0119	0.1677	0.4727	0.0961
	<i>xp</i> ²	0.6319	0.1934	0.4512	0.0099	0.1462	0.4968	0.0826
	<i>GEN</i>	0.6030	0.7257	0.4597	0.8492	0.6721	0.4916	0.8484
	<i>ETH</i>	0.3753	0.4307	0.1890	0.9584	0.4035	0.4475	0.5874
	<i>NE</i>	0.3755	0.0302	0.0021	0.0260	0.0013	0.1062	0.0037
	<i>SE</i>	0.6385	0.3132	0.1189	0.2694	0.0623	0.2722	0.2609
	<i>co</i>	0.7580	0.0970	0.0301	0.0038	0.0007	0.2145	0.0048
4. <i>S.M.L.T.S.</i>	<i>ed</i>	0.5397	0.0034	0.0000	0.0000	0.0000	0.0000	0.0000
	<i>xp</i>	0.0045	0.2095	0.3819	0.0000	0.0000	0.0000	0.0000
	<i>xp</i> ²	0.0025	0.0168	0.0690	0.0913	0.0099	0.0740	0.0014
	<i>GEN</i>	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	<i>ETH</i>	0.1760	0.7155	0.0306	0.0001	0.0000	0.0001	0.0000
	<i>NE</i>	0.0000	0.0000	0.0000	0.0827	0.0005	0.0041	0.0000
	<i>SE</i>	0.5876	0.0319	0.0016	0.0001	0.0000	0.0067	0.0000
	<i>co</i>	0.0560	0.7392	0.0069	0.0047	0.0000	0.0000	0.0000

Table A.1 (Continued)

Occupations	Coefficients	p value						
		q05 = q25	q05 = q50	q05 = q75	q25 = q50	q25 = q75	q50 = q75	q05 = q25 = q50 = q75
5. U.M.L.	<i>ed</i>	0.0722	0.0014	0.9870	0.0000	0.0002	0.0000	0.0000
	<i>xp</i>	0.0756	0.0019	0.0141	0.0005	0.0690	0.1030	0.0008
	<i>xp²</i>	0.1582	0.0062	0.0081	0.0004	0.0030	0.9967	0.0016
	<i>GEN</i>	0.0000	0.0000	0.0000	0.0000	0.0000	0.2687	0.0000
	<i>ETH</i>	0.2263	0.1737	0.0095	0.5710	0.0014	0.0002	0.0009
	<i>NE</i>	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	<i>SE</i>	0.4525	0.6682	0.9723	0.5113	0.1909	0.2977	0.5098
	<i>co</i>	0.4375	0.5996	0.7918	0.6124	0.0416	0.0301	0.1061

Source: Own estimations using PNAD/2011 data.

Appendix B. Quantile Mincer equations – the effects of explanatory variables

Figs. B1–B5.

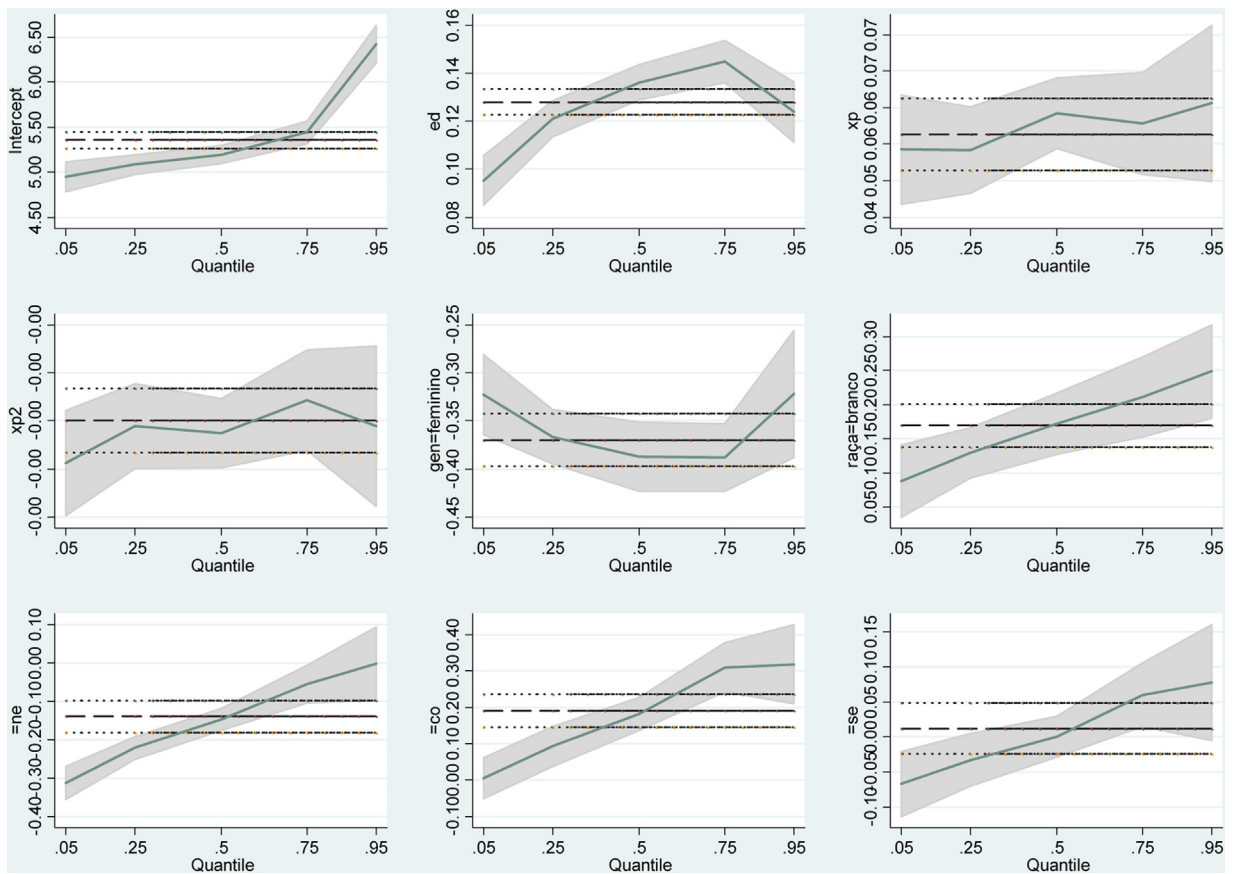


Fig. B1. Occupation 1.

Source: Own elaboration using PNAD/2011 data.

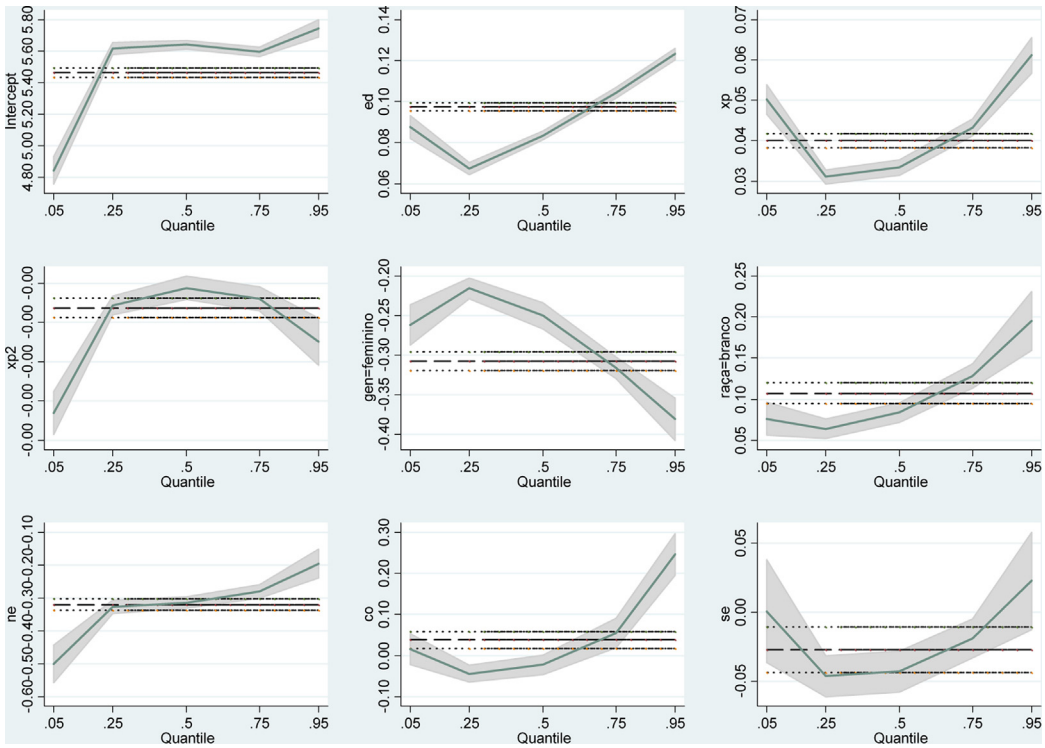


Fig. B2. Occupation 2.

Source: Own elaboration using PNAD/2011 data.

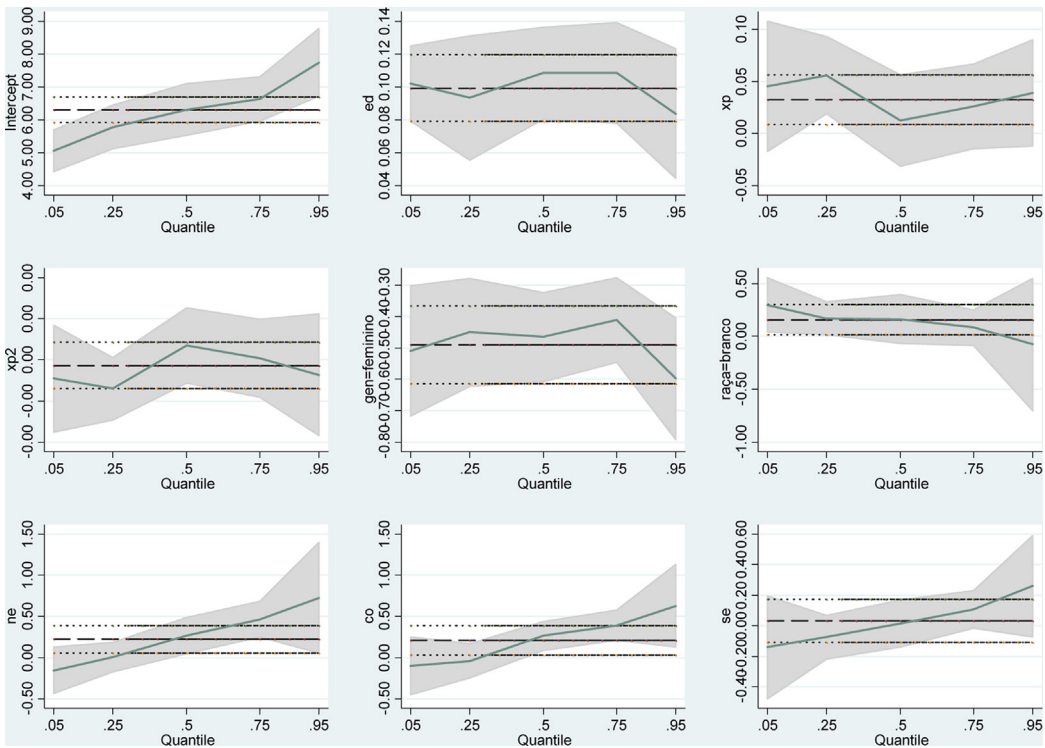


Fig. B3. Occupation 3.

Source: Own elaboration using PNAD/2011 data.

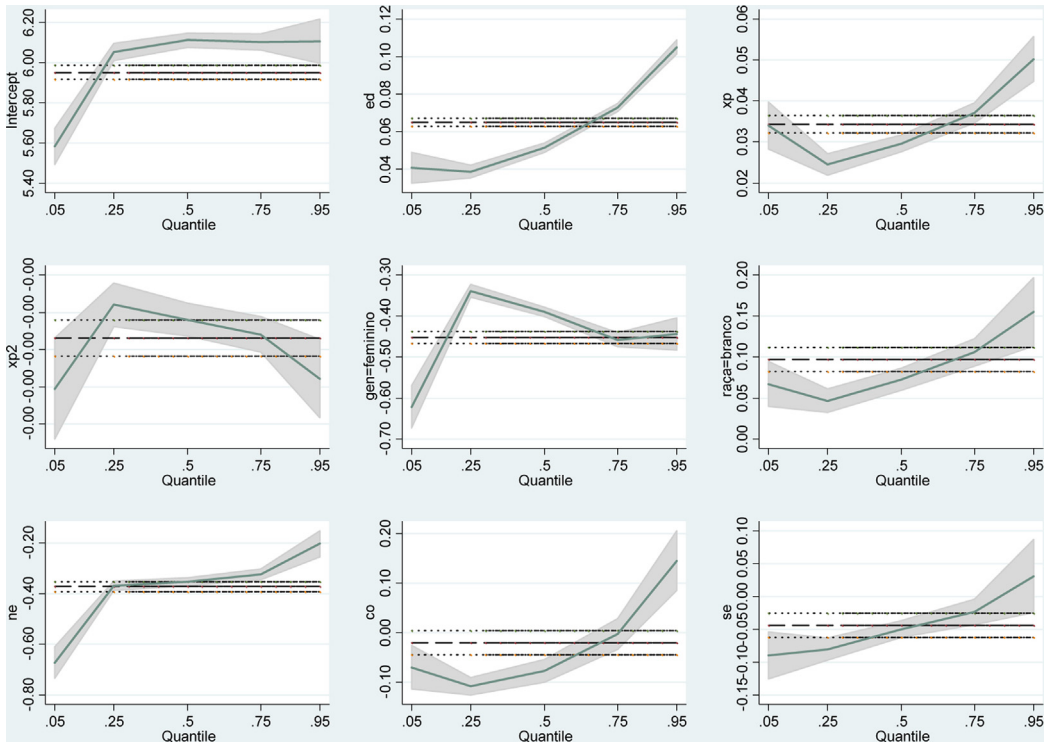


Fig. B4. Occupation 4.

Source: Own elaboration using PNAD/2011 data.

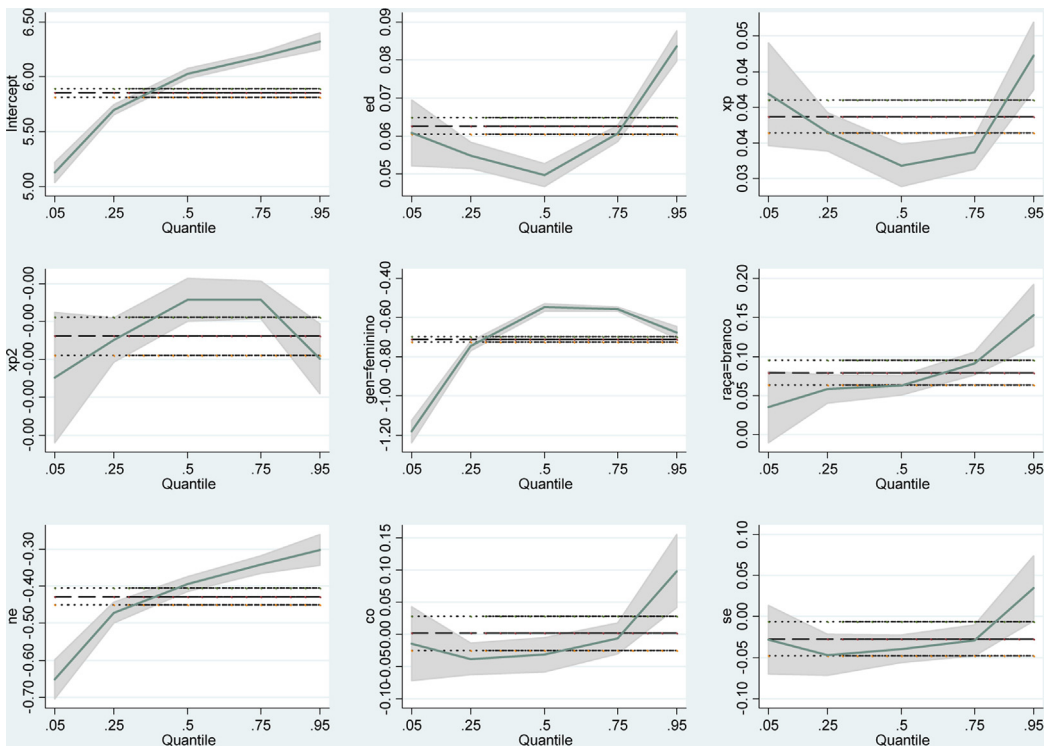


Fig. B5. Occupation 5.

Source: Own elaboration using PNAD/2011 data.

Appendix C. Coefficient effects (discrimination)

Figs. C1–C3.

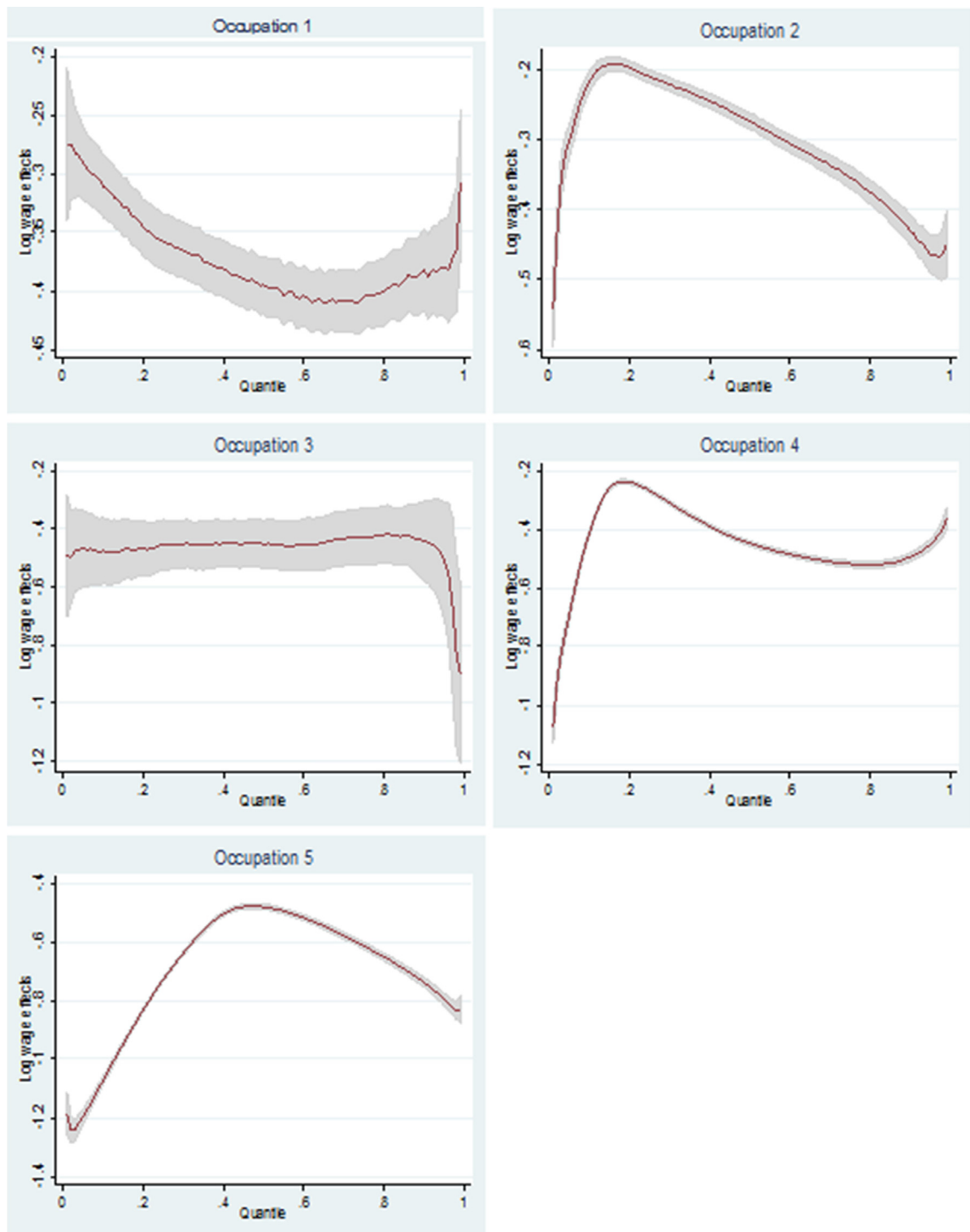


Fig. C1. Gender.

Source: Own elaboration using PNAD/2011 data.

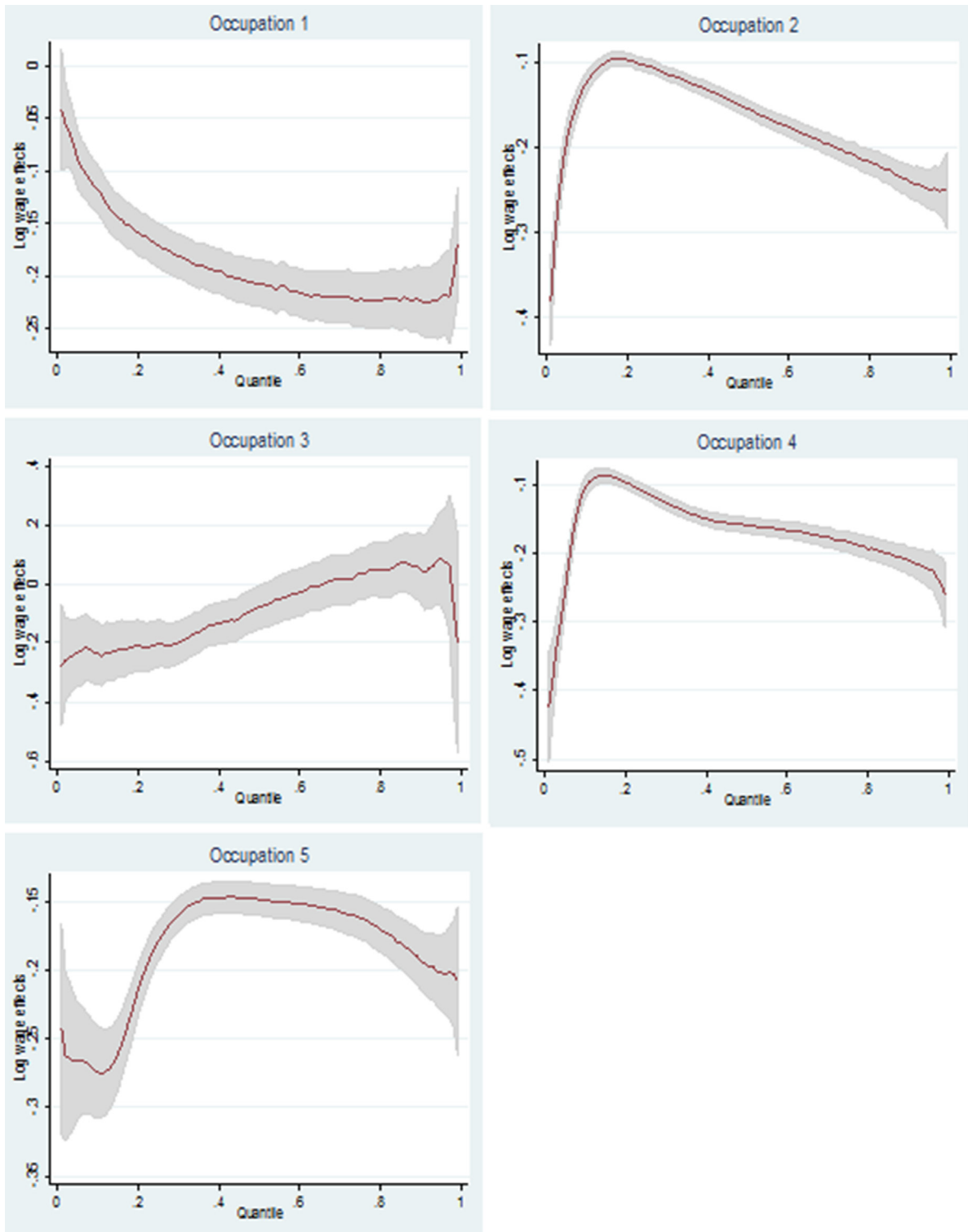


Fig. C2. Ethnicity.

Source: Own elaboration using PNAD/2011 data.

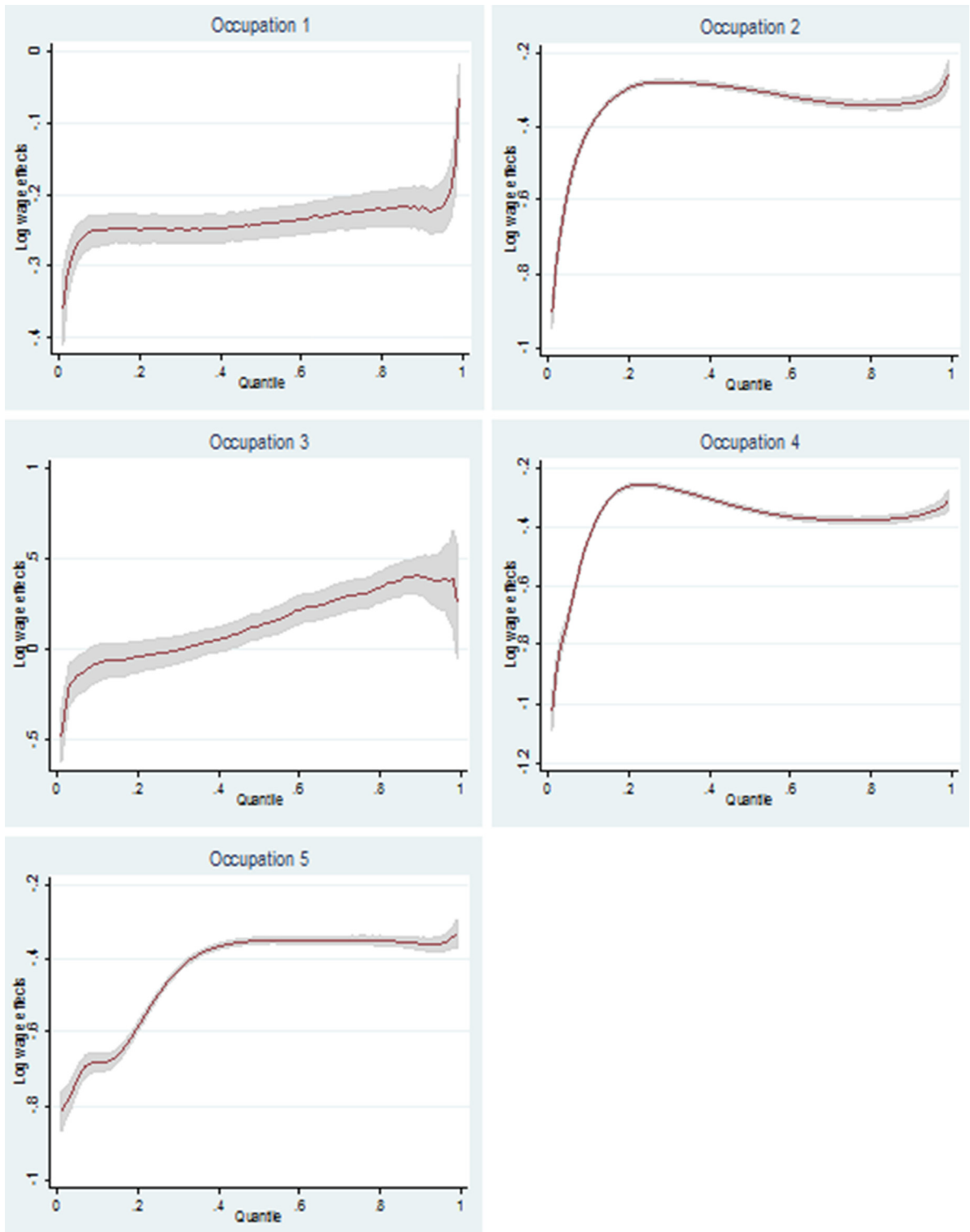


Fig. C3. Northeast region.

Source: Own elaboration using PNAD/2011 data.

Appendix D. Decomposition of differences in distribution

Figs. D1–D3.

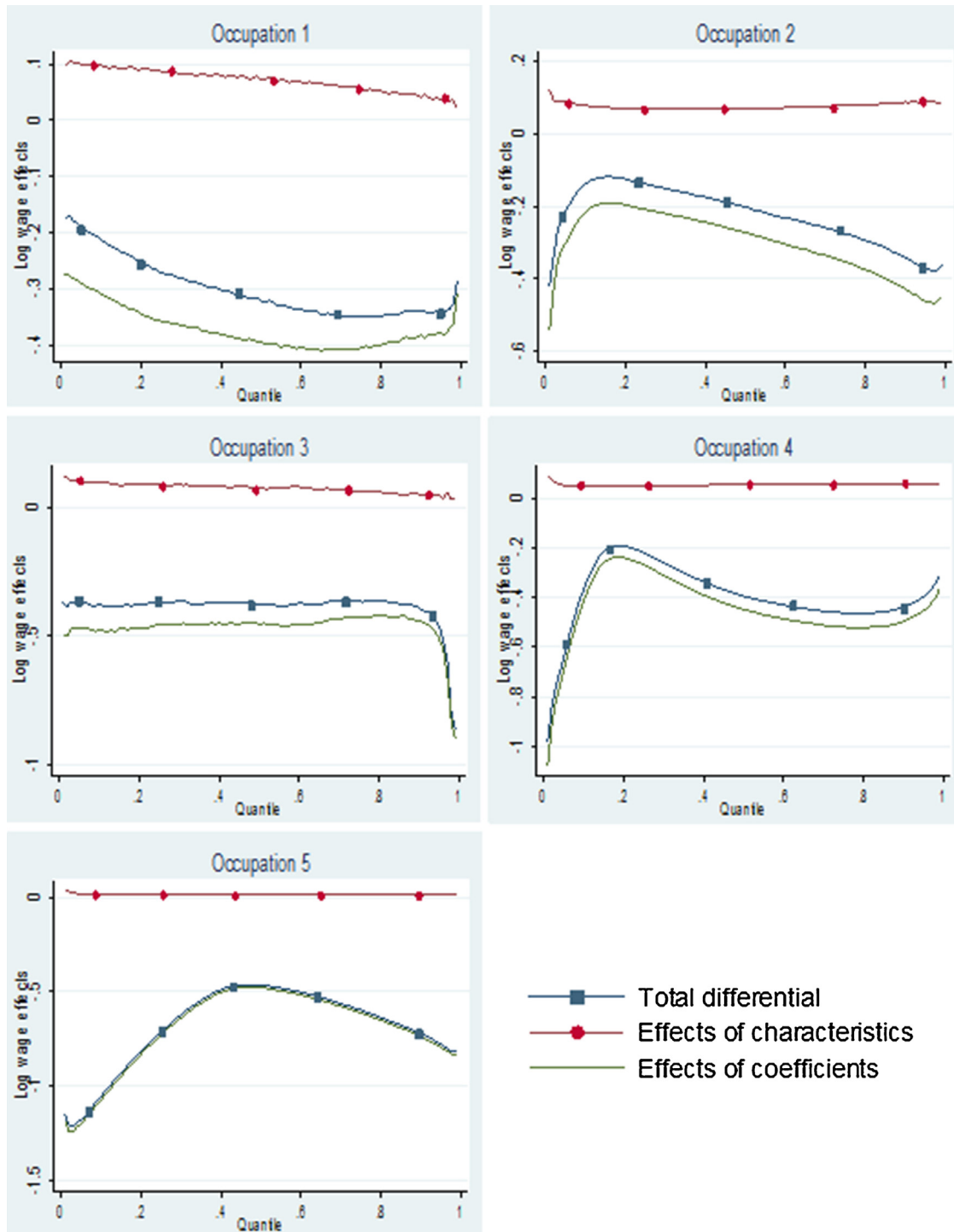


Fig. D1. Gender.

Source: Own elaboration using PNAD/2011 data.

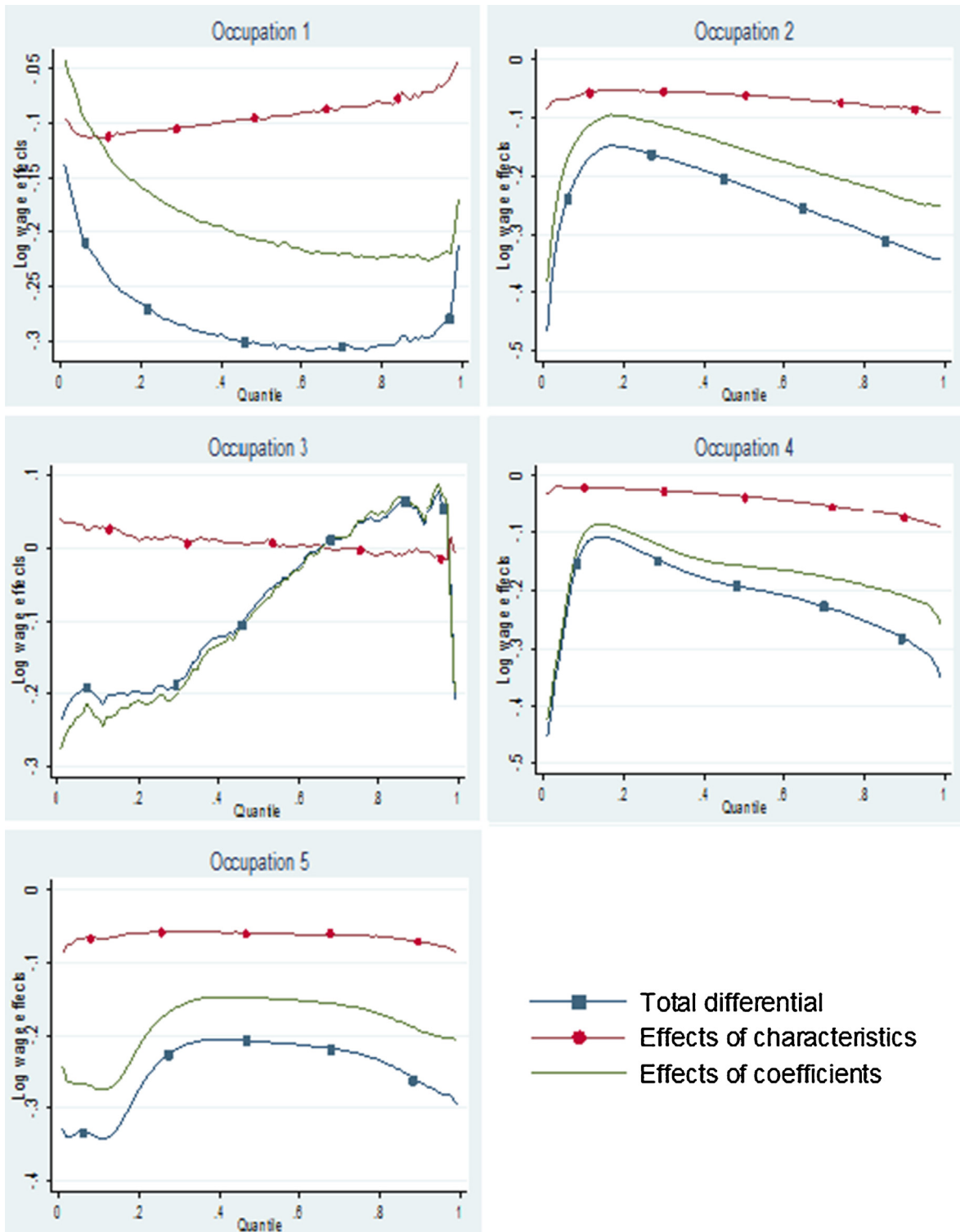


Fig. D2. Ethnicity.

Source: Own elaboration using PNAD/2011 data.

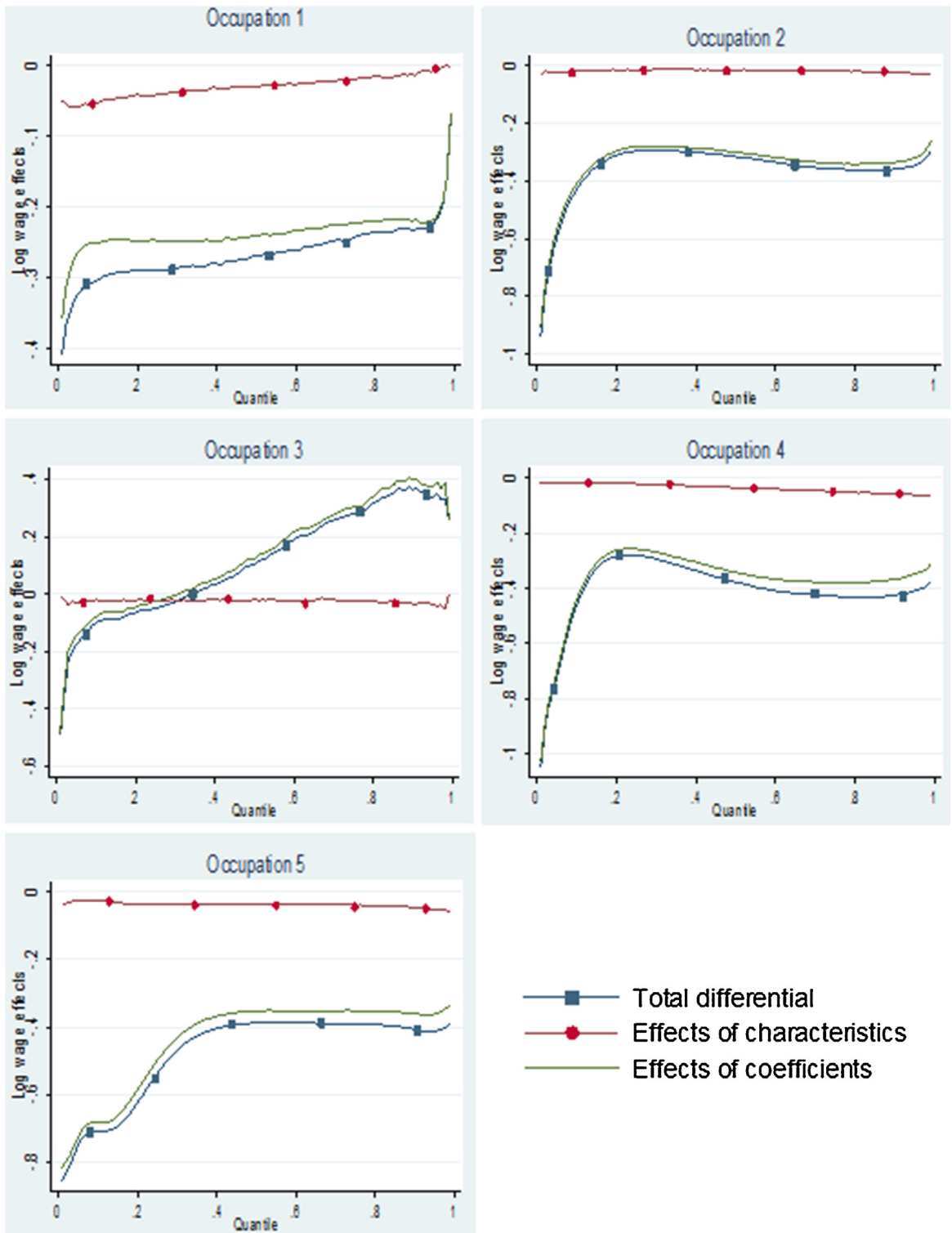


Fig. D3. Northeast region.

Source: Own elaboration using PNAD/2011 data.

References

- Albrecht, J., Bjorkland, A., Vroman, S., 2003. Is there a glass ceiling over Sweden? *J. Labor Econ.* 21, 145–177.
- Albrecht, J., van Vuuren, A., Vroman, S., 2009. Counterfactual distributions with sample selection adjustments: econometric theory and an application to the Netherlands. *Labor Econ.* 16 (4), 383–396.
- Arulampalam, W., Booth, A.L., Bryan, M.L., 2005. Is there a Glass Ceiling over Europe? Exploring the Gender Pay Gap across the Wage Distribution, Mimeo.
- Barros, Ricardo P. et al. 2007. A Queda Recente da Desigualdade de Renda no Brasil. In: Barros, Ricardo P., Foguel, Miguel N., Ulyssa, Gabriel. Desigualdade de Renda no Brasil: uma análise da queda recente. IPEA, Brasília, 1, 107–127.
- Ben-Porath, Y., 1967. The production of human capital and the life cycle of earnings. *J. Polit. Econ.* 75 (4), 352–365.
- Blau, F.D., Kahn, L., 1996. International differences in male inequality: institutions versus market forces. *J. Polit. Econ.* 104 (4), 791–837.
- Blinder, A.S., 1973. Wage discrimination: reduced form and structural estimates. *J. Hum. Resour.* 8 (4), 436–455.
- Buchinsky, M., 1994. Changes in the U.S. wage structure 1963–1987: an application of quantile regression. *Econometrica* 62, 405–458.
- Cameron, A.C., Trivedi, P.K., 2009. *Microeconometrics using Stata*. StataCorp LP, Texas.
- Chamberlain, G., 1994. Quantile regression, censoring and the structure of wages. In: Sins, C. (Ed.), *Advances in Econometrics*. Cambridge University Press, Cambridge, pp. 171–209.
- Coelho, D., et al., 2010. Regressão quantílica com correção para a seletividade amostral: Estimativa dos retornos educacionais e diferenciais raciais na distribuição de salários das mulheres no Brasil. *Pesquisa e Planejamento Econômico* 40 (1), 85–102.
- Dias, J., et al., 2013. Função de Capital Humano dos Estados Brasileiros: Retornos Crescentes ou Decrescentes da Educação? *Pesquisa e Planejamento Econômico* 43 (2), 333–379.
- Dinardo, J., Fortin, N., Lemieux, T., 1996. Labor market institutions and the distribution of wages, 1973–1992: a semiparametric approach. *Econometrica* 64 (5), 1001–1044.
- Erickson, R., Goldthorpe, J.H., 1992. The CASMIN Project and the American dream. *Eur. Soc. Rev.* 8, 283–306.
- Erickson, R., Goldthorpe, J.H., 1987. Commonality and variation in social fluidity in industrial nations. Part I: A model for evaluating the FJH hypothesis. *Eur. Soc. Rev.* 3, 54–77.
- Erickson, R., Goldthorpe, J.H., Portocarrero, L., 1979. Intergenerational class mobility in three western European societies: England, France and Sweden. *Br. J. Sociol.* 30, 415–430.
- Fitzenberger, B., Kurz, C., Mimeo 1997. *New Insights on Earnings Trends across Skill Groups and Industries in West Germany*. Universität Konstanz, Germany.
- Goldthorpe, J.H., Llewellyn, C., Payne, C., 1987. *Social Mobility and Class Structure in Modern Britain*, 2nd ed. Clarendon Press, Oxford.
- González, X., Miles, D., 2001. Wage inequality in a developing country: decrease in minimum wage or increase in education returns. *Empir. Econ.* 26, 135–148.
- Harmon, C., Hogan, V., Walker, I., 2003. Dispersion in the economic return to schooling. *Labor Econ.* 10 (2), 205–214.
- Heckman, J., et al., 2006. Earnings functions, rates of return and treatment effects: The Mincer equation and beyond. In: Hanushek, E.A., Welch, F. (Eds.), *Handbook of the Economics of Education*, Vol. 1. Elsevier Press, Amsterdam.
- Hout, M., Hauser, R.M., 1992. Symmetry and hierarchy in social mobility: a methodological analysis of the CASMIN model of class mobility. *Eur. Sociol. Rev.* 8 (3), 239–266.
- Kassouf, A.L., 1998. Wage gender discrimination and segmentation in the Brazilian labor market. *Econ. Apl.* 2 (2), 243–269.
- Koenker, R., Bassett, G.W., 1978. Regression quantiles. *Econometrica* 46 (1), 33–50.
- Lemieux, T., 2006. The “Mincer Equation”, thirty years after schooling, experience, and earnings. In: Grossbard, S. (Ed.), *Jacob Mincer: A Pioneer of Modern Labor Economics*. Springer Science, New York, NY.
- Machado, J.A.F., Mata, J., 2005. Counterfactual decomposition of changes in wage distributions using quantile regression. *J. Appl. Econometr.* 20 (4), 445–465.
- Martins, P.S., Pereira, P.T., 2004. Does Education reduce wage inequality? Quantile regression evidence from 16 countries. *Labor Econ.* 11, 355–371.
- Mincer, J., 1974. *Schooling, Experience, and Earnings*. Columbia University Press for the National Bureau of Economic Research (NBER), New York.
- Murphy, K., Welch, F., 1990. Empirical age-earning profiles. *J. Labor Econ.* 8 (2), 202–229.
- Nielsen, H., Rosholm, M., 2001. The public–private sector wage gap in Zambia? A quantile regression approach. *Empir. Econ.* 26, 169–182.
- Oaxaca, R.L., 1973. Male female wage differentials in urban labor markets. *Int. Econ. Rev.* 14 (3), 693–709.
- Paredes, D., 2012. The impact of the occupations and economic activities on the gender wage gap using a counterfactual framework. In: *Serie de Documentos de Trabajo en Economía, Working Paper 2012-08*. UCN, Chile.
- Pastore, J., 1979. *Desigualdade e Mobilidade Social no Brasil*. Editora da Universidade de São Paulo, São Paulo.
- Polachek, S.W., 2008. Earnings over the life cycle: The Mincer earnings function and its applications. *Found. Trends Microecon.* 4 (3), 165–272.
- Ribeiro, C.A.C., Scalón, M.C., 2001. Mobilidade de Classe no Brasil em Perspectiva Comparada. *Dados* 44 (01), 2001.
- Rica, S., Dolado, J.J., Llorens, V., 2005. Ceilings and Floors: Gender Wage Gaps by Education in Spain. IZA, Discussion Paper Series 1483, Bonn, 35 pp.
- Santos, R.V., Ribeiro, E.P., 2006. Diferenciais de Rendimentos entre Homens e Mulheres no Brasil revisitado: explorando o “Teto de Vidro”, Available from: http://ww2.ie.ufrj.br/eventos/seminarios/pesquisa/texto06_05_02.pdf
- Scalón, M.C., 1999. *Mobilidade Social no Brasil: Padrões e Tendências*. IUPERJ-UCAM, Rio de Janeiro, Revan.
- Schultz, T.W., 1961. Investment in human capital. *Am. Econ. Rev.* 51 (1), 1–17.
- Silva, N.V., Roditi, D., 1988. Et Plus Ça Change. Tendências Históricas da Fluidez Social no Brasil. In: Hasenbalg, C., Silva, N.V. (Eds.), *Estrutura Social, Mobilidade e Raça*. IUPERJ/Vértice, Rio de Janeiro.