Occupational segregation and wage disparities in Brazil – 1987-2006: An empirical assessment using quantile decomposition techniques

Paola Salardi*
University of Sussex

VERY PRELIMINARY DRAFT – DO NOT CITE WITHOUT PERMISSION

Abstract: In this paper, we attempt to provide a comprehensive picture, over the last two decades, on gender and racial wage gaps across the entire wage distribution and on the potential impact of gender and racial occupational segregation on wage determination, in the context of the Brazilian labour market. Drawing on a novel dataset, constructed by harmonizing occupational classification over twenty years, our analysis particularly focuses on the evolution of the role played by the female and non white occupational intensity on wage disparities. We employ mean and quantile regression analysis in order to investigate the role of female and non white occupational intensity at most selected points of the conditional wage distribution. Two different decomposition techniques proposed by Melly (2006) and by Firpo, Fortin and Lemieux (2009) are applied to investigate the determinants of differences in distribution using quantile regression. Finally, the proposed analysis covers the Brazilian labour market as a whole during the last two decades but it also disaggregates the analysis across formal and non-formal sectors.

* PhD student at the Department of Economics, University of Sussex, U.K. Email address: p.salardi@sussex.ac.uk.
This paper is **a very preliminary (and incomplete) draft** and it draws from my Phd thesis. It has been prepared for the 15th IZA European Summer School in Labor Economics to be hold in Buch/Ammersee, Germany, the 23th-29th of April 2012. Any comments and feedbacks are very welcome. All errors remain my own.

1. Introduction

There have been a huge range of studies of wage inequality and wage differentials over the last three decades. The vast majority of these studies focus on investigating wage disparities by employing the well-know Oaxaca (1973) and Blinder (1973) wage decomposition technique (OB decomposition, hereafter). It is a simple and very powerful tool that allows for disentangling the contribution of characteristics (the explained component) to the contribution of the returns to those characteristics (the unexplained component or wage structure effect).

However, it also has several limitations that have been documented in the literature. One important drawback is that it focuses only on average effects on wage gaps, and the estimation of only average treatment effects might overlook some key dynamics. Hence the restricted focus on averages may lead to a misleading or incomplete picture of the pattern of wage differentials across the entire wage distribution. A second limitation with most existing studies is that they do not draw clear connections between occupation segregation and wage discrimination, despite the fact that the two are likely to be closely interconnected.

With these two limitations in mind, the main goals of this paper are twofold. First, we estimate the evolution of gender and racial wage gaps in Brazil over the last two decades at different quantiles of the wage distribution. Second, while tracing the pattern of wage differentials, we focus particularly on the role played by the *female* and *non white occupational intensity* on gender and racial wage differentials, respectively. The latter analysis is made possible by our construction of a novel dataset that harmonized occupational codes from Brazilian household surveys over twenty years, and thus facilitates consistent analysis across occupations.

In order to achieve these two goals we apply two novel decomposition techniques developed by Melly (2006) and Firpo, Fortin and Lemieux (2009), which make it possible to decompose wage differentials into the effects of characteristics and the effects of coefficients at different quantiles of the wage distribution.

Along with the two main aims of this paper, the empirical analysis also takes account of several additional aspects. First, we look at both gender and racial wage differentials, and discuss similarities and differences between them. Second, we not only analyse the entire labour market but we also disaggregate the analysis between the formal and non formal sectors. Finally we adopt a temporal perspective to our analysis, as the period of interest covers two decades (from 1987 to 2006).

The structure of the paper is as follows. The next section presents a brief literature review, situating the contribution of this paper within the broader literature. Section 3 presents the data and provides an overview of gender and racial wage differentials over the wage distribution. Section 4 illustrates the empirical methodology by describing the identification strategy and the two quantile decomposition techniques adopted in this paper. Section 5 presents our findings and section 6 finally concludes.

2. Literature review

After the publication of seminal studies by Oaxaca (1973) and Blinder (1973), research on pay gaps in many developed and developing countries, both by gender and by race or ethnicities, has been prolific. Moreover, a large number of studies not only have applied this powerful methodology, but have also improved it in several respects. Several papers have sought to tackle the 'index number' problem (Cotton, 1988; Neumark, 1988; Oaxaca and Ransom, 1994) and to compute the variances for the decomposition components (Oaxaca and Ransom, 1998). Other papers have dealt with selection bias correction within the decomposition frameworks. This began with Dolton, Makepeace, and Van Der Klaauw (1989) and Neuman and Oaxaca (2004), while the most recent paper by Bourguignon, Fournier and Gurgand (2007) addresses the selection bias problem using a multinomial logit model.

Another important set of studies upgrade the OB decomposition technique by accounting for occupational structure. The seminal work by Brown et al (1980) introduced a modified version of the OB decomposition where the occupational attachment model is estimating using a multinomial logit, while Miller (1987) proposes estimation by ordered probit model. Reilly (1991) introduced the selection bias correction together with the occupational attachment model in order to estimate the occupational wage equations. In this set of studies the contribution of occupational segregation to wage gaps is thus estimated separately (see also Gill, 1994; Neuman and Silber, 1996; Appleton et al, 1999). Other works have aimed at estimating gender wage gaps by taking into account the 'degree of feminization' (see, among others, Macpherson and Hirsch, 1995; Baker and Fortin, 2003; Cotter, Hermsen and Vanneman, 2003), while some others explore inter-industry wage differentials (see, among others, Krueger and Summers, 1988; Fields and Wolff, 1995; Haisken De New & Schmidt, 1997; Horrance and Oaxaca, 2001). As the OB decomposition approach suffers from the absence of a direct estimation of individual productivity, an interesting approach that has been proposed is the analysis of wage differentials by employing employer-employee matching data (see, for example, Hellerstein, Neumark and Troske, 2002; Bayard, Hellerstein, Neumark and Troske, 2003, Hellerstein and Neumark, 2006; Hellerstein and Neumark, 2007).

Although each of these groups of studies tackles different limitations of the original OB decomposition method, they all rely on the estimation of wage gaps at the mean values. Going beyond the mean, namely to focus on more general counterfactual wage distributions, has been the subject of several studies in recent years (see Fortin, Lemieux and Firpo, 2011). The methodologies that go beyond the use of mean values include the residual imputation approach (Juhn, Murphy and Pierce, 1993), the weighted-kernel estimation (Di Nardo, Fortin and Lemieux 1996), the rank regression method (Fortin and Lemieux, 1998), methods based on estimating hazard functions (Donald, Green and Paarsch (2000) or methods based on parametric quantile estimation (such as Gosling, Maching and Meghir 2000 or Machado and Mata 2005). Melly (2005) has proposed a conditional quantile decomposition approach very similar to that of Machado and Mata (2005), while the more recent paper Chernozhukov, Fernandez-Val and Melly (2009) covers the modeling and estimation of a wide range of counterfactual conditional distributions. Finally, Firpo, Fortin and Lemieux (2009) have proposed a decomposition technique based on the recentered influence function of the statistics of interest, the RIF-regression approach.

In this paper we apply two of these techniques that go beyond the estimation based on mean values: the conditional quantile regression approach proposed by Melly (2006) and the RIF-regression method proposed by Firpo, Fortin and Lemieux (2009). We believe that employing these techniques in the context of the Brazilian labour market can provide a deeper understanding of specific dynamics related to wage differentials.

In analyzing gender and racial wage gaps for Brazil, this study builds on a large number of existing studies (see a review in Salardi, 2012). Some studies have accounted for occupational segregation while estimating wage differentials, following Brown et al (1980)'s reformulation of the OB decomposition (Ometto et al, 1999; Arcand and D'Hombres, 2004 or, more recently, Salardi, 2012). Several other studies have controlled for the selection bias problem, including Stecler et al (1992), Loureiro et al (2004), and Carvalho et al (2006), with a recent study using quantile regression with semi-parametric correction for sample selection á la Newey (1991) and Buckinsky (1998) (see Coelho, Veszteg and Soares, 2010). However very few studies have investigated wage gaps using quantile regression estimations. Duarte, Ferreira and Salvato (2003) study earning differentials between the Southeast and Northeast of Brazil using a semi-parametric approach based on DiNardo, Fortin and Lemieux (1996). Guimares, Cavalcanti and Neto (2006) apply the methodology developed by Machado and Mata (2005), which is based on a quantile regression technique, to similarly investigate differences between the Southeast and Northeast. Santos and Ribeiro (2006) also explore gender wage gaps using the Machado and Mata (2005) decomposition technique, but looking only at the labour market in a single year, 1999.

There is also an interesting branch of studies that link the issue of informality to the investigation of the Brazilian labour market discrimination. Among others, Birdsall and Behrman (1992), Tiefenhaler (1992) and Silva and Kassouf (2000) have estimated wage gaps by formal and non-formal labour markets. Carneiro and Henley (2001) have explored wage differentials between formal and informal sector additionally controlling for selection bias. More recent advances are provided by Cacciamali and Hirata (2005) and Cacciamali, Tatei and Rosalino (2009).

Against this background, to the best of the author's knowledge this paper makes several original contributions to the existing literature on Brazilian labour market wage discrimination. First, it explores the evolution over time of *both* gender *and* racial wage gaps across the entire wage distribution. Second, it looks at the evolution of gender and racial wage gaps over time. Third, it links the analysis of wage discrimination with issues related to occupational segregation by estimating the role of female and non white occupational intensity on wage differentials. Finally, it contributes to the analysis of informality within the Brazilian labour market by disaggregating the analysis between formal and non-formal sectors, and expanding upon the limited number of studies exploring differences between the two sectors.

3. Data and overview of wage gaps

For the analysis, we employ data at the micro-level from the national household survey for Brazil, the *Pesquisa Nacional por Amostra de Domicilio* (PNAD), covering the period from 1987 to 2006. The PNAD is collected by the national statistical office, the *Instituto de Geografia e Estatistica* (IBGE). It is one of the most comprehensive sources of socio-economic information on Brazilian households. The sample consists of workers aged between 15 and 65 years old who declare that they are working and for whom there are no missing observations for wages and occupational code. Unlike many studies of Brazil we consider the whole labour market by including civil servants, domestic workers and individuals involved in agricultural activities across all five regions of Brazil, in both urban and rural areas.

The primary advantage of this dataset is the availability of information on earnings and occupations over a prolonged period of time (two decades). The information related to earnings is consistently provided by the dataset and we compute the log of hourly earnings from the primary occupation. Dealing with occupational codes is more complex, as the PNAD employs a classification of occupations that varies across years and which, for the majority of years, is not directly comparable with the international classification provided by the ILO, the ISCO-08. We solve this consistency problem by employing a new harmonized occupational classification developed by Salardi (2012) and employed to study the evolution of occupational segregation in Brazil in Salardi (2011). This classification is harmonized and consistent over the two decades of interest (from 1987 to 2006) and consists of 80 different occupational categories at 3-digit level of specification (more details about the classification are provided in Salardi 2011, 2012).

Having harmonized and consistent classifications over time allows us to construct two variables of interest: female occupational intensity (FOCC) and the non white occupational intensity (NWOCC). They consist of the proportion of female (or non white) workers in each occupation. We compute these values at a 2-digit level of occupational classification, which includes 23 different occupational codes. In other words, these two variables tell us the degree of *femaleness* or *non whiteness* of each occupation.

On the other hand, the use of this dataset over such a prolonged period of time restricts the set of other information that is available for all years. Many variables that are commonly employed in the specification of wage equations, such as work experience, are not present in the earlier years of this dataset. For this reason, we employ an austere wage equation specification, which has nonetheless proven to have high explanatory power while benefitting from the temporal perspective granted by the use of two decades of data.

The dataset has a large sample size that varies from a labour force of roughly 98,000 observations in the first year (1987) up to roughly 150,000 in the last year (2006). The entire Brazilian labour force can then be further grouped in three main sectors: formal, informal and self-employed. The distinction between formal and non-formal labour markets is based on possession of a signed working card, the *carteira de trabalho*. We, also, choose to distinguish among non-formal activities between employees with no signed labour card and those who are self-employed, as these two categories show different trends. A comprehensive discussion of the similarities and differences between the three main labour market sectors can be found in Salardi (2011). Finally, in distinguishing workers by race we choose to aggregate the labour force into white and non-white categories, where the group of non-white workers considers both brown and black individuals. This avoids problematic definitional challenges, while nonetheless capturing an empirically important distinction (see more in Salardi, 2012).

Figure 1 reports the kernel density functions by gender and by race for the first and last years of period of analysis. Gender and racial disparities are clearly visible, as wage distributions for male and white workers are shifted to the right. On average, male and white workers earn more than female and non white workers, respectively. If we look at how the distributions have changed over time we notice an interesting pattern: the female and male wage distributions show significant convergence by 2006, while the wage distributions of white and non white workers in 2006 are similar to those in 1987. That is, while gender disparities appear to have consistently declined over time, racial wage differentials seem to have remained substantially unchanged.

[Figure 1 about here]

Disaggregating the sample into formal and non-formal sectors reveals similar trends across the different labour markets. There is an extended discussion of the evolution of wage gaps over time by gender and race, and by formal and non formal sectors, in Salardi (2012), while here the focus is on adding nuance to this story by focusing our attention to how wage gaps are distributed across the wage distribution. To this end, Figure 2 shows wage gaps, by both gender and race, over different quantiles of the wage distribution. We can clearly see that wage differentials by gender are considerably greater at the bottom of the wage distribution. By contrast, racial wage differentials increase as we move toward the top of the wage distribution. Over time we notice a considerable decline in gender wage gaps across the wage distribution, with the average value moving from 0.322 in 1987 to 0.05 in 2006; in the case of racial wage gaps the patterns remain fairly similar over time, with the average value moving from 0.489 in 1987 to 0.413 in 2006.

[Figure 2 about here]

If we disaggregate the analysis of wage gaps across quantiles into the formal and non-formal sectors, we see that gender wage differentials are severe at the bottom of wage distribution only in the case of non-formal sectors. By contrast, within the formal sector the gender gap seems to increase as we move toward the top of the wage distribution, with particularly large wage gaps at the very top of the distribution. Interestingly, and particularly for the informal sector, we record negative wage gaps in the upper half of the wage distribution, before observing large wage gaps at the very top of the distribution (see Figure 3a). Thus, the U-shape that we notice when looking at gender wage gaps over quantiles for the entire labour market disguises different patterns in the formal and non formal sectors: greater gender gaps within low-paid occupations occur primarily in non-formal sectors, while greater gender gaps within top occupations is a more prominent feature of formal sector activities.

Turning to racial wage gaps, we do not see large differences in patterns across sectors, as in all sectors racial wage gaps tend to increase as we move toward the top of the wage distribution (see Figure3b).

[Figure 3a and 3b about here]

Finally, Figure 4 presents a more general picture of both gender and racial wage gaps, presenting data at selected points within the wage distribution (specifically 0.1, 0.25, 0.5, 0.75 and 0.9), and for five years across the entire period (1987, 1992, 1997, 2002 and 2006). There are two main messages that emerge from these plots. First, gender wage gaps are more severe at the bottom, while racial wage gaps tend to increase as we move to the top of the wage distribution. Second, over time, both gender and racial differentials have consistently decreased, however the decrease is more pronounced for gender wage gaps.

[Figure 4 about here]

Given that the following decomposition analysis involves a series of mean and quantile regressions, exploring the relationship between a variety of covariates at different points in the wage distribution, it is useful to look briefly at the mean of the primary covariates for a few selected wage quantiles. In order to save space, we do not present tables of the means and standard deviations for all covariates across quantiles and years, but simply summarize the most important findings. While female and male workers are distributed relatively homogeneously across quantiles (especially in more recent years), there is a clear racial pattern, as the presence of non white workers declines within higher wage quantiles. Age and years of education increase as we move to higher quantiles, consistent with a positive relationship between earnings and human capital endowments. There are less workers living in urban areas at lower wage quantiles, confirming that rural workers have, on average, lower wages. Along the same lines, people working in the agricultural sector are more numerous at the bottom of the wage distribution, together with those working in the personal and restaurant services sector. Looking at the concentration of different occupations within different quantiles confirms that higher skilled jobs are, indeed, better paid.. When we look at the distribution of informality over wage quantiles, we discover an interesting story. Even if the presence in the formal sector is equal to roughly 45-46% over time, only 0.05% in 1987 and 0.008% in 2006 of formal workers are in the 10% bottom of the wage distribution.

Since the relationship between wage differentials and female and non white occupational intensity is of special interest, we describe patterns in occupations intensity in somewhat more detail. Our variable for female occupational intensity moves from an average equal to 37% in 1987 to 44% in 2006 and it is homogenously distributed over wage quantiles (although it is slightly higher at the bottom of the wage distribution in the earlier years). By contrast, non white occupational intensity moves from 47% in 1987 to 53% in 2006 but it consistently decreases as we move to the top quantiles in any given year. It broadly means that female-dominated occupations are located at different level of earnings while non white-dominated occupations are in fact characterized by low earnings.

Figure 5 further highlights how female and non white occupational intensity vary across wage quantiles. In both 1987 and 2006 there is no clear pattern for female occupational intensity, while non white occupational intensity steadily decreases as we move to the top of the wage distribution.

[Figure 5 about here]

Figure 6 plots average wages by gender or race at different levels of female or non white occupational intensity. Looking first at gender, we see little trend in the relationship between the two variables, as female dominated occupation are neither better not worse paid than male dominated professions, although on average male workers earn more than women independent of the degree of femaleness within occupations. The pattern by race is very different, as wages consistently decline as the non-white occupational intensity increases, while, as with the case of gender, white workers consistently earn higher wages within occupations, independent of the degree of non-whiteness.

[Figure 6 about here]

Finally, how can we summarise the linkages between gender and racial wage differentials, wage quantiles and female or non white occupational intensity? Gender differentials are more pronounced at the extreme of the wage distribution and particularly more severe within low-paid occupations, while racial wage gaps widen as we move to the top of the wage distribution. Women seem to be homogenously distributed across occupations, while non white individuals are mainly segregated in low-paid and low-skilled occupations. Although employed at any occupational levels, women seem to suffer by more severe wage gaps within low paid non-formal occupations and within the very top paid formal jobs, even if they are not an insignificant presence within the latter occupations. Non white workers tend to work in low-paid and low-skilled occupations and their wage difference with respect to white workers considerably widens within those occupations that are characterised by less presence of non whites and higher earnings. These figures seem to create space for the hypothesis of *sticky floors* within non-formal jobs and *glass ceilings* within formal jobs for female workers and *glass ceilings* for non white workers.

3. Empirical methodology

In this section we describe the quantile decomposition techniques that we apply in order to gain a more nuanced understanding of the descriptive statistics presented so far. First we discuss the identification strategy and the definition of the parameters of interest. Then we explain the conditional quantile decomposition technique developed by Melly (2006) and the RIF-regression method proposed by Firpo, Fortin and Lemieux (2009).

3.1 Identification strategy

Our investigation is ultimately interested in answering a counterfactual question: 'How much would female workers be paid if they were paid according to the wage structure for male workers?' Alternatively, from the racial perspective, 'How much would non white workers be paid if they were paid like white workers?'. We are thus seeking to compare observed wage structures with counterfactuals, which capture potential alternative wage structures. As such, our problem of the wage structure effect can be interpreted as a treatment effect and ultimately it can be linked to the program evaluation literature as extensively explained in Fortin, Lemieux and Firpo (2011).

In other words, we are interested in the effect that a binary variable, which is our treatment (gender or race in our case), exerts on a specific outcome, earnings. Using the notation adopted by Fortin, Lemieux and Firpo (2011), this binary treatment identifies two distinct groups, group A and group B, which, for our purposes, are female versus male workers and non white versus white workers. We can thus think of the effect of gender or race for each individual worker, $Y_{Bi} - Y_{Ai}$, as the individual treatment effect We can, in turn, interpret the difference between the average earnings of group B and the average earnings of group A, as the average treatment effect (ATE) from the impact evaluation literature, as follows:

$$ATE = E[Y_B] - E[Y_A] \qquad [1]$$

We know that moving from group A to group B is conceived to be "the treatment". Hence the observed average wages for group B and A are defined as $E[Y_B|D_B = 1]$ and $E[Y_A|D_A = 1]$ respectively.

The introduction of the counterfactual enables to disentangle the average treatment effects of the treated (ATT). In fact, by adding and subtracting the counterfactual, we obtain:

$$E[Y_B] - E[Y_A] = \{E[Y_B|D_B = 1] - E[Y_A|D_B = 1]\} + \{E[Y_A|D_B = 1] - E[Y_A|D_A = 1]\} \quad [2]$$

The first bracketed component on the right side identifies the ATT, namely the difference between the observed average wages of group B and the hypothetical wages that workers belonging to group B would have been paid if they belonged to group A. That is:

$$ATT = E[Y_B|D_B = 1] - E[Y_A|D_B = 1] \quad [3]$$

From equation [2] it appears now clear the link between the program evaluation literature and wage decomposition methodologies. In fact wage decomposition methodologies aim at investigating the extent to which wage differentials originate from differences in structure and differences in observed characteristics. The first bracketed term of equation 2 represents the differences in the returns of the observables, or differences in coefficients (or wage structure component), while the second bracketed term represents the differences in the observable characteristics. As pointed out by Fortin, Lemieux and Firpo (2011), the only difference between the two approaches lies in which component gets more attention: for the wage decomposition techniques, the differences in the observables is the key component, while for the program evaluation literature the wage structure effect or treatment effect of the treated (ATT) is central.

The choice of the reference group is arbitrary and it clearly depends on the structure of the researcher's problem. If we change the reference group in the above notation, we get a different counterfactual and equation [2] becomes:

$$E[Y_B] - E[Y_A] = \{E[Y_B|D_B = 1] - E[Y_B|D_A = 1]\} + \{E[Y_B|D_A = 1] - E[Y_A|D_A = 1]\} \quad [4]$$

Now, the second bracketed term identifies the treatment effect of the non-treated (ATNT), or, more intuitively, the difference between the hypothetical wages that workers belonging to group A would be paid if they were in group B, and the observed wages of workers belonging to group A. That is:

$$ATNT = E[Y_B|D_A = 1] - E[Y_A|D_A = 1] \quad [5]$$

Having thus presented the notation employed by Fortin, Lemieux and Firpo (2011) we conclude this section by re-formulating the notation to correspond with the research questions being investigated in this study. With respect to gender disparities, we have defined our research questions as: "what if female workers were paid according to male wage structure". Framed in this way, gender becomes our binary treatment, and, as such, group A now represents female workers, while group B now represents male workers. If now the generic outcome Y is defined as wage W, the average gender wage gap is defined as:

$$\Delta_{FM} = E[W_M] - E[W_F] \quad [6]$$

The counterfactual of interest is then $E[Y_M|D_F = 1]$ and we will focus on the differences in wage structure, defined as the treatment effect of the non-treated (ATNT):

$$ATNT = E[W_M|D_F = 1] - E[W_F|D_F = 1] \quad [7]$$

Similarly, in terms of racial disparities, we ask "what if non white workers were paid according to the wage structure of white workers". As such, the binary treatment is now race (identified as white/non-white) and the average racial wage gap is given by:

$$\Delta_{NWW}= E[W_W] - E[W_{NW}] \qquad [8]$$

The counterfactual of interest is $E[W_W|D_{NW} = 1]$, which captured the expected earnings of non white workers if they were paid according to the same wage structure as white workers. Again, the wage structure effect is provided by the average effect of the non-treated (ATNT) as follows:

$$ATNT = E[W_W|D_{NW} = 1] - E[W_{NW}|D_{NW} = 1] \qquad [9]$$

Having defined the problem in this way, it is worth briefly highlighting two issues about the likely effects, which are raised by Gardeazabal and Ugidos (2005). First, although gender or racial wage gaps could be negative in theory, in practice they have almost always been found to be positive, as men are on average paid more than women as well as white workers more than non white ones. Second, "the gender wage gap is not an upper bound on discrimination. When women are more productive than men, but yet are paid less than men, discrimination is greater than the gender wage gap."

Finally it is important to conclude by stressing that when we decompose wage differentials, we compute the contribution of several factors to observed outcomes but we are not necessarily identifying causal effects. Fortin, Lemieux and Firpo (2011) argue that the assumptions under which the wage structure effect could be interpreted as a causal effect are ultimately very stringent, for two reasons. First, the binary treatment defining the two distinct groups cannot, generally, be considered a choice in the case of gender or race. Second, the covariates are generally affected by the treatment variable. As a consequence, we cannot state that we are estimating the causal effect of the treatment while controlling for a set of exogenous characteristics, as these characteristics are not pre-treatment variables. Nonetheless, the identification of the contribution of different factors to observed wage differentials may remain useful in developing specific hypotheses, mechanisms or explanations.

Having thus specified the identification strategy, there are a variety of empirical methodologies that can be applied in order to compute the counterfactual of interest. The next two sub-sections overview the two approaches employed in this paper: the conditional quantile regression methodology (Melly, 2006) and the RIF- regression method (Firpo, Fortin and Lemieux, 2009).

3.2. Estimation of counterfactual distributions using quantile regression

In order to estimate the average treatment effect using the quantile methodology, we need to estimate the counterfactual quantile, Q_θ^C, following two steps: first we need to estimate the conditional distribution by quantile regression and then we estimate the unconditional distribution by integrating the conditional distribution over a range of covariates (Melly, 2006).

In the first step, the impacts of the characteristics on the conditional wage distribution can be estimated using a quantile regression framework (see Koenker and Bassett 1978; Koenker and Hallock 2001; Koenker 2005). This estimation procedure is formulated in terms of absolute rather than squared errors. The estimator is known as the Least Absolute Deviations (LAD) estimator.

In contrast to the OLS approach, the quantile regression procedure is less sensitive to outliers and provides a more robust estimator in the face of departures from normality (see Koenker (2005) and Koenker and Bassett (1978)). Quantile regression models may also have better properties than OLS in the presence of heteroscedasticity (see Deaton (1997)).

The conditional quantile function $Q_\theta(W|X)$ can be expressed using a linear specification as follows:

$$Q_\theta(W|X) = X_i{}'\beta_\theta \qquad \text{for each } \theta \in (0,1) \qquad [10]$$

Where W is the dependent variable and denotes wages, X_i represents the set of covariates for each individual i and β_θ are the different coefficient vectors that need to be estimated. These quantile regression coefficients can be interpreted as the returns to different characteristics at given quantiles of the wage distribution. It is important to note that we assume that all quantiles of W conditional on X are linear in X. We can then estimate the conditional quantile of W by linear quantile regression in each specific percentile of $\theta \in (0,1)$.

The conditional quantile function for group B would be:

$$Q_{B,\theta}(W_B|X_B) = X_{B,i}{}'\beta_{B,\theta} \qquad [11]$$

While for group A:

$$Q_{A,\theta}(W_A|X_A) = X_{A,i}{}'\beta_{A,\theta} \qquad [12]$$

The second step is necessary because the unconditional quantile is not the same as the integral of the conditional quantiles. In other words, the law of iterated expectations does not apply in the case of quantiles, so $Q_\theta(W) \neq E_X[Q_\theta(W|X)]$ where $Q_\theta(W)$ is the θth quantile of the unconditional distribution of wages and $Q_\theta(W|X)$ is the corresponding conditional quantile. To simplify by providing an example, if we focus on the quantile equal to 0.5, i.e. the median, we can simply say that the expectation of the conditional median does not produce the median of the marginal distribution.

As a consequence we need to know the entire conditional distribution of W given X in order to estimate the unconditional distribution.

Note that:

$$\theta = F_W(Q_\theta) = E[F_{W|X}\big(Q_\theta(W|X)\big) = \int F_{W|X}\big(Q_\theta(W|X)\big)dF_X(X) \qquad [13]$$

$F_W(Q_\theta)$ represents the conditional cumulative distribution of wages and the inverse of the distribution function $F_W^{-1}(\theta)$ is ultimately the quantile function.

Hence, by inverting the conditional quantile function we obtain the conditional distribution function. Then we obtain the unconditional distribution function by integrating the conditional distribution function over a range of covariates. Finally, by inverting the unconditional distribution function we obtain the unconditional quantiles of interest.

In our case, in order to obtain the key counterfactual quantile of interest, we need to invert the counterfactual distribution of interest, $Q_{B,\theta}^{C} = F_{W_B^C}^{-1}(\theta)$, which uses the distribution of the characteristics of group A with the wage structure of group B as follows:

$$F_{W_{B,\theta}^{C}}(W) = \int F_{W_{B,\theta}|X_B}(W|X)dF_{X_A}(X) \qquad [14]$$

Once the key counterfactual is estimated, we can perform the decomposition of wage gaps of the unconditional quantile function between groups B and A, denoted as:

$$\Delta_{\theta} = [Q_{B,\theta} - Q_{B,\theta}^{C}] + [Q_{B,\theta}^{C} - Q_{A,\theta}] \qquad [15]$$

The first bracketed term represents the effect of characteristics (or the quantile endowment effects) and the second represents the effect of coefficients (or the quantile treatment effects).

The conditional quantile regression methodology proposed by Melly (2006) is very similar to the decomposition technique proposed by Machado and Mata (2005). The Machado and Mata (2005) technique estimates components of the aggregate decomposition using simulation methods. The drawback is that it is computationally demanding. Melly (2006) demonstrates that if the number of simulations used in the Machado and Mata (2005) procedure goes to infinity, the decomposition technique by Melly (2006) is numerically identical. As a consequence, if one wants to use a large number of quantile regressions (e.g., 99, one for each percentile from 1 to 99), the Melly (2006) decomposition can be a more efficient option.[1] It is important to highlight that the Melly (2006) method assumes exogeneity for all covariates. Alternatively, one should explore instrumental variables or sample selection procedures. Finally, this type of conditional quantile decomposition technique does not allow for computing detailed decompositions. Methods based on conditional distributions that contemplate this option are further explored in Chernozhukov, Fernandez-Val and Melly (2009). An alternative method that estimates the effect of each covariate on the unconditional quantile has been recently proposed by Firpo, Lemieux and Fortin (2009), and it is the subject of the next sub-section.

3.3. Estimation of counterfactual distributions using RIF-regression

A new procedure proposed by Firpo, Fortin and Lemieux (2009) aims to estimate the impact of changing in distribution of covariates, X, on quantiles of the unconditional distribution of an outcome variable. It consists of running a simple regression where the outcome variable is replaced with a transformation of it,

[1] The stata command rqdeco, which is provided by Melly (2007), was used to compute the quantile endowment and treatment effects.

the (recentered) influence function (RIF). Although it can be applied to any distributional statistics of interest for which it is possible to compute an influence function, here we focus on the changes in the quantiles, denoted Q_θ, of the marginal unconditional distribution F_W.

As the statistics of interest in our case are quantiles, Q_θ, the influence function, $IF(W, Q_\theta)$, is defined as follows:

$$IF(W, Q_\theta) = (\theta - \mathbb{I}\{W < Q_\theta\})/f_W(Q_\theta) \qquad [16]$$

Where $\mathbb{I}\{\cdot\}$ is an indicator function and f_W is the density function of the marginal distribution of W evaluated at Q_θ.

Given that the RIF function, $RIF(W, Q_\theta)$, is equal to $Q_\theta + IF(W, Q_\theta)$, we then have the following formula:

$$RIF(W, Q_\theta) = Q_\theta + \frac{\theta - \mathbb{I}\{W < Q_\theta\}}{f_W(Q_\theta)} \qquad [17]$$

Hence, the RIF function can be computed easily in an OLS framework once we have computed the dummy variable $\mathbb{I}\{W < Q_\theta\}$ (which specifies whether the value of W is greater or smaller than Q_θ), and estimated the sample quantile Q_θ, as well as the density function f_W evaluated at Q_θ (generally computed using kernel density).

Then a value of transformed outcome variable is available for each observation and it can be used to estimate a simple OLS regression on a vector of covariates.[2] In the case of quantiles, the expected value of the RIF-regression model is viewed as an *unconditional* quantile regression. The coefficients of the unconditional quantile regression are computed for each group - group A and B if we keep the same notation as in previous sections -, and employed to compute the equivalent of the OB decomposition for each quantile as follows:

$$\Delta_\theta = (\bar{X}_B - \bar{X}_A)\hat{\gamma}_{B,\theta} + \bar{X}_A(\hat{\gamma}_{B,\theta} - \hat{\gamma}_{A,\theta}) \qquad [18]$$

Where the first term on the right side represents the differences in characteristics and the second term represents the differences in returns, which is the wage structure effect.

The primary advantage of this technique is that it estimates each individual covariate's effect at different quantiles of the wage distribution. This is significant, as few available techniques for estimating counterfactuals allow for such a detailed decomposition. In general, decomposition techniques on distributional functions different from the mean can rarely be employed to get a detailed decomposition. Machado and Mata (2005) provide a detailed decomposition of the wage structure, while the individual contribution of the binary variables, among the entire effect of the characteristics, is possible in the reweighted procedure proposed by Di Nardo, Fortin and Lemieux (1996).

---

[2] Examples of Stata ado file to implement the RIF-OLS methodology are available on Fortin's website http://www.econ.ubc.ca/nfortin/.

The primary limitation of this methodology lies in the linear approximation of a non-linear distributional function. This decomposition procedure provides only a first-order approximation of the composition effects and this approximation is not precise and produces approximation error. This issue is tackled further in Heywood and Parent (2009). A second limitation is that, at least for now, this methodology is built to estimate unconditional quantile regressions in the presence of exogenous covariates and does not consider the possible presence of endogeneity (Firpo, Fortin and Lemieux, 2009).

It is important to highlight that in this section we explain how to compute RIF function within an OLS framework. However Firpo, Fortin and Lemieux (2009) provide two alternative ways to estimate the average marginal effect. The RIF-logit estimates the marginal effect from a logit model while the RIF-NP is based on a purely nonparametric estimator.

Finally, it is useful to conclude this section by returning to the intuition behind this methodology. The key to the Firpo, Lemieux and Fortin (2009) methodology lies in the fact that the decomposition of quantiles is achieved by inverting proportions back into quantiles. Knowing that the cumulative distribution function links (unconditional) quantiles to their proportion of observations below each given quantile, we can obtain quantiles by dividing proportions by the density. In other words, this methodology estimates proportions that are needed to be inverted back in quantiles. In this sense, Firpo, Lemieux and Fortin (2009) methodology is very similar to the methodology proposed by Chernozhukov, Fernandez-Val and Melly (2009) to decompose a general distributional function. The latter, after estimating a model for proportions, inverts them back *globally* into quantiles, while the Firpo, Lemieux and Fortin (2009) methodology performs the inversion only *locally* (Fortin, Lemieux and Firpo, 2011).

3.4 Selectivity issues

We have presented two different methods to estimate quantile counterfactuals, though always based on the assumption of exogenous covariates. In reality, there could be several cases in which exogeneity fails, and in which the results would then suffer from self-selection problems or more general endogeneity problems. Following, Lemieux and Firpo (2011), we can consider three different cases: 1) different self-selection process within group A and group B; 2) self-selection into group A and group B; and 3) general endogeneity of the covariates.

The first case is possible particularly when the criteria that distinguish group A and group B are gender or race, as in this study. It is especially easy to imagine that women and men may have different decision processes that bring them into the labour market, while the same is certainly potentially true of different race or skin colour groups as well. In this case the unconfoundness (or ignorability) assumption does not hold, and the decomposition terms are not identified correctly. Machado (2009) invokes three different self-selection cases - selection based on observables, selection based on unobservables and bounds - and analyses possible solutions for each case. The second case occurs when individuals can decide whether to belong to group A or B. A proposed solution is the adoption of a control function, though this seems less likely to be relevant in this case owing to the nature of the binary categories. Finally, the third case refers to general endogeneity, which incurs when covariates are correlated with the error term. A standard solution to this problem is provided by instrumental variable methods. The investigation of self-selection and endogeneity issues and how to correct our empirical analysis in order to keep valid the identification of the decomposition components is subject to further improvements of this study.

4. Empirical findings

In this section we present the results in three stages. First, we present a set of quantile regressions, estimated at different quantiles of the wage distribution, from the pooled samples for the first and the last years of the period of interest. In estimating pooled regressions we are assuming that women and men, and non white and white workers, receive the same returns to characteristics. As previously, we pay particular attention to the importance of the female and non white occupational intensity. Second, we divide the samples and estimate quantile regressions by gender and by race separately. Finally, third, with these estimations, we implement the two different quantile decomposition techniques in order to identify how much of the gender and racial wage gaps estimated at different quantiles of the wage distribution can be attributed to differences in characteristics and how much can be attributed to differences in returns to those characteristics (or wage structure).

4.1 The effect of female and non white occupational intensity

In performing the pooled quantile regression analysis we attempt various different specifications of the wage equation, moving from a more austere to a more complete specification. In the most austere specification, the log of hourly wages is regressed on age, age squared, years of education, gender and race. We then add dummies for living in urban areas, living in each of the five main regions of Brazil and for being a formal worker, after which we add the variables for female (or, alternatively, non white) occupational intensity. Finally, the effect of this last key variable is estimated while also controlling for a set of dummies covering 9 economic sectors, 23 occupations at 2-digit level and the 27 federal states of Brazil.

To conserve space we only report the quantile regressions for the specification which includes female or non white occupational intensity and dummies for occupations (i.e. occupation effects) for both years. These regressions are presented in tables A1 and A2 in the appendix. The estimated coefficients show the expected effects: the variable for years of education is positive and strongly statistically significant and its effect increases as we move to higher quantiles. The same happens to the variables age and age squared, suggesting non-linear behaviour of this variable. Being a formal worker has a positive impact on the level of earnings, but this effect fades as we move to higher quantiles. Finally, the impact of being an urban worker is positive, and greater at the bottom of the distribution, suggesting that low-paid workers earn more in urban areas.

Turning to the impact of female and non-white occupational intensity, Tables 1a and 1b report the estimated coefficients for these two variables across different specifications, and for both 1987 and 2006.

[Table 1a and 1b about here]

Female occupational intensity (FOCC) exerts a negative impact on wages and this negative impact tends to become greater in absolute terms as we move toward the top of the wage distribution (see specification 'a' in tables 1a and 1b). However this larger effect at the top is reversed when we include dummies for occupations (specification 'b') and disappears if we control for economic activities as well (specification 'c').[3] In the case of non white occupational intensity (NWOCC), we see that the presence of non white workers has a negative effect on earnings, and this effect tends to be greater at the top of the distribution independently of whether we control for occupations and/or economic activities (see specifications 'e'-'h' in tables 1a and 1b).

When we look at female and male workers separately, we find that working in female-dominated jobs decreases earnings for female workers, particularly at the extremes of the wage distribution, even after controlling for the effect of occupations. By contrast, it has a positive effect on male wages, especially at the bottom of the wage distribution, though only when controlling for occupations (compare specification 'a' to 'b' in panel B in tables 1a and 1b). Turning to differences by race, for both non white and white workers, being employed in non white-dominated occupations means earning less, and this negative effect increases as we move toward the top of the distribution, independent of controlling for occupations or other effects (compare specifications 'e'-'h' in panel B in tables 1a and 1b. In sum, being employed in female-dominated occupations has a positive impact on male earnings but reduces earnings for female workers, particularly in the highest paid and lowest paid jobs. Being employed in non white-dominated occupations has a negative impact on wages, though relatively more among white workers and within better paid occupations.

4.2 Empirical findings from the Melly (2006) quantile decomposition

Before turning to the results when employing the Melly (2006) decomposition technique, it is useful to begin with results from the standard OB decomposition technique. Table 2 reports these decomposition results employing three alternative specifications. The first specification includes age, age

[3] Additionally controlling for dummies of 27 state of Brazil does not change the pattern (see specification 'd' in tables 1a and 1b).

squared, years of education, a dummy for formal workers, a dummy for urban workers, and dummies for the 5 main regions of Brazil. In the second specification we add the variable for female occupational intensity and occupations' effects (dummies for 23 occupations effects at the 2-digit level). Finally, in the third specification we add economic activities effects (dummies for 9 economic activities) and dummies for states of Brazil.

[Table 2 about here]

As is described in detail in Salardi (2012), at aggregate level we see a sharp decrease in both gender and racial wage gaps over time. Gender wage gaps, although smaller in magnitude, have declined much faster and considerably more. Strikingly, gender differences are overwhelmingly attributable to differences in returns to characteristics (or the wage structure effect, also called the 'unexplained' components in the decomposition) while the effect of characteristics (or 'explained' components) is generally negative, signalling that female workers have better endowments, particularly in educational attainment. By contrast, racial differences are largely attributable to differences in characteristics, as white workers have significantly greater endowments than non whites, though the returns to characteristics also remains positive, implying that the wage gaps persists even after accounting for differences in endowments. Finally, it is interesting to note that the inclusion of the occupational intensity variables (when moving from the first to the second specification) leads to a large change in the decomposition components of gender wage gaps, consistent with the hypothesis that female occupational intensity and occupational distribution are important factors. The impact of including these variables is also noticeable, but much more modest, in the case of racial gaps.

The detailed decomposition, which captures the contribution of each individual covariate of the estimated wage equations, explains these patterns further. Beginning with gender wage gaps, education accounts for the largest part of the impact of characteristics (explained component) on gender wage differentials, with a consistently negative and significant sign (see in panel A of table 2). Turning to the returns to characteristics (unexplained component), the role played by female occupational intensity stands out, as the extent of female occupational intensity has a strongly positive effect on gender differentials. In 1987 it accounted, by a large margin, for the largest part of the unexplained components and, while it has declined significantly over time, it remained strongly positive in 2006.

Turning to racial wage gaps (panel B of table 2), education again plays a central role in the explained component, however this time its contribution is, reflecting better endowments for white workers. When we move to returns to characteristics, both non white occupational intensity and occupations' effects account for a large portion of the overall pattern. The negative effect of non white occupational intensity implies better returns for non white workers within non white dominated occupations, though the positive contribution of occupations conveys that whites are employed in more rewarding jobs.

Overall, then, although female workers have better endowments than male workers, and hence should be paid more than their male colleagues, male salaries are, in fact, higher, owing to a large, positive, unexplained difference in returns to male characteristics. Notably, being a male worker within a female-dominated occupation appears to be particularly rewarded. In the case of racial differentials, white workers are paid more in large part because they have better endowments, and particularly better education. In addition, they benefit from large unexplained wage benefits (greater returns to characteristics), driven in large part by occupational structure, as non white-dominated occupations are significantly less rewarding. Finally, it is important to note the large effect of age in both the gender and racial decomposition results, particularly in accounting for differences in returns to characteristics. The message appears to be that experience is rewarded comparatively more among men and white workers.

Having reviewed the decomposition results at mean values, we now examine the results of the quantile regression decomposition of wage gaps, following Melly (2006). In what follows, we report only the results of the quantile decomposition exercise, which draws on the coefficients from the conditional quantile regressions.

We implement this methodology for both gender and racial wage differentials, and disaggregated into formal and non-formal labour markets. In order to retain the temporal perspective we apply the methodology to the first year (1987) and the last year (2006) of the period of interest. In the most comprehensive version of this study we perform the analysis for five years during the two decades of interest, however here we only report results for the first and last years due to constraints of space.

In the upper panels of tables 3a-3b through tables 6a-6b we report the quantile decomposition results following Melly (2006), while the lower panels report the RIF-regression decomposition results, which are discussed in the next sub-section. In addition, figures 7 and 8 plot the decomposition results over 99 percentiles of the wage distribution. Here we summarise the main findings.

Looking first at Figure 7, which plots data for the entire labour market, we see that in 1987 gender wage gaps were greater at the bottom of the wage distribution, and this was primarily attributable to the effects of the coefficients (or returns to characteristics). During the same period we also see increasing wage gaps at the very top of the wage distribution, though these were mainly the result of better characteristics for men in high income positions. These differences between the top and bottom of the distribution are quite striking, though we see that over time these considerable differentials have diminished and, as we can see from the plot for 2006, this is primarily thanks to a decline in the large effects of the coefficients, although better female endowments have contributed as well (see panel A of figure 7).

Turning to racial wage gaps, we again see that they are driven largely, but not exclusively, by differences in characteristics, which are relatively better for white workers. When we disaggregate the analysis into quantiles we see that the impact of both the characteristic and coefficient effects tends to increase as we move to the top of the distribution, while this situation has not improved over time (see panel B of figure 7).

[Figure 7 about here]

When we disaggregated the analysis into the formal and informal sectors, quite distinct patterns emerge. We begin with gender wage gaps, with the results reported in Figure 8a. In the formal labour market, we find that the effects of the coefficients are greater as we move toward the top of the distribution. The result has been consistently higher wage gaps within higher wage quantiles. That said, this effect this effect was more pronounced at the beginning of the period, as by 2006 the positive effect of the coefficients in increasing wage differentials within higher quantiles was largely offset by better female endowments. In contrast to the formal market, the non-formal labour markets show the opposite pattern. In the informal sector, the effects of the coefficients substantially higher at the *bottom* of the wage distribution, while the pattern is similar in the self-employed sector.

In the case of racial wage gaps, the results are quite different, as disaggregating the analysis reveals that patterns within the formal, informal and self-employed sectors are all quite similar, as reported in Figure 8b. This suggests less acute differences in labour conditions across the three sectors for non white and white workers.

[Figures 8a and 8b about here]

To summarize, we find gender wage differentials are mainly driven by the unexplained components, or wage structure effects, and particularly at the extremes of the wage distribution. These unexplained components, or wage structure effects, may be reflective of entrenched discrimination in the labour market. More positively, over time this gender wage gap has declined considerably, thanks primarily to a decline in the unexplained components. Notably, this wage structure component acts differently between the formal and non-formal labour markets. While it is higher at higher quantiles in the formal market, in the non-formal sectors the effect of the coefficients (or wage structure effect) is considerably greater at the bottom of the wage distribution. If we interpret these effects a comprising some sort of discrimination effect, the results suggest that gender discrimination may be most prevalent in better paid jobs within the formal labour market, but in lower paid jobs within the non-formal labour markets. Framing these findings in relation to key existing concepts in this field, the results suggest that if there is a *sticky floor* phenomenon for women, it is mainly occurring in the non-formal sectors. Turning to the formal sector, the apparently negligible gender wage gaps that we find in recent years disguise the fact that there remains a significant unexplained difference in wages within the higher wage quantiles, indicative of a discrimination effect, though it is largely offset, by better female endowments. This appears to be consistent with the continued existence of a *glass ceilings* phenomenon within the highest ranks of the formal sector.

Applying these same concepts to racial wage differentials, we see clearly that these differentials widen at higher wage quantiles due to both greater differences in characteristics in favour of white workers and unexplained higher returns to these characteristics, while neither pattern has improved over time. The continued importance of differences in returns to characteristics is consistent with the hypothesis of the existence of *glass ceilings* for non white workers.

4.3 Empirical findings from RIF-OLS decomposition

In the lower panels of the same set of tables (tables 3a-3b to 6a-6b) we report the results from the RIF-OLS regression decomposition methodology developed by Fortin, Lemieux and Firpo (2011). As was explained in the methodological section, the main advantage of this technique is the possibility of performing detailed decompositions across quantiles. This allows us to estimate the contribution of each individual covariate in determining wage differentials at different wage quantiles, either as part of the composition component (i.e. the effect of characteristics) or the wage structure component (i.e. the effect of coefficients). The decomposition results produced by the RIF-OLS methodology broadly coincide with those from the Melly (2006) technique, while adding additional nuance, thus reinforcing our broad confidence in the results.

The tables present the individual contribution of four key covariates to both the characteristics and coefficient components: age, years of education, female (or non white) occupational intensity (FOCC and NWOCC), and occupations' effects. Looking across the results, it is again clear that both education and occupational intensity play a crucial role in determining wage differentials, though in very different ways.

For gender wage gaps, education has a strong and negative effect on wage differentials across all of the decomposition results, covering the entire labour market, and the formal and non formal sectors separately. Its negative effect increases, in absolute terms, as we move to the top of the wage distribution again highlighting that education is the most important source of better female endowments, while this effect is even greater at higher wage quantiles.

Moving to the individual contributors to the coefficients component, the age variable shows a considerable impact. Its effect is positive, and higher at the top of wage distribution, meaning that men's work experience is rewarded more than that of women, particularly among high-paid professions. The female occupational intensity variable (FOCC) also plays a key role. It is consistently positive, and follows a sort of U-shape pattern across wage quantiles, as it is greater at the extremes of wage distribution. As was the case when using the Melly (2006) decomposition, the impact of female occupational intensity on wage structure effects is greatest for top-jobs within the formal labour market, while it is greatest for low-paid jobs within the non-formal sectors. In other words, the returns to working in female-dominated occupations are consistently higher for male workers, and these wage disparities are particularly pronounced among top jobs within the formal sector and low-paid positions within the informal sector.

Again, even if there are characteristics that we are not able to account for, such as ability, there is highly possible that a portion of the unexplained differences in gender wage gaps (the wage structure effect) are due to gender discrimination. This seems particularly likely in light of the fact that men's experience is rewarded more than women's in top positions and that men working in female-dominated occupations receive higher wages, particularly in top formal jobs and low paid informal occupations. This again suggests that women are subject to the phenomenon of *sticky floors* in non-formal occupations, as well as of *glass ceilings* in formal activities.

Turning to racial wage gaps, education also plays a key role in determining racial wage gaps though, in this case the effect is positive, and, again, greater at higher wage quantiles. Looking at the effects of the coefficients, the age variable makes a considerable positive contribution to wage differentials, especially in the middle of the wage distribution, as experience is more rewarded among white workers. In contrast to the case of gender wage gaps, non white occupational intensity has a generally negative impact on wages gaps, with a particularly dramatic effect at lower wage quantiles. Non white workers thus benefit from better returns to working in non white-dominated occupations, particularly within low-paid occupations. On the other hand, the occupations' effects contribute positively to wage differentials, and particularly at the very top of the wage distribution (0.99 quantile). Thus, while being employed in non white-dominated occupations reduces white wages within low-paid occupations, white workers are highly rewarded by their heavy representation in top-occupations.

In summary, the results from the RIF-OLS methodology indicate that racial wage differentials are primarily explained by differences in observed characteristics, and particularly by differences in educational attainments, while these differences tends to widen at higher wage levels. There are also small differences in returns to characteristics, though, interestingly, non white occupational intensity seems to contribute to positive discrimination within low-paid occupations. That said, the differences in returns tend to widen as we move to the top of the wage distribution, and there are very significant unexplained differences in returns at the top of the wage distribution. This is partially explained by the diminishing returns to work in non white-dominated occupations, as non white workers particularly benefit in low-paid jobs. However, there could be other factors that affect higher returns for white workers at the top of the distribution such as considerable higher returns for white hired in top-job positions. This could be evidence of a *glass ceilings* phenomenon affecting non white workers.

[Tables 3a-3b to tables 6a-6b about here]

5. Conclusions

The aim of this paper has been to contribute to the analysis of the evolution of gender and racial wage differentials in the Brazilian labour market, while making two particular contributions. First, we have moved beyond investigating wage differentials at mean values in order to consider wage differential at different points in the wage distribution. To this end we have employed two recent quantile decomposition techniques, developed by Melly (2006) and Firpo, Fortin and Lemieux (2009), in order to identify the explained and unexplained elements contributing to wage differentials at different points in the distribution. Second, within the decomposition analysis we have drawn on a novel dataset in order to focus attention on connections between occupational segregation and wage discrimination, focusing on the influence of female and non white occupational intensity on wage differentials. This conclusion briefly summarizes our main findings.

As a starting point, the decomposition results at the mean revealed that gender wage gaps are smaller than racial wage gaps, in large part because gender wage gaps have declined significantly over the last two decades. The considerable, and relatively constant, size of racial wage differentials is of obvious concern, while the sharp decline in gender wage gaps is encouraging. That said, the decomposition results provide a more nuanced picture of the underlying components of these trends. In the case of gender differentials, the shape decline in aggregate wage gaps has been driven to a significant degree by changes in characteristics, led by increasing female education, while unexplained variation, which is potentially indicative of direct wage discrimination, has been declining, but remains positive and significant. Interestingly, and consistent with the second goal of the paper, we find evidence that the latter is closely related to the question of occupational segregation, as men are more rewarded than women particularly when employed in female dominated occupations. In the case of racial differentials, lower wages for non whites are overwhelmingly the result of consistently lower endowments, again with education playing a leading role. Meanwhile unexplained differences in the wage structure are positive and significant, but lower than those related to gender based wage differential. These very different patterns suggest that the challenges of reducing wage differentials are quite different depending on whether the focus is on gender or race.

With these results as a baseline, decomposing the wage differentials at different quantiles revealed important differences across the wage distribution, particularly in relation to gender gaps. Gender wage differentials tend to show a sort of U-shaped pattern, indicating higher wage differentials at the extremes of the wage distribution. Again, these differentials are primarily the result of wage structure effects, which have declined considerably over time. In turn, this U-shaped pattern reflects differences in the impact of the wage structure component between the formal and non-formal labour markets: The wage structure effect is greater at higher quantiles in the formal market, while in the non-formal sectors the effect of coefficients is considerably greater at the bottom of the wage distribution. This suggests the existence of a *sticky floor* phenomenon for women working in non-formal sectors, while it suggests the existence of continued *glass ceilings* in the formal sector where, despite greater endowments than men, women continue to receive lower wages. Turning to racial wage differentials a single key message emerges across the formal and non-formal sectors: wage differentials tend to widen at higher wage quantiles, due to *both* larger differences in characteristics in favour of white workers and higher returns to those characteristics, while this pattern does not seem to have improved over time. Aside from suggesting the importance of policy to improve the endowments of non white workers, the continued existence of uneven returns supports the hypothesis of the existence of *glass ceilings* for non white workers.

Finally, by employing the RIF-OLS technique developed by Firpo, Fortin and Lemieux (2009) we gain additional insight the role of individual variables in accounting for wage gaps. Focusing first on the importance of characteristics, we find, consistent with the basic OB decomposition, that education is the major contributor to better female characteristics, while we can now see that this effect is particularly important as we move to the top of the wage distribution. Education is equally the most important characteristic in looking at racial wage gaps, though in that case is serves to increase wage differentials, as white workers have more education than no whites, while this effect expands at higher quantiles.

Turning attention to the effects of coefficients on gender wage gaps we find that men's experience is rewarded more than women's at the top of the wage distribution, while men working in female-dominated occupations are better paid than women, again particularly in top formal jobs and low paid informal occupations. These trends reinforce the apparent existence of *sticky floors* in non-formal occupations and of *glass ceilings* in formal activities. Looking at racial wage gaps, occupational intensity again plays an important role, though in the opposite direction, as non white workers receive higher wages in non white-dominated occupations, particularly within low-paid occupations. However, while occupation intensity thus favours non white workers in low-paid occupations, we see that the returns on occupations' effects contribute positively to wage differentials, with very large effects at the very top of the wage distribution. Thus, while being employed in non white-dominated occupations marginally depresses white wages within low-paid occupations, white workers are very highly rewarded by the choice of top-occupations. This would seem to be evidence of a *glass ceilings* phenomenon affecting non white workers.

Taken together these results provide a comparatively nuanced and disaggregated view of wage discrimination in Brazil, and of the connections between wage discrimination and occupational segregation (the latter of which is explored in much more detail in Salardi, 2011). These findings are suggestive of key areas of focus for interventions aimed at reducing wage differentials and of key areas of continued unexplained differences in wage structure, which are indicative of continuing discrimination in parts of the labour market. Finally, by treating gender and racial wage differentials side-by-side the analysis highlights certain commonalities, but also sharp differences that point towards differing challenges moving forward.

References

Appleton, S., J. Hoddinott, and P. Krishnan (1999, January). The gender wage gap in three African countries. Economic Development and Cultural Change 47 (2), 289-312.

Arcand, J.-L. and B. D'Hombres (2004). Racial discrimination in the Brazilian labour market: Wage, employment and segregation effects. Journal of International Development 16, 1053-1066.

Baker, M. and N. M. Fortin (2003). Occupational gender composition and wages in Canada, 1987-1988. Canadian Journal of Economics 34 (2), 345-376.

Bayard, K., J. Hellerstein, D. Neumark, and K. Troske (2003, October). New evidence on sex segregation and sex differences in wages from matched employee-employer data. Journal of Labor Economics 21 (4), 887-922.

Birdsall, N. and J. Berhman (1991). Why do males earn more than females in urban Brazil: Earning discrimination or job discrimination? In N. Birdsall and R. Sabot (Eds.), Unfair Advantage Labor Market Discrimination in Developing Countries. World Bank.

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. Journal of Human Resources 8 (4), 436-55.

Bourguignon, F., F. M. and M. Gurgand. (2007). Selection bias correction based on the multinomial logit model: Monte comparisons. Journal of Economic Surveys 21 (1), 174-205.

Brown, R. S., M. Moon, and B. S. Zoloth (1980). The Journal of Human Resources XV (1).

Buchinsky, M. (1998). The dynamics of changes in the female wage distribution in the U.S.A.: a quantile regression approach. Journal of Applied Econometrics 13 (1), 1-30.

Cacciamali, M. C. and G. I. Hirata (2005). A influencia da raca e do genero nas oportunidades de obtencao de renda uma analise da discriminacao em mercados de trabalho distintos: Bahia e Sao Paulo. Est. econ. 35 (4), 767-795.

Cacciamali, M. C., F. Tatei, and J. W. Rosalino (2009). Estreitamento dos diferenciais de salarios e aumento do grau de discriminacao, limitacoes da mensuracao padrao? Planejamento e politicas publica 33, 195-222.

Carneiro, F. G. and A. Henley (2001). Modelling formal vs. informal employment and earnings: Micro-econometric evidence for Brazil. mimeo.

Carvalho, A. P. d., M. C. Neri, and D. B. do Nascimento Silva (2006). Diferenciais de salarios por raca e genero no brasil: Aplicacao dos procedimentos de Oaxaca e Heckman em pesquisas amostrais complexas. Trabalho apresentado no XV Encontro Nacional de Estudos Populacionais, ABEP, realizado em Caxambffu-MG Brasil, de 18 a 22 de Setembro de 2006.

Chernozhukov, V., I. Fernandez-Val, and B. Melly (2009, May). Inference on counterfactual distributions. CeMMAP working papers CWP09/09, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Coelho, D., R. Veszteg, and F. V. Soares (2010). Regressao quantilico com correcao para a seletividade amostral: Estimativa dos retornos educacionais e diferenciais raciais na distribuicao de salarios das mulheres no brasil. Working Papers 1483, IPEA, Brasil.

Cotter, D. A., J. M. Hermsen, and R. Vanneman (2003). The effects of occupational gender segregation across race. Sociological Quarterly 44 (1), 17-36.

Cotton, J. (1988). On the decomposition of wage differentials. Review of Economics and Statistics 70 (2), 236-243.

Di Nardo, J., N. M. Fortin, and T. Lemieux (1996). Labor market institutions and the distribution of wages, 1973-1992: A semi-parametric approach. Econometrica 64 (5), 1011-1044.

Dolton, P.J., G. H. Makepeace and W. V. D. Klaauw (1989). Occupational choice and earnings determination: The role of sample selection and non- pecuniary factors. Oxford Economic Papers 41 (3), 573-594.

Donald, S. G., D. A. Green, and H. J. Paarsch (2000, October). Differences in wage distributions between Canada and the united states: An application of a flexible estimator of distribution functions in the presence of covariates. Review of Economic Studies 67 (4), 609-33.

Duarte, A., P. C. Ferreira, and M. Salvato (2003). Disparidades regionais ou educacionais. Technical report.

Fields, J. and E. N. Wolff (1995). Interindustry wage differentials and the gender wage gap. Industrial and Labor Relations Review 49 (1), 105-20.

Firpo, S., N. M. Fortin, and T. Lemieux (2009). Unconditional quantile regressions. Econometrica 77 (3), 953-973.

Fortin, N. M. and T. Lemieux (1998). Rank regressions, wage distributions, and the gender gap. Journal of Human Resources 33 (3), 610-643.

Fortin, N., T. Lemieux, and S. Firpo (2011, October). Decomposition Methods in Economics, Volume 4 of Handbook of Labor Economics, Chapter 1, pp. 1-102. Elsevier.

Gill, A. M. (1996). Incorporating the causes of occupational differences in studies of racial wage differentials. The Journal of Human Resources XXIX(1), 20-41.

Gosling, A., S. Machin, and C. Meghir (2000). The changing distribution of male wages in the U.K. The Review of Economic Studies 67 (4), 635-666.

Guimares, J. F., T. Cavalcanti, and R. S. Neto (2006). Accounting for labour income differences in Brazil: the role of human capital. Universidade Federal de Pernambuco, mimeo.

Haisken-DeNew, J. P. and C. M. Schmidt (1997). Interindustry and interregion differentials: Mechanism and interpretation. The Review of Economics and Statistics 79 (3), 516-521.

Hellerstein, J. and D. Neumark (2006). Using matched employer-employee data to study labor market discrimination. In W. M. Rodgers (Ed.), Handbook on the Economics of Discrimination. Edward Elgar Publishing.

Hellerstein, J. K. and D. Neumark (2007, January). Production function and wage equation estimation with heterogeneous labor: Evidence from a new matched employer-employee data set. In Hard-to-Measure Goods and Services: Essays in Honor of Zvi Griliches, NBER Chapters, pp. 31-71. National Bureau of Economic Research, Inc.

Hellerstein, J. K., D. Neumark, and K. R. Troske (2002). Market forces and sex discrimination. Journal of Human Resources 37 (2), 353-380.

Horrace, W. C. and R. L. Oaxaca (2001). Inter-industry wage differentials and the gender wage gap: An identification problem. Industrial and Labor Relations Review 54 (3), 611-618.

Juhn, C., K. M. Murphy, and B. Pierce (1993). Wage inequality and the rise in returns to skill. Journal of Political Economy 101 (3), 410-442.

Krueger, A. B. and L. H. Summers (1988). Efficiency wages and the inter-industry wage structure. Econometrica 56 (2).

Loureiro, P. R. A., F. G. Carneiro, and A. Sachsida (2004). Race and gender discrimination in the labor market: an urban and rural sector analysis for Brazil. Journal of Economic Studies 31 (2), 129-143.

Machado, J. A. and J. Mata (2005). Counterfactual decomposition of changes in wage distribution using quantile regression. Journal of Applied Econometrics 20, 445-465.

Macpherson, D. A. and B. T. Hirsch (1995). Wages and gender composition: Why do women's jobs pay less? Journal of Labor Economics 13 (3).

Melly, B. (2005). Decomposition of differences in distribution using quantile regression. Labour Economics 12, 577-90.

Melly, B. (2006). Estimation of counterfactual distributions using quantile regression mimeo, Swiss Institute for International Economics and Applied Economic Research(SIAW), University of St. Gallen.

Miller, P. W. (1987). The wage effect of the occupational segregation of women in Britain. The Economic Journal 97, 885-896.

Neuman, S. and J. Silber (1996). Wage discrimination across ethnic groups: Evidence from Israel. Economic Inquiry 34 (4), 648-61.

Neuman, S. and R. L. Oaxaca (2004). Wage decompositions with selectivity-corrected wage equations: A methodological note. Journal of Economic Inequality 2 (1), 3-10.

Neumark, D. (1988). Employers' discrimination behavior and the estimation of wage discrimination. 'Journal of Human Resources 23 (3), 279-295.

Newey, W. (1991). Two step series estimation of sample selection model. Mimeo, MIT.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. International Economic Review 14 (3), 693-709.

Oaxaca, R. L. and M. R. Ransom (1994). On discrimination and the decomposition of wage differentials. Journal of Econometrics 61 (1), 5-21.

Oaxaca, R. L. and M. R. Ransom (1998). Calculation of approximate variances for wage decomposition differentials. Journal of Economic and Social Measurement 24, 55-61.

Ometto, A. M. H., R. Hoffmann, and M. C. Alvaes (1999). Participacao da mulher no mercado de trabalho: Discriminacao em pernambuco e sao paulo. Revista Brasileira de Economia 53 (3), 287-322.

Reilly, B. (1991). Occupational segregation and selectivity bias in occupational wage equations: An empirical analysis using irish data. Applied Economics 23 (1), 1-7.

Salardi, P. (2011). The Evolution of Gender and Racial Occupational Segregation across Formal and non-Formal Labour Markets in Brazil – 1987 to 2006. Working Paper Series 3011, Department of Economics, University of Sussex.

Salardi, P. (2012). The evolution of occupational segregation and wage discrimination in Brazil over time. Phd Thesis in Economics. Department of Economics, University of Sussex. (work in progress).

Santos, R. V. and E. P. Ribeiro (2006). Diferenciais de rendimentos entre homens e mulheres no brasil revisitado: explorando o "`teto de vidro". Centro Universitario Newton Paiva/MG e PPGE/UFRGS.

Silva, N. d. D. V. and A. L. Kassouf (2000). Mercados de trabalho formal e informal: uma analise da discriminacao e da segmentacao. Nova Economia 10 (1).

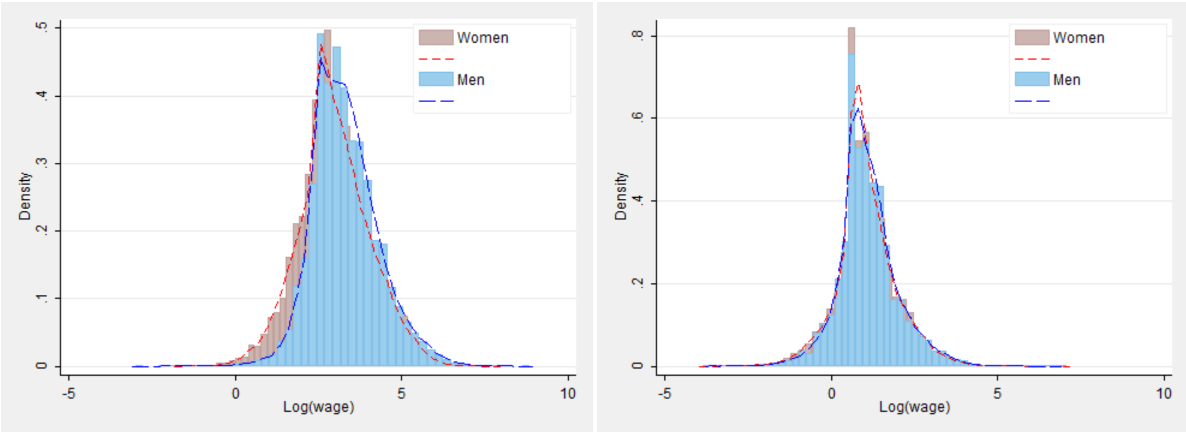
Stecler, M., J. B. Smith, J. A. Breslaw, and G. Monette (1992). Labor force behavior and earnings of Brazilian women and men. In G. Psacharopolous and Z. Tzannatos (Eds.), Case Studies of Women's Employment and Pay in Latin America. World Bank.

Tiefenthaler, J. (1992). Female labor force participation and wage determination in Brazil 1989. In G. Psacharopolous and Z. Tzannatos (Eds.), Case Studies of Women's Employment and Pay in Latin America. World Bank.

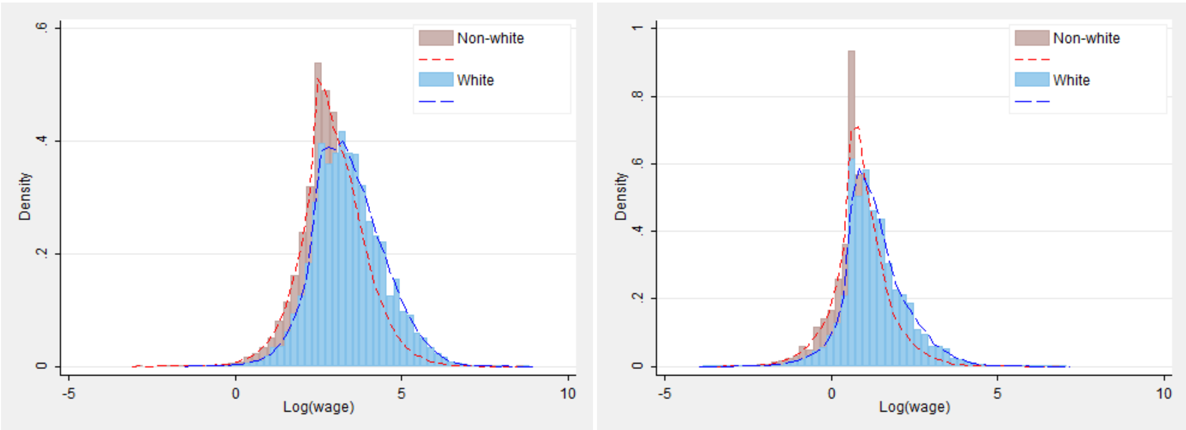
Figures and tables – to be inserted in the text

Figure 1: Kernel density of log hourly wage

Panel A - Kernel density of log hourly wage by gender – 1987 and 2006



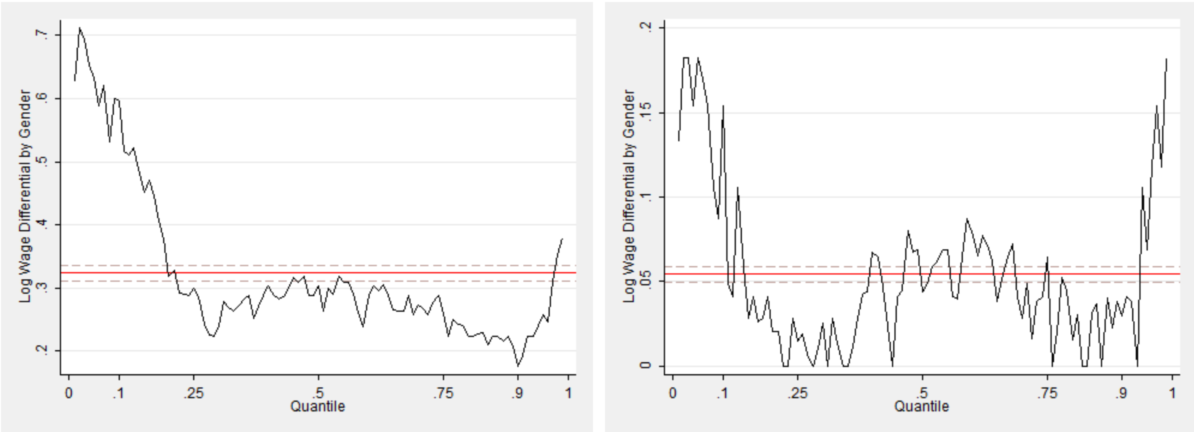
Panel B - Kernel density of log hourly wage by gender – 1987 and 2006



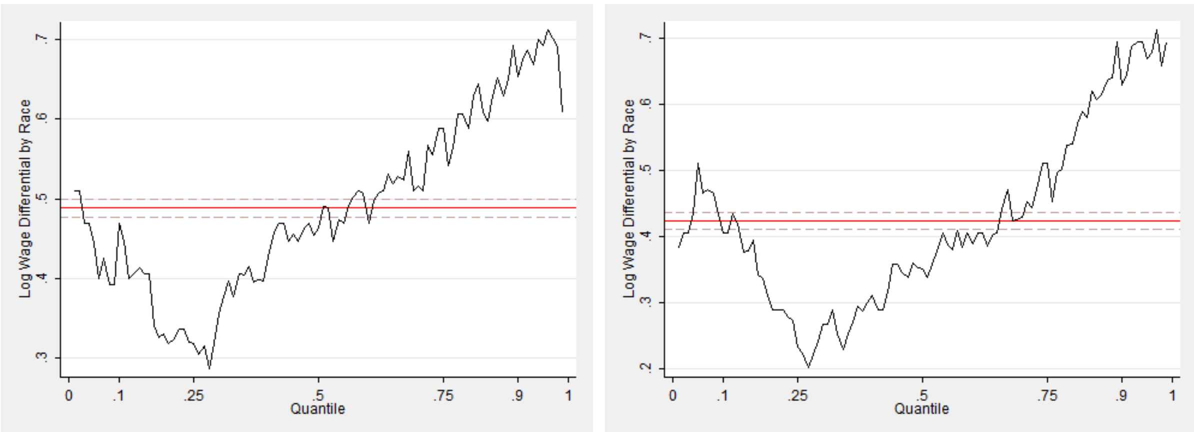
Source: Author’s computations using PNAD 1987-2006.

Figure 2: Wage differentials over wage quantiles

Panel A – Wage differentials by gender – 1987 and 2006



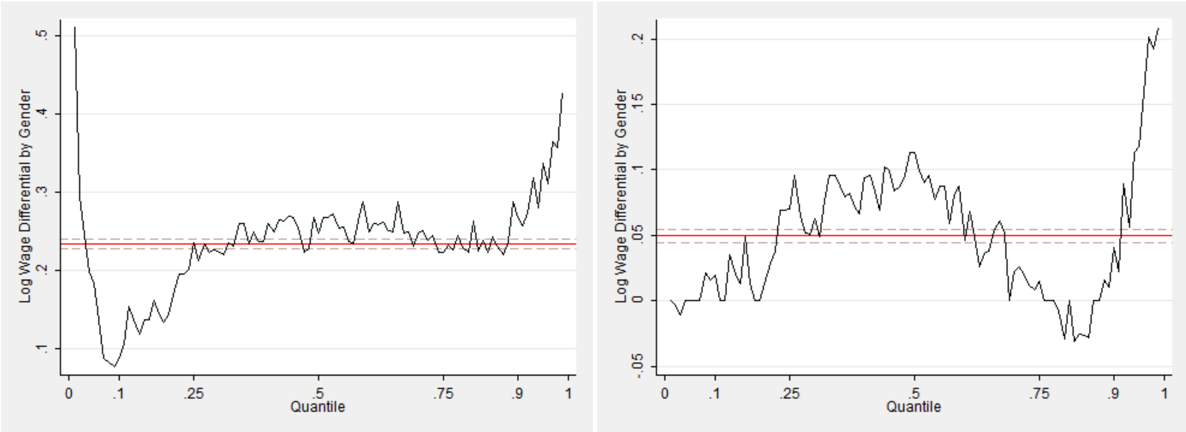
Panel B – Wage differentials by race – 1987 and 2006



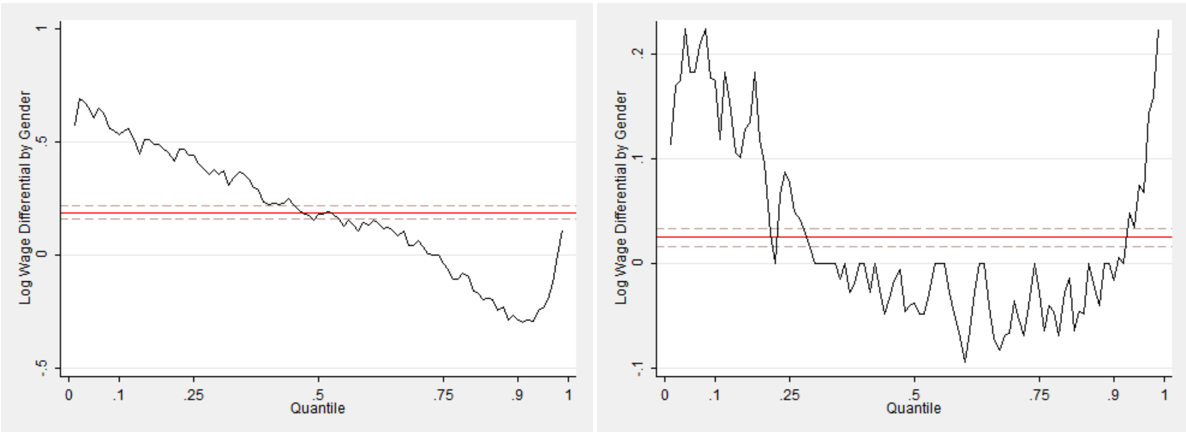
Source: Author's computations using PNAD 1987-2006.

Figure 3a: Wage differentials over wage quantiles by gender and disaggregated by formal and non-formal sectors

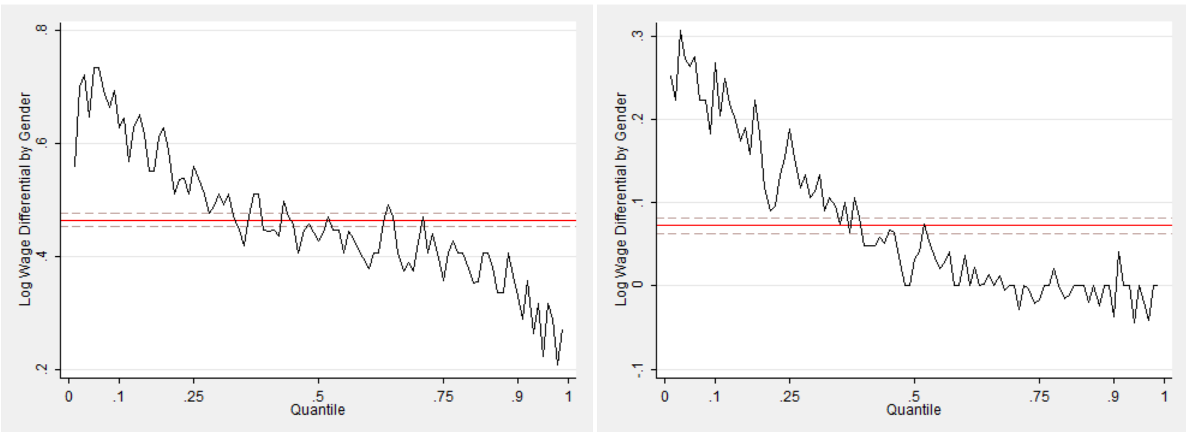
Panel A – Formal sector – 1987 and 2006



Panel B - Informal sector – 1987 and 2006



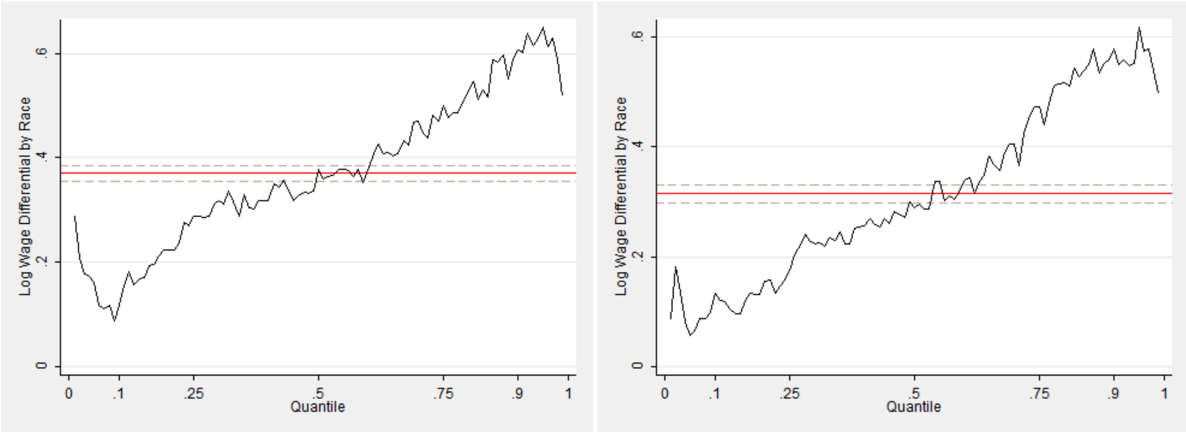
Panel C – Self-employed sector – 1987 and 2006



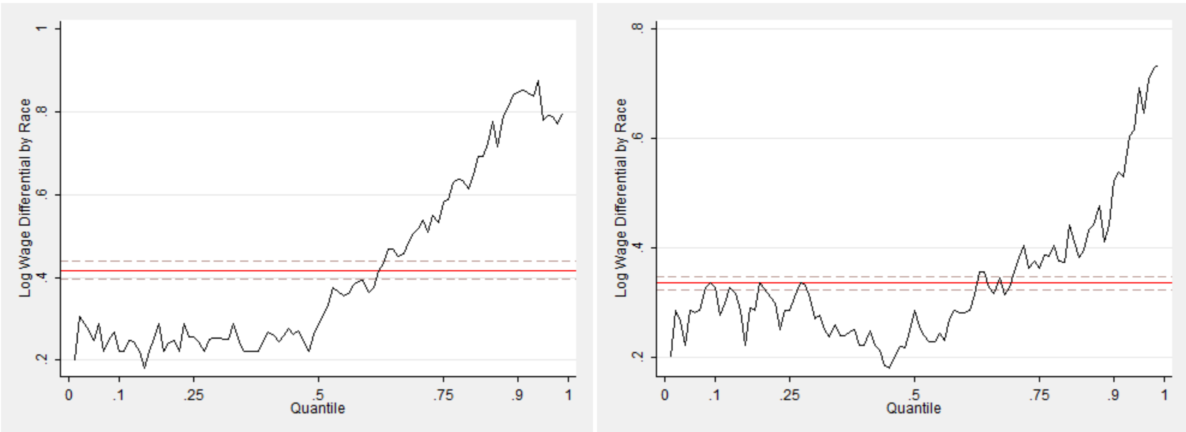
Source: Author's computations using PNAD 1987-2006.

Figure 3b: Wage differentials over wage quantiles by race and disaggregated by formal and non-formal sectors

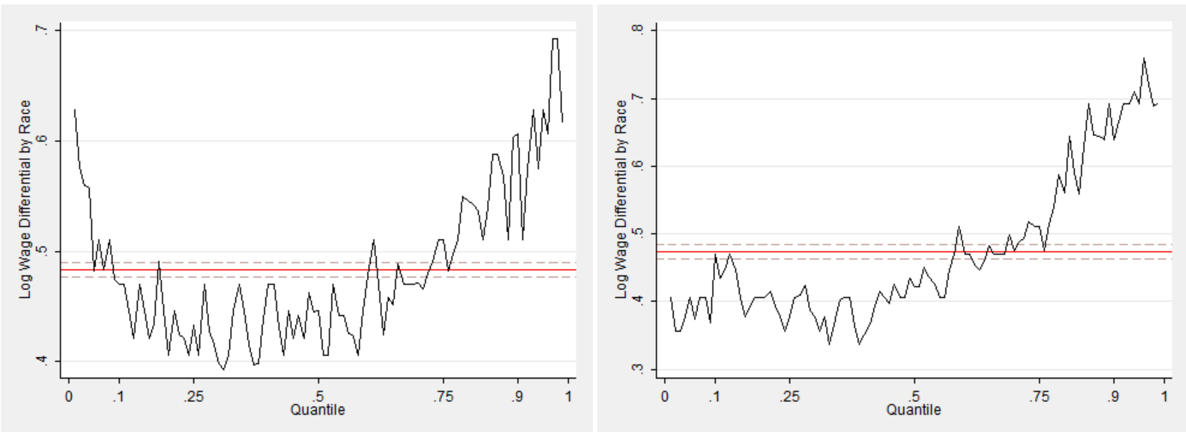
Panel A – Formal sector – 1987 and 2006



Panel B - Informal sector – 1987 and 2006



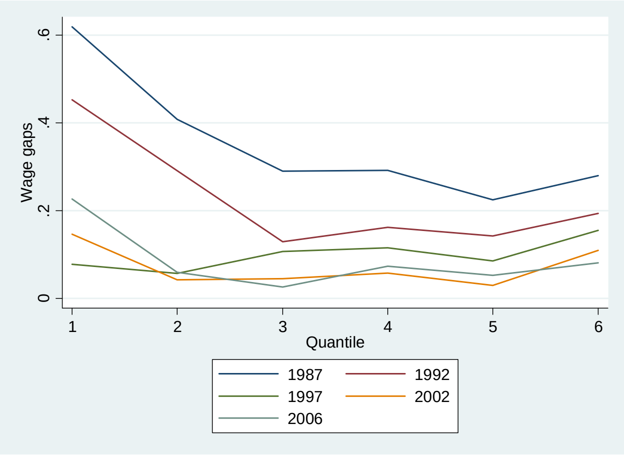
Panel C – Self-employed sector – 1987 and 2006



Source: Author’s computations using PNAD 1987-2006.

Figure 4: Evolution of wage gaps over time, all labour market

Panel A – Gender wage gaps



Panel B – Racial wage gaps

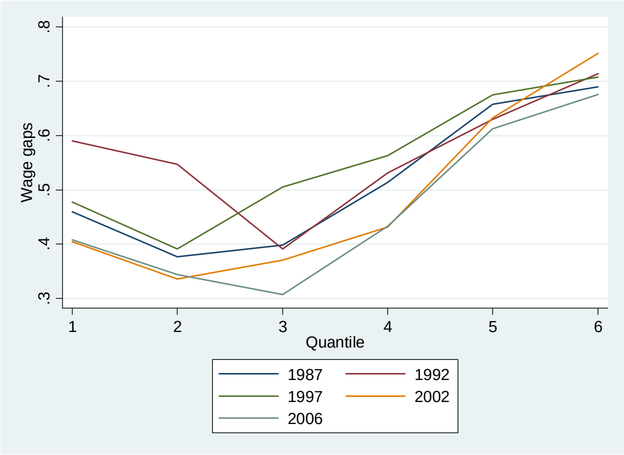
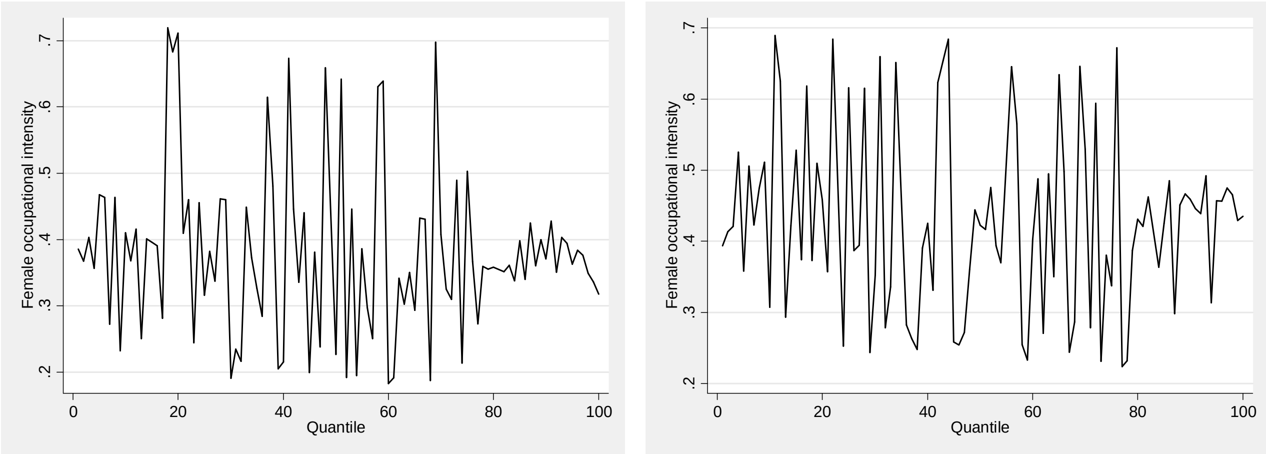
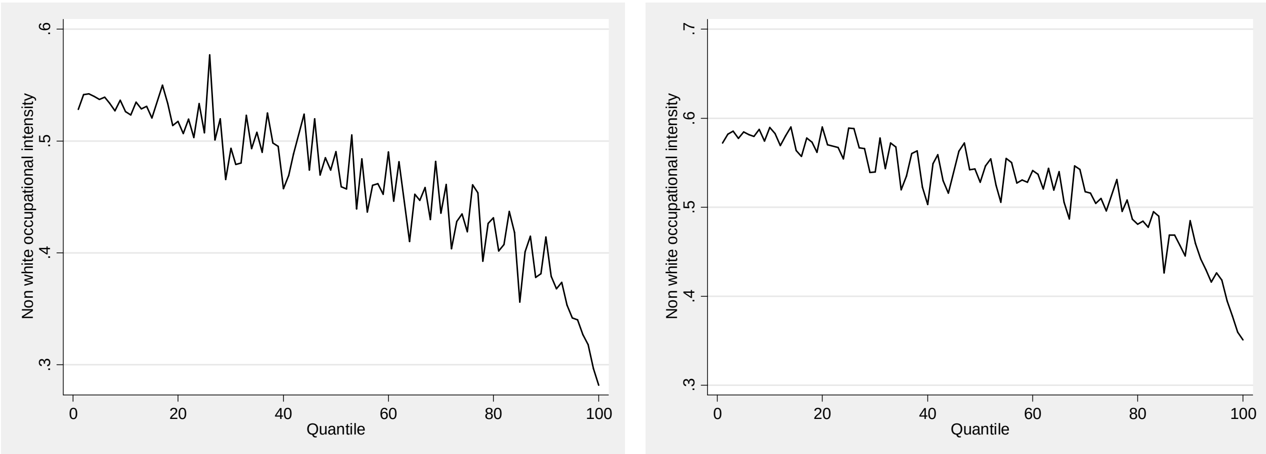


Figure 5: Female (non white) occupational intensity over wage quantiles

Panel A – Female occupational intensity in 1987 and 2006



Panel B – Non white occupational intensity in 1987 and 2006



Source: Author's computations using PNAD 1987-2006.

Figure 6: Average wages over female (or non white occupational intensity

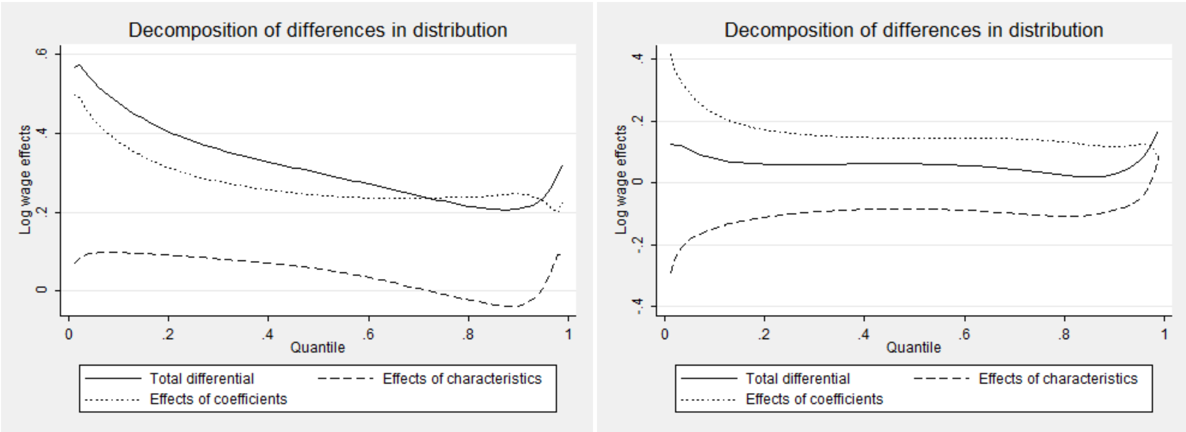
Panel A – Wages by gender over female occupational intensity in 1987 and 2006

Panel B – Wages by race over non white occupational intensity in 1987 and 2006

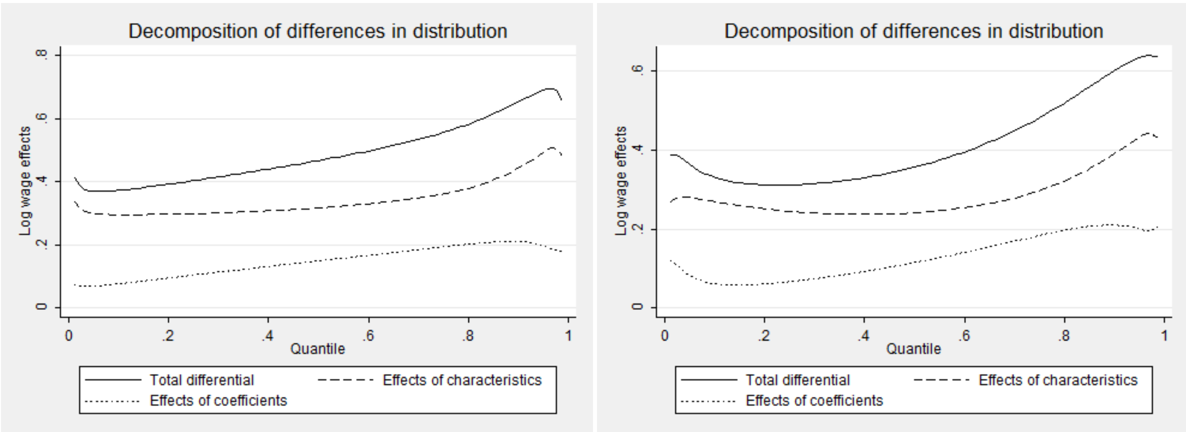
Source: Author's computations using PNAD 1987-2006.

Figure 7: Melly (2006) quantile decomposition results, all labor market

Panel A – Decomposition of gender wage gaps – 1987 and 2006



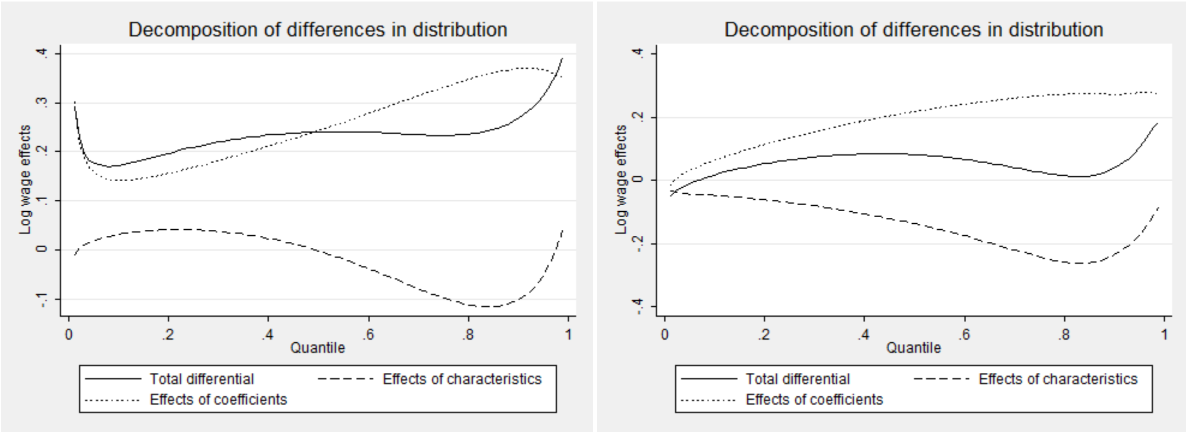
Panel B – Decomposition of racial wage gaps – 1987 and 2006



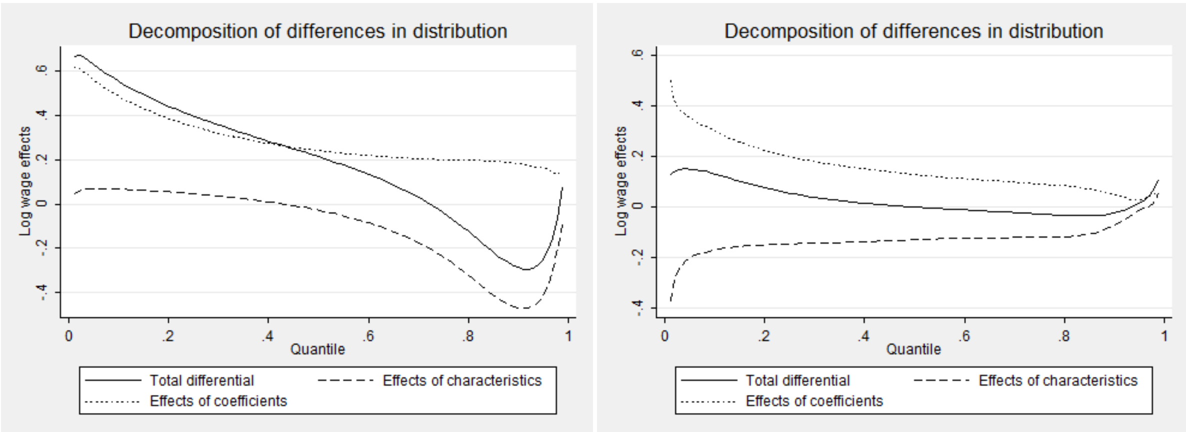
Source: Author's computations using PNAD 1987-2006.

Figure 8a: Melly (2006) quantile decomposition results of gender wage gaps, disaggregated by formal and non-formal sectors

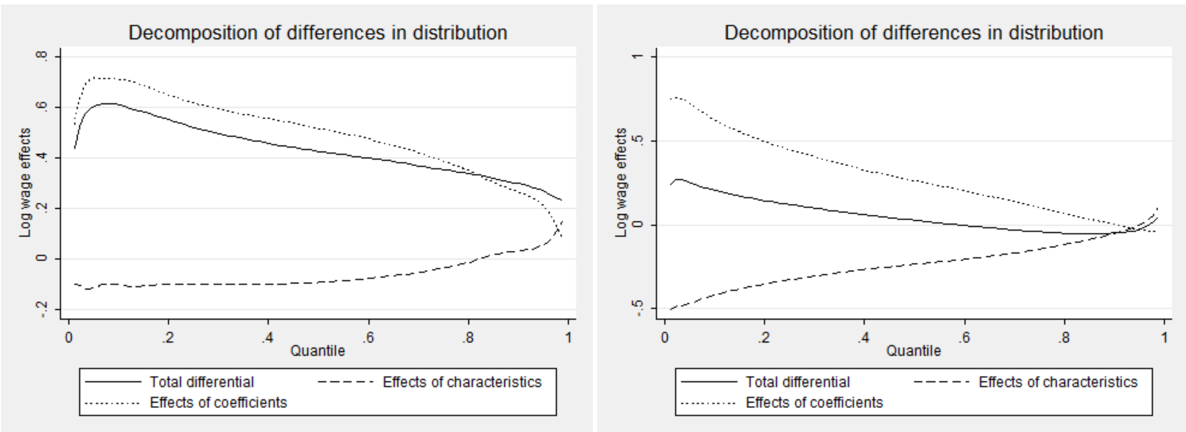
Panel A – Formal sector – 1987 and 2006



Panel B – Informal sector – 1987 and 2006



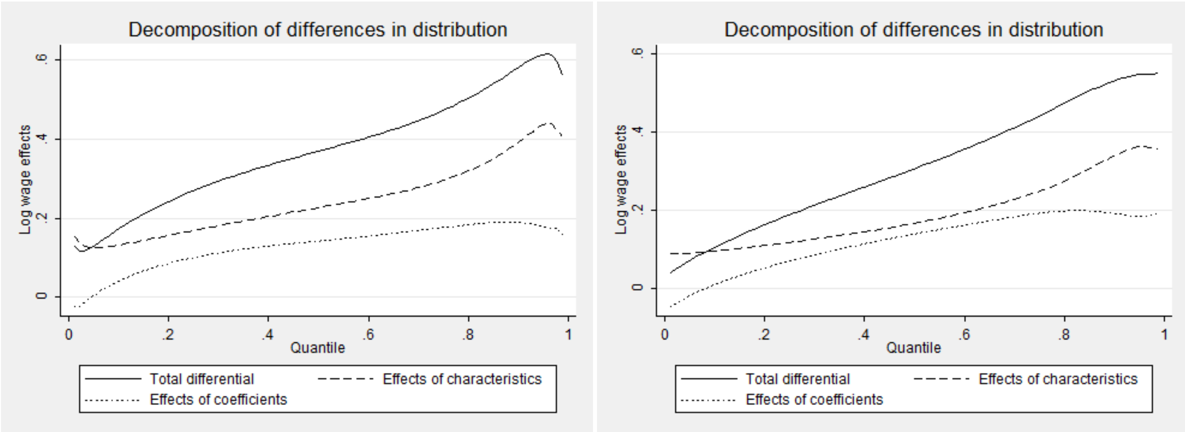
Panel C – Self-employed sector – 1987 and 2006



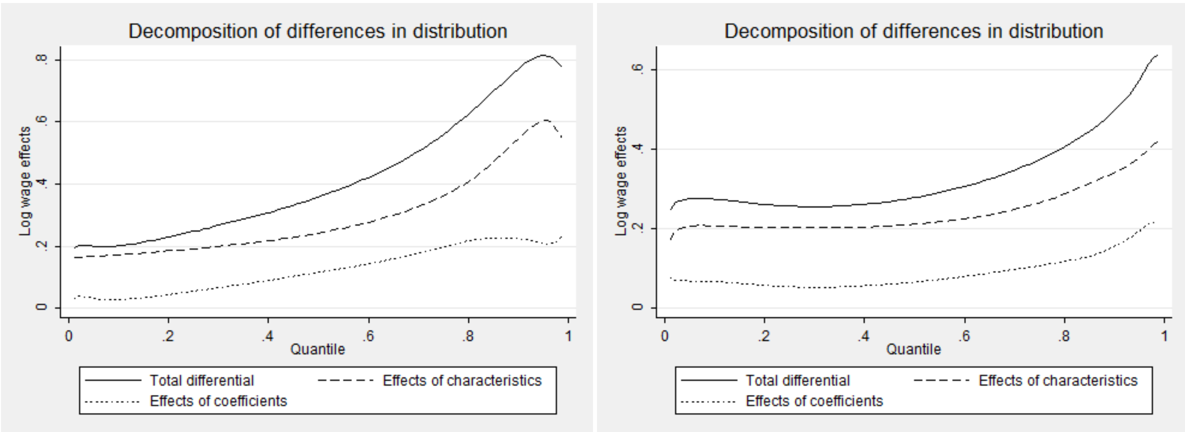
Source: Author’s computations using PNAD 1987-2006.

Figure 8b: Melly (2006) quantile decomposition results of racial wage gaps, disaggregated by formal and non-formal sectors

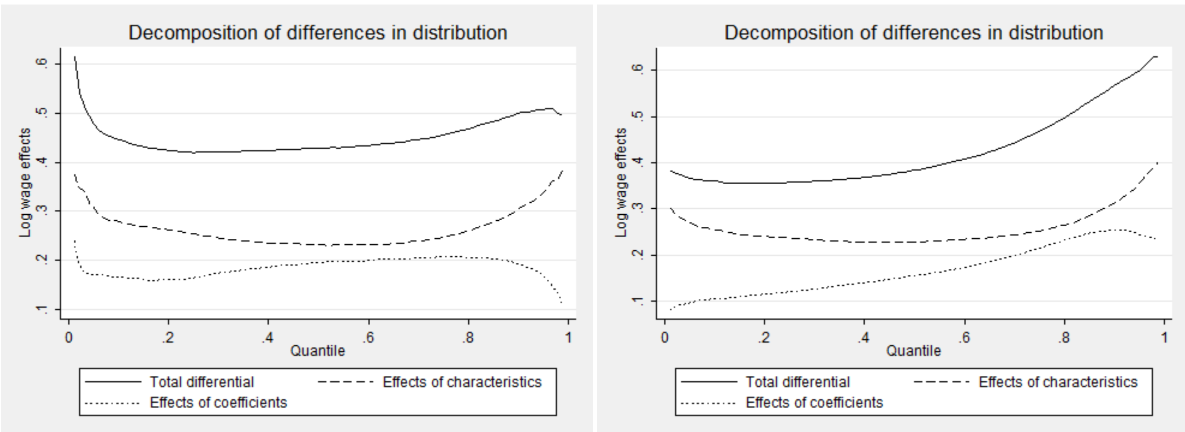
Panel A – Formal sector – 1987 and 2006



Panel B – Informal sector – 1987 and 2006



Panel C – Self-employed sector – 1987 and 2006



Source: Author's computations using PNAD 1987-2006.

Table 1a: Coefficients for female and non white occupational intensity from quantile regressions, pooled and by sub-samples – year 1987

	(1) mean	(2) 0.10	(3) 0.25	(4) 0.50	(5) 0.75	(6) 0.90
Panel A- Pooled sample						
focc [a]	-0.379***	-0.341***	-0.363***	-0.382***	-0.394***	-0.402***
	(0.010)	(0.015)	(0.012)	(0.011)	(0.013)	(0.018)
focc [b]	-0.186***	-0.200***	-0.188***	-0.176***	-0.160***	-0.153***
	(0.014)	(0.023)	(0.017)	(0.015)	(0.019)	(0.022)
focc [c]	-0.119***	-0.118***	-0.099***	-0.097***	-0.120***	-0.117***
	(0.015)	(0.025)	(0.018)	(0.017)	(0.019)	(0.026)
focc [d]	-0.099***	-0.104***	-0.081***	-0.069***	-0.082***	-0.122***
	(0.015)	(0.025)	(0.017)	(0.016)	(0.018)	(0.027)
nwocc [e]	-1.802***	-1.504***	-1.697***	-1.847***	-1.937***	-2.013***
	(0.029)	(0.046)	(0.032)	(0.031)	(0.036)	(0.048)
nwocc [f]	-0.467***	-0.228***	-0.539***	-0.670***	-0.706***	-0.592***
	(0.051)	(0.086)	(0.061)	(0.057)	(0.068)	(0.080)
nwocc [g]	-0.482***	-0.174*	-0.512***	-0.619***	-0.693***	-0.751***
	(0.053)	(0.095)	(0.064)	(0.058)	(0.065)	(0.088)
nwocc [h]	-0.392***	-0.046	-0.353***	-0.536***	-0.588***	-0.664***
	(0.052)	(0.081)	(0.062)	(0.054)	(0.064)	(0.093)
Panel B – By gender or race						
FEMALES						
focc [a]	-0.473***	-0.381***	-0.400***	-0.424***	-0.483***	-0.565***
	(0.015)	(0.025)	(0.019)	(0.017)	(0.018)	(0.024)
focc [b]	-0.401***	-0.440***	-0.407***	-0.347***	-0.352***	-0.406***
	(0.025)	(0.044)	(0.031)	(0.028)	(0.032)	(0.044)
focc [c]	-0.363***	-0.327***	-0.314***	-0.286***	-0.306***	-0.360***
	(0.028)	(0.050)	(0.036)	(0.030)	(0.034)	(0.051)
focc [d]	-0.280***	-0.238***	-0.227***	-0.221***	-0.266***	-0.298***
	(0.028)	(0.040)	(0.034)	(0.028)	(0.032)	(0.050)
MALES						
focc [a]	-0.262***	-0.212***	-0.273***	-0.325***	-0.290***	-0.230***
	(0.013)	(0.021)	(0.017)	(0.015)	(0.018)	(0.024)
focc [b]	0.095***	0.147***	0.100***	0.045**	0.038	0.045
	(0.019)	(0.030)	(0.024)	(0.023)	(0.027)	(0.035)
focc [c]	0.112***	0.186***	0.139***	0.052**	0.030	0.040
	(0.020)	(0.031)	(0.025)	(0.023)	(0.026)	(0.037)
focc [d]	0.097***	0.131***	0.112***	0.052**	0.049**	0.011
	(0.019)	(0.032)	(0.023)	(0.022)	(0.025)	(0.040)
NON WHITES						
nwocc [e]	-1.629***	-1.154***	-1.428***	-1.694***	-1.873***	-1.938***
	(0.044)	(0.065)	(0.048)	(0.049)	(0.057)	(0.072)
nwocc [f]	-0.224***	0.103	-0.232**	-0.591***	-0.541***	-0.345***
	(0.077)	(0.121)	(0.093)	(0.088)	(0.101)	(0.128)
nwocc [g]	-0.326***	0.049	-0.268***	-0.677***	-0.639***	-0.581***
	(0.078)	(0.120)	(0.085)	(0.087)	(0.100)	(0.127)
nwocc [h]	-0.147*	0.388***	-0.082	-0.408***	-0.461***	-0.501***
	(0.077)	(0.124)	(0.089)	(0.085)	(0.101)	(0.141)
WHITES						
nwocc [e]	-1.899***	-1.702***	-1.821***	-1.912***	-1.967***	-2.052***
	(0.038)	(0.059)	(0.043)	(0.041)	(0.044)	(0.066)
nwocc [f]	-0.691***	-0.495***	-0.695***	-0.753***	-0.936***	-0.819***
	(0.070)	(0.109)	(0.092)	(0.080)	(0.095)	(0.125)
nwocc [g]	-0.607***	-0.361***	-0.592***	-0.622***	-0.823***	-0.855***
	(0.072)	(0.114)	(0.091)	(0.084)	(0.092)	(0.118)
nwocc [h]	-0.600***	-0.271**	-0.558***	-0.625***	-0.789***	-0.913***
	(0.071)	(0.111)	(0.088)	(0.076)	(0.096)	(0.130)

Note: [a] control for male, white, age, age squared, years of education, 5 main geographical region of Brazil, urban, formal; [b] as specification ‘a’ plus control for 23 occupational codes at 2-digit level; [c] as specification ‘a’ plus control for 23 occupational codes at 2-digit level and 9 economic sectors; [d] as specification ‘a’ plus control for 23 occupational codes at 2-digit level, 9 economic sectors and 27 states of Brazil; [e] control for male, white, age, age squared, years of education, 5 main geographical region of Brazil, urban, formal; [f] as specification ‘e’ plus control for 23 occupational codes at 2-digit level; [g] as specification ‘a’ plus control for 23 occupational codes at 2-digit level and 9 economic sectors; [h] as specification ‘e’ plus control for 23 occupational codes at 2-digit level, 9 economic sectors and 27 states of Brazil.
Source: Author’s computations using PNAD 1987-2006.

Table 1b: Coefficients for female and non white occupational intensity from quantile regressions, pooled and by sub-samples – year 2006

	(1) mean	(2) 0.10	(3) 0.25	(4) 0.50	(5) 0.75	(6) 0.90
Panel A- Pooled sample						
focc3[a]	-0.093***	-0.089***	-0.093***	-0.086***	-0.101***	-0.125***
	(0.007)	(0.010)	(0.007)	(0.008)	(0.009)	(0.013)
focc3[b]	-0.043***	-0.031**	-0.025***	-0.036***	-0.066***	-0.107***
	(0.010)	(0.013)	(0.010)	(0.010)	(0.012)	(0.018)
focc3[c]	-0.051***	-0.047***	-0.036***	-0.032***	-0.051***	-0.082***
	(0.010)	(0.013)	(0.011)	(0.010)	(0.012)	(0.019)
focc3[d]	-0.044***	-0.047***	-0.025**	-0.020*	-0.043***	-0.078***
	(0.010)	(0.014)	(0.010)	(0.010)	(0.012)	(0.016)
nwocc3[e]	-1.753***	-1.187***	-1.343***	-1.649***	-2.003***	-2.304***
	(0.020)	(0.027)	(0.020)	(0.019)	(0.023)	(0.035)
nwocc3[f]	-0.818***	-0.473***	-0.646***	-0.834***	-1.083***	-1.152***
	(0.038)	(0.052)	(0.038)	(0.033)	(0.043)	(0.066)
nwocc3[g]	-0.780***	-0.411***	-0.596***	-0.778***	-0.981***	-1.130***
	(0.038)	(0.051)	(0.038)	(0.034)	(0.043)	(0.066)
nwocc3[h]	-0.772***	-0.429***	-0.569***	-0.772***	-0.960***	-1.085***
	(0.038)	(0.050)	(0.037)	(0.034)	(0.044)	(0.059)
Panel B – By gender or race						
FEMALES						
focc3 [a]	-0.179***	-0.053***	-0.066***	-0.086***	-0.203***	-0.422***
	(0.013)	(0.018)	(0.014)	(0.013)	(0.016)	(0.023)
focc3 [b]	-0.153***	-0.119***	-0.087***	-0.084***	-0.152***	-0.242***
	(0.018)	(0.023)	(0.017)	(0.016)	(0.021)	(0.031)
focc3 [c]	-0.181***	-0.173***	-0.111***	-0.109***	-0.149***	-0.233***
	(0.018)	(0.025)	(0.018)	(0.016)	(0.022)	(0.034)
focc3 [d]	-0.174***	-0.145***	-0.102***	-0.100***	-0.162***	-0.229***
	(0.018)	(0.024)	(0.018)	(0.017)	(0.021)	(0.033)
MALES						
focc3 [a]	-0.023**	-0.074***	-0.080***	-0.068***	-0.036***	0.029*
	(0.009)	(0.011)	(0.010)	(0.009)	(0.012)	(0.016)
focc3 [b]	0.018	0.048***	0.033**	-0.011	-0.044***	-0.080***
	(0.013)	(0.018)	(0.014)	(0.012)	(0.017)	(0.023)
focc3 [c]	0.050***	0.062***	0.048***	0.039***	0.022	-0.004
	(0.013)	(0.020)	(0.013)	(0.014)	(0.017)	(0.024)
focc3 [d]	0.054***	0.048**	0.058***	0.041***	0.029*	-0.006
	(0.013)	(0.020)	(0.014)	(0.013)	(0.017)	(0.022)
NON WHITES						
nwocc3 [e]	-1.476***	-0.889***	-1.027***	-1.335***	-1.787***	-2.177***
	(0.028)	(0.039)	(0.027)	(0.027)	(0.034)	(0.051)
nwocc3 [f]	-0.529***	-0.166**	-0.432***	-0.672***	-0.944***	-0.951***
	(0.054)	(0.073)	(0.058)	(0.048)	(0.057)	(0.085)
nwocc3 [g]	-0.506***	-0.216***	-0.392***	-0.644***	-0.882***	-0.950***
	(0.054)	(0.078)	(0.053)	(0.048)	(0.058)	(0.091)
nwocc3 [h]	-0.511***	-0.235***	-0.370***	-0.632***	-0.830***	-0.997***
	(0.054)	(0.074)	(0.059)	(0.046)	(0.061)	(0.078)
WHITES						
nwocc3 [e]	-1.923***	-1.452***	-1.588***	-1.843***	-2.069***	-2.304***
	(0.028)	(0.037)	(0.028)	(0.030)	(0.034)	(0.052)
nwocc3 [f]	-1.105***	-0.792***	-0.865***	-0.988***	-1.173***	-1.331***
	(0.055)	(0.074)	(0.054)	(0.057)	(0.067)	(0.099)
nwocc3 [g]	-1.050***	-0.729***	-0.767***	-0.900***	-1.093***	-1.359***
	(0.055)	(0.071)	(0.054)	(0.056)	(0.064)	(0.093)
nwocc3 [h]	-1.034***	-0.664***	-0.742***	-0.897***	-1.056***	-1.324***
	(0.054)	(0.071)	(0.052)	(0.049)	(0.066)	(0.093)

Note: [a] control for male, white, age, age squared, years of education, 5 main geographical region of Brazil, urban, formal; [b] as specification ‘a’ plus control for 23 occupational codes at 2-digit level; [c] as specification ‘a’ plus control for 23 occupational codes at 2-digit level and 9 economic sectors; [d] as specification ‘a’ plus control for 23 occupational codes at 2-digit level, 9 economic sectors and 27 states of Brazil; [e] control for male, white, age, age squared, years of education, 5 main geographical region of Brazil, urban, formal; [f] as specification ‘e’ plus control for 23 occupational codes at 2-digit level; [g] as specification ‘a’ plus control for 23 occupational codes at 2-digit level and 9 economic sectors; [h] as specification ‘e’ plus control for 23 occupational codes at 2-digit level, 9 economic sectors and 27 states of Brazil.
Source: Author’s computations using PNAD 1987-2006.

Table 2: Oaxaca-Blinder decomposition at the mean, all labour market – 1987 and 2006

Panel A – Gender wage gaps

	1st specification		2nd specification		3rd specification	
	1987	**2006**	**1987**	**2006**	**1987**	**2006**
Explained	-0.163	-0.182	-0.071	-0.156	-0.028	-0.172
s.e.	0.005	0.003	0.008	0.005	0.009	0.005
Unexplained	0.485	0.243	0.393	0.216	0.380	0.242
s.e.	0.005	0.004	0.008	0.005	0.009	0.005
Total gap	0.322	0.061	0.322	0.061	0.352	0.070
s.e.	0.007	0.005	0.007	0.005	0.007	0.005
Expl: age	0.023	-0.003	0.021	-0.003	0.019	-0.003
s.e.	0.002	0.001	0.002	0.001	0.002	0.001
Expl: edu	-0.157	-0.151	-0.115	-0.102	-0.098	-0.096
s.e.	0.004	0.002	0.003	0.002	0.003	0.002
Expl: focc			-0.042	-0.007	-0.045	-0.021
s.e.			0.009	0.005	0.009	0.005
Expl: occ			0.081	-0.026	0.104	0.005
s.e.			0.006	0.004	0.008	0.004
Unexp: age	0.151	0.206	0.083	0.175	0.123	0.166
s.e.	0.039	0.033	0.038	0.033	0.038	0.032
Unexp: edu	-0.101	-0.079	0.000	-0.024	-0.021	-0.028
s.e.	0.008	0.009	0.012	0.011	0.011	0.011
Unexp: focc			0.324	0.113	0.250	0.150
s.e.			0.020	0.014	0.022	0.014
Unexp:occ			-0.065	-0.133	-0.093	-0.156
s.e.			0.076	0.047	0.081	0.047

Panel B: Racial wage gaps

	1st specification		2nd specification		3rd specification	
	1987	**2006**	**1987**	**2006**	**1987**	**2006**
Explained	0.384	0.320	0.401	0.344	0.378	0.342
s.e.	0.006	0.004	0.006	0.004	0.006	0.004
Unexplained	0.105	0.093	0.088	0.068	0.084	0.067
s.e.	0.006	0.004	0.005	0.004	0.006	0.004
Total gap	0.489	0.413	0.489	0.413	0.462	0.409
s.e.	0.006	0.005	0.006	0.005	0.007	0.005
Expl: age	0.019	0.022	0.018	0.019	0.017	0.019
s.e.	0.002	0.002	0.002	0.001	0.002	0.001
Expl: edu	0.293	0.215	0.208	0.138	0.181	0.132
s.e.	0.004	0.003	0.004	0.002	0.003	0.002
Expl: nwocc			0.037	0.057	0.030	0.053
s.e.			0.004	0.003	0.004	0.003
Expl: occ			0.072	0.046	0.052	0.040
s.e.			0.004	0.003	0.004	0.003
Unexp: age	0.295	0.192	0.278	0.153	0.276	0.159
s.e.	0.036	0.032	0.035	0.031	0.035	0.031
Unexp: edu	0.091	0.236	0.076	0.140	0.073	0.137
s.e.	0.006	0.007	0.008	0.009	0.007	0.009
Unexp: nwocc			-0.231	-0.317	-0.226	-0.287
s.e.			0.051	0.042	0.052	0.042
Unexp:occ			0.121	0.291	0.107	0.272
s.e.			0.077	0.057	0.085	0.056

Note: the ‘explained’ component is also named the effect of characteristics or the composition effect; the ‘unexplained’ component is also named the effect of coefficients or wage structure effect.
Source: Author’s computations using PNAD 1987-2006.

Table 3a: Gender wage gap: quantile decomposition results, all labour market – 1987 and 2006

	Gender wage gap - year 1987							Gender wage gap - year 2006						
Quantile	**0.01**	**0.1**	**0.25**	**0.5**	**0.75**	**0.9**	**0.99**	**0.01**	**0.1**	**0.25**	**0.5**	**0.75**	**0.9**	**0.99**
Raw log gap	0.629	0.596	0.299	0.303	0.260	0.176	0.379	0.134	0.154	0.014	0.044	0.065	0.030	0.182
Decomposition method: Machado & Mata (2005) - Melly (2006)														
Explained	0.069	0.097	0.087	0.056	-0.009	-0.038	0.093	-0.291	-0.146	-0.101	-0.084	-0.105	-0.089	0.091
s.e.														
Unexplained	0.498	0.377	0.292	0.243	0.237	0.246	0.226	0.416	0.223	0.161	0.145	0.139	0.118	0.078
s.e.														
Total gap	0.567	0.475	0.379	0.299	0.228	0.208	0.319	0.125	0.077	0.060	0.062	0.035	0.029	0.168
s.e.														
Decomposition method: RIF-OLS regressions (Fortin, Lemieux and Firpo, 2011)														
Explained	-0.330	-0.067	-0.013	-0.037	-0.106	-0.133	0.036	-0.423	-0.174	-0.064	-0.072	-0.180	-0.314	-0.219
s.e.	0.052	0.011	0.008	0.011	0.015	0.021	0.031	0.028	0.009	0.004	0.005	0.008	0.013	0.026
Unexplained	1.042	0.651	0.310	0.342	0.358	0.336	0.348	0.570	0.336	0.071	0.120	0.250	0.318	0.416
s.e.	0.055	0.015	0.010	0.012	0.016	0.023	0.040	0.030	0.011	0.005	0.006	0.008	0.016	0.038
Total gap	0.712	0.583	0.298	0.305	0.252	0.203	0.384	0.148	0.162	0.007	0.048	0.070	0.004	0.197
s.e.	0.031	0.012	0.008	0.008	0.011	0.014	0.029	0.023	0.009	0.004	0.005	0.007	0.011	0.026
Detailed decomposition:														
Expl: age	0.006	0.010	0.012	0.021	0.027	0.035	0.046	-0.007	-0.008	-0.003	-0.005	-0.003	0.003	0.015
s.e.	0.004	0.002	0.001	0.002	0.002	0.003	0.004	0.002	0.001	0.001	0.001	0.002	0.002	0.003
Expl: edu	-0.033	-0.041	-0.052	-0.099	-0.159	-0.239	-0.242	-0.055	-0.056	-0.047	-0.083	-0.139	-0.206	-0.202
s.e.	0.007	0.002	0.002	0.003	0.004	0.007	0.014	0.007	0.002	0.001	0.002	0.003	0.005	0.011
Expl: focc	-0.437	-0.004	0.044	-0.023	-0.089	-0.092	0.027	-0.436	-0.084	0.043	0.066	-0.007	-0.049	0.033
s.e.	0.064	0.015	0.011	0.014	0.015	0.019	0.046	0.035	0.011	0.005	0.007	0.009	0.014	0.030
Expl: occ	0.166	-0.033	-0.006	0.085	0.136	0.180	0.218	0.100	0.001	-0.038	-0.032	-0.016	-0.050	-0.064
s.e.	0.033	0.009	0.007	0.010	0.013	0.019	0.037	0.019	0.007	0.004	0.005	0.007	0.012	0.027
Unexp: age	0.559	-0.664	-0.306	0.438	0.140	0.229	-0.078	-0.460	-0.668	0.136	0.497	0.412	0.360	-0.035
s.e.	0.307	0.100	0.058	0.052	0.062	0.080	0.179	0.284	0.096	0.033	0.036	0.048	0.078	0.200
Unexp: edu	-0.257	-0.119	-0.090	0.032	-0.086	0.167	0.436	-0.245	-0.281	-0.028	0.035	0.038	0.045	0.094
s.e.	0.057	0.022	0.014	0.015	0.021	0.033	0.086	0.074	0.027	0.010	0.012	0.017	0.030	0.084
Unexp: focc	0.894	0.620	0.456	0.199	0.096	0.265	0.687	0.712	0.338	0.063	-0.090	-0.016	0.257	0.638
s.e.	0.132	0.042	0.029	0.032	0.037	0.050	0.133	0.091	0.034	0.014	0.018	0.024	0.042	0.108
Unexp:occ	-0.677	-0.407	-0.296	0.076	0.237	-0.031	-1.140	-0.281	-0.149	0.024	0.194	0.106	-0.653	-0.859
s.e.	0.123	0.065	0.049	0.044	0.105	0.319	1.644	0.073	0.046	0.020	0.024	0.070	0.224	0.988

Note: s.e. for Melly (2006) decomposition results are computed via bootstrapping procedure and are not available yet (We are in the processing of computing them, they take weeks and the few already computed are all strongly statistically significant at 1%). Source: Author's computations using PNAD 1987-2006.

Table 3b: Racial wage gap: quantile decomposition results, all labour market – 1987 and 2006

	Racial wage gap - year 1987							Racial wage gap - year 2006						
Quantile	**0.01**	**0.1**	**0.25**	**0.5**	**0.75**	**0.9**	**0.99**	**0.01**	**0.1**	**0.25**	**0.5**	**0.75**	**0.9**	**0.99**
Raw log gap	0.511	0.470	0.318	0.463	0.588	0.654	0.608	0.383	0.405	0.231	0.349	0.511	0.629	0.693
Decomposition method: Machado & Mata (2005) - Melly (2006)														
Explained	0.337	0.296	0.299	0.317	0.361	0.442	0.480	0.270	0.269	0.245	0.241	0.297	0.390	0.430
s.e.														
Unexplained	0.073	0.077	0.105	0.150	0.195	0.211	0.176	0.119	0.061	0.067	0.115	0.184	0.211	0.208
s.e.														
Total gap	0.411	0.373	0.404	0.467	0.556	0.653	0.656	0.389	0.331	0.312	0.357	0.481	0.600	0.638
s.e.														
Decomposition method: RIF-OLS regressions (Fortin, Lemieux and Firpo, 2011)														
Explained	0.503	0.384	0.327	0.383	0.477	0.495	0.341	0.461	0.383	0.230	0.300	0.429	0.454	0.376
s.e.	0.029	0.008	0.006	0.006	0.008	0.011	0.022	0.025	0.007	0.003	0.004	0.006	0.008	0.021
Unexplained	0.007	0.018	0.006	0.078	0.130	0.228	0.260	-0.044	0.016	0.019	0.051	0.068	0.242	0.258
s.e.	0.051	0.014	0.007	0.007	0.009	0.012	0.030	0.042	0.011	0.004	0.005	0.006	0.009	0.026
Total gap	0.509	0.402	0.333	0.461	0.607	0.724	0.601	0.417	0.399	0.250	0.351	0.497	0.696	0.634
s.e.	0.031	0.011	0.007	0.008	0.010	0.013	0.031	0.023	0.009	0.004	0.004	0.007	0.011	0.028
Detailed decomposition:														
Expl: age	0.018	0.016	0.016	0.019	0.021	0.018	0.007	0.012	0.011	0.009	0.017	0.028	0.034	0.034
s.e.	0.003	0.002	0.002	0.002	0.002	0.002	0.003	0.002	0.001	0.001	0.001	0.002	0.002	0.003
Expl: edu	0.136	0.111	0.114	0.185	0.313	0.371	0.318	0.120	0.098	0.064	0.115	0.205	0.250	0.221
s.e.	0.015	0.005	0.003	0.004	0.006	0.009	0.019	0.012	0.004	0.002	0.002	0.004	0.005	0.013
Expl: nwocc	-0.020	0.018	0.032	0.017	0.047	0.094	0.098	-0.084	0.002	0.026	0.057	0.080	0.122	0.211
s.e.	0.026	0.007	0.005	0.006	0.007	0.010	0.025	0.021	0.005	0.003	0.003	0.005	0.008	0.027
Expl: occ	0.014	0.038	0.052	0.111	0.106	0.061	-0.011	0.113	0.052	0.025	0.035	0.083	0.066	-0.041
s.e.	0.023	0.007	0.005	0.006	0.007	0.010	0.022	0.022	0.006	0.003	0.003	0.005	0.008	0.022
Unexp: age	0.660	0.257	0.518	0.361	0.505	0.002	-0.693	0.645	-0.562	0.055	0.513	0.636	0.053	-0.541
s.e.	0.299	0.092	0.049	0.048	0.055	0.076	0.204	0.277	0.086	0.034	0.034	0.048	0.074	0.217
Unexp: edu	0.048	0.052	0.068	0.096	0.212	0.075	-0.330	0.175	0.043	0.033	0.183	0.351	0.204	-0.257
s.e.	0.040	0.014	0.008	0.010	0.013	0.022	0.073	0.059	0.019	0.008	0.009	0.014	0.023	0.075
Unexp: nwocc	-1.463	-0.123	-0.290	0.086	-0.391	-0.674	0.319	-1.097	-0.741	-0.202	-0.161	-0.103	-0.478	0.223
s.e.	0.352	0.108	0.063	0.072	0.087	0.130	0.408	0.330	0.097	0.040	0.043	0.066	0.116	0.427
Unexp:occ	0.992	0.006	0.039	-0.211	-0.336	-0.428	7.981	0.665	0.300	-0.001	-0.112	-0.461	0.437	7.046
s.e.	0.231	0.082	0.050	0.060	0.095	0.276	2.130	0.211	0.070	0.032	0.035	0.068	0.201	1.397

Note: s.e. for Melly (2006) decomposition results are computed via bootstrapping procedure and are not available yet (We are in the processing of computing them, they take weeks and the few already computed are all strongly statistically significant at 1%). Source: Author's computations using PNAD 1987-2006.

Table 4a: Gender wage gap: quantile decomposition results, formal sector – 1987 and 2006

	Gender wage gap - year 1987							Gender wage gap - year 2006						
Quantile	**0.01**	**0.1**	**0.25**	**0.5**	**0.75**	**0.9**	**0.99**	**0.01**	**0.1**	**0.25**	**0.5**	**0.75**	**0.9**	**0.99**
Raw log gap	0.511	0.087	0.235	0.247	0.223	0.268	0.427	0.000	0.020	0.070	0.113	0.015	0.041	0.208
Decomposition method: Machado & Mata (2005) - Melly (2006)														
Explained	-0.009	0.031	0.041	-0.002	-0.098	-0.103	0.041	-0.033	-0.048	-0.070	-0.137	-0.241	-0.235	-0.084
s.e.														
Unexplained	0.302	0.141	0.169	0.243	0.332	0.370	0.353	-0.014	0.065	0.135	0.218	0.268	0.272	0.267
s.e.														
Total gap	0.293	0.172	0.210	0.241	0.234	0.268	0.394	-0.047	0.018	0.065	0.082	0.026	0.037	0.183
s.e.														
Decomposition method: RIF-OLS regressions (Fortin, Lemieux and Firpo, 2011)														
Explained	-0.036	-0.007	-0.017	-0.060	-0.152	-0.149	0.034	-0.005	-0.012	-0.027	-0.094	-0.301	-0.456	-0.198
s.e.	0.023	0.008	0.012	0.013	0.019	0.028	0.028	0.009	0.004	0.005	0.006	0.011	0.020	0.030
Unexplained	0.502	0.073	0.255	0.314	0.379	0.375	0.332	-0.007	0.043	0.102	0.210	0.322	0.517	0.411
s.e.	0.044	0.010	0.013	0.014	0.020	0.031	0.047	0.012	0.005	0.006	0.007	0.012	0.024	0.049
Total gap	0.466	0.067	0.238	0.254	0.227	0.227	0.366	-0.012	0.031	0.075	0.116	0.020	0.061	0.213
s.e.	0.039	0.007	0.009	0.011	0.016	0.021	0.041	0.008	0.004	0.004	0.006	0.011	0.016	0.036
Detailed decomposition:														
Expl: age	0.001	0.008	0.020	0.029	0.040	0.051	0.057	-0.002	-0.004	-0.010	-0.016	-0.024	-0.030	-0.018
s.e.	0.002	0.001	0.002	0.003	0.004	0.004	0.006	0.001	0.001	0.001	0.002	0.003	0.004	0.004
Expl: edu	-0.044	-0.042	-0.088	-0.126	-0.217	-0.315	-0.233	-0.021	-0.026	-0.054	-0.092	-0.164	-0.254	-0.177
s.e.	0.006	0.002	0.004	0.004	0.007	0.011	0.019	0.003	0.001	0.002	0.002	0.004	0.007	0.013
Expl: focc	-0.047	0.023	0.017	-0.034	-0.068	-0.042	0.011	0.035	0.045	0.082	0.044	-0.044	-0.046	0.082
s.e.	0.040	0.013	0.015	0.014	0.015	0.020	0.040	0.014	0.005	0.006	0.007	0.010	0.018	0.033
Expl: occ	0.058	0.011	0.040	0.074	0.097	0.159	0.203	-0.015	-0.023	-0.037	-0.021	-0.058	-0.115	-0.079
s.e.	0.025	0.008	0.011	0.012	0.017	0.028	0.037	0.009	0.004	0.005	0.006	0.010	0.019	0.036
Unexp: age	-0.259	0.356	0.950	0.659	0.208	-0.163	-1.174	0.129	0.202	0.561	0.629	0.415	0.603	-0.843
s.e.	0.420	0.069	0.078	0.083	0.104	0.139	0.319	0.094	0.038	0.045	0.055	0.085	0.132	0.348
Unexp: edu	-0.508	0.008	0.127	0.013	-0.087	0.342	0.199	-0.037	0.019	0.091	0.042	-0.122	0.384	-0.195
s.e.	0.136	0.018	0.022	0.026	0.036	0.057	0.138	0.036	0.013	0.015	0.019	0.031	0.051	0.141
Unexp: focc	0.278	0.172	0.091	0.072	0.137	0.295	0.871	0.061	0.043	-0.098	-0.062	0.196	0.361	0.811
s.e.	0.151	0.028	0.033	0.035	0.041	0.058	0.159	0.035	0.014	0.017	0.021	0.031	0.050	0.139
Unexp:occ	-0.325	-0.102	0.054	0.187	0.194	-0.168	1.934	-0.047	0.010	0.119	0.231	0.219	-1.350	-1.124
s.e.	0.172	0.027	0.037	0.049	0.123	0.344	1.786	0.023	0.012	0.018	0.028	0.098	0.264	0.941

Note: s.e. for Melly (2006) decomposition results are computed via bootstrapping procedure and are not available yet (We are in the processing of computing them, they take weeks and the few already computed are all strongly statistically significant at 1%). Source: Author's computations using PNAD 1987-2006.

Table 4b: Racial wage gap: quantile decomposition results, formal sector – 1987 and 2006

	Racial wage gap - year 1987							Racial wage gap - year 2006						
Quantile	**0.01**	**0.1**	**0.25**	**0.5**	**0.75**	**0.9**	**0.99**	**0.01**	**0.1**	**0.25**	**0.5**	**0.75**	**0.9**	**0.99**
Raw log gap	0.288	0.118	0.288	0.377	0.500	0.606	0.519	0.383	0.405	0.231	0.349	0.511	0.629	0.693
Decomposition method: Machado & Mata (2005) - Melly (2006)														
Explained	0.154	0.132	0.169	0.227	0.295	0.391	0.404	0.088	0.096	0.118	0.167	0.248	0.338	0.356
s.e.														
Unexplained	-0.024	0.041	0.100	0.142	0.176	0.189	0.156	-0.047	0.010	0.071	0.139	0.193	0.192	0.196
s.e.														
Total gap	0.129	0.173	0.269	0.369	0.471	0.580	0.560	0.041	0.106	0.188	0.306	0.441	0.530	0.552
s.e.														
Decomposition method: RIF-OLS regressions (Fortin, Lemieux and Firpo, 2011)														
Explained	0.216	0.141	0.241	0.298	0.383	0.432	0.245	0.095	0.109	0.174	0.231	0.375	0.381	0.236
s.e.	0.020	0.006	0.008	0.009	0.011	0.016	0.028	0.008	0.003	0.004	0.005	0.009	0.012	0.021
Unexplained	0.073	-0.052	0.046	0.045	0.117	0.174	0.273	-0.011	-0.032	-0.025	0.069	0.109	0.186	0.233
s.e.	0.042	0.009	0.010	0.010	0.012	0.017	0.042	0.013	0.005	0.005	0.006	0.009	0.014	0.033
Total gap	0.290	0.089	0.286	0.343	0.500	0.606	0.518	0.084	0.077	0.149	0.300	0.484	0.567	0.470
s.e.	0.032	0.007	0.009	0.011	0.014	0.020	0.042	0.008	0.003	0.004	0.006	0.010	0.016	0.034
Detailed decomposition:														
Expl: age	-0.002	-0.003	-0.006	-0.008	-0.010	-0.012	-0.010	0.000	0.000	-0.001	0.001	0.006	0.012	0.010
s.e.	0.001	0.001	0.002	0.003	0.004	0.004	0.004	0.000	0.001	0.001	0.002	0.003	0.004	0.003
Expl: edu	0.083	0.051	0.101	0.162	0.260	0.331	0.259	0.030	0.030	0.060	0.107	0.201	0.235	0.151
s.e.	0.010	0.003	0.004	0.005	0.008	0.011	0.021	0.004	0.002	0.002	0.003	0.005	0.007	0.012
Expl: nwocc	0.020	0.027	0.033	0.032	0.065	0.105	0.044	0.005	0.013	0.024	0.040	0.050	0.122	0.154
s.e.	0.012	0.005	0.007	0.006	0.009	0.013	0.027	0.006	0.003	0.003	0.004	0.006	0.011	0.025
Expl: occ	0.022	0.028	0.065	0.088	0.073	0.022	0.009	0.015	0.018	0.030	0.046	0.101	0.042	-0.045
s.e.	0.014	0.005	0.007	0.007	0.009	0.014	0.024	0.006	0.003	0.003	0.004	0.007	0.011	0.021
Unexp: age	-0.303	0.388	0.567	0.364	0.301	-0.219	-1.016	-0.052	0.383	0.611	0.397	0.358	-0.182	-0.312
s.e.	0.322	0.066	0.076	0.079	0.090	0.133	0.332	0.097	0.039	0.044	0.054	0.078	0.131	0.342
Unexp: edu	-0.124	0.040	0.105	0.108	0.213	0.125	-0.186	0.043	0.078	0.171	0.226	0.513	0.303	-0.376
s.e.	0.069	0.012	0.015	0.018	0.024	0.041	0.122	0.029	0.011	0.012	0.015	0.024	0.043	0.116
Unexp: nwocc	0.203	-0.023	-0.176	-0.042	-0.307	-0.223	0.898	-0.056	-0.058	-0.100	0.259	0.094	-0.670	-0.138
s.e.	0.187	0.058	0.083	0.098	0.124	0.190	0.491	0.080	0.037	0.046	0.058	0.088	0.164	0.476
Unexp:occ	-0.248	-0.046	0.034	-0.087	-0.373	-0.438	3.690	0.008	-0.014	-0.044	-0.332	-0.652	0.449	4.173
s.e.	0.142	0.038	0.056	0.074	0.117	0.309	1.724	0.048	0.023	0.031	0.046	0.090	0.267	1.260

Note: s.e. for Melly (2006) decomposition results are computed via bootstrapping procedure and are not available yet (We are in the processing of computing them, they take weeks and the few already computed are all strongly statistically significant at 1%). Source: Author's computations using PNAD 1987-2006.

Table 5a: Gender wage gap: quantile decomposition results, informal sector – 1987 and 2006

	Gender wage gap - year 1987							Gender wage gap - year 2006						
Quantile	**0.01**	**0.1**	**0.25**	**0.5**	**0.75**	**0.9**	**0.99**	**0.01**	**0.1**	**0.25**	**0.5**	**0.75**	**0.9**	**0.99**
Raw log gap	0.575	0.535	0.442	0.182	-0.041	-0.287	0.113	0.113	0.175	0.077	-0.038	-0.028	-0.016	0.223
Decomposition method: Machado & Mata (2005) - Melly (2006)														
Explained	0.046	0.068	0.046	-0.028	-0.240	-0.469	-0.076	-0.368	-0.169	-0.145	-0.128	-0.118	-0.072	0.060
s.e.														
Unexplained	0.618	0.484	0.350	0.240	0.200	0.182	0.158	0.499	0.299	0.200	0.129	0.092	0.049	0.053
s.e.														
Total gap	0.664	0.552	0.396	0.213	-0.039	-0.287	0.081	0.131	0.131	0.055	0.001	-0.026	-0.023	0.113
s.e.														
Decomposition method: RIF-OLS regressions (Fortin, Lemieux and Firpo, 2011)														
Explained	0.307	-0.021	-0.047	-0.101	-0.256	-0.558	-0.010	-0.002	-0.093	-0.119	-0.105	-0.154	-0.183	-0.129
s.e.	0.111	0.022	0.015	0.016	0.028	0.062	0.122	0.058	0.020	0.013	0.009	0.016	0.025	0.074
Unexplained	0.274	0.555	0.486	0.246	0.272	0.255	0.200	0.131	0.266	0.188	0.069	0.162	0.174	0.386
s.e.	0.125	0.027	0.019	0.018	0.031	0.065	0.134	0.065	0.024	0.016	0.010	0.018	0.028	0.094
Total gap	0.581	0.534	0.439	0.145	0.016	-0.303	0.190	0.129	0.173	0.069	-0.036	0.008	-0.009	0.258
s.e.	0.042	0.017	0.013	0.012	0.022	0.031	0.064	0.031	0.015	0.011	0.006	0.011	0.017	0.058
Detailed decomposition:														
Expl: age	-0.015	-0.012	-0.010	-0.010	-0.013	-0.008	0.039	-0.037	-0.040	-0.029	-0.022	-0.040	-0.048	-0.060
s.e.	0.006	0.003	0.002	0.002	0.004	0.006	0.011	0.006	0.003	0.002	0.001	0.003	0.004	0.010
Expl: edu	-0.094	-0.048	-0.049	-0.072	-0.159	-0.324	-0.665	-0.047	-0.036	-0.039	-0.036	-0.078	-0.128	-0.272
s.e.	0.024	0.006	0.004	0.004	0.008	0.018	0.066	0.010	0.004	0.003	0.002	0.004	0.006	0.028
Expl: focc	0.222	0.023	0.035	0.025	-0.014	-0.028	0.257	0.187	0.073	0.053	0.037	0.045	0.060	0.235
s.e.	0.140	0.033	0.022	0.024	0.039	0.071	0.192	0.061	0.023	0.015	0.011	0.020	0.030	0.100
Expl: occ	0.226	0.042	0.003	-0.001	-0.019	-0.163	0.325	-0.076	-0.063	-0.075	-0.062	-0.055	-0.044	-0.018
s.e.	0.116	0.028	0.018	0.020	0.034	0.072	0.200	0.040	0.016	0.011	0.008	0.014	0.021	0.074
Unexp: age	1.072	-0.025	-0.405	-0.277	-0.266	0.929	0.751	-0.166	-0.525	-0.722	0.044	0.598	0.594	0.886
s.e.	0.336	0.123	0.088	0.074	0.111	0.157	0.410	0.323	0.132	0.083	0.044	0.072	0.110	0.440
Unexp: edu	0.073	-0.073	-0.078	-0.075	-0.293	-0.222	1.299	0.066	-0.163	-0.161	-0.011	0.101	0.219	0.930
s.e.	0.084	0.032	0.024	0.022	0.038	0.064	0.209	0.079	0.037	0.026	0.014	0.024	0.040	0.183
Unexp: focc	-0.140	0.386	0.351	0.542	0.267	-0.041	0.097	-0.283	0.191	0.140	0.054	0.009	0.000	0.868
s.e.	0.234	0.073	0.059	0.060	0.116	0.157	0.437	0.163	0.073	0.055	0.032	0.054	0.082	0.294
Unexp:occ	-0.311	-0.420	-0.264	-0.231	1.079	-0.448	-13.808	0.170	-0.065	0.051	0.074	-0.043	-1.238	-5.350
s.e.	0.226	0.132	0.117	0.112	0.174	1.374	4.197	0.126	0.094	0.101	0.058	0.152	0.453	3.009

Note: s.e. for Melly (2006) decomposition results are computed via bootstrapping procedure and are not available yet (We are in the processing of computing them, they take weeks and the few already computed are all strongly statistically significant at 1%). Source: Author's computations using PNAD 1987-2006.

Table 5b: Racial wage gap: quantile decomposition results, informal sector – 1987 and 2006

	Racial wage gap - year 1987							Racial wage gap - year 2006						
Quantile	**0.01**	**0.1**	**0.25**	**0.5**	**0.75**	**0.9**	**0.99**	**0.01**	**0.1**	**0.25**	**0.5**	**0.75**	**0.9**	**0.99**
Raw log gap	0.200671	0.223144	0.256429	0.287682	0.581862	0.847298	0.798508	0.202941	0.328504	0.287682	0.287682	0.364643	0.523729	0.733969
Decomposition method: Machado & Mata (2005) - Melly (2006)														
Explained	0.165	0.173	0.192	0.242	0.362	0.544	0.546	0.175	0.207	0.202	0.212	0.266	0.341	0.419
s.e.														
Unexplained	0.031	0.029	0.056	0.116	0.198	0.225	0.231	0.076	0.067	0.055	0.067	0.108	0.155	0.217
s.e.														
Total gap	0.195	0.202	0.248	0.359	0.559	0.769	0.776	0.250	0.274	0.257	0.278	0.373	0.496	0.636
s.e.														
Decomposition method: RIF-OLS regressions (Fortin, Lemieux and Firpo, 2011)														
Explained	0.296	0.225	0.254	0.308	0.571	0.580	0.489	0.299	0.294	0.253	0.252	0.296	0.397	0.357
s.e.	0.038	0.011	0.009	0.010	0.016	0.021	0.053	0.030	0.011	0.006	0.006	0.008	0.015	0.038
Unexplained	-0.019	0.007	-0.018	-0.051	0.010	0.248	0.256	-0.093	0.001	0.010	0.006	0.072	0.069	0.352
s.e.	0.077	0.021	0.014	0.011	0.016	0.019	0.057	0.056	0.020	0.011	0.007	0.010	0.015	0.045
Total gap	0.277	0.232	0.236	0.258	0.581	0.828	0.746	0.206	0.295	0.263	0.258	0.368	0.466	0.708
s.e.	0.049	0.017	0.013	0.012	0.019	0.026	0.066	0.032	0.015	0.009	0.008	0.011	0.018	0.053
Detailed decomposition:														
Expl: age	0.028	0.024	0.027	0.033	0.048	0.042	0.041	0.007	0.006	0.005	0.007	0.012	0.022	0.031
s.e.	0.006	0.003	0.003	0.004	0.005	0.005	0.008	0.003	0.003	0.002	0.002	0.003	0.004	0.007
Expl: edu	0.107	0.085	0.086	0.120	0.263	0.367	0.480	0.085	0.059	0.060	0.066	0.102	0.178	0.171
s.e.	0.029	0.008	0.006	0.007	0.012	0.017	0.054	0.017	0.006	0.004	0.003	0.005	0.009	0.025
Expl: nwocc	-0.011	-0.052	-0.048	-0.010	0.004	0.030	0.220	0.007	0.000	0.009	0.046	0.071	0.110	0.178
s.e.	0.038	0.013	0.011	0.011	0.016	0.022	0.095	0.022	0.010	0.006	0.006	0.009	0.020	0.062
Expl: occ	-0.004	0.067	0.095	0.104	0.217	0.151	-0.137	0.019	0.069	0.062	0.040	0.063	0.082	0.021
s.e.	0.035	0.013	0.011	0.011	0.017	0.023	0.082	0.026	0.011	0.006	0.006	0.009	0.019	0.056
Unexp: age	0.417	0.047	0.046	0.804	0.614	-0.448	-1.361	0.290	-0.338	-0.458	0.151	0.134	0.374	-1.027
s.e.	0.372	0.122	0.085	0.068	0.100	0.131	0.420	0.320	0.125	0.069	0.051	0.071	0.118	0.398
Unexp: edu	0.067	0.031	0.010	0.084	0.196	0.204	-0.051	0.142	-0.016	0.010	0.090	0.130	0.327	-0.087
s.e.	0.059	0.021	0.015	0.013	0.022	0.035	0.134	0.074	0.030	0.018	0.014	0.020	0.036	0.125
Unexp: nwocc	0.007	0.434	0.111	0.262	0.069	-0.588	-2.623	-0.128	-0.276	-0.105	-0.289	0.088	-0.194	1.404
s.e.	0.440	0.179	0.144	0.136	0.213	0.320	1.284	0.352	0.173	0.110	0.091	0.145	0.274	1.192
Unexp:occ	0.022	-0.228	-0.128	-0.285	-1.074	-0.662	0.858	0.088	0.031	0.022	-0.012	-0.008	0.096	8.433
s.e.	0.343	0.180	0.164	0.150	0.378	1.074	7.778	0.245	0.129	0.088	0.077	0.156	0.438	4.021

Note: s.e. for Melly (2006) decomposition results are computed via bootstrapping procedure and are not available yet (We are in the processing of computing them, they take weeks and the few already computed are all strongly statistically significant at 1%). Source: Author’s computations using PNAD 1987-2006.

Table 6a: Gender wage gap: quantile decomposition results, self-employed sector – 1987 and 2006

	Gender wage gap - year 1987							Gender wage gap - year 2006						
Quantile	**0.01**	**0.1**	**0.25**	**0.5**	**0.75**	**0.9**	**0.99**	**0.01**	**0.1**	**0.25**	**0.5**	**0.75**	**0.9**	**0.99**
Raw log gap	0.560	0.629	0.560	0.428	0.357	0.329	0.272	0.251	0.267	0.188	0.032	-0.017	-0.036	0.000
Decomposition method: Machado & Mata (2005) - Melly (2006)														
Explained	-0.097	-0.100	-0.097	-0.091	-0.033	0.033	0.154	-0.508	-0.418	-0.327	-0.236	-0.146	-0.052	0.102
s.e.														
Unexplained	0.536	0.709	0.618	0.517	0.388	0.267	0.081	0.750	0.623	0.447	0.262	0.103	0.003	-0.057
s.e.														
Total gap	0.439	0.609	0.521	0.427	0.354	0.300	0.234	0.242	0.205	0.119	0.026	-0.043	-0.049	0.045
s.e.														
Decomposition method: RIF-OLS regressions (Fortin, Lemieux and Firpo, 2011)														
Explained	-0.381	-0.237	-0.283	-0.204	-0.160	0.014	0.015	-0.328	-0.227	-0.158	-0.146	-0.200	-0.231	-0.206
s.e.	0.094	0.035	0.026	0.030	0.037	0.053	0.154	0.035	0.021	0.016	0.014	0.018	0.027	0.057
Unexplained	0.925	0.867	0.792	0.634	0.514	0.324	0.325	0.577	0.426	0.230	0.113	0.175	0.245	0.188
s.e.	0.103	0.041	0.031	0.033	0.040	0.058	0.164	0.046	0.028	0.020	0.017	0.022	0.034	0.088
Total gap	0.543	0.630	0.509	0.430	0.354	0.338	0.340	0.250	0.198	0.072	-0.033	-0.025	0.015	-0.017
s.e.	0.071	0.027	0.021	0.017	0.020	0.029	0.064	0.049	0.022	0.015	0.013	0.015	0.024	0.065
Detailed decomposition:														
Expl: age	-0.019	-0.011	-0.012	-0.008	-0.005	-0.001	0.011	0.000	0.003	0.003	0.005	0.008	0.012	0.019
s.e.	0.009	0.003	0.003	0.002	0.002	0.003	0.007	0.003	0.002	0.002	0.002	0.002	0.003	0.005
Expl: edu	-0.008	-0.012	-0.015	-0.020	-0.028	-0.040	-0.048	-0.046	-0.078	-0.089	-0.100	-0.136	-0.173	-0.155
s.e.	0.004	0.002	0.003	0.004	0.005	0.007	0.011	0.014	0.006	0.004	0.004	0.006	0.008	0.022
Expl: focc	-0.109	-0.137	-0.233	-0.169	-0.120	0.040	-0.163	-0.224	-0.209	-0.135	-0.045	-0.022	0.016	0.220
s.e.	0.160	0.069	0.047	0.050	0.057	0.079	0.255	0.051	0.037	0.027	0.023	0.029	0.041	0.088
Expl: occ	-0.223	-0.061	-0.014	-0.003	0.002	0.019	0.234	-0.090	0.055	0.072	0.007	-0.036	-0.072	-0.285
s.e.	0.085	0.048	0.034	0.033	0.037	0.051	0.139	0.045	0.027	0.019	0.016	0.020	0.029	0.069
Unexp: age	1.635	0.170	0.452	0.372	0.107	-0.100	-0.346	-0.090	-0.243	0.202	0.235	0.374	0.307	0.670
s.e.	0.880	0.297	0.195	0.140	0.149	0.217	0.508	0.639	0.258	0.155	0.117	0.131	0.189	0.534
Unexp: edu	-0.109	-0.048	-0.065	-0.066	-0.089	-0.162	0.066	-0.271	-0.172	-0.160	-0.136	-0.091	-0.086	-0.343
s.e.	0.070	0.028	0.024	0.021	0.028	0.050	0.138	0.109	0.047	0.031	0.025	0.031	0.052	0.167
Unexp: focc	-0.010	1.058	0.964	0.563	0.439	0.135	-0.013	0.544	0.424	0.201	0.086	0.267	0.322	-0.021
s.e.	0.273	0.152	0.109	0.087	0.098	0.136	0.409	0.163	0.098	0.069	0.058	0.073	0.113	0.331
Unexp:occ	0.162	-0.796	-0.654	-0.407	-0.126	1.657	-6.926	-0.398	-0.302	-0.149	0.041	0.092	-1.053	-0.370
s.e.	0.282	0.178	0.152	0.118	0.145	0.763	2.384	0.166	0.088	0.065	0.061	0.147	0.556	2.943

Note: s.e. for Melly (2006) decomposition results are computed via bootstrapping procedure and are not available yet (We are in the processing of computing them, they take weeks and the few already computed are all strongly statistically significant at 1%). Source: Author's computations using PNAD 1987-2006.

Table 6b: Racial wage gap: quantile decomposition results, self-employed sector – 1987 and 2006

	Racial wage gap - year 1987							Racial wage gap - year 2006						
Quantile	**0.01**	**0.1**	**0.25**	**0.5**	**0.75**	**0.9**	**0.99**	**0.01**	**0.1**	**0.25**	**0.5**	**0.75**	**0.9**	**0.99**
Raw log gap	0.629	0.470	0.433	0.446	0.511	0.606	0.616	0.405	0.470	0.377	0.421	0.511	0.639	0.693
Decomposition method: Machado & Mata (2005) - Melly (2006)														
Explained	0.374	0.279	0.254	0.232	0.248	0.305	0.384	0.302	0.254	0.237	0.228	0.252	0.312	0.402
s.e.														
Unexplained	0.239	0.166	0.165	0.196	0.207	0.193	0.111	0.081	0.106	0.121	0.155	0.214	0.254	0.227
s.e.														
Total gap	0.613	0.445	0.419	0.429	0.454	0.498	0.495	0.383	0.360	0.357	0.383	0.466	0.566	0.629
s.e.														
Decomposition method: RIF-OLS regressions (Fortin, Lemieux and Firpo, 2011)														
Explained	0.635	0.399	0.324	0.316	0.373	0.415	0.396	0.381	0.415	0.311	0.350	0.433	0.461	0.427
s.e.	0.072	0.021	0.014	0.013	0.015	0.021	0.050	0.042	0.018	0.010	0.009	0.012	0.017	0.052
Unexplained	-0.003	0.025	0.061	0.133	0.152	0.236	0.262	0.045	-0.035	0.016	0.069	0.104	0.273	0.261
s.e.	0.129	0.034	0.021	0.017	0.018	0.024	0.061	0.077	0.030	0.015	0.013	0.014	0.019	0.063
Total gap	0.632	0.423	0.385	0.450	0.525	0.650	0.658	0.426	0.380	0.327	0.419	0.537	0.734	0.688
s.e.	0.077	0.027	0.019	0.017	0.018	0.025	0.064	0.047	0.021	0.013	0.011	0.015	0.021	0.063
Detailed decomposition:														
Expl: age	0.016	0.011	0.009	0.008	0.008	0.006	0.005	0.022	0.025	0.018	0.021	0.024	0.028	0.054
s.e.	0.007	0.003	0.002	0.002	0.003	0.003	0.005	0.008	0.004	0.002	0.002	0.002	0.003	0.008
Expl: edu	0.158	0.125	0.139	0.177	0.240	0.283	0.308	0.101	0.143	0.117	0.132	0.170	0.177	0.200
s.e.	0.029	0.010	0.007	0.008	0.010	0.016	0.042	0.020	0.009	0.005	0.005	0.007	0.009	0.027
Expl: nwocc	-0.108	0.059	-0.012	-0.037	0.004	0.036	0.046	-0.011	0.016	0.025	0.049	0.097	0.140	0.101
s.e.	0.049	0.016	0.009	0.008	0.010	0.015	0.045	0.035	0.014	0.007	0.007	0.009	0.015	0.053
Expl: occ	0.068	-0.010	0.055	0.097	0.098	0.130	0.136	-0.002	-0.005	-0.004	0.016	0.056	0.085	0.135
s.e.	0.039	0.013	0.009	0.009	0.010	0.015	0.039	0.026	0.012	0.006	0.006	0.008	0.013	0.042
Unexp: age	2.136	0.479	0.177	0.127	0.614	0.575	1.018	1.141	0.444	0.080	0.241	0.087	-0.019	-0.274
s.e.	0.975	0.278	0.169	0.135	0.138	0.188	0.527	0.637	0.248	0.133	0.106	0.126	0.177	0.534
Unexp: edu	0.060	0.022	0.027	0.017	0.088	0.038	0.004	0.039	0.166	0.048	0.091	0.117	-0.031	-0.122
s.e.	0.058	0.022	0.016	0.015	0.018	0.031	0.100	0.084	0.034	0.020	0.017	0.023	0.036	0.124
Unexp: nwocc	-1.778	-0.435	0.263	0.190	-0.268	-0.550	-0.010	-0.606	-0.602	-0.471	-0.398	-0.933	-1.185	1.138
s.e.	0.855	0.333	0.187	0.161	0.177	0.256	0.863	0.670	0.250	0.139	0.113	0.147	0.253	0.933
Unexp:occ	1.416	0.069	-0.343	-0.310	-0.158	-1.837	3.704	0.625	0.527	0.324	0.146	0.180	1.721	7.893
s.e.	0.539	0.268	0.175	0.159	0.182	1.038	6.269	0.431	0.172	0.106	0.094	0.148	0.461	4.409

Note: s.e. for Melly (2006) decomposition results are computed via bootstrapping procedure and are not available yet (We are in the processing of computing them, they take weeks and the few already computed are all strongly statistically significant at 1%). Source: Author's computations using PNAD 1987-2006.

Appendix

Table A1: Pooled quantile regressions considering one selected specification, year 1987

Panel A - Mean and quantile regressions for all pooled sample with FOCC - year 1987

	(1) mean	(2) 0.10	(3) 0.25	(4) 0.50	(5) 0.75	(6) 0.90
male	0.324***	0.323***	0.304***	0.298***	0.309***	0.323***
	(0.006)	(0.010)	(0.008)	(0.007)	(0.009)	(0.010)
white	0.110***	0.104***	0.095***	0.103***	0.116***	0.129***
	(0.005)	(0.008)	(0.006)	(0.005)	(0.007)	(0.007)
age	0.087***	0.076***	0.080***	0.086***	0.092***	0.094***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
agesq	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
edu	0.090***	0.072***	0.078***	0.087***	0.093***	0.099***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
urban	0.198***	0.258***	0.212***	0.172***	0.158***	0.139***
	(0.007)	(0.012)	(0.009)	(0.008)	(0.010)	(0.011)
formal	0.108***	0.334***	0.213***	0.104***	0.013*	-0.078***
	(0.005)	(0.009)	(0.007)	(0.006)	(0.007)	(0.008)
focc	-0.186***	-0.200***	-0.188***	-0.176***	-0.160***	-0.153***
	(0.014)	(0.023)	(0.017)	(0.015)	(0.019)	(0.022)
Constant	1.541***	1.006***	1.385***	1.680***	1.916***	2.159***
	(0.036)	(0.063)	(0.047)	(0.041)	(0.050)	(0.058)
Main region effects	YES	YES	YES	YES	YES	YES
Occ. 2-digit effects	YES	YES	YES	YES	YES	YES
N	97679	97679	97679	97679	97679	97679
r2	0.554	0.3125	0.3084	0.3418	0.3677	0.3788

Note: * $p<0.10$, ** $p<0.05$, *** $p<0.01$; R2 for mean regressions, pseudo-R2 for quantile regressions.

Source: Author's computations using PNAD 1987-2006.

Panel B - Mean and quantile regressions for all pooled sample with NWOCC - year 1987

	(1) mean	(2) 0.10	(3) 0.25	(4) 0.50	(5) 0.75	(6) 0.90
male	0.360***	0.369***	0.338***	0.333***	0.338***	0.351***
	(0.006)	(0.009)	(0.007)	(0.006)	(0.007)	(0.009)
white	0.108***	0.100***	0.094***	0.102***	0.114***	0.125***
	(0.005)	(0.009)	(0.006)	(0.005)	(0.006)	(0.007)
age	0.088***	0.077***	0.081***	0.086***	0.092***	0.095***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
agesq	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
edu	0.089***	0.072***	0.077***	0.085***	0.092***	0.098***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
urban	0.196***	0.259***	0.211***	0.170***	0.156***	0.141***
	(0.007)	(0.012)	(0.008)	(0.008)	(0.009)	(0.011)
formal	0.118***	0.346***	0.223***	0.108***	0.019***	-0.074***
	(0.005)	(0.009)	(0.006)	(0.006)	(0.007)	(0.008)
nwocc	-0.467***	-0.228***	-0.539***	-0.670***	-0.706***	-0.592***
	(0.051)	(0.086)	(0.061)	(0.057)	(0.068)	(0.080)
Constant	1.579***	0.981***	1.427***	1.777***	2.004***	2.246***
	(0.037)	(0.068)	(0.046)	(0.043)	(0.051)	(0.059)
Main region effects	YES	YES	YES	YES	YES	YES
Occ. 2-digit effects	YES	YES	YES	YES	YES	YES
N	97679	97679	97679	97679	97679	97679
r2	0.553	0.3118	0.3081	0.3418	0.3679	0.3788

Note: * $p<0.10$, ** $p<0.05$, *** $p<0.01$; R^2 for mean regressions, pseudo-R^2 for quantile regressions.

Source: Author's computations using PNAD 1987-2006.

Table A2: Pooled quantile regressions considering one selected specification, year 2006

Panel A - Mean and quantile regressions for all pooled sample with FOCC - year 2006

	(1) mean	(2) 0.10	(3) 0.25	(4) 0.50	(5) 0.75	(6) 0.90
male	0.208***	0.196***	0.189***	0.189***	0.187***	0.195***
	(0.005)	(0.006)	(0.004)	(0.004)	(0.005)	(0.008)
white	0.097***	0.069***	0.078***	0.080***	0.103***	0.127***
	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)	(0.007)
age	0.059***	0.053***	0.053***	0.054***	0.058***	0.064***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
agesq	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
edu	0.066***	0.050***	0.052***	0.056***	0.065***	0.076***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
urban	0.072***	0.106***	0.085***	0.060***	0.042***	0.033***
	(0.006)	(0.008)	(0.006)	(0.006)	(0.007)	(0.011)
formal	0.203***	0.501***	0.309***	0.175***	0.070***	-0.042***
	(0.004)	(0.005)	(0.004)	(0.004)	(0.005)	(0.007)
focc	-0.043***	-0.031**	-0.025***	-0.036***	-0.066***	-0.107***
	(0.010)	(0.013)	(0.010)	(0.010)	(0.012)	(0.018)
Constant	-0.175***	-0.819***	-0.331***	0.141***	0.382***	0.488***
	(0.028)	(0.039)	(0.027)	(0.026)	(0.033)	(0.050)
Main region effects	YES	YES	YES	YES	YES	YES
Occ. 2-digit effects	YES	YES	YES	YES	YES	YES
N	148960	148960	148960	148960	148960	148960
r2	0.475	0.3012	0.2570	0.2745	0.3117	0.3351

Note: * p<0.10, ** p<0.05, *** p<0.01; R^2 for mean regressions, pseudo-R^2 for quantile regressions.
Source: Author's computations using PNAD 1987-2006.

Panel B - Mean and quantile regressions for all pooled sample with NWOCC - year 2006

	(1) mean	(2) 0.10	(3) 0.25	(4) 0.50	(5) 0.75	(6) 0.90
male	0.222***	0.208***	0.199***	0.204***	0.209***	0.228***
	(0.004)	(0.006)	(0.004)	(0.004)	(0.005)	(0.007)
white	0.092***	0.064***	0.074***	0.077***	0.097***	0.121***
	(0.004)	(0.005)	(0.004)	(0.003)	(0.004)	(0.007)
age	0.058***	0.053***	0.053***	0.053***	0.057***	0.063***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
agesq	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
edu	0.065***	0.050***	0.051***	0.055***	0.063***	0.074***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
urban	0.077***	0.109***	0.090***	0.061***	0.047***	0.038***
	(0.006)	(0.008)	(0.006)	(0.005)	(0.007)	(0.011)
formal	0.207***	0.502***	0.312***	0.179***	0.076***	-0.033***
	(0.004)	(0.005)	(0.004)	(0.003)	(0.005)	(0.007)
nwocc	-0.818***	-0.473***	-0.646***	-0.834***	-1.083***	-1.152***
	(0.038)	(0.052)	(0.038)	(0.033)	(0.043)	(0.066)
Constant	0.011	-0.744***	-0.180***	0.338***	0.627***	0.743***
	(0.030)	(0.041)	(0.030)	(0.026)	(0.034)	(0.053)
Main region effects	YES	YES	YES	YES	YES	YES
Occ. 2-digit effects	YES	YES	YES	YES	YES	YES
N	148960	148960	148960	148960	148960	148960
r2	0.477	0.3017	0.2582	0.2763	0.3137	0.3368

Note: * p<0.10, ** p<0.05, *** p<0.01; R^2 for mean regressions, pseudo-R^2 for quantile regressions.
Source: Author's computations using PNAD 1987-2006.