Wage Disparities and Occupational Intensity by Gender and Race in Brazil: An Empirical Analysis Using Quantile Decomposition techniques

Paola Salardi

University of Sussex

p.salardi@sussex.ac.uk

October 2012

JOB MARKET PAPER

Preliminary draft: Not to be quoted without permission

Abstract: In this paper, we attempt to provide a comprehensive portrait, over the last two decades, of gender and racial wage gaps across the entire wage distribution and of the impact of gender and racial occupational segregation on wage determination in the context of the Brazilian labour market. Adopting an occupational classification consistent over twenty years, our analysis particularly focuses on the evolution of the impact of female and non-white occupational intensity on wage disparities. We first employ quantile regression analysis in order to investigate the role of female and non-white occupational intensity at different points along the conditional wage distribution. We then apply two different decomposition techniques, proposed by Machado and Mata (2005) and Melly (2006) and by Firpo, Fortin and Lemieux (2009) in order to investigate the determinants of wage disparities at these different points in the wage distribution, and in order to understand how these determinants vary at different wage levels.

Keywords: Brazil, Gender, Race, Occupational Intensity, Wage Gap, Quantile Decomposition Technique.

JEL Classification: J15, J16, J31, J71, O54.

1. Introduction

There have been a range of studies on 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-known Oaxaca (1973) and Blinder (1973) wage decomposition technique (OB decomposition, hereafter). This is a simple and powerful tool that allows the disentangling of the contributions of differences in characteristics (the explained component) and differences in returns to those characteristics (the unexplained component or wage structure effect) to the wage gap to be quantified.

However, this technique also has several limitations that have been documented in the literature. One important drawback is that it focuses only on average effects, and this restricted focus may lead to a misleading or incomplete assessment if the effects of wages covariates vary across the wage distribution. A second limitation is that most of the existing studies do not make a clear connection between occupational segregation and wage discrimination, despite the fact that the two are likely to be closely related.

Thus, this paper has two major goals. 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. This allows us to decompose the determinants of these wage gaps, and their evolution, at each point in the wage distribution. Second, while tracing the pattern of wage differentials across the wage distribution, we focus particularly on the impact of female and non-white occupational intensity on gender and racial wage differentials respectively.

In order to achieve these two goals we apply two relatively new decomposition techniques, the first developed by Machado and Mata (2005) and Melly (2005, 2006) and the second developed by Firpo, Fortin and Lemieux (2009). Both techniques permit the decomposition of wage differentials into the effects of characteristics and the effects of coefficients at different quantiles of the wage distribution. Alongside the application of these techniques we are able to investigate the specific impact of female and non-white occupational intensity on earnings in two ways. We first explore the impact of female and non-white occupational intensity on wage determination at both mean values and at specific quantiles of the wage distribution. Having thus highlighted broad trends we are then able to investigate the role played by these variables within the detailed decomposition at specific wage quantiles that we estimate using the Firpo, Fortin and Lemieux (2009) methodology.

The empirical analysis presented makes two further contributions. First, we look at both gender and racial wage differentials, and discuss similarities and differences between them. Second, we adopt a longer temporal perspective to our analysis than has previously been possible, as the period of interest spans two decades (from 1987 to 2006).

Focusing first on the connections between occupational intensity and wage determination we find significant differences between the patterns by gender and by race, while uncovering novel patterns that do not appear in earlier research. Being employed in female-dominated occupations reduces wages for female workers, particularly in the highest paid jobs, while, by contrast, it has a positive impact on male wages, though only in low-paid jobs. Turning to racial dynamics, being employed in non-white dominated occupations has a negative impact on wages for all workers, though somewhat more among white workers. As with female occupational intensity, this negative impact is most pronounced within better paid occupations. These patterns have remained relatively stable over time, though with the magnitude of the effects actually increasing over time.

Turning to the main findings from decomposing the wage differentials at different quantiles, gender wage differentials tend to exhibit a U-shaped pattern, indicating higher wage differentials at the extremes of the wage distribution, which are primarily driven by wage structure effects. Over time the gender wage gap has declined considerably, owing primarily to a decline in these unexplained components. However, this decline has occurred primarily at the bottom of the wage distribution, while unexplained gender wage gaps have been more persistent at higher quantiles. Racial 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. This pattern does not appear to have changed over time. This suggests the existence of *sticky floors* and *glass ceilings* phenomenon for women and the existence of *glass ceilings* for non-white workers.

The RIF-OLS technique developed by Firpo, Fortin and Lemieux (2009) offers additional insights into the role of individual variables in accounting for pay gaps. For both groups we find that education is the primary contributor to differences in endowments, which favour women and white workers, and that this is particularly so at the top of the wage distribution. We further find that experience, as proxied by age, is more rewarded among male and white workers, and is thus an important unexplained contributor to observed wage gaps. Finally, we find divergent impacts of occupational structure on pay gaps. Within female dominated occupations women are paid significantly less than men, as noted earlier. By contrast, we find that non-white workers are comparatively better paid than white workers in non-white dominated occupations. However, we also find that white wages are significantly higher owing to the overall concentration of white workers in better paid professions, as non-white dominated occupations are, on average, significantly less well paid.

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 on this topic. Section 3 presents the data and provides an overview of gender and racial wage differentials at different points in the wage distribution. Section 4 discusses the identification strategy and then outlines the two quantile decomposition techniques to be employed. Section 5 presents our findings and section 6 offers some concluding remarks.

2. Literature review

After the publication of seminal studies by Oaxaca (1973) and Blinder (1973), the growth of research on wage gaps in developed and developing countries, both by gender and race (or ethnicity), has been prolific. A significant number of these studies have gone beyond applying the core methodology by also enhancing it in several respects. Several papers have sought to directly address the 'index number' problem (Cotton, 1988; Neumark, 1988; Oaxaca and Ransom, 1994). 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 issue using a multinomial logit model.

Another important set of studies extends the OB decomposition technique by accounting for occupational structure. The seminal work by Brown, Moon and Zoloth (1980) introduced a modified version of the OB decomposition where the occupational attachment model is estimated using a multinomial logit, while Miller (1987) proposes estimation by ordered probit model. Reilly (1991) introduced a selection bias correction in conjunction 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, Hoddinott and Krishnan, 1999). A strand of this literature has aimed at accounting for occupational segregation by investigating the 'degree of feminization', or in other words, the shares of females within each occupation. These include studies by Johnson and Solon (1986), Macpherson and Hirsch (1995) and Cotter, Hermsen and Vanneman (2003), which have investigated the role of feminization for the U.S. labour market; Lucifora and Reilly (1992) for the Italian labour market, and Baker and Fortin (2003) for Canada and the U.S. None of these studies have considered potentially similar dynamics when looking at the shares of non-white workers (or any other disadvantaged minorities).

Other studies have explored inter-industry wage differentials (see, among others, Krueger and Summers, 1988; Fields and Wolff, 1995; Haisken De New & Schmidt, 1997; Horrace and Oaxaca, 2001). Several recent studies have proposed strategies for the analysis of wage differentials by exploiting employer-employee matching data, in order to address the fact that the OB decomposition approach suffers from the absence of a direct measure of individual productivity (see, for example, Hellerstein, Neumark and Troske, 2002; Bayard et al, 2003, Hellerstein and Neumark, 2006; Hellerstein and Neumark, 2007). Finally, the OB decomposition has been extended to the decomposition of changes over time, as explained by Smith and Welch (1986) and subsequently Juhn, Murphy and Pierce (1991, 1993). [1] They have offered an extension that facilitates the decomposition of pay gaps between two points in time.

While these studies have tackled different limitations of the original OB decomposition method, they all rely on the estimation of wage gaps at the mean. Going beyond the mean, by focusing on more general counterfactual wage distributions, has been the subject of several studies in recent years (see Fortin, Lemieux and Firpo, 2011). Methodologies in this tradition include 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) and methods based on parametric quantile estimation (such as Gosling, Maching and Meghir (2000) and Machado and Mata (2005)). Melly (2005, 2006) has proposed a conditional[2] quantile decomposition approach that is very similar to that of Machado and Mata (2005), while Chernozhukov, Fernandez-Val and Melly (2009) cover the modelling and estimation of a wide range of

[1] The Juhn, Murphy and Pierce (1991) methodology has been subject to several criticisms, summarized by Yun (2007). Most notably, in using their decomposition methodology the residual component (i.e., unobservable prices and quantities) accounts for most of the growth in overall wage inequality. More recent literature has, by contrast, revealed a smaller role for residuals in explaining changes in wage distribution. For further discussion, see also Card and Di Nardo (2002) and Lemieux (2006).

[2] The use of the terminology 'conditional' and 'unconditional' quantile decomposition warrants a precise definition. The 'unconditional' quantile distribution is the distribution of a certain outcome Y at specific quantiles. The 'conditional' quantile distribution is the distribution of a certain outcome Y at specific quantiles conditional on a set of covariates X.

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 types of techniques in order to move beyond estimation based on mean values: the conditional quantile regression approach, as proposed by Machado and Mata (2005) and subsequently by Melly (2005, 2006), and the RIF-regression method suggested by Firpo, Fortin and Lemieux (2009). We argue that employing these techniques in the context of the Brazilian labour market can provide deeper insights into the nature of wage differentials.

In analyzing gender and racial wage gaps in Brazil, this study builds on a large number of existing. Some studies have accounted for occupational segregation while estimating wage differentials, following the Brown, Moon and Zoloth (1980) reformulation of the OB decomposition (see Ometto, Hoffmann and Alves, 1999; Arcand and D'Hombres, 2004, Salardi, 2012). Several other studies have addressed the selection bias problem, including Stecler et al (1992), Loureiro, Carneiro and Sachsida (2004) and Carvalho, Neri and Silva (2006). Further studies have linked the study of wage gaps to questions of labour market informality by estimating wage gaps while distinguishing between the formal and non-formal labour markets (Birdsall and Behrman, 1991; Tiefenthaler, 1992; Silva and Kassouf, 2000). This includes an effort by Carneiro and Henley (2001) to explore wage differentials between the formal and informal sectors while controlling for selection bias, as well as recent studies by Cacciamali and Hirata (2005) and Cacciamali, Tatei and Rosalino (2009).

However, few studies have investigated wage gaps for Brazil using quantile regression estimation. Santos and Ribeiro (2006) explore gender wage gaps using the Machado and Mata (2005) decomposition technique, but restrict the analysis to only a single year (1999). They report the presence of more severe differentials at the extremes of the wage distribution, which are driven primarily by unobserved factors. Madalozzo and Martins (2007) find a similarly non-linear pattern when employing a gender dummy in pooled quantile regressions.

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 of both gender and racial wage gaps over time across the entire wage distribution. Second, it looks at the evolution of gender and racial wage gaps over a longer time period than previously possible. Finally, it links the analysis of wage discrimination to the issue of occupational segregation by estimating the impact of female and non-white occupational intensity on wage differentials.

3. Data and overview of wage gaps

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. We consider a sample 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 codes. The dataset has a large sample size that varies from a labour force of roughly 98,000 observations in the first year (1987) to roughly 150,000 in the final year (2006).

The analysis presented here is, again, crucially dependent on the use of a harmonized and consistent over time occupational classification, which makes it possible to strengthen the analysis in several respects. The primary advantage of this dataset is the availability of information on earnings and comparable occupations over a protracted period of time (two decades). The information related to earnings is provided consistently within the original dataset and we compute the log of hourly earnings using data from the primary occupation. Dealing with occupational codes is more complex, as the raw PNAD dataset employs occupational classifications that vary across years and which, for the majority of years, are not directly comparable with the international classification provided by the ILO, the ISCO-08. We address this consistency problem by employing a new harmonized occupational classification developed in Salardi (2012). This classification is harmonized and consistent over the two decades of interest (from 1987 to 2006) and consists of 83 different occupational categories at the 3-digit level.

Harmonizing the occupational classifications over time allows us to construct two variables of interest: female occupational intensity (focc3) and the non-white occupational intensity (nwocc3). These variables capture the proportion of female (or non-white) workers in each occupation. We compute these values at a 3-digit level of occupational classification, which includes 83 different occupational codes. These two variables reflect the degree of *femaleness* (or feminization) or *non-whiteness* of each three-digit occupational group.

The primary drawback of using this dataset over such a prolonged period of time is that it restricts the nature of other information that is available for all years. For example, the variable for work experience is commonly employed in the specification of wage equations, but it is not present in the earlier years of the PNAD dataset. For this reason, we employ an austere wage equation specification, which has nonetheless proven to have high explanatory power (see Salardi, 2012).

Having reviewed the main features of the data employed in this paper, we now report some preliminary descriptive analysis. Figure 1 illustrates the distribution of wage gaps across the wage distribution by both gender and race (the plots to the left side are for the first year, 1987, and the plots to the right side are for the last year, 2006). We can clearly see that wage differentials by gender are considerably greater at the bottom end of the wage distribution and, interestingly, are widening at the top end in more recent years. By contrast, racial wage differentials widen as we move toward the top of the wage distribution. These preliminary descriptive figures appear to provide preliminary evidence of the existence of a dual phenomenon of *glass ceilings* for women and *sticky floors* for non-white workers.[3]

Figure 1 further highlights a sizeable decline in gender wage gaps over time across the wage distribution, with the average value moving from 0.322 in 1987 to 0.05 in 2006 (as indicated by the horizontal red lines). In the case of racial pay gaps the patterns remain fairly stable over time, with the average value moving from 0.489 in 1987 to 0.413 in 2006.

[Figure 1 about here]

Figure 2 provides a more general portrait of both gender and racial wage gaps, presenting data at selected points of the wage distribution (0.1, 0.25, 0.5, 0.75 and 0.9), for five years spanning the entire period (1987, 1992, 1997, 2002 and 2006). These plots reaffirm the key findings from figure 1. First, we again see that gender wage gaps are wider at the bottom of the wage distribution, while racial wage gaps tend to increase with progression up the wage distribution. Second, over time, both gender and racial differentials

[3] The concepts of 'glass ceilings' and 'sticky floors' are, in fact, closely related. Glass ceilings are invisible but concrete barriers that prevent career advancement and restrict minorities from reaching the best paying and most prestigious occupations, despite their characteristics. Sticky floors refer to women and minorities being trapped in low-paid, low-mobility jobs (Booth, Francesconi and Frank, 2003; De La Rica, Dolado and Llorens, 2005; Kee, 2006; Chi and Li, 2008).

have consistently decreased, however the contraction is considerably more pronounced for gender wage gaps (particularly those at the lower quantiles).

Given that our subsequent analysis explores the relationship between a variety of covariates and wage differentials at different points of the wage distribution, it is useful to look briefly at summary statistics for the key covariates. In order to conserve space we do not present tables of the means and standard deviations for all of the covariates across all 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 as we move to the higher wage quantiles. Age and years of education increase as we progress to higher quantiles, consistent with a positive relationship between human capital endowments and earnings. There are less workers living in urban areas within lower wage quantiles, confirming that rural workers have, on average, lower wages. Individuals working in the agricultural sector are more numerous at the bottom end of the wage distribution, together with those working in the personal and restaurant services sector. Examining the concentration of different occupations within different quantiles confirms that higher skilled jobs are better paid. When we look at the distribution of informality across wage quantiles, we find that although the formal sector represents roughly 45-46% of total employment over time, only 0.05% in 1987 and 0.008% in 2006 of formal workers are in the bottom 10% of the overall wage distribution.

[Figure 2 about here]

Since the relationship between wage differentials and female and non-white occupational intensity is of special interest, we now describe patterns related to occupational intensity in greater detail. Our variable for female occupational intensity moves from an average of 37% in 1987 to 44% in 2006 and it is fairly homogenously distributed over wage quantiles, although it is slightly higher at the bottom end of the wage distribution in earlier years. By contrast, non-white occupational intensity moves from 47% in 1987 to 53% in 2006, but in all years consistently decreases as we move toward the top quantiles. Overall, this implies that female dominated occupations are located comparatively homogenously across the wage distribution, while non-white dominated occupations are characterized by relatively low earnings.[4]

Figure 3 provides additional insights into how female and non-white occupational intensity vary across wage quantiles. The values of the female and non-white occupational intensity variables at different quantiles of the wage distribution are derived using a variation of the Machado and Mata (2005) approach, which is explained in detail in the methodological section. In simplified form, it consists of taking the mean of the observations drawn at random with replacement at different quantiles from each population sub-sample. In 1987 female occupational intensity is noticeably greater at the bottom end of the wage distribution. However, over time this pattern largely disappears, as in 2006 there is no clear pattern, with female occupational intensity noticeably lower between the 60^{th} and the 80^{th} percentiles, before increasing again at the top of the wage distribution. Meanwhile, we again see that the pattern for non-white occupational intensity is more homogeneous and stable over time. From panel B of figure 3, we observe that the degree of non-whiteness steadily decreases as we move to the top of the wage distribution.

Figure 4 plots average wages by gender and race at different levels of female and non-white occupational intensity. Looking first at gender, we see no obvious trend in the relationship between the two variables, as female-dominated occupations are neither better nor worse paid than male-dominated professions, although males earn more, on average, than females, independent of the degree of femaleness within occupations. The pattern by race is very different, as wages consistently decline as non-white occupational intensity increases, while, as with the case of gender, white workers consistently earn higher wages within occupations, independent of non-whiteness.

[Figure 3 about here]

To conclude this section it is useful to briefly summarise some key insights from this preliminary descriptive analysis. Gender differentials are more pronounced at the extremes of the wage distribution and are particularly wide within low-paid occupations. By contrast, racial wage gaps widen as we move to the top end of the wage distribution. Women appear

[4] Both female and non-white occupational intensity have, on average, increased over time (by 7 and 6 percentage points, respectively). However, female occupational intensity has increased more homogenously across occupations than non-white occupational intensity. These patterns are consistent with the findings about occupational segregation presented in Salardi (2012), where we found a sizeable decline in gender segregation but only a small contraction in racial segregation.

to be homogenously distributed across occupations, while non-white individuals appear to be concentrated in low-paid and low-skilled occupations. Thus, although woman are employed relatively homogenously across the wage distribution, they appear to suffer from more sizeable wage gaps within low paid occupations and, to a somewhat lesser extent, in the top paid jobs. Meanwhile, non-white workers tend to work in low-paid and low-skilled occupations, while wage gaps are most pronounced within occupations with higher earnings and a correspondingly lower presence of non-white workers. These figures are consistent with existence of both *sticky floors* and *glass ceilings* for female workers and *glass ceilings* for non-white workers. In the subsequent sections we explore these patterns in more detail by decomposing these gender and wage gaps over the entire wage distribution.

[Figure 4 about here]

4. Empirical methodology

This section outlines the quantile decomposition techniques to be employed, and proceeds in three parts. First, we discuss the identification strategy and the definition of the parameters of interest. We then explain the conditional quantile decomposition techniques developed separately by Machado and Mata (2005) and by Melly (2006). Finally, we present the RIF-regression method proposed by Firpo, Fortin and Lemieux (2009).

4.1 Identification strategy

Our analysis is ultimately aimed at answering a counterfactual question: ‘How much would female (non-white) workers be paid if they were rewarded according to the wage structure for male (white) workers?’ We are thus seeking to compare observed wage structures with counterfactuals, which capture alternative potential wage structures. As such, the problem of the wage structure effect can be interpreted as a treatment effect and ultimately linked to the programme evaluation literature, as recently explained in Fortin, Lemieux and Firpo (2011).[5]

[5] In this section we first re-state the identification strategy in terms of the programme treatment framework for mean pay gaps and then for the quantile framework, which is the primary subject of empirical investigation in this study.

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

$$ATE = E[W_B] - E[W_A] \tag{1}$$

The overall average treatment effect (ATE) is simply the difference between average wages if everybody where paid accordingly to the wage structure of group A and average wages if everybody where paid according to the wage structure of group B. Thus, we know that moving from group A to group B is interpreted to be "the treatment".

Now in reality we simply observe the actual average wages for group B and A defined as $E[W_B|D_B = 1]$ and $E[W_A|D_A = 1]$ respectively. We need now to link the observed average wage differential to the average treatment effect. The introduction of the counterfactual enables us to do so and ultimately to compute the average treatment effects of the treated (ATT). The counterfactual, $E[W_A|D_B = 1]$, represents the average wages if group B workers were paid according to the wage structure of group A. Thus, by adding and subtracting the counterfactual, we obtain:

$$\begin{aligned} E[W_B] - E[W_A] = \; & \{E[W_B|D_B = 1] - E[W_A|D_B = 1]\} + \\ & \{E[W_A|D_B = 1] - E[W_A|D_A = 1]\} \end{aligned} \tag{2}$$

The first bracketed term on the right-hand side of equation (2) represents differences in the returns to observable characteristics, or differences in coefficients (i.e., the wage structure component), while the second bracketed term represents differences in observable characteristics.

From equation (2) the link between the programme evaluation literature and wage decomposition methodologies becomes clear. Wage decomposition methodologies are designed to investigate the extent to which wage differentials originate from differences in structure and differences in observed characteristics. The first bracketed component on the right-hand side represents the wage structure component for the wage decomposition methodology literature and identifies the average treatment effects of the treated (ATT) in the context of the programme evaluation literature. That is:

$$ATT = E[W_B|D_B = 1] - E[W_A|D_B = 1] \qquad (3)$$

which is the difference between the observed average wages of group B, $E[W_B|D_B = 1]$, and the hypothetical wages that workers belonging to group B would have been paid if they belonged to group A, $E[W_A|D_B = 1]$ (i.e., the counterfactual).

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

$$\begin{aligned} E[W_B] - E[W_A] = {} & \{E[W_B|D_B = 1] - E[W_B|D_A = 1]\} + \\ & \{E[W_B|D_A = 1] - E[W_A|D_A = 1]\} \end{aligned} \qquad (4)$$

Now, the second bracketed term identifies the average treatment effect of the non-treated (ATNT), or, more intuitively, the difference between the hypothetical wages 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:

$$ATNT = E[W_B|D_A = 1] - E[W_A|D_A = 1] \qquad (5)$$

The average treatment effect of the non-treated (ATNT) is of particular importance because of the nature of the research questions investigated in this study. With respect to gender (racial) disparities, we have defined our research questions as follows: "what if female (non-white) workers were paid according to the male (white) wage structure". Thus, the wage structure effect for our purposes is provided by the average effect of the non-treated (ATNT).

Now we can extend this approach beyond the mean level by considering the quantile treatment effects. The overall θ^{th} quantile treatment effect (QTE) is:

$$F_{W_B}^{-1}(\theta) - F_{W_A}^{-1}(\theta) \tag{6}$$

where $F_{W_A}^{-1}(\theta)$ is the θ^{th} quantile of the wage distribution W_A. It is important here to note that $F_{W_A}(\theta)$ represents the wage cumulative distribution function for group A at the θ^{th} quantile; thus, its inverse, $F_{W_A}^{-1}(\theta)$, represents the quantile function.

We now need to introduce the counterfactual at quantile level, which will be equal to:

$$Q_\theta^C = F_{W_B}^{-1}(\theta | D_A = 1) = X_{A,i}{}'\beta_{B,\theta} \tag{7}$$

The quantile counterfactual, $F_{W_B}^{-1}(\theta | D_A = 1)$, represents the hypothetical quantile wage distribution that group B workers would have been paid if they belonged to group A at the θ^{th} quantile. As already observed for the mean values, by adding and subtracting the counterfactual to the quantile treatment effect (QTE), we can then isolate the θ^{th} quantile treatment effect on the treated (QTET) as follows:

$$F_{W_B}^{-1}(\theta | D_B = 1) - F_{W_A}^{-1}(\theta | D_B = 1) \tag{8}$$

And, correspondingly, the θ^{th} quantile treatment effect on the non-treated (QTENT) is:

$$F_{W_B}^{-1}(\theta | D_A = 1) - F_{W_A}^{-1}(\theta | D_A = 1) \tag{9}$$

Finally, it is important to note that what we identify and then estimate is the difference between the quantiles and not the quantile of the difference.

We conclude this section with few remarks important for both mean and quantile approaches. It is important to stress 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 bona fide pre-treatment variables.

Nonetheless, the identification of the contribution of different factors to observed wage differentials may remain useful in conducting tests for specific hypotheses, identifying important mechanisms or providing meaningful explanations for the unequal treatment phenomenon.

There are a variety of empirical methodologies that can be applied to compute the counterfactual of interest. The next two sub-sections provide an overview of the two approaches employed in this paper: the conditional quantile regression methodology proposed by Machado and Mata (2005) and further developed by Melly (2006) and the RIF-OLS regression method developed by Firpo, Fortin and Lemieux (2009).

4.2. Estimation of counterfactual distributions using quantile regression

In order to estimate the average treatment effect using the quantile regression methodology, we need to estimate the counterfactual quantile, $Q_{\theta}^{C} = X_{A,i}{}'\beta_{B,\theta}$. Machado and Mata (2005) and Melly (2005, 2006) propose two different but similar methodologies for computing the counterfactual quantile. Machado and Mata (2005) provide a simulation-based estimator where the counterfactual unconditional wage distribution is constructed from the generation of a random sample. Melly (2005, 2006) instead proposes estimating the unconditional distribution by integrating the conditional distribution over a range of covariates. In this section we will explain both methodologies in detail, but we begin by reviewing the basics of the quantile regression estimations.

Ultimately, both methods are based on the estimation of the conditional distribution by quantile regression. In adopting the quantile regression framework, the impacts of observable characteristics on the conditional wage distribution can be estimated (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:

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

where W is the dependent variable denoting log hourly 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 for the different θ^{th} quantiles. 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 for each specific percentile of $\theta \in (0,1)$.

The conditional quantile function for group B would be:

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

while for group A:

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

The next step is to construct the counterfactual unconditional wage distribution, $Q_\theta^C = X_{A,i}'\beta_{B,\theta}$, using estimates from the conditional quantile regressions. However this phase is complicated by the fact that 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 θ^{th} 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 say that the expectation of the conditional median does not produce the median of the marginal distribution.

In addressing this problem, Machado and Mata (2005) estimate the counterfactual unconditional wage distribution using a simulation-based technique. This technique consists of several steps:

1) generate a random sample of size m from a uniform distribution U[0,1] (invoking the probability integral transformation theorem);

2) for each group, estimate m different quantile regression coefficients, $\hat{\beta}_{A,\theta}$ and $\hat{\beta}_{B,\theta}$ respectively for group A and group B;

3) generate a random sample of size m with replacement from the empirical distribution of the covariates for each group, namely $X_{A,i}$ and $X_{B,i}$;

4) generate the counterfactual of interest by multiplying different combinations of quantile coefficients and distribution of observables between group A and group B after repeating this last step m times.

Standard errors for the estimated quantiles of the counterfactual distribution are computed using a bootstrapping technique proposed by Machado and Mata (2005). The alternative is to calculate analytical asymptotic standard errors as proposed by Albrecht, van Vuuren and Vroman (2009).

An alternative and simplified version of the Machado and Mata (2005) has been adopted in several applied studies. This method consists of estimating the quantile coefficients, $\hat{\beta}_{B,\theta}$, for a grid of values of θ and drawing random samples only for the covariates $X_{A,i}$ from the empirical distribution. Albrecht, Bjorklund and Vroman (2003) were the first to adopt this alternative version and it has subsequently been adopted by Autor, Katz, and Kearney (2005), Melly (2006) and Pham and Reilly (2007). With this simplified version, 100 observations are randomly drawn with replacement from each of the group A and group B sub-samples. Then each observation is ranked, thus representing a percentile point θ of the wage distribution. In this way, the full set of characteristics $X_{A,i}$ is retrieved. This process is replicated m times in order to obtain a sample of size m at each θ^{th} quantile. The mean characteristics of these observations at each quantile are used as realizations to construct the counterfactual. For the sake of completeness and comparison, we implement both the simplified and original versions of the Machado and Mata (2005) technique.

Because the conditional quantile function is not necessarily monotonic it might not be possible to invert it. In order to overcome this problem, Melly (2005, 2006) proposes integrating the entire conditional distribution function by integrating over the full set of covariates. 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.

From this starting point, we first we estimate the entire conditional distribution by quantile regression. We can then 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:

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

The standard errors can be obtained by bootstrapping the results. However, the bootstrapping technique is computationally demanding and time consuming and, as such, when datasets are very large this process can become an almost insurmountable exercise. For this reason, Melly (2005) constructs an analytical estimator of the asymptotic variance using the asymptotic results for the parametric estimator.[6]

Once the key counterfactual, $Q_{\theta}^{C} = X_{A,i}'\beta_{B,\theta}$, is estimated using either of these quantile techniques, we can perform the decomposition of wage gaps of the unconditional quantile function between groups B and A denoted as:

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

The first bracketed term represents the effect of characteristics (or the quantile endowment effects) and the second the effect of coefficients (or the quantile treatment effects). Note that the residual component asymptotically disappears, whereas it is still present when we implement the decomposition of the unconditional quantile wage gap using the Machado and Mata (2005) method as implemented by Albrecht, Bjorklund and Vroman (2003).[7]

Ultimately, 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

---

[6] The Stata command 'rqdeco' by Melly (2006) currently provides only for bootstrapping standard errors. The computation of these standard errors is very time-consuming: for example, estimating standard errors for the explained and unexplained components, as well as the total gap, for one quantile can take a week for a sample size of roughly 150,000 observations.

[7] In the case of the Machado and Mata (2005) technique as implemented by Albrecht, Bjorklund and Vroman (2003) we will report the conditional quantile wage gap and the unconditional quantile wage gap (or predicted gap) where the unconditional wage gap is the sum of the conditional wage gap and the residual.

decomposition using simulation methods, but with the drawback 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 provides a more efficient option. Finally, it is important to highlight that the Melly (2006) method assumes exogeneity for all covariates.

4.3. Estimation of counterfactual distributions using RIF-regression

An important limitation of the Machado and Mata (2005) and Melly (2006) decomposition techniques is that they do not allow for computing detailed decompositions that allow the computation of the effect of each covariate on the unconditional quantile wage distribution. Chernozhukov, Fernandez-Val and Melly (2009) discuss a variety of methods based on conditional distributions that attempt to address this limitation, while we focus here on an alternative method recently proposed by Firpo, Lemieux and Fortin (2009).

This method estimates the impact of changes in the distribution of covariates, X, on the quantiles of the unconditional distribution of an outcome variable. It consists of running a simple regression where the outcome variable is replaced by a transformed version, the (recentered) influence function (RIF). Although it can be applied to any distributional statistic of interest for which it is possible to compute an influence function, here we focus on the difference between 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:

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

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:

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

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 have 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 the transformed outcome variable is available for each observation and it can be used to estimate a simple OLS regression on a vector of covariates.[8] 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 are then used to compute the equivalent of the OB decomposition for each quantile as follows:

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

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. It is worth noting at this stage that while we have focused here on how to compute the RIF function within an OLS framework, Firpo, Fortin and Lemieux (2009) provide two alternative ways to estimate the unconditional quantile partial effect.[9] The RIF-logit estimates the marginal effect from a logit model while the RIF-NP is based on a non-parametric estimator.

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 for distributional functions that differ from the mean cannot be employed to get a detailed decomposition. An example is represented by the Di Nardo, Fortin and Lemieux (1996) technique where the individual contribution of the binary variables, among the entire set of characteristics, is estimated through a reweighted procedure.

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

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

[9] The unconditional quantile partial effect (UQPE) correspond to the following formula: $E[\frac{dE[RIF(W,Q_\theta)|X]}{dx}]$.

approximation of the composition effects and this approximation is not precise and may produce approximation errors. This issue is tackled further in Heywood and Parent (2009). A second limitation is that, at least for now, this methodology is based around the estimation of unconditional quantile regressions in the presence of exogenous covariates and does not consider the possible presence of endogeneity (Firpo, Fortin and Lemieux, 2009).

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 into their corresponding quantiles. In this sense, the 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).[10]

4.4 Selectivity issues

We have presented different methods to estimate quantile counterfactuals, though both are based on the assumption of exogenous covariates. In reality, the exogeneity assumption may fail in some cases, in which case the results could be biased by self-selection or more general endogeneity problems. Following Fortin, Lemieux and Firpo (2011), we can consider three different cases: 1) different self-selection processes within group A and group B; 2) self-selection into group A and group B; and 3) general endogeneity problems with respect to the covariates.

The first case is possible in our application as it is straightforward to imagine that women and men may have different decision processes that bring them into the labour market, while the same is potentially true of different racial groups as well. In this case the unconfoundness (or ignorability) assumption does not hold, and the decomposition terms are not identified correctly. Machado (2011) invokes three different self-selection cases

[10] Many approaches, such as the Machado and Mata (2005), Albrecht, Bjorklund and Vroman (2003) and Melly (2005), have proposed estimating and integrating the entire conditional distribution over a set of covariates to obtain the counterfactual unconditional distribution. The Firpo, Fortin and Lemieux (2009) methodology estimates the conditional distribution only at one point of the distribution at a time.

(selection based on observables, selection based on unobservables and bounds) and investigates possible solutions for each. 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 our binary categories. Finally, the third case refers to general endogeneity, which occurs 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 options for correcting our empirical analysis in order to permit a robust identification of the decomposition components, is thus potentially warranted, but is beyond the scope of this particular thesis. A few recent studies have attempted to account for sample selection when implementing quantile decomposition techniques (Albrecht, van Vuuren and Vroman, 2009; Nicodemo, 2009; Chzhen and Mumford, 2011; Chzhen, Mumford and Nicodemo, 2012). These studies have generally applied a semi-parametric adaptation of the Heckman parametric procedure for quantile wage regressions, as proposed by Buckinsky (1998). In particular, Albrecht, van Vuuren and Vroman (2009) first proposed an extension of the Machado and Mata (2005) technique which employs the semi-parametric Buckinsky (1998) method where a power series approximation for the selection term is estimated using the single-index model as proposed by Ichimura (1993).

However, any selection correction within a quantile framework suffers from significant challenges, together with the general issue of the validity of the instruments. These include the choice of the appropriate estimation method for the first stage (i.e., probit model versus non-parametric single index model) and the problem of the identification of the intercept of the wage equation, due to its conflation with the constant term associated with the power series approximation of the selection term (Andrews and Schafgans,1998; Buckinsky, 1998). While selection correction within decomposition techniques is acknowledged to be problematic to begin with, its application within a quantile framework is thus even more complex. At the same time, we tend to be confident of our uncorrected findings at the quantile level given that the mean decomposition results proved to be robust in surviving the selection correction process relatively unchanged. Ultimately, we thus focus this paper on estimating pay gaps at different quantiles of the wage distribution through the application of multiple techniques while leaving the selection correction within quantile decomposition techniques to further research, given that the analysis in this paper is already dense.

5. Empirical findings

Having outlined the relevant methodologies, we now present the results in three stages. First, we present a set of regressions, estimated at different quantiles of the wage distribution, for the pooled samples for the first and the last years of the data. In estimating pooled regression models we are assuming that women and men, and non-white and white workers, receive the same returns to their characteristics. We then divide the samples and estimate quantile regressions by gender and by race separately. As noted earlier, while presenting quantile regression estimates we pay particular attention to the impact of female and non-white occupational intensity on wage differentials at different points of the wage distribution.

With these regression estimates, we then implement the two different quantile decomposition techniques: the Machado and Mata (2005) and Melly (2006) quantile decomposition techniques and the RIF-OLS method developed by Firpo, Fortin and Lemieux (2009). These quantile decomposition techniques allow us 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 to differences in returns (or wage structure). Finally, we summarize the results from these different techniques, emphasizing both the similarities and differences across the alternative methods.

5.1 Quantile regression estimates: the effect of female and non-white occupational intensity

In performing the pooled quantile regression analysis we explore the use of various different specifications of the wage equation, moving from more austere to more ornate specifications. In the most austere specification the log of hourly wages is regressed on age, age squared, years of education, gender and race, as well as dummies for living in urban areas, living in each of the five main regions of Brazil, and for being a formal worker. In the second specification we then insert dummies for occupations and in the third the variables for female (or, alternatively, non-white) occupational intensity are included. Finally, the fourth and most complete specification includes dummies for occupational codes and the variable for occupational intensity.[11]

Figures 5 and 6 provide a graphical summaries by presenting the coefficients for the main covariates (male, white, education and occupational intensity) across different quantiles for the first year 1987 (plots to the left) and for the final year (plots to the right).

[Figure 5 about here]
[Figure 6 about here]

The male dummy shows different patterns depending on the equation specification and the year. From panel B in figure 5, we can see that when controlling for both female occupational intensity (focc3) and occupational dummies, the male dummy is always positive, and has a U-shaped pattern in 1987, indicating a greater impact on wages at the bottom and top of the wage distribution. By 2006 the male dummy remains positive, but has declined in magnitude, while the U shaped curve has disappeared entirely. Thus, by 2006 the disproportionate impact of gender on wages at the bottom and top of the wage distribution has disappeared.

By contrast, figure 6 reveals that the estimates for the white dummy increase steadily as we move toward the top of the wage distribution, and this pattern is stable over time even after controlling for occupational structure. Interestingly, while including occupational dummies exerts a noticeable impact on the male dummy estimates, it does not have any noticeable effect on the white covariate's coefficient.

Moving beyond the key variables, the estimated coefficients for age and education 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.[12] The same is the case for the age and age squared variables, suggesting a non-linear relationship for this variable. Both variables show a smaller impact on wages over time, though still with an increasing pattern as we move along the wage distribution. For the median regressions, the positive effect of one additional year of age for a 30 year old individual was roughly 3.5% in 1987 but had declined to roughly 1% in 2006. Interestingly, the impact of education declines

[11] We do not report the set of pooled quantile regressions for each equation specification. These regressions are available upon request from the author. Inter-quantile regression estimates are also available in order to test the statistical significance of differences across the main quantiles.
[12] Coelho, Veszteg and Soares (2010) have found similar patterns in a study that estimates the returns to education by likewise employing a quantile regression (in their case they also adopt a semi-parametric correction for sample selection á la Newey (1991) and Buckinsky (1998)).

by roughly 4 percentage points at the top of the wage distribution, and by about 2 percentage points at the bottom of the distribution, when occupational dummies are inserted in the wage equation alongside female occupational intensity. By contrast, when we insert non-white occupational intensity the impact of education immediately declines by 3 percentage points, while adding occupational dummies leads to only a further 1 percentage point decline. Thus, while controlling for non-white occupational intensity has a large impact on the estimate of the education covariate, as does the inclusion of occupation effects, the impact of including female occupational intensity does not have a similarly large effect.

Being a formal sector worker has a positive impact on the level of earnings, but this effect attenuates as we move to higher quantiles and, interestingly, it becomes negative within the top 10% of the wage distribution. 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.

Finally, we wish to look in slightly greater depth at the impact of female and non-white occupational intensity over time, as this represents an important contribution of this paper. To this end we explore the role of these variables not only through the pooled regressions, but also when separating the sample between female and male workers and non-white and white workers. Table 1 reports the estimated coefficients for these two variables across different quantiles and specifications for 1987 and 2006 respectively.

We begin by considering the impact of female and non-white occupational intensity at the mean, reported in the first column of table 1, in order to compare our basic results to those from previous studies. Female occupational intensity (focc3) has a negative impact on overall wages: a 10 percentage point increase in intensity decreases wages on average by roughly 4%. This impact is diminished when occupational controls are also included in the wage equation (for 1987 the coefficient declines from -.0379 to -0.186), while it also declines dramatically over time (from -0.186 in 1987 to -0.043 in 2006 when occupational controls are included). When we split the sample between females and males we see that the overall impact is an average of two contrasting effect: female occupational intensity exerts a negative impact on female wages but a positive one on male wages. However, over time we see these contrasting impacts converging, with the impact on female wages becoming less negative, and the impact on male wages approaching zero. Thus, a 10 percentage point increase in female occupational intensity decreased female wages by 4% in 1987 but by only 1.5% in 2006 (when controlling for occupation effects), while it increased male waged by roughly 1% in 1987 but had no significant impact in 2006.

[Table 1 about here]

Turning to the impact of non-white occupational intensity (nwocc3) the results are more straightforward, as an increasing proportion of non-white workers has a negative impact on wages for both white and non-whites. The magnitude of this negative effect is greater than the impact of female occupational intensity, and is somewhat larger for white workers than non-white workers. Thus, in 1987 a 10 percentage point increase in non-white occupational intensity decreases non-white wages by 2.2% and white wages by 7%. This effect appears to increase in more recent years, as the corresponding figures for 2006 are a 5.3% decline for non-white wages and an 11% decline for white wages.

Before looking at how these results differ at different points of the wage distribution, it is useful to compare these estimated semi-elasticities at the mean to findings from similar studies internationally as presented in table 2. Our estimates of female wage penalties during the 2000s are similar to those that existed in the U.S labour market in the late 1980s and in the 1990s, which generally lie between -0.15 and -0.20 (Johnson and Solon, 1986; MacPherson and Hirsch 1995; Cotter, Hermsen and Vanneman, 2003). By contrast, a similar study of the Canadian labour market found that there was no significant penalty for women working in female dominated occupations (Baker and Fortin, 2003), while a study of the Italian labour market found that females benefit from working in female-dominated occupations (Lucifora and Reilly, 1992). Interestingly, when we turn attention to the impact of female occupational intensity on male wages we find that our results are very different than those from more advanced economies. We find that men have historically benefitted from working in female-dominated occupations, though this effect has largely disappeared in recent years, while, in sharp contrast, previous results from Italy, Canada and the U.S. generally find that men's wages decline even more than female wages in female dominated occupations.

To the best of the author's knowledge there are no similar studies investigating the impact of the concentration within occupations of other minorities, making this the first study to have looked explicitly at the impact of non-white occupational intensity on wages. However, the study by Cotter, Hermsen and Vanneman (2003), noted above, does provide some indirect evidence, as they disaggregate their sample into different ethnic groups in measuring the impact of female occupational intensity on wages. They find that the negative effect is more severe for African American women and for all minorities among men. This appears generally consistent with our findings for Brazil that non-white occupational intensity has a strongly negative impact on wages.

[Table 2 about here]

Having contextualized our broad findings, we now move to exploring our results across different quantiles of the wage distribution. As was initially apparent in figures 5 and 6, we find that female occupational intensity (focc3) exerts a negative impact on wages, while this negative impact becomes greater in absolute terms as we move towards the top of the wage distribution. This larger effect at the top of the distribution is, moreover, even more pronounced in recent years, as can be seen by comparing panels A and B of table 1. In the case of non-white occupational intensity (nwocc3), we see that the presence of non-white workers has had a persistently negative effect on earnings over time, while this effect has been consistently greater at the top end of the pay distribution, independent of whether or not we control for other occupation effects.

Table 1 presents further results focussing on female and male workers (or white and non-white workers) separately. These results are displayed graphically in figures 7 and 8 and reveal the distinct impact of female and non-white occupational intensity on the different population sub-groups. Looking first at the results for female and male workers separately, we find that working in female-dominated jobs decreases earnings for female workers in all specifications and years, while this effect is particularly acute at the top of the wage distribution. The latter effect is strongest when we do not include occupational dummies (panel A of figure 7), while it holds only for 2006 when we add these occupational controls. Conversely, we find that female occupational intensity has a positive effect on male wages, though this effect is only at the bottom end of the wage distribution, and is only apparent when controlling for occupations (compare panel C with panel D in figure 7). That is, once we control for occupational effects, male workers seem to be positively affected by working in female-dominated occupations, particularly within low-paid occupations.

Turning to differences by race, employment in non-white dominated occupations reduces wages for both non-white and white workers, though the effect is slightly more pronounced for white workers. This negative effect increases, in absolute terms, as we move up the distribution, independent of whether we control for occupations, and the magnitude of the effect increases somewhat in recent years (for example, compare panel A with B and C with D in figure 8). The general pattern is relatively stable over time, though it is somewhat more pronounced when we do not control for occupations.

In sum, being employed in female-dominated occupations reduces earnings for female workers, particularly in the highest paid jobs. Interestingly, it has a positive impact on male earnings, but this is only consistently the case in low-paid jobs. Being employed in non-white dominated occupations has a negative impact on wages for all workers, though somewhat more for white workers. As with female occupational intensity, this negative impact is most pronounced within better paid occupations. Finally, these patterns have generally remained stable over time, while only the magnitude of the female occupational intensity variables has, on average, declined over time.

It is important to re-emphasize that many of these estimates represent a novel contribution to existing research, not only for Brazil but internationally, and thus highlight important, but previously overlooked, aspects of wage determination. These new insights fall into three broad categories. First, while several studies internationally have previously looked at the impact of female occupational intensity on wage determination, ours is the first to discover a positive impact on male wages. These results allude to the potential complexity of patterns of wage discrimination, while also pointing towards strikingly entrenched, and explicit, wage discrimination, as employers in female dominated occupations remain willing to pay higher wages to male employees. Second, this study is, to our knowledge, the first to have investigated the impact of non-white occupational intensity on wages. This is a gap in the earlier research, and our finding that non-white occupational intensity has a larger and more persistently negative impact on wages than does female occupational intensity speaks to the importance of this issue. Finally, this study is the first to have linked occupational intensity to wages not only at mean values but also across the entire distribution of earnings. We consistently see more pronounced negative connections at the extremes of the wage distribution, and particularly at the top end, and this provides an important insight into the nature of wage discrimination and particularly into the barriers confronted by minorities in these top positions.

[Figure 7 about here]

[Figure 8 about here]

5.2 Empirical findings from the Machado and Mata (2005) and Melly (2005, 2006) quantile decompositions

We now examine the results of the quantile regression decomposition of the wage gaps, following Machado and Mata (2005) and Melly (2006). In what follows, we report only the results of the quantile decomposition exercise, which exploits the coefficients from the conditional quantile regressions.

We implement both the Machado and Mata (2005) and Melly (2006) techniques, although they should provide asymptotically similar results. We also implement two different variations on the Machado and Mata (2005) technique. We thus first implement the simplified version of this simulation-based decomposition technique, following Albrecht, Bjorklund and Vroman (2003), in which we draw simulated samples only for the realizations of the covariates. In practice, we use 10,000 replications given that in the presence of the occupation effect a higher number of replications is likely to guarantee more realistic realizations for these occupational controls at different quantiles.

We then implement the original version of the Machado and Mata (2005) decomposition and finally the Melly (2006) decomposition.[13] In order to distinguish the implementation of the original version of the Machado and Mata (2005) methodology from the simplified version described above, we denote the original Machado and Mata (2005) version "á la Albrecht, van Vuuren and Vroman (2009)" in our tables. This notation reflects the fact that the implementation of this method relies heavily on the explanation of the methodology provided by Albrecht, van Vuuren and Vroman (2009), particularly in relation to sample selection correction.[14]

We implement these methodologies for both gender and racial pay differentials. In order to retain the temporal perspective we apply the methodology to the first year (1987) and the last year (2006) of our data.[15] In the upper panels of tables 3 and 4 we report the quantile decomposition results using the most complete wage equation specification (the 4th specification).[16] The first three panels report the Machado and Mata (2005) and Melly (2006)

[13] For the implementation of these techniques we adopt two Stata commands. The implementation of the Melly (2006) technique relies on the Melly (2006) Stata command 'rqdeco'. The current command is only able to compute the standard errors via bootstrapping, for which we employ 200 replications, while Melly (2005) provides the computation of the asymptotic variances. The implementation of the original Machado and Mata (2005) technique is conducted using the Stata command 'mmsel', recently released by Souabni (2012).

[14] The 'mmsel' command computes standard errors via a bootstrapping procedure, again set to 200 replications, although Albrecht, van Vuuren and Vroman (2009) provided the computation of analytical asymptotic standard errors. Interestingly, the standard errors using this command are much greater than those obtained using the bootstrapping procedure with 'rqdeco'.

[15] 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.

[16] We perform the decomposition analysis for each wage equation specification. However we discuss only the decomposition results for the 4th specification as reported in tables 2 and 3. The decomposition results for the other specifications are available upon request from the author.

aggregate decomposition results, while the lower panels report the RIF-regression decomposition results, which are discussed in sub-section 4.3. In addition, figure 9 plots the decomposition results over the percentiles of the wage distribution using the Melly (2005) technique.

Looking first at panel A in Figure 9, we see that in 1987 the gender wage gaps were greater at the bottom end of the wage distribution, declining as we move towards the top of the wage distribution before exhibiting a small increase in the highest 10% of the distribution. These wage gaps were primarily attributable to the effects of the coefficients (or returns to characteristics), which were significantly larger at the bottom of the wage distribution. By contrast, the small increase in wage gaps at the top end of the distribution is primarily explained by somewhat better characteristics for men in the higher wage jobs.

When we turn to the results for 2006 we see that the size of wage gaps has contracted over time, while differences across the wage distribution have also declined. Wage gaps have fallen most rapidly at the bottom end of the wage distribution, with this reduction explained primarily by a decline in the effects of the coefficients, although better female endowments have contributed as well. The result is that by 2006 there are only modestly higher wage gaps at the bottom of the wage distribution. When we look to the top of the distribution the pattern is quite different, as the effect of coefficients has decreased rapidly at the bottom end but considerably less so at the top end, with the statistically significant decreases of -0.33, -0.18 and -0.05 at the 10^{th}, 50^{th} and 90^{th} quantiles respectively (see table 3). The effect of coefficients has thus remained relatively stable in the upper part of the wage distribution, yielding an overall U-shaped pattern for both the effect of the coefficients and overall wage gaps. This U-shaped pattern is comparable to that noted in other studies for Brazil. Santos and Ribeiro (2006), for instance, find the existence of wider gender pay gaps at the extremes of the wage distribution, labelling these phenomena glass floors and glass ceilings (in the same spirit as de la Rica, Dolado and Llorens (2005)). Similar results have been also confirmed by Madalozzo and Martins (2007). Garcia Marquez, Ñopo and Salardi (2009) similarly detect a U-shaped pattern of unequal treatment in computing gender wage gaps in Brazil using an alternative non-parametric matching decomposition methodology (see also in Ñopo, 2012: 171). In addition, this U-shaped pattern has been similarly found in other South American countries, such as Chile and Colomb (Ñopo, 2012).

[Figure 9 about here]

Turning to racial pay gaps (see panel B of figure 9), we again see that, in contrast to gender gaps, they are driven largely, but not exclusively, by differences in characteristics, which are generally superior for the white workers. When we disaggregate the analysis into quantiles we see that the impact of both characteristic and coefficient effects tends to increase as we move to the upper part of the distribution, although this progression is particularly apparent for the effects of coefficients. As such, although racial pay gaps are generally the result of differences in characteristics, the sizeable increase in the gap at the top end of the distribution is explained to a large degree by the widening of the effects of coefficients at the top. This wider gap at the top of the distribution is also highlighted in Garcia Marquez, Ñopo and Salardi (2009) and Ñopo (2012: 276), and is potentially indicative of unequal treatment concentrated at the top of the wage distribution. In addition, the effects of coefficients do not appear to have improved at all over time, with the treatment effect at the 90th quantile increasing by 2 percentage points over time, from 0.157 to 0.175 (see table 4). This is an obvious policy concern.

In summary, we find that gender wage differentials are driven primarily by the unexplained components (or treatment effects) with particularly strong effects at the extremes of the wage distribution. These unexplained components may be reflective of entrenched gender-based discrimination in the labour market. More positively, over time the gender wage gap has declined considerably due primarily to a decline in these unexplained components. However, these declines have occurred primarily at the bottom end of the wage distribution, while unexplained gender wage gaps have been more persistent at higher quantiles. Framing these findings in relation to existing concepts in this field, the results suggest that there is a *sticky floor* phenomenon for women, but that it has reduced over time. Turning to the higher pay quantiles, there remain significant unexplained differences in wages, indicative of a discrimination effect. This is consistent with the continued existence of a *glass ceilings* phenomenon within the highest echelons of the Brazilian labour market.

Applying these same concepts to racial wage differentials, we see highly persistent differentials that widen at the higher wage quantiles. This is due to both differences in characteristics and unexplained higher returns to these characteristics among white workers. The continued importance of differences in returns to characteristics is consistent with the existence of *glass ceilings* for non-white workers in the Brazilian labour market.

[Table 3 about here]

[Table 4 about here]

5.3 Empirical findings from RIF-OLS decomposition

As discussed earlier, the primary advantage of the RIF-OLS decomposition technique developed by Firpo, Fortin and Lemieux (2009) is that it permits the computation of more detailed decompositions across quantiles. In particular, it allows us to estimate the contribution of each 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).

In order to provide context for these detailed results it is useful to begin by presenting results from the standard OB decomposition technique at average values. The OB decomposition for mean regressions is presented in tables 5a and 5b. This allows us to compare the detailed decomposition results from the RIF-OLS procedure to these mean results using the OB technique.[17] To this end, before turning to the RIF-OLS results we begin by reviewing the main findings from the aggregate OB decomposition analysis and then discussing the results of the detailed OB decomposition analysis.

At the aggregate level we see a decrease in both gender and racial wage gaps over time with gender wage gaps having declined much faster, despite being smaller in magnitude. Gender differences are, again, overwhelmingly attributable to differences in returns to characteristics (or the wage structure effect), while the effect of characteristics is generally negative, indicating that female workers have better endowments, particularly in their educational attainments.

By contrast, racial differences are largely attributable to differences in characteristics, as white workers have significantly greater endowments than non-whites. The returns to characteristics also remain positive, implying that there remain unexplained wage gaps even after accounting for differences in these endowments. Finally, it is interesting to note that the inclusion of occupational controls (either occupational dummies or female occupational intensity) leads to a large change in the decomposition components of gender wage gaps for the initial year 1987. This is consistent with the hypothesis that female occupational intensity and occupational distribution are important determinants of wage gaps, though the effect of including occupational controls is more muted in 2006 amidst the broader decline in gender

[17] When implementing detailed decomposition analysis, we encounter the problem of choosing a base group. Given our wage equation specification, the choice of the base categories for the occupational dummies is an obvious concern. We have tried several base categories and found that our main decomposition findings are not affected.

wage gaps and gender occupational segregation. The impact of including these variables is also noticeable, but much more modest, in the case of the racial pay gaps.

The detailed decomposition results explain these patterns further by capturing the contribution of each individual covariate in the estimated wage equations. 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 table 5a). Turning to the returns to characteristics (unexplained component), the role played by female occupational intensity stands out, as it has a strongly positive effect on gender differentials. In 1987 it accounted for the largest part of the unexplained components and, while it has declined significantly over time, it remained strongly positive by 2006. If we interpret the unexplained component as capturing labour market discrimination then this finding suggests that a large part of existing discrimination is rooted in higher rewards for males working in female-dominated jobs.

Turning to racial wage gaps (table 5b), education again plays a central role in the explained component, reflected in much better endowments for white workers. When we turn to the returns to characteristics, non-white occupational intensity and occupation effects together account for a large portion of the overall pattern. The negative effect of non-white occupational intensity implies better returns for non-white workers, relative to white workers within non-white-dominated occupations, while recalling our earlier analysis that non-white occupational intensity has an overall negative effect on wages. Meanwhile, the positive contribution of occupations conveys the fact that whites are employed in more rewarding jobs.

[Table 5a and 5b about here]

Overall, 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 these male characteristics. Notably, being a male worker within a female-dominated occupation appears to be particularly well rewarded. In the case of racial differentials, white workers are paid more in large part because they have better endowments, and particularly better educational levels. 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. If we interpret the impact of age as the possible role of work experience, the message appears to be that experience is rewarded comparatively better for men and white workers.

Having reviewed these findings from the detailed decomposition at the average level, we are able to more fully interpret the detailed decomposition results from the RIF-OLS regression decomposition methodology, reported earlier in the lower panels of tables 3 and 4. The first point to note is that the decomposition results produced by the RIF-OLS methodology broadly coincide with those from the Machado and Mata (2005) and Melly (2006) techniques, thus reinforcing confidence in the results. A discussion of the similarities and differences in the results across these different quantile decomposition methodologies is presented in the next sub-section.

Moving to the specific results, the tables present the individual contributions of four key covariates to both the characteristics and coefficient components: age, years of education, female (or non white) occupational intensity (focc3 and nwocc3), and occupation effects. Looking across the results, it is again clear that both education and occupational intensity perform a crucial role in determining wage differentials, though in distinctly different ways.

For gender, education has a strong and negative effect on wage differentials across all of the decomposition results covering the entire labour market. Its negative effect increases, in absolute terms, as we move to the top end of the wage distribution, again highlighting that education is the most important source of better female endowments, while this effect is greater at higher wage quantiles. Moving to the individual contributors to the coefficients component, the age variable exerts a sizeable impact. Its effect is positive, and higher at the top end of the wage distribution. If we again interpret age as the effect of work experience, we may conclude that men's work experience is rewarded more than that of women, particularly among high-paid jobs. The returns to education are also positive, indicating that while women are better educated, men receive consistently greater rewards to education, particularly in the higher quantiles. The returns to female occupational intensity (focc3) also play a key role here. It is always positive, and follows a U-shaped pattern across wage quantiles, as it is greater at the extremes of the distribution. The returns to occupation, meanwhile, are generally negative, and particularly so at the top end of the pay distribution in 2006. This pattern can be interpreted as indicating that female occupational outcomes, particularly among those in highly paid jobs, have been increasingly rewarded over time.

Turning to racial wage gaps, education again plays a key role in determining the magnitude of these gaps. In this case the effect is positive, while, as with gender, the effect is greater at the higher wage quantiles. Looking at the effects of the coefficients, there are higher returns to education for white workers, in addition to their already higher educational endowments, particularly as we move to the top end of the wage distribution. The age variable again makes a large positive contribution to the wage differentials, especially in the centre of the wage distribution, which we might again interpret as reflecting superior rewards to work experience for white workers.

In contrast to the case of gender wage gaps, the returns to non-white occupational intensity generally have a negative impact on wage gaps, with a particularly sizeable effect at lower wage quantiles. Non-white workers thus benefit from better returns to working in non-white dominated occupations, relative to white workers, particularly within low-paid occupations. On the other hand, the occupation effects contribute positively to wage differentials, and particularly strongly so at the very top of the pay distribution (0.99 quantile). Thus, while being employed in non white-dominated occupations reduces relative white wages within low-paid occupations, white workers are disproportionately rewarded by their heavy representation in the highly paid occupations.

In summary, the results when employing the RIF-OLS methodology, are broadly consistent with the mean regression analysis, while adding important insights into the role of key covariates at different points of the pay distribution. In the case of gender wage gap differentials, the large positive unexplained component is mitigated by the negative explained component, particularly so at the top of the distribution. Were it not for superior female endowments, largely in terms of education, the total wage gap would be significantly wider, particularly at the top of the pay distribution. Even if there are some characteristics that we are not able to control for in our analysis, such as innate ability, it is possible that a good portion of the sizeable unexplained differences in gender wage gaps (the wage structure effect) are due to gender discrimination. This seems likely in light of the fact that men's age is rewarded more than women's age in top positions and that men working in female-dominated occupations receive higher wages in both high and low paid occupations. This again suggests that women are subject to the dual phenomenon of *sticky floors* and *glass ceilings* in the Brazilian labour market.

On the other hand, racial wage differentials are overwhelmingly explained by differences in observed characteristics, with differences in educational attainments playing an important role and with these differences tending to widen at higher wage levels. Although wage differentials are generally explained by differences in characteristics, differences in returns have remained persistent over time, and are accentuated as we move to the top end of the wage distribution, where there remain significant unexplained differences in returns. For recent years these disproportionately large unexplained differences at the top of the pay distribution reflect three factors. First, non-white workers are more rewarded within low-paid jobs, thus reducing wage differentials at the bottom end of the distribution. Second, there are systematically higher returns to education at higher wage quantiles, while white workers are generally more educated. Finally, there are very high and positive returns to occupations at the very top of the distribution, implying considerably higher returns for those whites who disproportionately occupy highly paid positions (using the third specification, in which we do not include occupational dummies), this is reflected in a highly positive coefficient on non-white occupational intensity at the top of the wage distribution. This could be taken as providing genuine evidence of a *glass ceilings* phenomenon affecting non-white workers.

5.4 Comparing the different quantile decomposition techniques

We have now reported quantile decomposition results computed using several techniques, which we expected to provide generally similar outcomes. This sub-section compares the results from these different methodologies, focusing on the question of whether the estimated decomposition components are different across methods.

Tables 3 and 4 presented the core results computed by implementing the Machado and Mata (2005) á la Albrecht, Bjorklund and Vroman (2003), the Melly (2006) decomposition, the original version of the Machado and Mata (2005) decomposition and, finally, the RIF-OLS method with its detailed decomposition results.

Figures 10 and 11 plot the decomposition results across these methodologies for gender and racial wage gaps, respectively, looking separately at the explained component, the unexplained component and the total gap. These figures are based on the results reported in tables 3 and 4, which are computed using the most complete specification, which includes both occupational controls (occupational intensity and occupation effects).

The results using the Melly (2006) and Machado and Mata (2005) procedures are almost identical. Meanwhile, the results from the Machado and Mata (2005) procedure, implemented á la Albrecht, Bjorklund and Vroman (2003), are generally similar to the results using the RIF-OLS procedure, though with only some slight differences.

Where we notice differences between the methods, these tend to occur at the extremes of the wage distribution, and most notably at the 10th quantile for gender and the 90th quantile for race. By contrast, the median decomposition results are less likely to differ across methods. Overall, the similarity of the results across methods inspires much confidence that the broad results obtained are fairly robust across all procedures.

[Figure 10 about here]

[Figure 11 about here]

6. Conclusions

The paper has analysed the evolution of gender and racial wage differentials in the Brazilian labour market, while making two innovative contributions. First, we have moved beyond investigating wage differentials at mean values in order to consider wage differentials at different points of the wage distribution. To this end we have employed two recent quantile decomposition techniques (developed by Machado and Mata (2005) and Melly (2005, 2006) and by Firpo, Fortin and Lemieux (2009)), in order to isolate the endowment and treatment elements contributing to wage differentials at different points of the distribution. Second, within the decomposition analysis we have drawn on a harmonized dataset in order to focus attention on the relationship between occupational intensity and wage determination and discrimination.

The paper began by presenting a preliminary analysis of the relationship between occupational intensity and earnings differentials and this discussion yielded a number of relatively useful insights. In broad terms we find significant differences in the relationships between occupational intensity and earnings by gender and race. Being employed in female-dominated occupations reduces earnings for female workers, particularly in the highest paid jobs, while, in contrast, it exerts a positive impact on male earnings, though only in low-paid jobs. Turning to racial dynamics, being employed in non-white dominated occupations has a negative impact on wages for all workers, though somewhat more among whites. As with female occupational intensity, this negative impact is most pronounced within the better paid occupations. These patterns have, again, remained relatively stable over time, though with the magnitude of the effects actually increasing over time.

This preliminary analysis not only highlighted the importance of accounting for occupational intensity in assessing wage discrimination, but also provided important new research insights in its own right. First, this study finds that female occupational intensity has a negative impact on female wages but a positive impact on male wages in contrast with the existing literature. Given that earlier studies have all focused on more developed economies, the finding here may point toward a previously overlooked aspect of wage discrimination in less developed countries. Second, this study, to our knowledge, is the first to have investigated the impact of non-white occupational intensity on wages, which gives added importance to our finding that non-white occupational intensity has a larger and more persistently negative impact on wages than female occupational intensity. Finally, this study is the first to have linked occupational intensity to wage determination across the entire distribution of earnings, and highlighted the existence of significant variation particularly at the extremes of the distribution.

Turning to the decomposition analysis, we began with the results calculated at the mean, which revealed that gender wage gaps are smaller than racial wage gaps. This is in large part because gender wage gaps have declined significantly over the last two decades. The considerable and relatively stable magnitude of racial pay differentials is of obvious concern, while the sharp decline in gender wage gaps is somewhat encouraging. However, the detailed decomposition results provide a more nuanced portrait of the underlying components of these trends. In the case of gender differentials, the sharp decline in aggregate wage gaps has been driven to a significant degree by changes in characteristics, attributable primarily to increasing female education, while the unexplained component, which is potentially indicative of discrimination, has been declining but remains positive and statistically significant. Interestingly, and consistent with the second objective of the paper, we find evidence that the unexplained component is closely related to the question of occupational segregation, as men not only receive higher wages than women, but receive even more disproportionate returns when employed in heavily female dominated occupations.

In the case of racial differentials, lower wages for non-whites are primarily the result of persistently lower endowments, again with education playing a primary role. The unexplained differences in the wage structure are lower than those found for the gender-based wage differentials but remain positive, and have been highly persistent over time. These very different patterns suggest that the challenge of reducing wage differentials is quite different depending on whether the focus is on gender or race.

While these results provide a baseline, decomposing the wage differentials at different quantiles reveals important differences across the wage distribution, particularly in relation to gender pay gaps. Gender wage differentials tend to exhibit a U-shaped pattern, indicating higher wage differentials at the extremes of the pay distribution. Again, these differentials are primarily the result of wage structure effects, which remain positive despite having declined considerably over time, particularly at the bottom end of the pay distribution. This suggests the existence of a *sticky floor* phenomenon, while also revealing the existence of persistent *glass ceilings* for women.

Turning to racial wage differentials a single key message emerges: 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, and this pattern does not appear to have changed 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 insights into the role of individual variables in accounting for the wage gaps. Focusing first on the importance of characteristics, we find that education is the major contributor to better female characteristics, while we can now also see that this effect is particularly important as we move up the wage distribution. Education is also the most important characteristic in looking at racial wage gaps, though in that case it serves to increase wage differentials, as white workers possess more education than non-whites, while this effect increases at higher quantiles.

Turning attention to the effects of coefficients on gender wage gaps we find that male experience, as proxied by age is rewarded more than women's at the top of the pay distribution, while men working in female-dominated occupations are better paid than women, again particularly in top and low-paid occupations. These trends reinforce the apparent existence of *sticky floors* and of *glass ceilings* for female workers. 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 among low-paid occupations. However, while occupational intensity is seen to favour non-white workers in low-paid occupations, we see that the returns to occupations contribute positively to racial wage differentials, with very large effects at the very top of the pay distribution. Thus, while being employed in non white-dominated occupations marginally reduces white wages within low-paid occupations, white workers are very highly rewarded by their presence in top-occupations. This would seem to provide evidence for the presence 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 inter-connections between wage discrimination and occupational segregation. These results appear to be highly robust, as the main findings have remained essentially unchanged across a range of alternative quantile decomposition methodologies. These findings are suggestive of key areas of focus for interventions aimed at reducing wage differentials and the persistence of unexplained differences in wage structure is 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 exposes some differences that point towards differing challenges in moving forward and the potential need for distinct group-specific policy prescriptions.

References

Albrecht, J., A. Bjorklund, and S. Vroman (2003, January). Is there a glass ceiling in Sweden? *Journal of Labor Economics* 21 (1), 145-177.

Albrecht, J., A. van Vuuren, and S. Vroman (2009a, August). Counterfactual distributions with sample selection adjustments: Econometric theory and an application to the Netherlands. *Labour Economics* 16 (4), 383-396.

Andrews, D.W. and M.A. Schafgans (1998). Semiparametric estimation of the intercept model. *Review of Economic Studies* 65, 497–517.

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 labor market: Wage, employment and segregation effects. *Journal of International Development* 16, 1053-1066.

Autor, D. H., L. F. Katz, and M. S. Kearney (2005). Rising wage inequality: The role of composition and prices. Harvard Institute of Economic Research Working Papers 2096, Harvard - Institute of Economic Research.

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.

Booth, A. L., M. Francesconi, and J. Frank (2003, April). A sticky floors model of promotion, pay, and gender. *European Economic Review*, Elsevier, vol. 47(2), pages 295-322.

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

Brown, R. S., M. Moon, and B. S. Zoloth (1980). Incorporating occupational attainment in studies of male-female earnings differentials. *The Journal of Human Resources* 15 (1), 3-28.

Buchinsky, M. (1998). The dynamics of changes in the female wage distribution in the USA: 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 Caxambu-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.

Chi, W. and B. Li (2008, June). Glass ceiling or sticky floor? Examining the gender earnings differential across the earnings distribution in urban china, 1987-2004. *Journal of Comparative Economics* 36 (2), 243-263.

Chzhen, Y. amd K. Mumford and C. Nicodemo (2012). The Gender Pay Gap in the Australian Private Sector: Is Selection Relevant across the Wage Distribution?. IZA Discussion Papers 6558, Institute for the Study of Labor (IZA).

Chzhen, Y. and K. Mumford (2011). Gender gaps across the earnings distribution for full-time employees in Britain: Allowing for sample selection. *Labour Economics* 18 (6), 837-844.

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, Brasilia.

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.

de la Rica, S., J. J. Dolado, and V. Llorens (2005, January). Ceiling and floors: Gender wage gaps by education in Spain. IZA Discussion Papers 1483, Institute for the Study of Labor (IZA).

Deaton, A.S. (1997). The Analysis of Household Surveys: a Microeconometric Approach to Development Policy. Baltimore: John Hopkins/ World Bank.

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., M. G. H. 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 (2000a, 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.

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. (1994). Incorporating the Causes of Occupational Differences in Studies of Racial Wage Differentials. *Journal of Human Resources* 29(1), 20-41.

Gosling, A., S. Machin, and C. Meghir (2000, October). The changing distribution of male wages in the U.K. *The Review of Economic Studies* 67 (4), 635-66.

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.

Heywood, J. S. and D. Parent (2009). Performance pay and the white-black wage gap. Cahiers de recherche 0916, CIRPEE.

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.

Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single index models. *Journal of Econometrics* 58, 71–120.

Johnson, G. E. and G. Solon (1986). Estimates of the direct effects of comparable worth policy. *American Economic Review* 76 (5), 1117-25.

Juhn, C., K. M. Murphy and B. Pierce. (1991). Accounting for the Slowdown in Black-White Wage Convergence. In M. H. Kosters (ed.), Workers and Their Wages: Changing Patterns in the United States, 1991. AEI Press: Washington D.C., 107-143.

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.

Kee, H. J. (2006). Glass ceiling or sticky floor? Exploring the Australian gender pay gap. *Economic Record* 82 (259), 408-427.

Koenker, R. (2005). Quantile Regression. Econometric Society Monograph Series. Cambridge University Press.

Koenker, R. W. and G. J. Bassett, (1978, January). Regression quantiles. *Econometrica* 46 (1), 33-50.

Koenker, R. W. and K. F. Hallok (2001). Quantile regression: An introduction. *Journal of Economic Perspectives* 15 (4), 143-156.

Krueger, A. B. and L. H. Summers (1988). Efficiency waged 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.

Lucifora, C. and B. Reilly (1990). Wage discrimination and female occupational intensity. *Labour* 4 (2), 147-168.

Machado, C. (2011). Selection, heterogeneity and the gender wage gap. Job Market Paper.

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).

Madalozzo, R. and S. R. Martins (2007). Gender wage gaps: comparing the 80s, 90s and 00s in Brazil. *Revista de Economia e Administração* 6(2), 141-156.

Marquez Garcia, L.,H. Ñopo, and P. Salardi (2009, May). Gender and racial wage gaps in Brazil 1996-2006: Evidence using a matching comparisons approach. RES Working Papers 4626, Inter-American Development Bank, Research Department.

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. MIT, mimeo.

Nicodemo, C. (2009, January). Gender pay gap and quantile regression in European families. IZA Discussion Papers 3978, Institute for the Study of Labor (IZA).

Ñopo, H. (2012). New Century, Old Disparities: Gender and Ethnic Earnings Gaps in Latin America and the Caribbean. Washington D.C.: Inter-American Development Bank and World Bank.

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.

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.

Pham, T.-H. and B. Reilly (2007, October). The gender pay gap in Vietnam, 1993-2002: A quantile regression approach. *Journal of Asian Economics*, 18(5), 775-808.

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. (2012). An Analysis of Pay and Occupational Differences by Gender and Race in Brazil - 1987 to 2006. PhD Thesis. University of Sussex, U.K.

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).

Smith, J. P. and F. R. Welch (1989). Black economic progress after Myrdal. *Journal of Economic Literature* 27 (2), pp. 519-564.

Souabni, S. (2012). MMSEL: Stata module to simulate (counterfactual) distributions from quantile regressions (with optional sample selection correction). Statistical Software Components, Swansea University College of Business and Law.

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.

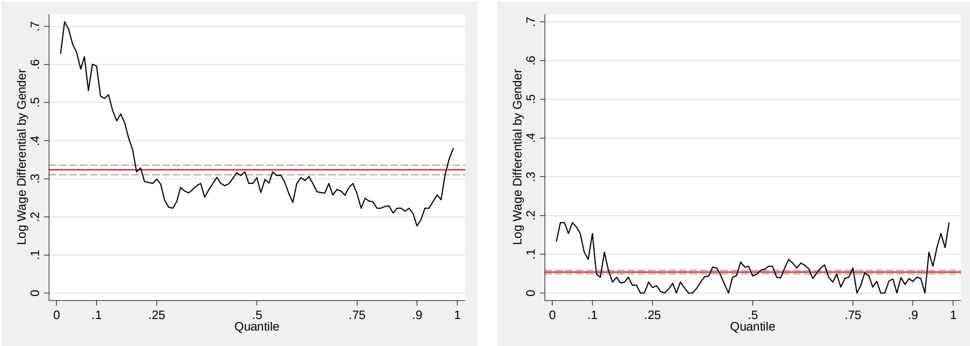
Tiefenthaler, J. (1992). Female labor force participation and wage determination in Brazil 1989. In G. Psacharopolous and Z. Tzannatos (Eds.), Case Studies of Women's Employment and Pay in Latin America. World Bank.

Yun, M.-S. (2009). Wage differentials, discrimination and inequality: A cautionary note on the Juhn, Murphy and Pierce decomposition method. *Scottish Journal of Political Economy* 56 (1), 114-122.

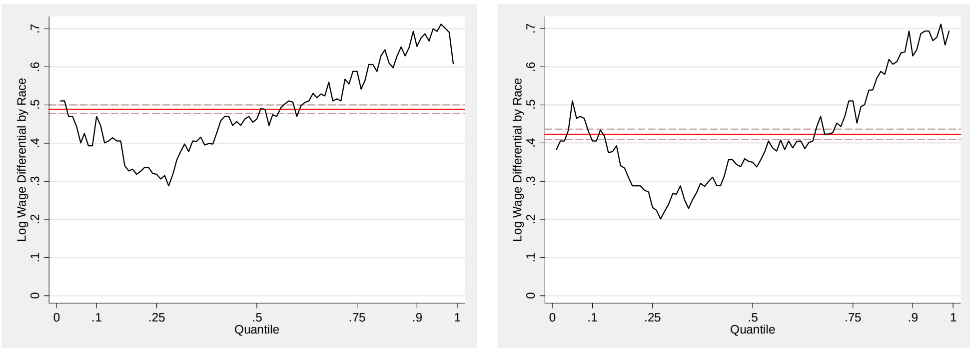
Figures and Tables to be inserted in the paper

Figure 1: Wage differentials over wage quantiles

Panel A – Wage differentials by gender, 1987 and 2006



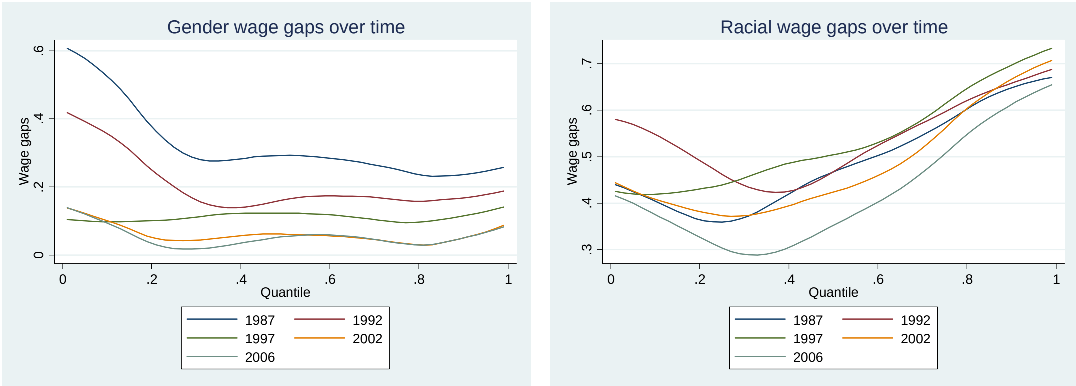
Panel B – Wage differentials by race, 1987 and 2006



Source: Author's computations using PNAD 1987 and 2006.
Note: the red horizontal lines represent the mean values for wage gaps. The wage differentials are the difference of the value of wages for each percentile computed separately for each sub-group.

Figure 2: Evolution of wage gaps over time, all labour market

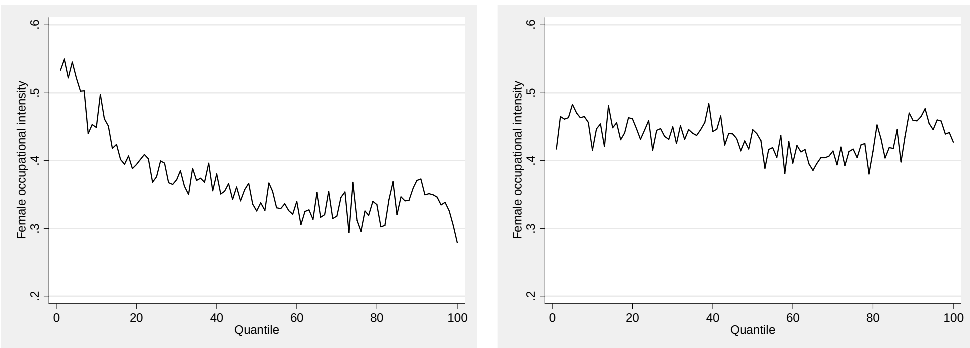
Panel A – Gender wage gaps **Panel B – Racial wage gaps**



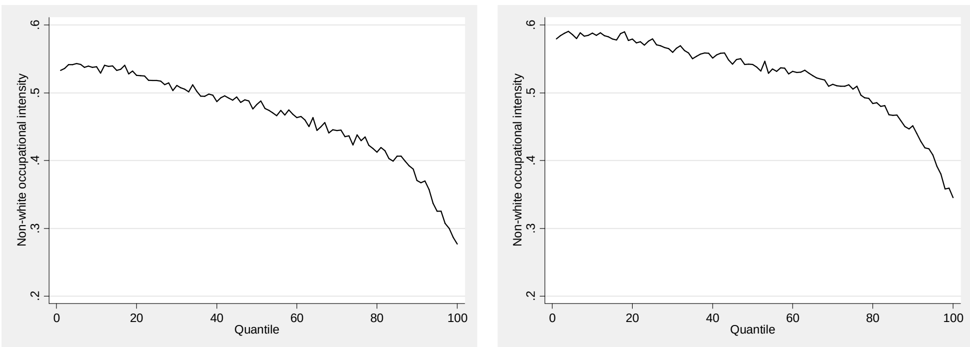
Source: Author's computations using PNAD 1987 – 1992 – 1997 – 2002 - 2006.

Figure 3: Occupational intensity over wage quantiles

Panel A- Female occupational intensity, 1987 and 2006



Panel B- Non-white occupational intensity, 1987 and 2006

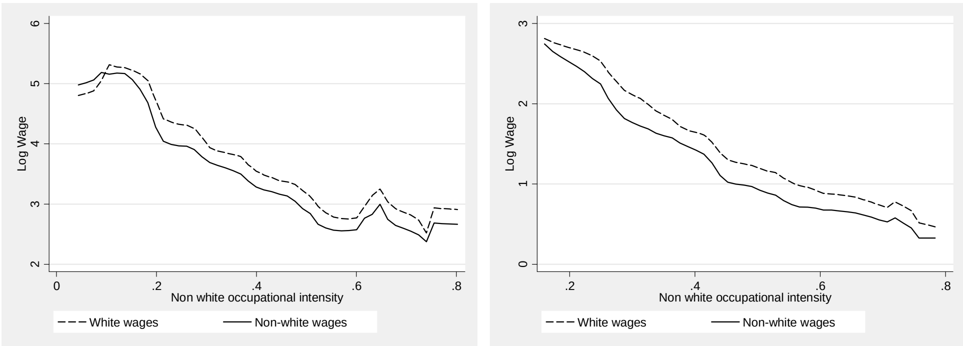


Source: Author's computations using PNAD 1987 and 2006.

Figure 4: Average wages over occupational intensity

Panel A - By gender, 1987 and 2006

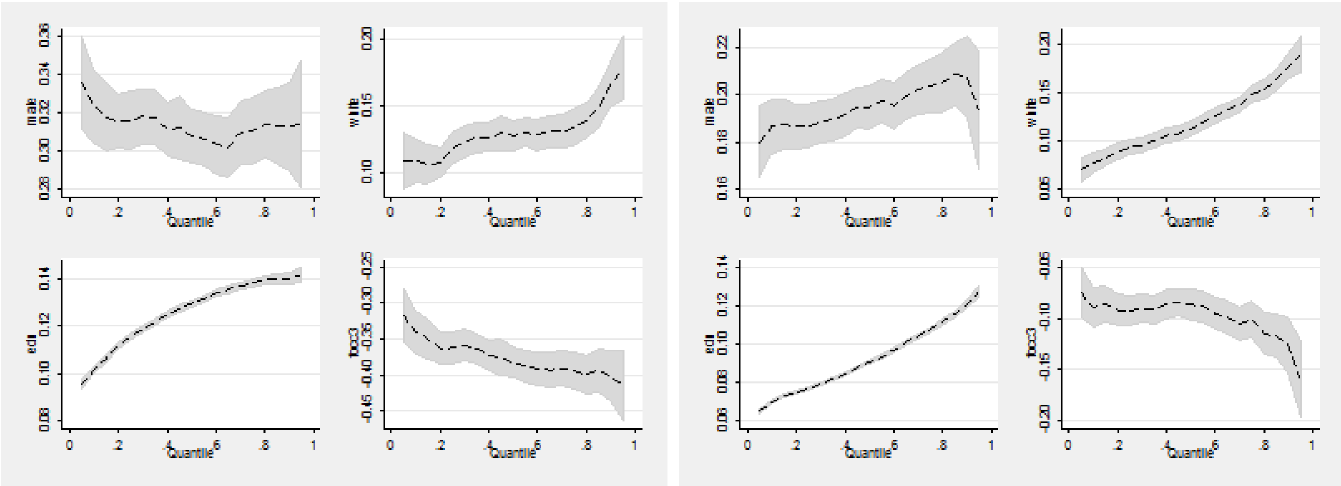
Panel B - By race, 1987 and 2006



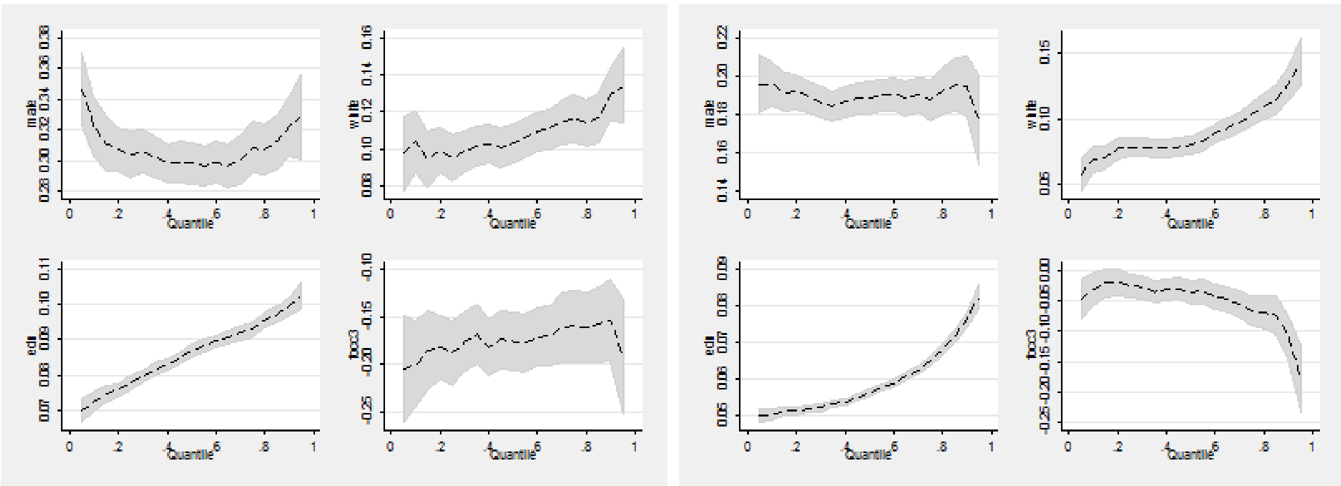
Source: Author’s computations using PNAD 1987 and 2006.

Figure 5: Main covariates' effect (including focc3) from pooled quantile regressions

Panel A – without occupations (using the 3rd specification), 1987 and 2006



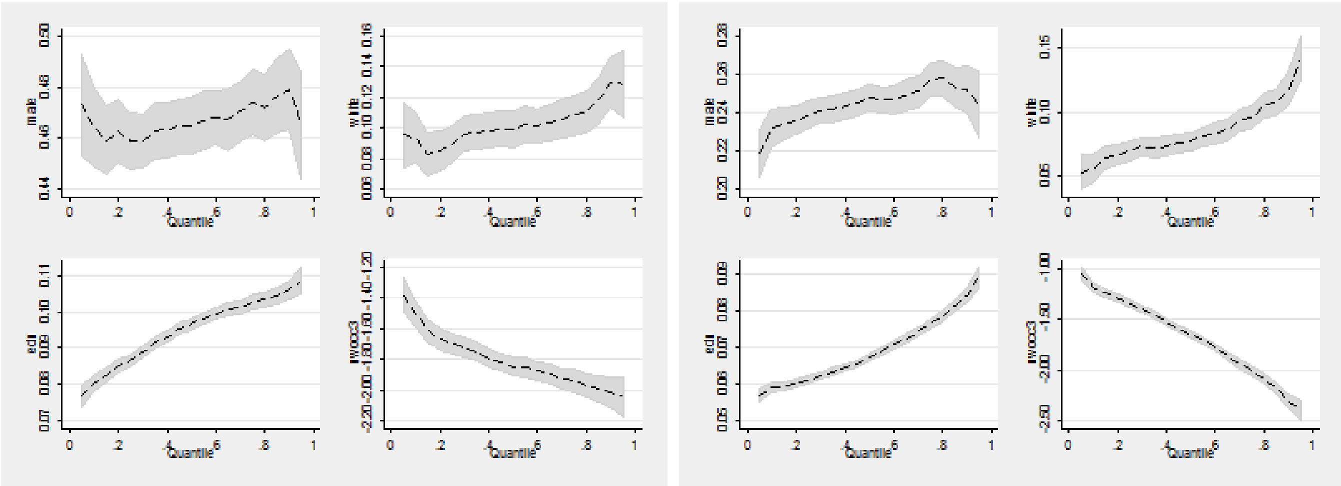
Panel B – with occupations (using the 4th specification), 1987 and 2006



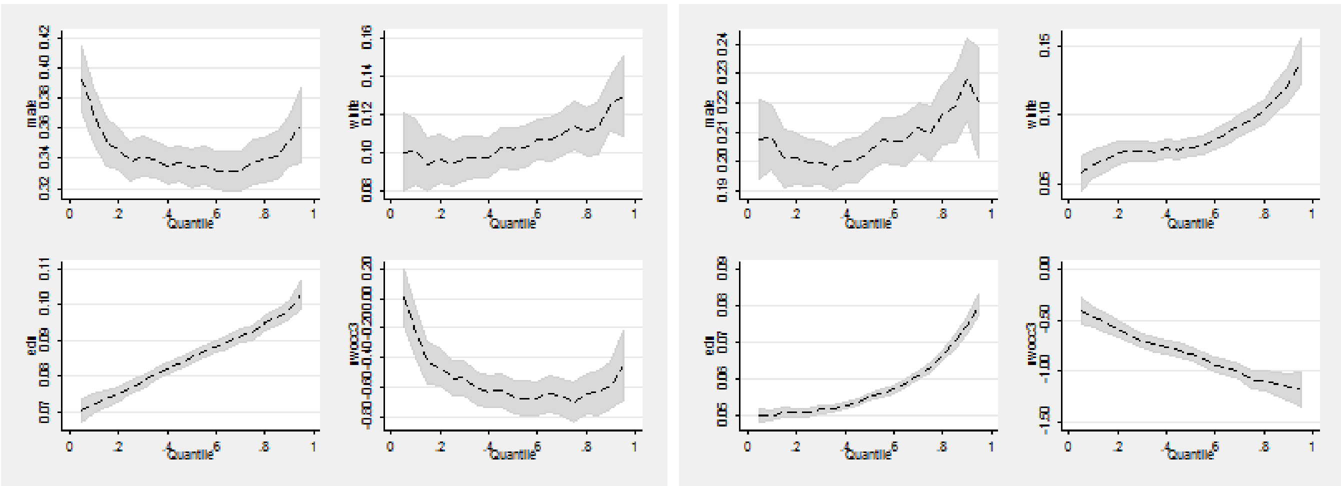
Source: Author's computations using PNAD 1987 and 2006. Note: Bootstrapped standard errors using 200 replications.

Figure 6: Main covariates' effect (including nwocc3) from pooled quantile regressions

Panel A –without occupations (using the 3rd specification), 1987 and 2006



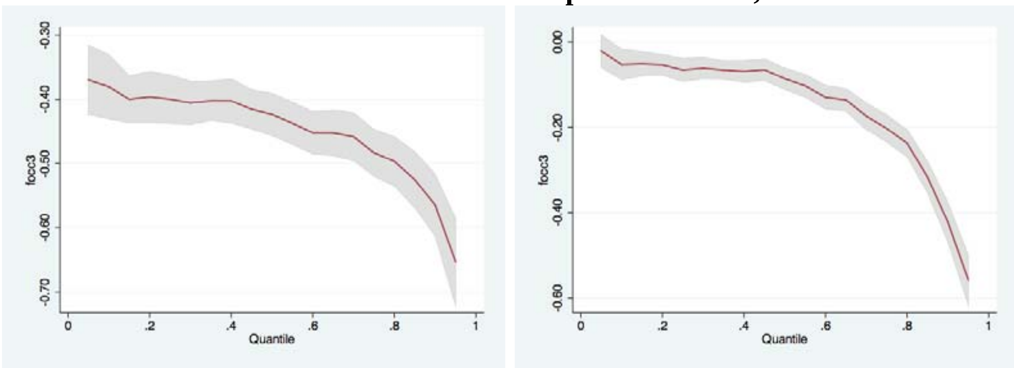
Panel B – with occupations (using the 4th specification), 1987 and 2006



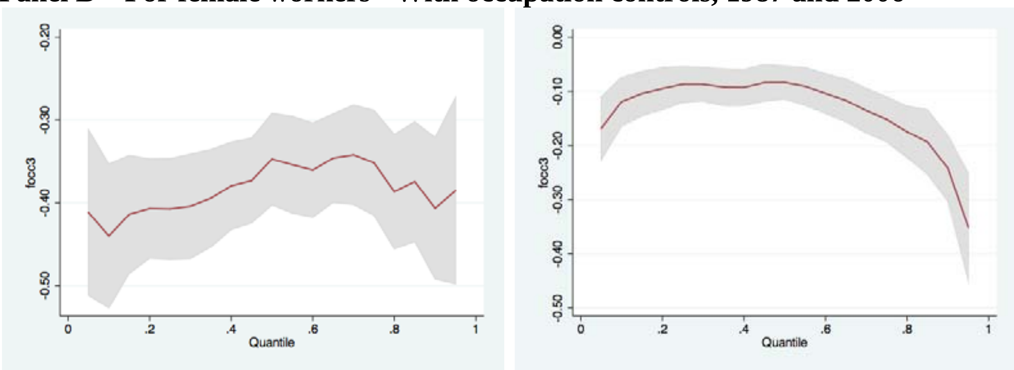
Source: Author's computations using PNAD 1987 and 2006. Note: bootstrapped standard errors using 200 replications.

Figure 7: The role of female occupational intensity

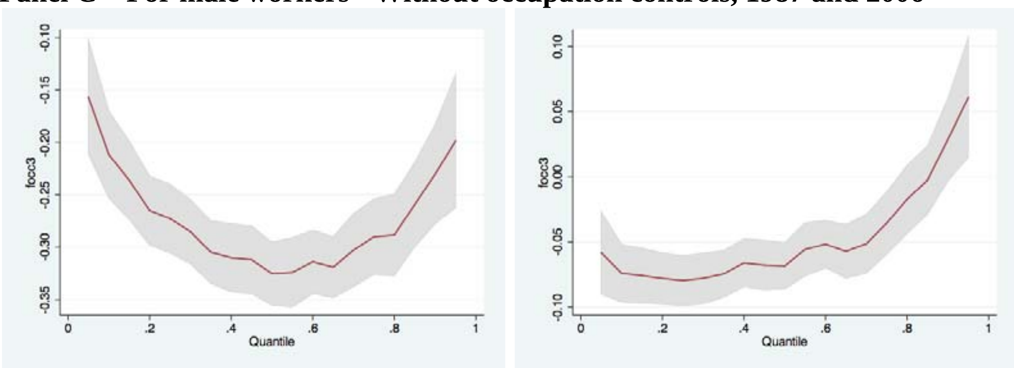
Panel A – For female workers - Without occupation controls, 1987 and 2006



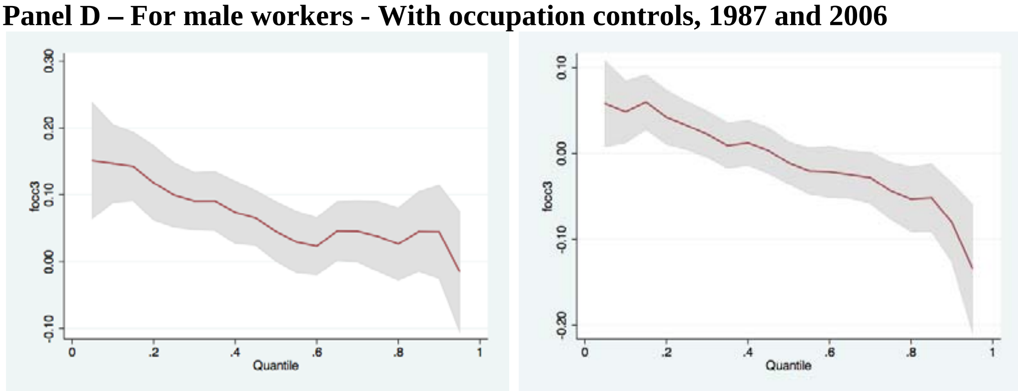
Panel B – For female workers - With occupation controls, 1987 and 2006



Panel C – For male workers - Without occupation controls, 1987 and 2006



Panel D – For male workers - With occupation controls, 1987 and 2006

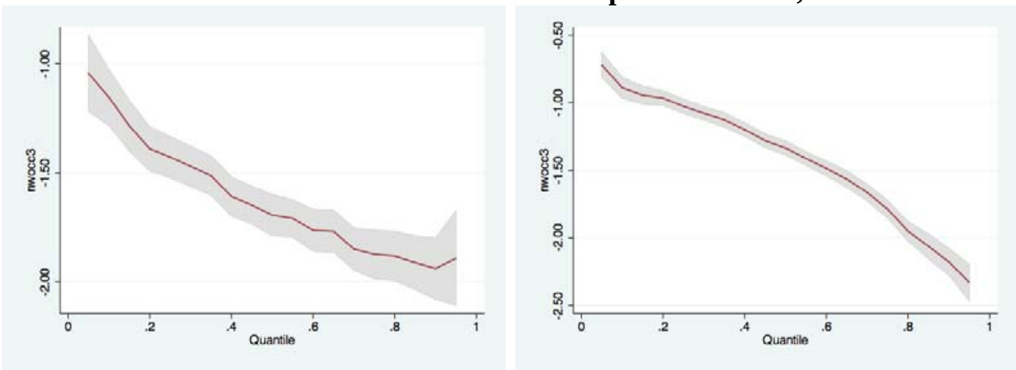


Source: Author's computations using PNAD 1987 and 2006.

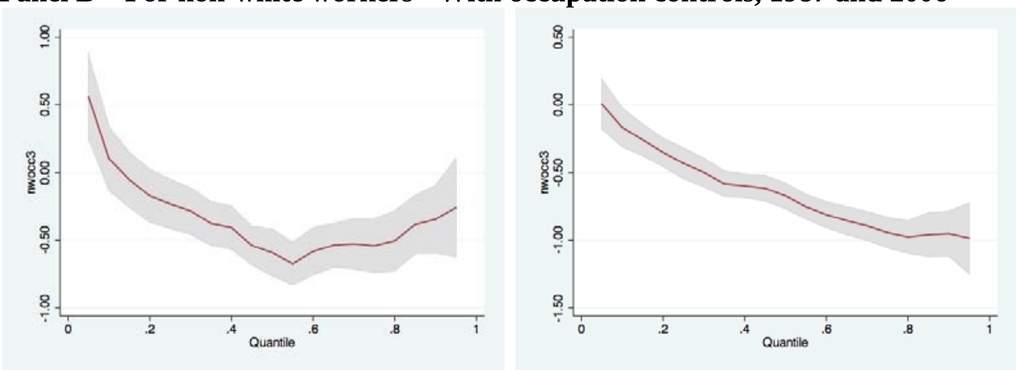
Note: Panels A and C correspond to the 3rd specification of the wage equation, while panels B and D correspond to the 4th specification.

Figure 8: The role of non-white occupational intensity

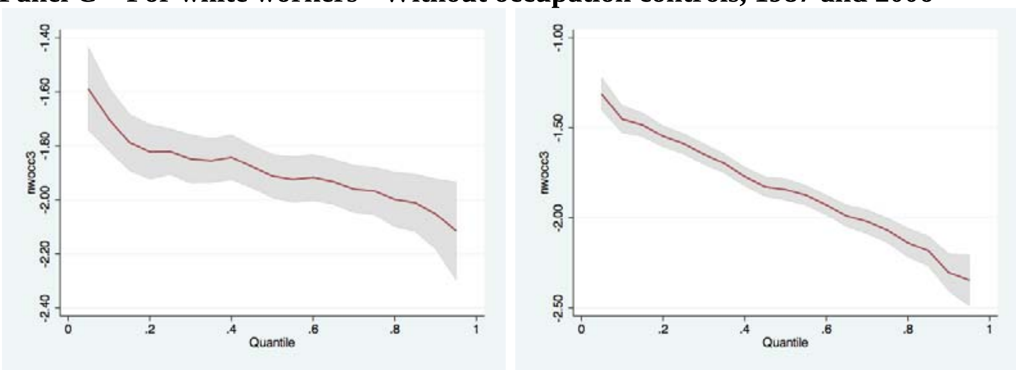
Panel A – For non-white workers - Without occupation controls, 1987 and 2006



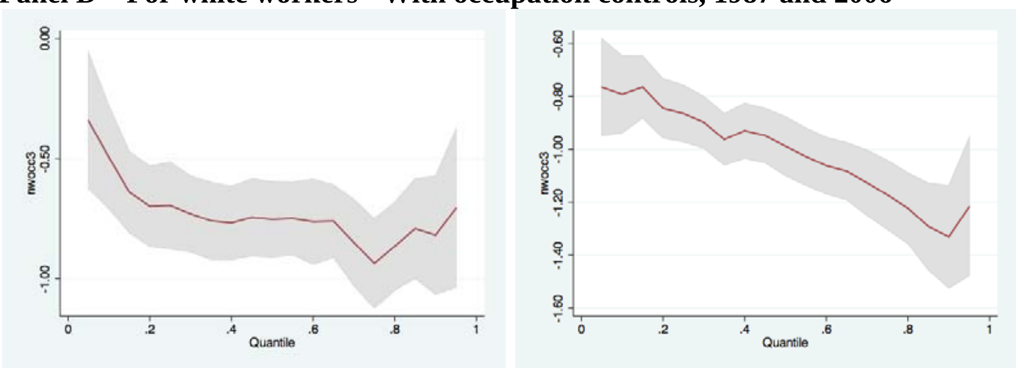
Panel B – For non-white workers - With occupation controls, 1987 and 2006



Panel C – For white workers - Without occupation controls, 1987 and 2006



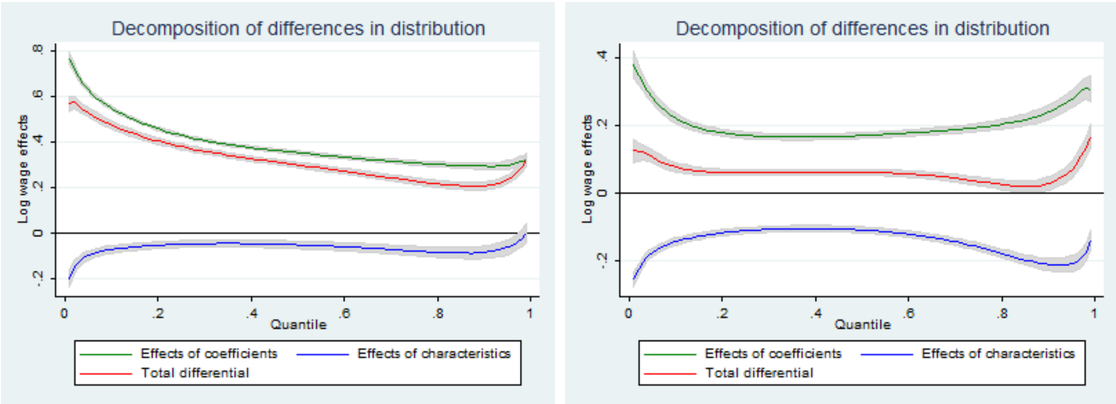
Panel D – For white workers - With occupation controls, 1987 and 2006



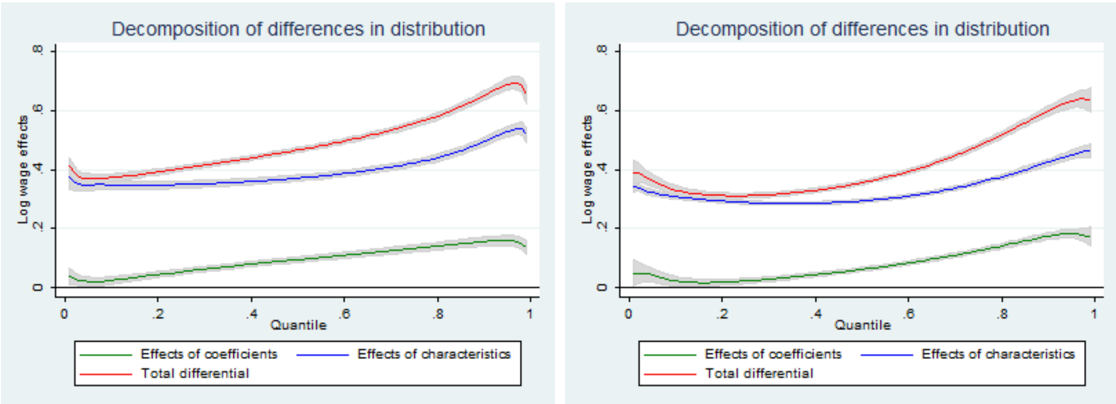
Source: Author's computations using PNAD 1987 and 2006.
Note: Panels A and C correspond to the 3rd specification of the wage equation, while panels B and D correspond to the 4th specification.

Figure 9: Melly (2006) quantile decomposition results (using the 4th specification)

Panel A – Gender wage gaps, 1987 and 2006



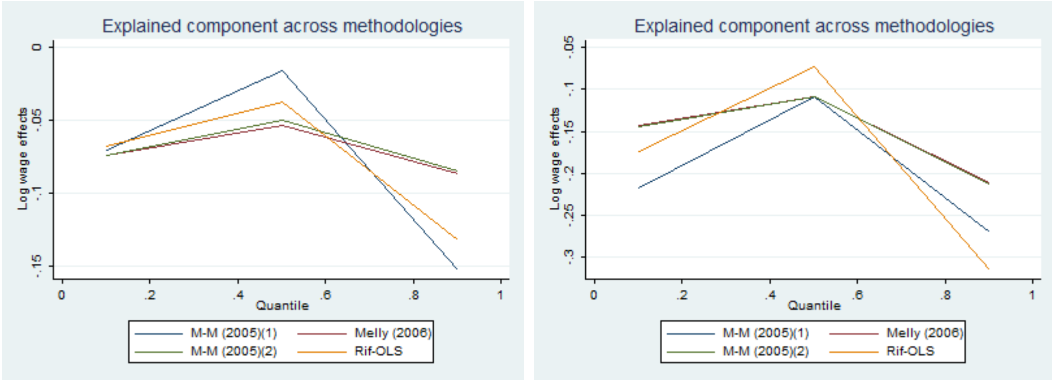
Panel B – Racial wage gaps, 1987 and 2006



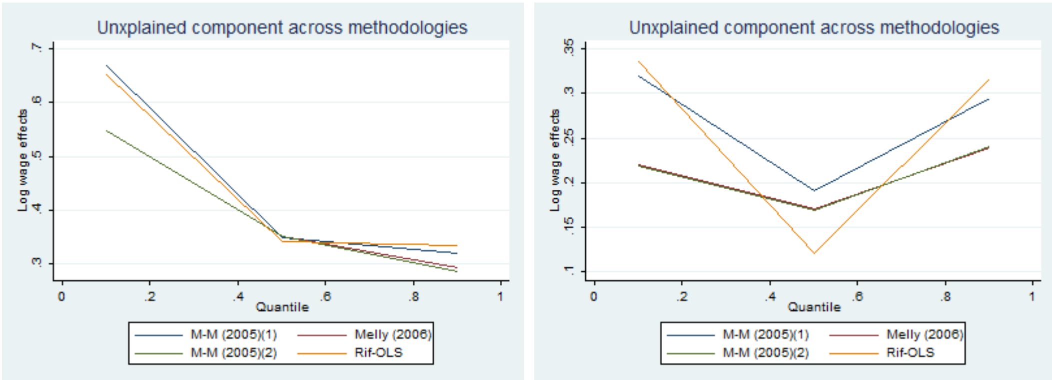
Source: Author's computations using PNAD 1987 and 2006.
Note: bootstrapped standard errors using 200 replications.

Figure 10: Comparing decomposition results across methodologies for gender gaps (using the 4th specification)

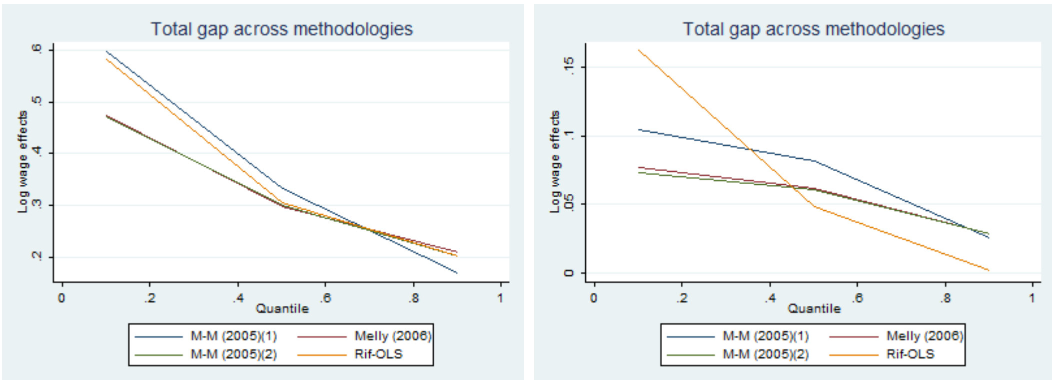
Panel A – Explained component (effect of characteristics)



Panel B – Unexplained component (effect of coefficients)



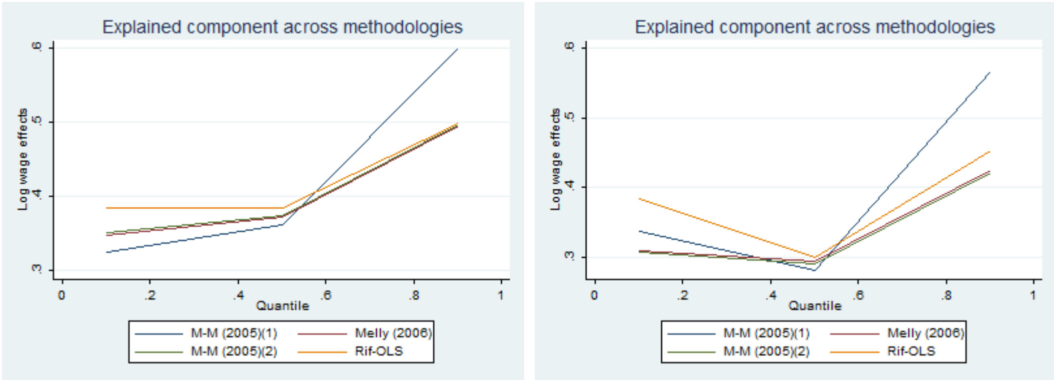
Panel C- Total gap



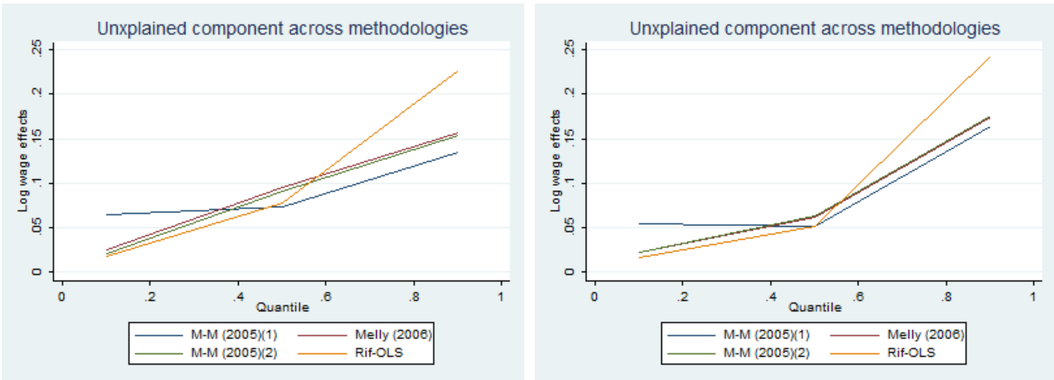
Source: Author's computations using PNAD 1987 and 2006.
Note: M-M (2005) (1) refers to Machado and Mata (2005) as Albrecht, Bjorklund and Vroman (2003); M-M (2005) (2) refers to Machado and Mata (2005) as Albrecht, van Vuuren and Vroman (2009).

Figure 11: Comparing decomposition results across methodologies for racial gaps(using the 4th specification)

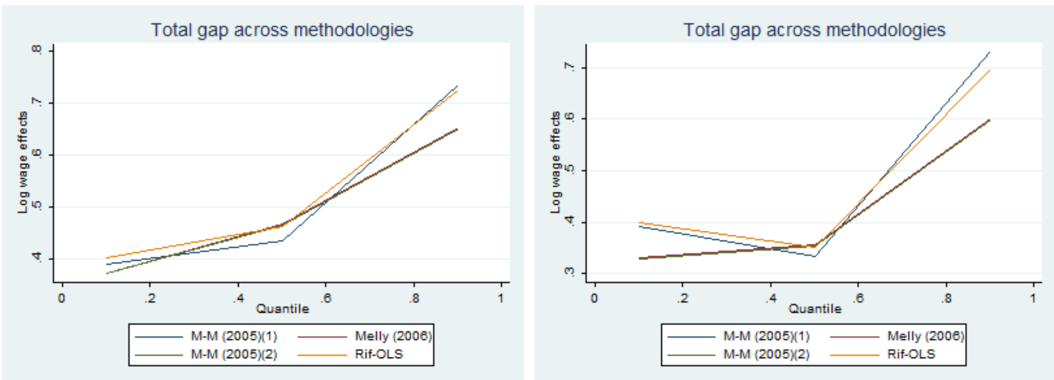
Panel A – Explained component (effect of characteristics)



Panel B – Unexplained component (effect of coefficients)



Panel C- Total gap



Source: Author's computations using PNAD 1987 and 2006.
Note: M-M (2005) (1) refers to Machado and Mata (2005) as Albrecht, Bjorklund and Vroman (2003); M-M (2005) (2) refers to Machado and Mata (2005) as Albrecht, van Vuuren and Vroman (2009).

Table 1: Semi-elasticities for female and non-white occupational intensity across different specifications and samples

Panel A - year 1987

	(1) mean	(2) 1	(3) 10	(4) 25	(5) 50	(6) 75	(7) 90	(8) 99
ALL SAMPLE								
focc3(a)	-0.379***	-0.252***	-0.341***	-0.363***	-0.382***	-0.394***	-0.402***	-0.408***
	(0.010)	(0.039)	(0.015)	(0.014)	(0.011)	(0.011)	(0.017)	(0.040)
focc3(b)	-0.186***	-0.195***	-0.200***	-0.188***	-0.176***	-0.160***	-0.153***	-0.166***
	(0.014)	(0.065)	(0.022)	(0.016)	(0.014)	(0.019)	(0.022)	(0.063)
nwocc3(a)	-1.802***	-1.146***	-1.504***	-1.697***	-1.847***	-1.937***	-2.013***	-1.776***
	(0.029)	(0.095)	(0.041)	(0.033)	(0.034)	(0.039)	(0.047)	(0.132)
nwocc3(b)	-0.467***	0.342	-0.228**	-0.539***	-0.670***	-0.706***	-0.592***	0.081
	(0.051)	(0.261)	(0.102)	(0.060)	(0.060)	(0.066)	(0.089)	(0.207)
FEMALE SAMPLE								
focc3(a)	-0.473***	-0.297***	-0.381***	-0.400***	-0.424***	-0.483***	-0.565***	-0.812***
	(0.015)	(0.060)	(0.029)	(0.021)	(0.019)	(0.019)	(0.028)	(0.087)
focc3(b)	-0.401***	-0.282**	-0.440***	-0.407***	-0.347***	-0.352***	-0.406***	-0.459***
	(0.025)	(0.112)	(0.045)	(0.032)	(0.029)	(0.035)	(0.042)	(0.118)
MALE SAMPLE								
focc3(a)	-0.262***	-0.095*	-0.212***	-0.273***	-0.325***	-0.290***	-0.230***	-0.053
	(0.013)	(0.055)	(0.020)	(0.017)	(0.016)	(0.017)	(0.030)	(0.071)
focc3(b)	0.095***	0.302***	0.147***	0.100***	0.045**	0.038	0.045	0.135
	(0.019)	(0.081)	(0.024)	(0.024)	(0.021)	(0.024)	(0.036)	(0.104)
NON-WHITE SAMPLE								
nwocc3(a)	-1.629***	-0.942***	-1.154***	-1.428***	-1.694***	-1.873***	-1.938***	-1.763***
	(0.044)	(0.142)	(0.060)	(0.051)	(0.053)	(0.058)	(0.076)	(0.180)
nwocc3(b)	-0.224***	1.083***	0.103	-0.232**	-0.591***	-0.541***	-0.345**	0.713**
	(0.077)	(0.374)	(0.128)	(0.099)	(0.086)	(0.103)	(0.138)	(0.311)
WHITE SAMPLE								
nwocc3(a)	-1.899***	-1.241***	-1.702***	-1.821***	-1.912***	-1.967***	-2.052***	-1.934***
	(0.038)	(0.112)	(0.056)	(0.040)	(0.048)	(0.053)	(0.057)	(0.154)
nwocc3(b)	-0.691***	-0.354	-0.495***	-0.695***	-0.753***	-0.936***	-0.819***	-0.661*
	(0.070)	(0.386)	(0.146)	(0.083)	(0.067)	(0.089)	(0.123)	(0.365)

Panel B - year 2006

	(1) mean	(2) 1	(3) 10	(4) 25	(5) 50	(6) 75	(7) 90	(8) 99
ALL SAMPLE								
focc3(a)	-0.093***	-0.019	-0.089***	-0.093***	-0.086***	-0.101***	-0.125***	-0.219***
	(0.007)	(0.019)	(0.011)	(0.007)	(0.007)	(0.009)	(0.014)	(0.055)
focc3(b)	-0.043***	0.006	-0.031**	-0.025***	-0.036***	-0.066***	-0.107***	-0.404***
	(0.010)	(0.033)	(0.013)	(0.009)	(0.010)	(0.012)	(0.018)	(0.058)
nwocc3(a)	-1.753***	-0.696***	-1.187***	-1.343***	-1.649***	-2.003***	-2.304***	-2.503***
	(0.020)	(0.053)	(0.024)	(0.022)	(0.021)	(0.025)	(0.042)	(0.113)
nwocc3(b)	-0.818***	-0.329***	-0.473***	-0.646***	-0.834***	-1.083***	-1.152***	-1.054***
	(0.038)	(0.115)	(0.056)	(0.041)	(0.035)	(0.052)	(0.067)	(0.233)
FEMALE SAMPLE								
focc3(a)	-0.179***	0.069	-0.053***	-0.066***	-0.086***	-0.203***	-0.422***	-0.867***
	(0.013)	(0.042)	(0.018)	(0.013)	(0.013)	(0.017)	(0.024)	(0.081)
focc3(b)	-0.153***	-0.110**	-0.119***	-0.087***	-0.084***	-0.152***	-0.242***	-0.609***
	(0.018)	(0.055)	(0.022)	(0.017)	(0.019)	(0.019)	(0.030)	(0.089)
MALE SAMPLE								
focc3(a)	-0.023**	-0.041	-0.074***	-0.080***	-0.068***	-0.036***	0.029	0.158**
	(0.009)	(0.029)	(0.012)	(0.009)	(0.008)	(0.012)	(0.020)	(0.063)
focc3(b)	0.018	0.093**	0.048***	0.033**	-0.011	-0.044**	-0.080***	-0.305***
	(0.013)	(0.041)	(0.015)	(0.015)	(0.013)	(0.018)	(0.021)	(0.078)
NON-WHITE SAMPLE								
nwocc3(a)	-1.476***	-0.517***	-0.889***	-1.027***	-1.335***	-1.787***	-2.177***	-2.459***
	(0.028)	(0.094)	(0.036)	(0.032)	(0.027)	(0.039)	(0.059)	(0.171)
nwocc3(b)	-0.529***	0.094	-0.166**	-0.432***	-0.672***	-0.944***	-0.951***	-0.994**
	(0.054)	(0.176)	(0.079)	(0.059)	(0.048)	(0.070)	(0.088)	(0.387)
WHITE SAMPLE								
nwocc3(a)	-1.923***	-0.870***	-1.452***	-1.588***	-1.843***	-2.069***	-2.304***	-2.479***
	(0.028)	(0.085)	(0.036)	(0.028)	(0.033)	(0.035)	(0.053)	(0.162)
nwocc3(b)	-1.105***	-0.797***	-0.792***	-0.865***	-0.988***	-1.173***	-1.331***	-1.262***
	(0.055)	(0.168)	(0.079)	(0.057)	(0.056)	(0.072)	(0.107)	(0.361)

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

Note: (a) 3rd specification; (b) 4th specification with occupational dummies.

Table 2: Overview of the main studies estimating the impact of femaleness on earnings (as semi-elasticities)

Authors	Country coverage	Time coverage	Dataset used	Results: semi-elasticities
Johnson and Solon (1986)	US	1978	1978 CPS (workers older than 16)	-0.244*** (women – w/o controls) -0.090*** (women – with controls) -0.343*** (men – w/o controls) -0.168*** (men – with controls)
MacPherson and Hirsch (1995)	US	1973-1993	1973-1993 CPS	-0.068 (women 1973/74) -0.101 (women 1977/78) -0.163 (women 1989) -0.174 (women 1993) -0.148 (men 1973/74) -0.186 (men 1977/78) -0.183 (men 1989) -0.190 (men 1993)
Cotter, Hermsen and Vanneman (2003)	US	1989	1990 PUMS (employed aged 25-54 in 1989)	-0.206*** (White females) -0.231*** (African Amer. females) -0.200*** (Hispanic Amer. females) -0.125*** (Asian females) -0.149*** (White males) -0.193*** (African Amer. males) -0.204*** (Hispanic Amer. males) -0.324*** (Asian males)
Lucifora and Reilly (1992)	Italy	1985	1985 Actual Earnings Survey *(Indagine sulle Retribuzioni di Fatto)*	0.01902 **(females) -0.3220*** (males)
Baker and Fortin (2003)	Canada and U.S.	1987-1988	1987 and 1988 LMAS for Canada and from the 1987 and 1988 CPS ORG for the United States	0.006 [n.s]. (women – Canada – 1987) -0.028 [n.s] (women – Canada – 1988) - 0.228*** (women – US – 1987) - 0.227*** (women – US – 1988) -0.13*** (men – Canada – 1987) -0.145*** (men – Canada – 1988) -0.022 [n.s] (men – US – 1987) -0.028 [n.s] (men women – US – 1988)

Source: Author’s compilation.

Table 3: Quantile decomposition results for gender wage gaps (using the 4th specification), 1987 and 2006

	1987			2006		
Quantile	**0.1**	**0.5**	**0.9**	**0.1**	**0.5**	**0.9**
Raw log gap	0.596	0.303	0.176	0.154	0.044	0.030
Decomposition method: Machado & Mata (2005) as Albrecht , van Vuuren and Vroman (2003)						
Explained	-0.071	-0.016	-0.152	-0.216	-0.109	-0.269
s.e.	0.014	0.008	0.018	0.007	0.003	0.008
Unexplained	0.670	0.349	0.321	0.320	0.191	0.295
s.e.	0.018	0.009	0.022	0.010	0.005	0.012
Total gap (conditional wages)	0.358	0.312	0.314	0.090	0.056	0.047
s.e.	0.005	0.005	0.005	0.003	0.003	0.003
Residual	0.241	0.021	-0.145	0.014	0.026	-0.021
Total gap (predicted wages)	0.599	0.333	0.168	0.104	0.082	0.026
s.e.	0.023	0.012	0.028	0.012	0.006	0.015
Decomposition method: Melly (2006)						
Explained	-0.074	-0.054	-0.086	-0.143	-0.108	-0.210
s.e.	0.010	0.009	0.013	0.006	0.005	0.009
Unexplained	0.549	0.352	0.294	0.220	0.170	0.239
s.e.	0.005	0.004	0.007	0.007	0.005	0.010
Total gap	0.475	0.299	0.208	0.077	0.062	0.029
s.e.	0.008	0.006	0.009	0.007	0.004	0.008
Decomposition method: Machado & Mata (2005) as Albrecht, van Vuuren and Vroman (2009)						
Explained	-0.074	-0.050	-0.084	-0.145	-0.108	-0.212
s.e.	0.024	0.026	0.038	0.027	0.019	0.039
Unexplained	0.547	0.350	0.285	0.218	0.169	0.241
s.e.	0.029	0.025	0.038	0.031	0.018	0.040
Total gap	0.473	0.300	0.201	0.073	0.060	0.029
s.e.	0.031	0.026	0.041	0.031	0.019	0.039
Decomposition method: RIF-OLS regressions (Firpo, Fortin and Lemieux, 2009)						
Explained	-0.068	-0.037	-0.131	-0.174	-0.072	-0.315
s.e.	0.011	0.011	0.021	0.009	0.005	0.013
Unexplained	0.651	0.342	0.334	0.336	0.120	0.317
s.e.	0.015	0.012	0.023	0.011	0.006	0.016
Total gap	0.583	0.305	0.203	0.162	0.048	0.002
s.e.	0.012	0.008	0.014	0.009	0.005	0.011
Expl: age	0.010	0.021	0.035	-0.008	-0.005	0.003
s.e.	0.002	0.002	0.003	0.001	0.001	0.002
Expl: edu	-0.041	-0.099	-0.240	-0.056	-0.083	-0.206
s.e.	0.002	0.003	0.007	0.002	0.002	0.005
Expl: focc3	-0.003	-0.023	-0.091	-0.084	0.065	-0.049
s.e.	0.015	0.014	0.019	0.011	0.007	0.014
Expl: occ	-0.033	0.085	0.180	0.001	-0.032	-0.051
s.e.	0.009	0.010	0.019	0.007	0.005	0.012
Unexp: age	-0.667	0.437	0.225	-0.670	0.496	0.351
s.e.	0.022	0.015	0.033	0.027	0.012	0.030
Unexp: edu	-0.120	0.032	0.168	-0.281	0.036	0.044
s.e.	0.042	0.019	0.023	0.039	0.013	0.024
Unexp: focc3	0.621	0.198	0.261	0.339	-0.090	0.251
s.e.	0.042	0.032	0.049	0.034	0.018	0.041
Unexp:occ	-0.408	0.075	-0.032	-0.149	0.195	-0.647
s.e.	0.065	0.044	0.319	0.046	0.024	0.224

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

Table 4: Quantile decomposition results for racial wage gaps (using the 4th specification), 1987 and 2006

	1987			2006		
Quantile	**0.1**	**0.5**	**0.9**	**0.1**	**0.5**	**0.9**
Raw log gap	0.470	0.463	0.654	0.405	0.349	0.629
Decomposition method: Machado & Mata (2005) as Albrecht , van Vuuren and Vroman (2003)						
Explained	0.326	0.362	0.600	0.337	0.281	0.567
s.e.	0.004	0.003	0.008	0.003	0.003	0.006
Unexplained	0.065	0.074	0.135	0.054	0.052	0.164
s.e.	0.013	0.006	0.011	0.009	0.004	0.008
Total gap (conditional wages)	0.491	0.486	0.504	0.402	0.392	0.461
s.e.	0.004	0.005	0.005	0.003	0.003	0.003
Residual	-0.101	-0.051	0.231	-0.010	-0.059	0.270
Total gap (predicted wages)	0.391	0.436	0.735	0.391	0.333	0.731
s.e.	0.013	0.007	0.013	0.009	0.005	0.009
Decomposition method: Melly (2006)						
Explained	0.348	0.372	0.496	0.308	0.294	0.425
s.e.	0.008	0.006	0.007	0.005	0.003	0.007
Unexplained	0.025	0.095	0.157	0.022	0.062	0.175
s.e.	0.006	0.005	0.007	0.008	0.004	0.007
Total gap	0.373	0.467	0.653	0.331	0.357	0.600
s.e.	0.007	0.004	0.008	0.006	0.004	0.009
Decomposition method: Machado & Mata (2005) as Albrecht, van Vuuren and Vroman (2009)						
Explained	0.352	0.374	0.497	0.307	0.290	0.422
s.e.	0.033	0.022	0.042	0.028	0.019	0.033
Unexplained	0.021	0.091	0.154	0.022	0.063	0.176
s.e.	0.030	0.021	0.034	0.028	0.017	0.031
Total gap	0.372	0.466	0.651	0.329	0.353	0.598
s.e.	0.032	0.022	0.040	0.029	0.018	0.037
Decomposition method: RIF-OLS regressions (Firpo, Fortin and Lemieux, 2009)						
Explained	0.384	0.384	0.499	0.383	0.300	0.454
s.e.	0.008	0.006	0.011	0.007	0.004	0.008
Unexplained	0.018	0.078	0.227	0.016	0.051	0.243
s.e.	0.014	0.007	0.011	0.011	0.005	0.009
Total gap	0.402	0.462	0.726	0.400	0.351	0.696
s.e.	0.011	0.008	0.013	0.009	0.004	0.011
Expl: age	0.016	0.019	0.018	0.011	0.017	0.034
s.e.	0.002	0.002	0.002	0.001	0.001	0.002
Expl: edu	0.111	0.185	0.371	0.098	0.116	0.250
s.e.	0.005	0.004	0.009	0.004	0.002	0.005
Expl: nwocc3	0.018	0.017	0.094	0.002	0.058	0.124
s.e.	0.007	0.006	0.010	0.005	0.003	0.008
Expl: occ	0.038	0.111	0.061	0.051	0.035	0.064
s.e.	0.007	0.006	0.010	0.006	0.003	0.008
Unexp: age	0.255	0.360	0.000	-0.563	0.510	0.035
s.e.	0.014	0.010	0.022	0.019	0.009	0.023
Unexp: edu	0.052	0.096	0.073	0.043	0.184	0.207
s.e.	0.029	0.016	0.020	0.030	0.012	0.021
Unexp: nwocc3	-0.125	0.083	-0.689	-0.745	-0.166	-0.514
s.e.	0.109	0.072	0.129	0.097	0.043	0.115
Unexp:occ	0.006	-0.210	-0.421	0.303	-0.109	0.464
s.e.	0.082	0.060	0.276	0.070	0.035	0.201

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

Table 5a: Detailed OB decomposition for gender wage gaps

	1987								2006							
	1	**t-test**	**2**	**t-test**	**3**	**t-test**	**4**	**t-test**	**1**	**t-test**	**2**	**t-test**	**3**	**t-test**	**4**	**t-test**
Explained	-0.163	-34.021	-0.049	-7.087	-0.050	-6.720	-0.071	-8.518	-0.182	-58.839	-0.152	-37.146	-0.174	-37.826	-0.156	-31.180
s.e.	0.005		0.007		0.008		0.008		0.003		0.004		0.005		0.005	
Unexplained	0.485	95.176	0.371	54.559	0.373	49.013	0.393	47.902	0.243	63.895	0.213	50.667	0.235	46.900	0.216	43.280
s.e.	0.005		0.007		0.008		0.008		0.004		0.004		0.005		0.005	
Total gap	0.322	46.014	0.322	46.014	0.322	46.014	0.322	46.014	0.061	13.152	0.061	13.152	0.061	13.152	0.061	13.152
s.e.	0.007		0.007		0.007		0.007		0.005		0.005		0.005		0.005	
Expl: age	0.023	10.810	0.021	10.947	0.023	10.810	0.021	11.000	-0.003	-1.929	-0.003	-2.333	-0.003	-1.857	-0.003	-2.333
s.e.	0.002		0.002		0.002		0.002		0.001		0.001		0.001		0.001	
Expl: edu	-0.157	-42.297	-0.115	-39.759	-0.160	-42.132	-0.115	-39.724	-0.151	-63.000	-0.102	-53.737	-0.152	-63.208	-0.102	-53.789
s.e.	0.004		0.003		0.004		0.003		0.002		0.002		0.002		0.002	
Expl: focc3					0.117	19.797	-0.042	-4.988					0.009	2.559	-0.007	-1.388
s.e.					0.006		0.009						0.003		0.005	
Expl: occ			0.061	11.472			0.081	12.641			-0.029	-9.767			-0.026	-7.027
s.e.			0.005				0.006				0.003				0.004	
Unexp: age	0.151	3.907	0.052	1.377	0.120	3.147	0.083	7.112	0.206	6.230	0.175	5.394	0.187	5.653	0.175	15.873
s.e.	0.039		0.038		0.038		0.012		0.033		0.033		0.033		0.011	
Unexp: edu	-0.101	-12.354	0.000	0.034	-0.079	-9.634	0.000	-0.007	-0.079	-8.630	-0.025	-2.270	-0.072	-7.859	-0.024	-1.960
s.e.	0.008		0.012		0.008		0.014		0.009		0.011		0.009		0.012	
Unexp: focc3					0.138	10.585	0.324	15.882					0.103	9.923	0.113	7.826
s.e.					0.013		0.020						0.010		0.014	
Unexp:occ			0.156	2.090			-0.065	-0.857			-0.064	-1.382			-0.133	-2.836
s.e.			0.074				0.076				0.046				0.047	

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

Note: We follow the same rationale as for the previous analysis. The 1st specification refers to the baseline specification with age, age squared, years of education, formal, urban and regional dummies. The 2nd specification includes occupational dummies while the 3rd specification includes female occupational intensity. The 4th and most complete specification adds both occupational controls.

Table 5b: Detailed OB decomposition for racial wage gaps

	1987								2006							
	1	**t-test**	**2**	**t-test**	**3**	**t-test**	**4**	**t-test**	**1**	**t-test**	**2**	**t-test**	**3**	**t-test**	**4**	**t-test**
Explained	0.384	69.873	0.399	72.473	0.409	74.309	0.401	72.982	0.320	81.923	0.338	86.538	0.353	90.590	0.344	88.231
s.e.	0.006		0.006		0.006		0.006		0.004		0.004		0.004		0.004	
Unexplained	0.105	18.714	0.091	16.759	0.080	14.618	0.088	16.241	0.093	21.651	0.075	18.317	0.059	14.095	0.068	16.683
s.e.	0.006		0.005		0.006		0.005		0.004		0.004		0.004		0.004	
Total gap	0.489	76.422	0.489	76.422	0.489	76.422	0.489	76.422	0.413	91.689	0.413	91.689	0.413	91.689	0.413	91.689
s.e.	0.006		0.006		0.006		0.006		0.005		0.005		0.005		0.005	
Expl: age	0.019	8.818	0.018	9.211	0.019	9.300	0.018	9.211	0.022	14.800	0.019	14.769	0.020	14.500	0.019	14.462
s.e.	0.002		0.002		0.002		0.002		0.002		0.001		0.001		0.001	
Expl: edu	0.293	73.350	0.211	58.500	0.224	62.194	0.208	57.889	0.215	73.966	0.142	59.000	0.156	65.125	0.138	59.870
s.e.	0.004		0.004		0.004		0.004		0.003		0.002		0.002		0.002	
Expl: nwocc3					0.102	42.417	0.037	9.737					0.099	54.889	0.057	19.552
s.e.					0.002		0.004						0.002		0.003	
Expl: occ			0.102	39.385			0.072	18.487			0.091	45.500			0.046	16.000
s.e.			0.003				0.004				0.002				0.003	
Unexp: age	0.295	8.206	0.278	7.842	0.284	7.994	0.278	37.013	0.192	6.022	0.158	5.042	0.193	6.179	0.153	18.000
s.e.	0.036		0.035		0.036		0.008		0.032		0.031		0.031		0.009	
Unexp: edu	0.091	15.724	0.078	10.427	0.060	8.600	0.076	6.759	0.236	32.764	0.148	17.459	0.137	17.163	0.140	12.972
s.e.	0.006		0.008		0.007		0.011		0.007		0.009		0.008		0.011	
Unexp: nwocc3					-0.134	-4.659	-0.231	-4.497					-0.246	-11.303	-0.317	-7.496
s.e.					0.029		0.051						0.022		0.042	
Unexp:occ			-0.013	-0.185			0.121	1.567			0.109	2.140			0.291	5.129
s.e.			0.072				0.077				0.051				0.057	

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

Note: We follow the same rationale as for previous analysis. The 1st specification refers to the baseline specification with age, age squared, years of education, formal, urban and regional dummies. The 2nd specification includes occupational dummies while the 3rd specification includes non-white occupational intensity. The 4th and most complete specification adds both occupational controls.