Wage Discrimination in Brazil: Inferences based on Unconditional Quantile Regressions

Paulo Roberto de Sousa Freitas Filho *

July 20, 2015

Abstract

Discrimination increases income inequality, which is a major problem in Brazil. In this paper we estimate the gender and racial discrimination, using the data from the 2010 Brazilian Censo and the method proposed by Firpo, Fortin and Lemieux (2009) - reweighing and recentered influence function regressions. This method overcomes the limitations of Oaxaca-Blinder decomposition, (see Barsky (2002)), and the decompositions proposed by Machado and Mata (2005), and Melly (2006). We also estimate the main components of the discrimination. Our results suggest that wage discrimination between males and females does not present sharp variations across the quantiles, but the racial discrimination increases along the quantiles of the wage distribution. The decomposition of the unexplained effects shows that education, experience and region are the most important components of the racial discrimination. We estimate the discrimination in each of the five regions of Brazil, and found evidences that the it is smaller in north and northeast than in other regions.

Key words: Labor discrimination, Discrimination, Quantile Regressions
JEL Classification: J7, J71, C21

Resumo

A discriminação aumenta a desigualdade de renda, que é um grande problema no Brasil. Neste trabalho, estimamos a discriminação de gênero e racial, utilizando dados do Censo Brasileiro de 2010 e o método proposto por Firpo, Fortin e Lemieux (2009) - regressões da função de influência recentrada e ponderada. Este método supera as limitações da decomposição de Oaxaca-Blinder , (ver Barsky (2002)), e das decomposições propostas por Machado e Mata (2005), e Melly (2006). Também estimamos os principais componentes da discriminação. Nossos resultados sugerem que a discriminação salarial entre homens e mulheres não apresenta variações acentuadas entre os quantis, mas a discriminação racial aumenta ao longo dos quantis da distribuição salarial. A decomposição dos efeitos inexplicáveis mostra que a educação, a experiência e a região são os componentes mais importantes da discriminação racial. Estimamos o a discriminação em cada uma das cinco regiões do Brasil e encontramos evidências de que ela é menor no Norte e Nordeste do que em outras regiões.

Palavras Chaves: Discriminação no mercado de trabalho, Discriminação, Regressões Quantílicas
Área 13 - Economia do Trabalho

*Department of Economics, Universidade Federal de Pernambuco.

1 Introduction

There are many evidences that racial and gender discrimination occurs in the labor market. Bertrand and Mullainathan (2003) performed a field experiment, responding to help-wanted ads in some U.S. newspapers with fictitious resumes. Each resume was assigned either a very African American sounding name or a very White sounding name. The fictitious people with white names received 50 percent more callbacks for interviews. Goldin and Rouse (1997) analyzed the effects of the use of "blind" auditions with a "screen" to conceal the identity of the candidate from the jury. The screen increased by 50% the probability a woman will be advanced out of certain preliminary rounds, and greatly increased the probability that a female contestant will be the winner in the final round. In 1970 - prior to the change in the audition policy - less than 5% of the musicians in the top five orchestras in the United States were female, in 1997 this share increased to 25%.

These authors succeeded in estimating discrimination in the hiring process. Nevertheless, it is difficult to estimate the wage discrimination in the labor market. Fortin, Lemieux and Firpo (2010) review many decomposition methods that can be used to estimate discrimination. They point out the advantages and limitations of each method. They explain that, at first, the decomposition methods consisted of decompositions of mean wages differentials. Later the methods evolved and were used to decompose variations, and finally the conditional distribution. They also described the reweighting methods, which are non-parametric.

Many authors have estimated the wage discrimination using the method proposed by Oaxaca (1973) and Blinder (1973). This method approximate the conditional expectations by best linear predictors. Barsky et al. (2002) showed that it may yield misleading results when the true conditional expectation functions are nonlinear.

The decomposition method proposed by Machado and Mata (2005) can be used to evaluate distributional effects, but these authors provided no econometric theory for the quantile regression decomposition estimators (Chernozhukov, Fernández-Val and Melly, 2012). This method is computationally demanding, so that if the data sets contains more than a few thousand observations, it becomes quite cumbersome. Unfortunately, the MM does not provide a way of performing the detailed decomposition for the composition effect (Fortin, Lemieux and Firpo (2010)). For example, when the proportion of skilled workers in the economy increases there is a change in the composition of the the groups of workers, and this change causes a reduction in the the inequality, and it is called a composition effect. For Fortin, Lemieux and Firpo (2010), this is a major drawback since the detailed decomposition of the composition effects into the contribution of each variable is always clearly interpretable. Another drawback of the MM decomposition is that it is only consistent if the right functional form is used for quantiles. Therefore, this method requires the choice of the right functional form for each and every quantiles, and making sure that the specification is correct is a very difficult empirical exercise.

Melly (2006) estimates the conditional distribution by parametric and nonparametric quantile regressions. The parametric quantile regression is an extension of the Oaxaca/Blinder decomposition of means to the full distribution. The nonparametric quantile regression is an efficient local-linear regression estimator for quantile treatment effects. This decomposition technique has one important limitation, it does not allow for computing detailed decompositions that allow the computation of the effect of each variable on the unconditional quantile wage distribution.

Firpo, Fortin and Lemieux (2009) developed the reweighing and recentered influence function regressions, and a decomposition technique that enables the analysis of the contribution of each explanatory variable after the estimation of the wage structure and composition effects (see Firpo, Fortin and Lemieux (2007)). Salardi (2012) is the first work that used this method to estimate wage discrimination in Brazil. One limitation of this work is that it fails to give detailed results of the Firpo, Fortin and Lemieux (2009) decomposition, since it presents results of a great number of decompositions (besides the standard Oaxaca-Blinder decomposition, it also presents the Brown, Moon and Zoloth (1980), Machado and Mata (2005), and Melly (2006) decompositions).

The main contribution of this paper is to present detailed results of the decomposition of the wage gap in Brazil between whites and non-whites, and males and females using the reweighing and recentered influence function regressions. We also break down the explained and unexplained differences in earnings between these groups into the contribution of each explanatory variable using a generalized Oaxaca-Blinder method, which do not require the linearity assumption. To our knowledge it is the only work that estimate the discrimination in each of the five regions of Brazil, using the method proposed by Firpo, Fortin and Lemieux (2009). This paper proceeds as following: the next section describes the method used in this work, the third section explains the results, the fourth section concludes and the appendix describes the discrimination in the regions of Brazil.

2 Methods

2.1 Data

The data used in the econometric analysis were obtained from the Brazilian Censo for the year 2010. The sample consists of 200000 observations of males and females between 40 and 49 years of age, with more than 8 years of schooling and positive income. The sample selection was made according to Angrist, Chernozhukov and Fernández-Val(2006). We used the following variables:

perw individual sampling weights

logw average log weekly wage, calculated as the log of reported monthly income from work divided by weeks worked

educ years of schooling

black indicator variable for race that assumes the value 1 for blacks, browns and native Americans, and 0 for whites and yellows.

female indicator variable for gender that assumes the value 1 for females and 0 for males.

reg indicator variable for region that assumes the value 1 for north and northeast, and 0 for other regions.

sit indicator variable for home location that assumes value 1 for rural location.

age age in years

exper potential experience, calculated as $age - educ - 6$

exper2 square of exper

actv1…actv20 dummies for sector of the economy, aggregated in 20 sector according to the Classificação Nacional de Atividades Econômicas Domiciliar 2.0 - CNAE-Domiciliar 2.0: agriculture, extrative industry, transformation Industry, electricity and gas, water and waste, construction, vehicle commerce and maintenance, transport, food, communication, finance, real state, professional and scientific, administration and other services, public administration, education, health and social services, arts and sports, other services, and domestic services. This last dummy variable is dropped from the model to avoid perfect multicolinearity (Wooldridge (2009)).

Panel (a) of figure 1 shows the density of the log of weekly wages of males and females. We can notice that the distribution of the wages is more right skewed for females than for males, which indicate the presence of gender discrimination. Analyzing the graphs of the wage distribution of whites and non-whites, as shown in panel (b) of figure 1, we can also presume that there is race discrimination.

In order to estimate and decompose the discrimination, we regress the log of weekly wage on education, experience, square of experience, female (or race) dummy, region, location and activity dummies, following the methodology described in the next section. We also estimate the discrimination in each region, to identify where the discrimination is located, and present the results in the appendix.

2.2 Reweighing and Recentered Influence Function Regressions

This section follows the notation from Firpo, Fortin and Lemieux (2007). Suppose that the sample of N individuals indexed by $i = 1, \dots, N$, is divided in two groups, 1 and 0. Therefore $N = N_1 + N_0$, where N_1 and N_0 are the number of individuals in each group. We denote the conditional probability that an individual i is in group 1 given $X = x$, as $p(x) = Pr[T = 1|X = x]$. This conditional probability is sometimes called the "propensity-score".

The wage depends on some observed variables X_i and on some unobserved variables $\varepsilon_i \in \mathbb{R}^m$ and is determined by wage structure functions

$$Y_{it} = g_t(X_i, \varepsilon_i), \qquad \text{for } t = 0, 1 \tag{1}$$

We can non-parametrically identify the distributions of $Y_1|T = 1 \sim^d F_1$ and of $Y_0|T = 0 \sim^d F_0$, from observed data on (Y, T, X). But we need more assumptions to identify the counterfactual distribution of $Y_0|T = 1 \sim^d F_C$. The counterfactual distribution F_C is the one that would have prevailed under the wage structure of group 0, but with the distribution of observed and unobserved characteristics of group 1. Consider these three distributions conditional on X: $Y_1|X, T = 1 \sim^d F_{1|X}$, $Y_0|X, T = 0 \sim^d F_{0|X}$ and $Y_0|X, T = 1 \sim^d F_{C|X}$.

Let ν_1,ν_0 and ν_C be a functional (variance, median, quantile, Gini, etc.) of the conditional joint distribution of $(Y_1, Y_0)|T$, and $F\nu$ is a class of distribution functions such that $F \in F\nu$ if $||\nu(F)|| < +\infty$. The difference in the ν's between the two groups is the difference in wages measured in terms of the distributional statistic ν, and is called the ν-overall wage gap.

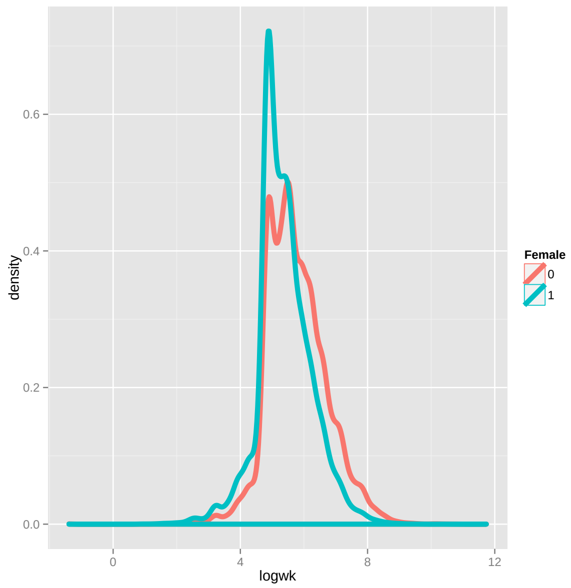
$$\Delta_O^\nu = \nu(F_1) - \nu(F_0) = \nu_1 - \nu_0 \tag{2}$$

We can decompose equation (2) in two parts, using the fact that X can be unevenly distributed across groups, then

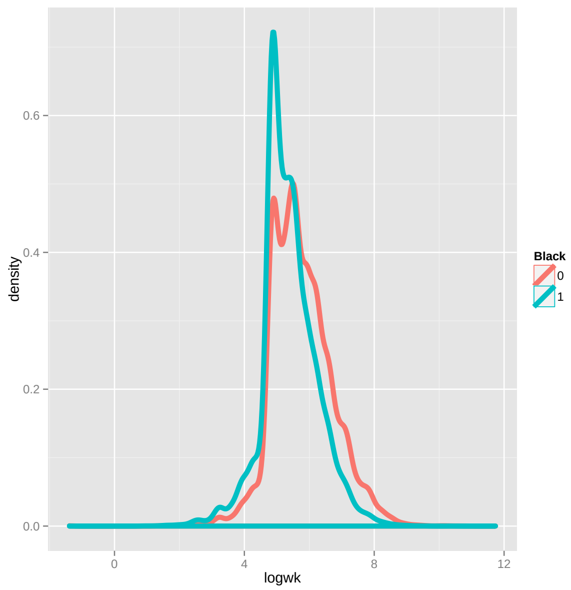
$$\Delta_O^\nu = (\nu_1 - \nu_c) + (\nu_c - \nu_0) = \Delta_S^\nu + \Delta_X^\nu \tag{3}$$

Figure 1: Densities of the log of weekly wage by group

(a) Density of the log of weekly wages of males and females



(b) Density of the log of weekly wages of whites and non-whites (black=1)



Where the first term Δ_S^ν corresponds to the effect on ν of a change from $g_1(\cdot,\cdot)$ to $g_0(\cdot,\cdot)$ keeping the distribution of $(X,\varepsilon)|T=1$ constant. This is called the wage structure effect or the unexplained difference effect. With no other restrictions, the second term Δ_X^ν corresponds to the effect of changes in distribution from the one of $(X,\varepsilon)|T=1$ to that of $(X,\varepsilon)|T=0$, keeping the "wage structure" $g_0(\cdot,\cdots)$ constant. This is called the composition effect or the explained difference effect.

Firpo, Fortin and Lemieux (2007) use the reweighing and recentered influence function regressions to analyze how polarization of U.S. male wages that took place between the late 1980s and the mid 2000s was affected by factors such as de-unionization, education, occupations and industry changes. Given the total change in the distribution of wages Δ_O^ν, they tried to answer two questions. The first was, what is the change in the distribution of wages between the late 1980s the mid 2000s, that are associated with the change in the wage structure (Δ_S^ν or "wage structure"), if worker characteristics were the same in the two periods. The second was, what is the change in the distribution of wages in these two periods, that are associated with the change in worker characteristics (Δ_X^ν or "composition effect"), keeping the wage structure constant.

Here we need the conditional independence assumption, that is, we assume that ε is independent of T given X. This implies that Δ_X^ν only reflect changes in the distribution X. This assumption is also crucial for identification of F_C and ε_C. We do not need any assumption about the format of $g_0(\cdot,\cdot)$ and $g_0(\cdot,\cdot)$, but if we had imposed some assumption on the format of these function then we could have relaxed the conditional independence assumption. In the Oaxaca-Blinder decomposition it is assumed that $g_1(X,\varepsilon)=X^T\beta_1+\epsilon_1$, $g_0(X,\varepsilon_0)=X^T\beta_0+\epsilon_0$, and that

$$E[\epsilon_t|X,T=t]=0 \tag{4}$$

The assumption that the functions $g_1(\cdot,\cdot)$ and $g_0(\cdot,\cdot)$ are linear can be plausible in many cases, but the assumption of exogeneity described by equation (4), is more difficult to be accepted, since if any variable that affects wages (like ability) is missing in the model, we cannot affirm that this assumption is valid.

We can identify the parameters of interest under the common assumptions of Ignorability (sometimes called unconfoundedness) and Overlapping Support (or Common Support). The Ignorability assumption should be analyzed in each specific case, as it is more plausible in some cases than in others. In our specific case, it states that the distribution of the unobserved explanatory factors in the wage determination is the same across groups 1 and 0, once we condition on a vector of observed components. Formally, the Ignorability assumption is: Let (L,X,ε) have a joint distribution. For all x in $\mathcal{X}$, ε is independent of T given $X=x$.

The Overlapping Support assumption requires that there be an overlap in observable characteristics across groups, in the sense that there no value of x in X such that it is only observed among individuals in group 1. In gender wage gap decompositions where some of the detailed occupations are only held by men or by women, this assumption is not valid, but in our case the we consider only 21 sectors (or types of occupation), therefore the Overlaping Support assumption is plausible. Formally, the Overlaping Support assumption is: For all x in $\mathcal{X}$, $p(x)=Pr[T=1|X=x]<1$. Futhermore, $Pr[T=1]>0$.

2.2.1 Step 1: Reweighing

The distributions F_0,F_1 and F_C can be estimated non-parametrically using the weights:

$$\omega_1(T) \equiv \frac{T}{p} \qquad \omega_0(T) \equiv \frac{1-T}{1-p} \qquad \omega_C(T,X) \equiv \left(\frac{p(X)}{1-p(X)}\right) \cdot \left(\frac{1-T}{p}\right)$$

where $p(X) = Pr[T = 1|X = x]$ is the proportion of people in the combined population of two groups that is in group 1, given that those people have $X = x$, and p is the unconditional probability. $\omega_1(T)$ and $\omega_0(T)$ transform features of the marginal distribution of Y into features of the conditional distribution of Y_1 given $T = 1$, and of Y_0 given $T = 0$, respectively. $\omega_C(T)$ transforms features of the marginal distribution of Y into features of the counterfactual distribution of Y_0 given $T = 1$. By identifying F_C we can identify the functional $\nu(F_C)$ (variance, median, quantile, Gini, etc.), and from equations (2) and (3) we can identify and Δ_S^ν and Δ_X^ν. The following result permits us to identify F_C.

Theorem 1 (Inverse Probability Weighing) *Under assumptions of Ignorability and Overlapping Support:*

a. $F_t(y) = \mathbb{E}[\omega_t(T) \cdot \mathbb{1}\{Y \leq y\}] \qquad t = 0, 1$

b. $F_C(y) = \mathbb{E}[\omega_C(T,X) \cdot \mathbb{1}\{Y \leq y\}]$

By weighing the observations with the inverse probabilities of belonging to group 0 or 1 given T, as stated in theorem 1, we can obtain the distribution and the functionals of the distributions, which enable us to obtain Δ_S^ν and Δ_X^ν, as the following result states:

Theorem 2 (Identification of Wage Structure and Composition Effects) *Under assumptions of Ignorability and Overlapping Support:*

a. Δ_S^ν *and* Δ_X^ν *are identifiable from the data on* (Y, T, X)*;*

b. *if* $g_1(\cdot,\cdot) = g_0(\cdot,\cdot)$ *then* $\Delta_S^\nu = 0$

c. *if* $F_{X|T=1} \sim F_{X|T=0}$*, then* $\Delta_X^\nu = 0$

Part b of Theorem 2 states that when there are no group differences in the wage determination functions, then we should find no wage structure effects (unexplained effects), while part c states that if there are no group differences in the distribution of the covariates, there will be no composition effects.

We now explain how to estimate the weighting functions. The distributional statistics ν_1, ν_0 and ν_C can be computed directly from the appropriately reweighted samples. The three weighting functions we are interested in are $\omega_1(T)$, $\omega_0(T)$, and $\omega_C(T,X)$. The first two weights are are estimated by:

$$\hat{\omega}_1(T) = \frac{T}{\hat{p}} \quad \text{and} \quad \hat{\omega}_0(T) = \frac{1-T}{1-\hat{p}}$$

, where $\hat{p} = N^{-1}\sum_{i=1}^{N} T_i$. The third weight is estimated by

$$\hat{\omega}_C(T,X) = \frac{1-T}{\hat{p}} \cdot \left(\frac{\hat{p}(X)}{1-\hat{p}(X)}\right)$$

, where $\hat{p}(\cdot)$ is an estimator of the true probability of being in group 1 given X. For details of the parametric and the non-parametric approaches to estimate this probability, see Firpo, Fortin and Lemieux (2007). We used the following normalization to have weights summing up to one:

$$\hat{\omega}_1^*(T_i) = \frac{\hat{\omega}_1(T_i)}{\sum_{j=1}^N \hat{\omega}_1(T_j)} = \frac{T_i}{N \cdot \hat{p}}$$

,

$$\hat{\omega}_0^*(T_i) = \frac{\hat{\omega}_0(T_i}{\sum_{j=1}^N \hat{\omega}_0(T_j)} = \frac{1 - T_i}{N \cdot (1 - \hat{p})}$$

,

$$\hat{\omega}_C^*(T_i, X_i) = \frac{\hat{\omega}_C(T_i}{\sum_{j=1}^N \hat{\omega}_C(T_j)} = \frac{(1 - T_i) \cdot \left(\frac{\hat{p}(X_i}{1 - \hat{p}(X_i)}\right)}{\sum_{j=1}^N (1 - T_j) \cdot \left(\frac{\hat{p}(X_i}{1 - \hat{p}(X_i)}\right)}$$

.

We estimate ν_1, ν_0 and ν_C by replacing the CDF by the empirical distribution function: $\hat{\nu}_t = \nu(\hat{F}_t)$, $t = 0, 1$ and $\hat{\nu}_C = \nu(\hat{F}_C)$, where

$$\hat{F}_t(y) = \sum_{i=1}^N \hat{\omega}_t{}^*(T_i) \cdot \mathbb{1}\{Y_i \leq y\}, t = 0.1$$

$$\hat{F}_C(y) = \sum_{i=1}^N \hat{\omega_C}^*(T_i, X_i) \cdot \mathbb{1}\{Y_i \leq y\}$$

.

Another way to get estimates of a distribution statistic $\hat{\nu}$ is to use standard software programs, to compute distributional statistics directly from the observations on Y, weighted using the appropriate weighting factor. We estimate the composition and the wage structure effects by using the estimated distributional statistics as $\hat{\Delta}_S^\nu = \hat{\nu}_1 - \hat{\nu}_C$ and $\hat{\Delta}_X^\nu = \hat{\nu}_C - \hat{\nu}_0$.

In this paper we use quantiles as distributional measures for the decomposition of wage distributions. In decompositions of the gender wage gap, they are used to differentiate the effects of the discrimination in the middle of the distribution from its impact in the tails. To carry out the decomposition of the median, we first estimate me_t, $t = 0, 1$ and me_C by reweighting as $\hat{me}_t = argmin_q \sum_{i=1}^N \hat{\omega}_t(T_i) \cdot |Y_i - q|, t = 0, 1$ and $\hat{me}_C = argmin_q \sum_{i=1}^N \hat{\omega}_C(T_i) \cdot |Y_i - q|$. The estimators for the wage gaps are computed as: $\hat{\Delta}_O^{me} = \hat{me}_1 - \hat{me}_0$, $\hat{\Delta}_S^{me} = \hat{me}_1 - \hat{me}_C$ and $\hat{\Delta}_X^{me} = \hat{me}_C - \hat{me}_0$.

2.2.2 Step 2: Application of the UQR methodology

Let $\nu = \nu(F)$ be a general functional. The influence function is

$$IF(y; \nu, F) = lim_{\epsilon \to 0} \frac{(\nu(F_\epsilon) - \nu(F))}{\epsilon}$$

, where $F_\epsilon(y) = (1 - \epsilon)F + \epsilon\delta_y$, $0 \leq \epsilon \leq 1$, and where δ_y is a distribution that only puts mass at the value y. It can be shown that $\int_{-\infty}^{\infty} IF(y; \nu, F) dF(y) = 0$.

We use a recentered influence function: $RIF(y; \nu, F) = \nu(F) + IF(y; \nu, F)$ whose expectation is the original ν:

$$\int_{-\infty}^{\infty} RIF(y; \nu, F) dF(y) = \int_{-\infty}^{\infty} (\nu(F) + IF(y; \nu, F)) dF(y) = \nu(F)$$

.

The τ-th quantile of the distribution F is defined as the functional $q_\tau = Q(F,\tau) = \inf\{y|F(y) > \tau\}$, and its influence function is:

$$IF(y;q_\tau) = \frac{\tau - \mathbb{1}\{y \leq q_\tau\}}{f_y(q_\tau)}$$

.

The rescaled influence function of the τ-th quantile is

$$RIF(y;q_\tau) = q_\tau + IF(y;q_\tau) = q_\tau + \frac{\tau - \mathbb{1}\{y \leq q_\tau\}}{f_y(q_\tau)}$$

.

The influence function of the median is,

$$IF(y;me) = \frac{\frac{1}{2} - \mathbb{1}\{y \leq me\}}{f_y(me)}$$

.

The rescaled influence function is

$$RIF(y;me) = me + \frac{\frac{1}{2} - \mathbb{1}\{y \leq me\}}{f_y(me)}$$

.

In order to estimate the linear RIF-regressions, we compute the rescaled influence function for each observation by plugging the sample estimate of the median, $\hat{me}$, and estimating the density at the sample median, $\hat{f}(\hat{me})$. For the median of $Y_1|T=1$, we would use

$$\widehat{RIF}(y;me_1) = \hat{me}_1 + \left(\hat{f}_1(\hat{me}_1)\right)^{-1} \cdot (\frac{1}{2} - \mathbb{1}\{\leq \hat{me}_1\})$$

where $\hat{f}_1(\cdot)$ is a consistent estimator for the density of $Y_1|T=1$, $f_1(\cdot)$. Kernel methods can be used to estimate the density. We use the guaussian kernel function with halfwidth of kernel equals to 0.06. The RIF-regressions are then estimated by replacing the usual dependent variable, Y , by the estimated value of $RIF(y;me_1)$. The resulting regression coefficients are

$$\hat{\gamma}_t^{me} = \left(\sum_{i=1}^{N} \hat{\omega}_t(T_i)X_iX_i'\right)^{-1} \cdot \sum_{i=1}^{N} \hat{\omega}(T_i)X_i\widehat{RIF}(Y_i;me_t), \quad t=0,1$$

$$\hat{\gamma}_C^{me} = \left(\sum_{i=1}^{N} \hat{\omega}_C(T_i,X_i)X_iX_i'\right)^{-1} \cdot \sum_{i=1}^{N} \hat{\omega}_C(T_i,X_i)X_i\widehat{RIF}(Y_i;me_C)$$

.

Then, we get:

$$\Delta_S^{me} = \mathbb{E}[X,T=1]^T \cdot (\hat{\gamma}_1^{me} - \hat{\gamma}_C^{me})$$

,

$$\Delta_X^{me} = (\mathbb{E}[X|T=1] - \mathbb{E}[X|T=0])^T \cdot \hat{\gamma}_0^{me} + \hat{R}^{me}$$

, where $\hat{R}^{me} = \mathbb{E}[X|T=1]^T \cdot (\hat{\gamma}_C^{me} - \hat{\gamma}_0^{me})$

3 Results

In this section we show the decomposition of the wage gap between males and females, and whites and non-whites in Brazil. We decompose the wage differentials in explained and unexplained effects. The unexplained effects give an estimate of the discrimination. The figures show the decomposition of the wage differentials from the 5th until the 95th quantile of the wage distribution, and the estimates are made for each five quantiles in this range.

Panel *a* of figure 2 displays the total wage differential between males and females, the explained effect and the unexplained effects. The wage discrimination between males and females does not present sharp variations across the quantiles of the wage distribution. It is greater in the 5th and the 35th quantiles. Table 1 displays the wage gap decomposition and the decomposition of the unexplained effects for selected quantiles of the wage distribution. As we use log of weekly wage as the depend variable of our estimation, all values are in log of weekly wages. As we can see in this table, the gender discrimination varies around its mean value of 0.04 across the quantiles.

Panel *b* of figure 2 displays the decomposition of the unexplained effects. Activity is the greater component of the unexplained differences. This means that there are activities in which women receive smaller wages than men. These results suggest that the gender discrimination is not generalized to all activities, for if it were true, the activity would not be a important component of the unexplained effects. If, for example, education were the main component of the discrimination, then we would conclude that in general women with the same educational level than men would receive a smaller wage.

These results differ from Salardi (2012), who found evidences that education is the primary contributor to differences in endowments, and that experience is an important contributor to the unexplained wage gap between male and female and white and non-white workers, using the RIF-OLS technique developed by Firpo, Fortin and Lemieux (2009). Bartalotti (2007), using the Machado-Mata decomposition, found evidences that the gender discrimination in Brazil increases smoothly from the lower quantiles until the 85th, and it increases sharply thereafter. Our results suggest that the gender discrimination in Brazil in greater in the 35th quantile, presents some sharp variations in the lower quantiles, and decreases in the upper quantiles.

Figure 3 is similar to figure 2, and presents the estimates of the race discrimination. Panel *b* of figure 3 shows that race discrimination increases along the quantiles of the wage distribution and it is greater than gender discrimination. The total wage gap also is greater between whites and non-whites than between males and females, and this is more evident in the upper quantiles. Table 1 shows that the total wage gap is -0.062 in the 10th quantile, and increases to 0.212 in the 90th quantile.

Our results are similar to those found by Álvarez (2013), who analyzed the race wage gap in Brazil using the Melly (2006) decomposition, and data from the PNAD for the years 2001 and 2011. The graphs of the race wage gap and discrimination have a U-shape (specially in 2011), with the minimum point located at the 20th quantile, but the discrimination increases in the upper quantiles. He found that the explained and the unexplained effects were roughly equal, therefore approximately half of the wage gap occurs because of race discrimination.

The decomposition of the unexplained effects is displayed in panel *b* of figure 3. Education, experience and region are the most important components of the race discrimination. Education and experience become more important, while region becomes less important in the upper quantiles. Table 1 shows that in the 10th quantile the components of the discrimination are very similar, but education, experience and region stand out in the quantiles above the 50th quantile.

These results imply that the non-whites which are on the higher quantiles of the wage dis-

Figure 2: Gender Discrimination and Decomposition of Unexplained Effects

(a) Wage Gap Decomposition

(b) Decomposition of Unexplained Effects

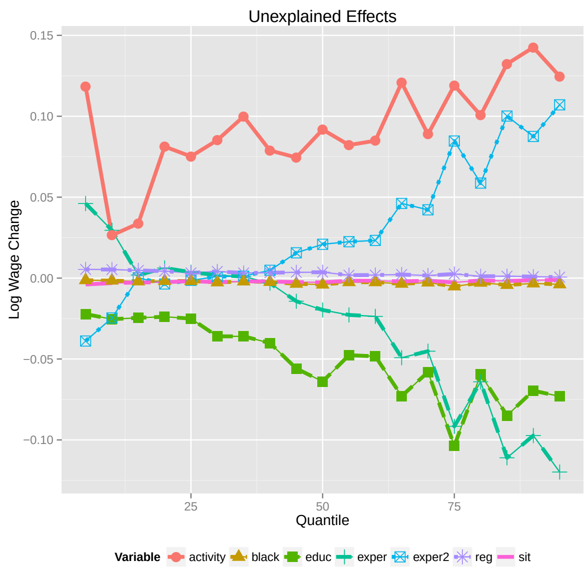
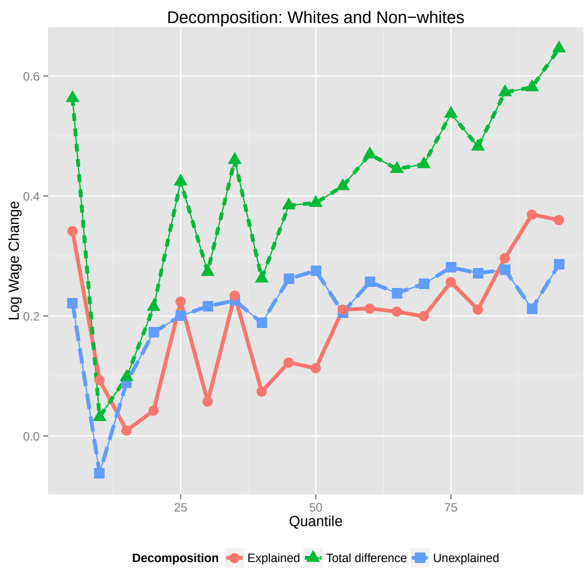


Figure 3: Race Discrimination and Decomposition of Unexplained Effects

(a) Wage Gap Decomposition



(b) Decomposition of Unexplained Effects

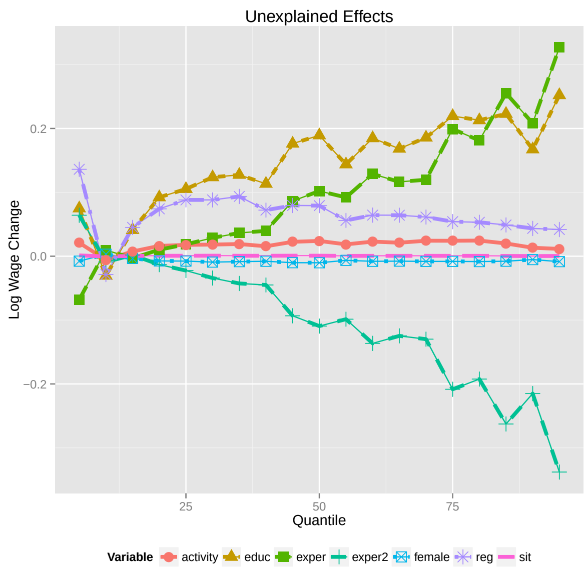


Table 1: Decomposition Results

	Males and Females					Whites and non-whites				
Decomposition of The Wage Gap										
Quantile	**10**	**25**	**50**	**75**	**90**	**10**	**25**	**50**	**75**	**90**
Total Wage Gap	0.104	0.406	0.481	0.363	0.396	0.031	0.424	0.388	0.537	0.581
	(0.007)	(0.003)	(0.006)	(0.009)	(0.008)	(0.002)	(0.004)	(0.006)	(0.006)	(0.007)
Unexplained	0.007	0.052	0.026	0.003	0.059	-0.062	0.200	0.275	0.281	0.212
	(0.003)	(0.002)	(0.004)	(0.005)	(0.004)	(0.001)	(0.002)	(0.003)	(0.004)	(0.003)
Explained	0.098	0.354	0.455	0.360	0.337	0.093	0.224	0.113	0.256	0.369
	(0.006)	(0.003)	(0.005)	(0.008)	(0.008)	(0.002)	(0.004)	(0.005)	(0.005)	(0.006)
Decomposition of Unexplained Effects										
educ	-0.025	-0.025	-0.064	-0.104	-0.070	-0.031	0.105	0.189	0.219	0.167
	(0.001)	(0.001)	(0.002)	(0.004)	(0.003)	(0.001)	(0.002)	(0.003)	(0.003)	(0.003)
exper	0.030	0.004	-0.020	-0.092	-0.097	0.010	0.019	0.102	0.199	0.208
	(0.003)	(0.001)	(0.002)	(0.006)	(0.007)	(0.002)	(0.004)	(0.006)	(0.009)	(0.010)
exper2	-0.025	-0.001	0.021	0.085	0.088	-0.009	-0.022	-0.109	-0.208	-0.215
	(0.003)	(0.001)	(0.002)	(0.006)	(0.006)	(0.002)	(0.004)	(0.006)	(0.008)	(0.009)
black	-0.002	-0.002	-0.004	-0.005	-0.003					
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)					
female						0.003	-0.008	-0.010	-0.008	-0.005
						(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
reg	0.005	0.003	0.004	0.003	0.001	-0.029	0.088	0.079	0.054	0.043
	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
sit	-0.003	-0.002	-0.003	-0.003	-0.001	0.000	0.001	0.001	0.001	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
actv	0.026	0.075	0.092	0.119	0.143	-0.006	0.018	0.024	0.024	0.013
	(0.016)	(0.006)	(0.011)	(0.015)	(0.012)	(0.002)	(0.005)	(0.005)	(0.005)	(0.004)

Notes: Dependent variable: Log of weekly wages. Standard errors are in parenthesis. The unexplained effects corresponds to discrimination.

tribution have a smaller return to education than the whites. Region is a important component of the discrimination, therefore in some regions the discrimination is greater than in others. The region dummy is 1 for north and northeast and 0 for other regions, and these two regions have a great portion of non-whites. Thus it is likely that the discrimination is smaller in north and northeast than in the other regions, where non-whites are minority. To verify this, we estimate the gender and race discrimination for each of the five regions of Brazil. The results are shown in figures 4 and 5 in the appendix. Figure 4 shows that gender discrimination is very small in north and northeast regions, and it is greater in the other regions. Figure 5 shows that race discrimination is smaller in north and northeast than in other regions, but in this two regions race discrimination is greater than gender discrimination.

4 Conclusions

In this paper we decomposed the wage gap in Brazil between whites and non-whites, and males and females using the reweighing and recentered influence function regressions. The wage discrimination between males and females does not present sharp variations across the quantiles of the wage distribution. It is greater in the 5th and the 35th quantiles. Our results suggest that gender discrimination is not generalized to all activities, since activity is the main component of the unexplained effects. We also found evidences that gender discrimination is very small in north and northeast regions, and it is greater in the other regions.

The racial discrimination increases along the quantiles of the wage distribution and it is greater than gender discrimination. The decomposition of the unexplained effects shows that education, experience and region are the most important components of the race discrimination. This means that whites have a greater return to education and experience than non-whites, and that the discrimination is greater in some regions than in others. The estimation of the racial discrimination for each of the five regions of Brazil shows that race discrimination is smaller in north and northeast than in other regions. This occurs because non-whites are minority in south, southeast and midwest, therefore it is more likely that the discrimination is greater in these regions than in north and northeast.

One limitation of this paper is that it does not identify the activities where gender and race discrimination occurs. Another limitation is that it does not discuss the problem of sensitivity to the choice of omitted baseline category (see Oaxaca and Ransom (1999)). To avoid perfect multicollinearity, one of the dummy variables is omitted in the regression equation. This variable represents the baseline category, for the coefficients of the remaining dummy variables are interpreted as deviations from this variable. The results of the Oaxaca-Blinder decompositions are sensitive to the researcher's choice of the omitted baseline category. Although, this paper fails to analyze this problem, it is possible that this problem occurs because of the linearity assumption of the Oaxaca-Blinder decomposition, and this assumption is not required in our estimates, therefore it is not necessary the perform the adjustment proposed by Gardeazabal and Ugidos (2004).

Brazil is one of the most unequal countries in the world. Racial and gender discrimination may be important factors contributing to this inequality. Although some policies are being created to reduce the inequality of opportunity - in 2004 the Universidade Federal de Brasília was the first public university to adopt a quota system to increase the number of non-whites students - there are many more actions to be implemented to reduce the discrimination and inequality that are present in every region of this country. Bertrand and Mullainathan (2003) argue that training alone may not be enough to alleviate the barriers raised by discrimination, since blacks living in the U.S. with the same qualification as whites, have a lesser probability of receiving callbacks for interviews, after responding to help-wanted ads. We hope that the insights on the subject provided by this paper may stir up the debate about discrimination, and that in a near future, non-whites may have the same access as whites to education, job interviews and receive the same return to education and experience.

References

[1] ÁLVAREZ, L. G (2013) *Women in the Workforce: A Decomposition of Female Racial Wage Gaps in Brazil*, Master Thesis, Lund University - School of Economics and Management.

[2] ANGRIST,J.,CHERNOZHUKOV V., FERNÁNDEZ-VAL,I.(2006) Quantile Regression under Misspecification. *Econometrica*, Vol. 74, No. 2, 539–563

[3] BERTRAND, Marianne, MULLAINATHAN,Sendhil (2003) Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination *NBER Working Paper* No. 9873

[4] BLINDER, A.S. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *The Journal of Human Resources* 8: 436-455.

[5] BARSKY, R. BOUND,J. CHARLES, K. and LUPTON,J. (2002), Accounting for the Black-White Wealth Gap: A Nonparametric Approach, *Journal of the American Statistical Association*, 97(459), 663-673.

[6] BARTALOTTI,O. A. C. (2007) *Discriminação Salarial por Cor e Gênero Revisitada: Uma abordagem de decomposição contrafactual utilizando regressões quantílicas*. Master Thesis, FGV: São Paulo.

[7] BROWN, R. S. MOON,M. and ZOLOTH,B. S. (1980). Incorporating occupational attainment in studies of male-female earnings differentials. *The Journal of Human Resources* 15 (1), 3-28.

[8] CHERNOZHUKOV V., FERNÁNDEZ-VAL I., MELLY B. (2012) 'Inference on counterfactual distributions" The Institute for Fiscal Studies Department of Economics, UCL cemmap working paper CWP05/12

[9] FIRPO S., FORTIN, N. LEMIEUX T., (2009) Unconditional Quantile Regressions *Econometrica*, Vol. 77, No. 3, 953–973.

[10] FIRPO S., FORTIN, N. LEMIEUX T., (2007) Decomposing Wage Distributions using Recentered Influence Function Regressions, Unpublished Manuscript, University of British Columbia.

[11] FORTIN, N. LEMIEUX T., FIRPO S. (2010) "Decomposition Methods in Economics" NBER Working Paper Series.

[12] GARDEAZABAL,J. UGIDOS, A. (2004). More on Identification in Detailed Wage Decompositions. *Review of Economics and Statistics*, 86(4), 1034–1036.

[13] GOLDIN, C., ROUSE,C. (1997) Orchestrating Impartiality: The impact of "Blind" Audintions on Female Musicians *NBER Working Paper* N^{o}5903

[14] MACHADO, J., MATA J., (2005) Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression *Journal of Applied Econometrics* 20: 445– 465.

[15] MELLY, B (2006) "Estimation of counterfactual distributions using quantile regression" Swiss Institute for International Economic s and Applied Economic Research (SIAW), University of St. Gallen

[16] OAXACA, R. (1973), Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review* 14: 693-709.

[17] OAXACA,R.L., RANSOM,M.R. (1999). Identification in Detailed Wage Decompositions. *Review of Economics and Statistics*, 81(1), 154–157.

[18] SALARDI, P. (2012), *An Analysis of Pay and Occupational Differences by Gender and Race in Brazil - 1987 to 2006*, Phd Thesis, University of Sussex.

[19] WOOLDRIDGE, Jeffrey M (2009). Introductory econometrics: a modern approach. Cengage Learning.

Appendix

Figure 4: Decomposition of the regional wage gap: Males and Females

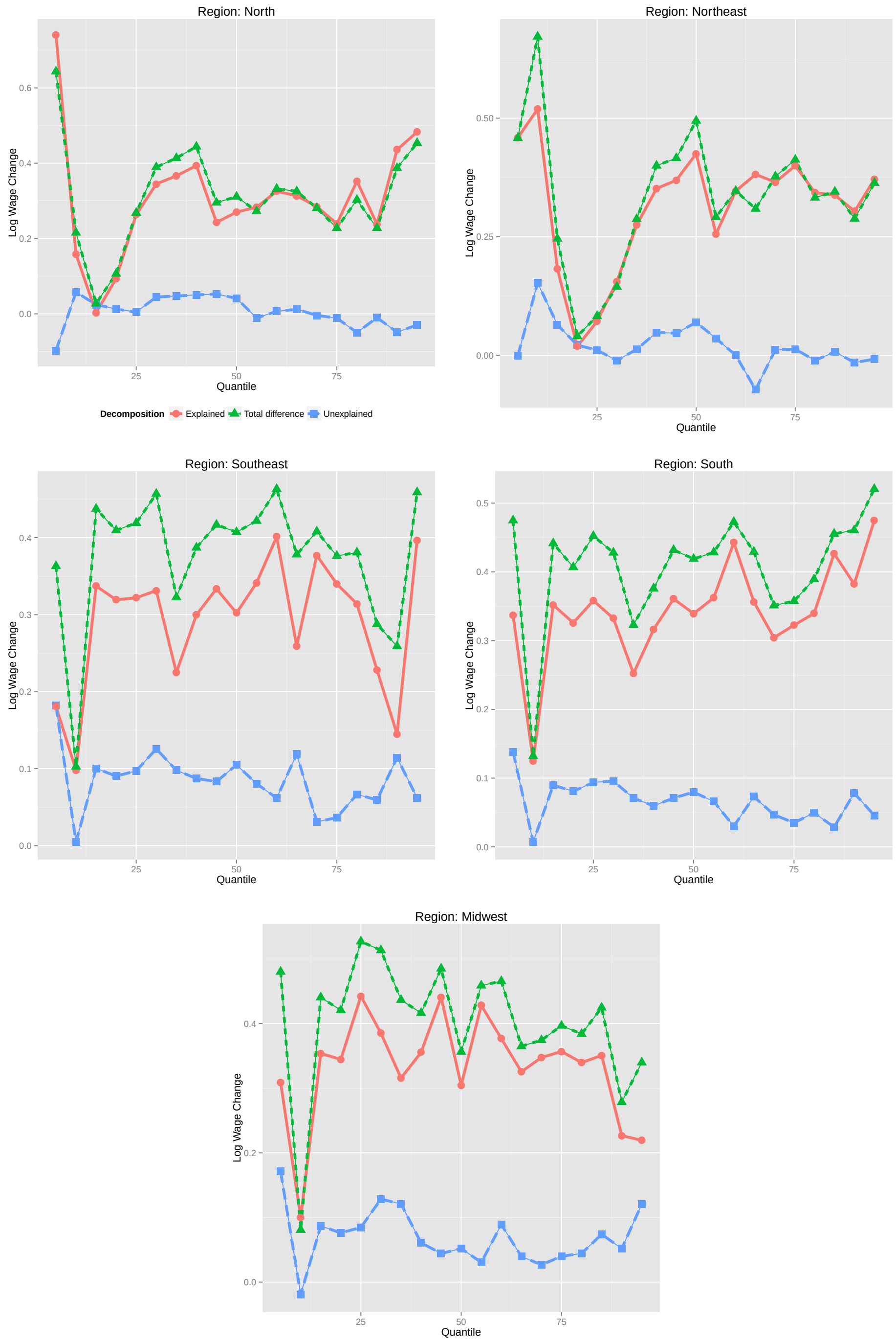


Figure 5: Decomposition of the regional wage gap: Whites and Non-whites