An Analysis of Pay and Occupational Differences by Gender and Race in Brazil - 1987 to 2006

Paola Salardi

Thesis submitted for the Degree of Doctor of Philosophy

September, 2012

Department of Economics

University of Sussex

WORK NOT SUBMITTED ELSEWHERE FOR EXAMINATION

I hereby declare that this thesis has not been, and will not be, submitted in whole or in part to another University for the award of any other degree.

Signature:…………………………………………………….

UNIVERSITY OF SUSSEX

PAOLA SALARDI

DEGREE OF DOCTOR OF PHILOSOPHY

An Analysis of Pay and Occupational Differences by Gender and Race in Brazil - 1987 to 2006

SUMMARY

This thesis investigates the magnitude and evolution of gender and racial occupational segregation and wage gaps in Brazil from 1987 to 2006. First, we provide the construction of a new harmonized and temporally consistent re-classification of the occupational codes using the Brazilian household survey, the PNADs. This new occupational classification permits an examination of the evolution of the Brazilian occupational structure over a protracted period of time.

Second, we examine the occupational structure in Brazil assessing both the extent and trends in gender and racial based occupational segregation. We use several well-known indices of segregation (Duncan and Duncan, 1955; Moir and Selby-Smith, 1979; Karmel and Maclachlan, 1988; Silber, 1989) and focus on the evolution over time of the occupational segregation across formal and non-formal labour markets. An attempt is made to assess the main forces driving changes in occupational segregation over time by employing a decomposition of the segregation measures developed by Deutsch, Flueckiger and Silber (2009).

Third, we investigate the magnitude and evolution of gender and racial pay gaps in Brazil by employing several decomposition techniques. Together with the standard Oaxaca-Blinder decomposition, we apply the Brown, Moon and Zoloth (1980) decomposition technique, which allows us to account for the impact of occupational segregation on the wage gap. We explore the impact of the selection process on our decomposition results by employing different parametric corrections (the Heckman (1979) and Lee (1983) corrections). Several sensitivity checks are also implemented and

alternative correction methods investigated such as the non-parametric imputation method by Olivetti and Petrongolo (2008) and the local wage gap estimation by Machado (2011).

Fourth, we attempt to provide a comprehensive portrait of gender and racial wage gaps across the entire wage distribution while exploring the impact of gender and racial occupational segregation on wage determination in the Brazilian labour market. Our analysis particularly focuses on the evolution of the impact of female and non-white occupational intensity on wage outcomes and disparities. We 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), to investigate the determinants of wage disparities at these different points in the wage distribution and to understand how these determinants vary across the wage distribution.

Finally, we offer some concluding remarks, discuss the limitation of the research and provide an agenda for future research on the themes investigated in this thesis.

Acknowledgement

This piece of work has been more than a thesis, but a true journey during which many things have happened around me and inside of me. I have learned to be stronger, more tolerant and more accurate. I have begun to learn the meaning of perseverance and resilience along with the ability to be focused while continuing to cultivate the spirit of joy and playfulness. All of these have been possible thanks to the people that have surrounded me throughout this journey.

My first thanks goes to my supervisors. Thanks to Julie Litchfield and Richard Dickens, for their support, guidance, understanding and encouragement. Particular thanks go to Barry Reilly for being such an inspiring and challenging guide during these years, always present even during the most difficult moments. None of this would have been possible without their constant support.

Outside of my supervisors, I am also very grateful to Hugo Ñopo and Patricia Justino: I have learned a lot from you during our intense work experiences together. The rest of the faculty here at Sussex have also always been of great support. I would particularly like to thank my Phd colleagues, for being friends more than colleagues, both the 'old guard', Marinella, Javier, Stefania, Marta, Chiara, Jairo, Gemechu, Giulia, Outi, Antonio, and the new one, Matteo, Elsa, Kalle, Mattia, Hector and Edgar, among others.

The circle of friends that have inundated me with love during these years are scattered in many places around the word. Thanks to Gaia and Franz, Shandana and Miguel, Nick and Sarah, Gustavo and Mariela, Julian, Vega, Marianna, and Giovanna.

Very special thanks go to my 'magical' friends, for your constant presence and teachings. Together we know that separation is only an illusion. Thanks to Claudio, Elvira, Ada, Silvia, Sonia, Monica, Paola, Genni, Letizia, Lisa, Lorenza, the rest of the 'crew' and the beautiful places we have been and will be together.

I could not be more grateful to my parents and to my three angels, my sisters Stefania and Lucia for their love, support and understanding from a distance and, of course, the sweet and furry one, Lavanda.

And finally a special thanks to Wilson for being such a unique, light-hearted and loving partner, friend, colleague, supporter; in one word, the love of my life. Thanks for the special adventures that we have had together during these years, and "just think ... of all the beautiful, playful, things we will do together when it is done!"

Unless someone like you

Cares a whole awful lot,

Nothing is going to get better.

It's not.

Tables of Contents

Chapter 1

Introduction

This thesis investigates the magnitude and evolution of gender and racial occupational segregation and wage gaps in Brazil over the period from 1987 to 2006. Occupational segregation and wage discrimination are among the central themes of labour economics, but within developing countries there remain significant research gaps, which this thesis attempts to address.

The lack of attention to this issue has meant not only that comparatively few estimates of occupational segregation from developing countries are available, but also that most existing studies of wage discrimination have failed to take account of the potential importance of occupational segregation in explaining observed pay outcomes. The core contribution of this thesis lies in providing a detailed analysis of occupational segregation over a protracted period of time, and then in linking occupational segregation to the analysis of wage discrimination over the same time period.

The analysis focuses on documenting broad trends over time and on investigating the underlying determinants of these trends using a variety of decomposition techniques. The aim of the decomposition techniques is to assess whether patterns of occupational segregation and wage discrimination can be explained by differences in the characteristics of different groups of workers or reflect unobservable and unexplained differences in occupation attachment and wages. The former would suggest unequal access to factors that determine employment (e.g., education), while the latter potentially captures the existence of a discrimination animated by gender or race.

The foregoing represents a novel contribution to the relevant literature, but the thesis goes further by disaggregating the analysis by both race and gender, and into the formal, informal and self-employed sectors. The existing research on labour market discrimination has focused primarily on gender based segregation and discrimination, with less attention paid to its racial dimension. These trends have rarely been considered jointly, and by doing so in this thesis we are able to draw attention to both commonalities and differences, and thus towards emphasize equally the unique challenges associated with combating racial and gender discrimination respectively.

In addition to the disaggregation by race and gender, most of the empirical analysis presented in this thesis considers trends at the aggregate level, but also these patterns within the formal, informal and self-employed sectors of the labour market. It is reasonable to anticipate potentially significant differences in these trends between the formal and non-formal sectors, particularly given the large size of the non-formal sectors in Brazil and elsewhere in the developing world, but most previous analysis has not taken these potential differences fully into account. In exploring these distinctions this thesis reports evidence of significant differences across employment sectors, particularly in the case of gender based segregation and discrimination, and argues for the importance of addressing divergent labour market trends across these sectors.

A central requirement in addressing these issues is high quality data, and the relative absence of such data collected consistently over long periods of time has been an important barrier to research of this nature in developing countries. The Brazilian national household survey, the *Pesquisa Nacional por Amostra do Domicilios* (PNAD), provides a valuable starting point, as it contains data on a wide range of labour market variables over more than two decades. However, the PNAD underwent a major change in the classification of occupations after 2001, which has previously made it impossible to study occupational structure and segregation over a period that straddles this definitional change. Given this challenge, a major contribution of this thesis lies in the construction of a harmonized re-classification of occupational codes from the PNAD surveys from 1987 to 2006 (inclusive), which not only permits compatible analysis over a long period of time than previously possible, but also aligns the codes with the international system of occupational classifications, the ISCO-08, to facilitate international comparisons.

This thesis focuses on the Brazilian case for both pragmatic and conceptual reasons. The pragmatic reason is that Brazil offers the availability of the PNAD dataset. This data source contains a wide array of labour market variables and spans an extended time period. Most importantly, it contains adequately sized and nationally representative samples. It thus offers a unique opportunity to investigate occupational segregation and wage discrimination in significant detail.

Brazil also provides an ideal setting for this research given the combination of high levels of diversity and inequality and the existence of significant changes in political, economic and labour market conditions over time. We are thus able to look not only at the dimensions of gender and racial inequality in the labour market, but also at the ways in which these conditions have evolved against the backdrop of rapid economic reform and political liberalization initiated at the end of the 1980s.

The country is among the most economically unequal in the world, and there are important gender and racial dimensions to this inequality. As with many South American countries, women in Brazil have historically played a comparatively smaller role than men in the labour market, while also being subject to high levels of occupational segregation and wage discrimination. However, the past two decades have seen changes in the role of women and the rapid entry of more women into the labour force. This has been reflected in changes in both the composition of the labour market and in the nature of earnings patterns.

These broad trends in female labour market participation have been relatively widely recognized, but the importance of race to labour market outcomes has been less well studied. In part this reflects the fact that Brazil has historically denied the existence of racial inequality, instead trumpeting the existence of “racial democracy” and a truly multi-racial society. However, despite this historical neglect, the past two decades have witnessed an increasing acknowledgement of significant racial discrimination in the Brazilian labour market, coupled with government policies aimed at curbing discrimination. Against this historical background, there is significant value in further investigating the character of racial segregation and discrimination in the labour market, and also in comparing racial segregation to gender segregation, in order to highlight commonalities and differences between these two phenomena.

We focus on both occupational segregation and wage discrimination, but it is our focus on the importance of occupations that is relatively distinctive, as it has been a comparatively neglected topic in empirical studies of developing countries. This is surprising, as occupational segregation is likely to be both a cause and consequence of economic inequality and social unrest, as well as being a potentially important contributor to overall trends in wage discrimination.

In contrast to the relative lack of attention to occupational segregation, there is an extensive literature on the role of particular industries in explaining labour market phenomena of interest (e.g., efficiency wages, monopoly rents, unionization effects, compensating wage differentials etc.) (e.g., Krueger and Summers (1988) and Fields and Wolff (1995)). However, by focusing on industries this literature concentrates on where people do their jobs, as opposed to the type of jobs they actually do. There has been an increased interest in the role of occupation in determining wage inequality and particularly in relation to skill-biased technological change and the role of international trade for off-shoring (e.g., see Firpo, Fortin and Lemieux (2011) and cited references therein).

A primary motivation for an emphasis on occupations relates to the fact that the jobs individuals do, rather than the industry branch within which they do it, coincide more closely with their interests and concerns. In addition, occupational categories better capture meaningful groups within the labour force with similar qualifications, wages and conditions of work. As such, labour market trends related to occupations are likely to be of more immediate relevance to workers themselves.

A secondary motivation relates to the fact that a focus on occupations permits a better understanding of social and economic outcomes, and these outcomes are most closely connected to occupational attainment. When individuals choose employment they more often choose a type of occupation rather than a particular industry. There are almost no barriers to entry to particular industries, but there are significant barriers to entry to certain occupations (e.g., high-skilled occupations). As researchers, we are interested in the mechanisms that shape occupational attainment, as this is the more meaningful job outcome, and one which has obvious social and economic consequences including occupational segregation. In fact, an early definition of social class, proposed by the British sociologist John Goldthorpe, was based on criteria that were closely related to occupational positions. The Goldthorpe class schema is based on eleven classes grouped into three main clusters—the service class, the intermediate class, and the working class (Goldthorpe, 2000).

Occupational attainment and occupational segregation are central to understanding labour market outcomes and equity. If certain groups face barriers to entering particular occupations this may shape both their earnings potential, and the broader range of professional and social options available to them. Moreover, occupational segregation – reflected in the shares of different population sub-groups in particular occupations – may be an important component of more frequently studied wage discrimination. As such, it is of both academic and policy interest to better understand patterns of occupational segregation, as well as whether that occupational segregation stems from differences in endowments, such as education and skills, or reflects unexplained factors (e.g., discrimination) – that is differences in the ability of similarly qualified individuals to access particular occupations.

The thesis is organized into eight separate chapters. The next chapter, Chapter 2, provides an overview of the key features of the political, socio-economic and labour market environment in Brazil. It highlights key institutional features of Brazil, it traces major political and economic developments over recent decades and provides a preliminary account of the socio-economic context focussing on high levels of inequality, and the gender and racial dimensions to these disparities. The chapter then provides a more detailed overview of key features of the labour market including key disparities in outcomes by gender and race, and contains a brief discussion of the evolution of anti-discrimination legislation (ADL) in the country in order to contextualize the broader trends in occupational segregation and wage discrimination documented.

Chapter 3 provides a detailed description of the dataset employed in the empirical analysis, the *Pesquisa Nacional por Amostra do Domicilios* (PNAD). We first discuss the rationale for relying on the PNAD relative to other available data sources. We then focus on describing the temporal coverage, sample size, sampling procedure and general structure. Finally, we provide a description of the key variables available that are relevant to this study.

Chapter 4 begins by describing the construction of a new harmonized re-classification of occupations for the Brazilian national household surveys (the PNAD) from 1987 to 2006. The creation of a compatible re-classification of the occupational codes represents a key contribution of this thesis, as it overcomes discontinuities generated by changes in the original PNAD classification. Given our focus on occupational segregation, and its relationship to wage discrimination, this re-classification is essential to the investigation of labour market trends over a longer period than previously been possible, while making these findings internationally comparable. The remainder of the chapter is devoted to investigating occupational structure in Brazil over time, examining trends disaggregated by both gender and race and separately within the formal, informal and self-employed labour markets.

Chapter 5 presents an analysis of the evolution of gender and racial occupational segregation over time. It first assesses the magnitude of occupational segregation both by gender and race using several well-known indices of segregation, and explores trends in segregation separately within the formal, informal and self-employed sectors. The analysis of occupational segregation is then further disaggregated by several key characteristics of the labour force in order to identify specific demographic, educational, sectoral and spatial patterns. Finally, the chapter presents an initial discussion of possible determinants of occupational segregation over time by exploiting a decomposition technique developed by Deutsch, Flueckiger and Silber (2009).

Chapter 6 investigates the magnitude and evolution of gender and racial wage gaps in the Brazilian labour market over time with particular reference to the role of occupational segregation. We begin by employing the standard Oaxaca (1973) and Blinder (1973) decomposition as well as the variant proposed by Brown, Moon and Zoloth (1980). The Brown, Moon and Zoloth (1980) decomposition technique is significant in that it incorporates the role of occupational segregation when decomposing earnings differentials, thus drawing a link to the analysis presented in the previous chapter and addressing an important limitation of most existing studies. We control for selection bias by applying the parametric Heckman (1979) and Lee (1983) correction methods to both decomposition techniques, and also experiment with alternative correction methods proposed by Olivetti and Petrongolo (2008) and Machado (2011) in order to further assess the robustness of our key results. The analysis is conducted for the overall labour market with further analysis conducted for the formal and non-formal sectors.

In Chapter 7 we attempt to provide a comprehensive portrait of gender and racial wage gaps across the entire wage distribution. We thus move beyond the standard Oaxaca and Blinder (1973) decomposition by estimating the evolution of gender and racial wage gaps in Brazil over the last two decades at different quantiles of the wage distribution using decomposition methods recently developed by Machado and Mata (2005) and Melly (2005, 2006), and by Firpo, Fortin and Lemieux (2009). We also focus particularly on the impact of female and non-white occupational intensity on gender and racial wage differentials respectively. This emphasis provides a novel contribution to existing research for developing countries. As with previous chapters we also briefly disaggregate the analysis between the formal and non-formal sectors in order to highlight any divergent trends.

Finally, Chapter 8 summarizes the contribution of the research, offers some concluding remarks, discusses the limitations and provides an agenda for future research on the themes investigated in this thesis.

Chapter 2

The Context

This chapter lays out the context for the analysis to follow by providing a description of key features of the political, socio-economic and labour market environment in Brazil. This thesis focuses on the Brazilian labour market, with an emphasis on occupational segregation and wage discrimination, these issues are best understood in relation to a broader country context. Thus, this chapter is divided into five sections. The first section provides a broad historical background, highlighting the key characteristics of the Brazilian Republic, alongside the key political and economic developments in recent decades. The second section provides a more detailed account of the socio-economic context, highlighting the geographic, economic and ethnic diversity in the country, with a focus on high levels of inequality, and the gender and racial dimensions to these disparities. The third section turns in greater detail to the Brazilian labour market, flagging key features of the labour market including disparities in outcomes by gender and race. The fourth section discusses the evolution of anti-discrimination legislation in the country in order to help contextualize the broader trends in occupational segregation and wage discrimination. Finally, the last section provides some conclusions.

2.1 Historical and Political Background

Brazil is among one of the largest countries in the world and occupies half of the entire land mass of South America. In 2006, which is the final year covered by this research, the Brazilian population was estimated at 188 million, making Brazil the fifth largest nation in the world, while according to the World Bank[1] Brazil is also the seventh largest economy in the world with a GDP of US$2.2 trillion in 2011.

[1] http://www.worldbank.org/en/country/brazil.

Politically, Brazil achieved independence from Portuguese rule in 1822, after which it passed through a period of Monarchical rule until the declaration of a Republic in 1889. Much of Brazil's political history has been punctuated by military rule, but civilian administrations have been in place since 1985. This transition to civilian rule is reflected in the current Constitution (the country's seventh) which was established in 1988 as the product of a two-year process that followed the handover of power by the military after 21 years of military dictatorship. Under the current constitution Brazil is a federal republic formed of 27 states, one of which is the *Districto Federal* where the capital Brasilia is situated.

The new era of democracy began in 1985 during a period of high inflation and economic stagnation, with the 1980s known as the "lost decade" as countries across South America experienced economic and political turmoil. In contrast, the 1990s came to be known as the "decade of the reforms", combing three major macroeconomic policies: financial and trade openness, fiscal adjustment and stabilization. The Collor government began the process of trade reform at the end of the 1980s, while the Cardoso administration advanced economic stabilization, which was consolidated after the last devaluation plan (the *Plano Real)* undertaken in 1994, and was followed by fiscal adjustment in 1998 (Baer, 2008: 99).

Following this period of liberalization and stabilization, a new era for Brazil began with the election of President Luiz Inácio Lula da Silva in 2002 with a clear agenda for economic growth, equity and social inclusion. It was a period of significant adverse external economic instability, which posed critical social and economic risks, but the administration succeeded in building credibility through tight fiscal and monetary policies and economic growth reached 5.2% in 2005, the highest in ten years. Since then Brazil has experienced a period of relatively stable economic growth, low inflation and fiscal balance. This growth has been accompanied by a range of social initiatives, including an increase in the minimum wage, the *Fome Zero* (Zero Hunger) program, designed to give each Brazilian three meals a day, and *Bolsa Família* (Family Allowance) program, which has focused on cash transfers as part of the *Fome Zero* network (Baer, 2008: 163).

In 2011 Lula was succeeded as president by Dilma Vana Rousseff, the country's first female President, and a symbol for many of the changing role of women in the country. Despite the successes of the previous decade, she faces several important challenges including the appreciation of the Brazilian currency, a need for new infrastructure investment, problems in the education system and, most relevant to this thesis, urgent problems of social inequality.

2.2 The socio-economic context and extent of inequality

While Brazil is among the largest countries in the world, and has achieved significant economic gains over the past two decades, it is equally notable for its striking diversity. Brazil is home to remarkable geographical and climatic variations as well as a hugely diverse population: indigenous tribes, white descendants of Europeans, a black population that arrived during the era of slavery, and a large Asian population that has grown in the wake of successive waves of immigration. While this diversity has the potential to provide the foundation for a dynamic and multicultural nation, such diversity also provides fertile conditions for social and economic inequalities, and Brazil is well-known for its very high levels of income inequality (see among others, Langoni, 1973; Hoffman, 1989; Litchfield, 2001; Barros et al., 2006).

In 2002, Brazil was the eighth most unequal country in the world, based on Gini Index calculations conducted by the UNDP, which found a Brazilian Gini value close to 0.6 (UNDP, 2002). However, this ranking in many ways understates the uniquely high levels of Brazilian inequality. The six most unequal countries in the ranking are all very small African countries, with economies less that 0.1% the size of Brazil's, while the only large country more unequal than Brazil is South Africa, where inequality is the product of several decades of state sanctioned discrimination under apartheid. This inequality is reflected in the existence of widespread poverty, as in 2001 more than one-fifth of the population was still living on less than US$2 a day (World Bank, 2001). Notably, high levels of inequality appear to be closely linked to issues of gender and race, as women and non-whites in Brazil are disproportionately represented at the bottom end in the income distribution, as is more generally the case throughout Latin America and the Caribbean (Ñopo, 2012).

However, while Brazil is characterized by high levels of inequality and poverty, it has also experienced significant improvements over the past two decades, with the last decade in particular buoyed by its improving economic performance. The share of the population living on less than US$2 a day declined from 21% in 2003 to 11% in

2009 (World Bank, 2011). During the same period the Gini index declined by 4.4 points from 0.591 in 2002 to 0.547 in 2009. It is this combination of very high levels of inequality and significant changes over time that makes Brazil a particularly interesting case in which to investigate the importance of gender and race in shaping broader patterns of inequality. More importantly, while both women and non-whites have historically been overrepresented at the lower end of the income distribution, the specific challenges confronted by these groups differ significantly, and thus provide an interesting basis for comparison.

Looking first at gender inequality the existing research has documented the fact that in Brazil women represent the majority of the poor, female labour market participation is significantly lower than that for men and, most notably, those women that do have a job receive disproportionately lower salaries and limited social protection (Soares and Inaki, 2002; Wajnman and Rios-Neto, 2000b). Strikingly, women represent 98% of the Brazilian paid domestic workforce, and only 40% of these 12 million women enjoy any kind of employment benefits or social security (ILO, 2010; UNIFEM, 2006: 66). Women also represent the majority of informal sector workers, and the majority of unremunerated workers within households are involved in agricultural activities. These inequalities are still more pronounced if we focus specifically on non-white women. Ribeiro (2008), for example, reports that 32% of Brazilian households are headed by females, while 39% of these are headed by non-white women. Among these non-white female headed households 18% have failed to complete primary education, while 52% report being victims of violence and sexual abuse.

Similar degrees of inequality are in evidence in areas beyond the labour market. Maternal health remains a major challenge in Brazil, where the maternal mortality rate declined from 96 to 67 deaths per 100,000 live births between 1995 and 2005, thus falling far short of the Millennium Development Goal of reducing maternal mortality by 75%.[2] These broader trends are reinforced by enormous regional differences, with better health care outcomes concentrated in the south, while other regions have outcomes closer to those of low-income countries in Africa. Within the political sphere, although the right to vote was extended to women in 1934 (UNIFEM, 2006: 36) and women represent more than a half of the Brazilian electorate (UNIFEM, 2006: 39), they currently occupy less than 10% of elected positions. This leaves Brazil among those

[2] WHO (2011), Global Health Observatory (GHO), Brazil country profile available at http://www.who.int/gho/countries/bra/country_profiles/en/.

countries with the lowest female shares in public office in the world, in contrast to the political power of the Brazilian feminist movement outside of electoral politics (UNIFEM, 2006: 41).

These gender differences are rooted in cultural and social norms in Brazil, but changes in attitudes and in access to opportunities have led to declining disparities over time, mirroring similar trends in many Latin American and Caribbean countries (Ñopo, 2012). The 1991 Census revealed that Brazilian women had surpassed men in terms of years of education attained (Beltrão, 2003), while their participation in the labour market has also consistently increased over these two decades. According to the ILO,[3] female labour market participation among women aged 25-54 increased from 51% in 1990 to 72.5% in 2010. Socially, although Brazil is one of the most Catholic countries in the world, 80 percent of Brazilian women of childbearing age use some form of contraception, and the size of Brazilian families is getting smaller, with the fertility rate dropping to 1.9 children per woman by 2010 (World Bank, 2011).

The political and public position of women is also changing. The *Relação Anual de Informações Sociais* (Rais), which is an important firm survey undertaken by the Ministry of Labour, reports that 24% of the 42,276 CEO positions in Brazil were held by women in 2006 (UNIFEM, 2006: 73). The current CEO of *Petrobras-Petróleo Brasil,* the largest company in Brazil and the fourth largest company in the world, is a woman, Maria das Graças Silva Foster.[4] Finally, as noted earlier, Dilma Vana Rousseff is currently the 36th President of Brazil, and the first woman to ever hold the office.

Turning to racial inequalities, we see similarly large disparities, though the details of these inequalities are different from gender. Brazil is a particularly interesting country in which to investigate these racial inequalities, as it has the second largest African-origin population in the world after Nigeria, with almost 150 million Brazilians having African roots (Lovell, 1994). This is a legacy of the long history of slavery, abolished in 1888 by the *Lei Áurea* ("Golden Law").[5] However, unlike other countries involved in the slave trade, where racial divisions have remained relatively explicit, these boundaries are much more fluid in Brazil, which is more often described as a multi-racial, rather than bi-racial, society (Telles, 2006). This reflects widespread

---

[3] ILO (2011), The Key Indicator of the Labour Market (KILM) dataset available at http://kilm.ilo.org/kilmnet/.

[4] Time article (2012), The 100 Most Influential People in the World, Read more: http://www.time.com/time/specials/packages/article/0,28804,2111975_2111976_2111991,00.html#ixzz25akle03w

[5] Brazil was the last Western nation to abolish slavery.

"miscegenation"[6] in the wake of slavery, which contributed to the blurring of group boundaries. Bailey (2002) argues that this blurring of racial divisions facilitated the emergence of an ideology of "racial democracy",[7] which created a common belief that Brazil does not have the racial tensions that exist elsewhere, which has, in turn, contributed to the comparative absence of research studies on racial discrimination.

This popular denial of the existence of racism as a social phenomenon is contradicted by the reality of Brazilian socio-economic conditions, as there remain significant differences in socio-economic outcomes for white and non-white groups. The non-white population – including those identified as "black" and "brown" - is almost half of the entire population, but is significantly overrepresented at the bottom end of the income distribution.[8] Significant discrimination and segregation is similarly apparent in the labour market, with Osorio (2008) ascribing these patterns to a combination of weaker educational outcomes, poorer initial conditions and socio-historical constructs about race.

Divergent outcomes for the non-white population are illustrated across a wide range of socio-economic dimensions, including childhood opportunities, educational attainment and labour market prospects. A study by the sociologist Telles (1992a) captures residential segregation by race, with members of the black and brown populations more likely to be poor than whites. Lovell (2000) highlights similar regional segregation, with the non-white population concentrated in the less dynamic and more economically depressed North and the North-East regions. Lovell and Wood (1998) highlight consistently lower life expectancy for non-whites, higher school enrolment rates among whites, and consistently lower levels of aggregate educational attainment among non-white Brazilians. In the same vein, a study by Beltrão (2003) of educational patterns from 1940 to 2000 highlights a clear ethnic hierarchy with Asian and white populations at the top and mixed race, black and Native American populations in the lower half of the hierarchy. These educational disparities are again highlighted by Gradin (2007), who argues that the considerable racial poverty gap is primarily explained by differences in observed human capital assets, and particularly

---

[6] The myth of racial democracy is based on the miscegenation phenomenon (i.e., the mixing of different races that involved Europeans, African-origin population and indigenous people). Miscegenation played a central role in creating a multi-racial society in Brazil (Skidmore, 1985, 1993; Telles, 2006).

[7] Freye (1933) formalized the ideology of racial democracy in his famous novel entitled "Casa Grande e Senzala".

[8] According to 1999 PNAD, 54% of Brazilians were white while 39.9% brown and 5.4% black and in most recent years the proportion of non-white population has increased. African-origin Brazilians are viewed as the second biggest African nation in the entire world (Henriques, 2001).

educational attainments, resulting from unequal access to high-quality education. Osorio (2008) documents a corresponding lack of social mobility among black Brazilians over the last three decades, with black individuals entering the labour market and into less ambitious and lucrative occupations, disproportionately located in the informal labour sector. Beltrão and Teixeira (2004) echo this finding, highlighting the fact that professions with less social and economic prestige are more likely to be dominated by non-whites.

2.3 The Brazilian labour market

We see significant inequality of both opportunity and outcomes across a wide range of socio-economic indicators. Labour market inequalities are central to understanding these broader outcomes as they act as both a reflection and cause of broader economic disparities. As such, it is useful to present a more detailed discussion of the features of the Brazilian labour market, as a precursor to the more detailed analysis to come. At a broader level, the Brazilian labour market exhibits very distinctive but also contradictory features. It is characterized by high job turnover, a low unemployment rate, a low minimum wage, significant inequality of earnings, and a high level of litigation in labour courts. In addition, there is also a large informal labour market (World Bank, 2002a).

According to the World Bank (2011), there was a total Brazilian labour force of 62 million in 1990, which increased to 95.6 million in 2006, the final year of this study and 101.5 million in 2010. This rapid increase in the labour force has been accompanied by a rising labour force participation among women, which increased from 44.5% in 1990 to 58.7% in 2006, while labour force participation among men declined from 85.3% to 81.6%. These data not only cover a period of changing labour force participation but of significantly broader economic change Brazil has moved from being a closed and public-sector dominated economy to a more open and more private oriented economy, while experiencing rapid economic growth exceeding 5% of per capita GDP annually over the past decade (Baer, 2008: 179).

Given these changes, and the pace of economic growth, it is important to consider the broader labour market changes that have occurred. Similar to many countries around the world over the course of its history Brazil has shifted progressively from the agricultural sector to the industrial sector with, most recently, a greater focus on the services sector. There has been a significant expansion of the service sector over the past two decades, as it has absorbed much of the expanding labour force, and increased its share from 51.6% in 1987 to 60.7% in 2010. During the same period the agriculture and industrial workforces have contracted respectively from 24.6% to 17% and from 23.4% to 22.1% (World Bank, 2011).

The macroeconomic reforms undertaken in Brazil, beginning in the late 1980s, radically altered the nature of labour demand and the level of wages and employment. These changes reflect the effects of three major reforms enacted during the 'decade of the reforms': openness, price stabilization and fiscal adjustment. Price stabilization and fiscal adjustment are held to have had quite different effects, including lower wage flexibility and an increase in the evasion of income and labour taxes (World Bank, 2002a).

Research has suggested that trade liberalization has had three main consequences: an increase in the price of the non-tradable goods, increased productivity in tradable goods and a rise in informality (Ferreira, Leite and Wai-Poi, 2007; Green, Dickerson and Abache 2001; Goldberg and Pavcnik 2003; World Bank, 2002a). There has been significant attention to the impact of trade liberalization on informality. Bosch, Goni and Maloney (2007) explored several factors including trade liberalization, which help to explain the expansion of the informal sector over the last two decades. However, research findings have not been universal in this area, with Goldberg and Pavcnik (2003) finding no impact of trade liberalization on informality. In an interesting addition to earlier arguments, a recent paper by Gaddins and Pieters (2012) has also explored the impact of liberalization on female participation. They argue that declining trade protection was associated with rising female labor force participation and employment with a lag of about two years. They attribute this to both 'push' and 'pull' factors, including a sectoral shift to services, increased male unemployment and increased economic insecurity.

These influences highlight the close relationship between macroeconomic reforms and labour market outcomes, and it has been argued that, given these close relationships, macroeconomic reforms should be accompanied by broader changes in labour market policies. The World Bank, for example, suggested lowering severance costs and strengthening income support programs, reforming collective bargaining and minimum wage legislation, reforming payroll taxes and social security and improving the enforcement of labour regulations (World Bank, 2002a). However, despite the close links between macroeconomic and labour market policies, in practice the dynamism experienced in the realm of macroeconomic reforms has not been replicated in relation to labour policies, since Brazil is yet to enact and enforce many of the reforms suggested. This is largely explained by a very institutionalized labour market and obsolete legislation. This legislation is difficult to alter because changes in labour laws often demand an amendment to Constitutional laws, which are more difficult to implement. The most important source of labour legislation, the *Consolidacão das Leis do Trabalho* (the Consolidated Labour Code, CLT), has remained essentially unaltered for the last 60 years.

Against this general reform background, a closer look at the quantity, quality and composition of the labour supply, highlights significant changes over time, and correspondingly important challenges on the horizon. First, the working age population has been growing faster than employment from the beginning of the 1990s onwards, generating a problem given insufficient job creation. This is reflected in a declining overall labour market participation rate, though this reflects not only the increasing relative size of the working age population but also changes in the characteristics of the economically active population (World Bank, 2002a). On the one hand, individuals are entering the labour market later, as they opt to invest more in education in response to new opportunities - clearly a good sign for the country. On the other, part of the decline in the economically active population appears to reflect an increase in the reservation wage, as individuals do not consider work more rewarding than alternative pursuits (World Bank, 2002a). It is important to highlight that the decline in participation appears to be concentrated among male workers, as female participation has increased consistently over this period, with the ratio of female to male labour participation rates increasing from 52.16% in 1990 to 71.93% in 2006 (World Bank, 2011).

Alongside these broad changes in participation rates, there have been significant changes in the allocation of the workforce across different economic sectors. There has been a sizeable contraction in the agricultural workforce, though this has occurred relatively more for male workers, with a decline of 3.2 percentage points for women and 8.9 percentage points for men. The shares of men and women working in the industrial sector has been stable over time, implying that the decrease in the share of the workforce in agriculture has been exactly offset by an increase in the share of both men and women employed in the services sector. While there has thus been a large increase in the share of men employed in the services sector, it is still women who work primarily in the service sector, with almost three-quarters of service sector employment female in most recent years.

Just as there have been shifts in patterns of labour force participation, there have also been changes in labour quality over time. The demand for more highly skilled workers has increased since the 1980s, as reflected in higher returns to educations and particularly to higher education. Arriagada and Ziderman (1992) estimated a rate of return to vocational education of roughly 22 percent, which was broadly similar to that for academic education. Strauss and Thomas (1996) present and summarize evidence of high positive returns to schooling in Brazil, as well as of a non-linearity in these returns, as the returns to secondary education are higher than the returns to primary. More recently, Curi and Menezes-Filho (2007) find an overall positive elasticity equal to 0.3 between school attainment and earnings. Freguglia et al (2011) take a step back and investigate the impact of expanding public expenditure on education. The study reports, not surprisingly, that expenditure on education increase the propensity of individuals to undertake more education, which, in turn, allows those individuals to shift into the more skilled occupations.

However, despite increasing returns to higher education, Binelli, Meghir and Menezes-Filho (2009) find that the actual progression of students into college has been declining. They show that this is not the result of a lack of highly skilled job opportunities, as the demand for highly-skilled workers has been steadily increasing. Instead, the primary cause of declining educational progression after high school appears to be the contraction of the resources invested in intermediate level schools. Brazil is thus facing a situation in which there is an increasing need for highly-skilled workers, but in which the actual workforce is less educated than required, in significant part due to insufficient spending on education and training (Baer, 2008: 172).

Finally, alongside these broad patterns of change in the labour market, it is important to highlight a number challenges that are rooted in the more particular institutional features of the Brazilian labour market. The World Bank (2002b) provides a useful summary of a range of such issues, including high levels of litigation, the need for more effective minimum wage legislation, the role played by unions and the degree of informality of the labour market. While it is beyond the scope of this study to present a detailed discussion of all of these issues, the first and last merit additional comment.

High levels of litigation brought before the system of labour courts is among the most burdensome characteristic of the Brazilian labour market. The labour market is very dynamic, with a high rate of firing and hiring, however the obsolete labour legislation has meant that a huge range of labour disputes are brought before the labour courts, where all labour disputes must be resolved. Brazilian labour courts had to manage a remarkable two million complaints in 2000 compared to 17,000 in the U.S (World Bank, 2002a). This is a symptom of a highly institutionalized labour market with obsolete legislation, but this legislation remains difficult to change because, as noted earlier, it requires amendments to the Constitutional law.

The other particularly striking and important feature of the Brazilian labour market is the persistently high levels of informality. Carneiro (1997) has argued that in 1990 about one-half of the economically active population was employed in informal activities. Soares (2004) similarly calculated that in 1999 only 14 out of 36 million private sector workers were in the formal sector. Over the past two decades the aggregate size of the informal sector has remained relatively stable, but this aggregate picture disguises significant compositional changes over time. Bosch, Goni and Maloney (2007) documented a substantial increase in the size of the informal private sector in urban areas, which they estimate to have increased by 10 percentage points during the 1990s. In addition, several studies have documented an increase in the 'degree of formalization' of certain occupations and sectors, and most notably among domestic workers and agricultural workers.[9] Ultimately, Ramos and Ferreira (2005) argue that the finding of an increase in informality in Brazil depends heavily on restricting the analysis to metropolitan areas, while a broader focus yields a more complex pattern.

There is broad agreement that the informal sector is large, and also significant evidence of differing patterns in different segments of the labour market. However, there remain significant questions that demand more precise analysis. This begins with the fact that defining the "informal sector" is itself difficult, so much so that in a recent study Henley, Arabheibani and Carneiro (2009) compare three different definitions of informality and report that only 40% of cases are classified as informal across all three

[9] See Fonseca & Rayp (2011) on the formalization of agricultural workers and ILO (2010) on the formalization of the domestic workers.

definitions. Given the importance to this study of distinguishing labour market trends in the formal and informal sectors, it is thus important to briefly discuss these definitional challenges.

Under Brazilian law, as defined in the Consolidated Labour Code (CTL) and the Constitution, employment as a formal worker requires a *carteira de trabalho* (working card) signed by the employer. The majority of studies on informality in Brazil treat all those with a signed working card as formal (Carneiro, 1997; Soares, 2004; Ulyssea, 2005). Any workers without such a card are considered informal based on a strict interpretation of the law. This involves grouping together self-employed workers and informal wage workers, as both are *sem carteira de trabalho assinada* (without a 'signed working card') and both might not pay social security contributions. However, this aggregation is potentially misleading as these two categories have very distinctive characteristics. While informal wage workers differ from formal waged employees primarily in terms of the absence of a signed labour card, the self-employed category is more heterogeneous, as it includes both the unskilled and highly-skilled workers defined as 'own account'.

In this study we opt to distinguish between self-employed workers and informal wage workers. If we disaggregate the labour force into employers, employees and the self-employed we find that in 2000 4.5% of the workforce was an employer, 30.3% were self-employed and the remainder employees. While the share of employers is relatively stable over time, the self-employed share is increasing slightly over time, having reached 31.7% in 2006 (World Bank, 2011). In this study we exclude employers and focus on a combination of employees (with and without a signed working card) and the self-employed. The "formal sector" thus includes all employees who declare that they possess a signed working card for their current primary occupation, the "informal sector" includes all wage earners who do not possess a signed working card and we retain a separate category for the "self-employed sector". In the subsequent analysis in later chapters, we frequently refer to the latter two categories collectively as the "non-formal sector" but by retaining two distinct categories we are also able to highlight important differences between the informal sector and the self-employed. Given the centrality of these questions of informality to this thesis we return to these questions of definition and measurement in greater depth in chapter 5.

2.4 The role of the anti-discrimination legislation in Brazil

Having highlighted not only the broad labour market trends but also significant labour market inequalities by gender and race it is useful to conclude the discussion with a brief review of Brazilian anti-discrimination legislation. This reflects the fact that anti-discrimination legislation may play an important role in shaping patterns of labour market inequality, alongside broader macroeconomic reforms and cultural changes.

In reviewing this legislation it is important to note that under the Brazilian federal system anti-discrimination laws are issued at each of the federal, state and municipal levels. Enforcement of legislation may thus occur at both state and federal levels. At the federal level, judicial powers are exercised by the Federal Supreme Court (*Supremo Tribunal Federal*), the Superior Court of Justice (*Superior Tribunal de Justiça*), five regional courts and a group of specific courts (i.e., for electoral, labour, military issues). This broad structure is mirrored at the state level, with each state judiciary headed by a State Court of Justice. One of the consequences of this federal structure is that understanding the role of anti-discrimination legislation is comparatively complex, as both legislative realities and enforcement vary significantly across states despite a common set of federal laws.

Against this institutional background we can look at anti-discrimination laws targetted at gender and race in turn, while a longer list of relevant legislation is contained in table A1 in the Appendix. Efforts to advance legislation to support gender equity goals are relatively recent, and are connected to the broader international women's movement, as reflected in the World Conference on Women held Mexico in 1975, Nairobi in 1985 and Beijing in 1995. There earliest significant legislative acts were the creation of the National Committee for Women's Rights in 1985, which was followed by the inclusion of a range of important anti-discrimination clauses in 1988. These initial steps were accompanied by the ratification of important international treaties and followed by a variety of additional steps to further outlaw gender discrimination, including the possibility of discrimination against pregnant women. More recently, the position of State Secretary for Women's Rights (*Secretaria de Estado dos Direitos da Mulher*) was created in 2002 and subsequently transformed into the Special Secretary for the Policies for Women in 2003 (*Secretaria Especial de*

*Políticas para as Mulhere*s, SPM), which oversees a comparably titled secretariat. This was, in turn, followed by the creation of a National Plan of Policies for Women (*Plan Nacional de Políticas para las Mujeres*, PNPM) in 2004, and a follow up plan in 2008. These political institutions have played an important role in recent Brazilian political debate and reflect increased government commitment to addressing gender discrimination (UNIFEM, 2006).

The history of legislation aimed at addressing racial discrimination has a somewhat longer and more complex history, while there has similarly been increasing government action since the promulgation of the 1988 Constitution. After the abolition of slavery in 1888, Brazil did not take any measures to organize the millions of former slaves in order to aid their integration into society. This was, in some sense, a precursor to the adoption and wider acceptance of the ideology of racial democracy, with Brazil held to be free of racial segregation and discrimination. This ideology was widely accepted throughout Brazil's history, but it has been increasingly appreciated that Brazil has long had a double identity: a formal identity, in which everyone is Brazilian first, without any racial distinction, and a concrete, material identity, in which skin colour and ethnic identity (i.e. as a white, brown, black or indigenous person) have profoundly shaped individual behaviour and the array of opportunities available (Santos and Silva, 2006).

Progress towards the recognition of this reality dates to the Alfonso Arinos Act of 1951 (Law No. 1390/51), which officially recognized the existence of racial discrimination. However, it was only later, in 1988 that the new Constitution designated racial discrimination as a crime, and even then this occurred only after a protracted debate. Supporters successfully argued that although half of the Brazilian population was black, or had black ancestors, blacks still suffered from the denial of their full rights of citizenship (IACHR, 1997: 132). The inclusion of a reference to racial discrimination in the Constitution established it as the most important legislative document aimed at addressing racial discrimination.[10]

[10] The third article of the Constitution affirms its commitment to promoting the well-being of all citizens without any distinctions (where race and skin colour are included in such prejudices). Article 4 says that "the Federative Republic of Brazil follows specific principles in its international relations wherein the repudiation of terrorism and racism is mentioned". Article 5 claims that "all citizens are equal before the law regardless of class and that all Brazilians and foreign nationals residing in the country are guaranteed the inviolable rights to life, liberty, equality, security and property as set out below: XLI - The law will punish all acts of discrimination against fundamental rights and liberties and XLLI – Racism is a crime for which bail is not available and which is punishable by imprisonment pursuant to law" (IACHR, 1997: 133).Finally the sixth article of the Constitution guarantees social and economic rights and section XXX

Since 1988 the basic guarantees contained in the Constitution have been reinforced through a number of subsequent laws aimed at modifying, improving and strengthening implementation of these anti-discrimination goals, and through the ratification of international declarations. This effort has been further supported by the adoption of a number of affirmative action policies aimed at proactively addressing historical discrimination.

In 1989 the Caò Law No. 7716/89, or "Act against Racism", was enacted, as complementary legislation to the criminal code. It deals with crimes resulting from prejudices based on race or skin colour. The Caò Law is the most important legislative act after the Constitution of 1988 in the fight against racism.[11] Four year later, in 1992, the Brazilian government ratified the American Convention of Human Rights, also known as the "Pact of San Josè", which had originally been promulgated in 1969. In 1995 Brazil subsequently authorized an inspection by the Inter-American Commission of Human Rights, which had first been requested in 1989. On 20th November of the same year, the Inter-ministerial Group for Appreciation of the Black Population was created by Presidential decree. Soon after, on 20th March 1996, the Ministry of Labour created the Working Group for the Elimination of Discrimination in the Workplace, (*Grupo de Trabajo para la eliminación de la discriminación en el empleo y la ocupación,* GTEDEO), by decree, while ILO Convention 111 was also ratified. Finally, in 1996 the first National Human Rights Program (*Programa Nacional de Dereitos Humanos*, PNDH), was established.

Reflecting an increasing government engagement with issues of racial discrimination, Brazil was one of the principal actors during the UN World Conference against Racism in Durban in 2001. Brazil subsequently renewed its commitment by creating the PNDH II in 2002 along with a comprehensive national affirmative action plan. This rapid legislative action continued in 2003 with the creation, at Ministerial rank, of the Special Secretariat for Policies to Promote Racial Equality (*Secretaria de Políticas de Promoción de la Igualdad Racial*, SEPPIR), which has been followed by the creation of the National Plan for the Promotion of Racial Equity in 2009 (Plan Nacional de Promoción de la Igualdad Racial- PLANAPIR).

---

specifically prohibits differences in wages, job tasks and selection criteria based on gender, age, skin colour or civil status.

[11] Several laws to modify and put in force the Caò Law of 1989, such as the Federal Law No. 8081/90, the law No. 8882/94 and the law No. 9459/97, which establishes punishment for the crimes arising from discrimination on the basis of race and colour but also religion and nationality.

Ultimately, after a long history of neglect, the past two decades have witnessed accelerated legislative action at the national level aimed at addressing racial discrimination, and the legacies of such discrimination. This has been reflected not only in anti-discrimination legislation, but also in the ratification of international conventions and in the development of affirmative action programs and specialized institutions. It is also reflected in growing efforts aimed at the enforcement of these rules using workplace inspections.[12] Given the complex institutional structures of a large federal country like Brazil these enforcement efforts remain difficult but a particularly important part of the country's anti-discrimination effort.

2.5 Conclusions

This chapter provides the background to the analysis of occupational segregation and wage discrimination in subsequent chapters. Ultimately, Brazil is a particularly interesting country in which to investigate these questions owing to the combination of extremely high levels of inequality, but also a rapid pace of economic, political and institutional change since the 1980s. Despite an ideology that long interpreted Brazil as a multi-racial society largely unaffected by discrimination, in practice Brazil continues to have among the highest levels of inequality in the world, and research has clearly documented a connection between race and poverty. There is similarly quite clear evidence of highly divergent opportunities for women and men. These patterns of inequality are intimately tied to labour market outcomes, as existing inequality creates divergent access to occupational opportunities, while discrimination in the labour market risks creating deeper entrenched patterns of inequality. These patterns cut across the formal and informal sectors of the economy, though the large informal sector is host to somewhat divergent patterns and significantly less regulation. However, while Brazil remains associated to exceptionally high levels of inequality, there are signs of change. The past two decades have witnessed a rapid movement of more women into the labour force, indicating a change to earlier patterns, and ostensibly

[12] Always in 1995, the Ministry of Labour established the "nucleos" for promoting equal opportunities within the framework of the programme "Brazil, Gender and Race – United for Equal Opportunities". Those "nucleos" are specialized units at state level that operate to promote gender and race equality at work by being involved in public actions that go from awareness-enhancing to labour inspections and dialogue with various enterprises (Alexim, Cappellin and Letierre, 2005).

declining barriers to female occupational opportunities. During the same period, and beginning with the 1988 Constitution, there has also been much greater official attention to these issues and significant changes in legislation designed to address existing patterns of both gender and race based inequality and discrimination. Many of these policies have explicitly targetted the labour market by strengthening anti-discrimination rules. These broad trends of continuity and change provide the background for a more in-depth exploration of trends in labour market segregation and discrimination over time that are explored in more detail in the subsequent empirical chapters of this thesis.

Appendix to Chapter 2

Table A1: Notable Federal Anti-Discrimination Legislation and International Conventions

- 1951: "Alfonso Arinos" Act, Law No. 1390/51, which recognized the existence of racial discrimination in Brazil.
- 1969: American Convention on Human Rights, "Pact of San Josè", adopted in Costa Rica, but not ratified by Brazil
- 1985: Law No.7437/85, which reinforced the Arinos Act, Law No. 1390/51.
- 1985: Law No. 7353/1985 to create the National Comittee for Women's Rights (*Consejo Nacional de los Derechos de la Mujer*, CNDM). Later reaffirmed in 1988 by decree No. 96895/88.
- 1988: The Federal Constitution, which contains multiple articles prohibiting discrimination, the most important of which are 5(X), 3(IV) and 5(XLI).
- 1989: Inspection Request from the Inter-American Commission on Human Rights. Initially not granted by the Brazilian government.
- 1989: Federal Law No.7716/89, known as Lei "Cao" or "Act Against Racism of Jan 5", and introduced as Complementary Legislation to the Criminal Code. Provides for imprisonment of two to five years for those who deny or prevent someone's employment due to racial discrimination.
- 1990: Federal Law No. 8081/90 which made all discriminatory acts based on race, religion, colour or ethnic and national origin illegal.
- 1992: Ratification of the American Convention on Human Rights.
- 1994: Cardoso Presidency begins
- (1994: Law No. 8882/94 which reinforced the previous law No. 7716/90.
- Federal Law 9029/95, which sets forth a penalty of one to two years imprisonment for an employer or prospective employer that requests any medical certification that a female employee or prospective employee is not pregnant.
- 1995: November 20th Presidential Decree to create the Inter-ministerial Group for Appreciation of the Black Population
- 1995: Brazil authorizes inspection by the Inter-American Commission on Human Rights, originally requested in 1989.
- 1996: Decree issued on March 20 by the Ministry of Labour to create the Working Group for the Elimination of Discrimination in the Workplace (*Grupo de Trabajo para la eliminación de la discriminación en el empleo y la ocupación*, GTEDEO) and the related Interdisciplinary Working Group (*Grupo de Trabajo Multidisciplinario*, GTM).
- 1996: September 3rd Decree creates the Permanent Group for Women's Work (*Grupo permanente del Trabajo de la Mujer*, GPTM) with the goal of identifying interventions for the Ministry of Labour relative in order to improve female participation in the labour market.
- 1996: Creation of the National Human Rights Program (*Programa Nacional de Dereitos Humanos*, PNDH). The PNDH reflected Brazil's commitment to the Vienna Declaration and Programme of Action adopted at the 1993 World Conference on Human Rights.
- 1996: Creation of the Executive Group for the Repression of Forced Labour (*Grupo Executivo de Repressão ao Trabalho Forçado*, GERTRAF)
- (1997: Law No.9455/97 of April 7, which typifies torture crimes of torture referred to racial issue.)
- 1998: Federal Decree 2682/98, which puts into force International Labor Organization ("ILO") Convention 168 dealing with the "promotion of employment and protection against unemployment".
- (1998: Law No.6165/98 which deals with the legalization of lands of descendents of "quilombolos" in all territory.).

- 1999: Federal Law 7,783/99, which provides that if a strike is not ruled abusive, workers involved in the strike may not be dismissed.
- 1999: Federal Law 9,799/99, which prohibits any form of female discrimination in the workplace.
- 2001: UN World Conference against Racism, Durban, in which Brazil played a leading role
- Federal Law No. 10244/ 2001, allowing women to work overtime. This reflected a change to Article 376 of the existing Code of the Labour Laws (*La Codificación de las Leyes del Trabajo*, CLT).
- 2002: Creation of the Second National Human Rights Program (PNDH II), on May 13.
- 2003: Lula Presidency begins
- 2003: Law No. 10678/2003 creating the Special Secretary of the President of the Republic on Policies for the Promotion of Racial Equality (*Secretaria de Políticas de Promoción de la Igualdad Racial*, SEPPIR).
- 2003: Decrees No. 4885/2003 and 4919/2003, creating the National Committee for the Promotion of Racial Equality (*Consejo Nacional de Promoción de la Igualdad Racial*, CNPIR).
- 2003: Decree No. 4625/ 2003 creating the Special Secreatary on Policies for Women (*Secretaría Especial de Políticas para la Mujer*, SPM).
- 2003: Decree No. 4773/2003 modify the earlier1985 law creating the National Committee for Women's Rights (*Consejo Nacional de los Derechos de la Mujer*, CNDM). Later modified again in 2004 by Decree No. 5273/04).
- 2003: UN Special Rapporteur, Mrs. Asma Jahangir, visits Brazil as part of preparing a report to the UN Human Rights Commission, delivered in March 2004
- 2005: Decree 5390/05 approving the National Plan of Policies for Women (*Plan Nacional de Políticas para las Mujeres*, PNPM).
- 2005: Decree No. 5397/05 strengthening the National Council to Combat Discrimination (*Consejo Nacional de Combate a la Discriminación*, CNCD).
- 2007: Decree No. 6269/07 modifying Decree 5390/05 and approving the National Plan of Policies for Women (*Plan Nacional de Políticas para las Mujeres*, PNPM) .
- 2008: Decree No. 6509/08 (ammending Decree No. 4885/03) on the composition, structure, competencies and functions of the National Committee for the Promotion of Racial Equality (*Consejo Nacional de Promoción de la Igualdad Racial*, CNPIR)
- 2008: Decree No. 6572/08 (ammending Decree No. 5390/05) on the National Plan of Policies for Women (Plan Nacional de Políticas para las Mujeres, PNPM).
- 2008: Decree No. 6387/08 approving the Second National Plan of Policies for Women (*II Plan Nacional de Políticas para las Mujeres*, II PNPM).
- 2009: Decree No. 6872/09 approving the National Plan for the Promotion of Racial Equality (*Plan Nacional de Promoción de la Igualdad Racial*, PLANAPIR).
- 2009: Decree No. 7,037 of 21 December creating of the Third National Human Rights Program (PNDH-3), which was approved in 2010 (Decree No. 7177/2010).
- 2010: Law No. 12,288, of 20 June, creating the Statute of Racial Equality.

Source: Author's compilation.

Chapter 3

A Description of the Data Source Used

A central challenge facing any study of the evolution of occupational segregation and wage discrimination in developing countries over time is the nature and quality of the data available. The data, at a minimum, must provide a large and representative sample and contain detailed information on occupations, wages and personal characteristics consistently collected over time. Brazil provides an attractive setting in which to conduct such a study given the existence of a nationally representative household survey, the *Pesquisa Nacional por Amostra do Domicilios* (PNAD) collected annually and covering the period from 1987 to 2006. This chapter provides an introduction to this data source, highlighting its core features, its advantages and some of its limitations for the purposes of our research.

3.1 Potential Alternative Datasets

The *Pesquisa Nacional por Amostra do Domicilios* (PNAD) covers the period from 1973 up to now and is collected annually by the national statistical office, the *Instituto de Geografia e Estatistica* (IBGE). The survey is generally conducted in the last quarter of the year, between the end of September and beginning of October, with data subsequently made available on the IBGE website accompanied by additional documentation, including a coding manual and variable definitions. It is among the most comprehensive source of information about the socio-economic characteristics of Brazilian households, as it covers the entire country over two decades and provides reasonably detailed data. However, there are several alternative Brazilian datasets available, and it is useful to briefly highlight its advantages relative to other options given the primary objectives of this study. The most plausible alternatives are the *Pesquisa de Padrões de Vida* (PPV), the *Pesquisa Mensal de Emprego* (PME), the

Census and the *Relação Anual de Informações Sociais* (RAIS). However, for our purposes the PNAD provides the best combination of detailed labour market information and geographic and temporal coverage.

The *Pesquisa de Padrões de Vida* (PPV) has a very wide range of socio-economic variables, but the geographical and temporal coverage is limited. It has been conducted in collaboration with the World Bank exclusively in the North-East and South-East regions and only in the years 1996 and 1997. The *Pesquisa Mensal de Emprego* (PME) is a monthly survey and provides detailed information on labour market conditions. This includes information related to labour market activity, employment conditions, nominal and real incomes, classes of workers and the possession of formal contracts. However, while the survey provides rich information, it covers only a small number of urban areas (in particular, the metropolitan areas of Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo and Porto Alegre), which, although highly populated, are unlikely to provide a comprehensive portrait of the overall national trends. The *Relação Anual de Informações Sociais* (RAIS) is an annual census of all firms and their employees. This annual report contains detailed information about each employee (including wages, hours worked, education, age, tenure, gender) and each firm (including industry, region, size, establishment type), including a unique identifier for each employee, each firm and each establishment. It is a potentially powerful tool because it can be matched with other Brazilian firm datasets, including, for example, the *Pesquisa Industrial Annual* (PIA). More importantly, it has a panel structure as it follows the same firms and their employees over time. However, because the dataset covers only regularly registered firms, it is restricted to the formal sector and formal employees, and thus excludes the large and sizeable informal sector. Finally, the Census constitutes the only source of information about life conditions of the population across all municipalities and sub-divisions. However, because of the high costs of data collection it is conducted only every ten years.

3.2 Basic Structure of the PNAD

The PNAD is generally comprised of two questionnaires: a core questionnaire that is relatively compatible over time and a supplementary questionnaire that varies from year to year in order to investigate special issues of interest. These supplementary questionnaires have included surveys focussing on youth education and training (PNAD 2007), health and health expenditure (PNAD 2003, 2008), food security (PNAD 2004-2009), migration, child labour and social programs (PNAD 2006), social behaviour (the use of the internet and mobile phones in PNAD 2005 and 2008, victimization and access to justice in PNAD 2011).

Ultimately, our focus is on information contained in the core questionnaire, which permits the analysis of trends over time. The core questionnaire includes extensive information about the households of respondents: geographic location of the household, structure and type of household, relationships among household members, primary sources of household income and characteristics of the dwelling, including ownership, size, building materials, access to facilities and main assets. This household information is accompanied by detailed information about the characteristics of individual respondents, including gender, age, ethnicity as well as more detailed information about, *inter alias*, educational history and migration. Most importantly for this thesis, the questionnaire includes a comprehensive employment section containing details about both current and previous activities, including an individual's occupation, working hours, the size of firm, position in firm, duration of employment or unemployment, types of income, both in cash and in kind (i.e., employment and business income, income from savings, investments and private insurance and social security transfer), payment of social contributions, possession of a working card, participation in union activity and so forth. Finally, there are also sections on employment characteristics for individuals aged between five and nine years, and a section on women's fertility history.

3.3 Temporal Coverage and Sample Size

The PNAD survey was first conducted in 1973, but the range of information was relatively restricted, in part as a reflection of the political situation during the military dictatorship. It was only with the transition to democracy that the survey began in 1987 to consistently collect information on ethnicities, which explains the decision to focus on the years from 1987 onwards in this study. The survey has been implemented every year since 1987 with the exception of the years in which the census was conducted (1991, 2000, 2010) and in 1994 due to the political and economic upheaval associated with the implementation of the *Plano Real* stabilization reforms.

While the PNAD survey is undertaken annually it is not a panel data, as the same individuals are not followed over time. It thus derives its utility instead from the large sample sizes generally available, though the sample size and list of variables has altered considerably over the years. The sample size has increased continuously over time from an initial size of approximately 290,000 individuals up to 400,000 individuals in more recent years from between 70,000 and 140,000 households.

Over time the number of variables on which information is available has also increased. This has included the expansion of questions about consumer durables at the household level as well as the collection of additional information related to the migration and labour market experiences of individuals. The range of variables is particularly limited in the first three years of our dataset (i.e., from 1987-89) when the number of variables was approximately 130. Beginning in the 1990s the number of variables increased rapidly and reached the current level of 338 by 2001.

3.4 Sampling Procedure

The sampling procedure adopted in collection of the PNAD data is a three-level multi-stage sampling procedure that incorporates municipalities, census sectors and households. The first stage involves the choice of municipalities, of which there are 5000 in Brazil. The municipalities are divided into three groups: metropolitan areas, auto-representative and non-auto-representative. The first two types of municipality are exclusively capital municipalities, metropolitan municipalities and municipalities with very high population densities and are automatically included in the sample. The majority of municipalities fall into the non-auto-representative municipalities group, and these are selected based on a probability proportional to their size as reflected in the most recent population census. The second stage of the sample procedure refers to the selection of the Census sectors within each municipality which are identified and selected according to population proportions of the Census. Finally, in the final stage of the survey process households are randomly sampled from within each Census sector.

Within this sampling procedure no household is ever interviewed more than once, as they are removed from the PNAD register after been interviewed.

The adoption of this three-level multi-stage sampling procedure rather than a simple random selection of households from across the entire nation reflects an effort to reduce costs given Brazil's large geographical and population size and the desire to implement the survey annually. Logistical and cost challenges also explain the exclusion of several rural areas in the Northern region from the survey. These account, however, for less than 3% of the total population. The excluded areas include the states of Rondônia, Acre, Amazonas, Roraima, Pará and Amapá, which are all remote areas in the Amazon with very low population densities and which, moreover, are potentially dangerous for interviewers.

3.5 Description of Key Variables

While there is a wide range of socio-economic information contained in the survey, for the purposes of this study the most crucial variables are those related to ethnicity/race and to the coding of occupations for each individual. Both pose potential challenges, and have motivated particularly choices as well as dictating the occupational code harmonization described in the next chapter.

The fact that Brazil is a multi-racial society has complicated the development of racial classifications, which have altered over the life of the survey. As noted earlier, the PNAD has included a question about race and skin colour from its inception in 1987, and it initially adopted classification based on the Census. This classification included *brancos* (whites), *pretos* (black people), *amarelos* (Asians), and *pardos* (brown people, which included *mulatos*, *caboclos*, *cafuzos*, *mamelucos* and *mestiços*), while in 1992 a separate category was created from indigenous people distinct from their earlier categorization as *pardos*. However, despite these seemingly clear categorizations, there is a need to consider the actual existence of a "colour continuum", which can make classification ambiguous and subjective. Indeed, in practice there is evidence that the racial classifications adopted by Brazilians tend to be influenced by their socio-economic conditions (Lovell and Wood, 1998; Wood, 1991). For example, Telles and Lim (1998) find less income inequality using data based on self-classification than when using data based on an interviewer's classification, thus providing a possible example of how money "whitens". The consequence is that self-assigned racial classifications might change across time as people get wealthier, while data based on self-classification may yield different results than that based on interviewer classifications.

This issue is likely to be particularly acute for those generally classified as 'brown', as they may re-classify themselves as 'white' or 'black' depending on their social and economic status. This results in an underestimation of white incomes when self-classifications are applied. The existence of a multi-racial society, and a corresponding skin colour continuum, makes Brazilian data on race less reliable than is perhaps the case for other countries. The methodological choice characterizing most of the empirical work on this topic is to aggregate all non-white groups into one category. For the purposes of this thesis, individuals are assigned to either the 'white' or 'non-white' group. Throughout the thesis, we employ the commonly recognized and understood term "racial" to denote skin tone. However, while we employ this term for the sake of simplicity, the term "skin-tone" is arguably more accurate, as the Brazilian population is generally held not to be classifiable into ethnicities, as explained in the previous chapter.

Even more crucial is the variable identifying occupational codes, as it provides the basis for drawing inferences about occupational segregation and wage discrimination over time. However, the classification used in the PNAD survey is problematic in two key respects. The classification varies over time, while for the majority of available years it is not directly comparable with the international classification provided by the ILO, the ISCO-08.

It is this inconsistency in the classification of occupations that necessitates the construction of a new harmonized re-classification of occupational codes. We have thus constructed a new re-classification of 83 occupational codes at 3-digit level, 25 occupational codes at 2-digit level and nine occupational codes at 1-digit level, all of which are compatible over time, internationally comparable, and coherent with the different classifications adopted within the survey itself. This novel contribution permits the analysis of the evolution of the occupational structure and occupational distribution over a longer period than has previously been possible, while also providing a consistent and comparable basis for the analysis of occupational segregation (in chapter 5) and its role in determining wage differentials (in chapters 6 and 7). The details of this re-classification are discussed in detail in the next chapter.

3.6 Conclusions

The PNAD dataset is relatively unique among developing countries in providing extensive representative data on both individual characteristics and employment outcomes across the entire country for a large sample and over an extended period of time. It thus renders Brazil, an ethnically diverse country, a particularly rich setting for studying trends in occupational segregation and wage discrimination across both the racial and gender divides. However, while the PNAD dataset provides a useful and important basis for the empirical analysis undertaken for the thesis, its use is also subject to limitations in relation to the classification of occupations. In addition, specific care should be exercised particularly in relation to both the racial classifications and the informality definitions used. Having thus provided the background to the data, the next chapter provides a detailed account of the construction of the harmonized data for the occupational codes used in this study, and a detailed discussion of the construction of other key variables subsequently used in the empirical analysis.

Chapter 4

A Profile of the Brazilian Occupational Structure Using a New Harmonized Classification of Occupational Codes

4.1 Introduction

The objectives of this chapter are twofold. First, we construct a new harmonized classification of occupations for the Brazilian national household surveys, the *Pesquisa Nacional por Amostra do Domicilios* (PNAD) from 1987 to 2006. Second, we empirically investigate occupational structure in Brazil over time, examining trends disaggregated by both gender and race and separately for the formal, informal and self-employed labour markets.

The creation of a compatible re-classification of the occupational codes represents a key contribution of this chapter and of the thesis more generally. By overcoming the discontinuities generated by changes in the original PNAD classification, it renders feasible the exploration of labour market trends over a longer period than previously possible, and permits the use of categories that facilitate international comparison. Furthermore, not only do we adopt this harmonized re-classification to analyze the evolution of occupational structure over time, we also use it in subsequent chapters to study the evolution of occupational segregation and to account for the impact of occupational segregation in shaping patterns of wage discrimination over time. To the best of the author's knowledge, this is the first study that investigates trends in occupational structure, occupational segregation and wage discrimination over a protracted period of time thanks to the availability of this new harmonized, detailed and consistent re-classification of occupational codes.

Our comprehensive analysis of the structure of the Brazilian labour market identifies several key characteristics. First, over the last two decades the Brazilian labour market has witnessed a large increase in female participation in the labour market, with the share of female workers in the workforce increasing from roughly 36% in 1987 to 43% by 2006. The composition of the labour market by race is also changing, as non-white workers have comprised the majority of the workforce since 2003.

Second, the size of the informal sector has remained relatively constant as a share of the entire labour force at the national level between 1987 and 2006, with informality concentrated in metropolitan areas, consistent with the previous findings reported by Ramos and Ferreira (2005). However, despite the constant share of the informal sector in the overall labour market, the proportion of women, and particularly non-white women, in the informal sector has increased significantly over time.

Third, turning to differences in the occupational distribution across gender and race, the data reveal that the occupational distribution between white and non-white workers is broadly similar. By contrast, there are major differences in occupational distributions between men and women. Interestingly, the degree of concentration increases slightly when we restrict the analysis to the informal sector, which may reflect a less diversified informal labour market. Finally, over time we observe a steady decline in the concentration of both gender and racial shares in many of the most concentrated occupations.

Fourth, despite these modest changes the overall occupational distribution has remained remarkably stable over the last two decades, despite increased female and non-white participation in the labour market. That is, the proportion of male and female, and white and non-white, workers in individual occupations has been surprisingly stable over time. However, this does not imply that the occupational structure has remained unaltered over time, as we see a persistent expansion of the tertiary sector. What the stability of the occupational distribution reveals is that female and non-white workers entering the labour market have primarily joined occupations that were already female and non-white dominated, and disproportionately so in the tertiary sector.

Fifth, there are also noteworthy trends in the employment of different groups in high-skilled and less-skilled occupations. Looking at the issue of race, white-dominated occupations tend to be highly-skilled, while non-white-dominated occupations tend to be less-skilled. More surprising is that the most female-dominated occupations are also the highly skilled occupations. On the surface this would seem to present evidence that women have unhindered access to high-skilled formal sector employment. But there is also a more subtle, and more likely, explanation: women face higher barriers to entering formal sector occupations than men. As a result, women in the formal sector are generally highly-skilled, and concentrated in highly-skilled occupations, while on the whole women lacking high levels of education continue to be disproportionately confined to the informal sector. These barriers to formal sector employment are, in turn, particularly acute for non-white women, whose participation in the informal sector is increasing most rapidly, particularly in the personal services sector (e.g., as housekeepers).

This chapter is divided into five sections. The next section reviews various prior attempts to harmonize Brazilian occupational classifications. The third section then describes the construction of our re-classification. Section four provides a comprehensive analysis of the Brazilian occupational structure, looking first at broad trends in the composition of the Brazilian labour force, disaggregated by sector (formal, informal and self-employed) and by major population groups (gender and race) and then turning to more detailed analysis of trends in the occupational distribution. Finally, the fifth section offers some concluding remarks.

4.2 Previous attempts to harmonize the classification

Most studies of occupational segregation and wage discrimination in Brazil exploit based on the *Pesquisa Nacional por Amostra do Domicilios* (PNAD), but these studies are plagued by the existence of a major break in the data on occupations given the radical change in the way occupations have been classified since 2001. The result is that it has been difficult to conduct studies of the evolution of occupational segregation in Brazil over a protracted period of time.[13] In order to overcome this problem, this study is based on a novel re-classification and harmonization of the occupational codes from successive PNAD surveys, thus allowing for the analysis of trends in occupational structure and segregation over a longer period, and in greater detail, than has previously been possible. Part of the reason that such a thorough harmonization effort has not

[13] While de Oliviera (2001) analyses occupational segregation over the relatively long period from 1981 to 1999, this was only possible because the study focused on years prior to the radical change in the PNAD classification of occupational codes in 2001 (de Oliveira, 2001).

previously been undertaken is that the majority of empirical studies investigating the Brazilian labour market have focused on economic sectors,[14] or major labour categories (formal/informal, private/public occupations, self-employed/employer/employee), rather than specific occupational groups, thus avoiding the need for occupational data harmonization.

Among the relatively few existing studies that have focused on specific occupational groups, a number of solutions have emerged to deal with the break in the PNAD data. Two studies employ the classification of occupations proposed by the national household survey, but both are restricted to periods during which there was no deep restructuring of the codes. De Oliveira (2001) explores occupational segregation by gender using the PNAD from 1981 to 1999, which includes 357 occupational codes at the 3-digit, 59 at 2-digit and seven at 1-digit level. Machado, de Oliveira and Carvalho (2003) aggregate the occupational codes into 67 groups at 2-digit level in order to obtain a consistent classification using the PNAD from 1981 to 2001.

Alternatively, several studies have attempted to re-classify the occupational codes collected from different Brazilian datasets, though never with the level of detail envisaged in this chapter. Barros, Machado and Mendonca (1997) use the Brazilian monthly labour survey, the *Pesquisa Mensal do Emprego* (PME), from 1983 to 1993 and aggregate the occupational codes into 19 occupational groups consistent over the ten years. Using Census data, Lovell (1994, 2000, 2006) groups occupational codes into six hierarchical occupational categories, three white collar and three blue collar.[15] The most detailed compatible re-classification of occupational codes using Census data is provided by Lago (2006) who proposes 25 occupational codes that are compatible across the 1991 and 2000 Brazilian Censuses.

The most detailed and comparable re-classification of occupational codes to the re-classification exercise undertaken here is that of Osorio (2008), who similarly draws on the PNAD national household survey and constructs 46 occupational codes at 2-digit level over the period 1986 to 2006. This effort possesses some commonalities with that reported here, but our re-classification, while covering the same 20 years of the PNAD dataset, is more detailed and is harmonized with the most recent international

[14] Some studies using the variable for economic sectors are Arabsheibani, Carneiro and Henley (2003), Arcand and D'Hombres (2004), Arias, Yamada and Tejerina (2004), Arbache (2001), Green, Dickerson and Arbache (2001), Campante, Crespo and Leite (2004), Guimares (2006), Soares (2000), Azzoni and Servo (2002), Ferreira, Leite and Wai-Poi, (2007).

[15] The same re-classification has been used in the Lovell and Wood (1998) study.

classification of occupations provided by the International Labour Organization (ILO), the ISCO-08.[16]

Aside from these academic efforts, the Brazilian national commission for classification, the *Commisão Nacional de Classificação* (CONCLA) has also prepared a re-classification that recodes the official national classification of occupations (CBO-94) and the official classification of occupations used by the Census in order to make both compatible with the international classification standard ISCO-08.[17] A study released by Muendler et al (2004) provides detailed discussion of the mapping from CBO-94 to the international standard classification, ISCO-88. However, the official national classification of occupations by CONCLA does not address the need for a compatible re-classification of PNAD's occupational codes over the entire period 1987-2006 as it does not deal with the distortion of the time series by changes in the occupational codes for the PNADs starting in 2002. This change saw the PNAD's occupational classification move from a 3-digit to a 4-digit classification, which is very similar to the international framework provided by ILO. Furthermore, although the classifications in the PNADs were relatively stable prior to 2002, we have identified several minor changes after 1992 that have also needed explicit attention.

4.3 The Construction of the new Classification of Occupational Codes

If we focus on changes in the occupational classification used by the PNAD over time, the original PNAD datasets can be grouped into three waves (although the first two groups are very similar):

- 1st group PNAD 1987-1990: occupations are reported in a 3-digit classification with 367 different codes;
- 2nd group PNAD 1992-2001: occupations are reported in a 3-digit classification with 381 different codes;
- 3rd group PNAD 2002-2006: occupations are reported in a 4-digit classification with 489 different codes.

[16] The ISCO-08 has been released in 2008. The old-version ISCO-88 has been constructed in 1988. The new version is very similar to the old one, only several occupational categories have been added.
[17] http://www.ibge.gov.br/concla/cl_corresp.php?sl=3

Our classification adopts the 3-digit categorization,[18] used by the international classification, ISCO-08, released by the ILO, as presented in table A1 of the appendix. The major groups are:

- Major Group 1: Legislators, senior officials, and managers;
- Major Group 2: Professionals;
- Major Group 3: Technicians and associate professionals;
- Major Group 4: Clerks;
- Major Group 5: Service workers and shop and market sales workers;
- Major Group 6: Skilled agricultural and fishery workers;
- Major Group 7: Craft and related trades workers;
- Major Group 8: Plant and machine operators and assemblers;
- Major Group 9: Elementary occupations.
- Major Group 0: Armed forces.

Table A1 provides a detailed breakdown of the re-classification, laying out the correspondence between occupational codes from the three different groups of PNAD surveys and the equivalent ISCO-08 classification. In simplified terms, arriving at this re-classification has involved a two-stage process. It was first necessary to translate the Portuguese occupational label and attempt to match them to the closest English equivalent from the international classification. This has generally involved aggregating multiple PNAD occupational codes into a single occupational code from the international classification. The second stage has been to then look in detail at the individual earnings and educational characteristics of workers within each aggregated category in order to detect possible mismatches. Given the changes in the PNAD occupational classification over time, an important part of this exercise has been ensuring that the aggregation decisions taken with respect to each distinct group of PNAD surveys produced a re-classification that was consistent over time, and across those initial points of discontinuity.

While we do not dwell on the details of how each category has been harmonized and validated, it is worth highlighting some particularly important challenges and choices involved in arriving at the final re-classification. We look at each Major Group within the ISCO-08 classification in turn.

[18] Our re-classification of occupational codes for the PNAD datasets, compatible from 1987 to 2006, is available in Stata do file format on request to the author.

Beginning with Major Group 1 ('Legislators, Senior Officials and Managers) we encounted a significant challenge in distinguishing between the two categories of managers -"(12) Corporate managers" and "(13) General managers" – which are designed to capture the managers of large firms and small enterprises, respectively. The distinction is straightforward in the most recent PNAD surveys (3rd group PNAD 2002-2006), which clearly distinguished managers and employers within large enterprises (over five employees) from the managers of small enterprises. However, this distinction does not exist in the dataset prior to 2001, which draws a distinction between "employers" (*empregador*) and "managers" (*dirigente*) but without any reference to firm size. Because of the consequent difficulties in distinguishing between the two groups of managers, we merge categories 12 and 13 in the subsequent analysis.

Moving to Major Group 2 ('Professionals'), the classification is generally straightforward, as we find, across all three groups of PNAD surveys, that those within identified occupations have relatively homogeneous levels of education. However, in a small number of cases, the occupational descriptions offered in the PNAD are too generic to make a clear judgement about whether they should be classed as professionals. In these cases, we have inferred the appropriate classification by looking at average levels of education, and these decisions have been relatively clear cut. Thus, for example, workers in "166-Women assisting births" (166-*Parteira*) have only average levels of education, and are thus not classified as nursing professionals. The only category in which this approach has proven ambiguous is "245 - Writers and creative or performing artists", as we find heterogeneous levels of education within this category. In this case we opt to classify them as professionals, given that it is plausible that professionals within this occupation may nonetheless have varied levels of education.

Substantial changes were needed in dealing with Major Group 6 ("Skilled agricultural and fishery workers"), where it has been necessary to re-classify most of the categories because of incompatibility between the ISCO-08 and the PNAD classifications. The ISCO-08 makes a primary distinction between workers and producers, while the PNAD includes a division in the agricultural sector between "workers" and "own-account producers". However, within the PNAD it is not possible to distinguish whether "animal producers" and "crop growers" are "own account producers" or "workers". We thus choose to treat crop and animal producers as own-account producers, while we treat crop and animal workers as "not own-account" and re-label this category "agricultural workers". We do not encounter similar difficulties for forestry and fishery workers, thus leaving a four part re-classification for the Major Group:

- Agricultural workers - own account excluded
- Crop and animal producers - own account
- Forestry and related workers
- Fishery workers, hunters and trappers

Finally, there is a potential problem in dealing with occupations recorded under Major Groups 7, 8 and 9 by ISCO08. Within the ISCO-08 classification, these Major Groups distinguish between specific types of occupations that cut across a range of economic activities (for example, mining, construction or manufacturing). Thus, Major Group 7 is for "workers", Major Group 8 is for "plant-operators" and "machine-operators" and Major Group 9 is for "labourers" or "elementary occupations". By contrast, the PNAD often conflates these types of worker together by economic activity, thus making it very difficult to distinguish between labourers, plant-operators, machinery-operators and workers within any given economic activity. This is particularly the case in the earlier years, which provide a less disaggregated set of occupational categories. The only feasible solution has been to aggregate many occupational codes from the official ISCO-08, owing to the difficulty of identifying PNAD occupational codes to correspond to some of the occupations identified in the ISCO-08 classification. For example, among "elementary occupations" (Major Group 9), we were only able to include "street vendors" and "transport," as it was impossible to confidently distinguish other labourers based on the PNAD classification.

In taking the decisions necessary to reconcile the PNAD and ISCO-08 classification, the most important challenge has been to ensure that within the new classification we are grouping occupations appropriately, such that there is sufficient homogeneity within each occupational grouping. To this end, we have analysed key features of the occupational groups that are aggregated together in transitioning to the ISCO-08 classification to ensure this homogeneity. In this process we have particularly focused on average earnings from primary employment and both the average and modal levels of educational attainment. Key to understanding the challenges associated with this process is the fact that this re-classification involves transforming the more profession-based Brazilian classification system, CBO-94, into the more skill-oriented international system, ISCO-08 (Muendler et al, 2004).

Ultimately, we include all of the occupations that are part of the public sector, and exclude only the armed forces and two categories of poorly-defined occupations, leaving us with a new classification that contains 83 occupational codes at 3-digit level, 25 at 2-digit level and 9 at 1-digit level. Our re-classification of Brazilian occupational codes is consistent over time and compatible with international standards. It thus offers the potential for significantly more detailed analysis over a longer time period of occupational structure and segregation than has previously been possible for Brazil.

4.4 The Brazilian Occupational Structure

Having completed this re-classification of Brazilian occupational codes, it is now possible to present extensive new data on the evolution of occupational structure in Brazil over time. What follows focuses in particular on highlighting differences based on gender and race, with a further emphasis on differences across the formal, informal and self-employed sectors. Throughout the analysis, we consider the entire national labour market (i.e., all five regions of Brazil and both urban and rural areas). As noted earlier, this is important, as conditions vary across the country and, as such, the analysis of specific regions or metropolitan areas risks capturing trends that do not reflect the national patterns.

We opt for a more inclusive approach, while acknowledging that our findings are averages at the national level – an issue that is addressed later in the chapter by looking separately at regional and state level trends. We initially examine broad trends in the composition of the Brazilian labour force, disaggregated by sector (formal, informal and self-employed) and by major population groups (gender and race). Having outlined these broad trends, the analysis then turns to looking at more detailed trends in the occupational distribution of Brazil.

4.4.1 Aggregate Labour Market Trends

For our analysis, we divide the entire labour market into the formal, informal and self-employed sectors. The formal sector comprises private sector employees with signed labour cards, domestic workers with signed labour cards, and civil servants. The informal sector comprises private sector employees and domestic workers without a signed labour card. Given that our sample covers both urban and rural areas, agricultural workers both with and without signed labour cards are included in the formal and informal sectors respectively. We choose to keep self-employed workers separate from informal workers due to differences in their composition across the two sectors. We exclude military forces and have also excluded employers. While we explored the possibility of following Bosch, Goni and Maloney (2007) in using the ILO threshold to distinguish formal and informal sector employers (with those enterprises with less than five employees considered informal), this method appears to be problematic in our case, as the data reveal that at least 50% of small firm employers pay social security contributions and, as a consequence, should not be considered informal workers. Finally, our sample excludes workers who are not remunerated or for whom the wages variable is missing, placing them in a 'no wage' category. As the exclusion of 'no wage' observations is likely to result in an underestimate of the non-formal sectors, we perform a sensitivity analysis (presented later in the next chapter) to see if accounting for these 'no wage' observations has a significant effect on our estimates of informality and segregation. Table 1 reports the number of observations, as well as their relative share of the sample, for each of the excluded categories in the dataset.

Table 1: Data Exclusions in Defining the Final Sample

	1987		1992		1997		2002		2006	
Original sample	298,287	%	316,026	%	344,874	%	381,545	%	405,265	%
- age<15 & age>65	120,544	40.41%	121,555	38.46%	124,404	36.07%	130,201	34.12%	133,421	32.92%
- missing occupational code	66,709	22.36%	68,327	21.62%	81,827	23.73%	90,161	23.63%	92,001	22.70%
- 'no wage' observations	6,613	2.22%	17,382	5.50%	16,333	4.74%	17,167	4.50%	18,396	4.54%
- employers	3,834	1.29%	4,683	1.48%	5,621	1.63%	6,549	1.72%	7,396	1.82%
- military forces	1,605	0.54%	1,733	0.55%	1,874	0.54%	1,565	0.41%	1,708	0.42%
Final sample	98,982		102,346		114,815		135,902		152,343	

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

Note: The exclusion of individuals younger than 15 and older than 65 limits the sample to the economically active population. The exclusion of observations with missing occupational codes excludes those who are out of the labour force, those who are unemployed and a small group of employed individuals who failed to report their occupations. Finally, the 'no wage' category includes both observations with missing wages and workers who are not remunerated (for further explanation see subsection 3.3.2 in the next chapter).

4.4.1.1 Distribution of Workers between the Formal, Informal and Self-Employed Sectors

Having carefully defined our categories, we now examine at the composition of the entire labour force divided into formal, informal and self-employed sectors. Although we consider all 20 years of data from 1987 to 2006 in the analysis, for the sake of brevity we report only five representative years (1987, 1992, 1997, 2002 and 2006) in this chapter.[19]

Table 2: The composition of the final sample across sectors

	1987	1992	1997	2002	2006
All labour market					
# of obs.	98,982	102,346	114,815	135,902	152,343
of which:					
Women	36.03%	37.98%	39.09%	40.70%	41.77%
Non-whites	46.67%	47.21%	47.70%	49.90%	53.27%
Formal sector					
# of obs.	45,065	48,540	52,306	60,680	71,235
as % of the total	45.53%	47.43%	45.56%	44.65%	46.76%
of which:					
Women	34.15%	38.83%	41.40%	42.54%	42.69%
Non-whites	40.61%	40.55%	41.86%	43.42%	47.73%
Informal sector					
# of obs.	28757	27726	32045	40436	43287
as % of the total	29.05%	27.09%	27.91%	29.75%	28.41%
of which:					
Women	42.27%	43.12%	44.35%	45.61%	47.75%
Non-whites	54.25%	57.77%	57.27%	58.07%	61.02%
Self-employed					
# of obs.	25160	26080	30464	34786	37821
as % of the total	25.42%	25.48%	26.53%	25.60%	24.83%
of which:					
Women	32.26%	30.95%	29.61%	31.78%	33.20%
Non-whites	48.88%	48.37%	47.67%	51.70%	54.81%

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

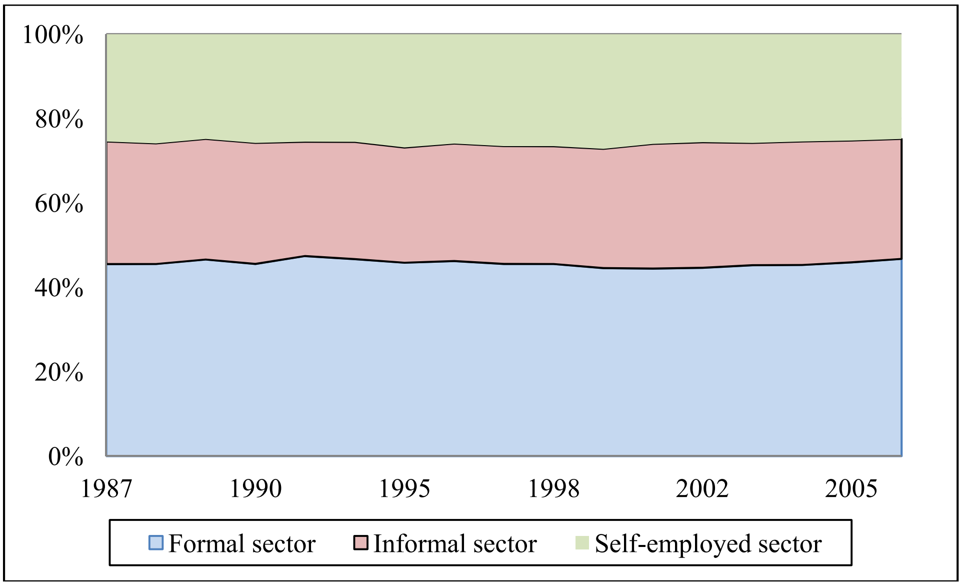
As presented in table 2, at an aggregate level the Brazilian formal and informal labour markets appear to possess two main features. First, the informal and self-employed sectors cover more than half of the entire sample across all 20 years. Second, the distribution of workers across these three sectors has remained broadly immutable over time. The formal sector has increased by only 1.23 percentage points during the

[19] As was explained in the previous chapter this start date is not arbitrary, but is because the PNAD dataset only re-introduced information on race/skin colour in 1987. From that start date in 1987 we simply select these five years at regular intervals in order to be brief while being representative of the whole period. Analysis for all 20 years is available on request from the author.

last two decades, moving from 45.53% in 1987 to 46.76% in 2006. On the other hand, both the informal and self-employed sectors have each declined by 0.6 percentage points.

The evolution of the formal, informal and self-employed sectors over time is displayed in figure 1. The absence of an increase in informal sector activities at the national level is in line with previous research by Ramos and Ferreira (2005). While several other studies have reported increasing informality,[20] this increase is a more restricted phenomenon concentrated mainly among private employees in metropolitan areas, especially in the South-East region. Part of the explanation for the quite different trend when looking at the national level is that our sample accounts for agricultural and domestic workers, both of which have experienced an increase in the "degree of formalization" of their occupations over time.[21]

Figure 1: Shares of formal and non-formal sectors over time - as percentage of total labour force



Source: Author's computations using PNAD 1987 - 2006.
Note: 1991, 1994 and 2001 missing years.

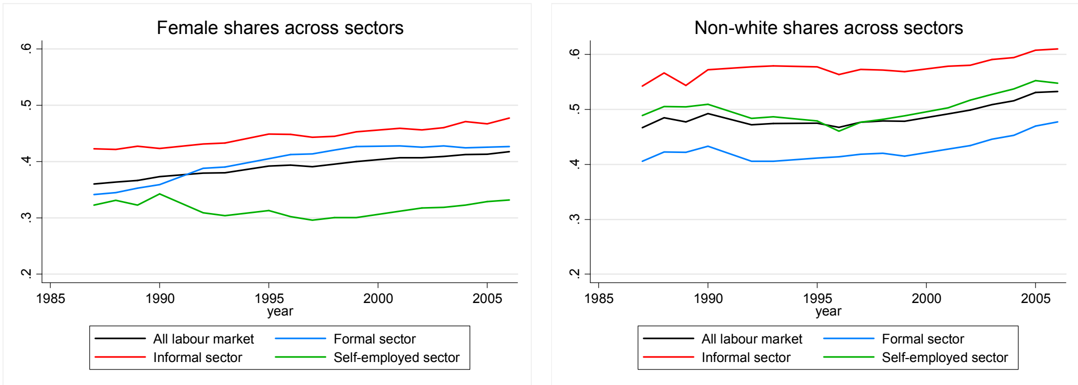
Turning to trends by gender and race, figure 2 reveals that although male and white workers have traditionally dominated the Brazilian labour market, the presence of

[20] See, for example, Carneiro (1997) and Bosch, Goni and Maloney (2007). Carneiro (1997) reports an increase in informality when looking at the metropolitan area of Sao Paulo. Bosch, Goni and Maloney (2007) find that the informal sector has increased by 10 percentage points between 1985 and 2002 when considering only the private sector in six metropolitan areas.

[21] In our dataset, the share of formal agricultural workers increased from 12% to 18% between 1987 and 2006, while formal domestic workers increased from 19% in 1992 to 27% in 2006 (see Fonseca & Rayp (2011) on the formalization of agricultural workers and ILO (2010) on the formalization of domestic workers).

women and non-white workers has increased steadily over time (by 5.74 and 6.59 percentage points respectively). Looking specifically at women, despite the rapid increase in the share of women in the labour market, they still comprise over slightly above 40% of the total labour force and are disproportionately employed in the informal sector. In the case of race, the share of non-white workers in the labour force has increased sharply over time, such that it exceeds the share of white workers from 2003 onward. That said, the employment of non-white individuals varies widely across sectors, as they remain underrepresented in the formal sector, while the informal sector has always been dominated by non-whites.

Figure 2: Female and non-white shares across all labour markets over time



Source: Author's computations using PNAD 1987 - 2006.
Note: 1991, 1994 and 2001 missing years.

Having captured the growing role of women and of non-white individuals in the employed labour force, and their continued overrepresentation in the informal sector, we can provide a more nuanced portrait by considering four distinct groups: white men, non-white men, white women and non-white women. The results of this exercise are presented in figure 3.

We note that men comprise the strong majority of the employed labour force, with a decreasing role of white relative to non-white men. The representation of women in the employed labour force has increased in both the formal and informal sectors, but with important differences by race. The increased presence of white females in the employed labour force is almost exclusively within the formal sector. Meanwhile, non-white female workers record the largest increase, of roughly six percentage points, in both the formal and informal sectors. Finally, it is interesting to note that while non-white individuals represent an increasing share of employment in both non-formal sectors, there are important differences by gender. Increasing non-white employment in the informal sector is driven overwhelmingly by non-white women, while in the self-employed sector it is non-white men who have increased their share of total employment by roughly five percentage points.

Figure 3: The gender/racial composition across sectors

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

4.4.1.2 Trends in Labour Market Participation Rates

Having noted these broad trends in the composition of the employed labour force, it is useful to also briefly explore the trends in participation rates (computed as the ratio of the labour force the number of working age individuals) across the four major population groups of interest. Overall, we find that across the entire labour force, the participation rate rises by 5.47 percentage points between 1987 and 2006 (from 56.6% in 1987 to 62% in 2006), as reported in table 3. This is driven primarily by increases in female participation, which increased by 13.4 percentage points (moving from 38.8% in 1987 to 52.2% in 2006), while there was little difference between whites and non-whites, with participation rates rising at similar rates. This increasing female participation in the Brazilian labour market has already been documented in several studies (see, among others, Tzannatos (1999), Soares & Inaki (2002) and World Bank (2002a)).

More interesting trends are apparent in the differences between the formal and informal sectors. We find that increasing participation in the formal sector is driven primarily by white female workers, while in the informal sector non-white women is the key group. Thus, although increasing participation in the labour market is primarily attributable to women, the situations of white and non-white women have been markedly different, with white women largely joining the formal sector, while increasing participation among non-white women has been more heavily concentrated in the informal sector.[22]

Table 3: Changes over time of participation rates across sectors

	All labour market	Formal sector	Informal sector	Self-employed sector
All workers	**5.48%**	0.77%	-1.95%	-0.15%
Female workers	**13.46%**	3.88%	0.32%	0.02%
Male workers	-3.01%	-2.53%	-4.38%	-0.30%
Non-white workers	5.47%	1.01%	-2.74%	-0.74%
White workers	5.45%	1.29%	-1.79%	0.30%

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

4.4.2 Labour Market Trends, Disaggregated by Characteristics

Having described aggregate labour market trends, we now present the analysis of trends over time in female and non-white participation disaggregated by individual characteristics of the labour force, in order to add further nuance to the narrative. These trends are not only of interest in their own right, but also provide the necessary context for the analysis of occupational segregation to follow in the next chapter. We consider several characteristics, including labour force characteristics (age, educational attainment), geographic characteristics (region of residence, urban/rural differences) and different sectors of economic activity. When considering age, we define three age groups: young (aged 15-29), adult (aged 30-49) and mature (aged 50-65). For educational attainment we divide the labour force among illiterate workers, workers that completed compulsory school only, and more educated workers, with more than a compulsory school degree. With respect to geography we consider the five main Brazilian regions (North, North-East, South-East, South and Central-West) as well as the spatial division between urban and rural areas. Finally, we present analysis based on a standard three sector grouping of economic (or branch) activities (primary,

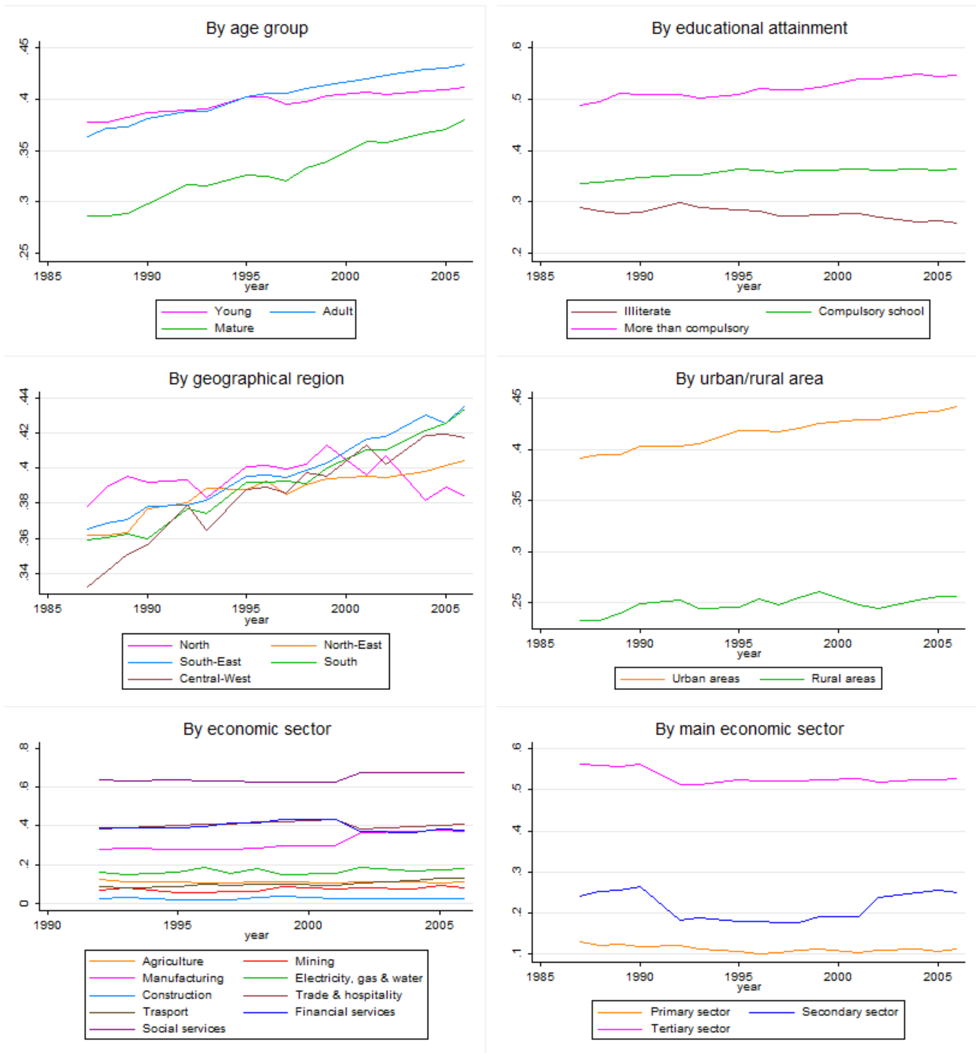
[22] Participation rates over the entire period are available on request from the author.

secondary and tertiary sectors) as well as a detailed breakdown of different economic activities: a) agricultural, forestry and fishing activities (hunting is included as well), b) mining, c) manufacturing, d) services related to electricity, gas and water provision, e) construction, f) trade activities and services related to hospitality and tourism, g) transport and storage activities, h) financial services (including insurance services) and real estate and, finally, i) social services (including health and education).

4.4.2.1 Trends in Female Participation over time

Figure 4a presents the evolution of the share of female workers as a percentage of the total labour force disaggregated by age, years of education, geographic region, urban or rural residence, and economic sectors.

Figure 4a: The evolution of the female share over time across several characteristics of the labour force (as gender ratio of each category sub-group)



Source: Author's computations using PNAD 1987- 2006.
Note: 1991, 1994 and 2001 missing years.

Looking first at age, we find, as expected, growing labour force participation among young and adult women with respect to their male peer group, but also find that the share of mature women in the labour force has risen rapidly, as women now remain in the labour market much longer, as documented in Wajnman, de Oliveira and de Oliveira (2004). With respect to years of education, the share of female workers with more than compulsory school attainment has increased, though they have always represented more than half of the total educated labour force. The message is that well educated women are much more likely to be engaged in the labour force. With respect to economic activities, the presence of women is predominantly in the tertiary sector and is especially confined to the social services, trade and hospitality and financial services sectors. Finally, female participation has increased sharply in the South and South-East regions, and primarily in urban areas.

4.4.2.2 Trends in Non-White Participation over time

Figure 4b reports the corresponding trends for non-white workers. Labour force participation of non-white workers has increased across all age groups, and particularly among young workers, with young non-white workers comprising more than half of all young people in the labour force after 2000. Non-white workers represent the predominant share of the illiterate labour force, at an average of 70%, but their share of the more educated workforce is also increasing. Among those that have attained only compulsory schooling, the share of non-white workers has increased from 47% to 59%, while their share of the workforce with more than compulsory education has similarly increased from 20% to 34% - rising educational attainment among non-white Brazilians is also documented in Osorio (2008). With respect to their distribution across economic sectors and activities, non-white workers represent a dominant share of the labour force in the primary sector, followed by the secondary sector, while non-white participation has increased steadily across all three sectors over time. At a more detailed level of disaggregation, non-white workers are most heavily concentrated in agricultural activities and in the construction and mining sectors (primarily non-white men) followed by trade and hospitality and social services (primarily non-white women). The North and North East regions record the largest share of non-white workers, while these workers also represent the majority of the rural labour force over time. The fact that the share of non-white workers in the North and North-East has remained relatively stable, despite the aggregate rise of non-white participation in the labour market, may be evidence of migration across regions during this period.[23]

Figure 4b: The evolution of the non-white share over time across several characteristics of the labour force (as racial ratio of each category sub-group)

Source: Author's computations using PNAD 1987-2006.
Note: 1991, 1994 and 2001 missing years.

[23] Brito and de Carvalho (2006) and Gomes Braga (2006) explore the features of internal migration in Brazil, and report evidence of migration towards urban areas and towards the southern regions, particularly during the 1990s. On the other hand, new work by Pochmann (2007), titled "A nova geoeconomia do emprego" reports a different trend in recent years, as new regions, such as Amazonas, Mato Grosso e Goiás (among others, primarily in the Central-West and North regions) have replaced the South and the South-East regions as the primary recipients of internal migrants.
(See also http://www.unicamp.br/unicamp/unicamp_hoje/ju/fevereiro2007/ju349pag03.html).

4.4.3 Trends in Occupational Distribution

Having looked at the representation of different groups in the labour force as a whole and disaggregated by major labour force characteristics, we now examine occupational distribution – that is, the jobs in which members of different population sub-groups are primarily employed. This brief review provides a very preliminary and descriptive account of the broad patterns of occupational segregation. What we find is that although there has been an increase in female and non-white participation in the labour market, the occupations in which these groups are primarily employed have remained relatively stable over time. However, while the share of female and non-white workers in individual occupations has remained relatively stable, certain occupations have expanded their overall share of the labour force, with the tertiary sector as a whole expanding significantly during the period under study (World Bank, 2002). This implies that a large part of the increase in female and non-white labour force participation has been within the same occupations in which they have been employed during the past two decades, which have become an increasingly important part of the overall labour force.

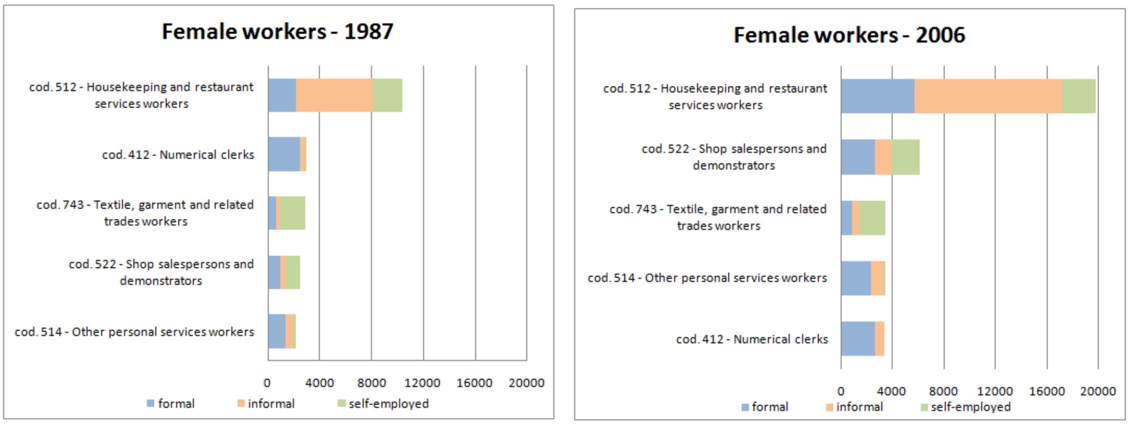
4.4.3.1 Distribution of Workers across Occupations, By Gender and Race

We begin by looking at the five occupational categories (at the 3-digit level) in which each population group (female and male workers, non-white and white workers) is most heavily employed, comparing the structure of employment between 1987 and 2006. We observe three broad features from the data. First, as noted above, all of the distributions are quite stable over time, despite the rapid entry of women and non-whites into the labour force. Second, again as noted above, employment in the tertiary sector has expanded most rapidly overall, as a large share of female and non-white entry into the labour market has been in the tertiary sector and into occupations in which they were already well represented. Finally, we find that the structure of the male labour force is noticeably different from the female labour force. Women are primarily employed in housekeeping, and in the service sector in general, while men are more frequently employed in agriculture and as workers. By contrast, the non-white and white labour forces are broadly similar, differing primarily in the low presence of white workers in agricultural activities, particularly in more recent years.

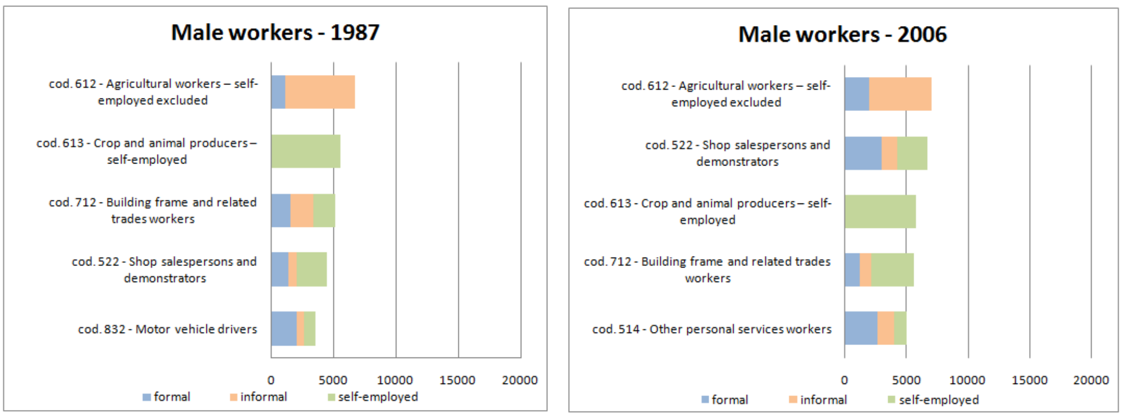
The occupational distribution disaggregated by gender (reported in Figure 5a) reveals that in 2006 women were primarily employed as housekeeping and restaurant services workers (cod. 512), and as shop salespersons and demonstrators (cod. 522).[24] The former was also the primary sector of employment in 1987, while the employment of women as shop salespersons and demonstrators has assumed greater importance over time. It is noteworthy that a significant share of women in both occupations are employed informally (informal sector and self-employed), while we see similarly high levels of informality among textile and related workers (cod. 743) and, to a lesser extent, among other personal services workers. Turning to male employment, the primary employment sectors for men remain agricultural workers (cod. 612), crop and animal producers (cod. 613), as well as shop salespersons and demonstrators (cod. 522). Interestingly, a significant share of building frame and related trades workers (cod. 712) are informal and self-employed.

Figure 5a: Distribution of female and male workers in occupations

Panel A – Female workers



Panel B – Male workers



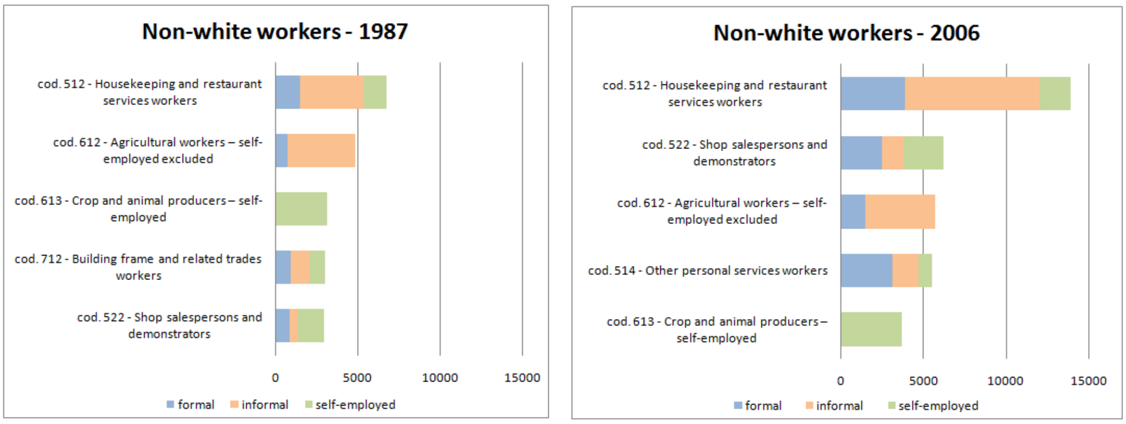
Source: Author's computations using PNAD 1987 and 2006.
Note: Panels A and B capture the number of women and men, respectively, employed in each of the top five most numerous occupations, disaggregated across the formal, informal and self-employed sectors.

[24] The 'cod. 512' includes domestic workers generally, and not only those employed in housekeeping services in hotels and other commercial enterprises.

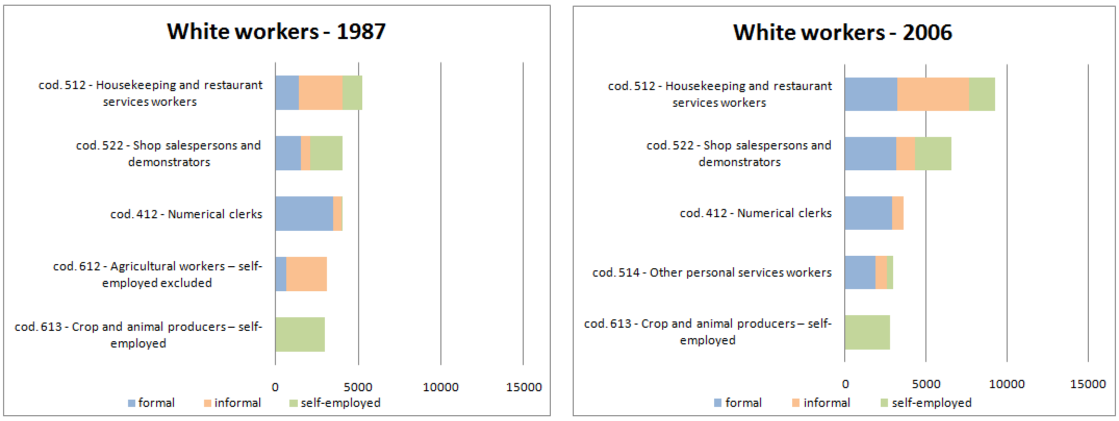
Turning to the occupational distribution disaggregated by race (reported in Figure 5b), we find that that for non-whites housekeeping and restaurant services employment (cod. 512) is predominant, and is also characterized by very high levels of informal sector employment. Non-white workers are also heavily employed as shop salespersons and demonstrators (cod. 522), and in agricultural activities (cod. 612 and cod. 613), both of which have high shares of informal sector and self-employed workers.

Figure 5b: Distribution of non white and white workers in occupations

Panel A – Non-white workers



Panel B – White workers



Source: Author's computations using PNAD 1987 and 2006.
Note: Panels A and B capture the number of women and men, respectively employed in each of the top five most numerous occupations, disaggregated across the formal, informal and self-employed sectors.

This general overview, which focuses exclusively on the top five sectors of employment for each group, concludes with two more general observations about the labour force distribution. First, consistent with Figures 5a and 5b, we find that women tend to be concentrated in a relatively small number of occupations relative to men, but we do not see a similar pattern among non-whites, who are more homogenously distributed across the occupational structure. Second, we find some evidence that non-formal employment tends to be concentrated in a smaller number of occupations than formal sector employment. This is suggestive of less diversified informal and self-employed sectors, but the difference with the formal sector is less than we might have anticipated. The absence of a larger difference is in line with the Bosch, Goni and Maloney (2007) hypothesis that informality does not simply exist in marginal sectors, but frequently expands across all sectors of the labour market.

4.4.3.2 Most Dominated Occupations by Gender and Race

While the previous section highlighted the distribution of different population groups across the labour market we now turn to the identification of occupations that are dominated overwhelmingly by particular population groups. Thus, Tables 4a and 4b highlight the occupations in which a particular group (i.e., gender and race respectively) comprises the largest share of the total workforce. This is achieved by dividing the number of women or men (non-whites or whites) in a given occupation by total employment in that occupation. This gives us a list of the most female-, male-, non-white- and white-dominated occupations, this time reporting occupations at 2-digit level in order to provide a meaningful set of findings.

Looking first at the occupational structure by gender we uncover two key points: first, there is a high level of concentration[25] in particular occupations, with female dominated occupations including roughly about 15% men. Interestingly, male-dominated occupations are almost entirely composed of male workers, as women comprise less than five per cent in these occupations. Second, while levels of concentration are high in both the formal and non-formal sectors, the specific occupations vary, particularly for women. In other words, female-dominated occupations in the formal labour market differ from female-dominated occupations in the non-formal labour market.

In 2006 the most female-dominated occupations were, in rank order, teaching associate professionals (cod. 33), teaching professionals (cod. 23), customer services clerks (cod. 42), life science and health associate professionals (cod. 32) and life science

---

[25] It is important to bear in mind that concentration and segregation are two different concepts. Quoting Siltanen, Jarman and Blackburn (1995), "concentration refers to the representation of one sex *within occupations*, segregation refers to the separation of the two sexes *across occupations"*.

and health professionals (cod. 22). The pattern does not change over time, but there are interesting differences between the formal and non-informal sectors. In the informal sector, personal services workers (cod. 51) partially substitute for customer services clerks (cod. 42), while in the self-employed sector other craft and related trades workers (cod. 74) enter into the classification. Due to the characteristics of these occupations, there is more room for informality in the case of personal services activities or craft workers than is the case for clerical occupations. The most male-dominated occupations are, in rank order, extraction and building trades workers (cod. 71), labourers in mining, construction, manufacturing and transport (93) and drivers and mobile plant operators (83), with only modest differences evident between the formal and informal sectors.

Table 4a: The most female and male dominated occupations

Panel A - Women as a Percent of the Labour Force by 3-digit occupational category

1987			2006		
All labour market					
33	93.45%	Teaching associate professionals	33	82.80%	Teaching associate professionals
42	83.13%	Customer services clerks	23	77.21%	Teaching professionals
23	77.74%	Teaching professionals	42	75.19%	Customer services clerks
32	76.01%	Life science and health associate professionals	32	74.89%	Life science and health associate professionals
51	69.03%	Personal and protective services workers	22	65.14%	Life science and health professionals
Formal sector					
33	93.13%	Teaching associate professionals	33	85.15%	Teaching associate professionals
42	83.26%	Customer services clerks	23	77.99%	Teaching professionals
32	77.73%	Life science and health associate professionals	42	77.96%	Customer services clerks
23	71.85%	Teaching professionals	32	77.03%	Life science and health associate professionals
22	50.85%	Life science and health professionals	22	67.03%	Life science and health professionals
Informal sector					
33	94.80%	Teaching associate professionals	33	82.36%	Teaching associate professionals
23	83.83%	Teaching professionals	51	79.49%	Personal and protective services workers
51	83.35%	Personal and protective services workers	23	74.37%	Teaching professionals
42	82.43%	Customer services clerks	32	74.37%	Life science and health associate professionals
32	75.92%	Life science and health associate professionals	42	70.25%	Customer services clerks
Self-employed sector					
51	85.43%	Personal and protective services workers	33	73.24%	Teaching associate professionals
33	83.46%	Teaching associate professionals	74	72.60%	Other craft and related trades workers
23	76.32%	Teaching professionals	23	65.96%	Teaching professionals
74	74.83%	Other craft and related trades workers	51	59.51%	Personal and protective services workers
73	55.63%	Precision, handicraft, craft printing and related trades workers	22	56.89%	Life science and health professionals

Panel B - Men as a Percent of the Labour Force by 3-digit occupational category

1987			2006		
All labour market					
83	99.72%	Drivers and mobile plant operators	71	99.44%	Extraction and building trades workers
71	99.37%	Extraction and building trades workers	93	99.28%	Labourers in mining, construction, manufacturing and transport
72	95.53%	Metal, machinery and related trades workers	83	98.52%	Drivers and mobile plant operators
93	95.24%	Labourers in mining, construction, manufacturing and transport	72	95.55%	Metal, machinery and related trades workers
81	92.49%	Stationary plant and related operators	61	88.46%	Skilled agricultural and fishery workers
Formal sector					
93	100%	Labourers in mining, construction, manufacturing and transport	83	99.09%	Drivers and mobile plant operators
83	99.96%	Drivers and mobile plant operators	71	98.91%	Extraction and building trades workers
71	99.23%	Extraction and building trades workers	93	98.76%	Labourers in mining, construction, manufacturing and transport
72	94.01%	Metal, machinery and related trades workers	72	93.85%	Metal, machinery and related trades workers
81	88.73%	Stationary plant and related operators	61	87.46%	Skilled agricultural and fishery workers
Informal sector					
83	99.65%	Drivers and mobile plant operators	93	99.54%	Labourers in mining, construction, manufacturing and transport
71	99.25%	Extraction and building trades workers	71	99.53%	Extraction and building trades workers
72	98.40%	Metal, machinery and related trades workers	83	98.21%	Drivers and mobile plant operators
81	95.24%	Stationary plant and related operators	72	97.56%	Metal, machinery and related trades workers
93	94.07%	Labourers in mining, construction, manufacturing and transport	61	89.57%	Skilled agricultural and fishery workers
Self-employed sector					
81	100%	Stationary plant and related operators	12	100%	Corporate managers
71	99.60%	Extraction and building trades workers	11	100%	Legislators and senior officials
83	99.19%	Drivers and mobile plant operators	71	99.67%	Extraction and building trades workers
72	98.98%	Metal, machinery and related trades workers	93	99.43%	Labourers in mining, construction, manufacturing and transport
82	98.42%	Machine operators and assemblers	72	98.55%	Metal, machinery and related trades workers

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

Turning to the race-based occupational structure, the key features are somewhat different. Most evidently, levels of concentration in individual occupations are not as pronounced. Whereas the most male-dominated occupations employ almost exclusively male workers (between 93.6% and 99.9%), the most white-dominated occupations comprise roughly 81% white workers. Turning to specific occupations, the most non-white-dominated occupations are, again in rank order, skilled agricultural and fishery workers (cod. 61), extraction and building trades workers (cod. 71) and sales and services elementary occupations (cod. 91). Meanwhile, the most white-dominated occupations are physical, mathematical and engineering science professionals (cod. 21), life science and health professionals (cod. 22), corporate managers (cod. 12) and other professionals, such as business and legal professionals (cod. 24). Consistent with the data presented so far, white-dominated occupations tend to be relatively high-skilled, while non-white individuals dominate lower skilled occupations. The latter pattern is, in turn, somewhat more pronounced in the informal sector, where non-whites make up a higher proportion of the total workforce in the occupations in which they are most dominant.

Ultimately, three issues warrant particular attention. First, as noted above, concentration of the labour force along gender lines is substantially more pronounced than concentration along racial lines. Looking at the extremes of the occupational distribution, we find there are four occupations in which women comprise less than 5% of the total labour force consistently over time, while there are not any occupations in which non-whites constitute either less than 10% or greater than 90% of the workforce. Second, we find that female-dominated occupations are generally more-skilled (such as teaching) than male dominated jobs (such as extraction and building trade workers). Along the same lines, non-white dominated occupations are generally less skilled occupations than those that are white dominated. Finally, while this analysis has focused on those occupations that are dominated by individual population groups, we do see a general decline over time, across all occupations, in the degree to which individual groups are concentrated in particular jobs. This trend towards greater homogeneity in the representation of female and non-white labourers in the labour force is an important issue to which attention turns during the analysis of the determinants of changes in different measures of segregation over time.

Table 4b: The most non-white and white dominated occupations

Panel A - Non-whites as a Percent of the Labour Force by 3-digit occupational category

1987			2006		
All labour market					
93	80.36%	Labourers in mining, construction, manufacturing and transport	61	66.23%	Skilled agricultural and fishery workers
82	57.93%	Machine operators and assemblers	71	64.84%	Extraction and building trades workers
61	57.70%	Skilled agricultural and fishery workers	91	63.38%	Sales and services elementary occupations
71	57.52%	Extraction and building trades workers	51	61.18%	Personal and protective services workers
51	55.53%	Personal and protective services workers	93	59.93%	Labourers in mining, construction, manufacturing and transport
Formal sector					
93	69.70%	Labourers in mining, construction, manufacturing and transport	61	67.10%	Skilled agricultural and fishery workers
71	55.13%	Extraction and building trades workers	91	66.67%	Sales and services elementary occupations
61	53.44%	Skilled agricultural and fishery workers	71	62.57%	Extraction and building trades workers
51	52.63%	Personal and protective services workers	51	58.37%	Personal and protective services workers
82	52.26%	Machine operators and assemblers	82	52.60%	Machine operators and assemblers
Informal sector					
93	82.96%	Labourers in mining, construction, manufacturing and transport	61	74.20%	Extraction and building trades workers
82	64.72%	Machine operators and assemblers	71	70.46%	Skilled agricultural and fishery workers
71	63.46%	Extraction and building trades workers	93	65.44%	Labourers in mining, construction, manufacturing and transport
61	62.37%	Skilled agricultural and fishery workers	51	65.18%	Personal and protective services workers
73	60.69%	Precision, handicraft, craft printing and related trades workers	91	64.84%	Sales and services elementary occupations
Self-employed sector					
73	65.00%	Precision, handicraft, craft printing and related trades workers	91	63.16%	Sales and services elementary occupations
82	61.67%	Machine operators and assemblers	71	62.30%	Extraction and building trades workers
81	55.09%	Stationary plant and related operators	42	61.80%	Customer services clerks
71	54.38%	Extraction and building trades workers	93	61.36%	Labourers in mining, construction, manufacturing and transport
61	53.88%	Skilled agricultural and fishery workers	61	59.54%	Skilled agricultural and fishery workers

Panel B - Whites as a Percent of the Labour Force by 3-digit occupational category

1987			2006		
All labour market					
22	82.23%	Life science and health professionals	21	77.81%	Physical, mathematical and engineering science professionals
21	80.98%	Physical, mathematical and engineering science professionals	22	75.36%	Life science and health professionals
23	74.71%	Teaching professionals	12	70.46%	Corporate managers
24	74.45%	Other professionals (business and legal)	13	68.44%	General managers
12	73.87%	Corporate managers	24	66.66%	Other professionals (business and legal)
Formal sector					
22	80.87%	Life science and health professionals	21	76.78%	Physical, mathematical and engineering science professionals
21	80.58%	Physical, mathematical and engineering science professionals	22	74.37%	Life science and health professionals
24	77.04%	Other professionals (business and legal)	24	71.75%	Other professionals (business and legal)
23	75.95%	Teaching professionals	12	71.70%	Corporate managers
12	75.40%	Corporate managers	13	68.90%	General managers
Informal sector					
22	81.15%	Life science and health professionals	21	74.40%	Physical, mathematical and engineering science professionals
21	79.55%	Physical, mathematical and engineering science professionals	22	72.56%	Life science and health professionals
23	72.60%	Teaching professionals	12	66.06%	Corporate managers
11	71.81%	Teaching professionals	13	65.23%	General managers
34	69.36%	Legislators and senior officials	24	61.11%	Other professionals (business and legal)
Self-employed sector					
23	89.47%	Teaching professionals	11	100%	Legislators and senior officials
22	85.86%	Life science and health professionals	12	100%	Corporate managers
21	85.07%	Physical, mathematical and engineering science professionals	21	84.69%	Physical, mathematical and engineering science professionals
24	74.95%	Other professionals (business and legal)	22	80.92%	Life science and health professionals
41	74.00%	Office clerks	13	77.46%	General managers

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

4.4.4. Robustness checks

Although the preceding section has been largely descriptive, the ranking of the most male-, female-, white- and non-white-dominated jobs, reported in tables 4a and 4b, is nonetheless sensitive to the particular methodology employed. The results reported above are based on ranking, in decreasing order, the frequencies of the relevant sub-groups of the population in the total labour force. In order to check the robustness of these results, we assess three alternative ranking methodologies.

The first additional criterion adopts the ratio of total female workers over total labour force (*F/T*) as the threshold. The female-dominated occupations can be considered all occupations for which the F_i/T_i is greater than *F/T* plus 10% and the male-dominated occupations are all occupations whose frequencies is lower than *F/T* minus 10% (Flückiger and Silber, 1999). All remaining occupations are defined as mixed occupations. The same exercise can be conducted for non-white- and white-dominated occupations by using the ratio of total non-white workers over total labour force (*B/T*).

An alternative criterion called Oppenheimer's approach (Oppenheimer, 1979) exploits the ratio of female workers over male workers by occupation, the so-called gender ratio (F_i/M_i). An occupation is considered disproportionately-female if its gender ratio is greater than one and disproportionately-male if its gender ratio is smaller than 0.25, otherwise it is defined as well-represented. The same can be applied to race, by looking at the racial ratio as the ratio of non-white workers over white workers by occupation (B_i/W_i).

Finally, we also consider the marginal matching methodology that uses a different approach to categorize occupations (Blackburn, Jarman and Siltanen, 1993). By gender, the procedure requires that first all occupations are ranked by decreasing order of F_i/M_i. The first occupations whose total accumulated number of workers is equal to the total female labour force are categorized as female occupations. Ranking occupations in decreasing order of F_i/M_i, the female occupations are all occupations for which the cumulative of F_i+M_i is equal to F, where *i* are the occupations. Hence, the total number of workers in female occupations, T_f, is equal to the total female labour force, *F*. As a consequence, the total number of workers in male occupations, T_m, is equal to the total male labour force, *M*. Once again, we can apply the same methodology to jobs using racial groups, where the total number of workers, T_b, is equal to the non-white labour force, B, and white-occupations, whose total number of workers, T_w, is equal to the white labour force, W.

Table 5: Robustness check for the definition of most dominated occupations

Our method (baseline) - group-specific share as percent of the employed labour force, ordered by decreasing values, only first five occupation kept.

Female-dominated		**Male-dominated**		**Non-white-dominated**		**White-dominated**	
1987	**2006**	**1987**	**2006**	**1987**	**2006**	**1987**	**2006**
33	33	83	71	93	61	22	21
42	23	71	93	82	71	21	22
23	42	72	83	61	91	23	12
32	32	93	72	71	51	24	13
51	22	81	61	51	93	12	24

1st alternative criterion (Flückiger and Silber, 1999) - male-dominated: Fi/Ti<F/T-10%; mixed:F/T-10%<Fi/Ti<F/T+10%; female-dominated Fi/Ti>F/T+10%. The same applied to race.

Female-dominated		**Male-dominated**		**Non-white-dominated**		**White-dominated**	
1987	**2006**	**1987**	**2006**	**1987**	**2006**	**1987**	**2006**
13	12	12	21	13	61	11	11
23	22	21	31	61	71	12	12
32	23	31	61	71	91	21	13
33	32	34	71	82		22	21
41	33	61	72	93		23	22
42	41	71	73			24	23
51	42	72	81			34	24
74	51	81	82			41	34
	74	82	83			42	41
		83	93				
		93					

Note: All occupations of our baseline are considered with exception for non-white dominated: 51 is not included in both years and 93 is not included in 2006.

2nd alternative criterion (Oppenheimer, 1979) - disproportionately-female: Fi/Mi>1; well-represented: 0.25<Fi/Mi<1; disproportionately-male: Fi/Mi<0.25. The same applies to race.

Female-dominated		**Male-dominated**		**Non-white-dominated**		**White-dominated**	
1987	**2006**	**1987**	**2006**	**1987**	**2006**	**1987**	**2006**
13	12	21	61	13	33	21	
23	22	31	71	51	51	22	
32	23	61	72	61	61		
33	24	71	81	71	71		
42	32	72	82	81	72		
51	33	81	83	82	74		
74	41	82	93	93	83		
	42	83			91		
	51	93			93		
	74						

Note: All occupations of our baseline are considered with exception for white dominated: 23, 24, 12 not included in 1987 and 21, 22, 12, 13, 24 non included in 2006

3rd alternative criterion (Blackburn, Jarman and Siltanen 1993) - marginal matching methodology, see text.

Female-dominated		**Non-white-dominated**	
1987	**2006**	**1987**	**2006**
33	33	93	61
42	23	82	71
23	42	61	91
32	32	71	51
51	51	51	93
74	22	81	82
13	12	13	83
	24		33

Note: All occupations included. It is important to highlight that once the female-dominated occupations are defined, the male-dominated are the ones left over. The same approach applies to race.

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

By applying these additional criteria we obtain the patterns reported in table 5 which appear broadly similar to our previous findings. This is particularly the case for gender. By race, the rankings are slightly less clear-cut. This is likely attributable to the fact that white and non-white workers are distributed more homogenously across occupations so the categorizations become sensitive to the criteria adopted.

4.5 Conclusions

The first half of this chapter has been devoted to introducing and describing the construction of a new harmonized occupational classification based on the PNAD datasets. The existing PNAD datasets employ occupational classifications that are neither consistent over time nor compatible with the internationally recognized system, the ISCO-08. By contrast, the new harmonized re-classification for this study covers 83 occupations at the 3-digit level, employs a consistent occupational classification across 20 years, and is harmonized with the international classification system. This has required not only matching existing occupational codes from the PNAD datasets to their appropriate international classification, but also carefully checking the new classification for consistency. This has meant ensuring that the occupational groupings drawn from the PNAD, and being aggregated together, are both sufficiently homogeneous within groups, and consistent over time given changes in the classification system used by the PNAD surveys. The creation of this new harmonized dataset makes it possible to analyse labour market trends over a significantly longer period than has previously been possible, and provides benchmark for incorporating data beyond the 20 years covered here. It also provides the basis for more direct international comparisons.

In this chapter we have constructed a new re-classification of the occupational codes in order to explore the evolution of the Brazilian occupational structure disaggregated by both gender and race and with a focus on differences between the formal, informal and self-employed sectors over time. Over the last two decades the Brazilian labour market has exhibited a large increase in female participation, generating an increase in the female share of the workforce from 36% in 1987 to 42% in

2006. This increase is well-documented[26] and this pattern is common for many South American countries where the gap between male and female participation has narrowed more than in any other region in the developing world.[27] The composition of the labour force by race is also changing, as the non-white share has continued to increase over time and in 2003 non-white workers became the majority of the workforce.[28]

As a share of the entire labour force, the informal sector has remained relatively constant at the national level between 1987 and 2006, with informality concentrated in metropolitan areas, as reported by Ramos and Ferreira (2005). During the period under study the female proportion in the informal sector has increased significantly, and this is particularly true for non-white women. It may be that key features of informal sector employment, such as its flexibility, lower commitment to long-term job positions and higher turnover, are well suited to female labour supply given the nature of female preferences and tastes. On the other hand, the informal market may exploit the lack of choice available to less skilled female workers, and particularly to non-white female workers, who are disproportionately employed in personal services (such as housekeepers).

Despite increasing female and non-white participation in the labour market, the overall occupational distribution for the entire labour market has remained surprisingly stable over the last two decades. That is, the distribution of men, women, white and non-white workers within individual occupations has not undergone a dramatic change over time. Moreover, although non-formal employment tends to be concentrated in a relatively smaller number of occupations when compared to formal employment, differences in the occupational distribution between the formal, informal and self-employed sectors are generally small. This finding is consistent with previous studies that have argued that the rise of informality in metropolitan areas has occurred primarily within industries (Ramos and Ferreira, 2005; Bosch, Goni and Maloney, 2007). Although the occupational distribution has remained relatively stable over time this disguises underlying changes, particularly the persistent expansion of the tertiary sector. The combination of a rapidly expanding tertiary sector and a relatively stable

[26] For the analysis of the Brazilian female participation in the labour market see Wajman and Rios Neto (2000a, 2000b) and references there in.

[27] For more details see: ILO, *Global Employment Trends for Women, Brief* (Geneva, 2007); website: http://www.ilo.org/trends.

[28] Whether the number of non-white individuals entering the labour force has really increased over time or whether the number of individuals among work force that report themselves as non-white population is increasing cannot be determined, as in the PNAD dataset for the race/skin colour categorization is self-reported. This finding should thus be interpreted with caution.

occupational distribution reflects the fact that female and non-white entrants to the labour market have disproportionately joined occupations that were already female and non-white dominated, particularly in the tertiary sector.

Turning to differences in occupational distribution by gender and race, the data reveal that the occupational distribution between white and non-white workers is relatively similar, with much lower levels of concentration than is the case by gender. On the surface, at least, this appears to signal relatively equal access to occupations for non-white workers, at least based on a 3-digit classification. By contrast, there are major differences in occupational distribution between men and women. Interestingly, the degree of concentration increases slightly when we restrict analysis to the informal sector, which may reflect a less diversified informal labour market. While the fact that female and male occupational distributions are highly concentrated may suggest unequal opportunities, we also need to take into account differences in tastes and preferences among female and male workers. In other words, there are some jobs that are male-dominated not as a result of discrimination but because women are not inclined towards those particular jobs, and vice-versa (see also the discussion in Bertrand, 2010). As a result, while occupational concentration appears less severe by race than by gender, racial occupational concentration may actually be a more serious problem, as it cannot be as easily explained by differences in tastes and preferences. That said, we observe a clear and encouraging decline over time in concentration by both gender and race in the majority of the most concentrated occupations.

There are also noteworthy trends in the employment of different groups in high and less-skilled occupations. With respect to race, white-dominated occupations tend to be highly-skilled, while non-white-dominated jobs tend to be less-skilled. More surprising is that the most female dominated occupations are highly skilled occupations, whereas men tend to dominate less-skilled occupations. This would appear to provide evidence that women have access to high-skilled formal sector employment, and thus do not face significant labour market discrimination. Yet, there is also an alternative interpretation. In a study using 1980 data Telles (1992b) claimed that *"[...] education and race are more frequently used in screening women's than men's entrance into the formal sector"*. His inference is that women will face barriers to entry into the formal labour market if they lack a high level of education or are non-white, whereas men do not face similar barriers. As such, the high concentration of women in high-skilled formal sector jobs may reflect not unfettered access to opportunities, but the fact that opportunities for women in the formal sector are constrained outside of these highly skilled professions. At a superficial and descriptive level the data support this hypothesis: women, and particularly non-white women, who are employed in less-skilled jobs are more likely than men to be employed in the informal sector, suggesting that such women may face greater barriers to entering formal employment. This would potentially account for the fact that the participation of non-white women is growing particularly rapidly in the informal sector, often in the personal services sector as housekeepers. Although there has been some 'formalization' of this type of occupation, it remains a likely candidate for informality.[29]

[29] NYT article http://www.nytimes.com/2011/05/20/world/americas/20brazil.html?_r=1

Appendix to Chapter 4

Table A1: Classification of occupational codes

ISC O88	MAJOR, SUB-MAJOR, MINOR AND UNIT GROUPS	PNAD ’02-‘06	PNAD ’92-‘01	PNAD ’87-‘90
	MAJOR GROUP 1: LEGISLATORS, SENIOR OFFICIALS AND MANAGERS			
11	**Legislators and senior officials**			
111	Legislators	1111; 1112; 1113	20/21	20/21
112	Senior government officials	1122; 1123	-	-
113	Traditional chiefs and heads of villages	1130	-	-
114	Senior officials of special-interest organisations	1140	-	-
12	**Corporate managers**			
121	Directors and chief executives	1210; 1219	30/39 40	30/39 40
122	Production and operations department managers	1220	-	-
123	Other department managers	1230	-	-
13	**General managers**			
131	Managers of small enterprises	1310; 1320	1/12 15	1/12 15
	MAJOR GROUP 2: PROFESSIONALS			
21	**Physical, mathematical and engineering science professionals**			
211	Physicists, chemists and related professionals	2131; 2132; 2133; 2134	121 123	121 123
212	Mathematicians, statisticians and related professionals	2111; 2112	171 172 124 125 203	171 172 124 125 203
213	Computing professionals	2121; 2122; 2123; 2124; 2125	173	173
214	Architects, engineers and related professionals	2140; 2141; 2142; 2143; 2144; 2145; 2146; 2147; 2148; 2149	101 102 103 104 112	101 102 103 104 112
		2011; 2012		
		2021		
22	**Life science and health professionals**			
221	Life science professionals	2211	141 142 143	141 142 143
		2221		
222	Health professionals (except nursing)	2231; 2232; 2233; 2234	122 144 151 152	122 144 151 152
223	Nursing and midwifery professionals	2235; 2236; 2237	153 154	153 154
23	**Teaching professionals**			
231	College, university and higher education teaching professionals	2340	211 212	211 212
232	Secondary education teaching professionals	2321	213	213
		2330		
233	Primary and pre-primary education teaching professionals	2311; 2312; 2313	214	214
234	Special education teaching professionals	2391; 2392	219	219
235	Other teaching professionals	2394	221	221
24	**Other professionals**			
241	Business professionals	2521; 2522; 2523; 2524; 2525	182 183	182 183
		2531		
242	Legal professionals (Lawyers and Judges)	2410; 2412; 2419	231/233	231/233
		2421; 2422; 2423		
243	Archivists, librarians and related information professionals	2612; 2613; 2614	291 292	291 292
244	Social science and related professionals	2511; 2512; 2513; 2514; 2515; 2516	181 201 202 204 205	181 201 202 204 205

245	Writers and creative or performing artists	2611; 2615; 2616; 2617	261 111 271 272 273 275 276 278 279	261 111 271 272 273 275 276 278 279
		2621; 2622; 2623; 2624; 2625; 2627		
246	Religious professionals	2631	251 252	251 252
247	Public service administrative professionals	n.a.	n.a.	n.a.
	MAJOR GROUP 3: TECHNICIANS AND ASSOCIATE PROFESSIONALS			
31	**Physical and engineering science associate professionals**			
311	Physical and engineering science technicians	3111; 3112; 3113; 3114; 3115; 3116; 3117	131 133 113 112 401/406 589	131 133 113 112 401/406 589
		3001; 3003		
		3011; 3012		
		3121; 3122; 3123		
		3131; 3132; 3134; 3135; 3136; 3137		
		3141; 3142; 3143; 3144; 3146; 3147		
		3161; 3162; 3163		
		3189		
		3191; 3192		
312	Computer associate professionals	3171; 3172	194	193
313	Optical and electronic equipment operators	n.a.	n.a.	n.a.
314	Ship and aircraft controllers and technicians	2151; 2152; 2153	711 721 722 723 741 742 761	711 721 722 723 741 742 761
		3411; 3412; 3413		
		3421; 3422; 3423; 3424; 3425; 3426		
315	Safety and quality inspectors	3516; 3517	571 51 588	571 51 588
		3522; 3523; 3524; 3525		
		3911; 3912		
32	**Life science and health associate professionals**			
321	Life science technicians and related associate professional	3210; 3211; 3212; 3213; 3214	302	302
		3201		
322	Modern health associate professionals	3221; 3223; 3224; 3225	132 161 163 164 165 167 168	132 161 163 164 165 167 168
		3231; 3232		
		3241; 3242		
		3250; 3251; 3252; 3253		
		3281		
323	Nursing and midwifery associate professionals	3222	162	162
324	Traditional medicine praticioners and faith healers	n.a.	n.a.	n.a.
33	**Teaching associate professionals**			
331	Primary education teaching associate professionals	3311; 3312; 3313	215 216	215 216
332	Pre-primary education teaching associate professionals	3321; 3322	217	217
333	Special education teaching associate professionals	3331	218	218
334	Other teaching associate professionals	3341	222	222
34	**Other associate professionals**			
341	Finance and sales associate professionals	3531; 3532	643	643
342	Business services agents and trade brokers	3541; 3542; 3543; 3544; 3545; 3546; 3547; 3548	641 642 644 645 646 631/633	641 642 644 645 646 631/633
343	Administrative associate professionals	3511; 3512; 3513; 3514	191 192 241 242 243 244	191 192 241 242 243 244

344	Customs, tax and related government associate professionals	3515	50 918	50 918
345	Police inspectors and detectives	3518	868	858
346	Social work associate professionals	n.a.	n.a.	n.a.
347	Artistic, entertainment and sports associate professionals	3711; 3712; 3713	280 274 281 282 283 277 831/834	280 274 281 282 283 277 831/834
		3721; 3722; 3723		
		3731; 3732		
		3741; 3742; 3743		
		3751		
		3761; 3762; 3763; 3764; 3765		
		3771; 3772; 3773		
348	Religious associate professionals	n.a.	n.a.	n.a.
	MAJOR GROUP 4: CLERKS			
41	**Office clerks**			
411	Secretaries and keyboard-operating clerks	4121; 4122; 4123	56 59 57 64	56 59 57 64
		4101		
412	Numerical clerks	4110	52 53 58 60 65	52 53 58 60 65
		4131; 4132		
		4102		
413	Material-recording and transport clerks	4141; 4142	54 55	54 55
414	Library, mail and related clerks	4151; 4152	61 62 771 772 775	61 62 771 772 775 776
419	Other office clerks	n.a.	n.a.	n.a.
42	**Customer services clerks**			
421	Cashiers, tellers and related clerks	4201	603 912 193	603 912
		4211; 4212; 4213; 4214		
		4231		
		4241		
422	Client information clerks	4221; 4222; 4223	63 774 773	63 774 773
	MAJOR GROUP 5: SERVICE WORKERS AND SHOP AND MARKET SALES WORKERS			
51	**Personal and protective services workers**			
511	Travel attendants and related workers	5111; 5112	712 752	712 752
512	Housekeeping and restaurant services workers	5121	801/808 811/818 821/825 926	801/808 811/818 821/825
		5131; 5132; 5133; 5134		
		5151; 5152		
		5161; 5162; 5165; 5166; 5167; 5169		
		5101; 50102		
514	Other personal services workers	5141; 5142	166 826 842 844 845 915 916 919 920 293	166 826 842 844 845 915 916 919 920 293
		5191; 5192; 5198; 5199		
515	Astrologers, fortune tellers and related workers	n.a.	n.a.	n.a.
516	Protective services workers	5171; 5172; 5173; 5174	841 843 869 913 917	841 843 859 913 917
		5103		
52	**Models, salespersons and demonstrators**			
521	Fashion and other models	n.a.	n.a.	n.a.
522	Shop salespersons and demonstrators	5211	851 852 601 602	601 602
		5221		
		5231		
		5201		
523	Stall and market salespersons and demonstrators	5241	604 605	604 605
	MAJOR GROUP 6: SKILLED AGRICULTURAL AND FISHERY WORKERS			

61	**Skilled agricultural and fishery workers**			
611	Market gardeners and crop growers	n.a.	n.a.	n.a.
612	Agricultural workers - own account excluded -MODIFIED	6201	304 305	304 305
		6210		
		6229		
		6239		
613	Crop and animal producers - own account	6110	301	301
		6129		
		6139		
614	Forestry and related workers	6329	331/336	331/336
		6301		
615	Fishery workers, hunters and trappers	6319	321/322	321/322
	MAJOR GROUP 7: CRAFT AND RELATED TRADES WORKERS			
71	**Extraction and building trades workers**			
711	Miners, shotfirers, stone cutters and carvers	7111; 7112; 7113; 7114	341 345 351 361 371 381 391 578	341 345 351 361 371 381 391 578
		7121; 7122		
		7101		
712	Building frame and related trades workers	7152; 7153; 7154; 7155	511/513 519	511/513 519
		7102		
713	Building finishers and related trades workers	7161; 7162; 7163; 7164; 7165	515/518 521 587	515/518 521 587
		7156; 7157		
		7151		
714	Painters, building structure cleaners and related trades workers	7166	514 520	514 520
		7170		
72	**Metal, machinery and related trades workers**			
721	Metal molders, welders, sheet-metal workers, structural-metal preparers, and related trades workers	7221; 7222; 7223; 7224	411/415 426/428 581	411/415 426/428 581
		7231; 7232; 7233		
		7241; 7242; 7243; 7244; 7245; 7246		
		7201		
722	Blacksmiths, tool-makers and related trades workers	7211; 7212; 7213; 7213; 7214; 7215	429/431 416/422	429/431 416/422
723	Machinery mechanics and fitters	7250; 7251; 7252; 7253; 7254; 7255; 7256; 7257	423/425 921 923	423/425 921 923
		7202		
		8201		
		8211; 8212; 8213; 8214		
		8221		
		9101; 9102; 9109		
		9111; 9112; 9113		
		9131		
		9141; 9142; 9143; 9144		
		9151; 9152; 9153; 9154		
		9191; 9192; 9193		
724	Electrical and electronic equipment mechanics and fitters	7301	501/508	491/498
		7311; 7312; 7313		
		7321		
		9501; 9502; 9503		

		9511; 9513		
		9531		
		9541; 9542; 9543		
73	**Precision, handicraft, craft printing and related trades workers**			
731	Precision workers in metal and related materials	7401	572 573	572 573
		7411		
		7421		
		7501; 7502		
		7519		
732	Potters, glass-makers and related trades workers	7521; 7522; 7523; 7524	561/564	561/564
		8231; 8232; 8233		
		8281		
		8202		
733	Handicraft workers in wood, textile, leather and related materials	7601; 7602; 7603; 7604; 7605	441/450 452 461 462 585	441/450 452 461 462 585
		7610; 7611; 7612; 7613; 7614; 7618		
		7620; 7621; 7622; 7623		
		8311		
		8321		
		8339		
		8301		
734	Craft printing and related trades workers	7606	451 551/557	451 551/557
		7660; 7661; 7662; 7663; 7664		
		7686; 7687		
74	**Other craft and related trades workers**			
741	Food processing and related trades workers	8411; 8412; 8413; 8416; 8417	531/545 579 580	531/545 579 580
		8421; 8423; 8429		
		8484; 8485		
		8491; 8492; 8493		
		8401		
742	Wood treaters, cabinet-makers and related trades workers	7711	481/486 489 490 577	481/486 489 490 577
		7721		
		7731; 7732; 7733; 7734; 7735		
		7741		
		7751		
		7764		
		7771; 7772		
		7701		
743	Textile, garment and related trades workers	7630; 7631; 7632; 7633	470/476 487 488	470/476 487 488
		7650; 7652		
		7681; 7682; 7683		
744	Pelt, leather and shoemaking trades workers	7640; 7641; 7642; 7643	477 478 479	477 478 479
		7651; 7653; 7654		
	MAJOR GROUP 8: PLANT AND MACHINE OPERATORS AND ASSEMBLERS			
81	**Stationary plant and related operators**			
811	Mining and mineral-processing-plant operators	n.a.	n.a.	n.a.
812	Metal-processing plant operators	n.a.	n.a.	n.a.
813	Glass, ceramics and related plant operators	n.a.	n.a.	n.a.
814	Wood-processing- and papermaking-plant operators	n.a.	n.a.	n.a.

815	Chemical-processing-plant operators (and machine operators)	8110; 8111; 8112; 8113; 8114; 8115; 8116; 8117; 8118	574/576 586	574/576 586
		8121		
		8101; 8102; 8103		
		8131		
		8181		
816	Power-production and related plant operators (and machine operators)	8611; 8612	509 583 724 744 753 922	499 583 724 744 753 922
		8621; 8622; 8623; 8624; 8625		
		8601		
817	Industrial robot operators	n.a.	n.a.	n.a.
82	**Machine operators and assemblers**			
821	Metal- and mineral-products machine operators	n.a.	n.a.	n.a.
822	Chemical-products machine operators	n.a.	n.a.	n.a.
823	Rubber- and plastic-products machine operators	n.a.	n.a.	n.a.
824	Wood-products machine operators	n.a.	n.a.	n.a.
825	Printing-, binding- and paper-products machine operators	n.a.	n.a.	n.a.
826	Textile-, fur- and leather-products machine operators	n.a.	n.a.	n.a.
827	Food and related products machine operators	n.a.	n.a.	n.a.
828	Assemblers	n.a.	n.a.	n.a.
829	Other machine operators not elsewhere classified	7801	582 584 911 914 924	582 584 911 914 924
		7811; 7813; 7817		
		7820; 7821; 7822		
		7841; 7842		
		8711		
83	**Drivers and mobile plant operators**			
831	Locomotive engine drivers and related workers	7826	743 745 746 761	743 745 746 761
		7831		
832	Motor vehicle drivers	7823; 7824; 7825	751	751
833	Agricultural and other mobile plant operators	7828	303	303
		6410; 6420; 6430		
834	Ships' deck crews and related workers	7827	725 726 727 731 732	725 726 727 731 732
		7832		
	MAJOR GROUP 9: ELEMENTARY OCCUPATIONS			
91	**Sales and services elementary occupations**			
911	Street vendors and related workers	5242; 5243	13 14 611/617 621	13 14 611/617 621
916	Garbage collectors and related labourers	n.a.	n.a.	n.a.
92	**Agricultural, fishery and related labourers**			
921	Agricultural, fishery and related labourers	n.a.	n.a.	n.a.
93	**Labourers in mining, construction, manufacturing and transport**			
931	Mining and construction labourers	n.a.	n.a.	n.a.
932	Manufacturing labourers	n.a.	n.a.	n.a.
933	Transport labourers and freight handlers	9911; 9912; 9913; 9914	925 762	925 762
		9921; 9922		
	MAJOR GROUP 0: ARMED FORCES			
100	Armed forces	0100; 0200; 0300	861/867	851/857
		0401; 0403		
		0411; 0413		
		0501; 0503		
		0511; 0513		

998	**Mal definidas**	9988	927	926
999	**Nao declarada**	9999	928	927

Source: Author's computations using PNADs.
The categories highlighted in yellow have been omitted because it was not possible to find the related occupations across all Brazilian datasets. They are "247 – Public service administrative professionals", "313 – Optical and electronic equipment operators", "324 – Traditional medicine practitioners and faith healers", "346 – Social work associate professionals", "328 – Religious associate professionals", "419 - Other office clerks", "515 - Astrologers, fortune tellers and related workers" and "521 - Fashion and other models", "611 – Market gardeners and crop growers", "811 – Mining and mineral-processing-plant operators", "812 – Metal-processing plant operators", "813 – Glass, ceramics and related plant operators", "814 – Wood-processing and papermaking- plant operators", "817 – Industrial robot operators", "821 – Metal- and mineral-products machine operators", "822 – Chemical-products machine operators", "823 – Rubber- and plastic-products machine operators", "824 – Wood-products machine operators", "825 - Printing-, binding- and paper-products machine operators", "826 – Textile-, fur- and leather-products machine operators", "827 – Food and related products machine operators", "828 – Assemblers", "916 – Garbage collectors and related laborers", "Agricultural, fishery and related laborers", "931 – Mining and construction laborers", "932 – Manufacturing laborers".

Chapter 5

The Evolution of Occupational Segregation by Gender and Race in Brazil – 1987 to 2006

5.1 Introduction

Occupational segregation represents one of the core themes in the labour economics literature and it has been the subject of many theoretical and empirical studies over several decades.[30] However, despite the centrality of occupational segregation to any understanding of labour market outcomes, studies of occupational segregation have been very rare for developing countries. One of the reasons for the absence of studies in such countries has been the absence of sufficiently detailed and reliable data over time. This chapter seeks to address this gap in the existing research with a focus on Brazil, and makes three principle contributions. First, it assesses the magnitude of occupational segregation both by gender and by race using several well-known indices of segregation, while exploring trends in segregation separately within the formal, informal and self-employed sectors. Second, the analysis of occupational segregation is conducted not only on an aggregate basis, but also disaggregated by several key characteristics of the labour force in order to identify specific demographic, educational, sectoral and spatial patterns. Finally, having presented the description of trends in labour market structure and occupational segregation, we present an initial discussion of possible determinants of occupational segregation over time. To that end we apply a decomposition technique developed by Deutsch, Flueckiger and Silber (2009) designed to assess the main forces driving changes in occupational segregation over time.

[30] See, for example, Anker (1997), Reardon and Firebaugh (2002), Fryer (2010) for theories of occupational segregation by both gender and ethnicity and King (1992), Charles and Grusky (1995) and Watts (1997), among others, for influential empirical studies of these questions.

The analysis of this question in the past has been constrained by data availability, as a revision of the classification of occupational codes used in the Brazilian household survey, the *Pesquisa Nacional por Amostra do Domicilios* (PNAD), at the beginning of this decade has prevented consistent comparison over time. A study by de Oliveira (2001) is the only study to date to have analysed segregation across a detailed range of occupations in Brazil over a fairly protracted period of time (from 1981 to 1997). However, that study does not cover the last decade and focuses only on gender segregation, thus neglecting the importance of the racial dimension. A study by Telles (1994) is similarly the only one to have investigated occupational inequality by race in Brazil, with a specific focus on exploring the impact of industrialization on racial inequality, but that analysis is based on only three occupational groups (Telles, 1994). A key contribution of this chapter is to compute levels of both gender and racial occupational segregation using a detailed occupational classification that is valid for the entire period 1987-2006. This is made possible by employing the harmonized re-classification of occupational codes described in the previous chapter, and makes this the first study, to the author's knowledge, to explore both gender and race based occupational segregation over such an extended time period.

Our analysis reveals that while gender segregation is significantly greater than racial segregation, it has fallen more rapidly over the last two decades. The persistence of racial segregation, which cannot be easily explained by differences in preferences and tastes, is a potentially worrying trend. When we disaggregate the analysis between the formal and informal sectors we find a more rapid decline in both gender and racial segregation in the formal sector, while racial segregation in particular has experienced a negligible decline the informal sector. The implementation of the Shapley decomposition proposed by Deutsch, Flueckiger and Silber (2009) offers further insight, as we find that the decline in both gender and race based segregation is primarily the result of the more homogenous representation of women and non-whites within occupations. This is a significant finding, as it is sometimes suggested that declining occupational segregation is simply the result of increasing labour market participation by women and non-whites, and thus not reflective of broaders changes in labour market outcomes. However, we find that this simple explanation does not hold. The entry of women and non-whites into the labour market has, if anything, increased segregation, as many new entrants to the labour force have joined traditionally more segregated occupations, which have increased in size over time. The aggregate decline in segregation is thus driven by the general improvement in composition within individual occupations, which represents a more 'real', and encouraging, change.

While the chapter is primarily focused on presenting novel descriptive data about the evolution of the Brazilian labour market, an important related question is whether policy reforms over the past 20 years have contributed to reducing occupational segregation, and whether they have exerted a differential impact on the formal and non-formal (i.e., informal and self-employed sectors) labour markets in Brazil. It is important to stress that the aim of this chapter is not to establish clear causality or to estimate the impact of institutional and macroeconomic reforms (or exogenous shocks) on occupational segregation. The primary aim is to describe the phenomenon of occupational segregation and its evolution across the formal and non-formal sectors. This, though, makes it possible to then suggest some tentative propositions about the possible impact of reform on patterns of occupational segregation.

In recent decades several institutional and macroeconomic shocks have affected Brazil and its labour markets. These reforms have included the establishment of a new Constitution in 1988, structural economic reforms in the 1990s, including significant trade reform, and negative external shocks at the end of the last decade. These reforms have impacted the Brazilian labour market in terms of inter-sectoral composition, the degree of competitiveness, labour market flexibility, the level social protection for workers. Economic reforms and shocks are thus likely to have contributed to shaping levels of, and changes in, occupational segregation.

Alongside these broader economic changes, since the late 1980s the Brazilian government has introduced a range of anti-discrimination legislation (ADL, hereafter) aimed at reducing occupational discrimination and, by extension, segregation, as discussed in detail in Chapter 2. Unfortunately, while exploring the role of these reforms on the Brazilian labour market is interesting, it is difficult to disentangle the effect of legislative reforms from broader economic and cultural changes. Various approaches have previously been adopted in an attempt to disentangle the effects of legislation, including the adoption of time dummies to account for the passage of specific anti-discrimination legislation (see, among others, Ashenfelter, 1970; Zabalda and Tzannatos, 1985) or relying on an ad hoc quasi-experiment exploiting variation in the passage of anti-discrimination laws across time and states (Neumark and Stock, 2006). In the Brazilian case the use of time dummies is likely to be relatively unreliable due to the difficultly of isolating the effect of new legislation from other economic and institutional shocks that occurred concurrently. On the other hand, the quasi-experiment approach, which exploits variation across time and states, is potentially very appealing but does not appear to be feasible due to the lack of data the case of Brazil.

Given the limitations of available approaches we aim to gain insight into the potential impact of ADL indirectly by investigating occupational segregation separately in the formal and non-formal sectors. The formal sector provides the framework for regulated labour markets to function, and it is in the formal sector that ADL is expected to have the greater impact. As a consequence, we would expect different outcomes in terms of gender and racial differentials across the formal, informal and self-employed sectors. It is important to again stress that our aim is to provide a descriptive analysis, and not to establish causal relationships between segregation and institutional or macroeconomic reforms. However, it is with the larger questions in mind that we seek to identify the forces driving changes in occupational segregation.

The remainder of the chapter is structured as follow. Section 2 reviews existing studies that investigate both gender and race based occupational segregation, as well as studies that investigate informality in the Brazilian context. Section 3 presents the analysis of occupational structure over time using our newly constructed occupational classification, and presents data disaggregated by both gender and race and also across the formal, informal and self-employed sectors. Section 4 presents the analysis for gender and race based occupational segregation by applying several well-known indices. The analysis is undertaken over time across the formal, informal and self-employed sectors, while we also explore sectoral and spatial patterns of segregation. Section 5 presents an analytical decomposition of changes in both gender and racial occupational segregation over time. Finally, the last section concludes by proposing factors that may explain observed differences between the formal and non-formal sectors.

5.2 Literature Review

Because this chapter describes differences in occupational structure and segregation across the formal and non-formal labour markets, it draws on and bridges the economics literature dealing with both segregation and informality. This section consequently begins by reviewing important studies of occupational structure and segregation, while highlighting those that have specifically focused on Brazil. It then turns to the issue of informality in the Brazilian context and looks at two issues in turn: important debates about the definition of the concept of informality, and the findings of the most important empirical studies on informality in Brazil.

5.2.1 Studies of Occupational Segregation

Empirical studies of occupational segregation tend to analyse the phenomenon by employing tools drawn from the study of income inequality in order to construct new measures of segregation to assess the extent of segregation and its trend over time. Given that segregation is a core theme within labour economics, studies measuring the extent of occupational segregation in individual countries have been surprisingly uncommon, and focused primarily on gender rather than racial segregation. A careful review of the literature has uncovered 29 studies that focus exclusively on measuring country-specific occupational segregation, of which 23 focus on gender[31] and only six on racial segregation.[32] In addition to these individual country studies, a few adopt a cross-country perspective. Charles and Grusky (1995), Blackburn, Jarman and Siltanen (1993) and Deutsch and Silber (2005) provide cross-country studies focusing on subsets of OECD countries, while Melkas and Anker (1997) provide a comparison across the set of Nordic countries. Semyonov and Jones (1999) analyze data from 56 nations to study both occupational (horizontal) segregation and hierarchical inequality. Most notably, none of the empirical work examines segregation in developing countries, with

[31] For the U. S. see Albelda (1986), Blau and Hendricks (1979), Baunach (2002), Cotter, Hermsen, and Vanneman (2003), Hutchens (1991, 2004), King (1992) and Watts (1995); for United Kingdom see Hakim (1992, 1993) and Watts (1998); for Australia see Lewis (1982), Moir and Selby-Smith (1979) and Karmel and Maclachlan (1988); for Ireland see Reilly (1991); for Israel see Neuman (1994, 1998), for Switzerland see Deutsch, Flueckiger and Silber (1994) and Flueckiger and Silber (1999), for Brazil see de Oliveira (2001), for Spain see Mora & Ruiz-Castillo (2003), for Mexico see Calonico and Ñopo (2007), for Colombia see Isaza-Castro and Reilly (2011).

[32] Among previously cited works, Albelda (1986), King (1992) and Neuman (1994, 1998) also explore racial occupational segregation. In addition we find two studies only on racial segregation in the US, such as Boisso et al (1994) and Maume (1999).

the exception of Deutsch et al (2005), who explore gender segregation in Costa Rica, Ecuador and Uruguay, and Anker, Melkas and Korten (2003), who analyze cross-country variation in occupational segregation in a sample comprising both developed and developing countries.

Focusing on Brazil, theoretical and empirical research that examines wage discrimination is extensive, but work looking at occupational segregation has been limited. To the best of the author's knowledge, the only empirical study that measures gender based occupational segregation in Brazil is de Oliveira (2001). This study suggests that over two decades gender based occupational segregation has declined. The author reports a decrease in the Duncan index by three percentage points between 1987 and 1999 using a 3-digit occupational classification (moving from 63.29 to 60.06). This finding is driven primarily by increased female participation in the labour market, while female workers remain heavily concentrated in certain jobs (de Oliveira, 2001).

5.2.2 Understanding the Brazilian Informal Sector

One of the drawbacks of all of these existing studies is that they do not distinguish between the formal and non-formal sectors. In practice, paying specific attention to the issue of informality is important for a number of reasons. First, it highlights the degree of risk exposure, as individuals working in the informal sector do not benefit from social protection or regulation. Second, it is also important to understand whether informality flourishes since it may offer a more flexible alternative for economic activity. Finally, different trends between the formal and informal sectors may reveal something about the impact of labour market regulations in shaping patterns of segregation.

5.2.2.1 Defining Informality

Defining informality is a difficult task because of the complexity of the phenomenon. On the one hand, the informal sector offers employment to micro-entrepreneurs, families engaged in small businesses and to vulnerable and unskilled workers. On the other, it can also be viewed as a site for unregulated and illegal activities that evade taxation. The first conceptualization of informality is attributable to Hart (1971, 1972) who used the concept of informality to refer to small businesses primarily characterized by rudimentary technologies. Since then many other definitions of informality have been used by researchers, generally following either a "structuralist approach" or a "neo-liberal" approach (Carneiro, 1997). Focusing specifically on South American countries Gasparini and Tornarolli (2007) consider two alternative definitions of informality. The first definition focuses on the productive aspects of the activity and defines informal activities as small-scale, family-based and low-technology activities. The second focuses primarily on the legalistic and social protection aspects of informality. They conclude that the latter is possibly a more appropriate definition for informality in the South American context.

For Brazil, the first significant work that investigates the concept of informality is attributed to Cacciamali (1982). The author highlights two essential issues: first, the difference between informality and the shadow economy (that is, the distinction between small businesses and other illegal and tax-avoiding activities); second, how the formal and informal, registered and unregistered, economies may influence each other. Some of the early studies of Brazilian informality focused on wage workers without labour contracts, self-employed individuals, employers earning up to a certain portion of the minimum wage,[33] unpaid family workers, and domestic service workers (see Jatoba, 1987; and Gatica, 1989 cited in Carneiro, 1997). Other studies adopted a definition of informality based on the payment of social security contributions (see Cacciamali, 1988; Telles, 1992). In a more recent paper, Henley, Arabheibani and Carneiro (2009) compare three different definitions of informality centred on: i) contract status, based on the possession of a signed labour card; ii) social security status, based on contributions to a social security institution; and iii) formal sector activity, based on employment within a firm with more than five employees. They find that only 40% of cases are classified as informal across all three definitions of informality.

Despite these definitional challenges and uncertainties, most empirical studies have defined informal workers as those without signed work cards, the *carteira de trabalho* (Carneiro, 1997; Soares, 2004; Ulyssea, 2005). That said, this apparent working consensus on the definition of informal workers, based on possession of a signed labour card, can disguise the fact that there remain many difficulties in establishing a consistent categorization of informal sector workers.

As a starting point, some studies restrict the analysis to private sector employees and divide them into formal and informal workers (i.e. with and without signed labour cards). By contrast, others include the self-employed and workers in small firms in the

[33] In Jatoba (1987) self-employed individuals earning up to two times the minimum wage and employers earning up to five times the minimum wage are considered informal workers.

informal sector alongside unregistered workers. The inclusion of self-employed workers as part of the informal sector has become generally accepted, but a more recent literature also emphasises the potential benefits of considering informal employment and self-employment as two separate categories (Maloney, 2004, Fields, 1990, 2005 cited in Almeinda and Carneiro, 1997).

Aside from this disagreement over who should be included in the informal sector, there are also data challenges in operationalising any definition. This is particularly true with respect to employers, as it is very difficult to disentangle the formal status of small employers. The ILO considers employers with less than five employees to be in the informal sector, [34] and this definition is adopted by Bosch, Goni and Maloney (2007) in their study of the Brazilian informal sector. However, this threshold varies from country to country (see the discussion in Bosch, Goni and Maloney, 2007), since information on the number of employees might be missing and some small firm employers may be formal according to other metrics, such as the payment of social contributions.

5.2.2.2 Evidence on the Size and Characteristics of the Brazilian Informal Sector

Notwithstanding differences in definitions, there is consensus that the Brazilian informal market is sizeable. Carneiro (1997) reports that in 1990 about one-half of the economically active population was employed in informal activities. According to Urani (1996) approximately 49% of workers possessed a *carteira de trabalho* in 1995. Soares (2004) claims that in 1999 only 14 out of 36 million private sector workers were in the formal sector.

Building on these claims, Bosch, Goni and Maloney (2007) estimate that the informal private sector in urban areas increased by 10 % during the 1990s. However, Ramos and Ferreira (2005) argue that this increase in informality in Brazil depends heavily on restricting the analysis to metropolitan areas. They investigate the evolution of informality both in metropolitan areas alone and for the country as a whole and report that informality is primarily a metropolitan phenomenon, with a high prevalence in the manufacturing sector. Although they find that the rapid expansion of the service sector included a significant degree of informality, they conclude that changes in the manufacturing sector have been the primary cause of the overall increase in informality

[34] See the ILO definition in the official ILO website at http://stats.oecd.org/glossary/detail.asp?ID=1350.

and also explain the concentration of increased informality in urban areas. In terms of its spatial distribution, they find that informality has risen more acutely in the North-East and South-East regions rather than homogenously across the entire country. As such, limiting the analysis to certain metropolitan areas or to selected regions of the country may distort the findings, particularly by only capturing the elimination of formal occupations without considering the creation of new ones.

Bosch, Goni and Maloney (2007) explore several factors, including trade liberalization and rigidities arising from the Constitutional reforms,[35] which may help to explain the expansion of the informal sector in certain areas over the last two decades. They conclude that trade liberalization has had a small effect, while institutional reforms affecting the labour market have provided the main impetus.[36] By contrast, Paes de Barros and Corseuil (2001) focus on the impact of labour market regulations and find no evidence of any effect of the extent of informality.[37] Thus, they argue that changes in labour market outcomes can be attributed primarily to the macroeconomic developments in the economy. However, Goldberg and Pavcnik (2003) analyse the impact of trade liberalization, measured by trade exposure and tariff changes, and again find no evidence of any effects on the extent of informality. Alongside these somewhat inconclusive findings, it is also important to bear in mind other factors that may affect the size and expansion of the informal sector, including reforms in the public health sector and the large-scale migration of workers from rural areas to urban/metropolitan areas, primarily in the South-East of Brazil.

Although, these studies draw a clear distinction between the formal and informal sectors, it is important to also note the extent to which these two sectors are interconnected. In the work noted above, Bosch, Goni and Maloney (2007) confirm the

[35] The constitutional changes that affect the labour market analyzed by Bosch, Goni and Maloney (2007) are specifically union power, firing costs and overtime (the reduction of the legal limit of working hours per week).

[36] They estimate the impact using four specifications of labour market outcomes (creation and destruction rates of formal jobs, the size of formal sector and industry formality differentials) to trade variables and constitutional reforms variables using static and dynamic regression model. They find that trade variables explain less than 5% of the informality movements. The remainder is largely explained by constitutional variables.

[37] They focus only on separation rates and find no evidence of the impact of regulation on labour. To date, the most comprehensive work relating these changes to the functioning of the labour market was undertaken by Paes de Barros and Corseuil (2001) who find that separation rates decreased after the Constitutional changes for short employment spells and increased for longer spells, but find inconclusive results on the impacts on flows into informality from the formal sector. Bosch, Goni and Maloney (2007) use four different variables (job creation rate, destruction rate and size of formal sector and industry formality differentials) as a dependent variable to assess the impact of constitutional changes on the size of formal and informal sector and indeed the rate of job destruction in the formal sector is the least satisfactory.

conclusions of a model developed by Fiess, Fugazza and Maloney (2008) that suggests that the informal sector should not be considered inferior to the formal one and can be understood as an attractive alternative for more flexible and unregulated business opportunities. They find that both formal and informal labour markets are highly pro-cyclical and strictly interrelated: most transitions from the formal to the informal sector occur *within particular industries*, implying that the increase in informality is not widely attributable to structural changes in different economic sectors.[38] Additional studies by Maloney (1999, 2004) similarly oppose the traditional dualistic view that conceptualises the informal sector as a parallel labour market that acts like a “shock absorber” for the formal economy, thus exhibiting anti-cyclical behaviour. In fact, in many developing countries the informal sector displays pro-cyclical behaviour and formal and informal labour markets are found to be well integrated, growing during economic booms and contracting during periods of lethargic economic performance. This conception suggests a more positive view of informal markets, which may contribute to minimizing social instability during economic hardship and may offer employment opportunities outside the influence of government regulation (Carneiro, 1997).[39]

Apart from studying trends in the size of the informal sector over time, Ulyssea (2005) highlights several issues relating the nature and composition of the informal sector that are worth considering. First, although most studies claim that increased education increases the probability of participating in the formal sector, many studies report that the level of educational attainment of informal workers has increased noticeably over time. There is thus no consensus about whether returns to education are higher in the formal sector for all types of worker. Second, while women are identified as being overrepresented within the informal labour force in all studies, racial issues are ignored in almost all of the literature on Brazilian informality. This is an important motivation of the focus here on both gender and race based trends in occupational structure and segregation. Finally, a full understanding of the informal sector requires attention to high levels of turnover, and shorter job tenures, as these features are more typical of informal than formal occupations.

[38] The limited role of structural changes in justifying the expansion of the informal sector is also acknowledged in Ramos and Ferreira (2005).

[39] In his study, Carneiro (1997) argues that the growth of the Brazilian informal sector may reflect excessive intervention by the government.

There is only a very small body of literature that has looked at the connections between gender, race and informality, with the most notable being the work of Telles (1992). He argues that rapidly increasing female participation in the labour market was not being accompanied by a commensurate increase in opportunities with respect to both employment and remuneration. He further noted the tendency to ignore racial issues in the literature on Brazilian informality, reflecting an assumption that the hiring process is “colour blind”. Empirically, Telles (1992) highlighted the leading role played by female and non-white workers in the informal sector from the beginning of the 1980s. More significantly, and as noted in the previous chapter, he argued that women’s occupational opportunities in the formal sector were constrained by both education and race, making less educated women and non-white women more likely to participate in the informal sector. This, he argued, was reflected in the fact that female dominated occupations that required a low level of education were more likely to be informal when compared to male dominated occupations that required similar levels of education. These trends have been broadly similar over time. For example, according to Abramo (2004) in 2001 71.2% of white women and 76.2% of non-white women engaged in domestic service work did not possess a work card. This heavy representation of non-white women in informal domestic service work is thus an important indicator of the potential connections between race, gender and informality.

Finally, alongside this literature focused on the extent, nature and determinants of informality, lies research that has analysed possible interventions aimed at reducing informality. Holk (2002) investigates the effects of employment protection in Brazil, concluding that the decreasing rate of hiring in the formal market decreases employment in the formal sector, while increases in the separation rate in the informal market decreases employment opportunities in the informal market. Ulyssea (2010) claims that the focus should be on the level of employment protection rather than on the level of enforcement, in particular by decreasing the cost of entry into the formal sector. On the other side of the argument, Almeida and Carneiro (2007) stress the importance of enforcing labour market legislation because an increase in enforcement actions tends to decrease the demand for formal labour, increase its supply and consequentially decrease the supply of informal labour.

Having reviewed the most relevant studies of occupational segregation and informality, several important gaps in the literature that this study seeks to address can be summarized. First, most studies of occupational segregation focus on developed economies, particularly the US and UK. Second, most of these studies focus on gender segregation with fewer studies focusing on racial segregation. Third, studies of occupational segregation have not systematically considered differences across the formal, informal and self-employed sectors. Looking specifically at Brazil, to the best of the author's knowledge, there is only one study that investigates gender based occupational segregation, though not across the formal, informal and self-employed sectors, while there are no empirical studies investigating race based occupational segregation. This research tries to fill this vacuum by providing an investigation of both gender and racial occupational segregation across the formal, informal and self-employed labour markets over a more extended time period than has previously been possible in studies of the Brazilian labour market.

5.3 Gender and Racial Occupational Segregation

In the previous chapter we provided a detailed overview of the occupational structure in Brazil over time, focusing on differences by gender and race across the formal, informal and self-employed sectors. We now turn to a more formal analysis of levels of occupational segregation, and of trends in occupational segregation. As was explained at the outset, we define occupational segregation in terms of differences in the distribution of women and men, or non-whites and whites, across occupations (James and Taeuber, 1985; Anker, 1998).

In order to measure occupational segregation, we rely on several well-known indices of segregation. In the next section we introduce and describe these alternative measures and discuss their strengths and weaknesses. We then apply these measures to the Brazilian case in order to assess the magnitude and the evolution over time of both gender and racial occupational segregation. As with the previous chapter, we analyse both the aggregate labour market and individual trends in the formal, informal and self-employed sectors. Finally, after applying a series of robustness checks to these preliminary findings, we explore how these measures of occupational segregation differ by characteristics of the labour force, including age, educational attainment, geographic location and economic sector. The analysis here provides new insights into labour market trends in Brazil, and also sets the stage for the subsequent section, which investigates the possible determinants of changes in occupational segregation over time.

5.3.1 Measures of segregation

The analysis of occupational segregation is undertaken using five different indices of segregation: the Duncan index (I_D), the Moir and Selby-Smith index (I_{MSS}), the Karmel and Maclachlan index (I_{KM}), the Gini segregation index (I_G) and the Marginal Matching index (I_{MM}). Our motivation in applying this wide range of measures is twofold. First, to understand the extent to which results may be dependent on the particular measures employed, and, second, to gain some insight into which measure is optimally suited to our purpose.

The dissimilarity index, or Duncan index (Duncan and Duncan, 1955), is one of the most widely used measures of segregation and is given by the formula:

$$I_D = \frac{1}{2}\sum_{i=1}^{n}\left|\frac{F_i}{F} - \frac{M_i}{M}\right| \qquad \text{with } i=1,2,\ldots,n \tag{1}$$

where F_i and M_i are the number of female and male workers in the i^{th} occupation and F and M are the total number of women and men in the labour force.[40] The use of the 'modulus' indicates that this index is computed as one half of the sum of the absolute differences between the ratio of women in occupation i to the total female labour force and the ratio of men in occupation i to the total male labour force. The index is generally interpreted as measuring the proportion of the female workforce that would be required to shift between occupations in order to equalize female and male representation across occupations. The main weakness is that redistributing the female workforce in order to achieve zero segregation would inevitably result in a change in the occupational structure. Furthermore, this index assigns equal weights to each occupation independent of its relative size (i.e., its share in the total workforce). Watts (1998) claims that the Duncan index fails to show occupation invariance, but it is invariant to the gender composition of the labour force.

There are a large number of studies that have applied the Duncan index in order to measure occupational segregation. These include, among others, Butler (1987) and Hutchens (1991) and King (1992) for the U.S., Watts (1998) for the U.K., Reilly (1991)

[40] The formulas reported in this section refer to gender segregation. In order to compute the indices for racial segregation F and F_i have to be re-defined for non-white workers and M and M_i for white ones.

for Ireland, Deutsch, Flueckiger and Silber (1994) for Switzerland and Neuman (1998) for Israel. Similarly, cross-sectional analysis using the Duncan index has been carried out by Charles and Grusky (1995) and Deutsch et al (2005), among others. However, despite the popularity of the Duncan Index, several other measures of occupational segregation have been proposed in the literature in an effort to address criticisms faced by the Duncan index.[41]

Modification of the dissimilarity index have been developed by Moir and Selby-Smith (1979) and Karmel and Maclachlan (1988).[42] The Moir and Selby-Smith index is given by the following formula:

$$I_{MSS} = \frac{1}{2}\sum_{i=1}^{n}\left|\frac{T_i}{T} - \frac{F_i}{F}\right| \tag{2}$$

where *T=F+M*, and thus captures the total workforce, comprising both female and male workers.

The Karmel and Maclachlan index denotes the total labour force that would need to be relocated, with replacement, in order to reach zero segregation while retaining the initial occupational structure and overall female and male shares of the workforce.

$$I_{KM} = \frac{1}{2}\sum_{i=1}^{n}\left|a\frac{M_i}{T} - (1-a)\frac{F_i}{T}\right| \tag{3}$$

where $a = \frac{F}{T}$ is the share of females in the total labour force.

The Gini segregation index is defined by the following formula:

$$I_G = \frac{1}{2}\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{M_i M_j}{M\ M}\left|\frac{F_i}{M_i} - \frac{F_j}{M_j}\right|}{\frac{F}{M}} \tag{4}$$

As Silber (1989) notes, the G-segregation index is equal to the Gini Index of the female-male ratio where the weights are the shares of each occupation in the total male

---

41 Additional alternative measures of occupational segregation that are not explored in this study are Theil, 1967; Lewis, 1982; Kakwani, 1994; Charles and Grusky, 1995.

42 Applications of the Moir and Selby-Smith (1979) index are Moir and Selby-Smith (1979) itself and Kakwani (1994) on Australian labour market. While for Karmel and Maclachlan (1988) index we can cite Karmel and Maclachlan (1988) itself and Kakwani on Australia, Watts (1998) on the U.K. and Deutsch et al (2005) on Costa Rica, Ecuador and Uruguay.

workforce.[43] Assuming a segregation curve which can be defined as a cumulative distribution of the proportion of women in every occupation, the index is given by twice the area lying between the segregation curve and the equi-distribution line given by the 45 degrees diagonal. This is similar in spirit to the Lorenz curve in the inequality literature.

Finally, an interesting way to control for marginal changes is provided by the marginal matching procedure developed by Blackburn, Jarman and Siltanen (1993). This procedure first classifies occupations as either female or male occupations based on their gender ratio $\frac{F_i}{M_i}$. In practice, the total female workforce should be equal to the number of workers employed in female occupations, $F = T_F$ and the total male workforce should be equal to the number of workers employed in male occupations, $M = T_M$. Hence, the Marginal Matching measure is given by the following formula:

$$I_{MM} = \frac{[F_F M_M - F_M M_F]}{FM} = \frac{F_F}{M_F + F_F} - \frac{F_M}{M_M + F_M} = \frac{F_F}{T_F} - \frac{F_M}{T_M} \quad (5)$$

This measure captures the extent to which occupations divided into female and male occupations vary together.[44]

5.3.2 Results: Measures of Occupational Segregation over Time

We now turn to the results. Each index is computed independently for both gender and racial segregation, as well as separately for the formal, informal and self-employed sectors. As the objective of this exercise is to compare these measures of segregation over time and across different sectors, we have paid particular attention to ensuring the statistical validity of such comparisons. Specifically, the statistical significance of changes in occupational segregation are assessed using non-parametric tests of statistical significance, through bootstrapping the standard errors based on 500 replications, following the approach suggested by Efron (1979) and subsequently popularized by Efron and Tibshirani (1991, 1993).[45] The bootstrapping method

---

[43] Applications of the Gini segregation index are Butler (1987) for the U.S., Silber (1989) for France and Neuman (1998) for Israel.

[44] The Marginal Matching index has been applied by Blackburn, Jarman and Siltanen (1993) and Blackburn and Jarman (2005).

[45] Further applications of segregation indices were developed by Deutsch, Fluckinger and Silber (1994) and Flückiger and Silber (1999). Alternatively, parametric estimation of the standard errors is also feasible. For the Kakwani (1994) class of β-segregation indices, a framework to test the racial

estimates the distribution of the segregation measure by resampling with replacement in order to create multiple estimates of the statistics. These distributions are then used to construct confidence intervals around the original point estimates and ultimately to establish standard errors (Boisso et al, 1994).

Table 1 provides the results of computing the different measures of occupational segregation, along with their bootstrapped standard errors.[46] All indices of segregation have been computed using our harmonized 3-digit occupational classification.[47] In order to assess the importance of changes in the segregation measures across sectors and over time, we calculate the statistical significance of the changes within the formal, informal and self-employed sectors in any given year, as well as over time, using five reference years (viz., 1987, 1992, 1997, 2002 and 2006). Tables A1 and A2 in the appendix report the tests of mean differences among sectors and over time using the standard parametric two sample t-tests. The use of t-tests here is justified on the basis of the 'law of large numbers'. However, non-parametric tests yielded identical findings. We discuss the statistical significance of these differences, and provide an explanation of our findings, in the discussion to follow.

Across all indices of segregation, we notice three main chataceristics. First, gender occupational segregation is always considerably greater than racial occupational segregation - roughly three times greater. Second, gender segregation is generally more severe in the informal and self-employed sectors than in the formal sector. Third, overall levels of segregation are declining, though this decline has been much more pronounced for gender, while patterns vary in important ways between the formal and informal sectors.

---

segregation in occupations that has the advantage of being based on the Kakwani index is proposed. This statistic does not allow us to test statistical significance over time, however, which is of some importance in the current context.

[46] Outcomes for these segregation indices over all years of the dataset and across all sub-groups of the population are available on request from the author.

[47] Figures on occupational segregation using occupational codes at the 2-digit level are also available. The patterns are very similar, but the extent of segregation is, on average, smaller. The more detailed the occupational categorization the greater is the outcome from any measures of segregation.

Table 1: Indices of segregation

	All labour market					Formal sector					Informal sector					Self-employed sector				
	1987	**1992**	**1997**	**2002**	**2006**	**1987**	**1992**	**1997**	**2002**	**2006**	**1987**	**1992**	**1997**	**2002**	**2006**	**1987**	**1992**	**1997**	**2002**	**2006**
Duncan index																				
gender	0.605	0.612	0.601	0.571	0.565	0.556	0.577	0.572	0.532	0.513	0.692	0.729	0.710	0.648	0.653	0.647	0.620	0.605	0.624	0.617
s.e.	0.002	0.002	0.002	0.002	0.002	0.004	0.004	0.003	0.003	0.003	0.004	0.004	0.004	0.003	0.004	0.004	0.004	0.004	0.004	0.004
race	0.199	0.200	0.199	0.195	0.191	0.193	0.192	0.195	0.175	0.177	0.197	0.153	0.163	0.183	0.191	0.134	0.153	0.160	0.153	0.149
s.e.	0.003	0.003	0.003	0.003	0.002	0.004	0.004	0.004	0.004	0.004	0.005	0.005	0.005	0.005	0.005	0.006	0.006	0.006	0.005	0.005
Moir & Selby-Smith index																				
gender	0.387	0.379	0.366	0.338	0.329	0.366	0.353	0.336	0.305	0.294	0.400	0.415	0.395	0.353	0.341	0.438	0.428	0.426	0.425	0.412
s.e.	0.002	0.002	0.002	0.001	0.001	0.003	0.003	0.002	0.002	0.002	0.003	0.003	0.003	0.003	0.002	0.003	0.004	0.003	0.003	0.003
race	0.106	0.106	0.104	0.098	0.089	0.115	0.114	0.114	0.099	0.092	0.090	0.065	0.070	0.077	0.075	0.068	0.079	0.083	0.074	0.067
s.e.	0.002	0.002	0.001	0.001	0.001	0.003	0.003	0.002	0.002	0.002	0.003	0.002	0.002	0.002	0.002	0.003	0.003	0.003	0.002	0.002
Karmel & Maclachlan index																				
gender	0.277	0.286	0.285	0.275	0.275	0.229	0.268	0.270	0.258	0.249	0.328	0.345	0.341	0.319	0.324	0.270	0.255	0.244	0.267	0.272
s.e.	0.002	0.002	0.002	0.001	0.001	0.002	0.005	0.002	0.002	0.002	0.004	0.004	0.004	0.003	0.002	0.005	0.007	0.004	0.003	0.003
race	0.099	0.099	0.099	0.098	0.095	0.093	0.092	0.095	0.086	0.088	0.097	0.074	0.080	0.089	0.091	0.066	0.076	0.079	0.076	0.073
s.e.	0.002	0.001	0.001	0.001	0.001	0.002	0.002	0.002	0.002	0.002	0.003	0.003	0.002	0.002	0.002	0.003	0.003	0.003	0.003	0.002
Gini segregation index																				
gender	0.776	0.785	0.775	0.743	0.735	0.713	0.738	0.725	0.701	0.678	0.819	0.845	0.838	0.810	0.810	0.823	0.794	0.764	0.756	0.759
s.e.	0.003	0.002	0.004	0.002	0.002	0.005	0.004	0.003	0.003	0.003	0.006	0.005	0.005	0.003	0.003	0.004	0.007	0.004	0.004	0.003
race	0.265	0.267	0.261	0.263	0.262	0.259	0.262	0.259	0.247	0.244	0.249	0.203	0.215	0.239	0.256	0.196	0.224	0.218	0.221	0.218
s.e.	0.003	0.003	0.003	0.003	0.003	0.005	0.005	0.005	0.004	0.004	0.006	0.007	0.006	0.006	0.006	0.007	0.006	0.006	0.006	0.006
Marginal matching index																				
gender	0.580	0.586	0.556	0.544	0.515	0.344	0.519	0.484	0.430	0.455	0.678	0.704	0.700	0.615	0.589	0.626	0.569	0.513	0.325	0.489
s.e.	0.024	0.011	0.011	0.008	0.018	0.033	0.022	0.023	0.019	0.014	0.010	0.010	0.010	0.012	0.012	0.032	0.012	0.014	0.020	0.020
race	0.137	0.167	0.134	0.180	0.177	0.172	0.152	0.176	0.156	0.157	-0.207	-0.367	-0.411	-0.390	-0.304	-0.246	-0.049	-0.086	0.116	0.078
s.e.	0.046	0.020	0.031	0.009	0.010	0.011	0.025	0.013	0.013	0.016	0.054	0.043	0.048	0.177	0.343	0.059	0.017	0.105	0.051	0.034

Source: Author’s computations using PNAD 1987 – 1992 – 1997 – 2002 – 2006.
Note: Standard errors boostrapped with 500 replications.

In 2006 the Duncan index between female and male workers was 0.565, which is much greater than the Duncan index of 0.191 for race. This means that in 2006 more than half of female workers and one fifth of non-white workers would have needed to be reallocated in order to equalize representation across occupations. This difference between gender and racial segregation is present across all five segregation measures.

Turning to differences between the formal, informal and self-employed sectors, the Duncan index by gender in 2006 is equal to 0.513 if we restrict the analysis to the formal sector, while it is 0.653 for the informal sector. This implies that the formal sector records 9.2% less gender segregation than the labour market as a whole, while the informal sector has gender segregation 15.5% higher than the average value for the entire labour market. Differences in levels of segregation across sectors are generally statistically significant at the 5% level.

The pattern is somewhat different when examining racial segregation, as in both 1987 and 2002 there was no statistically significant difference in racial segregation between formal and informal sectors. During the 1990s racial segregation in the formal sector was, in fact, somewhat higher than in the informal sector, but this trend was reversed by the beginning of the 2000s. Finally, during the 1990s levels of racial segregation in the informal and self-employed sectors were not statistically different from each other. These patterns of statistical significance hold across the Duncan, Karmel and Machlachlan and Gini segregation indices, as shown in table A1 in the appendix.

Although gender segregation is more severe, over time the situation is improving more rapidly for women than for non-white workers. Using the Duncan index, gender segregation decreased by 6.5% between 1987 and 2006, while racial segregation declined by 4.2%. Focusing on gender segregation, we report an initial increase in segregation at the beginning of the 1990s, but this increase is negligible and not always statistically significant across the different segregation measures (see table A2). More importantly, we note decreases in all of the segregation measures between the beginning of the 1990s and 2006, and these changes are always statistically significant. In the case of racial segregation, the trend is somewhat different and is particularly sensitive to issues of informality, as racial segregation has decreased primarily in the formal sector.

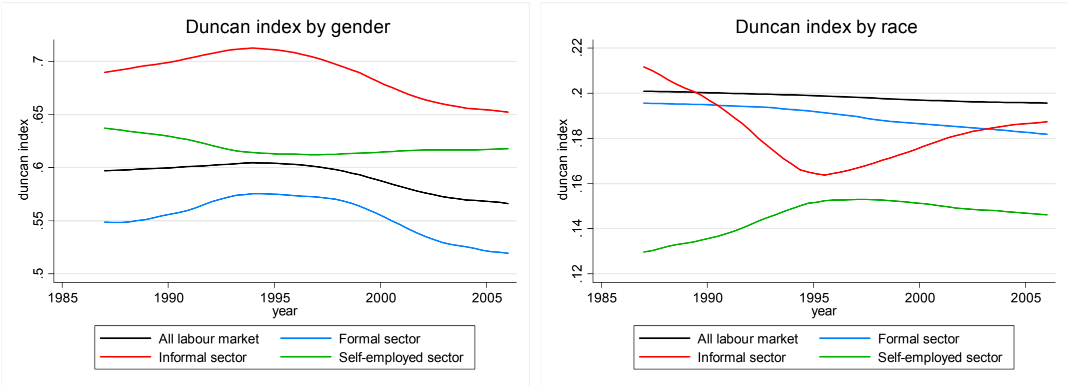
Looking at the Duncan indices, if we consider only the formal sector racial segregation has decreased slightly faster (8.2%) than gender segregation (7.7%). Conversely, in the informal sector gender segregation has been decreasing faster (5.5%)

than racial segregation (3.1%). Meanwhile, we see a quite clear pattern in which both gender and racial segregation have declined more rapidly in the sector than in the non-formal sectors.[48] In fact, in the informal sector we see a somewhat surprising increase in racial segregation since the mid-1990s, and this unexpected trend is confirmed in using all of the segregation indices. This increase since the mid-1990s is statistically significant across all of the segregation measures and results in an overall decline in racial segregation in the informal sector between 1987 and 2006 that is not statistically different from zero (see table A3). For example, between 1992 and 2006 gender segregation decreases by 10.4% in the informal sector using the Duncan index (or by 4.1% using the Gini), while racial segregation increases by 24.8% using the Duncan during the same period (and 26.1% using the Gini). While table 1 reports these measures of occupational segregation every five years, these trends can be seen in full in figure 1, which plots the evolution of these measures over time.

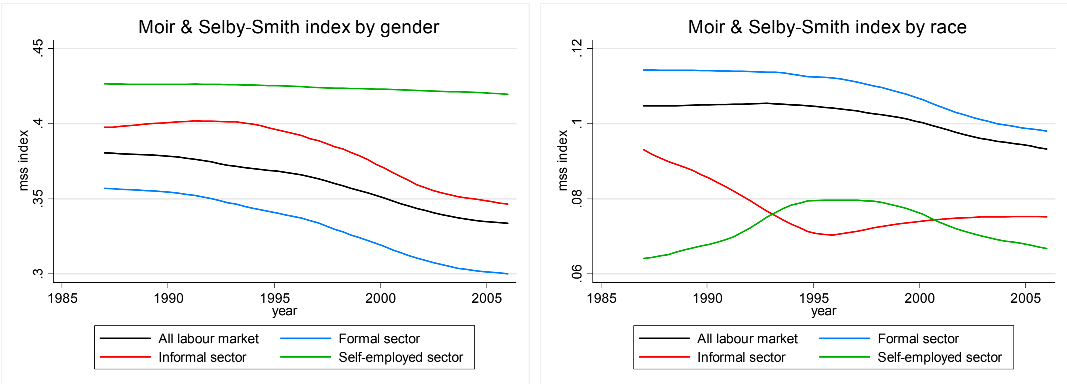
[48] One might expect a more rapid decline in segregation in the formal sector, particularly in the case of gender segregation, to be driven by a rapid expansion of female employment in the public sector, but this is not the case. We re-estimate the Duncan index by disaggregating the formal sector into public and private components and we find comparable results across both groups. For the public sector we find that the Duncan index was equal to 0.51 in 1992 and to 0.45 in 2006 while for the private sector it was equal to 0.54 in 1992 and to 0.48 in 2006. While we find somewhat higher gender segregation in the private sector, as expected, we find a similar level of decline (0.06) across both sectors. We produced these computations beginning in 1992 owing to the unavailability of the public sector variable for earlier years of the dataset.

Figure 1: Indices of segregation by gender and race over time

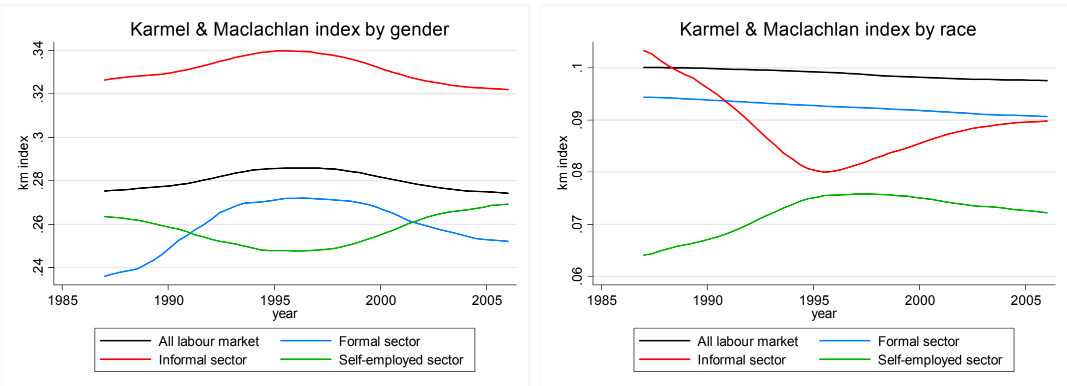
Panel A – Duncan and Duncan index



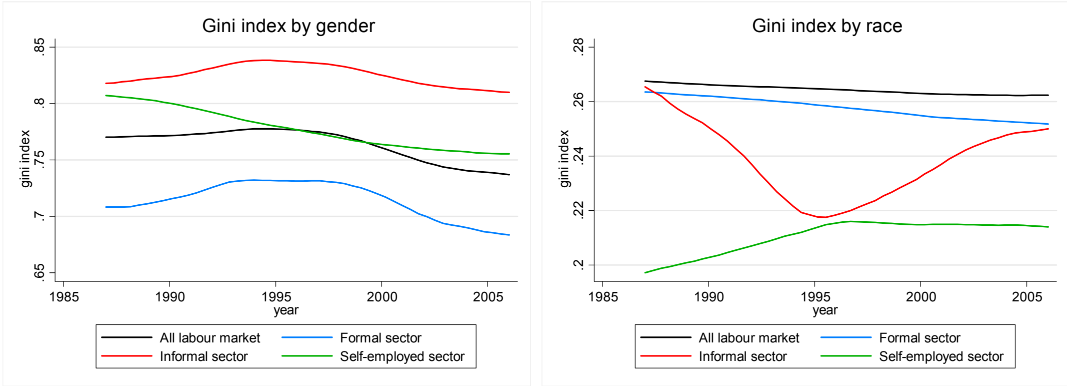
Panel B – Moir and Selby-Smith index



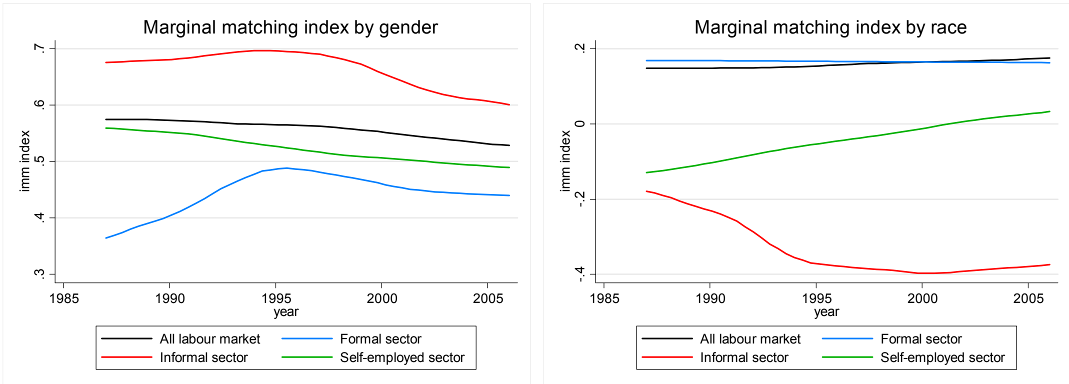
Panel C – Karmel and Maclachlan index



Panel D – Gini segregation index



Panel E – Marginal matching index



Source: Author's computations using PNAD 1987 - 2006.
Note: 1991, 1994 and 2001 missing years.

Having summarized broad trends that cut across the different segregation indices it is also useful to comment on differences in levels and trends between the various segregation indices.[49] Most noticeably, the Gini segregation index generates the highest figures among all of the indices with values in 2006 of 0.735 and 0.262 for gender and race, respectively. Instead of looking at mean deviations, as is the case for Duncan-type indices, the Gini index uses mean differences to measure the dispersion of the occupational distribution. Thus, segregation appears to be a more severe problem when focusing on compositional differences among all occupations together, which is the case using the Gini Index, than when focusing on how gender and racial ratios differ from the overall composition of the workforce within each occupation individually, as in the Duncan index case.

[49] In general, the Duncan index measure cannot be directly compared to the Moir and Selby-Smith and the Karmel and Maclachlan indices, as the indices have different upper bounds. The Duncan index has an upper bound equal to 1 (as is the case for the Gini index and the Marginal Matching index), while the upper bound of the Moir and Selby-Smith index is equal to M/T and the upper bound for the Karmel and Maclachlan is equal to 2(F/T)M/T.

Using the Moir and Selby-Smith index the outcomes are very similar to the Duncan and Duncan index, but with smaller magnitudes and somewhat larger changes over time. The explanation for these modest differences lies in subtle differences between the two indices. When measuring gender segregation using the Duncan and Duncan index the actual occupational distribution is compared to an ideal distribution in which there is an identical distribution of female and male workers. In slight contrast, the Moir and Selby-Smith index compares the actual distribution to an ideal distribution in which the distribution of female workers between occupations is equal to the female share of the labour force (Moir and Selby-Smith, 1979). As a consequence, segregation declines faster when using the Moir and Selby-Smith index owing to the actual increase in the number of female workers in the labour force.

When using the Karmel and Maclachlan index the estimated level of segregation is dramatically lower in aggregate terms, though the patterns of change are broadly consistent with the other indices. Again, an explanation can be found in how these indices are constructed. The Duncan index calculates the number of female workers that would need to be moved without replacement, and thus allows for changes in the occupational distribution. The Karmel and Maclachlan index measures the number of female workers that would need to be shifted with replacement in order to yield zero segregation, and thus retains the relative size of each occupation and the overall size of the labour force (Watts, 1998). In table 1, we can see that when the Duncan index is decreasing, the Karmel and Maclachlan index tends to remain constant and in some cases even increases slightly. This implies that although the female-male differential has narrowed, the increasing number of women entering the labour market has meant that the proportion of workers that would need to shift occupations in order to eliminate segregation has not changed, or has increased slightly (Karmel and Maclachlan, 1988).

The final index is the Marginal Matching index, developed by Siltanen, Jarman and Blackburn (1995), which is designed to to overcome the sensitivity of the other indices to changes in the female proportion of the labour force and in worker shares in the female-dominated occupations. The Marginal Matching index classifies occupations as female or male dominated at the outset in order to ensure gender composition and occupational invariance. The drawback of this approach is that the index suffers from high variability. Holding factors constant, our calculations reveal a substantial decrease in gender segregation, defined as the difference between the ratio of female workers working in “female” occupations to total employment in “female” occupations employment and the ratio of female workers working in “male” occupations to total employment in “male” occupations. The most divergent result, when compared to other indices, is the negative values for racial segregation in the informal and self-employed sectors.[50] This implies that the ratio of non-white workers working in white occupations to total white occupation workers is greater than the ratio of non-white workers working in non-white occupations to total non-white occupation workers. The main reason is the very high proportion of non-white male workers employed in the self-employed sector.

Having considered these five indices, the remainder of the analysis is focused on the results using the Duncan Index. The Duncan index is the most intuitive measure to interpret and also the most commonly used in the literature, which facilitates comparison with earlier research both in Brazil and elsewhere. There is nonetheless significant value in having considered all five indices. Perhaps most importantly, our confidence in the results is significantly reinforced by the fact that, despite significant differences in methodology, the results using these alternative measures are broadly comparable, particularly in capturing trends over time. As such, there does not appear to be any major risk in focusing on the Duncan Index. In addition to the confidence that emerges from the broad similarities in the results, the alternative methods have also served to provide an additional nuance to the overall narrative.

5.3.3 Robustness checks

Having presented these results, we now turn to two important robustness checks before proceeding further. The goal is to better understand the impact of our new classification on estimates of occupational segregation, and to ensure that choices made in the construction of the new classification have not significantly biased the results presented here.

---

[50] The negative sign when employing the Marginal Matching indices to measure racial segregation for the informal and self-employed sectors is related to the fact that the majority of the workers in these non-formal labour markets are non-whites. This highlights the importance of carefully interpreting the results of these segregation measures, particularly when women or non-whites are the majority of the labour force. Most notably, when looking at gender segregation men are always the majority of the labour force across formal and non-formal labour markets, in the case of racial segregation, the share of non-whites is smaller than whites in the formal sector both in the informal and self-employed sectors.

5.3.3.1 Comparing our Findings to the Duncan Index Computed Using the Original Occupational Codes

The findings on occupational segregation reported in the previous section are based on the new classification of occupational codes that we have constructed. As explained earlier, we constructed a harmonized classification of occupational codes at the 3-digit level in order to overcome changes in the classification of occupational codes for the PNAD datasets over time, and particularly after 2001. It is thus important to ensure that our findings represent actual changes in the distribution of workers across occupations, and are not an artefact of the methodology employed in constructing our new harmonized occupational classification.

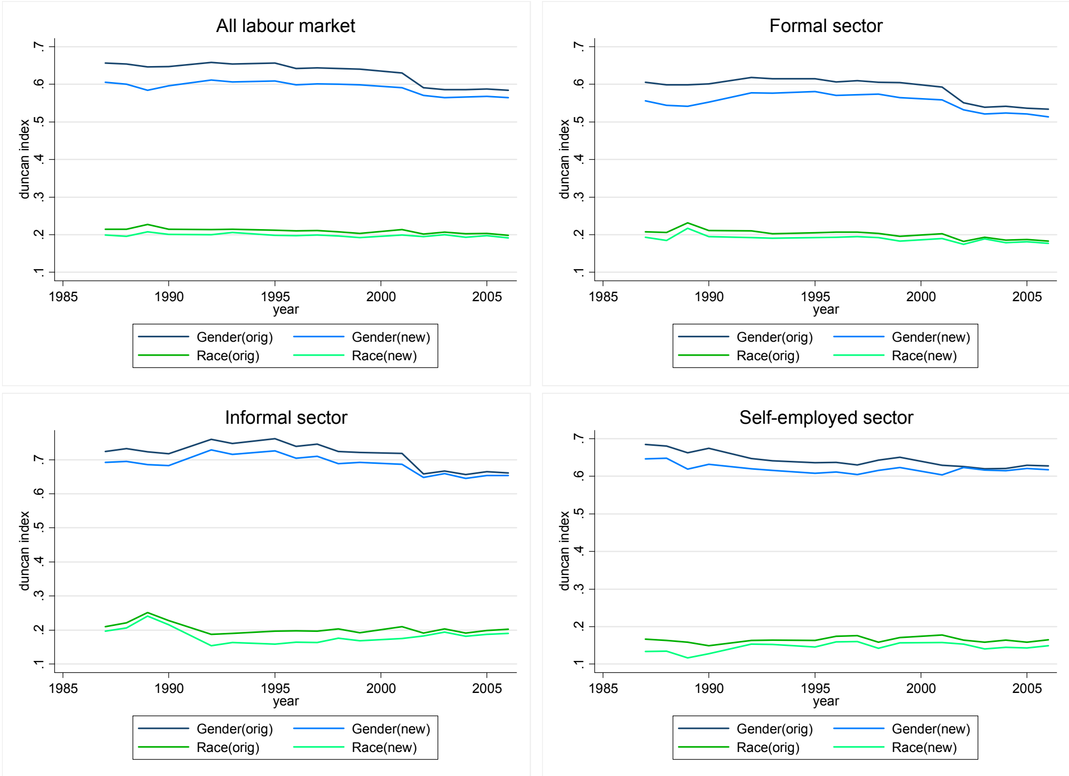
As a partial check against this possibility, we compare our results to results computed using the original classification, focusing exclusively on the Duncan index for the purposes of illustration.[51] There are obvious problems in computing the Duncan Index using the original classification, given that the number of occupation codes varies over time, leading to inconsistencies in the data. Despite these imperfections, we consider the original classification the most appropriate basis for comparison. We thus adopt the original 3-digit classification until 2001 and then, beginning in 2002, we re-aggregate the original 4-digit classification to the 3-digit level in order to make the analysis as compatible as possible over time without imposing any other adjustments on the data.

Figure 2 reports the evolution of the Duncan index by both gender and race in four panels, capturing the entire labour market and the formal, informal and self-employed sectors separately. The first notable observation is that the calculations using the two alternative occupational classifications follow the same trend over time, thus reinforcing our confidence in the reliability of the results. However, a discrepancy between the Duncan indices computed using the original and new classifications is visible in all cases, with occupational segregation greater when based on the original (non-compatible) classification. In addition, there is a considerable downward jump between the years 2001 and 2002 in the first three panels, and this jump is particularly pronounced when employing the original occupational classification.

---

[51] While we present results only for the Duncan Index here for the sake of brevity, the conformity between the results using the original and new classifications is consistent across all of the other methods as well. The volatility observed in employing the Marginal Matching index similarly persists when employing the original data.

Figure 2: Robustness check - Comparing occupational segregation computed using the original and the new classification of the occupational codes



Source: Author's computations using PNAD 1987 - 2006.
Note: 1991, 1994 and 2001 missing years. This robustness check is performed using the Duncan index.

Both features can be readily explained. The Duncan index is sensitive to the number of occupational categories used in its computation. The more categories there are, the greater the computed level of segregation. The original 3-digit level classification for the PNAD datasets between 1987 and 2001 included between 350 and 430 codes. Beginning in 2002, there are 459 codes at the 4-digit level and 175 codes at the 3-digit level. Given that our new re-classification includes roughly 80 codes, the gap between the segregation computed using the original and new classifications is likely to be driven largely or entirely by the difference in the number of codes. The jump between 2001 and 2002 follows the same logic, as the passage from 354 codes down to 175 codes would be expected to lead to a decline in the index value.

Overall, the outcome of this comparison of the two occupational classifications is reassuring. The smaller number of codes in our re-classification is immediately more comparable with any analysis employing the official ILO ISCO-08 classification, which includes roughly 200 occupational categories at the 3-digit level. Although it is impossible for us to cover all two hundred occupations contained in the international classification, our classification is nonetheless far more compatible than the original.

Perhaps more importantly, our re-classification succeeds in smoothing the jump that we observe in the data between 2001 and 2002 when using this original classification. The motivation for the new classification was precisely to create a more consistent data series before and after 2002, and the comparative results presented here suggest that this objective has been achieved.

5.3.3.2 The impact of excluding 'no wage' observations

As noted earlier, it has been necessary to exclude observations for which the wages variable is missing, but this risks underestimating the magnitude of the non-formal labour markets and altering the reported estimates of occupational segregation. It is thus important to make an effort to understand the likely implications of this exclusion for the results presented here.

In order to understand the likely impact we first need to disaggregate 'no wage observations' into two categories: workers that are not remunerated and respondents that randomly failed to report their wages (missing wages in the strict sense). We begin with the latter group, who comprise, on average, only 1.4% of the entire sample (see panel A of table 2), and we check whether missing wage observations are randomly distributed across occupations and across the formal, informal and self-employed sectors. We confirm that the observed profile of those with missing wage observations is broadly similar to the entire sample, which, coupled with the limited number of such observations, suggests that there is little risk of any undue influence on the results.

The issue is more complicated in the case of workers who are not remunerated, as they represent a considerable share of the sample, at 9.5% on average (shown in panel A of table 2 and highlighted graphically in the left hand graph in panel A of figure 3). More importantly, 'not remunerated' workers are non-random, and generally report employment in own-production, own-construction or as a member of the household, primarily in the agricultural sector. Furthermore, 'not remunerated' workers are overwhelmingly women and primarily non-white, as evidenced by the increase in the female and non-white shares of the labour force if they are included in the sample (see the right hand graph in panel A of figure 3).

Table 2: Accounting for 'no wage' observations

Panel A – Sample size including 'no wage' obs.

	1987	1992	1997	2002	2006
Formal sector					
# of obs.	45218	48884	52774	61194	72098
Informal sector					
# of obs.	28983	28286	32486	40940	43781
Self-employed					
# of obs.	25484	27642	31578	35974	38822
Total (1)					
# of obs.	99685	104812	116838	138108	154701
Of which random missing wage obs.					
# of obs.	703	2466	2023	2206	2358
%	0.67%	2.06%	1.55%	1.44%	1.38%
PLUS 'not remunerated' workers (2)					
# of obs.	5801	14694	14070	14602	15655
%	5.50%	12.30%	10.75%	9.56%	9.19%
Total as (1)+(2)					
# of obs.	105486	119506	130908	152710	170356

Panel B – Comparison of non-formal sectors including/excluding 'no wage' obs.

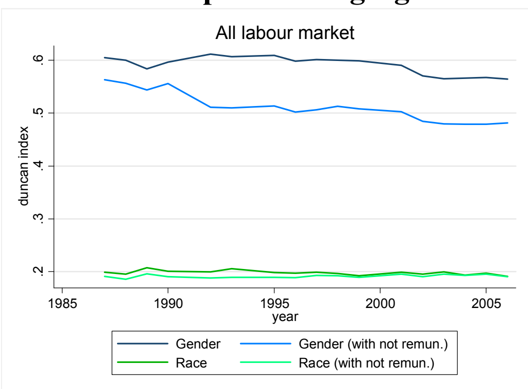
	1987	1992	1997	2002	2006
Shares of non-formal sectors (excluding 'no wage' obs.)					
%	54.64%	53.36%	54.83%	55.69%	53.40%
Share of non-formal sectors (including 'no wage' obs.)					
%	57.13%	59.10%	59.69%	59.93%	57.68%

Source: Author's computations using PNAD 1987 – 1992 – 1997 – 2002 – 2006.
Note: In general the 'no wage' observations are the sum of workers who are not remunerated and cases in which the wages variable is missing. The number of 'no wage' observations reported here is smaller than the total number of 'no wage' observations in the dataset, reported in the previous chapter, as the figures here exclude employers and military forces.

In their earlier study of Brazilian informality, Ramos and Ferreira (2005) find that the exclusion of 'not remunerated' workers from the analysis leads both to an underestimate in the size of the informal sector and to a change in the observed trend over time. Specifically, they find that the level of informality is relatively constant when 'not remunerated' workers are excluded, but increasing over time when included. Our findings are very similar, as we find that if we add 'not remunerated' workers to the sample the overall estimate of the size of the informal sector increases, while the trend shifts from showing a slight decrease in informality to showing a constant level of informality over time. A comparison of the shares of the non-formal sectors when 'no wage' observations are excluded/included is presented in panel B of table 2. When they are excluded we find a slight decrease among non-formal sectors of 2.2%, while if they are included we find that the share of the non-formal sectors is constant over time, and larger than in the baseline case, at an average of 58.7% of the entire labour market instead of 54.4%.

Figure 3: Robustness check – Controlling for the exclusion of 'no wage' observations
Panel A- Size of the sample

Panel B – Occupational segregation



Source: Author's computations using PNAD 1987 - 2006.
Note: 1991, 1994 and 2001 missing years. This robustness check is performed using the Duncan index.

Turning finally to the impact of excluding 'not remunerated' workers on our measures of segregation, we compute the Duncan index using a sample that includes 'not remunerated' workers. The comparison is reported in panel B of figure 3. We find that the inclusion of 'not remunerated' work in the analysis results in a decrease in our measures of segregation by both gender and race, though the overall trends are largely unchanged. This is what we would expect given that the majority of 'not remunerated' workers are actually female and non-white.[52]

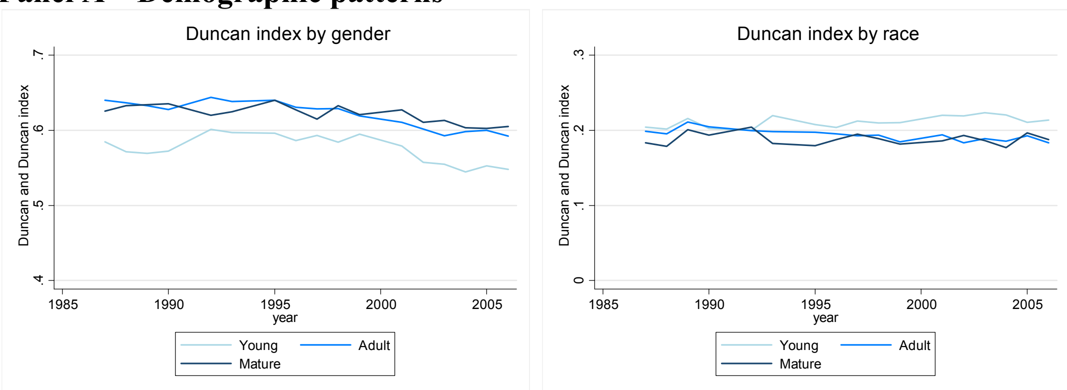
[52] In 1987 gender segregation was equal to 0.60 when excluding 'not remunerated' workers and declines to 0.56 when they are included - a 7% decrease. In the same year racial segregation was equal to 0.199 and decreases to 0.191 with the inclusion of 'not remunerated' workers - a 4% decrease.

5.3.4 Exploring demographic, educational, sectoral and spatial patterns

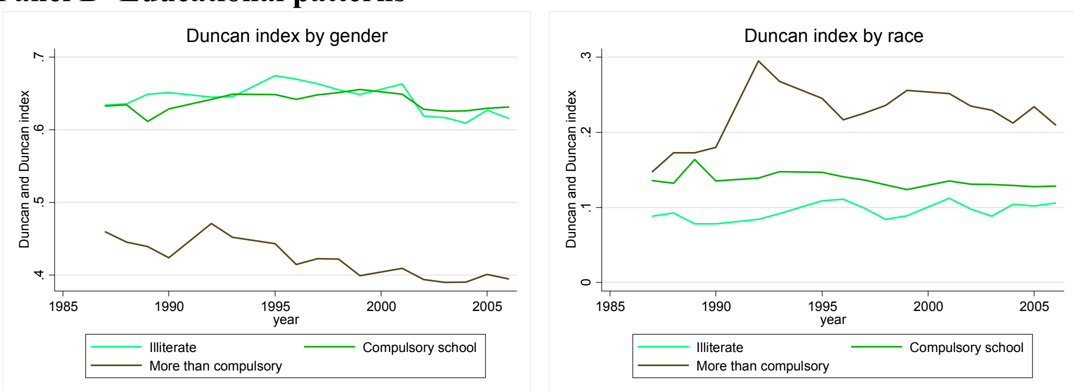
We have so far conducted our analysis of the Brazilian occupational structure, and the evolution of occupational segregation, at the national level and across the formal, informal and self-employed sectors. However, there are likely to be significant additional insights to be obtained by employing a still more disaggregated approach. Brazil is a huge country characterized by important differences across regions, between urban and rural areas, and across economic sectors. For example, in describing the evolution of informality over time, Ramos and Ferreira (2005) stress that the apparently dramatic rise in informality reported in some studies is, in fact, restricted to some metropolitan areas and to some sectors of economic activity. Similarly, given rapid economic, political and social changes over time it is reasonable to expect that patterns of occupational segregation may vary significantly for different age groups or for those with differing levels of education.

Figure 4: Occupational segregation disaggregated by characteristics of the labour force

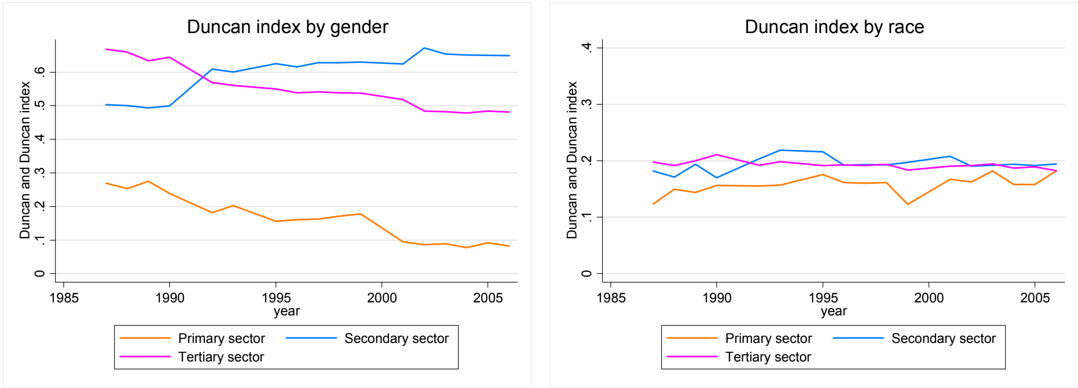
Panel A – Demographic patterns



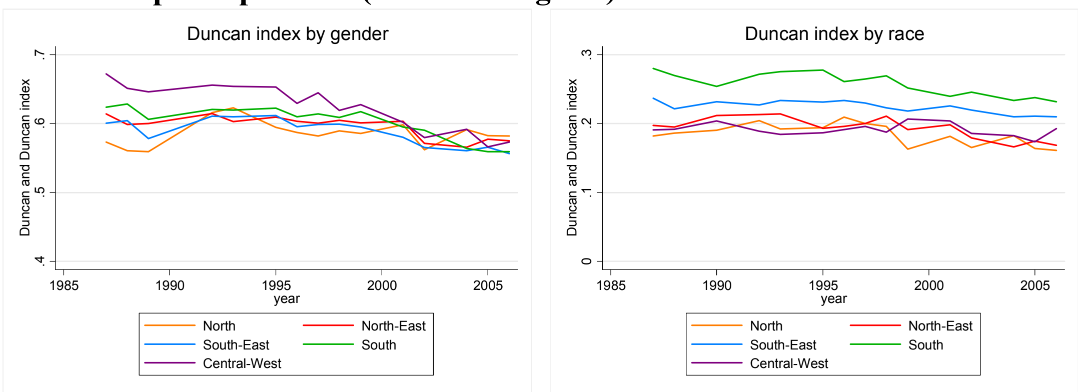
Panel B- Educational patterns



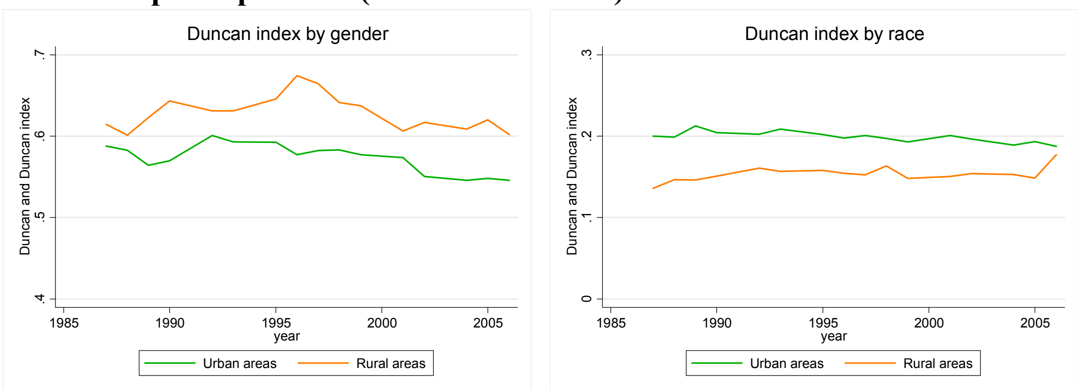
Panel C – Sectoral patterns (three main economic sectors)



Panel D – Spatial patterns (five main regions)



Panel E – Spatial patterns (urban/rural areas)



Source: Author's computations using PNAD 1987 - 2006.
Note: 1991, 1994 and 2001 missing years.

We thus follow the approach adopted by Ramos and Ferreira (2005) in studying informality and examine patterns of occupational segregation, disaggregated by sectoral and spatial attachment, and also according to demographic and educational characteristics. We calculate the Duncan index, the most widely used of the segregation indexes, at the three-digit level over time for each of the sub-groups of interest in order to disentangle different trends in the data.

Our main findings can be summarized as follows. In terms of demographic patterns, gender segregation is higher among older workers, while racial segregation is, somewhat surprisingly, higher among younger workers. With respect to educational patterns, gender segregation is lower among those with more education, but racial segregation is higher among those with more education. Although this may appear surprising, it is in line with evidence that female workers are, on average, more educated than men while non-white workers are less educated than whites, and are thus particularly poorly represented in high education occupations. Equally of interest, while aggregate data reveal an overall pattern of declining segregation in the country, the disaggregated analysis suggests that this decrease in segregation has occurred primarily among more educated workers, while for less educated workers levels of segregation have remained relatively stable.

Moving to sectoral patterns, we find that segregation has increased in the secondary sector, particularly by gender, while in the tertiary sector both gender and racial segregation have declined over time. In regard to the spatial patterns of segregation, we see that segregation has declined relatively homogenously across all regions. That said, there has been a particularly dramatic decline in gender segregation in the Central-West region, while racial segregation has been declining everywhere, but remains strikingly high in the South-East and South regions when compared to elsewhere in the country. Finally, gender segregation is higher in rural areas while racial segregation is higher in urban areas. These primary trends are displayed in figure 4. These results are now in greater detail.

5.3.4.1 Demographic patterns

The trends over time in occupational segregation across different age groups are displayed in panel A of the figure 4, above. As expected, gender segregation is higher among the adult and elderly labour force. Encouragingly, it is lower among the younger generation, while it has been decreasing across all age cohorts. The demographic patterns for racial segregation are very different. Segregation is higher among the younger generation and there is no clear pattern of decline over time. Although we do not report the plots disaggregated by formal, informal and self-employed sectors,[53] it is

[53] Trends in occupational segregation by gender and race disaggregated by the main characteristics of the labour force (age, education, region of residence, economic sector) and by formal/informal/self-employed sectors are available on request from the author.

important to note that the higher level of racial segregation reported for the younger generation appears to be limited to the non-formal sectors, while in the formal sector segregation is lower for the younger generation. Overall the higher level of gender segregation among elderly workers is consistent with a pattern of declining levels of gender segregation over time, while the higher level of racial segregation among younger workers, though more surprising, may reflect higher segregation among new, and less experienced, workers, primarily in the non-formal sectors.

5.3.4.2 Educational patterns

We have also explored how segregation varies depending on the level of educational attainment, which could also proxy for the degree of skill required for different jobs. As with the analysis by age groups, we notice very different patterns for gender and racial segregation (see panel B of figure 4). Gender segregation is much lower among the most educated/skilled workers than it is for illiterate workers or workers with only compulsory school education. Moreover, gender segregation has only decreased over time among more educated workers, and has remained relatively stable for less educated workers. In the case of racial segregation we find almost exactly the reverse. Racial segregation is found to be higher among more educated workers. This reflects the fact that non-white workers continue to be predominantly employed in unskilled and low-skilled occupations, leading to limited representation, and thus a high level of occupational segregation, particularly in occupations that require greater education and skills. Finally, it is important to note that while patterns of segregation disaggregated by educational attainment levels are very different for the two groups, racial segregation nonetheless remains lower in absolute terms even among the more highly educated.

5.3.4.3 Sectoral patterns

Turning to the analysis of segregation by sectors of economic activity, we first examine trends across the three major economic sectors, and then turn to a more detailed breakdown of economic activities. Looking first at trends over time in the three main sectors (panel C of figure 4) we find that the secondary sector records the highest levels of segregation by both gender and race. Interestingly, there has been no significant decline in either gender or racial segregation. The absence of any decline in racial segregation in the secondary sector is surprising given the significant entry of non-whites into the sector over time, with the non-white share having increased from just below 45% to around 53% over time. One possible explanation is a particularly sharp division between top-jobs for white workers and unskilled jobs for non-white workers within these industries. By contrast, gender segregation has remained relatively stable as female participation in the secondary sector has remained fairly low (at roughly 24-25%), while there seems to have been little change in the occupations in which women are employed.

The tertiary sector has slightly lower levels of both gender and racial segregation, while both have declined gradually over time. Within the tertiary sector women and non-whites comprise the majority of the labour force over the period of interest, and, as such, declining segregation appears to reflect declining concentration within occupations. Although both women and non-whites have continued to enter the tertiary sector rapidly, leading to its growing share in the labour force as a whole, men have also continued to enter the tertiary sector, thus contributing to a declining segregation.[54]

Finally, there is significantly less segregation in the primary sector, with gender segregation declining over time, and racial segregation either stable or increasing. Sharply lower segregation in the primary sector, particularly by gender, is a reminder of the importance of focusing on the entire labour market, and not simply urban areas, when examining patterns of occupational segregation.

Not surprisingly, when we analyse sectoral patterns using the more detailed classification of economic activities, we find quite dramatic differences in patterns of segregation. Gender segregation is generally higher in the construction, mining and transport sectors, while racial segregation is higher in sectors such as mining, electricity, gas and water services and social services. The nature of these two types of segregation is, however, very different. In sectors in which gender segregation is highest, like construction, mining and transport, the high level of segregation results from the fact that these jobs are overwhelmingly male dominated. By contrast, mining, electricity, gas and water services and social services exhibit a high degree of racial segregation not because they are dominated entirely by white or non-white workers, as concentration is relatively low. Instead, segregation in these cases is driven by the fact that within these

[54] The tertiary sector covered 43% of the entire economy in 1987 and had grown to 67% by 2006. The female share of the tertiary sector declined from 56% in 1987 to 52% in 2006, while the non-white share increased, moving from 46.4% to 51% over this same period.

sectors whites are concentrated in the top jobs while non-whites are concentrated in lower-skilled jobs.

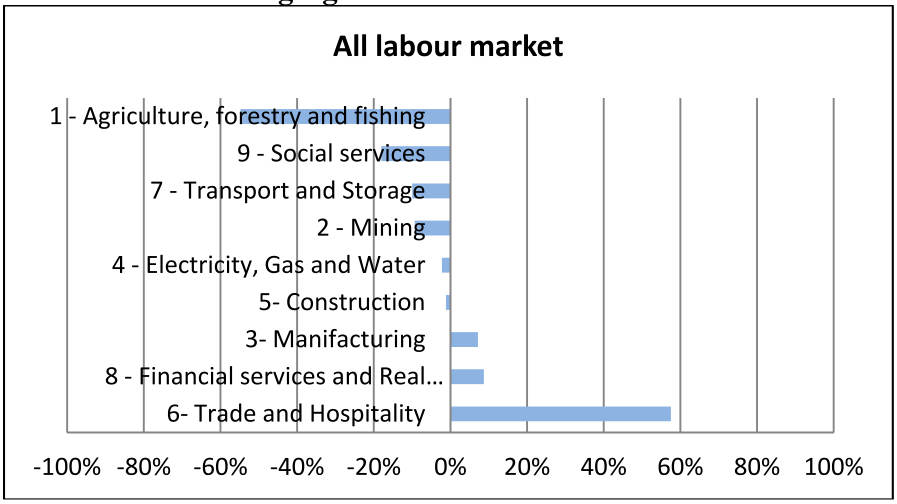
In order to better understand the evolution of segregation over time across several economic activities, figure 5 reports changes in the Duncan index by sectors of economic activity between the first and last years of the period being analysed. Although not reported here, a similar analysis has been conducted disaggregating the formal, informal and self-employed sectors,[55] and several key discrepancies that emerge are noted below.

In panel A of figure 5 we see that the sectors that have experienced the greatest decline in gender based segregation are the primary (i.e., agriculture, forestry and fishing) and the social services sectors. The decrease in the primary sector occurs primarily in the informal and self-employed sectors, while the decrease in the social services sector is common across both the formal and non-formal labour markets. In general, the decrease in segregation appears to reflect rising female participation in these sectors of economic activity. While we saw above that segregation has declined in the tertiary sector as a whole, we see increases in gender based segregation in sub-sectors of the tertiary sector: trade and hospitality and financial services. Interestingly, within the trade and hospitality sector increasing segregation is limited to non-formal labour markets, while within financial services the increase in segregation is concentrated in the formal labour market. The former seems to reflect the continued entry of women into informal sector jobs in the trade and hospitality sectors, while the rise of gender segregation over time in the financial sector reflects the entry of more men than women into financial services. Finally, the overall labour market recorded an increase in gender segregation in the manufacturing industry.

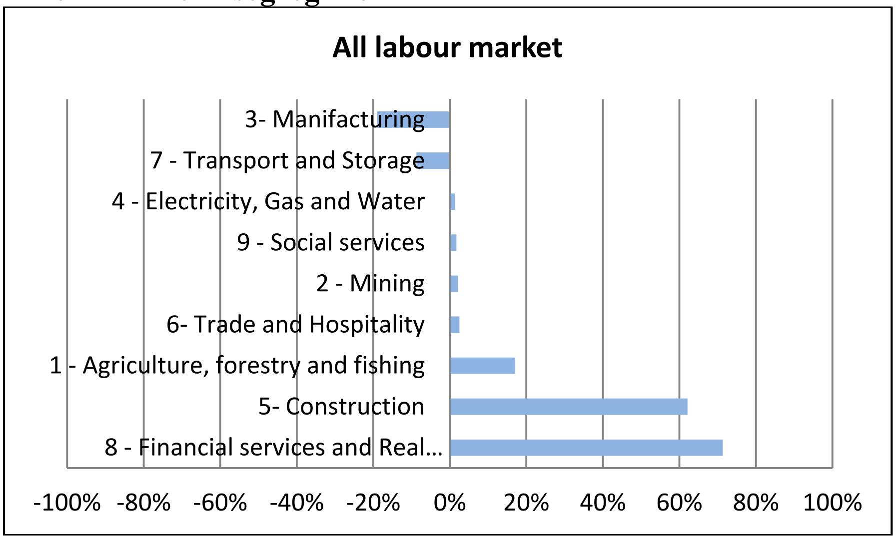
[55] The same charts, disaggregated into the formal, informal and self-employed sectors, are available upon request from the author.

Figure 5: Changes in the Duncan Index Between 1992 and 2006 by Sector of Economic Activity

Panel A – Gender segregation



Panel B – Racial segregation



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

The most significant decline in racial segregation is in the manufacturing sector, with this trend common to both the formal and non-formal labour markets (see figure 5 panel B). We also note a decline in segregation in the transport sector. On the other hand, both the construction and financial services sectors have experienced large increases in segregation, the latter possibly the result of a rapid increased entry of white male workers. Finally, we see a number of sectors in which a modest aggregate change in segregation disguises contrasting trends between the formal and informal sectors, with increasing segregation largely restricted to the informal sector, while segregation is declining in the formal sector. The mining sector exhibits an overall increase in segregation, with segregation decreasing in the formal and self-employed sectors but consistently increasing in the informal sector. In similar fashion, the trade and hospitality sector reports increasing segregation overall, with segregation decreasing in both the formal and informal sectors, but increasing substantially among the self-employed. Finally, social services has experienced an overall decrease in racial segregation, but this is restricted to the formal sector, while in both the informal and self-employed sectors a substantial increase is registered.

5.3.4.4 Spatial patterns

The analysis of spatial patterns provides additional insights into the evolution of Brazilian segregation. Spatial patterns can be investigated by looking at regional differences, difference between states or differences between urban and rural areas. In what follows we briefly report differences across regions, and between rural and urban areas, before focusing on differential patterns across states, which provide the most interesting insights.

Beginning with regional patterns, the Central-West is the region with the highest level of gender segregation but it has also recorded the most rapid decline over time, such that there is now a reasonable degree of homogeneity across the five regions. In the case of racial segregation we see much more pronounced regional differences, as segregation has remained markedly higher in the South and South-East than elsewhere in the country, despite a significant decline over time in the South in particular. More generally, although change has proceeded at different rates across the regions, segregation has been declining over time in all five regions (see panel D of figure 4).

Turning to differences between urban and rural areas, we observe that rural areas are subject to higher levels of gender segregation than urban areas, while racial segregation is more pronounced in urban areas. The fact that gender segregation is higher in rural areas is surprising given earlier evidence that segregation in the primary sector, including agriculture, is much lower than in other sectors. The inference to be drawn appears to be that there are extremely high levels of segregation in rural, non-agricultural, occupations. Finally, both gender and racial segregation register a steady decline in urban areas over time, while the pattern for rural segregation is less clear.

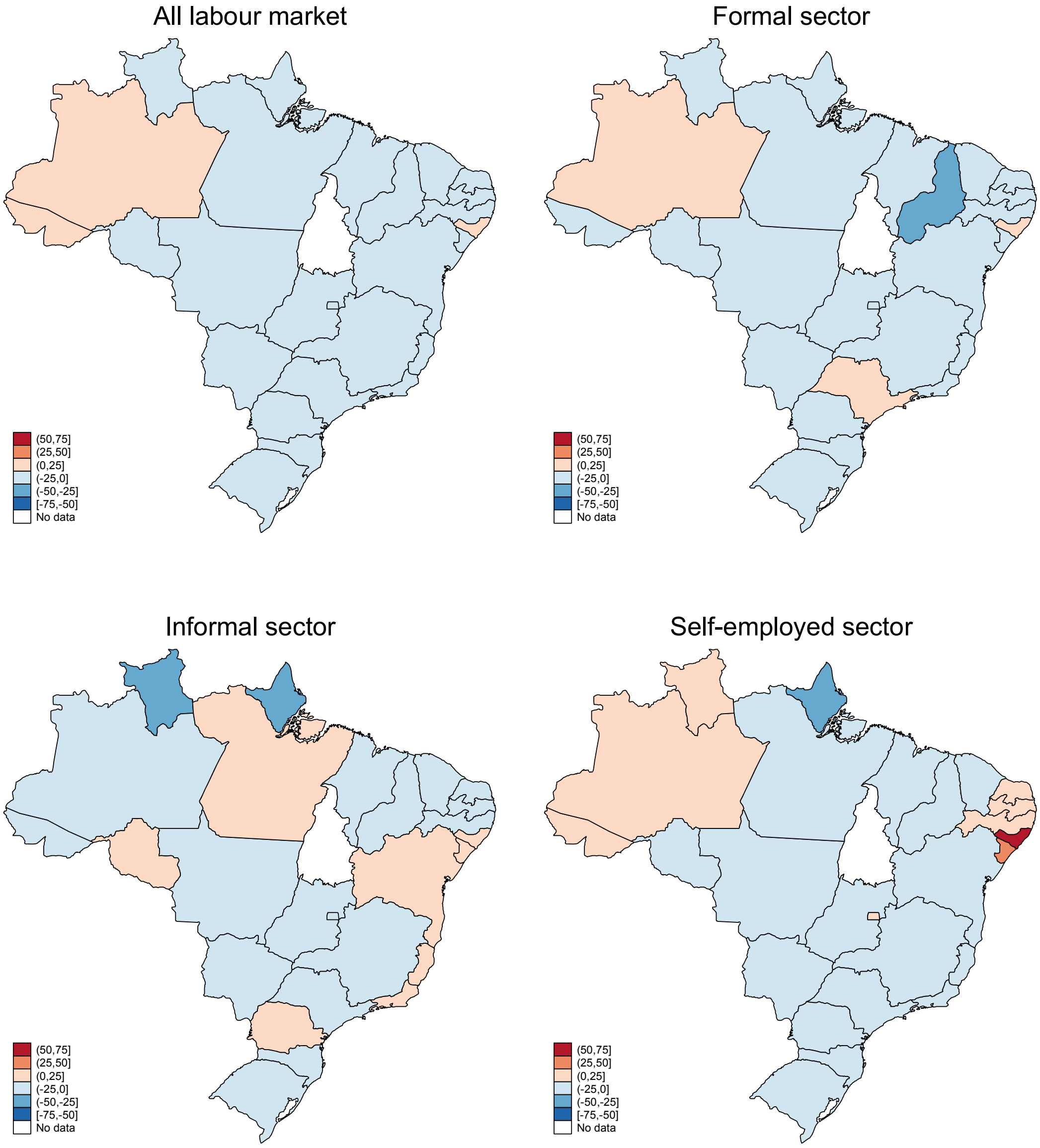
Having highlighted these basic regional and geographic trends, we now turn to differences among Brazilian states, the *Unidade de Federação*,[56] and figure 6 reports changes in segregation over time in each of the states individually. Focusing initially on gender segregation within the entire labour market, we find that, despite declining segregation in each of the regions, gender segregation has actually increased in three states in the North and North-East regions, Alagoas, Amazonas and Acre. On the other hand, and consistent with the regional patterns, the states of the Central-West region (Mato Grosso do Sul and Distrito Federal in particular) experienced the most steady decline in segregation, albeit from a very high initial level. It is potentially interesting to note that in the Distrito Federal, where there are more government offices and where governmental policy might be expected to be strongest, gender segregation has declined most quickly. If we restrict the analysis to formal labour markets the patterns remain broadly similar, with the notable exception the state of Sao Paolo, which experienced an increase in segregation within the formal sector. By contrast, if we focus on the informal and self-employed sectors there are a considerable number of states that have experienced an increase in gender segregation.

With respect to racial segregation, the most striking feature of the data is that while five states have experienced an overall increase in racial segregation this increase is almost entirely concentrated in the non-formal sectors of the economy, as only Amazonas records an increase in racial segregation in the formal sector. Indeed, if we focus on the non-formal sectors we find some large increases in segregation. This is particularly the case in the states of Espirito Santo and Goias in the informal sector and in the states of Piaui, Sergipe, as well as the Distrito Federal, for the self-employed sector.

[56] As described in chapter 2, Brazil is a Federal Republic comprising 27 states that can be grouped into five main geographical regions. These are the North, the North-East, the South-East, the South and the Central-West.

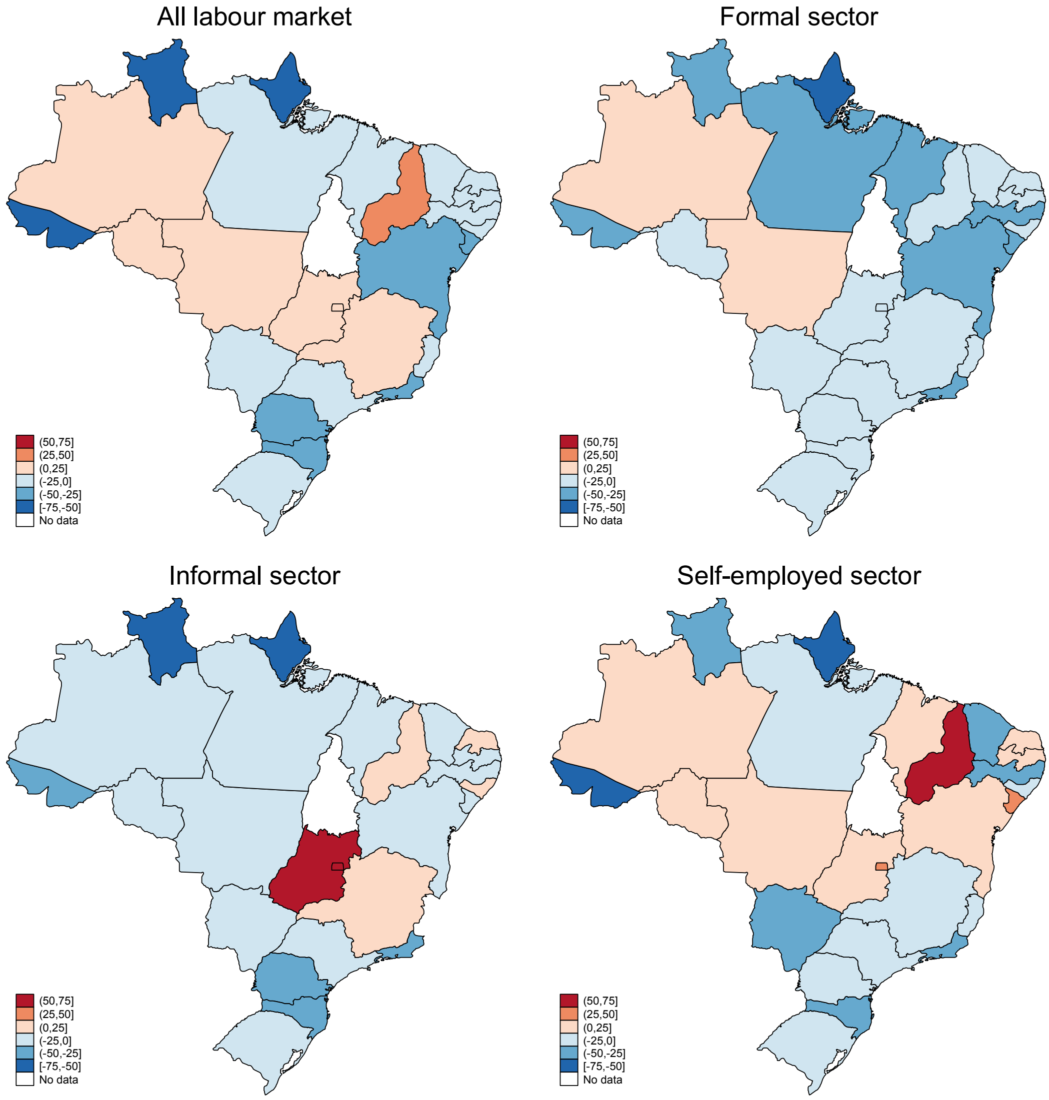
Figure 6: Changes in the Duncan Index Between 1987 and 2006 by State (*Unidade de Federação*)

Panel A – Gender segregation



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

Panel B – Racial segregation



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

These large variations in spatial outcomes provide revealing information, but are also noteworthy because they may represent a way of thinking in a more subtle way about the determinants of segregation. The need for further study is strongly suggested by the sharply differing patterns, as there does not seem to be a simple narrative of a generalized change in norms and a corresponding reduction in segregation equally across the country. Instead, there seem to be more complex factors at work here. A particularly intriguing possibility for future research may lie in exploiting differences across states in order to explore the impact of labour market legislation, or more specifically anti-discrimination legislation, in a way similar to that employed elsewhere by Neumark and Stock (2006). That is, interesting quasi-experiments could be undertaken by exploring differences in the timing, content and enforcement of anti-discrimination laws and quota systems at the state level. In the first case, research could focus on differences in the timing and content of anti-discrimination laws or quota systems that have been introduced by states in addition to federal laws. These laws might prove to be independently important in shaping outcomes, or might indicate states that are more politically committed to gender or racial equality more generally. In the second case, research could also focus on differences across states in the implementation of federal anti-discrimination laws or quotas in order to see whether these laws have played any role in decreasing occupational segregation. While measuring enforcement poses an obvious challenge it might be possible to focus on direct indicators of enforcement – for example, hiring according to the quota system in the branches of the federal government or the enforcement of quotas – or to focus on the relevance of broader good governance indicators at the state level, as they may be a useful proxy for enforcement as reflected in the quality of the rule of law. Obviously the feasibility of these proposed investigations depends on the availability of data. For example, this would require a list of relevant laws and quota systems across states and over time or good governance indicators not only at the country level but broken down state by state. At the moment this type of information is not available in readily usable form, and a major investment would be required to assemble the relevant data, rendering such analysis significantly beyond the scope of this study.[57] Despite these challenges, this is a clearly a potentially fruitful direction for future research.

[57] Aside from the list of federal ADL presented earlier, we have so far been able to identify a more disaggregated collection of laws (at federal, state and municipal levels) on ADL focused specifically on racial issues until 1998, but we have yet to identify sources for ADL from more recent years or focused on gender issues (Silva Jr., 1998). In the absence of such sources compiling data on legislation across 27 states is likely to be an extremely difficult task.

5.4 Decomposition of changes in segregation over time

We have so far provided an analysis of the evolution of occupational segregation over time in Brazil. This marks an important step forward in our understanding of these questions, as, in contrast to earlier studies, our analysis has drawn on more complete and reliable data over a longer period of time, and implemented a more careful analysis of occupational segregation. This analysis has revealed an array of important patterns in occupational segregation that have not been fully recognized previously, and which are summarized in greater detail in the conclusions to this chapter.

What remains is to seek a better understanding of the underlying forces that have driven these changes in occupational segregation over time. This is, of course, a hugely complicated task, and identifying the many social and economic factors that have contributed to observed outcomes is well beyond the scope of this study. Instead, the goal here is to use decomposition methodologies to disentangle the broad forces that have contributed to changes in occupational segregation. In particular, the subsequent analysis seeks to understand whether declining occupational segregation is driven by a more homogenous composition by gender and race *within* each occupation, by changes in the occupational structure itself (occupation weights), or by changes in the sub-population shares (gender or racial composition) of the labour force. This is particularly crucial given that the Duncan index, like other indices of segregation, is sensitive to changes in occupational weights (i.e. occupational structure) and to changes in gender and racial shares of the labour force.

We thus employ the decomposition methodology proposed by Deutsch, Flueckiger and Silber (2009), which combines the Karmel and MacLachlan (1988) decomposition and the concept of the Shapley value in order to distinguish between three three different sources of variation in occupational segregation (see also Shorrocks (1999) and Sastre and Trannoy (2002)). We find that the decline in both gender and racial segregation is driven primarily by the increasingly equitable representation of female and non-white workers within individual occupations. By contrast, broader changes in occupational structure have had a more mixed effect, as they have had only a small impact on trends within the formal sector, while they have tended to contribute to higher segregation in the informal sector. This reflects the increased relative size of more segregated occupations, and the fact that new female and non-white entrants to the labour force have largely entered historically female and non-white dominated informal occupations. This striking difference in outcomes between the formal and informal sectors is an important finding to which is returned in the conclusions. Ultimately, the overall downward trend in occupational segregation reflects the fact that the increase in segregation generated by changes in broad occupational structure is offset by the general improvement in the composition of the labour force within individual occupations. This is a particularly important finding as one might make the error of attributing declining segregation solely to the entry of women or non-whites into the labour force, while these results point towards more profound changes in the extent of segregation. What follows first describes the decomposition methodology before presenting the detailed findings.

5.4.1 The methodology

The decomposition methodology proposed by Deutsch, Flueckiger and Silber (2009) aims to decompose changes over time in segregation measures into three main components. First, segregation can change over time because of the changes in the relative weights of different occupations. Second, segregation can change over time because of variation in the sub-population (gender or racial) composition of the total labour force. Finally, segregation may change over time because of variation in the sub-population composition within each occupation. This latter source of variation is also defined as 'net segregation', or variation in the 'internal structure', because it is separate from variation that can occur 'in the margins', which are given by changes in the relative weights of occupations or in the shares of the sub-populations in the labour force. The sum of these three sources of variation (i.e., the internal structure and the two components of the margins) is defined as 'gross variation' in the occupational segregation literature.[58]

Following the Deutsch, Flueckiger and Silber (2009) derivation, it is possible to decompose the change in a segregation index over time as follows:

$$\Delta I = I_v - I_p \tag{6}$$

[58] The Shapley decomposition by Deutsch, Flueckiger and Silber (2009) is inspired by the decomposition technique proposed in Karmel and Maclachlan (1988). In fact, they decompose the segregation index into the mixed effects (gender, occupation and gender by occupation) and the composition effect. These are similar to the variations due to 'the margins' and to the 'internal structure', respectively, in the Deutsch, Flueckiger and Silber (2009). The important innovation in the Shapley decomposition is the absence of an interaction term or residual from the decomposition.

where I_v and I_p represent the indices for the final and initial periods of time respectively. If we apply the concept of the Shapley decomposition following Deutsch, Flueckiger and Silber (2009), the total variation, defined as 'gross variation' in segregation over time, can be decomposed as follows:

$$\Delta I = f(\Delta m, \Delta is) = C_{\Delta m} + C_{\Delta is} \tag{7}$$

where $C_{\Delta m}$ and $C_{\Delta is}$ represent the two main components of the decomposition, the component of the change due to variation in the 'margins' and the component of the change due to variation in the 'internal structure' (or 'net segregation') and they are

$$C_{\Delta m} = \frac{1}{2} f(\Delta m) + \frac{1}{2}[f(\Delta m, \Delta is) - f(\Delta is)] \tag{8}$$

and

$$C_{\Delta is} = \frac{1}{2} f(\Delta is) + \frac{1}{2}[f(\Delta m, \Delta is) - f(\Delta m)] \tag{9}$$

The contribution of these components can be re-expressed also as follows:

$$C_{\Delta m} = \frac{1}{2}\{[I(s) - I(p)] + [I(v) - I(w)]\} \tag{10}$$

and

$$C_{\Delta is} = \frac{1}{2}\{[I(w) - I(p)] + [I(v) - I(s)]\} \tag{11}$$

where the set of matrices employed in the above equations are obtained by interacting both the margins and internal structure of the segregation matrices from which the two indices I_v and I_p can be drawn. The two initial matrices are P and V and we need to compare them to derive matrix S, which has the internal structure of P but the margins of V. In the same way, matrix W can be derived with the internal structure of matrix V and the margins of matrix P simply by inverting the process.

In order to explain the derivation, let's start by considering the matrix P. This matrix has the ratio ${T_{ij}}/{T}$ in its internal structure where T_{ij} is the number of individuals in occupation *i* from the sub-population *j* and *T* is the total number of workers. The margins of matrix P are defined by $p_{i\cdot} = {}^{T_i}/_{T}$ and $p_{\cdot j} = {}^{T_j}/_{T}$ which are respectively the horizontal margins (occupational structure) and the vertical margins (shares of the sub-populations).

To derive the matrix S, we need to multiply all elements of P by the ratio ${}^{v_{i\cdot}}/_{p_{i\cdot}}$ and obtain an intermediate matrix X. Its elements need to be multiplied by the ratio ${}^{v_{\cdot j}}/_{x_{\cdot j}}$ to obtain a new matrix Y and so on. After several iterations, the matrix will converge to the matrix S with the internal structure of P and the margins of V (see Deming and Stephan, 1940). As already noted, we could also start with the matrix V and, by applying the same procedure, end up with the matrix W that has the internal structure of matrix V and the margins of matrix P.

Now, the proposed decomposition permits us to decompose the variation in the margins into components due to the variation in the occupational structure and the shares of the sub-populations. In other words, we have

$$C_{\Delta m} = C_{\Delta h} + C_{\Delta t} \tag{12}$$

where $C_{\Delta h}$ represents the contribution from changes in occupational structure and $C_{\Delta t}$ represents the contribution from changes in the shares of sub-populations in the total labour force. Using the same procedure as before we can express these two components as follows:

$$C_{\Delta h} = \frac{1}{2}\frac{1}{2}[\{[I(l) - I(p)] + (I(s) - I(k)]\} + \{[I(v) - I(c)] + (I(f) - I(w)]\}] \tag{13}$$

and

$$C_{\Delta t} = \frac{1}{2}\frac{1}{2}[\{[I(k) - I(p)] + (I(s) - I(l)]\} + \{[I(v) - I(f)] + (I(c) - I(w)]\}] \tag{14}$$

In order to derive these components we need to define additional matrices (see Deutsch, Flueckiger and Silber (2009) for the detailed construction of these matrices):

- matrix L with the internal structure of P, the horizontal margins of V and the vertical margins of P;
- matrix K with the internal structure of P, the horizontal margins of P and the vertical margins of V;

- matrix F with the internal structure of V, the horizontal margins of V and the vertical margins of P;
- matrix C with the internal structure of V, the horizontal margins of P and the vertical margins of V.

Through this decomposition, we are then able to decompose the change between two periods into:

$$\Delta I = C_{\Delta is} + C_{\Delta h} + C_{\Delta t} \qquad (15)$$

where $C_{\Delta is}$ represents the variation due to changes in the sub-population shares within occupations (the net segregation or changes in internal structure), $C_{\Delta h}$ represents the variation due to changes in the occupational structure of the labour markets (i.e. the weights of each occupation) and, finally, $C_{\Delta t}$ represents changes in the sub-population shares of the total labour force (i.e. gender or racial composition of the labour force).

5.4.2 Empirical Findings across the Formal and non-Formal Sectors

We perform the decomposition of changes in gender and racial segregation using the Duncan index between two periods: the initial period, comprising the years 1987, 1988, 1989, 1990 and 1992, and the final period, comprising the years 2002 to 2006. We aggregate the first and last five years periods in order to have a sufficient number of observations to implement the decomposition separately across the formal, informal and self-employed labour markets, as well as disaggregated by key characteristics of the labour force.[59] This aggregation does not appear to be problematic, as changes in occupational distribution within the conflated years are relatively modest. Finally, we also compute bootstrapped standard errors for the overall changes in occupational segregation, as well as for the components of these changes, using draws from 500 random samples in order to test the statistical significance of the point estimates for each component. The findings of the decomposition of changes in the Duncan index over time across the formal, informal and self-employed labour markets are reported in table 3 and depicted in figure 7.

[59] For the sake of brevity, the Shapley decomposition results for the Karmel and Maclachlan index and the Gini segregation index are presented in the appendix of this chapter, given that the statistically significant results are largely unchanged relative to the results presented here (see table A3, A4, A5 and A6 in the appendix).

Table 3: Shapley decomposition of changes in Duncan index over time across sectors

	I_p	I_v	ΔI	$C_{\Delta is}$	$C_{\Delta m}$ (1) + (2)	$C_{\Delta h}$ (1)	$C_{\Delta t}$ (2)
All labour market							
gender	0.5983	0.5624	-0.0359	-0.0343	-0.0016 n.s.	-0.0080	0.0064
s.e.	0.0011	0.0009	0.0014	0.0035	0.0032	0.0032	0.0006
			100.00%	95.72%	4.28%	22.89%	-18.60%
race	0.1981	0.1928	-0.0053	-0.0262	0.0209	0.0199	0.0010
s.e.	0.0014	0.0011	0.0017	0.0043	0.0041	0.0041	0.0001
			100.00%	493.30%	-393.30%	-374.41%	-18.89%
Formal sector							
gender	0.5526	0.5167	-0.0359	-0.0367	0.0008 n.s.	-0.0105	0.0113
s.e.	0.0018	0.0014	0.0023	0.0054	0.0053	0.0055	0.0012
			100.00%	102.25%	-2.25%	29.14%	-31.38%
race	0.1891	0.1768	-0.0123	-0.0096 n.s.	-0.0027 n.s.	-0.0032 n.s.	0.0004
s.e.	0.0021	0.0017	0.0026	0.0065	0.0058	0.0058	0.0001
			100.00%	77.71%	22.29%	25.80%	-3.51%
Informal sector							
gender	0.6949	0.6492	-0.0458	-0.0896	0.0438	0.0551	-0.0113
s.e.	0.0020	0.0016	0.0025	0.0103	0.0101	0.0113	0.0021
			100.00%	195.66%	-95.66%	-120.29%	24.63%
race	0.1985	0.1835	-0.0150	-0.0604	0.0454	0.0423	0.0030
s.e.	0.0023	0.0020	0.0030	0.0083	0.0078	0.0077	0.0002
			100.00%	401.82%	-301.82%	-281.75%	-20.07%
Self-employed sector							
gender	0.6300	0.6099	-0.0201	-0.0452	0.0251	0.0255	-0.0004 n.s.
s.e.	0.0020	0.0019	0.0027	0.0145	0.0143	0.0143	0.0005
			100.00%	225.12%	-125.12%	-126.98%	1.85%
race	0.1294	0.1400	0.0106	-0.0266	0.0372	0.0359	0.0013
s.e.	0.0027	0.0022	0.0034	0.0089	0.0083	0.0083	0.0002
			100.00%	-249.96%	349.96%	338.11%	11.85%

Source: Author's computations using PNAD 1987,1988, 1989, 1990, 1992 and 2002, 2003, 2004, 2005, 2006.
Note: the initial period comprises 1987-1988-1989-1990-1992 and the final period comprises 2002-2003-2004-2005-2006. Standard errors bootstrapped with 500 replications. n.s. indicates those components that are not statistically significant at 5%.

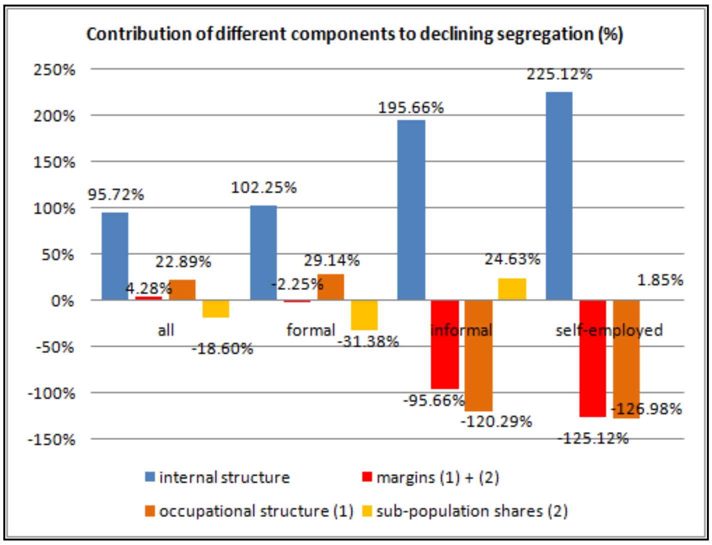
In general, we observe that the decline in both gender and racial segregation, which is also called the 'gross variation', is driven overwhelmingly by 'variations in the internal structure', also called 'net variation' in segregation – that is, by declining concentration by gender and race within individual occupations. The contribution of the internal structure component is almost always statistically significant, as shown in table 3. By contrast, we find that the impact of 'variations in the margins' – that is, changes in occupational structure (occupation weights) and in the share of different population sub-groups in the overall labour force - is to increase levels of occupational segregation. However, the 'variations in the margins' component warrants more careful analysis, as its two components behave differently across the formal and non-formal labour markets. We look first at the formal, informal and self-employed sectors separately, and then consider the labour market as a whole.

In the formal sector, when looking at both gender and racial segregation, changes in occupational structure (occupation weights) have contributed to declining segregation, although the effect is only statistically significant in the case of gender. On the other hand, the non-formal sectors show the opposite pattern, as changes in occupation weights are the main source of upward pressure on levels of both racial and gender segregation, with a particularly dramatic effect in the case of racial segregation. This pattern is particularly pronounced for the self-employed sector, where the increase in segregation caused by changes in occupation weights completely offsets the decline in occupational segregation resulting from variations in the internal structure of occupations, leading to an aggregate increase in racial segregation for self-employed workers. All of these estimates for the non-formal sector are statistically significant. The last component, changes in the sub-population shares, generally contributes to an increase in segregation, with the exception of gender segregation in the informal sector, where the increase in female participation contributes positively to reducing gender segregation (the same occurs in the self-employed sector, but the component is not statistically significant in that case). That said, this component is generally comparatively small in magnitude.

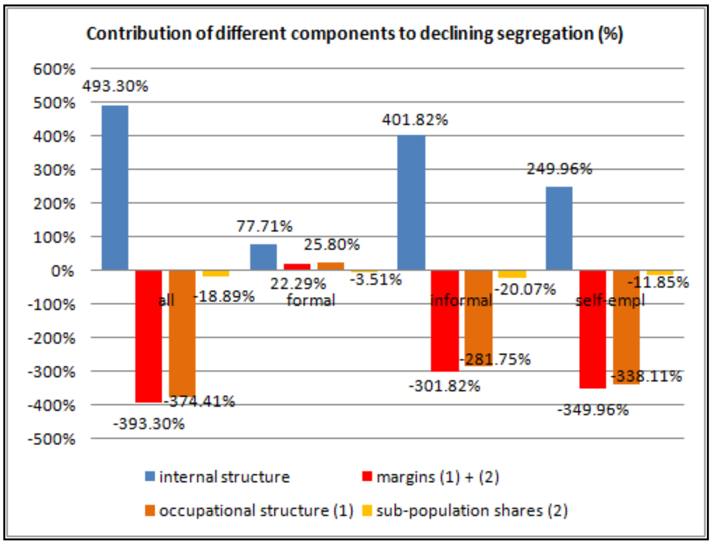
When we combine the formal, informal and self-employed sectors and look at the labour market as a whole, we see that the aggregate effect of changes in the occupational structure differs between gender and racial segregation. Focusing first on gender segregation, we see that the aggregate effect of changes in the occupational structure is to reduce segregation, as the effect in the formal sector (reducing segregation) outweighs the effect that we observe in the informal sector (increasing segregation). This is consistent with the fact that women recorded a larger increase in participation into the formal sector (from 34% to almost 43%) than into the informal sector (from 42.2% to 47.5%), while the formal sector is less segregated than the informal sector, thus implying that the growth of the formal sector is likely to result in a less segregated occupational structure overall. Turning to racial segregation, the results are more straightforward: changes in the internal structure reduce segregation, as is the case in each sector on its own, while the 'variations in the margins' component increases segregation, consistent with the fact that this component increases segregation in the non-formal labour markets and is statistically insignificant in the formal sector.

Figure 7: Contribution of different components to declining segregation (%)

Panel A – Gender segregation



Panel B – Racial segregation



Source: Author's computations using PNAD 1987,1988, 1989, 1990, 1992 and 2002, 2003, 2004, 2005, 2006.

We have now laid out these broad findings on the determinants of gender and racial segregation, but it still remains to reconcile these findings from the decomposition exercise with the descriptive data on occupational structure and segregation presented earlier.

The most important finding is that the primary driver of falling occupational segregation is variation in the internal structure, and this is consistent with the descriptive data presented in chapter 4. Looking first at gender segregation, the data presented in table 4a of chapter 4 reveals that almost all of the most female dominated occupations experienced a decreasing female share over time, as men increasingly entered these jobs. For example, 93.45% of teaching associate professionals were women in 1987, while this share had fallen to 82.8% in 2006; customer services clerks moved from having a female share of 83.13% in 1987 to 75.19% in 2006. This pattern is, in fact, common across a wider range of female dominated jobs. Interestingly, the pattern is somewhat more mixed when we look at the most male dominated occupations, as some of these occupations have remained almost exclusively male over time. That is, a small number of male dominated occupations have remained closed to women (e.g., drivers and mobile plant operators, extraction and building trades workers, metal and machinery related trades workers), while other male dominated occupations have witnessed a significant increase in the entry of female workers (e.g., physics, engineers and sales persons).

Turning to racial segregation, variations in internal structure are equally important, though the sources of this variation are slightly different, as revealed in table 4b of chapter 4. Among occupations historically dominated by non-whites, the share of non-white labourers has declined in some areas, with the non-white share of mining, construction, manufacturing and transport falling from 80.36% in 1987, to only 59.93% in 2006. However, other non-white dominated professions have seen little change in their composition over time, in part because the extent of segregation in these occupations is comparatively low given that non-white workers are extremely dominant in only a very small number of occupations in Brazil. As such, it appears that another important source of declining intra-occupation segregation has been the growing share of non-whites in erstwhile white dominated professions, as this is common across almost all of the occupations listed in table 4b of chapter 4 (e.g., life science and health professionals and teachers).

While variations in internal structure have thus driven declining segregation, we find that changes in the margins have on average increased occupational segregation. By far the most important trend is the fact that across both gender and racial segregation changes in occupation weights have contributed to increasing levels of segregation, though this effect is concentrated entirely in the non-formal labour markets. The implication is that within informal labour markets relatively concentrated occupations have grown larger over time, though this has not been the case in the formal sector.

In examining gender segregation we observe, for example, that the occupation group “Personal and protective services workers” (cod 51), which is female dominated

(69.03% in 1987 and to 65.1% in 2006), has experienced a rapid increase in size, and is concentrated in the informal sector. Figure 5a in chapter 4 depicts the number of female workers that have entered these professions over time, and the expansion of some occupations that are part of this group (namely cod.512, 514, 522) is clearly discernible, particularly for housekeepers and restaurant workers (cod.512) in the informal sector. This means that female concentrated occupations have expanded more than others, and largely in the non-formal sectors, leading to an increase in aggregate occupational segregation. We find similar patterns in terms of racial segregation, where the overall impact of variation in occupation weights in increasing segregation is relatively greater. As with gender segregation, we see that the most rapidly growing occupations, namely “Personal and protective services workers” (Cod 51) and non-self-employed agricultural occupations (cod. 612), have high and increasing shares of non-white workers and have had their growth concentrated largely in the informal labour market.

Finally, the data suggest, somewhat counter-intuitively, that, after accounting for the other trends discussed so far, the increasing share of women and non-whites in the labour force has contributed to increasing segregation, implying that new entrants to the labour force have disproportionately entered occupations in which women and non-whites, respectively, were already dominant. This is evident, for example, in the continued entry of women, and particularly non-white women, into the already female-dominated housekeeping profession. That said, the magnitude of these effects is comparatively small. The most notable findings, in terms of magnitude, relate to gender segregation. Here we see that the entry of women into the formal labour force appears to have significantly increased segregation. This implies that women have, to a significant degree, entered the formal market in traditionally female dominated jobs, while, as we noted above, rarely entering the most male dominated occupations. By contrast, in the informal sector the growing share of women in the labour force has modestly reduced occupational segregation, suggesting that women have, to a greater degree, also entered jobs that have not traditionally been dominated by women in the past.

5.4.3 Empirical Findings Disaggregated by Characteristics of the Labour Force

We also decompose changes in gender and racial segregation disaggregated by several key characteristics of the labour force. Table 4 reports the estimated components of the decomposition together with their bootstrapped standard errors. As many of these results are not statistically significant, we comment only on the statistically relevant results.

When we divide the labour force by age group we find that the contribution of changes in occupational structure to increasing gender segregation is particularly true for elderly people. In contrast, in the case of racial segregation there are no differences across age groups with changes in the occupational structure always increasing racial segregation. These findings are all statistically significant. The negative contribution of changes in the occupational structure is so strong among young people that it offsets the positive effect of improvements in the internal structure with the result that young people have experienced an aggregate increase in racial segregation.

Looking at educational attainment, and again focusing only on the statistically significant components, we find that among the more educated workers gender segregation decreases because of the positive contribution of changes in internal structure, while the negative contribution from changes in occupational structure is negligible. Conversely, racial segregation increases among the well educated because the contribution of changes in internal structure is small and completely offset by the large negative contribution of changes in occupational structure. This finding is consistent with the fact that in more educated and skilled jobs there is a prevalence of white workers.

When looking at sectoral patterns, we note that changes in occupational structure contribute to increasing gender segregation only in the broadly defined secondary sector. This reflects the fact that the occupations that have grown fastest within the secondary sectors, including food processing, wood treaters, textile processing related jobs and machine and plant operators and assemblers, are all highly gender segregated occupations. In terms of racial segregation, the only statistically significant decomposition effect is within the tertiary sector, where changes in internal structure have strikingly reduced segregation.

When we focus on spatial patterns, and separate rural and urban areas, we notice that gender segregation has decreased in both rural and urban areas, driven by the positive contribution of changes in internal structure, particularly in urban areas, while the negative contribution of changes in occupational structure is negligible. In regard to racial segregation, we record a consistent decrease in urban areas driven by a clear improvement in the internal structure within occupations.[60]

Table 4: Shapley decomposition of changes in Duncan index over time disaggregated by characteristics

	I_p	I_v	ΔI	$C_{\Delta is}$	$C_{\Delta m}$ (1) + (2)	$C_{\Delta h}$ (1)	$C_{\Delta t}$ (2)
By age							
young							
gender	0.5791	0.5500	-0.0291	-0.0324	0.0033 n.s.	-0.0010 n.s.	0.0043
s.e.	0.0017	0.0015	0.0023	0.0064	0.0063	0.0061	0.0007
			100.00%	111.16%	-11.16%	3.52%	-14.68%
race	0.2019	0.2157	0.0138	-0.0205	0.0343	0.0337	0.0006
s.e.	0.0021	0.0019	0.0028	0.0076	0.0072	0.0072	0.0001
			100.00%	-148.09%	248.09%	243.86%	4.24%
adult							
gender	0.6354	0.5926	-0.0427	-0.0411	-0.0016 n.s.	-0.0067 n.s.	0.0051
s.e.	0.0017	0.0013	0.0021	0.0046	0.0043	0.0049	0.0021
			100.00%	96.21%	3.79%	15.71%	-11.92%
race	0.1995	0.1854	-0.0141	-0.0288	0.0148	0.0133	0.0014
s.e.	0.0022	0.0017	0.0027	0.0059	0.0053	0.0053	0.0001
			100.00%	204.88%	-104.88%	-94.75%	-10.14%
elderly							
gender	0.6171	0.6002	-0.0169	-0.0610	0.0441	0.0430	0.0011
s.e.	0.0035	0.0024	0.0045	0.0107	0.0104	0.0105	0.0021
			100.00%	361.70%	-261.70%	-255.04%	-6.65%
race	0.1887	0.1839	-0.0048	-0.0494	0.0446	0.0437	0.0009
s.e.	0.0037	0.0029	0.0048	0.0108	0.0100	0.0100	0.0001
			100.00%	1029.06%	-929.06%	-909.65%	-19.41%
By education							
illiterate							
gender	0.6419	0.6139	-0.0280	-0.1177	0.0897	0.0815	0.0082
s.e.	0.0030	0.0041	0.0048	0.0220	0.0221	0.0224	0.0016
			100.00%	420.45%	-320.45%	-291.08%	-29.37%
race	0.0767	0.0939	0.0172	-0.0001 n.s.	0.0172 n.s.	0.0172 n.s.	0.0001 n.s.
s.e.	0.0034	0.0047	0.0059	0.0141	0.0137	0.0137	0.0001
			100.00%	-0.34%	100.34%	100.01%	0.32%
compulsory school							
gender	0.6295	0.6275	-0.0020 n.s.	-0.0314	0.0294	0.0249	0.0045
s.e.	0.0014	0.0012	0.0018	0.0067	0.0064	0.0065	0.0006
			100.00%	1577.01%	-1477.01%	-1249.55%	-227.45%
race	0.1400	0.1283	-0.0117	-0.0144 n.s.	0.0027 n.s.	0.0018 n.s.	0.0009
s.e.	0.0017	0.0015	0.0023	0.0088	0.0085	0.0085	0.0001
			100.00%	123.03%	-23.03%	-15.31%	-7.72%
more than compulsory school							
gender	0.4438	0.3810	-0.0628	-0.0869	0.0241	0.0274	-0.0033
s.e.	0.0041	0.0028	0.0050	0.0123	0.0113	0.0113	0.0007
			100.00%	138.37%	-38.37%	-43.60%	5.24%
race	0.1747	0.2178	0.0432	-0.0381	0.0812	0.0897	-0.0084
s.e.	0.0051	0.0032	0.0058	0.0110	0.0106	0.0108	0.0011
			100.00%	-88.24%	188.24%	207.74%	-19.50%

60 Gradin, del Río and Alonso-Villar (2011) suggest an alternative way of exploring the determinants of occupational segregation by exploiting variation across states, rather than variation over time. To do that, the portion of segregation explained by the specific distribution of characteristics in a specific state (so-called conditional segregation) is separated from unexplained segregation using a propensity matching score technique. While this approach is not adopted here, it is indicative of the potential for further research into the determinants of occupational segregation (see also Gradin, 2010, and Alonso-Villar, del Río and Gradín, 2010).

By main economic sectors							
primary sector							
gender	0.2419	0.0818	-0.1601	-0.0343 n.s.	-0.1258	-0.1297	0.0039
s.e.	0.0050	0.0056	0.0074	0.0336	0.0337	0.0341	0.0015
			100.00%	21.41%	78.59%	81.03%	-2.44%
race	0.1429	0.1666	0.0238	-0.0027 n.s.	0.0265 n.s.	0.0262 n.s.	0.0003
s.e.	0.0036	0.0036	0.0052	0.0344	0.0340	0.0340	0.0001
			100.00%	-11.48%	111.48%	110.27%	1.21%
secondary sector							
gender	0.5156	0.6542	0.1386	-0.0145 n.s.	0.1531	0.1526	0.0005
s.e.	0.0023	0.0018	0.0030	0.0086	0.0084	0.0085	0.0002
			100.00%	-10.47%	110.47%	110.08%	0.39%
race	0.1779	0.1911	0.0132	-0.0005 n.s.	0.0137 n.s.	0.0124 n.s.	0.0012
s.e.	0.0024	0.0024	0.0033	0.0113	0.0108	0.0108	0.0002
			100.00%	-3.42%	103.42%	94.27%	9.14%
tertiary sector							
gender	0.6291	0.4792	-0.1499	-0.0954	-0.0545	-0.0513	-0.0032
s.e.	0.0016	0.0012	0.0021	0.0065	0.0065	0.0064	0.0003
			100.00%	63.63%	36.37%	34.22%	2.15%
race	0.1977	0.1868	-0.0109	-0.0377	0.0268	0.0263	0.0006
s.e.	0.0020	0.0013	0.0024	0.0078	0.0077	0.0077	0.0001
			100.00%	346.79%	-246.79%	-241.53%	-5.26%
By rural and urban areas							
rural areas							
gender	0.6222	0.6171	-0.0051	-0.0648	0.0597	0.0842	-0.0244
s.e.	0.0033	0.0027	0.0042	0.0174	0.0172	0.0166	0.0019
			100.00%	1275.74%	-1175.74%	-1656.87%	481.13%
race	0.1429	0.1532	0.0103	-0.0020 n.s.	0.0123 n.s.	0.0126 n.s.	-0.0003 n.s.
s.e.	0.0033	0.0030	0.0047	0.0178	0.0172	0.0173	0.0003
			100.00%	-19.40%	119.40%	122.34%	-2.93%
urban areas							
gender	0.5807	0.5475	-0.0332	-0.0378	0.0046 n.s.	0.0003 n.s.	0.0043
s.e.	0.0014	0.0010	0.0017	0.0034	0.0032	0.0032	0.0007
			100.00%	113.81%	-13.81%	-0.82%	-12.99%
race	0.2021	0.1918	-0.0104	-0.0285	0.0181	0.0172	0.0010
s.e.	0.0016	0.0012	0.0020	0.0044	0.0040	0.0040	0.0001
			100.00%	274.91%	-174.91%	-165.41%	-9.50%

Source: Author's computations using PNAD 1987,1988, 1989, 1990, 1992 and 2002, 2003, 2004, 2005, 2006.

Note: the initial period comprises 1987-1988-1989-1990-1992 and the final period comprises 2002-2003-2004-2005-2006. Standard errors bootstrapped with 500 replications. n.s. indicates those components that are not statistically significant at 5%.

5.5 Comparison with International Patterns of Segregation

The new occupational re-classification introduced in this study not only permits us to study the evolution of segregation over a protracted period of time but also permits an international comparison with studies that have relied on classifications similar to, or the same as, the international ISCO-08. This has not previously been possible when working with Brazilian data. We thus compare the estimates of gender and racial segregation reported here with some existing empirical studies listed in table 5.

Before making these comparisons, several issues are worth noting. First, there is limited work in this field for developing countries. Second, none of the existing work for either developing or developed economies exploits the Shapley decomposition. As such, we get no sense from the existing literature of what drives changes in segregation, with the exception of the initial paper by Deutsch, Flueckiger and Silber (2009), which explores the case of Switzerland. Our findings thus represent an important contribution in advancing our understanding of occupational segregation along both dimensions. Finally, a degree of caution is also needed in making international comparisons. Because of the difficulties in harmonizing the occupational codes over time with the international classification, we have a total of 83 occupational codes at 3-digit level, which is considerably less than the official international classification, which includes roughly 220 codes at the 3-digit level. Because the segregation measures are dependent on the number of occupations, we need to consider the number of occupations employed, rather than simply the level of disaggregation (3-digit level), when making comparisons.

As an initial reference point we compare our findings to those of de Oliveira (2001), who has conducted the only similar study of Brazilian occupational segregation, but over a more limited period than that used in this study. We find that our estimates for the Duncan index, during the corresponding years from 1987-1999, lie within the results reported by de Oliveira (2001) using 3-digit and 2-digit level classifications respectively. This is true both in term of the magnitude of segregation, and in terms of the extent of decline over time. This is as we would expect, given that our 3-digit classification has 80 codes, and thus is between the number of codes in these two sets of results.

Having confirmed that our basic results are in comport with those reported elsewhere for Brazil, we can compare these Brazilian trends with those found elsewhere using similar occupational classification. Isaza-Castro and Reilly (2011) have explored the evolution of gender segregation from 1986 to 2004 in Colombia, which is another large South American country. We when compare these results we find that Brazilian gender segregation is higher than in Colombian and has also decreased more slowly over the same time period. Deutsch et al (2005) have also explored the evolution of gender segregation in other Central or South American countries - Costa Rica, Ecuador and Uruguay – over a slightly earlier period from 1989-1997. During this period there was only a negligible decline in Brazilian gender segregation, whereas Deutsch et al (2005) report a sharp decline in gender segregation over the same period across their sample of countries. Anker, Melkas and Korten (2003) present similar results for Costa Rica, Ecuador and Uruguay while also comparing several OECD countries – France, the U.S. and Spain. They find that these countries experienced surprisingly similar declines in gender segregation over the 1990s, with an average decline in the Duncan index of approximately -0.033. By contrast, Brazil experienced a much more modest decline in the Duncan Index of only -0.006 during the 1990s. That said, Brazil experienced a much more rapid decline in gender segregation beginning at the end of the 1990s, such that from the 1990 to 2006 the Duncan Index declined by -0.032. It is thus unclear whether Brazil has simply experienced less decline in gender segregation than many of its neighbours, or whether these changes simply began somewhat later, perhaps reflecting the speed of economic reform in the early 1990s.

Irrespective of whether there has been less change, or simply change that has come later, gender segregation in Brazil appears to have remained consistently higher than in many developing and developed countries. Charles and Grusky (2005) report estimates of the Duncan index for many developed countries in 1986 and they are consistently lower than our estimates for Brazil in 1987 (0.605 at 3-digit level). The same pattern is confirmed for estimates during the 1990s. Using either the Duncan Index or the Gini Index Brazilian gender segregation during the 1990s is higher than in all of the developed countries considered in a cross-country study by Deutsch and Silber (2005).

If we turn attention to racial segregation there are significantly fewer existing studies, and we focus on a comparison with the U.S. case using results reported by King (1992) and Hirsch and MacPherson (2003). Both studies report levels of racial segregation for men and women separately, and report levels of racial segregation significantly higher than those reported here for Brazil, though with the gap diminishing over time. Thus, in 1988 Hirsch and MacPherson (2003) find Duncan Indexes by race for men and women, respectively, of 0.298 and 0.284. By contrast the figures for Brazil were 0.182 and 0.226 in 1988. By 1998 Hirsch and MacPherson (2003) report that these figures have declined to 0.284 and 0.241 respectively, while in Brazil we see an increase in racial segregation for men, from 0.182 to 0.198, and a decrease for women that still lags behind the rate of decline observed in the U.S. While lower levels of racial segregation than those observed in the U.S. are encouraging, the comparison again highlights the disconcerting persistence of such segregation in Brazil over time.

Table 5: Comparable International Results

Gender segregation studies						
de Oliveira (2001)						
Country	**Index**	**Year**	**Value**	**No. of occ.**	**Change 1987-1999**	**Comparison with findings reported here for Brazil**
Brazil	Duncan index	1981	0.666	3-digit (257)	-0.032	Duncan index equal to 0.605 in 1987 and to 0.591 in 1999. Change over the period of -0.014.
	Duncan index	1987	0.633	3-digit (257)		
	Duncan index	1999	0.601	3-digit (257)		
	Duncan index	1981	0.578	2-digit (59)	-0.011	
	Duncan index	1987	0.543	2-digit (59)		
	Duncan index	1999	0.532	2-digit (59)		
Isaza-Castro and Reilly (2011)						
Country	**Index**	**Year**	**Value**	**No. of occ.**	**Change 1986-2004**	**Comparison with findings reported here for Brazil**
Colombia	Duncan index	1986	0.55	82	-0.051	Duncan index equal to 0.605 in 1987 and to 0.566 in 2004. Change over the period of -0.039.
	Duncan index	2004	0.499	82		
	Gini index	1986	0.746	82	-0.045	Gini index equal to 0.776 in 1987 and to 0.734 in 2004. Change over the period of -0.042.
	Gini index	2004	0.701	82		
	K&M index	1986	0.259	82	-0.015	K&M index equal to 0.277 in 1987 and 0.273 in 2006. Change over the period of -0.004.
	K&M index	2004	0.244	82		
Deutsch et al (2005)						
Country	**Index**	**Year**	**Value**	**No. of occ.**	**Change 1989-1997**	**Comparison with findings reported here for Brazil**
Costa Rica	Duncan index	1989	0.57	2-digit	-0.03	Duncan index equal to 0.607 in 1993 and to 0.6012 in 199. Change over the period equal to -0.006.
Costa Rica	Duncan index	1993	0.56	2-digit		
Costa Rica	Duncan index	1997	0.54	2-digit		
Ecuador	Duncan index	1989	0.58	2-digit	-0.04	
Ecuador	Duncan index	1992	0.54	2-digit		
Ecuador	Duncan index	1997	0.54	2-digit		
Uruguay	Duncan index	1989	0.56	2-digit	-0.01	
Uruguay	Duncan index	1993	0.57	2-digit		
Uruguay	Duncan index	1997	0.55	2-digit		

Country	Index	Year	Value	No. of occ.	Change	Comparison with findings reported here for Brazil
Anker, Melkas and Korten (2003)						
Country	**Index**	**Year**	**Value**	**No. of occ.**	**Change 1990-2000**	**Comparison with findings reported here for Brazil**
U.S.	Duncan index	2000	0.463	104	-0.034	Duncan index equal to 0.596 in 1990 and to 0.590 in 2001. Change over the period of -0.006.
France	Duncan index	1999	0.554	119	-0.036	
Spain	Duncan index	2000	0.528	78	-0.03	
Costa Rica	Duncan index	2001	0.526	55	-0.04	
Ecuador	Duncan index	2000	0.498	75	-0.038	
Uruguay	Duncan index	1996	0.53	71	-0.022	
Charles and Grusky (2005)						
Country	**Index**	**Year**	**Value**	**No. of occ.**		**Comparison with findings reported here for Brazil**
Switzerland	Duncan index	1986	0.399	45		Duncan index equal to 0.605 in 1986.
Sweden	Duncan index	1986	0.412	45		
U.K.	Duncan index	1986	0.444	45		
Turkey	Duncan index	1986	0.405	45		
Japan	Duncan index	1986	0.241	45		
Germany	Duncan index	1986	0.389	45		
U.S.	Duncan index	1986	0.366	45		
Greece	Duncan index	1986	0.302	45		
Deutsch and Silber (2005)						
Country	**Index**	**Year**	**Value**	**No. of occ.**		**Comparison with findings reported here for Brazil**
Switzerland	Duncan index	1997	0.55	ISCO 2-digit		Duncan index average between 1990 and 1999 equal to 0.603.
Finland	Duncan index	1990	0.54	ISCO 2-digit		
Norway	Duncan index	1990	0.51	ISCO 2-digit		
Sweden	Duncan index	1990	0.52	ISCO 2-digit		
France	Duncan index	1997	0.52	ISCO 2-digit		
Hungary	Duncan index	1993	0.51	ISCO 2-digit		
Luxemburg	Duncan index	1992	0.52	ISCO 2-digit		
Poland	Duncan index	1994	0.54	ISCO 2-digit		
Spain	Duncan index	1993	0.44	ISCO 2-digit		
U.K.	Duncan index	1989	0.48	ISCO 2-digit		
Switzerland	Gini index	1997	0.69	ISCO 2-digit		Gini index average between 1990 and 1999 equal to 0.776.
Finland	Gini index	1990	0.66	ISCO 2-digit		
Norway	Gini index	1990	0.68	ISCO 2-digit		
Sweden	Gini index	1990	0.66	ISCO 2-digit		
France	Gini index	1997	0.63	ISCO 2-digit		
Hungary	Gini index	1993	0.64	ISCO 2-digit		
Luxemburg	Gini index	1992	0.68	ISCO 2-digit		
Poland	Gini index	1994	0.67	ISCO 2-digit		
Spain	Gini index	1993	0.59	ISCO 2-digit		
U.K.	Gini index	1989	0.68	ISCO 2-digit		
Racial segregation studies						
King (1992) - both gender and racial segregation						
Country	**Index**	**Year**	**Value**	**No. of occ.**		**Comparison with findings reported here for Brazil**
U.S.	Duncan index (Black and white women)	1988	0.293	159		Duncan index by race for women is equal to 0.23 in 1988 while for men it is equal to 0.18.
U.S.	Duncan index (Black and white men)	1988	0.293	159		
U.S.	Duncan index (Black women and men)	1988	0.609	159		Duncan index by gender for non-whites is equal to 0.63 in 1988 while for whites it is equal to 0.58.
U.S.	Duncan index (White women and men)	1988	0.6	159		
Hirsch and Macpherson (2003) - only racial segregation						
Country	**Index**	**Year**	**Value**	**No. of occ.**	**Change 1988-1998**	**Comparison with findings reported here for Brazil**
U.S.	Duncan index (men)	1973	0.384	Census occ. codes	-0.014	Duncan index by race for men in 1988 equal to 0.182 and to 0.198 in 1998. Change over time is equal to +0.016.
U.S.	Duncan index (men)	1988	0.298	Census occ. codes		
U.S.	Duncan index (men)	1998	0.284	Census occ. codes		
U.S.	Duncan index (women)	1973	0.374	Census occ. codes	-0.043	Duncan index by race for women in 1988 equal to 0.226 and to 0.211 in 1998. Change over time is equal to -0.015.
U.S.	Duncan index (women)	1988	0.284	Census occ. codes		
U.S.	Duncan index (women)	1998	0.241	Census occ. codes		

Source: Author’s compilation.

5.6 Conclusions

In this chapter we have investigated the magnitude and evolution over time of gender and racial occupational segregation using several well-known indices of segregation. We additionally explore these trends in segregation separately for the formal, informal and self-employed sectors, following the earlier analysis of occupational structure. The analysis has been conducted on a national basis but also disaggregated across key characteristics of the labour force to identify specific demographic, educational, sectoral and spatial patterns. Finally, we investigate possible determinants of occupational segregation by applying a decomposition technique developed by Deutsch, Flueckiger and Silber (2009).

We conclude by summarizing our main findings. We find that gender segregation is significantly greater than racial segregation. However, we also find that gender segregation has fallen more rapidly over the last two decades than racial segregation, which has been surprisingly stable despite increased participation of non-white workers into the labour force. Whereas both groups have experienced declining segregation in the formal sector, gender segregation has also declined in the informal sector, but racial segregation has experienced only a negligible decline in this sector.

In examining these patterns of segregation over time we have been able to disaggregate outcomes by key individual, geographic and sectoral characteristics, yielding some additional insights. In terms of demographic patterns, the overall decline in gender segregation is clearly reflected in higher levels of segregation among older workers. More surprisingly, we find that racial segregation is higher among younger workers, which is an even starker indicator of the relative persistence of racial segregation, and the apparent barriers faced by less experienced non-white workers. In regard to patterns disaggregated by educational attainment, gender segregation is lower among the more highly educated, while racial segregation is higher among more educated groups, which is surprising and potentially indicative of potent employment barriers confronting non-white groups. Interestingly, although racial segregation is higher for more educated individuals, both types of segregation have declined more rapidly over time among the better educated, while for the less educated the level of segregation remains relatively stable. The overall finding from these disaggregated results appears to be that while gender segregation is much higher in absolute terms, various features of racial segregation are persistent with the trends are significantly less promising for the less educated.

Moving to sectoral patterns, we find that the secondary sector has witnessed an increase in occupational segregation, particularly by gender, while the opposite is the case in the primary sector, where gender segregation has declined significantly. The narrative in the tertiary sector for both types of segregation is of declining gradually and steadly over time. Turning to spatial patterns, at the regional level we see a relatively homogeneous decline in segregation across all regions, though with two noteworthy features. First, we observe a particularly dramatic decline in gender based segregation in the Central-West region, where the heavy presence of government offices might plausibly have an impact on the effectiveness of anti-discrimination policies. Second, although race-based segregation is declining everywhere, it remains much higher in the South-East and South regions, where socio-economic disparities between non-whites and whites are deeper. Finally, we find major differences between rural and urban areas, as gender segregation is comparatively higher in rural areas, particularly outside the primary sector, while racial segregation is comparatively higher in urban areas.

The most novel set of results emerge from the application of the Shapley decomposition proposed by Deutsch, Flueckiger and Silber (2009) to shed some light on the forces driving declining segregation. In simplified terms, the Shapley decomposition aims to distinguish between two types of change: changes in the "internal structure" of individual occupations, representing a more equitable participation of different groups within individual occupations, and changes in the "margins", which capture changes in the measure of segregation resulting from changes in occupational structure or changes in the share of different groups in the overall labour force. This is a hugely important distinction and our results suggest that the decline in both gender and race based segregation is primarily the result of the more homogenous representation of women and non-whites within occupations. In fact, contrary to a commentary that attributes declining measures of occupational segregation to the rapid entry into the labour force of previously underrepresented groups, our results reveal that changes in occupational structure and the entry of new groups into the labour force have contributed to increasing segregation, with many new entrants to the labour force joining traditionally more segregated occupations. Our aggregate results are driven by the fact that the increase in segregation generated by these latter two trends is offset by the general improvement in composition within individual occupations. This represents a more ‘real’ decline in segregation, and is thus a very encouraging finding from a social perspective.

This narrative remains incomplete, however, and as such we conclude with an important suggestion for future research aimed at understanding the determinants of occupational segregation. As has already been noted, a particularly interesting question in this field is the role of ADL in reducing segregation. Several aspects of our findings are suggestive of such a connection, most notably the fact that the past two decades have seen a more rapid decline in gender segregation in the formal sector than in the non-formal sectors. This is what we would expect if ADL is effective.

We gain additional insight into this process from the findings of the Shapley decomposition. The negative contribution of changes in occupational structure features in the non-formal labour markets only, suggesting that highly segregated occupations have had expanded primarily in the informal labour market, whereas such segregation has been more constrained in the formal sector.

However, while the findings across the formal and informal sectors are consistent with an impact of ADL on segregation, establishing clear causation is a significantly more challenging. As such, this remains an area in which there are significant avenues for future research building on the suggestive results presented here. The most intriguing possibility, noted earlier, is that it may be possible to adopt an econometric strategy that exploits differences over time and across states in order to conduct an ad hoc quasi experiment looking at the role of state level differences in either the passage of ADL or the enforcement of federal anti-discrimination laws. This follows research conducted across U.S. States by Neumark and Stock (2006) that employs this broad approach. However, such research requires significant additional data on variation in laws, enforcement and governance over time and across states, and such data are currently unavailable, to the best of author’s knowledge. Given the scope of the current chapter, this must remain for now part of the agenda for future research.

Appendix to Chapter 5

Table A1: Test of mean differences across sector

Duncan index - gender	1987	1992	1997	2002	2006	**Duncan index - race**	1987	1992	1997	2002	2006
Formal vs Informal	-23.688	-27.956	-26.329	-24.061	-29.268	Formal vs Informal	-0.580$^{\text{n.s.}}$	5.649	4.981	-1.316$^{\text{n.s.}}$	-2.412
Formal vs Self-Empl.	-15.109	-7.671	-5.994	-17.677	-20.439	Formal vs Self-Empl.	8.139	5.443	5.203	3.430	4.707
Informal vs Self-Empl.	7.559	18.254	18.217	4.631	6.827	Informal vs Self-Empl.	7.920	0.048$^{\text{n.s.}}$	0.518$^{\text{n.s.}}$	4.267	6.283

Moir & Selby-Smith index - gender	1987	1992	1997	2002	2006	**Moir & Selby-Smith index – race**	1987	1992	1997	2002	2006
Formal vs Informal	-7.997	-15.094	-15.587	-14.237	-14.843	Formal vs Informal	6.862	13.985	13.711	7.546	6.697
Formal vs Self-Empl.	-16.131	-16.610	-21.489	-31.625	-33.221	Formal vs Self-Empl.	11.722	8.853	7.943	7.477	8.728
Informal vs Self-Empl.	-8.294	-2.735	-6.829	-17.565	-19.019	Informal vs Self-Empl.	5.498	-3.819	-3.770	0.889$^{\text{n.s.}}$	2.459

Karmel & Maclachlan index - gender	1987	1992	1997	2002	2006	**Karmel & Maclachlan index - race**	1987	1992	1997	2002	2006
Formal vs Informal	-19.498	-11.791	-15.621	-18.038	-23.908	Formal vs Informal	-1.111$^{\text{n.s.}}$	5.448	4.899	-1.135$^{\text{n.s.}}$	-0.947
Formal vs Self-Empl.	-6.991	1.487$^{\text{n.s.}}$	6.110	-2.519	-5.961	Formal vs Self-Empl.	7.675	4.783	4.748	2.979	5.013
Informal vs Self-Empl.	8.374	10.894	17.366	12.412	13.408	Informal vs Self-Empl.	7.917	-0.401$^{\text{n.s.}}$	0.238$^{\text{n.s.}}$	3.654	5.277

Gini segregation index – gender	1987	1992	1997	2002	2006	**Gini segregation index – race**	1987	1992	1997	2002	2006
Formal vs Informal	-13.031	-16.286	-19.371	-26.223	-32.868	Formal vs Informal	1.262$^{\text{n.s.}}$	6.929	5.715	1.069$^{\text{n.s.}}$	-1.775
Formal vs Self-Empl.	-16.732	-7.177	-7.590	-11.459	-18.163	Formal vs Self-Empl.	7.135	4.533	5.249	3.531	3.828
Informal vs Self-Empl.	-0.634$^{\text{n.s.}}$	5.870	11.788	11.368	11.652	Informal vs Self-Empl.	5.535	-2.280	-0.421$^{\text{n.s.}}$	2.153	4.848

Marginal matching index – gender	1987	1992	1997	2002	2006	**Marginal matching index - race**	1987	1992	1997	2002	2006
Formal vs Informal	-9.615	-7.798	-8.713	-8.153	-7.273	Formal vs Informal	6.912	10.335	11.905	3.076	1.342
Formal vs Self-Empl.	-6.149	-2.050	-1.089$^{\text{n.s.}}$	3.738	-1.408	Formal vs Self-Empl.	6.915	6.647	2.465	0.747$^{\text{n.s.}}$	2.106
Informal vs Self-Empl.	1.594$^{\text{n.s.}}$	8.733	11.049	12.295	4.292	Informal vs Self-Empl.	0.481$^{\text{n.s.}}$	-6.839	-2.814	-2.749	-1.107$^{\text{n.s.}}$

Source: Author's computations using PNAD 1987 – 1992 – 1997 – 2002 – 2006.
Note: z-tests reported in the table; $^{\text{n.s.}}$ not statistically significant; all other z-tests are statistically significant at 5%.

Table A2: Test of mean differences over time for each sector

	All labour market						Formal sector						Informal sector						Self-employed sector				
Duncan index - gender																							
	1987	1992	1997	2002	2006		1987	1992	1997	2002	2006		1987	1992	1997	2002	2006		1987	1992	1997	2002	2006
1987						1987						1987						1987					
1992	1.838 n.s.					1992	3.846					1992	6.357					1992	-4.257				
1997	-1.225 n.s.	-3.142				1997	3.075	-0.926 n.s.				1997	3.078	-3.373				1997	-6.881	-2.575			
2002	-10.780	-12.585	-10.129			2002	-4.646	-9.172	-8.499			2002	-8.247	-15.129	-11.727			2002	-3.880	0.572 n.s.	3.279		
2006	-12.861	-14.666	-12.344	-2.021		2006	-8.342	-13.221	-12.671	-4.006		2006	-7.202	-14.016	-10.629	1.055 n.s.		2006	-4.998	-0.547 n.s.	2.145	-1.176 n.s.	
Duncan index - race																							
	1987	1992	1997	2002	2006		1987	1992	1997	2002	2006		1987	1992	1997	2002	2006		1987	1992	1997	2002	2006
1987						1987						1987						1987					
1992	0.205 n.s.					1992	-0.188 n.s.					1992	-5.820					1992	2.391				
1997	-109.677	-0.262 n.s.				1997	0.330 n.s.	0.520 n.s.				1997	-4.616	1.384 n.s.				1997	3.210	0.819 n.s.			
2002	-111.811	-1.164 n.s.	-0.928 n.s.			2002	-3.211	-2.966	-3.668			2002	-1.962	4.100	2.793			2002	2.544	0.007 n.s.	-0.867 n.s.		
2006	-121.453	-2.242	-2.030	-1.053 n.s.		2006	-2.975	-2.722	-3.448	0.394 n.s.		2006	-0.904 n.s.	5.335	4.032	1.165 n.s.		2006	1.984	-0.597 n.s.	-1.487 n.s.	-0.650 n.s.	
Moir and Selby-Smith index - gender																							
	1987	1992	1997	2002	2006		1987	1992	1997	2002	2006		1987	1992	1997	2002	2006		1987	1992	1997	2002	2006
1987						1987						1987						1987					
1992	-3.109					1992	-3.538					1992	3.378					1992	-1.906 n.s.				
1997	-8.865	-5.496				1997	-8.247	-4.960				1997	-1.059 n.s.	-4.495				1997	-2.528	-0.493 n.s.			
2002	-21.042	-17.509	-12.916			2002	-17.615	-14.744	-9.451			2002	-11.643	-14.967	-10.852			2002	-2.729	-0.627 n.s.	-0.128 n.s.		
2006	-25.996	-22.307	-18.036	-4.765		2006	-20.683	-18.048	-12.864	-3.931		2006	-15.005	-18.341	-14.307	-3.224		2006	-5.708	-3.355	-3.011	-2.977	
Moir and Selby-Smith index - race																							
	1987	1992	1997	2002	2006		1987	1992	1997	2002	2006		1987	1992	1997	2002	2006		1987	1992	1997	2002	2006
1987						1987						1987						1987					
1992	-0.239 n.s.					1992	-0.284 n.s.					1992	-7.501					1992	2.471				
1997	-0.951 n.s.	-0.722 n.s.				1997	-0.456 n.s.	-0.150 n.s.				1997	-6.228	1.614 n.s.				1997	3.523	1.091 n.s.			
2002	-3.831	-3.672	-3.042			2002	-4.971	-4.479	-4.633			2002	-4.242	4.055	2.475			2002	1.423 n.s.	-1.265 n.s.	-2.433		
2006	-8.270	-8.246	-7.781	-4.647		2006	-7.373	-6.733	-7.108	-2.271		2006	-4.939	3.272	1.666	-0.836 n.s.		2006	-0.300 n.s.	-3.163	-4.371	-2.026	
Karmel & Maclachlan index - gender																							
	1987	1992	1997	2002	2006		1987	1992	1997	2002	2006		1987	1992	1997	2002	2006		1987	1992	1997	2002	2006
1987						1987						1987						1987					
1992	3.557					1992	7.026					1992	2.789					1992	-1.663 n.s.				
1997	3.058	-0.548 n.s.				1997	13.178	0.337 n.s.				1997	2.079	-0.779 n.s.				1997	-4.005	-1.416 n.s.			
2002	-1.165 n.s.	-5.291	-4.769			2002	9.203	-1.880	-4.242			2002	-1.759 n.s.	-5.246	-4.446			2002	-0.414 n.s.	1.570 n.s.	4.786		
2006	-1.228 n.s.	-5.265	-4.750	-0.125 n.s.		2006	6.377	-3.489	-7.280	-3.028		2006	-0.891 n.s.	-4.480	-3.631	1.314 n.s.		2006	0.246 n.s.	2.099	5.638	0.921 n.s.	
Karmel & Maclachlan index – race																							
	1987	1992	1997	2002	2006		1987	1992	1997	2002	2006		1987	1992	1997	2002	2006		1987	1992	1997	2002	2006
1987						1987						1987						1987					
1992	0.209 n.s.					1992	-0.312 n.s.					1992	-6.263					1992	2.419				
1997	0.045 n.s.	-0.170 n.s.				1997	0.601 n.s.	0.894 n.s.				1997	-4.879	1.590 n.s.				1997	3.213	0.822 n.s.			
2002	-0.713 n.s.	-0.944 n.s.	-0.783 n.s.			2002	-2.846	-2.416	-3.491			2002	-2.246	4.212	2.717			2002	2.665	0.154 n.s.	-0.704 n.s.		
2006	-1.997	-2.264	-2.119	-1.326 n.s.		2006	-2.023	-1.605 n.s.	-2.690	0.986 n.s.		2006	-1.832 n.s.	4.842	3.322	0.506 n.s.		2006	2.000	-0.607 n.s.	-1.490 n.s.	-0.804 n.s.	

Gini segregation index - gender

	1987	1992	1997	2002	2006
1987					
1992	2.287				
1997	-0.293 n.s.	-2.288			
2002	-8.675	-14.855	-6.985		
2006	-10.757	-17.819	-8.717	-2.945	

	1987	1992	1997	2002	2006
1987					
1992	3.746				
1997	1.997	-2.471			
2002	-1.931 n.s.	-7.499	-5.597		
2006	-5.731	-12.373	-11.090	-5.579	

	1987	1992	1997	2002	2006
1987					
1992	3.226				
1997	2.526	-0.869			
2002	-1.253 n.s.	-5.706	-4.942		
2006	-1.306 n.s.	-5.803	-5.041	-0.076 n.s.	

	1987	1992	1997	2002	2006
1987					
1992	-3.737				
1997	-10.750	-3.813			
2002	-12.527	-4.881	-1.444 n.s.		
2006	-12.671	-4.663	-1.031 n.s.	0.498 n.s.	

Gini segregation index - race

	1987	1992	1997	2002	2006
1987					
1992	0.497 n.s.				
1997	-0.721 n.s.	-1.229 n.s.			
2002	-0.491 n.s.	-1.031 n.s.	0.277 n.s.		
2006	-0.731 n.s.	-1.286 n.s.	0.053 n.s.	-0.247 n.s.	

	1987	1992	1997	2002	2006
1987					
1992	0.347 n.s.				
1997	-0.023 n.s.	-0.389 n.s.			
2002	-1.693	-2.077	-1.777		
2006	-2.232	-2.635	-2.364	-0.513 n.s.	

	1987	1992	1997	2002	2006
1987					
1992	-4.974				
1997	-3.859	1.283 n.s.			
2002	-1.058 n.s.	4.035	2.862		
2006	0.933 n.s.	6.151	5.035	2.075	

	1987	1992	1997	2002	2006
1987					
1992	2.923				
1997	2.362	-0.652 n.s.			
2002	2.647	-0.365 n.s.	0.297 n.s.		
2006	2.433	-0.700 n.s.	-0.020 n.s.	-0.331 n.s.	

Marginal matching index - gender

	1987	1992	1997	2002	2006
1987					
1992	0.222 n.s.				
1997	-0.910 n.s.	-1.937 n.s.			
2002	-1.427 n.s.	-3.055	-0.922 n.s.		
2006	-2.155	-3.346	-1.957 n.s.	-1.444 n.s.	

	1987	1992	1997	2002	2006
1987					
1992	4.416				
1997	3.480	-1.111 n.s.			
2002	2.253	-3.046	-1.790		
2006	3.068	-2.475	-1.087 n.s.	1.010 n.s.	

	1987	1992	1997	2002	2006
1987					
1992	1.758 n.s.				
1997	1.512 n.s.	-0.248 n.s.			
2002	-4.086	-5.817	-5.578		
2006	-5.685	-7.438	-7.196	-1.538 n.s.	

	1987	1992	1997	2002	2006
1987					
1992	-1.677 n.s.				
1997	-3.276	-3.110			
2002	-7.987	-10.329	-7.619		
2006	-3.642	-3.425	-0.965 n.s.	5.731	

Marginal matching index – race

	1987	1992	1997	2002	2006
1987					
1992	0.587 n.s.				
1997	-0.066 n.s.	-0.903 n.s.			
2002	0.915 n.s.	0.621 n.s.	1.451 n.s.		
2006	0.840 n.s.	0.458 n.s.	1.331 n.s.	-0.234 n.s.	

	1987	1992	1997	2002	2006
1987					
1992	-0.734 n.s.				
1997	0.182 n.s.	0.832 n.s.			
2002	-0.964 n.s.	0.126 n.s.	-1.097 n.s.		
2006	-0.815 n.s.	0.152 n.s.	-0.944 n.s.	0.046 n.s.	

	1987	1992	1997	2002	2006
1987					
1992	-2.315				
1997	-2.843	-0.687 n.s.			
2002	-0.987 n.s.	-0.124 n.s.	0.119 n.s.		
2006	-0.279 n.s.	0.182 n.s.	0.309 n.s.	0.221 n.s.	

	1987	1992	1997	2002	2006
1987					
1992	3.188				
1997	1.321 n.s.	-0.345 n.s.			
2002	4.622	3.079	1.727 n.s.		
2006	4.724	3.344	1.478 n.s.	-0.629 n.s.	

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

Note: z-tests reported in the table; [a] not statistically significant; all other z-tests are statistically significant at 5%.

Table A3: Shapley decomposition of changes in Karmel and Maclachlan index over time across sectors

	I_p	I_v	ΔI	$C_{\Delta is}$	$C_{\Delta m}$ **(1) + (2)**	$C_{\Delta h}$ **(1)**	$C_{\Delta t}$ **(2)**
All labour market							
gender	0.2785	0.2725	-0.0060	-0.0163	0.0103	-0.0038	0.0141
s.e.	0.0005	0.0004	0.0007	0.0016	0.0015	0.0015	0.0004
			100.00%	271.41%	-171.41%	65.06%	-236.47%
race	0.0989	0.0963	-0.0026	-0.0131	0.0105	0.0099	0.0005
s.e.	0.0007	0.0006	0.0008	0.0022	0.0021	0.0021	0.0000
			100.00%	502.48%	-402.48%	-381.38%	-21.10%
Formal sector							
gender	0.2540	0.2527	-0.0013	-0.0174	0.0161	-0.0050	0.0210
s.e.	0.0009	0.0007	0.0011	0.0026	0.0025	0.0026	0.0007
			100.00%	1358.70%	-1258.70%	389.60%	-1648.30%
race	0.0920	0.0877	-0.0043	-0.0047	0.0005	-0.0016	0.0020
s.e.	0.0010	0.0008	0.0013	0.0032 n.s.	0.0029 n.s.	0.0029 n.s.	0.0001
			100.00%	110.65%	-10.65%	36.67%	-47.32%
Informal sector							
gender	0.3397	0.3231	-0.0166	-0.0442	0.0277	0.0271	0.0005 n.s.
s.e.	0.0010	0.0008	0.0013	0.0051	0.0050	0.0056	0.0010
			100.00%	267.24%	-167.24%	-164.01%	-3.23%
race	0.0978	0.0883	-0.0095	-0.0294	0.0198	0.0206	-0.0008
s.e.	0.0011	0.0010	0.0014	0.0040	0.0038	0.0038	0.0001
			100.00%	308.11%	-208.11%	-216.22%	8.11%
Self-employed sector							
gender	0.2768	0.2674	-0.0094	-0.0198	0.0105 n.s.	0.0112 n.s.	-0.0007 n.s.
s.e.	0.0011	0.0010	0.0014	0.0064	0.0063	0.0063	0.0009
			100.00%	212.03%	-112.03%	-119.61%	7.58%
race	0.0647	0.0696	0.0049	-0.0133	0.0182	0.0179	0.0003
s.e.	0.0013	0.0011	0.0017	0.0044	0.0041	0.0041	0.0001
			100.00%	-269.14%	369.14%	363.93%	5.21%

Source: Author's computations using PNAD 1987,1988, 1989, 1990, 1992 and 2002, 2003, 2004, 2005, 2006.
Note: the initial period comprises 1987-1988-1989-1990-1992 and the final period comprises 2002-2003-2004-2005-2006. Standard errors bootstrapped with 500 replications. n.s. indicates those components that are not statistically significant at 5%.

Table A4: Shapley decomposition of changes in Gini index over time across sectors

	I_p	I_v	ΔI	$C_{\Delta is}$	$C_{\Delta m}$ (1) + (2)	$C_{\Delta h}$ (1)	$C_{\Delta t}$ (2)
All labour market							
gender	0.7713	0.7355	-0.0358	-0.0434	0.0076	0.0045	0.0031
s.e.	0.0009	0.0008	0.0012	0.0028	0.0026	0.0027	0.0002
			100.00%	119.84%	-19.84%	-11.32%	-8.52%
race	0.2648	0.2610	-0.0038	-0.0358	0.0320	0.0303	0.0017
s.e.	0.0015	0.0013	0.0019	0.0053	0.0050	0.0050	0.0001
			100.00%	951.65%	-851.65%	-807.16%	-44.50%
Formal sector							
gender	0.7154	0.6834	-0.0320	-0.0538	0.0218	0.0131	0.0087
s.e.	0.0016	0.0013	0.0021	0.0049	0.0047	0.0047	0.0004
			100.00%	168.17%	-68.17%	-40.86%	-27.31%
race	0.2572	0.2446	-0.0126	-0.0156	0.0030 $^{n.s.}$	0.0023 $^{n.s.}$	0.0007
s.e.	0.0023	0.0019	0.0031	0.0081	0.0074	0.0074	0.0001
			100.00%	123.55%	-23.55%	-17.87%	-5.69%
Informal sector							
gender	0.8275	0.8083	-0.0191	-0.0835	0.0644	0.0657	-0.0014
s.e.	0.0023	0.0013	0.0026	0.0102	0.0100	0.0110	0.0015
			100.00%	436.52%	-336.52%	-343.80%	7.28%
race	0.2455	0.2426	-0.0030	-0.0823	0.0793	0.0763	0.0030
s.e.	0.0028	0.0024	0.0036	0.0094	0.0089	0.0089	0.0002
			100.00%	2751.10%	-2651.10%	-2549.46%	-101.63%
Self-employed sector							
gender	0.8067	0.7473	-0.0595	-0.0636	0.0041 $^{n.s.}$	0.0041 $^{n.s.}$	0.0000 $^{n.s.}$
s.e.	0.0019	0.0017	0.0025	0.0119	0.0121	0.0120	0.0003
			100.00%	106.92%	-6.92%	-6.98%	0.05%
race	0.1890	0.2044	0.0154	-0.0255	0.0409	0.0394	0.0015
s.e.	0.0030	0.0026	0.0040	0.0102	0.0095	0.0095	0.0001
			100.00%	-165.83%	265.83%	255.96%	9.87%

Source: Author's computations using PNAD 1987,1988, 1989, 1990, 1992 and 2002, 2003, 2004, 2005, 2006.

Note: the initial period comprises 1987-1988-1989-1990-1992 and the final period comprises 2002-2003-2004-2005-2006. Standard errors bootstrapped with 500 replications. $^{n.s.}$ indicates those components that are not statistically significant at 5%.

Table A5: Shapley decomposition of changes in Karmel and Maclachlan index over time disaggregated by characteristics

	I_p	I_v	ΔI	$C_{\Delta is}$	$C_{\Delta m}$ (1) + (2)	$C_{\Delta h}$ (1)	$C_{\Delta t}$ (2)
By age							
young							
gender	0.2736	0.2655	-0.0081	-0.0155	0.0073	-0.0005 n.s.	0.0078
s.e.	0.0008	0.0007	0.0011	0.0030	0.0030	0.0029	0.0005
			100.00%	189.85%	-89.85%	6.03%	-95.89%
race	0.1009	0.1073	0.0064	-0.0102	0.0166	0.0168	-0.0002
s.e.	0.0010	0.0009	0.0014	0.0038	0.0036	0.0036	0.0001
			100.00%	-159.98%	259.98%	263.56%	-3.58%
adult							
gender	0.2981	0.2903	-0.0078	-0.0197	0.0118	-0.0032 n.s.	0.0150
s.e.	0.0009	0.0007	0.0011	0.0022	0.0021	0.0023	0.0011
			100.00%	251.15%	-151.15%	41.05%	-192.20%
race	0.0993	0.0927	-0.0066	-0.0144	0.0078	0.0066	0.0011
s.e.	0.0011	0.0008	0.0013	0.0029	0.0027	0.0027	0.0001
			100.00%	218.01%	-118.01%	-100.81%	-17.20%
elderly							
gender	0.2568	0.2796	0.0227	-0.0269	0.0496	0.0193	0.0304
s.e.	0.0017	0.0012	0.0022	0.0047	0.0047	0.0046	0.0013
			100.00%	-118.42%	218.42%	84.76%	133.66%
race	0.0942	0.0919	-0.0022	-0.0247	0.0224	0.0218	0.0006
s.e.	0.0019	0.0015	0.0024	0.0054	0.0050	0.0050	0.0001
			100.00%	1097.89%	-997.89%	-970.42%	-27.48%
By education							
illiterate							
gender	0.2626	0.2394	-0.0232	-0.0473	0.0241	0.0326	-0.0085
s.e.	0.0015	0.0020	0.0025	0.0089	0.0090	0.0089	0.0011
			100.00%	204.01%	-104.01%	-140.63%	36.62%
race	0.0329	0.0378	0.0049	0.0001 n.s.	0.0048 n.s.	0.0071 n.s.	-0.0023
s.e.	0.0015	0.0019	0.0024	0.0058	0.0056	0.0057	0.0002
			100.00%	1.15%	98.85%	146.62%	-47.77%
compulsory school							
gender	0.2840	0.2901	0.0060	-0.0143	0.0204	0.0114	0.0090
s.e.	0.0007	0.0006	0.0009	0.0031	0.0029	0.0030	0.0006
			100.00%	-236.51%	336.51%	187.73%	148.78%
race	0.0699	0.0631	-0.0069	-0.0071 n.s.	0.0003 n.s.	0.0009 n.s.	-0.0006
s.e.	0.0009	0.0007	0.0012	0.0044	0.0042	0.0042	0.0000
			100.00%	103.89%	-3.89%	-12.93%	9.03%
more than compulsory school							
gender	0.2219	0.1889	-0.0330	-0.0433	0.0103 n.s.	0.0136	-0.0033
s.e.	0.0020	0.0014	0.0025	0.0061	0.0056	0.0056	0.0004
			100.00%	131.18%	-31.18%	-41.34%	10.15%
race	0.0601	0.0952	0.0351	-0.0153	0.0504	0.0348	0.0156
s.e.	0.0018	0.0014	0.0022	0.0044	0.0043	0.0042	0.0006
			100.00%	-43.47%	143.47%	99.17%	44.29%
By main economic sectors							
primary sector							
gender	0.0521	0.0160	-0.0361	-0.0068 n.s.	-0.0293	-0.0266	-0.0027
s.e.	0.0011	0.0011	0.0015	0.0068	0.0068	0.0070	0.0005
			100.00%	18.81%	81.19%	73.77%	7.42%
race	0.0692	0.0768	0.0075	-0.0013 n.s.	0.0088 n.s.	0.0124 n.s.	-0.0035
s.e.	0.0018	0.0017	0.0025	0.0162	0.0160	0.0161	0.0003
			100.00%	-17.22%	117.22%	164.24%	-47.02%
secondary sector							
gender	0.1902	0.2442	0.0539	-0.0054 n.s.	0.0593	0.0566	0.0027
s.e.	0.0010	0.0010	0.0014	0.0032	0.0032	0.0031	0.0009
			100.00%	-10.00%	110.00%	104.94%	5.05%
race	0.0883	0.0955	0.0072	-0.0002 n.s.	0.0075 n.s.	0.0062	0.0013
s.e.	0.0012	0.0012	0.0016	0.0056	0.0054	0.0054	0.0001
			100.00%	-3.08%	103.08%	85.64%	17.44%
tertiary sector							

gender	0.3117	0.2391	-0.0726	-0.0475	-0.0252	-0.0255	0.0003
s.e.	0.0008	0.0006	0.0011	0.0032	0.0032	0.0032	0.0001
			100.00%	65.32%	34.68%	35.11%	-0.43%
race	0.0985	0.0934	-0.0051	-0.0188	0.0137	0.0131	0.0006
s.e.	0.0010	0.0007	0.0012	0.0039	0.0039	0.0039	0.0000
			100.00%	368.33%	-268.33%	-256.51%	-11.83%
By rural and urban areas							
rural areas							
gender	0.2275	0.2516	0.0241	-0.0246	0.0487	0.0327	0.0160
s.e.	0.0015	0.0014	0.0020	0.0067	0.0067	0.0064	0.0009
			100.00%	-102.06%	202.06%	135.82%	66.24%
race	0.0705	0.0735	0.0031 $^{n.s.}$	-0.0009 $^{n.s.}$	0.0040 $^{n.s.}$	0.0061 $^{n.s.}$	-0.0021
s.e.	0.0016	0.0014	0.0023	0.0086	0.0083	0.0084	0.0002
			100.00%	-29.98%	129.98%	199.17%	-69.19%
urban areas							
gender	0.2782	0.2686	-0.0097	-0.0183	0.0087	0.0001 $^{n.s.}$	0.0085
s.e.	0.0007	0.0005	0.0008	0.0016	0.0016	0.0016	0.0004
			100.00%	189.80%	-89.80%	-1.24%	-88.56%
race	0.1004	0.0959	-0.0046	-0.0142	0.0096	0.0086	0.0011
s.e.	0.0008	0.0006	0.0010	0.0022	0.0020	0.0020	0.0000
			100.00%	311.12%	-211.12%	-187.19%	-23.93%

Source: Author's computations using PNAD 1987,1988, 1989, 1990, 1992 and 2002, 2003, 2004, 2005, 2006.

Note: the initial period comprises 1987-1988-1989-1990-1992 and the final period comprises 2002-2003-2004-2005-2006. Standard errors bootstrapped with 500 replications. $^{n.s.}$ indicates those components that are not statistically significant at 5%.

Table A6: Shapley decomposition of changes in Gini index over time disaggregated by characteristics

	I_p	I_v	ΔI	$C_{\Delta is}$	$C_{\Delta m}$ (1) + (2)	$C_{\Delta h}$ (1)	$C_{\Delta t}$ (2)
By age							
young							
gender	0.7569	0.7043	-0.0526	-0.0557	0.0031 n.s.	0.0005 n.s.	0.0026
s.e.	0.0014	0.0014	0.0020	0.0061	0.0061	0.0061	0.0002
			100.00%	105.94%	-5.94%	-0.91%	-5.03%
race	0.2657	0.2829	0.0172	-0.0327	0.0499	0.0486	0.0013
s.e.	0.0024	0.0021	0.0031	0.0091	0.0086	0.0087	0.0001
			100.00%	-190.13%	290.13%	282.42%	7.72%
adult							
gender	0.7906	0.7629	-0.0277	-0.0417	0.0140	0.0075	0.0065
s.e.	0.0014	0.0011	0.0018	0.0042	0.0039	0.0040	0.0006
			100.00%	150.68%	-50.68%	-27.22%	-23.46%
race	0.2721	0.2548	-0.0173	-0.0363	0.0190	0.0167	0.0022
s.e.	0.0025	0.0019	0.0031	0.0068	0.0061	0.0061	0.0001
			100.00%	209.71%	-109.71%	-96.83%	-12.88%
elderly							
gender	0.7713	0.7668	-0.0044	-0.0522	0.0478	0.0483	-0.0005 n.s.
s.e.	0.0031	0.0021	0.0038	0.0100	0.0096	0.0095	0.0009
			100.00%	1178.29%	-1078.29%	-1090.24%	11.95%
race	0.2540	0.2492	-0.0048	-0.0620	0.0572	0.0562	0.0010
s.e.	0.0042	0.0034	0.0055	0.0127	0.0120	0.0120	0.0002
			100.00%	1302.96%	-1202.96%	-1180.92%	-22.04%
By education							
illiterate							
gender	0.7840	0.7590	-0.0250	-0.1022	0.0772	0.0697	0.0075
s.e.	0.0027	0.0035	0.0043	0.0211	0.0210	0.0211	0.0011
			100.00%	408.68%	-308.68%	-278.56%	-30.12%
race	0.1124	0.1325	0.0200	-0.0017 n.s.	0.0217 n.s.	0.0215 n.s.	0.0002 n.s.
s.e.	0.0042	0.0052	0.0067	0.0180	0.0176	0.0176	0.0001
			100.00%	-8.34%	108.34%	107.21%	1.13%
compulsory school							
gender	0.7971	0.7914	-0.0058	-0.0347	0.0290	0.0277	0.0012
s.e.	0.0011	0.0010	0.0015	0.0082	0.0080	0.0082	0.0002
			100.00%	601.11%	-501.11%	-479.51%	-21.60%
race	0.1774	0.1711	-0.0063	-0.0168	0.0105	0.0094	0.0011
s.e.	0.0020	0.0017	0.0027	0.0104 n.s.	0.0101 n.s.	0.0101 n.s.	0.0001
			100.00%	266.43%	-166.43%	-148.59%	-17.83%
more than compulsory school							
gender	0.5969	0.5435	-0.0535	-0.0951	0.0416	0.0428	-0.0012 n.s.
s.e.	0.0063	0.0028	0.0070	0.0119	0.0099	0.0099	0.0013
			100.00%	177.87%	-77.87%	-80.13%	2.26%
race	0.2598	0.3085	0.0487	-0.0599	0.1086	0.1206	-0.0120
s.e.	0.0061	0.0037	0.0069	0.0157	0.0144	0.0144	0.0012
			100.00%	-123.01%	223.01%	247.67%	-24.66%
By main economic sectors							
primary sector							
gender	0.3450	0.1401	-0.2049	-0.0423 n.s.	-0.1626	-0.1673	0.0047
s.e.	0.0056	0.0060	0.0080	0.0388	0.0382	0.0385	0.0011
			100.00%	20.65%	79.35%	81.65%	-2.30%
race	0.1658	0.1781	0.0123	-0.0052 n.s.	0.0175 n.s.	0.0176 n.s.	-0.0001 n.s.
s.e.	0.0038	0.0037	0.0054	0.0337	0.0331	0.0331	0.0002
			100.00%	-42.33%	142.33%	143.26%	-0.93%
secondary sector							
gender	0.6762	0.8246	0.1483	-0.0258	0.1741	0.1739	0.0002
s.e.	0.0020	0.0014	0.0024	0.0073	0.0071	0.0071	0.0001
			100.00%	-17.38%	117.38%	117.24%	0.14%
race	0.2424	0.2649	0.0225	0.0001	0.0224	0.0197	0.0027
s.e.	0.0027	0.0027	0.0038	0.0144 n.s.	0.0138 n.s.	0.0138 n.s.	0.0002
			100.00%	0.37%	99.63%	87.64%	11.99%

tertiary sector							
gender	0.7803	0.6491	-0.1311	-0.0856	-0.0455	-0.0378	-0.0077
s.e.	0.0014	0.0012	0.0018	0.0047	0.0047	0.0046	0.0004
			100.00%	65.30%	34.70%	28.86%	5.85%
race	0.2589	0.2454	-0.0134	-0.0431	0.0296	0.0282	0.0014
s.e.	0.0022	0.0016	0.0027	0.0091	0.0089	0.0089	0.0001
			100.00%	320.72%	-220.72%	-210.36%	-10.37%
By rural and urban areas							
rural areas							
gender	0.7301	0.7295	-0.0006 n.s.	-0.0621	0.0615	0.0832	-0.0217
s.e.	0.0041	0.0027	0.0050	0.0176	0.0177	0.0174	0.0015
			100.00%	10160.69%	-10060.69%	-13614.19%	3553.50%
race	0.1898	0.2160	0.0263	-0.0011 n.s.	0.0274 n.s.	0.0269 n.s.	0.0004
s.e.	0.0037	0.0036	0.0054	0.0218	0.0211	0.0212	0.0002
			100.00%	-4.22%	104.22%	102.63%	1.58%
urban areas							
gender	0.7630	0.7267	-0.0363	-0.0481	0.0118	0.0074	0.0044
s.e.	0.0011	0.0009	0.0014	0.0028	0.0027	0.0027	0.0002
			100.00%	132.55%	-32.55%	-20.48%	-12.07%
race	0.2673	0.2589	-0.0084	-0.0339	0.0254	0.0234	0.0020
s.e.	0.0018	0.0013	0.0023	0.0050	0.0045	0.0045	0.0001
			100.00%	402.37%	-302.37%	-278.10%	-24.27%

Source: Author's computations using PNAD 1987,1988, 1989, 1990, 1992 and 2002, 2003, 2004, 2005, 2006.

Note: the initial period comprises 1987-1988-1989-1990-1992 and the final period comprises 2002-2003-2004-2005-2006. Standard errors bootstrapped with 500 replications. n.s. indicates those components that are not statistically significant at 5%.

Chapter 6

An Analysis of Gender and Racial Wage Gaps with particular Reference to the Role of Occupational Segregation

6.1 Introduction

This chapter investigates both the magnitude and evolution of gender and racial wage gaps in the Brazilian labour market over recent decades. We employ two alternative wage decomposition techniques to isolate the primary components of these wage gaps and to take of the impact of occupational segregation on broader trends. The analysis is conducted for the overall labour market and is then enhanced by disaggregating the analysis into the formal and non-formal sectors.

In Brazil the investigation of both gender and racial gaps is of important interest, as both remain significant but have also exhibited sharply different trends over time. In the case of gender, increased female labour market participation, and the changing role of women both within and outside the family, has been reflected in changes in both the composition of the labour market and earnings patterns. In the last two decades, the rate of female participation has increased from roughly 43% in 1987 to 56% in 2006, based on our figures computed using PNAD survey data.[61] In the case of race, the importance of racial discrimination has often been neglected in the Brazilian context, but it continues to exert a persistent impact on pre- and post-labour market conditions.[62] Ultimately, the trends in gender and racial dynamics, combined with macroeconomic

[61] These figures are similar to findings from other studies and reports by international organizations. Arabsheibani, Carneiro and Henley (2003) report that the economic activity rate for women increased from 43.3% in 1988 to 54.4% in 1999. The ILO reports that labour market participation among women has been increasing in many developing countries, with the level in Brazil increasing from 57.5% in 1995 to 58.4% in 2001 (see "ILO:2003-2004, Key indicators of the Labour Market", 2003).

[62] The ideological framework of "racial democracy" developed by Freye (1933) has created the common belief that Brazil does not suffer from racial discrimination and this has lead to relative neglect among policy-makers and researchers (Bailey, 2002).

and institutional changes[63] experienced in Brazil over the last two decades, have shaped changes in the magnitude of gender and racial pay gaps as well as broader trends in occupational structure.

In order to analyse these wage gaps we employ the standard Oaxaca (1973) and Blinder (1973) decomposition as well as a variant proposed by Brown, Moon and Zoloth (1980). The Brown, Moon and Zoloth (1980) decomposition technique is significant in that it takes the role of occupational segregation into account when decomposing earnings differentials. Specifically, whereas the Oaxaca and Blinder technique only decomposes wage gaps into the endowment component and the wage structure component, the Brown, Moon and Zoloth (1980) technique further decomposes wage gaps into within and between occupation components. We can thus identify the portion of wage differentials that is due to unexplained differences in wage returns within occupations, what we call "vertical segregation", and the portion of wage differentials that is due to differences in wage rewards across occupations, which we call "horizontal segregation".

In order to provide an overview of the evolution of gender and racial wage gaps over time, while accounting for the role of occupational segregation, we employ decomposition techniques that focus on mean values. This provides a clear picture of the overall patterns. However, there has recently been fruitful research into decomposition methods that go beyond the average values (Fortin, Lemieux and Firpo, 2011) and we will correspondingly employ some of these quantile decomposition techniques, which decompose the results at different points in the wage distribution, in chapter 7.

Alongside the two decomposition methods implemented in this chapter we also control for selection bias by applying parametric correction methods by Heckman (1979) and Lee (1983). This allows us to compare the results emerging from the uncorrected decomposition results with the findings from the selection-corrected decomposition techniques. However, there are well known challenges associated with correcting for selection when employing decomposition methods. There are difficulties in findings appropriate instruments for the selection process, while there is also a more

[63] The establishment of a new Constitution in 1988 and trade reforms in the 1990s along with external negative shocks at the end of the last decade exerted a profound impact on the Brazilian labour market in terms of inter-sectoral composition, the degree of competitiveness, its flexibility and the level of workers' social protection (Bosch, Goni and Maloney, 2007). For the impact of changes in legislation, see Paes de Barros and Corseuil (2001) and for the effect of trade liberalization, see Goldberg and Pavcnik (2003).

specific literature highlighting the fact that selection correction suffers from a lack of robustness (Manski, 1989) and introduces ambiguities within the decomposition framework (Neuman and Oaxaca 2003). We furthermore find differences across the two selection correction methods that are indicative of the presence of unobservable heterogeneity attached to the selection process, as highlighted by Machado (2011). In response to these concerns, we employ two additional selection correction methods – the non-parametric imputation method as developed by Olivetti and Petrongolo (2008) and the local wage gap estimation by Machado (2011) – in order to further check the robustness of our results.

Turning to a brief summary of the results, we find that both gender and racial differentials are lower than they were at the end of the 1980s, but it is only the gender wage gaps which have contracted sharply over time. The rapid decrease in gender wage gaps is primarily attributable to changes in the wage structure, which captures unobserved factors, including discriminatory behaviour. In contrast, the small decrease in racial wage gaps is entirely the result of modest improvements in the observed characteristics of non-white individuals.

The Brown, Moon and Zoloth (1980) decomposition technique further reveals that for gender wage gaps the reduction in differences in wage structure has occurred primarily within occupations (i.e., reduced vertical segregation has been the main contributor to improvements in gender pay differentials). Horizontal segregation has played a much more modest role, and actually favouring female workers over time. Turing to racial wage gaps we observe that the small improvements in pay differentials are almost entirely explained by decreased differences in observed characteristics across occupations (i.e. the explained inter-occupational component). Overall, both horizontal and vertical segregation are smaller in magnitude than their values for gender differentials, but they are also more persistent over time.

When we apply the parametric selection methods we find that selection corrected gender wage gaps (i.e., gender wage offer gaps) are greater than observed wage gaps. This reflects the existence of positive female selection into the labour market, which inflates average female wages above what they would be in the absence of selection. By contrast, we find less clear-cut trends for racial wage gaps. The Heckman procedure yields selection corrected racial wage gaps that are smaller than observed wage gaps, owing primarily to negative selection among non-white workers, and particularly among non-white men. However, this trend is reversed for the Lee correction, which finds positive selection among non-white workers by virtue of giving greater weight to positive selection within the formal sector.

While the different methodologies thus yield different estimates of the effects of selection, this is not a major concern, because the main findings from the uncorrected decomposition analysis survive the selection process. However, we do find that the magnitudes of the selection term differentials are particularly large for gender pay gaps, which, coupled with the more general concerns associated with these selection-corrected decomposition techniques, argues for additional sensitivity checks. We thus apply the non-parametric method for selection corrections by Olivetti and Petrongolo (2008) and the local wage gap estimation by Machado (2011). Ultimately, we find broadly consistent patterns for the evolution of both gender and racial wage gaps across all of these selection correction methods, which significantly enhance our confidence in these core results.

The remainder of this chapter is structured as follows. The next section reviews the relevant theoretical and empirical literature. The third section presents a description of the data, while the fourth introduces the relevant methodologies. In the fifth and sixth sections we describe our main findings using the baseline methodologies and then again after correcting for selection bias. The seventh section provides some sensitivity analysis through selecting more restricted samples and employing alternative estimation methods. The last section offers some concluding remarks.

6.2 Literature review

Discrimination is a complex phenomenon, and economists have found it useful to distinguish between pre-entry discrimination and post-entry discrimination. Pre-entry discrimination, or pre-market discrimination, occurs before entering the labour market and can operate in many different ways. For example, it may occur within households, within communities through the transmission of valuable skills or through differential access to formal education. A symptom of pre-entry discrimination is employment discrimination, as the likelihood of being employed depends upon a individual’s group affiliation.

Post-entry discrimination, often called market-discrimination, refers to discrimination occurring after entering the labour market. There are many possible mechanisms that can result in discrimination, including, for example, prejudice, imperfect information, imperfect competition or institutional discrimination. Two widely studied symptoms of market-discrimination are wage discrimination and occupational segregation. Wage discrimination occurs when the earnings of equally productive workers systematically differ for different groups. Occupational segregation occurs when the distribution of occupations for a certain group is very different from the distribution for another group (see chapter 12 in Ehrenberg and Smith (2009) and chapters 11-12 in Laing (2011)).

In what follows a review of the relevant literature is presented. We begin by reviewing the most important theories of labour market discrimination, which highlight possible causes and implications of discriminatory behaviour. We then review key methodologies that can be used to decompose earnings differentials. Finally, we conclude by reviewing the most relevant studies of gender and racial wage discrimination within the Brazilian labour market.

6.2.1 Theories of labour market discrimination

While there is a wide variety of theories that aim to understand the causes and mechanisms of gender and racial discrimination in labour markets, we follow the Ehrenberg and Smith (2009, chapter 12) classification of theoretical models based on three different sources of discrimination: personal prejudice, statistical prejudgment and the presence of non-competitive forces.[64]

Models based on the presence of *personal prejudice* are generally defined as neoclassical non-stochastic models, and are based on Becker's (1971) "Economics of Discrimination," which introduced the theory of "employer's taste for discrimination". In these models employers have preferences for hiring certain types of worker and they attach a cost to hiring those from groups that they discriminate against. This prejudice, which may equally be vested in customers or other employees, generates unequal access to employment opportunities.

The second general source of discrimination is *statistical prejudgment* and this underpins both competitive and stochastic models of discrimination. Statistical

[64] A more complete discussion of these different theories for discrimination are also presented in Cain (1987), Darity and Mason (1998) and Altonji and Black (1999).

prejudgment refers to situations in which employers use information on group affiliation to discern individual characteristics, and the seminal work in this area was provided by Phelps (1972) and Arrow (1973).

While statistical discrimination is based on the existence of imperfect information, alternative theories of discrimination are based on the presence of imperfect competition. These non-competitive forces may include dual labour markets, the presence of a crowding phenomenon, search-related monopsony and collusive behavior. The presence of a crowding phenomenon originated from the belief that occupational segregation is caused by a deliberate crowding policy in order to keep wages low. The theory of dual labour markets, developed by Piore (1972), is based on the existence of two separate markets, a primary and a secondary sector, with different wages, labour conditions and opportunities (see also Marshall, 1974). Alternatively, the main source of restricted mobility might lie in the cost to employees of job searches: this is the case of search-related monopsony, which is based on the monopsonistic model of firm behaviour (Manning, 1996, 2003).

While the preceding theories can be used to explain a variety of forms of discrimination, Akerlof (1997) has developed a theory of social distance that specifically aims to explain occupational segregation. He argues that occupational segregation can be explained by externalities that are factors in processes of social decision. Externalities are important either when people try to distance themselves (the status model) or to move themselves closer (the conformity model). He then shows that this model can lead to class stability.

There are also a range of more specific theories that have sought to specifically explain either gender or racial/ethnic occupational segregation. From a feminist/gender perspective, Anker (1997) proposed that wage gaps and occupational segregation grounded in lower levels of human capital accumulation for women may be linked to the broader subordinate position of women in society and within households. Closely related devaluation theories suggest that it is a matter of “status composition”: occupations filled with women are perceived to be of lower value than those occupations undertaken by male workers (see Cotter, Hermsen and Vanneman, 2003). Turning to racial discrimination, Fryer (2010) proposed the importance of peer dynamic and identity models. Certain peer dynamics may create a disincentive to invest in particular behaviours if they are rejected by a social peer group (most famously in the case of “acting white”). In the case of identity models, individuals engage in a selffulfilling vicious circle where the shift in preferences avoids achievement (see also Akerlof and Kranton, 2000).

Finally, while all of the preceding theories implicitly treat wage discrimination and occupational segregation separately, Baldwin, Butler and Johnson (2001) are unique in developing a theory that aims to explain wage discrimination while explicitly incorporating the impact of occupational segregation. They extend Becker's model by adding occupational segregation as a determinant of wage gaps through integrating a sorting function into the traditional earnings equation. The occupational sorting function generates an occupational hierarchy based on the number of women in lower occupations, male wages in lower occupations and the male distaste for female management, with the last factor based on the intuition that men dislike female management more than the presence of females in their workplace.

There are thus a wide range of theories that seek to account for wage and occupational discrimination, and while it is not the primary objective of this review, it is worth including a brief discussion of debates about the limits of alternative theories. Over the medium and long run competitive forces should eliminate discrimination induced by an employer's prejudice. Co-workers' and consumers' discrimination should result in completely segregated occupations, but not wage discrimination. In particular, market theories can explain the presence of industrial segregation but not pure occupational segregation (Arrow, 1998). If statistical discrimination does not derive from prejudice but from imperfect information then the employer should acquire more accurate information about the individuals, and not simply their group affiliation, as it becomes available: the immutability of imperfect information is rarely satisfactory (Darity and Mason, 1998). Competitive models thus struggle to adequately explain how discrimination persists.

As such, the persistence of labour market discrimination is expected to derive mainly from either non-competitive forces or a slow adjustment or response to competitive forces. In particular, the monopsonistic approach can explain discrimination in the long run because it is based on the idea that the employer has non-negligible power in determining wages. Moreover, it can explain how equal pay legislation may increase wages without depressing employment (Manning, 2003). As an alternative, theories of social distance are able to explain the existence of occupational segregation without passing through the analysis of the additional cost attached to the minority group but through the tendency of the individual to conform (Akerlof, 1997; Reardon and Firebaugh, 2002).

Ultimately, this review is intended simply to highlight the possible sources of discrimination and possible mechanisms for the propagation and persistence of discrimination. The goal here is not to attempt to empirically test these different propositions, though the reader may bear these alternative models in mind as we discuss our findings.

6.2.2 Wage decomposition techniques

Wage differentials have been widely studied in the context of the Brazilian labour market. The majority of these studies decompose the differentials using the Oaxaca-Blinder (OB) decomposition technique (Oaxaca, 1973; Blinder, 1973). This regression-based decomposition[65] allows wage gaps to be separated into differences in observed characteristics (the “endowment” component) and differences in the returns to these observed characteristics (the “treatment” component). The treatment component permits researchers to identify the portion of earnings differentials that is attributable to differences in wage structure, which can be interpreted as capturing both differences in unobserved characteristics and the possibility of pure discriminatory behaviour. Since the seminal work by Oaxaca and Blinder, a variety of studies have sought to enhance the basic decomposition technique. Juhn, Murphy and Piece (1991) enriched the OB decomposition technique by adding a temporal dimension in order to better understand the wage gap residual term. Blau and Kahn (1996) proposed a very similar procedure in which temporal variation is replaced by cross-country variation. The flourishing development of techniques for decomposing earnings differentials is well documented in a recent review by Fortin, Lemieux and Firpo (2011).

Among these alternative techniques very few studies have managed to link the analysis of wage discrimination directly to the question of occupational segregation, despite the fact that differences in occupational distribution can represent an important source of differences in earnings. A group of studies that have sought to draw this connection have investigated the role of gender occupational segregation on gender

[65] The OB decomposition technique is a regression-based technique that exploits explanatory factors. There are alternative methods to decompose earnings differentials that apply additively decomposable indices (see for example Bourguignon and Ferreira (2005) and Bourguignon, Ferreira, and Leite (2008)). In addition, an alternative non-parametric technique developed by Ñopo (2008) is the marginal matching method that has been applied on the Brazilian labour market by Marquez Garcia, Ñopo and Salardi (2009).

wage gaps by considering the proportion of female workers within each occupation, i.e., the 'degree of feminization' (or femaleness of occupations). Among others, important contribution have been made by Johnson and Solon (1986), Macpherson and Hirsch (1995) and Cotter, Hermsen and Vanneman, (2003) for the U.S. labour market, Lucifora and Reilly (1992) for the Italian labour market and Baker and Fortin (2003) for the Canadian labour market, which they compare to the U.S. market. None of these studies has sought to extend this logic to studying the impact of racial occupational segregation on wage gaps.

More generally, in order to account for the impact of occupational segregation on wage discrimination, Blinder (1973) suggested use of occupational dummies to control for occupational distribution. However, jobs might not be randomly distributed, when subject to some sort of discriminatory mechanism, and if so the dummy variable approach is inadequate (Meng and Miller, 1995; Miller and Volker, 1995; Miller, 1987; Reilly, 1991). A seminal study by Brown, Moon and Zoloth (1980) was the first to more fully incorporate occupational structure into the decomposition of earnings differentials. The key insight behind the BMZ (1980) decomposition technique is that occupational structure affects wage discrimination and, as a consequence, wage differentials can be decomposed into within- and between-occupation components. Their approach uses a multinomial logit (MNL) model to predict occupational attainment in order to estimate a counterfactual occupational distribution. Therefore, the occupational attachment process is endogenously determined in the first step. Other studies (Neuman and Silber, 1996; Liu, Zhang and Chong, 2004) have generally adopted a similar multinomial logit approach, though Miller (1987) relied on an ordered probit to compute the counterfactual occupational distribution.[66]

A very important issue in implementing these decomposition methods is the need to correct for selectivity bias within the earnings functions. Neuman and Oaxaca (2003, 2004) proposed several specifications that incorporate the differentials in selection effects into the wage gap decomposition components. To the best of our knowledge, Reilly (1991) was the first paper that contributed to this literature by accounting for the effect of occupational segregation on wage gaps while

[66] This decomposition technique has also been employed to study differences in employment rates (Blackaby et al, 1998; Altonji and Blank, 1999).

simultaneously controlling for selection bias into different occupations using the Lee (1983) procedure.[67]

6.2.3 Studies of Brazilian wage differentials

Having described these alternative methodologies we now review existing wage decomposition analysis for the Brazilian labour market. We focus here on the results most relevant to this study, while a comprehensive review of empirical studies of Brazilian gender and racial wage differentials that apply the OB decomposition techniques can be found in Marquez Garcia, Ñopo and Salardi (2009).

Among studies that focus on both gender and racial issues, one of the most complete studies is that of Soares (2000). Between 1987 and 1998 he finds that racial wage gaps were, on average, greater than gender wage gaps, while gender wage gaps declined over time but racial pay gaps remained broadly constant. Consistent with this overall trend, Lovell (2000, 2006) reports that prior to the 1980s gender wage gaps were, on average, greater than racial wage gaps. Turning to the decomposition results, Soares (2000) finds that female pay differentials are driven primarily by pure wage differentials, while both black women and men suffer additionally from lower human capital accumulation and insertion discrimination. These general findings have been confirmed by the empirical literature on Brazilian wage gaps (see, among others, Lovell, 1994; Cavalieri and Fernandez, 1998; Arabsheibani, Carneiro and Henley, 2003; Arias, Yamada and Terejina, 2004). Even when adopting a non-parametric marginal matching method developed by Ñopo (2008), Marquez Garcia, Ñopo and Salardi (2009) find that racial wage gaps are more pronounced than those along the gender divide and that the former have decreased less rapidly than the latter. They find that over the last decade the total gender wage differential has shrunk by 25% of its initial value, while the total racial wage gap has decreased by 18%.[68]

A recent study by Madalozzo (2010) investigated the connection between occupational segregation and wage gaps by treating occupational attachment as an exogenous process (i.e., by simply adding dummies for occupations into the earnings

[67] An earlier study by Dolton, Makepeace and van der Klaauw (1989) corrected earnings functions for participation in different occupations using a Lee type correction.

[68] Marquez Garcia, Ñopo and Salardi (2009) analyse gender and racial wage gaps in Brazil over the last decade. In their study, wage gaps are computed as differentials between relative wages. They found that not only racial wage gap is greater than gender one, but also it has decrease at a slower pace. Gender wage gap changed from 0.522 in 1996 to 0.391 in 2006, while racial wage gap contracted from 0.961 to 0.787.

equations, as originally suggested by Blinder (1973)). To the best of our knowledge, there are only two studies of the Brazilian labour market that have analysed wage gaps by simultaneously accounting for endogenously determined occupational segregation. Ometto, Hoffmann and Alves (1999) focus on comparing between- and within-occupation gender wage differentials for two states of Brazil, Pernanbuco and São Paulo, the latter of which is wealthier than the former. They find that the within-occupation component (also called intra-occupational) is dominant for the less wealthy Pernanbuco, while both the inter- and intra-occupational components are relevant for the case of São Paulo. Arcand and D'Hombres (2004) analyse the effect of occupational segregation on racial wage gaps among adult male workers, distinguishing between categories for 'brown' and 'black' skin tone. They find that differences in endowments are the primary explanation for wage differentials. The effect of occupational segregation is found to be small, but higher for black workers, accounting for 5% and 8% of the total wage gap for brown and black workers respectively.

An element that has yet to be explored by these studies is the potential importance of differences between the formal and non-formal labour markets in Brazil. As described in the previous chapters, the informal sector involves almost half of the employed labour force (Bosch, Goni and Maloney, 2007). Given that the formal sector provides scope for regulated labour markets to function, we might expect different outcomes in terms of earnings differentials between the formal and non-formal sectors.

Relatively few studies have investigated wage differentials across the formal and non-formal sectors. Birdsall and Berhman (1991) analysed gender wage gaps in the 1970s and found that gender wage differentials in the formal sector were largely attributable to unobservable characteristics while this was true to a much smaller extent in the informal sector. Whereas women working in the formal labour market were relatively more educated, wage differentials in the informal sector were largely driven by education and household variables. A similar study by Tiefenthaler (1992) found that the situation at the end of the 1980s was very similar to that which prevailed in the 1970s. In a more recent study, Silva and Kassouf (2000) find that gender wage gaps in the formal and informal sectors have declined over the last two decades. While they find that gender pay gaps are greater in the formal than the informal sector, they find that they are also decreasing faster than in the latter sector.[69] Cacciamali and Hirata

[69] It is important to note minor differences in the way that the different studies group workers among formal and non-formal activities. Birdshall and Behrman (1991) divide workers into formal, informal and

(2005) and Cacciamali, Tatei and Rosalino (2009) explore differences across the formal and informal sectors not only by gender but also by skin colour. They also find the extent of discrimination is greater within the formal sector for both women and non-whites.

Finally, an important element in many of these studies is an effort to correct for potential selectivity issues in the wage equation estimation. The majority of the papers that investigate differences between the formal and informal labour markets have corrected for selectivity in participation in these two different sectors (see Birdsall and Berhman, 1991; Tiefenthaler, 1992; Silva and Kassouf, 2000; Carneiro and Henley, 2001). On average, they report evidence of positive selection. This holds true even when a multinomial choice model is used for the selectivity correction, as in the case of Stecler et al (1992). Carneiro and Henley (2001) provide evidence that there are earnings advantages through being employed in both the formal and informal sectors and hence informal employment might be a desirable form of employment in Brazil in some cases, as in many other South American countries.

Two relatively recent studies highlight the potential complexity of the selection process. Loureiro, Carneiro and Sachsida (2004) detect a change between 1992 and 1998 from negative to positive selection, particularly for female workers and for rural male workers (while urban male workers always yield positive selection effects). Findings from Carvalho, Neri and Silva (2006) are particularly complex, as they report that correcting for selection bias results in a lower discrimination effect when comparing white women and men, but a larger discrimination effect when comparing non-white women with white men. In the first case the treatment is dominant while in the second the endowment effect is dominant. The explanation for these findings may lie in the fact that we would expect to estimate lower coefficients once we control for selection bias. The effects of these coefficients may have been overestimated, as the unobservable variables may capture ability and motivation. So, intuitively, in the Oaxaca decomposition the component that refers to estimated returns to characteristics should contract.

Collectively these studies shed significant light on the various dimensions of wage differentials in Brazil. Building on this existing research, the objective of our

---

domestic work if females and into formal and informal if males. Tiefenthaler (1992) divides the entire sample of workers into formal, informal and self-employed, while Silva and Kassouf (2000) into formal, informal and employers.

analysis is to bring together the disparate elements of this literature in a more complete and comprehensive fashion. Thus, the analysis that follows decomposes the components of both gender and racial earnings differentials, and explores changes in both aggregate wage gaps and in their components over the past two decades. In order to further deepen the analysis it takes into account the impact of occupational segregation in explaining wage differentials, while the analysis is also further disaggregated into the formal and non-formal sectors, to reflect potentially different labour market dynamics. Finally, we look at all of the decomposition results both with and without corrections for selection bias in order to assess the robustness of the reported findings.

6.3 Data description

We employ secondary 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 (see chapter 3). The sample employed in this chapter includes workers aged between 15 and 65 years who declare that they are working, while excluding those for whom there are missing observations for either their wages or their occupational codes. The wage used is the log of hourly earnings from the primary employment expressed in nominal terms (in current national currency, the Brazilian *reais*). When we seek to account for the selection process later in the chapter we incorporate information about those in the adult population who declare themselves to be unemployed or out of the labour force.

The occupational codes variable is central to our analysis. As highlighted in previous chapters, the original PNAD occupational classification varies across years, and for the majority of available years is not directly comparable with the international classification provided by the ILO, the ISCO-08. For this reason, we have constructed a classification that is harmonized and consistent over time, employing the occupational codes described and adopted in the previous chapters. This classification has 25 different occupational categories at the 2-digit level. However, for the analysis here we need to re-aggregate some of them in order to guarantee a minimum number of observations for each occupation-specific wage equation.[70] As such, the ultimate number of occupations that we consider is slightly different for the gender and racial wage analysis and across the formal and informal sectors.

The availability of an occupational classification that is harmonized over time enables us to construct an occupational intensity variable, which is also a key feature of our analysis. The female occupational intensity variable (focc3) is computed as the ratio of female workers within occupations, while the non-white occupational intensity variable (nwocc3) is the ratio of non-white workers within occupations. Since these variables are constructed based on the occupational classification at the 3-digit level they cover 83 different occupational codes. As noted earlier, occupational intensity variables have been used previously to capture occupational segregation and its effect on earnings in various studies of female occupational intensity (or degree of feminization). However, to the best of our knowledge the impact of non-white occupational intensity has never been similarly explored. In this chapter we begin to explore this relationship, while we extend the analysis in the next chapter, as this is a comparatively underexplored area and we are thus able to provide useful insights into these relationships.

Table 1 presents the size and composition of our sample for the first and last years of the period of interest (i.e., 1987 and 2006). Aside from the fact that the sample size has increased considerably over time, we can see a steady increase in the labour force over time driven primarily by rising female participation among both white and non-whites. We also see a rise in unemployment, as greater labour market participation has generated not only an increase in the share of the population that is employed, but also a persistent increase in the share that is unemployed.

The previous chapters detailed the evolution of labour market participation across the formal and non-formal sectors, and among all four population sub-groups (i.e. female, male, non-white and white workers). The only point that bears re-emphasizing here is that increasing female participation has occurred primarily in the formal sector (a 5.6 percentage point increase), to some extent in the informal sector (a 2.2 percentage point increase) and only marginally in the self-employed category (only one third of a percentage point).

[70] The number of occupations is hence reduced to a minimum of 16 categories and a maximum of 22 depending on whether we are considering the formal, informal or self-employed labour markets (with the self-employed sector featuring the most restricted number of occupations).

Table 1: Size and composition of the sample

PANEL A – Sample of year 1987

	female	%	male	%	non white	%	white	%	total	%
formal	15,203	0.175	29,400	0.389	18,168	0.243	26,435	0.302	44,603	0.275
informal	12,011	0.138	16,448	0.218	15,516	0.207	12,943	0.148	28,459	0.175
self-empl.	8,116	0.093	17,044	0.226	12,299	0.164	12,861	0.147	25,160	0.155
unemployed	1,810	0.021	2,653	0.035	2,182	0.029	2,281	0.026	4,463	0.027
labour force	37,140	0.428	65,545	0.868	48,165	0.644	54,520	0.623	102,685	0.632
out of the LF	49,719	0.572	9,974	0.132	26,682	0.356	33,011	0.377	59,693	0.368
total	86,859	0.535	75,519	0.465	74,847	0.461	87,531	0.539	162,378	

PANEL B – Sample of year 2006

	female	%	male	%	non white	%	white	%	total	%
formal	30,174	0.231	40,657	0.346	33,824	0.251	37,007	0.326	70,831	0.285
informal	20,573	0.157	22,431	0.191	26,292	0.195	16,712	0.147	43,004	0.173
self-empl.	12,558	0.096	25,262	0.215	20,730	0.154	17,090	0.151	37,820	0.152
unemployed	10,492	0.080	7,791	0.066	10,954	0.081	7,329	0.065	18,283	0.074
labour force	73,797	0.564	96,141	0.818	91,800	0.681	78,138	0.689	169,938	0.684
out of the LF	56,948	0.436	21,391	0.182	43,029	0.319	35,310	0.311	78,339	0.316
total	130,745	0.527	117,532	0.473	134,829	0.543	113,448	0.457	248,277	

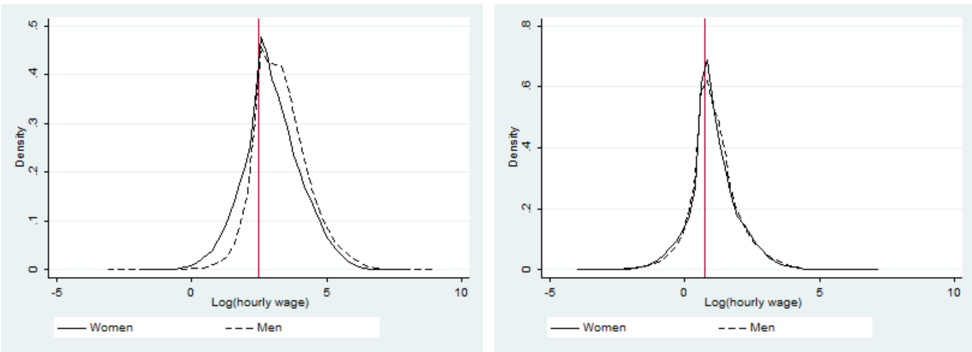
Source: Author's computations using PNAD 1987 and 2006.

In order to provide additional descriptive data about our samples, table A1 in the appendix reports key covariates for each population sub-group in 1987 and 2006 along with t-tests results. The first important message relates to education, as females have a clear advantage in educational attainment, which has modestly increased over time. Meanwhile, non-white workers have a significant educational disadvantage compared to white workers, but this disadvantage has diminished over time. A second important issue relates to patterns of occupational segregation. We can see that the level of gender concentration in occupations is much higher than racial concentration. That noted, we also see a modest decline in female occupational intensity within the female sample and an increase in female occupation intensity within the male sample. In contrast, the degree of non-whiteness has increased over time for both non-white and white samples.

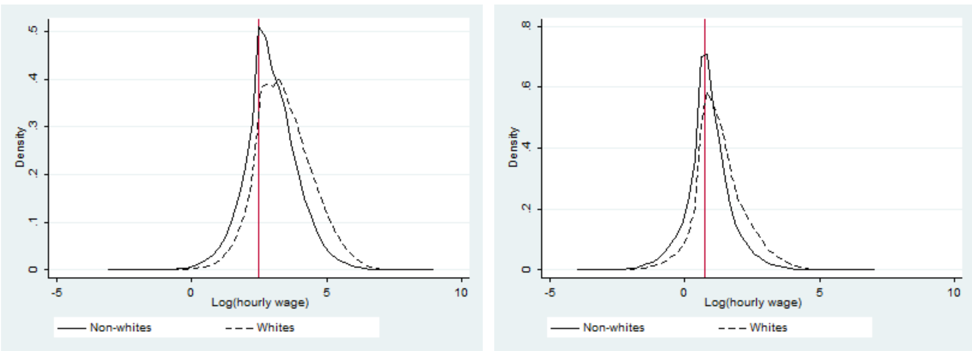
Moving closer to our core interest in the wage distribution, figure 1 plots the density distribution of wages by gender and race (for both panels, the graph to the left refers to the first year, 1987, while the graph to the right refers to the last year, 2006). Beginning with gender differentials, the male wage distribution is shifted clearly to the right in 1987, but by 2006 that differential has almost completely disappeared. On the other hand the difference between the non-white and white distributions has remained almost unchanged over the past two decades, with the white wage distribution located to the right. The plots in figure 1 also reveal that the wage distributions have become more "peaked" over time, which signifies a lower variance, or less dispersion, in wage values. While not reported here, disaggregation into the formal, informal and self-employed sectors reveals very similar patterns, with the exception that the wage distributions for the formal sector appear "flatter", indicating a wider variety of wage values across formal occupations than across non-formal occupations.

Figure 1: Kernel density of log hourly wage

Panel A – By gender, 1987 and 2006



Panel B – By race, 1987 and 2006



Source: Author's computations using PNAD 1987 and 2006.
Note: The vertical red lines represent the nominal minimum wage values for each year.

These plots provide a preliminary indication of the evolution of wage differentials over time, as we can see a visible improvement in gender differentials than race differentials. However, the implementation of wage gap decomposition techniques will provide more detailed portrait of these trends and of the mechanisms underlying them.

6.4 Methodology

In this section we first present a detailed description of the Oaxaca-Blinder (1973) and Brown, Moon and Zoloth (1980) decomposition techniques. We then discuss the problem of selection in estimating wage equations and subsequently propose alternative specifications that incorporate selection effect differentials into the decomposition formulae.

6.4.1 Oaxaca-Blinder (1973) and Brown, Moon and Zoloth (1980) decomposition techniques

We begin by outlining the OB wage decomposition and then move to the Brown, Moon and Zoloth (1980) variant, which takes occupational segregation into account. In order to estimate the wage gap decomposition, we first need to estimate two separate earnings equations. The underlying assumption is that the two groups of interest vary both in their characteristics and in the rewards to those characteristics. For simplicity we denote these two groups type-A and type-B workers. These represent, in turn, men and women for gender wage gaps and white and non-white workers for racial wage gaps.[71]

The specification of the wage equation follows the seminal work of Mincer (1974), where the log of wages is a function of a set of wage determining characteristics, which comprise controls for human capital. The group-specific wage equation to be estimated is:

$$lnW_i^k = X_i^{k\prime}\beta^k + \varepsilon_i^k \tag{1}$$

where X_i^k is the matrix of the observed characteristics for group k , β^k is the vector of the unknown parameters for group k and ε_i^k is its random error term. The superscript k refers to either group A or group B.

The standard OB methodology decomposes wage differentials as follows:

$$\overline{lnW}^A - \overline{lnW}^B = \bar{X}^{A\prime}\hat{\beta}^A - \bar{X}^{B\prime}\hat{\beta}^B = (\bar{X}^A - \bar{X}^B)'\hat{\beta}^A + \bar{X}^{B\prime}(\hat{\beta}^A - \hat{\beta}^B) \tag{2}$$

[71] The empirical validity of separating of the data in order to estimate the wage equation is empirically tested using a Wald test.

where the 'bars' denote mean values and 'hats' the OLS coefficient estimates.

Formula (2) above shows how average wage gaps between group A and group B can be decomposed into two components, the endowment effect and the treatment effect, or, alternatively, the explained and unexplained components, respectively. The endowment effect refers to differences in observed characteristics, while the treatment effect refers to differences in the returns to these characteristics. The latter component, often called the wage structure effect, is sometimes taken to capture the effect of unequal treatment (or discrimination) in the labour market.

A well-known problem related to the OB decomposition is the "index number" problem. Decomposition results vary depending of the choice of which category is selected as the reference category in the decomposition formula. More general approaches to wage decomposition that try to overcome the "index number" problem have been proposed by Neumark (1988) and by Oaxaca and Ransom (1988, 1994). These more general techniques estimate a non-discriminatory wage structure from a pooled sample.[72] However, we proceed with our decomposition within a discriminatory setting, with our counterfactual representing the average wage that type-B workers would earn if they were paid as type-A workers, namely $\bar{X}_B{}'\hat{\beta}_A$. As a consequence, the discriminatory wage differentials are attributed to underpayment of the subordinate group (type-B group) rather than an overpayment of the dominant one (type-A workers). In other words, we are assuming discrimination against women and non-white workers, rather than positive discrimination or nepotism in favour of men and white workers.

A second issue is related to the choice of the omitted category for dummy and categorical variables. The "omitted category" problem refers to the fact that the contribution of dummy and categorical variables to the unexplained components varies depending on the choice of the reference groups in the regression model (Jones, 1983; Oaxaca and Ransom, 1999). The choice is completely arbitrary and results in different unexplained components, which ultimately represent different estimated

[72] A previous study by Dolton and Makepeace (1985) proposes a comparison between the whole distribution in the case of discrimination and the whole distribution in the case of no discrimination. They derive a wage distribution density in the absence of discrimination.

counterfactuals. It is difficult to provide a unified solution to the "omitted category" problem.[73]

Building on the standard OB decomposition, the technique developed by Brown, Moon and Zoloth (1980) takes occupational structure into account in decomposing wage gaps. Their approach consists of incorporating a model of occupational attainment into the estimation of wage gaps.

Following their method, we first need to estimate the counterfactual occupational distribution. Let p_j^A and p_j^B be the proportion of type-A and type-B workers in the sample that are employed in occupation *j*, with *j=1...m*. We denote $\hat{p}_j^*$ the counterfactual occupational distribution, which is the proportion of type-B workers in the sample who would be employed in occupation *j* if they benefitted from the same occupational structure as type-A workers.

In order to compute the counterfactual distribution, we estimate a model for occupational attachment using the multinomial logit model (MNL).[74] Assuming that the probability $p_{i,j}$ of attachment of the *i* individual to the *j* occupational category is determined by the vector *Z*, the MNL model of occupational attainment may be represented as:

$$p_{i,j} = \frac{\exp(Z_i \gamma_i)}{1+\sum_{j=1}^{m-1}(Z_i \gamma_i)} \tag{3}$$

where the base outcome *j* is set equal to 0 as required by the Theil normalization. Estimates from (3) are then use to construct the counterfactual occupational distribution, $\hat{p}_j^*$. Since, in our case, it measures the proportion of type-B workers who would work in occupation *j* if the type-A occupational distribution were imposed, it is constructed by

[73] The "omitted category" problem is not an issue for what is proposed in this chapter as we do not present disaggregated (or "detailed") decomposition analysis. However, we will explore this problem in the next chapter where we do provide detailed decomposition estimates.

[74] Miller and Volker (1985) model occupational attainment through both unordered and order discrete choice models. Miller (1987) measures the effect of occupational segregation on wages by applying a multinomial ordered probit model to estimate sectoral attachment. Finally, Reilly (1991) provides empirical estimates taking into account sectoral attachment via multinomial logit models and correcting wage equations for occupational selection à la Lee (1983). There are a few other empirical studies that have employed this approach, and which look at gender differentials. For instance, Neuman and Silber (1996) using Israeli data and Appleton, Hoddinott and Khrisnan (1999) for three African countries, Cotè d'Ivoire, Ethiopia and Uganda. Specifically for Brazil, as already noted in the previous section, there is one study of gender differentials by Ometto, Hoffman and Alves (1999) and one for racial wage gaps by Arcand and D'Hombres (2004).

using the estimated coefficients of the MNL model for occupational attainment for type-A workers and the individual realizations of type-B workers.

The construction of the occupational distribution employs our harmonized classification of occupational codes. This classification allows for a large number of occupations, as we were able to construct a set of three-digit occupational codes. However, in order to achieve feasible estimation of the simulated occupational distributions, we choose to rely on the occupational codes aggregated to the two-digit level, which captures 22 different occupations.

The second step is the estimation of a set of occupation-specific wage equations expressed as follows:

$$lnW_{i,j}^{k} = X_{i,j}^{k}{}'\beta_{i,j}^{k} + \varepsilon_{i,j}^{k} \tag{4}$$

where the superscript *k* denotes the group, type-A or type-B workers, and the subscript *j* denotes the j^{th} occupation. Then we use the occupational distributions, p_j^A and p_j^B, the estimated counterfactual occupational distribution, $\hat{p}_j^*$, and the occupation-specific wage equations to decompose the overall average wage gaps between group A and B as follows:

$$\overline{lnW^A} - \overline{lnW^B} = \sum_j \bar{p}_j^B \left(\bar{X}_j^A - \bar{X}_j^B\right)'\hat{\beta}_j^A + \sum_j \bar{p}_j^B \bar{X}_j^B{}'\left(\hat{\beta}_j^A - \hat{\beta}_j^B\right) + \sum_j \bar{W}_j^A\left(\bar{p}_j^A - \hat{p}_j^*\right) + \sum_j \bar{W}_j^A\left(\hat{p}_j^* - \bar{p}_j^B\right) \tag{5}$$

The first term on the right-side of the equation represents the portion due to differences in endowments between type-A and type-B workers within occupations, defined as the “explained intra-occupational component”. The second term represents the difference in returns to characteristics between type-A and type-B workers within occupations, the “unexplained intra-occupational component”. These first two terms are similar to the traditional OB decomposition. The novelty of the Brown, Moon and Zoloth (1980) decomposition lies in the last two terms, which capture the effect of sectoral attachment in the wage decomposition formula. The third term captures differences in average qualifications across occupations (the explained component of the allocation of workers or “explained inter-occupational component”) and the fourth term captures differences in the structure of the occupational achievement and represents the portion of the wage gap due to occupational segregation (i.e., the "unexplained inter-occupational component").

As with the standard OB decomposition, we construct the counterfactual wage by taking the characteristics of type-B workers and applying the wage structure of type-A workers. In the case of gender wage differentials, this means that the counterfactual is constructed using female characteristics and the male wage structure, while for racial wage gaps it is constructed using non-white characteristics and the white wage structure. This reflects an assumption that the male and white wage structures would prevail in the absence of discrimination. This, though, does not resolve the "index number" problem. Appleton, Hoddinott and Krishnan (1999) propose the adoption of a pooled sample wage structure when employing the BMZ decomposition, following the strategy adopted by Oaxaca and Ransom (1994) when employing the standard OB decomposition, but because of the assumption made above, we do not follow their approach.

Finally, a note about terminology. We employ the term "vertical segregation" to refer to the unexplained intra-occupational component, which captures unexplained differences in wage returns within occupations. On the other hand, we employ the term "horizontal segregation" to refer to the estimated unexplained inter-occupational component, as it defines the extent of differences in wage rewards across occupations. The first concept captures wage disadvantage within occupations, the later captures wage disadvantage across occupations.[75]

6.4.2. Correcting for selection

The techniques discussed so far do not attempt to address the selectivity bias problem, but in the presence of selection OLS estimation can yield biased and inconsistent estimators (Gronau, 1974; Heckman, 1979). The most common technique

[75] It is important to be extremely clear about definitions, as these definitions of horizontal and vertical segregation sometimes differ within the economics literature. Blackburn, Brooks and Jarman (2001) argue that segregation measures are often conceptualized as overall measures, capturing both the horizontal and vertical dimensions. In a later study, Blackburn and Jarman (2005) equate vertical segregation with gender inequality and horizontal segregation with differences in occupational structure, exclusive of occupational inequality. Along the same lines, Semyonov and Jones (1998) distinguished between nominal and hierarchical segregation and argued that occupational segregation should not be confused with occupational inequality. In other studies, vertical (hierarchical) segregation has referred to the 'glass-ceiling' phenomenon, which focuses on the presence of certain groups (for example women) in top jobs (see Maume (1999) and Anker, Melkas and Korten (2003), among others). Finally, Merkas and Anker (1997) defined horizontal segregation as differences in employment across occupations, and vertical segregation as differences in positions within occupational groups. These two concepts tend to be equivalent when employing a sufficiently detailed number of occupations.

to correct for selection bias has been developed by Heckman (1979). This procedure is valid when the selection bias is generated by a binary distinction between participating and not participating in the labour market. However, if participation in the labour market is a more complex problem the Heckman procedure provides a more restrictive methodology. In other words, the participation decision may not be a simple binary choice between participating and not participating, but may involve multiple choices related, for example, to participating in different occupations, different economic activities or different sectors. In the Brazilian case, it seems particularly apposite that the participation decision might be shaped by the presence of formal and non-formal labour markets.

Given this possibility, Lee (1983) has proposed a more general approach to the issue. The Lee procedure is a two-step procedure that exploits estimates from the MNL model, rather than the probit, to construct the selection correction terms. Using the estimates obtained from a reduced form MNL we can compute a set of correction terms with one for each mutually exclusive category as follows:

$$\lambda_{i,j} = \frac{\phi(Z_{i,j})}{p_{i,j}} \tag{6}$$

This selection term, also called the inverse of the Mills ratio (IMR), is computed for each individual *i* and for each of the mutually exclusive categories *j*. It is very similar to the terms computed using the Heckman two-step approach, with the primary difference that the cumulative distribution function in equation (6) is based on the predicted probabilities using the logistic MNL cumulative distribution function.

The earnings equations corrected for selectivity bias can then be re-espressed as:

$$lnW_{i,j}^{k} = X_{i,j}^{k}\beta_{j}^{k} - \theta_{j}^{k}\lambda_{i,j}^{k} + \varepsilon_{i,j}^{k} \tag{7}$$

where θ_j^k are the unknown selection parameters, one for each *k* group (type-A and type-B workers), and $\lambda_{i,j}^{k}$ is the selection variable (IMR) computed in this application using the Lee (1983) procedure.

An important property of the OLS procedure is that the estimates are computed using mean values from the data, which allows the decomposition to incorporate the differentials in selection terms. However, Neuman and Oaxaca (2003, 2004) have discussed the ambiguities that selection correction introduces in decomposition analysis, focusing on how to interpret differences in the selection terms. The estimated parameter, θ_j^k, is the product of the standard deviation of the errors in the wage equation and the correlation between the errors in the wage equations and in the participation equation. It can therefore be tricky to define whether the selection effect differentials should be part of the component capturing differences in endowments or consigned to the wage structure effect.[76] In their studies, Neuman and Oaxaca propose several ways to incorporate the differentials in selection terms into the wage gap decomposition. We choose to adopt the straightforward decomposition in which the selection terms differential is kept separate from the endowment and wage structure components:

$$\overline{lnW}^A - \overline{lnW}^B = (\bar{X}^A - \bar{X}^B)'^{\hat{\beta}^A} + \bar{X}^{B\prime}\left(\hat{\beta}^A - \hat{\beta}^B\right) + (\hat{\theta}^A\bar{\lambda}^A - \hat{\theta}^B\bar{\lambda}^B) \qquad (8)$$

If the selection effects are netted out of the overall wage gap, the resultant differential is referred to as a “wage offer gap” (Reimers, 1983; Neuman and Oaxaca 2004).

Reimers (1983) adopts a very similar decomposition formula where the selectivity correction is applied to a non discriminatory wage structure as developed by Oaxaca and Ransom (1994). We can consequently derive a similar formula for the Brown, Moon and Zoloth (1980) wage gaps decomposition when correcting for selection:

$$\overline{lnW}^A - \overline{lnW}^B = \sum_j \bar{p}_j^B\left(\bar{X}_j^A - \bar{X}_j^B\right)'\hat{\beta}_j^A + \sum_j \bar{p}_j^B \bar{X}_j^{B\prime}\left(\hat{\beta}_j^A - \hat{\beta}_j^B\right) +$$
$$\sum_j \overline{W}_j^A\left(\bar{p}_j^A - \hat{p}_j^*\right) + \sum_j \overline{W}_j^A\left(\hat{p}_j^* - \bar{p}_j^B\right) + \sum_j \bar{p}_j^B\left[\hat{\theta}_j^A\bar{\lambda}_j^A - \hat{\theta}_j^B\,\bar{\lambda}_j^B\right] \qquad (9)$$

where the last term on the right-side of the equation captures the differences in the selection terms.

Several empirical studies have adopted the Lee correction, with Stecler et al (1992) the first Brazilian study to correct for selection with multiple outcomes. In his

[76] Neuman and Oaxaca (2003) also claim that “The same issue arises with respect to group differences in the IMR which reflect nonlinear group differences in the determinants of selection and in the probit coefficients.”

study he considers a selection equation where the possible outcomes are non-participation, participation as an employee or self-employed participation.

Moving to non-Brazilian studies, Dolton, Makepeace, van der Klaauw (1989) estimate a simultaneous model of occupational attainment, wage determination, and occupational status by gender in which selectivity corrections are included in the wage and occupational status equations. Selectivity corrections are made for the labour force participation of women while occupational selectivity corrections are made for both men and women. An innovative alternative procedure is proposed by Reilly (1991). He applies the Lee procedure in order to correct for selection when employing the Brown, Moon and Zoloth (1980) decomposition approach. The novelty is that the estimates from the MNL model for occupational attainment are used both to construct the Lee selection term and to construct the counterfactual occupational distribution for the Brown, Moon and Zoloth (1980) decomposition technique.

The approach here is slightly different. As in Reilly (1991), we employ the Lee procedure to correct for selectivity in the BMZ decomposition. However we estimate two separate MNL models. The first MNL model is employed as the selection equation and estimates the likelihood of participating in the formal, informal and self-employed sectors. The second MNL model is used to construct the counterfactual occupational distribution. Thus by estimating the likelihood of working in a specific occupation, it predicts occupational attachment. The MNL estimates can potentially be invalidated if the property of independence from irrelevant alternatives (IIA) is not satisfied. However, Bourguignon, Fourier and Gurgand (2007) show how the MNL model can provide a good correction for the wage equation even if the IIA hypothesis is actually violated.

6.5 Empirical Findings

In this section we present our main empirical findings when applying the decomposition techniques, both with and without taking account of occupational segregation. We first analyse the estimates from the earnings equations, and then present the baseline results when applying the standard OB decomposition. We then report our findings when taking account of occupational segregation using the Brown, Moon and Zoloth (1980) decomposition method (BMZ, hereafter), highlighting key changes in the results. Finally, we conclude this section by presenting additional robustness checks for the standard OB wage gaps decomposition.

6.5.1 Estimates from the earnings equations

Central to implementing regression-based decomposition techniques is the estimation of the earnings equations of interest, and we begin by describing the main features of these estimates. Following the majority of empirical studies of Brazil, the dependent variable in all of the wage equations is the log of hourly wages. This is preferred to the alternative of using monthly wages and a control for hours worked, which introduces potential endogeneity problems. While we believe that hourly wages are a more reliable dependent variable, we show at the end of this section how and why the standard OB decomposition results change if monthly wages are used instead.

We explore four different specifications for these wage equations. In the first specification the wage equation is estimated using the baseline controls for gender, skin colour, age and age squared, years of education, a dummy for living in urban or metropolitan areas, a dummy for working in the formal sector and dummies for living in each of the five main Brazilian regions (North, North-East, South-East, South and Central-West). It is important to note that the race dummy is included in the wage equations estimated using the samples disaggregated by gender, while the gender dummy is included in the wage equations disaggregated by race. The three subsequent specifications include controls for occupational structure to the baseline specification. The second specification adds dummies for the occupational codes at the 2-digit level, while the third specification adds the continuous variable for occupational intensity (focc3 or nwocc3). The final specification includes both dummies for occupational codes and the occupational intensity variable.[77]

We begin with an austere specification and then progressively incorporate the different measures for occupational structure in order to compare the role of occupational structure in the OB decomposition results, which treat occupational structure as an exogenous process, to the results when using the BMZ method, in which occupational attachment is determined endogenously. A further motivation for using an austere wage equation specification is to ensure compatibility in the specifications, as we have less rich data during the earlier years of our sample. For example, the PNAD started to collect variables on work experience only in the 1990s. To assess our specifications we perform several checks at the end of this section to see whether direct measures of work experience change the estimates for the earnings equations, and subsequently the OB decomposition results.

Finally, all of the regressions are performed using a large sample of workers (see table 1). We include workers aged 15-65 working across all economic sectors, both private and public and living in both rural and urban areas. We include both formal and non-formal labour markets, given that we are also interested in exploring differences between them. We only exclude those in the military forces and employers. Our motivation for retaining a very large sample is motivated by an interest in capturing trends for the aggregate Brazilian labour market, unlike other studies, which have focused only on particular segments of the labour market. There is equally a more practical motivation, as the BMZ decomposition method requires us to estimate occupation-specific wage equations for each sub-group of the population (i.e. female, male, non-white and white workers). At the end of the section we present robustness checks investigating the implications of retaining this large sample for the results, as we experiment with restricting the sample to only workers working more than 35 hours per week ('full-time' workers),[78] dropping the self-employed sector and finally also dropping workers in agricultural activities.

---

[77] In order to account for the occupational structure, the most complete specification includes occupational dummies in addition to female or non-white occupational intensity. There are many possible reasons to opt for the inclusion of occupational dummies in addition to occupational intensity variables. One is that these occupational dummies might capture other occupation-specific effects that are not captured by the intensity measure (e.g., the role of compensating differentials). Another possibility is that the occupational dummies could capture barriers to entry/exit in certain occupations.

[78] Using the OECD definition, 'full-time' workers are those working more than 35 hours per week. As we will see caution is required in interpreting any analysis of Brazilian data when including this variable, as in Brazil part-time employment is closely associated with precarious employment.

As a representative example, tables A2 and A3 in the appendix report a set of wage equations using the OB decomposition for the first and last years of interest (1987 and 2006) across all four wage equation specifications.[79] In table A2 we report the regression estimates used to implement the OB decomposition for the gender wage gaps. In columns (1) and (5) of panel A, we see that in 1987 one additional year of education (edu) increased wages by 13.7% for women and by 12% for men. Over time the returns to education seem to have diminished slightly. In 2006, one additional year of education increased wages by 10.5% for women and by 9.6% for men (see columns (1) and (5) in panel B).[80] Controlling for occupational structure by including dummies decreases the returns to education by roughly 3-5 percentage points, but we do not see the same decline when we control instead for female occupational intensity.

These estimated returns to education are in line with those reported in studies from this literature. Strauss and Thomas (1996) found returns of 10-13% for men and 11-18% for women in the South region in 1982, with higher returns for both men and women in the North-East (10-16% for men and 10-21% for women). Arriagada and Ziderman (1992) estimated a rate of return to vocational education of roughly 22 percent, which was broadly similar to that for academic education. Ferreira and Barros (1999) found decreasing returns to education over time, though with similar magnitudes to earlier studies, at 12.9% in 1985 and 8% in 1996. While pointing to a non-linear relationship between earnings and education, they also claim that this relationship has become more convex over time. The "convexification" of returns is similarly highlighted by Binelli (2008), who examines both Brazil and two other Latin American countries, including Mexico and Colombia.[81]

Interestingly the female occupational intensity variable (focc3) behaves very differently between the female and male specifications. For female workers in 1987, a

---

[79] To conserve space, we do not report the estimates for the 22 occupation-specific wage equations or for the MNL estimates for occupational attainment that are required for the BMZ decomposition analysis.

[80] Estimations of the returns to education should be interpreted cautiously, as there are many potential sources of bias. These may include, for example, the omission of ability and motivation variables, or aspects of marriage market returns, among others (Berhman and Birdsall, 1983). However, in a comprehensive study of the relationship between education and earnings, Card (1999) observed that the average returns to education once you control for ability, and the endogeneity of education, are not different from the estimates emerging from a standard OLS wage equation estimated using a cross-section (the "upward ability bias" is estimated to be around 10%).

[81] Binelli (2008) further notes that in 2002 the wage premium between college and intermediate education was greater than one hundred percent in Brazil and Colombia, while it was around 60 percent in Mexico. Although there is evidence of a convex pattern for returns to education, our specification for the wage equation only considers a linear relationship for education and a non-linear relationship for age. We have, though, experimented with adding the squared years of education and the decomposition results are unaltered.

10 percentage point increase in female occupational intensity decreases their wages by roughly 4% when we also include occupational dummies (column (4) of panel A). For male workers, the same change in female occupational intensity increases their wages by 0.95% (column (8) of panel A). Over time the impact of female occupational intensity declines considerably but a 10 percentage points increase in female occupational intensity is found to decrease female wages by 1.5% (see column (4) in panel B), though it no longer has a positive effect on male wages. These preliminary results appear to point toward an important dimension of wage determination. A more detailed and complete discussion of these issues is reserved for chapter 7.

In table A3 we report the estimates used to implement the OB decomposition for racial wage gaps. Returns to education are higher for white workers than non-white: one additional year of education increases white wages by 13.4% and non-white wages by 11.3%. As in the case of gender we see that the returns to education have diminished over time. Interestingly, in contrast to the gender case, the impact of education decreases whether we control for occupation or for non-white occupational intensity (nwocc3). This difference likely reflects the fact that non-white occupational intensity follows the pattern of educational attainment (i.e., non-white dominated professions are also more unskilled jobs, while, by contrast, female-dominated professions are, if anything, more skilled). Consistent with this pattern, non-white occupational intensity affects wages negatively for both groups, and particularly for white workers. It is interesting to note that while the negative impact of female occupational intensity has diminished over time, the negative impact of non-white occupational intensity appears to have increased in recent years. When controlling for both occupations and occupational intensity a 10 percentage points increase in non-white occupational intensity decreases non-white hourly pay by 5.3% and white pay by 11% in 2006, whereas the equivalent figures in 1987 were only 2.4% and 6.9%, respectively.

In sum, returns to education are higher for female and white workers, though this impact has diminished over time and is attenuated if we control for occupational dummies or non-white occupational intensity. Female-occupational intensity seems to negatively affect female wages, but not male wages, and its decrease over time is statistically significant at the 1% level.[82] By contrast, non-white occupational intensity negatively affects both non-white and white wages, with a stronger negative effect on

[82] The t-test for the decrease over time in the negative impact of female occupational intensity on female wages is equal to -8.26.

white workers, while this negative effect appears to have increased in more recent years.[83]

6.5.2 The standard Oaxaca-Blinder (1973) decomposition results

We now focus on the OB decomposition results. We first estimate gender and racial wage gaps across the four equation specifications in order to explore the sensitivity of the decomposition results to changes in the specifications. Tables 2 and 3 report these findings, again for the initial and final years only, while the results are also reported for the entire labour market as well as for the formal and non-formal sectors separately.

Gender wage gaps are explained primarily by the treatment component, which is positive and larger than the endowment component (table 2). In fact, the endowment component is negative and significant across all specifications, and generally increasing over time. This indicates that women have consistently better human capital endowments than men, which would lead us to expect a negative gender wage gap in the absence of any treatment effect. These findings are in comport with results from similar research conducted by Marquez Garcia, Ñopo and Salardi (2009) who employ a non-parametric matching method and similarly find that the unexplained component was dominant (see also Ñopo, 2012: 168).

If we look at the impact of changes in the specification, revealed in subsequent columns, we see that adding controls for occupational structure has a significant impact in 1987, but that this impact has declined significantly by 2006. Focusing on the first row of table 2, the endowment component in 1987 increases by 0.11 log points from column (1) to column (2), while in 2006 it increases by only 0.03 log points. The treatment component decreases by almost exactly the same amount. We see similar patterns when we control for female occupational intensity (in column (3)), indicating that the occupational structure (proxied by occupational dummies and female occupational intensity) was a more important explanation of gender differentials in earlier years, likely reflecting the higher degree of occupational segregation in the labour market in this earlier time period.

In table 3 we repeat the same exercise for racial wage gaps. In this case both the endowment and treatment components are positive and significant across all

[83] The t-tests for the increase over time in the negative impact of non-white occupational intensity are equal to 3.244 and 4.65 for non-white and white wages, respectively.

specifications, but it is the endowment effect that exerts the dominant role in explaining racial differentials. Again, Marquez Garcia, Ñopo and Salardi (2009) report similar results when using a non-parametric matching methodology (see also Ñopo 2012: 273). Adding controls for occupational structure alters the results slightly, with a small increase in the endowment effect and negligible decrease in the treatment effect. This is perhaps not surprising given that white and non-white workers are distributed relatively homogenously across occupations (see the discussion in previous chapters).

In both tables we see important changes over time. Most notably, gender wage gaps have decreased considerably faster than racial wage gaps. The gender differential in the entire labour market moved from approximately 0.32 in 1987 to 0.06 in 2006, amounting to a decrease of roughly 0.26 log points, while the racial differential moved from 0.49 to 0.41, amounting to a decrease of only 0.08 log points. Thus, the gender wage gaps contracted by 80%, while the racial wage gaps only fell by 16% over this period.

While racial differentials were roughly 35% greater than gender differentials in 1987 (0.49 vs. 0.32) the much faster decline in gender wage gaps meant that by 2006 racial wage gaps were 85% larger than gender wage gaps (0.41 vs. 0.061). Moreover, while a decreasing treatment component has been the primary cause of the rapid decrease in gender differentials over time, the smaller decline in racial gaps has been driven by changes in the endowment component, while the treatment component remains largely unchanged. The rapid decrease in the gender wage gap is mainly due to changes in wage structure (which includes, among other unobserved factors, discriminatory behaviour) while the small decrease in racial wage gaps is almost entirely due to improved characteristics, such as educational attainment, among non-white individuals.

Table 2: Oaxaca-Blinder wage decomposition across different specifications, year 1987 and 2006 – GENDER WAGE GAPS

	Year 1987									Year 2006							
	(1)		(2)		(3)		(4)			(1)		(2)		(3)		(4)	
	1st spec.	t-test	2nd spec.	t-test	3rd spec.	t-test	4th spec.	t-test		1st spec.	t-test	2nd spec.	t-test	3rd spec.	t-test	4th spec.	t-test
All labour market									**All labour market**								
Endowment effect	-0.163	***	-0.049	***	-0.050	***	-0.071	***	**Endowment effect**	-0.182	***	-0.152	***	-0.174	***	-0.156	***
s.e.	0.0048		0.0069		0.0075		0.0083		**s.e.**	0.0031		0.0041		0.0046		0.005	
Treatment effect	0.485	***	0.371	***	0.373	***	0.393	***	**Treatment effect**	0.243	***	0.213	***	0.235	***	0.216	***
s.e.	0.0051		0.0068		0.0076		0.0082		**s.e.**	0.0038		0.0042		0.005		0.005	
Total gap	0.322	***	0.322	***	0.322	***	0.322	***	**Total gap**	0.061	***	0.061	***	0.061	***	0.061	***
s.e.	0.007		0.007		0.007		0.007		**s.e.**	0.0046		0.0046		0.0046		0.0046	
Formal sector									**Formal sector**								
Endowment effect	-0.177	***	-0.058	***	-0.064	***	-0.068	***	**Endowment effect**	-0.188	***	-0.166	***	-0.168	***	-0.152	***
s.e.	0.0064		0.0091		0.0088		0.0099		**s.e.**	0.004		0.0054		0.0053		0.0059	
Treatment effect	0.413	***	0.294	***	0.300	***	0.304	***	**Treatment effect**	0.240	***	0.218	***	0.220	***	0.204	***
s.e.	0.0066		0.0086		0.0087		0.0095		**s.e.**	0.0047		0.0052		0.006		0.0058	
Total gap	0.236	***	0.236	***	0.236	***	0.236	***	**Total gap**	0.052	***	0.052	***	0.052	***	0.052	***
s.e.	0.0088		0.0088		0.0088		0.0088		**s.e.**	0.0057		0.0057		0.0057		0.0057	
Informal sector									**Informal sector**								
Endowment effect	-0.270	***	-0.163	***	-0.095	***	-0.157	***	**Endowment effect**	-0.182	***	-0.157	***	-0.162	***	-0.126	***
s.e.	0.008		0.0147		0.0148		0.0166		**s.e.**	0.0047		0.0076		0.0087		0.0103	
Treatment effect	0.461	***	0.353	***	0.286	***	0.347	***	**Treatment effect**	0.211	***	0.187	***	0.191	***	0.156	***
s.e.	0.0096		0.0151		0.0156		0.017		**s.e.**	0.0069		0.0085		0.0101		0.011	
Total gap	0.191	***	0.191	***	0.191	***	0.191	***	**Total gap**	0.029	***	0.029	***	0.029	***	0.029	***
s.e.	0.012		0.012		0.012		0.012		**s.e.**	0.0077		0.0077		0.0077		0.0077	
Self-employed sector									**Self-employed sector**								
Endowment effect	-0.099	***	-0.122	***	-0.150	***	-0.187	***	**Endowment effect**	-0.193	***	-0.157	***	-0.215	***	-0.185	***
s.e.	0.009		0.0161		0.0185		0.0228		**s.e.**	0.0066		0.0093		0.0105		0.0118	
Treatment effect	0.558	***	0.581	***	0.609	***	0.645	***	**Treatment effect**	0.273	***	0.237	***	0.295	***	0.265	***
s.e.	0.0121		0.0181		0.0204		0.0244		**s.e.**	0.0101		0.0118		0.0133		0.0138	
Total gap	0.458	***	0.458	***	0.458	***	0.458	***	**Total gap**	0.080	***	0.080	***	0.080	***	0.080	***
s.e.	0.0152		0.0152		0.0152		0.0152		**s.e.**	0.0116		0.0116		0.0116		0.0116	

Source: Author's computations using PNAD 1987 and 2006.
Note: Columns (1), (2), (3) and (4) correspond to the different specifications of the wage equation, as explained in section 5.1.

Table 3: Oaxaca-Blinder wage decomposition across different specifications, year 1987 and 2006 – RACIAL WAGE GAPS

	Year 1987									Year 2006							
	(1)		(2)		(3)		(4)			(1)		(2)		(3)		(4)	
	1st spec.	t-test	2nd spec.	t-test	3rd spec.	t-test	4th spec.	t-test		1st spec.	t-test	2nd spec.	t-test	3rd spec.	t-test	4th spec.	t-test
All labour market									**All labour market**								
Endowment effect	0.384	***	0.399	***	0.409	***	0.401	***	**Endowment effect**	0.320	***	0.338	***	0.353	***	0.344	***
s.e.	0.0055		0.0055		0.0055		0.0055		**s.e.**	0.0039		0.0039		0.0039		0.0039	
Treatment effect	0.105	***	0.091	***	0.080	***	0.088	***	**Treatment effect**	0.093	***	0.075	***	0.059	***	0.068	***
s.e.	0.0056		0.0054		0.0055		0.0054		**s.e.**	0.0043		0.0041		0.0042		0.0041	
Total gap	0.489	***	0.489	***	0.489	***	0.489	***	**Total gap**	0.413	***	0.413	***	0.413	***	0.413	***
s.e.	0.0064		0.0064		0.0064		0.0064		**s.e.**	0.0045		0.0045		0.0045		0.0045	
Formal sector									**Formal sector**								
Endowment effect	0.262	***	0.288	***	0.297	***	0.291	***	**Endowment effect**	0.210	***	0.236	***	0.247	***	0.240	***
s.e.	0.0071		0.0072		0.0072		0.0072		**s.e.**	0.0048		0.0048		0.0049		0.0049	
Treatment effect	0.108	***	0.082	***	0.073	***	0.079	***	**Treatment effect**	0.102	***	0.075	***	0.064	***	0.071	***
s.e.	0.0071		0.0066		0.0069		0.0066		**s.e.**	0.0049		0.0045		0.0047		0.0045	
Total gap	0.370	***	0.370	***	0.370	***	0.370	***	**Total gap**	0.311	***	0.311	***	0.311	***	0.311	***
s.e.	0.0082		0.0082		0.0082		0.0082		**s.e.**	0.0055		0.0055		0.0055		0.0055	
Informal sector									**Informal sector**								
Endowment effect	0.376	***	0.378	***	0.396	***	0.378	***	**Endowment effect**	0.259	***	0.285	***	0.293	***	0.288	***
s.e.	0.0098		0.0099		0.0099		0.0099		**s.e.**	0.0063		0.0064		0.0064		0.0064	
Treatment effect	0.042	***	0.039	***	0.021	**	0.040	***	**Treatment effect**	0.074	***	0.048	***	0.040	***	0.045	***
s.e.	0.0094		0.0089		0.0093		0.0089		**s.e.**	0.0077		0.0073		0.0074		0.0073	
Total gap	0.418	***	0.418	***	0.418	***	0.418	***	**Total gap**	0.333	***	0.333	***	0.333	***	0.333	***
s.e.	0.0115		0.0115		0.0115		0.0115		**s.e.**	0.0079		0.0079		0.0079		0.0079	
Self-employed sector									**Self-employed sector**								
Endowment effect	0.336	***	0.350	***	0.352	***	0.350	***	**Endowment effect**	0.354	***	0.362	***	0.381	***	0.371	***
s.e.	0.0112		0.0113		0.0112		0.0113		**s.e.**	0.0087		0.0088		0.0087		0.0088	
Treatment effect	0.150	***	0.136	***	0.133	***	0.136	***	**Treatment effect**	0.104	***	0.096	***	0.078	***	0.087	***
s.e.	0.0134		0.013		0.0133		0.013		**s.e.**	0.011		0.0106		0.0109		0.0107	
Total gap	0.486	***	0.486	***	0.486	***	0.486	***	**Total gap**	0.458	***	0.458	***	0.458	***	0.458	***
s.e.	0.0137		0.0137		0.0137		0.0137		**s.e.**	0.0105		0.0105		0.0105		0.0105	

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

Note: Columns (1), (2), (3) and (4) correspond to the different specifications of the wage equation, as explained in section 5.1.

When we disaggregate the results between the formal, informal and self-employed sectors some additional insights emerge. With respect to gender differentials, wage gaps in the formal sector are somewhat wider than in the informal sector, but have also decreased somewhat faster. These changes have also occurred slightly differently across the two sectors. Gender gaps in the informal sector have benefitted from a larger decrease in the wage structure effect, while the advantage in endowments held by women in the informal sector has declined over time. Looking at racial differentials, we see, by contrast, that they are wider in the informal sector, where they have also decreased more sharply over time, driven primarily by a declining endowment effect. Finally, the self-employed sector exhibits the highest differential for both gender and race, reflecting the highest degree of heterogeneity across occupations.

Trends over time are revealed in greater detail in figure 2, which traces the evolution of wage gaps, and their components, at five-year intervals. While gender wage gaps have decreased consistently, panel B of figure 2 reveals a modestly inverted U-shaped pattern for racial gaps, as there was an increase in racial differentials in the 1990s. This is in line with worsening racial occupational segregation during the same period, as reported in the previous chapter.

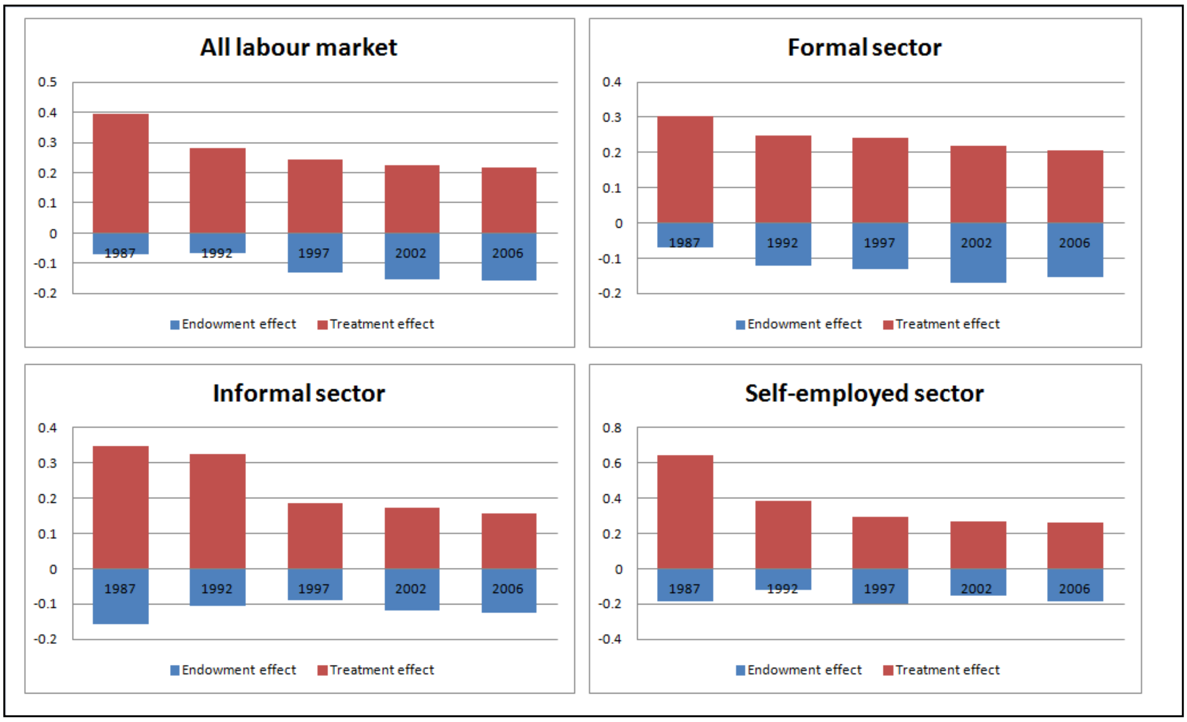
In sum, the standard OB decomposition reveals that gender differentials are explained primarily by the treatment effects (i.e., differences in the wage structure or unexplained component), while racial differentials are explained by the endowment effects (i.e., differences in observed characteristics or explained component). These broad empirical findings are in line with similar empirical studies for Brazil, including Soares (2000) and Marquez Garcia, Ñopo and Salardi (2009). Gender pay gaps are smaller than racial pay gaps and have decreased much faster than racial wage gaps over time, with this rapid decrease primarily due to changes in wage structure, while the small decrease in racial pay gaps is due entirely to the role of endowment effects. When we disaggregate the analysis into the formal and non-formal sectors we discover that gender gaps are greater in the formal than in the informal sector, while racial gaps are greater in the informal than the formal sector.

These differences between the formal and informal sectors are particularly interesting and we thus conclude this section with a discussion of what may be driving these results. Interestingly, while the total wage gaps are larger in the formal sector by gender and in the informal sector by race the seemingly divergent results reflect a common underlying pattern. For both race and gender the treatment component of wage gaps is larger in the formal sector, while for both race and gender the endowments component is larger in the informal sector. The total gap is larger in the formal sector by gender because the positive difference in the treatment effects between the formal and informal sectors is larger than the negative difference in the endowment effects. The reverse is true by race, which explains the higher total gap in the informal sector. These divergent weights on the treatment and endowment effects are, in turn, not necessarily surprising, as gender wage gaps are, in aggregate, primarily explained by treatment effects, while racial gaps are primarily the result of endowment effects.

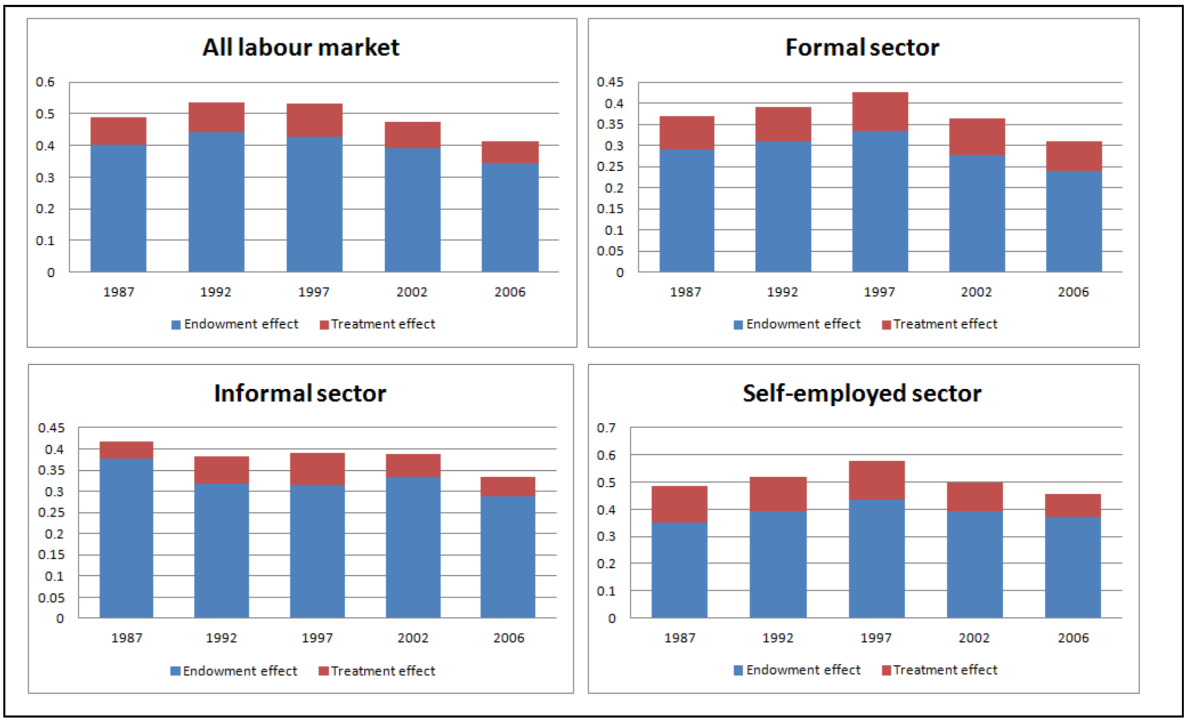
Given these underlying patterns, the more interesting question is not the distinction between gender and race, but what explains the fact that the treatment component is consistently higher in the formal sector, while the endowment component is consistently higher in the informal sector. Turning first to the endowment component, the most intuitive explanation is that this reflects the diversity of the informal sector, which is home to the majority of low-skilled female and non-white workers. We noted evidence in the previous chapter that low-skilled female and non-white workers appear to face barriers to entering the formal sector, and the findings here are consistent with such a pattern, as female and non-white endowments lag further behind in this sector. The existence of larger treatment components in the formal sector is not as intuitive a pattern, but tells us that unexplained wage discrimination is consistently higher in the formal sector. Taken together these patterns paint a somewhat troubling picture, suggesting that female and non-white workers face barriers to entering the formal sector (reflected in higher endowment components in the informal sector) and, once there, still face comparatively high levels of unexplained wage discrimination (reflected in higher treatment components in the formal sector).

Figure 2: Oaxaca-Blinder decomposition over time

PANEL A – Gender wage gaps



PANEL B – Racial wage gaps



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

6.5.3 The Brown, Moon and Zoloth (1980) decomposition results

Tables 4 and 5 present the results when employing the BMZ decomposition technique. Since this decomposition takes occupational segregation into account we now have four decomposition components, two intra-occupational components and two inter-occupational components. Alongside the results using the BMZ decomposition technique, we report at the bottom of each panel the earlier results when employing the 4th OB decomposition specification, which included controls for occupational structure. This allows us to compare the results using the two techniques.

Panel A of table 4 reports the gender wage gap results for the entire labour market. We find that the positive gender wage gap is almost entirely driven by the unexplained intra-occupational component, while this component has declined over time. This tells us that gender differentials are explained primarily by differences in the returns to observed characteristics within each occupation, which we have termed "vertical segregation". The explained inter-occupational component also figures prominently in the results, and becomes increasingly negative over time. This tells us that differences in endowments exist primarily across occupations and favour female workers, implying that female workers should gain higher wages as a result of their occupational allocation. Interestingly, the male advantage in endowments within occupations (given by the explained intra-occupational component) is small and disappears over time. Finally, the unexplained inter-occupational component is small, though not negligible, and decreases over time and has become mildly negative in recent years. This captures the difference in the structure of occupational achievement, which we have termed "horizontal segregation".

Table 4: Brown, Moon and Zoloth (1980) decomposition for gender wage gaps

PANEL A – ALL LABOUR MARKET

	1987	t-test	1992	t-test	1997	t-test	2002	t-test	2006	t-test
Explained intra-occupational component	0.023	***	0.041	***	-0.019	***	-0.021	***	-0.017	***
s.e.	0.00397		0.003842		0.00277		0.001727		0.001511	
Unexplained intra-occupational component	0.370	***	0.296	***	0.247	***	0.218	***	0.205	***
s.e.	0.007708		0.007789		0.006281		0.004894		0.004402	
Explained inter-occupational component	-0.110	***	-0.123	***	-0.147	***	-0.128	***	-0.126	***
s.e.	0.000901		0.001001		0.000959		0.000815		0.000796	
Unexplained inter-occupational component	0.039	***	-0.004		0.032	***	-0.004	***	-0.005	*
s.e.	0.005397		0.005464		0.004411		0.002882		0.00247	
Oaxaca-Blinder decomposition										
Explained component	-0.071	***	-0.068	***	-0.131	***	-0.154	***	-0.156	***
s.e.	0.008306		0.008102		0.007004		0.005602		0.004984	
Unexplained component	0.393	***	0.279	***	0.245	***	0.223	***	0.216	***
s.e.	0.008227		0.00805		0.00683		0.005548		0.005012	
Total gap	0.322	***	0.212	***	0.114	***	0.069	***	0.060	***
s.e.	0.006953		0.006639		0.00591		0.00518		0.004626	

PANEL B – FORMAL SECTOR

	1987	t-test	1992	t-test	1997	t-test	2002	t-test	2006	t-test
Explained intra-occupational component	-0.012	***	-0.032	***	-0.055	***	-0.037	***	-0.037	***
s.e.	0.003587		0.003379		0.002352		0.001965		0.001554	
Unexplained intra-occupational component	0.281	***	0.250	***	0.258	***	0.220	***	0.210	***
s.e.	0.008653		0.009652		0.008058		0.006081		0.005191	
Explained inter-occupational component	-0.088	***	-0.132	***	-0.136	***	-0.114	***	-0.106	***
s.e.	0.001364		0.001729		0.001416		0.001083		0.000956	
Unexplained inter-occupational component	0.056	***	0.042	***	0.041	***	-0.021	***	-0.014	***
s.e.	0.007186		0.008148		0.006883		0.00388		0.003148	
Oaxaca-Blinder decomposition										
Explained component	-0.068	***	-0.121	***	-0.132	***	-0.170	***	-0.152	***
s.e.	0.009935		0.0102		0.009		0.0071		0.0059	
Unexplained component	0.304	***	0.249	***	0.240	***	0.218	***	0.204	***
s.e.	0.009546		0.0103		0.0088		0.0068		0.0058	
Total gap	0.236	***	0.128	***	0.108	***	0.048	***	0.052	***
s.e.	0.008771		0.008		0.0077		0.0066		0.0057	

PANEL C – INFORMAL SECTOR

	1987	t-test	1992	t-test	1997	t-test	2002	t-test	2006	t-test
Explained intra-occupational component	-0.009		-0.060	***	-0.052	***	-0.073	***	-0.073	***
s.e.	0.006466		0.007361		0.004417		0.003004		0.002897	
Unexplained intra-occupational component	0.367	***	0.334	***	0.203	***	0.196	***	0.189	***
s.e.	0.015776		0.018021		0.012502		0.009448		0.008617	
Explained inter-occupational component	-0.170	***	-0.059	***	-0.076	***	-0.084	***	-0.091	***
s.e.	0.002826		0.001354		0.001204		0.001127		0.00124	
Unexplained inter-occupational component	0.003		0.000		0.018	*	0.011	*	0.003	
s.e.	0.014056		0.014378		0.010352		0.006991		0.006047	
Oaxaca-Blinder decomposition										
Explained component	-0.157	***	-0.107	***	-0.091	***	-0.120	***	-0.126	***
s.e.	0.0166		0.018		0.014		0.0111		0.0103	
Unexplained component	0.347	***	0.324	***	0.185	***	0.172	***	0.156	***
s.e.	0.017		0.0193		0.0148		0.0117		0.011	
Total gap	0.191	***	0.217	***	0.094	***	0.052	***	0.029	***
s.e.	0.012		0.0109		0.0093		0.0084		0.0077	

PANEL D – SELF-EMPLOYED SECTOR

	1987	t-test	1992	t-test	1997	t-test	2002	t-test	2006	t-test
Explained intra-occupational component	0.088	***	0.000		-0.022	***	-0.040	***	-0.062	***
s.e.	0.00953		0.006431		0.004457		0.003032		0.003372	
Unexplained intra-occupational component	0.588	***	0.397	***	0.290	***	0.258	***	0.249	***
s.e.	0.019407		0.017401		0.014591		0.012612		0.011719	
Explained inter-occupational component	-0.101	***	-0.103	***	-0.174	***	-0.147	***	-0.157	***
s.e.	0.00207		0.002083		0.002279		0.002193		0.002367	
Unexplained inter-occupational component	-0.142	***	-0.044	***	-0.017	**	0.017	***	0.020	***
s.e.	0.014266		0.011878		0.009016		0.006509		0.005975	
Oaxaca-Blinder decomposition										
Explained component	-0.187	***	-0.119	***	-0.201	***	-0.151	***	-0.185	***
s.e.	0.0228		0.0213		0.0178		0.0127		0.0118	
Unexplained component	0.645	***	0.387	***	0.297	***	0.270	***	0.265	***
s.e.	0.0244		0.0233		0.0191		0.0148		0.0138	
Total gap	0.458	***	0.268	***	0.097	***	0.119	***	0.080	***
s.e.	0.0152		0.0146		0.0139		0.0126		0.0116	

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

Table 5: Brown, Moon and Zoloth (1980) decomposition for racial wage gaps

PANEL A – ALL LABOUR MARKET

	1987	t-test	1992	t-test	1997	t-test	2002	t-test	2006	t-test
Explained intra-occupational component	0.159	***	0.223	***	0.219	***	0.201	***	0.169	***
s.e.	0.00278		0.003388		0.002851		0.002442		0.002178	
Unexplained intra-occupational component	0.089	***	0.095	***	0.106	***	0.085	***	0.068	***
s.e.	0.005292		0.005812		0.004938		0.004453		0.004059	
Explained inter-occupational component	0.198	***	0.168	***	0.165	***	0.151	***	0.141	***
s.e.	0.001219		0.001126		0.001033		0.000954		0.000941	
Unexplained inter-occupational component	0.043	***	0.053	***	0.045	***	0.042	***	0.040	***
s.e.	0.00053		0.000582		0.000568		0.000475		0.000449	
Oaxaca-Blinder decomposition										
Explained component	0.401	***	0.440	***	0.429	***	0.392	***	0.344	***
s.e.	0.0055		0.0055		0.005		0.0044		0.0039	
Unexplained component	0.088	***	0.097	***	0.103	***	0.082	***	0.068	***
s.e.	0.0054		0.0058		0.005		0.0045		0.0041	
Total gap	0.489	***	0.537	***	0.532	***	0.474	***	0.413	***
s.e.	0.0064		0.0062		0.0056		0.0049		0.0045	

PANEL B – FORMAL SECTOR

	1987	t-test	1992	t-test	1997	t-test	2002	t-test	2006	t-test
Explained intra-occupational component	0.082	***	0.116	***	0.141	***	0.115	***	0.091	***
s.e.	0.003212		0.003711		0.003181		0.002611		0.002128	
Unexplained intra-occupational component	0.079	***	0.086	***	0.089	***	0.084	***	0.070	***
s.e.	0.006437		0.006817		0.005962		0.005009		0.004251	
Explained inter-occupational component	0.154	***	0.135	***	0.141	***	0.127	***	0.115	***
s.e.	0.001238		0.00116		0.001091		0.001007		0.000923	
Unexplained inter-occupational component	0.055	***	0.055	***	0.056	***	0.041	***	0.039	***
s.e.	0.000729		0.000692		0.000746		0.000622		0.000577	
Oaxaca-Blinder decomposition										
Explained component	0.291	***	0.311	***	0.334	***	0.278	***	0.240	***
s.e.	0.0072		0.0067		0.0066		0.0056		0.0049	
Unexplained component	0.079	***	0.080	***	0.092	***	0.085	***	0.071	***
s.e.	0.0066		0.0069		0.0061		0.0052		0.0045	
Total gap	0.370	***	0.391	***	0.425	***	0.363	***	0.311	***
s.e.	0.0082		0.0076		0.0073		0.0063		0.0055	

PANEL C – INFORMAL SECTOR

	1987	t-test	1992	t-test	1997	t-test	2002	t-test	2006	t-test
Explained intra-occupational component	0.134	***	0.189	***	0.185	***	0.178	***	0.144	***
s.e.	0.004052		0.005586		0.004515		0.003855		0.003524	
Unexplained intra-occupational component	0.048	***	0.066	***	0.076	***	0.058	***	0.041	***
s.e.	0.008526		0.010446		0.008139		0.007471		0.007206	
Explained inter-occupational component	0.210	***	0.093	***	0.102	***	0.115	***	0.112	***
s.e.	0.002388		0.001709		0.001703		0.001652		0.001676	
Unexplained inter-occupational component	0.026	***	0.035	***	0.028	***	0.038	***	0.038	***
s.e.	0.000939		0.001118		0.001039		0.000801		0.000877	
Oaxaca-Blinder decomposition										
Explained component	0.378	***	0.318	***	0.315	***	0.333	***	0.288	***
s.e.	0.0099		0.0089		0.0079		0.0071		0.0064	
Unexplained component	0.040	***	0.065	***	0.076	***	0.053	***	0.045	***
s.e.	0.0089		0.0106		0.0085		0.0076		0.0073	
Total gap	0.418	***	0.384	***	0.391	***	0.387	***	0.333	***
s.e.	0.0115		0.0107		0.0093		0.0084		0.0079	

PANEL D – SELF-EMPLOYED SECTOR

	1987	t-test	1992	t-test	1997	t-test	2002	t-test	2006	t-test
Explained intra-occupational component	0.220	***	0.275	***	0.281	***	0.268	***	0.250	***
s.e.	0.007503		0.00866		0.007309		0.006423		0.006034	
Unexplained intra-occupational component	0.126	***	0.123	***	0.149	***	0.107	***	0.087	***
s.e.	0.01305		0.013889		0.012091		0.011223		0.010593	
Explained inter-occupational component	0.071	***	0.044	***	0.089	***	0.056	***	0.067	***
s.e.	0.001543		0.001505		0.001774		0.001565		0.001619	
Unexplained inter-occupational component	0.035	***	0.052	***	0.028	***	0.034	***	0.025	***
s.e.	0.001533		0.001649		0.001363		0.001122		0.000995	
Oaxaca-Blinder decomposition										
Explained component	0.350	***	0.394	***	0.434	***	0.394	***	0.371	***
s.e.	0.0113		0.011		0.0106		0.0094		0.0088	
Unexplained component	0.136	***	0.126	***	0.145	***	0.105	***	0.087	***
s.e.	0.013		0.0137		0.0121		0.0112		0.0107	
Total gap	0.486	***	0.520	***	0.578	***	0.499	***	0.458	***
s.e.	0.0137		0.0127		0.0121		0.011		0.0105	

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

Panels B, C and D in table 4 repeat the same exercise disaggregated into the formal, informal and self-employed sectors. Three main findings emerge from this disaggregation. First, the informal sector is unique in that the explained inter-occupational component becomes less negative over time. The female advantage in endowments has been contracting within informal sector occupations over time, such that by 2006 it was roughly equal in magnitude to the pattern in the formal sector, similar to the results using the standard OB decomposition. Second, the unexplained intra-occupational component (vertical segregation) is initially higher in the informal sector but decreases more rapidly than in the formal sector. Finally, the unexplained inter-occupational component (horizontal segregation) is essentially absent in the informal sector.

Table 5 reports the BMZ decomposition results for racial differentials. In panel A of table 5, we see that racial wage gaps are explained primarily by the explained intra- and inter-occupational components. Interestingly, over time we see a transition from the explained inter-occupational to the intra-occupational component as the primary determinant of the wage gaps, as the former has declined somewhat over time, while the later has, if anything, increased slightly. Although the occupational distribution for non-white and white workers has thus become more homogenous over time, there remain important differences in endowments, which account for persistent differences in occupational attachment. While horizontal and vertical segregation are a more modest component of the total gap (on average about 25% of the total) they have remained stable over time.

When we disaggregate the decomposition into the formal, informal and self-employed sectors we observe slightly divergent patterns. The informal sector most closely mirrors trends for the entire labour market, whereas in the formal sector the inter-occupational explained component plays a more prominent role, while in the self-employed sector the intra-occupational explained component is important. While the unexplained components are relatively small across the sectors, it is worth noting a modest increase in the inter-occupational unexplained component (horizontal segregation) within the informal sector.

To summarize, we find that gender wage gaps are explained primarily by vertical segregation (unexplained intra-occupational component), though this component has steadily decreased over time. Differences in observed characteristics occur primarily across occupations and tend to favour female workers, which, if not for vertical segregation, would be expected to result in higher female wages. Finally, horizontal segregation appears to be a minor issue and over time modestly favours female workers. These trends are broadly replicated across the formal and non-formal sectors, though we do find that vertical segregation has decreased faster in the informal sector, while horizontal segregation has acted more strongly in favour of females in the formal sector.

In the case of racial wage gaps, we find that they are driven primarily by differences in observed characteristics both within and across occupations. The small overall improvement over time is almost entirely explained by decreasing differences across occupations, while observed advantages for white individuals remain high and stable within occupations. Both horizontal and vertical segregation are relatively small, but are more persistent over time. In particular, horizontal segregation has decreased over time in the formal sector but has increased slightly in the informal sector.

The finding that gender wage gaps reflect sizeable vertical and horizontal segregation is in line with previous studies of Brazil. Most notably, in a study of Pernanbuco and São Paulo Ometto, Hoffmann and Alves (1999) found that the unexplained intra-occupational component was dominant for Pernanbuco while both the unexplained intra- and inter-occupational components were sizeable for São Paulo. Turning to racial pay gaps, Arcand and D'Hombres (2004) found a segregation component of 5-8% of the entire racial wage gap, depending on whether they were considering 'brown' or 'black' workers in the comparison to white wages. Our estimates are larger, as we find that total segregation (the sum of horizontal and vertical segregation) is equal on average to 25%. Our much larger estimates are likely to reflect the fact that we consider both female and male workers, while Arcand and D'Hombres (2004) restrict their study to racial wage gaps among males only.

We can similarly link our findings about the unexplained intra- and inter-occupational components of wage discrimination to findings about occupational segregation from the previous chapter. Beginning with gender wage gaps, we find sizeable vertical segregation (the unexplained intra-occupational component) that has declined significantly over time and this mirrors the high level of gender occupational segregation, and its significant decline over time, reported in the previous chapter. Turning to racial wage gaps, we find negligible unexplained intra- and inter-occupational components, and the comparatively small magnitude of both vertical and horizontal segregation is consistent with the comparatively modest level of racial occupational segregation reported in the previous chapter. At the same time, the negligible decline in unexplained racial wage gaps over time follows a similar trend to the persistence of racial occupational segregation highlighted in the previous chapter. Finally, and particularly strikingly, the increase in horizontal segregation over time in the informal sector, which can be seen in table 5, adds additional information to our finding in the previous chapter of increasing occupational segregation in the informal sector during the 1990s.

An alternative way of linking segregation to our current findings from the BMZ decomposition is to adopt an approach used by Reilly (1991) and compare the levels of occupational segregation measured using the actual and the counterfactual occupational distribution. To do that we need to re-compute the Duncan index using the counterfactual occupational distribution adopted here in order to decompose pay gaps using the BMZ decomposition technique. Then, we compare these counterfactual values with the original ones provided in table 1 of chapter 5.

As already explained in section 6.4.1, the counterfactual occupational distribution is obtained by fitting the estimated male coefficients from the occupational attainment MNL model to the female realizations of the explanatory variables. The Duncan index for gender segregation falls from 0.605 (0.565) in 1987 (2006) to 0.087 (0.099) in 2006. Thus, we observe a decline of gender segregation roughly equal to 92%. For racial segregation, the Duncan index drops from 0.199 (0.191) to 0.137 (0.123) in 1987(2006) reflecting a reduction of roughly one-third.

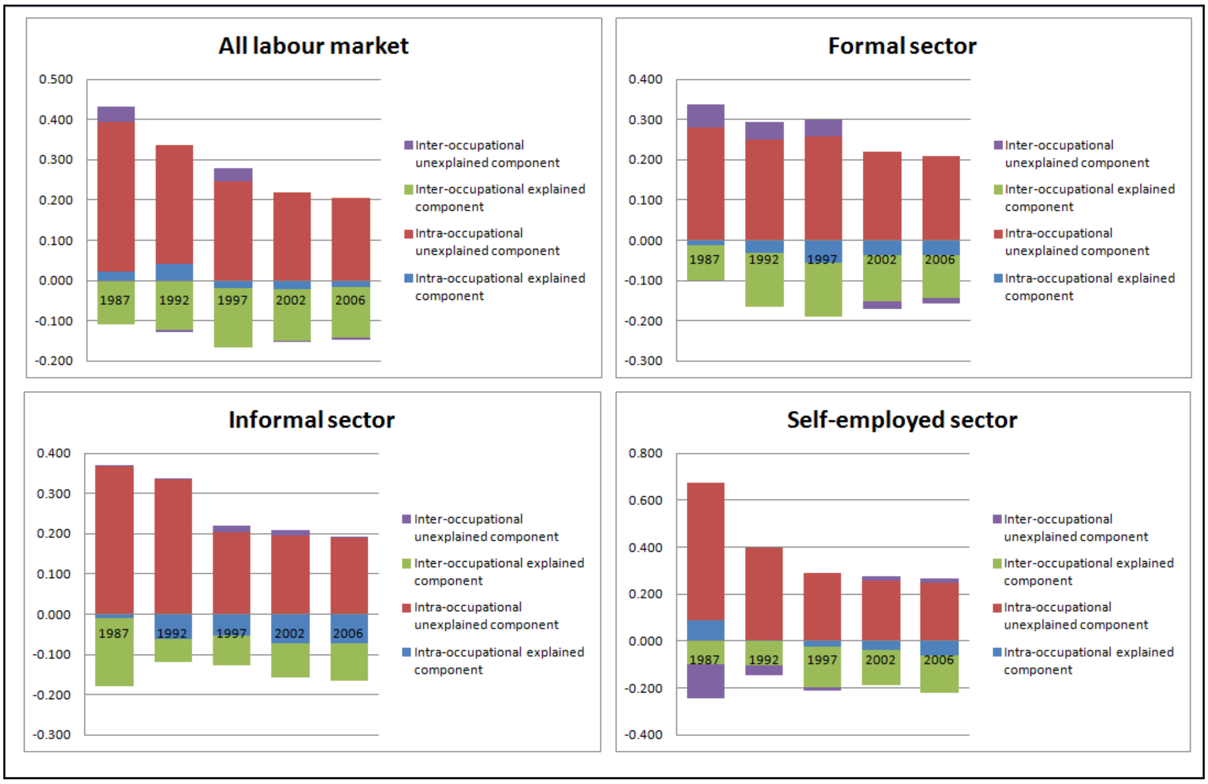
The greater decline observed for gender segregation when using the counterfactual occupational distribution confirms, using a different method, that gender segregation is mainly attributable to unequal treatment. In contrast, the smaller decline of racial segregation exhibited when using the counterfactual occupational distribution corroborates that racial segregation is mainly attributable to differentials in the set of observable characteristics.

Finally, if we compare the results using the BMZ decomposition technique to those using the standard OB technique we note some modest differences, while the overall trends remain largely unchanged. In regard to gender wage gaps, the BMZ technique yields a slightly larger unexplained component during the first half of the time period, but this trend is reversed in more recent years, with the unexplained component is smaller than for the OB technique. Turning to racial wage gaps, the BMZ technique yields a total unexplained component that is generally 30-40% greater than the OB technique over time. These differences, while small, are far from negligible, though the more obvious benefit of the BMZ technique lies in its ability to distinguish inter- and intra-occupational wage gap effects.

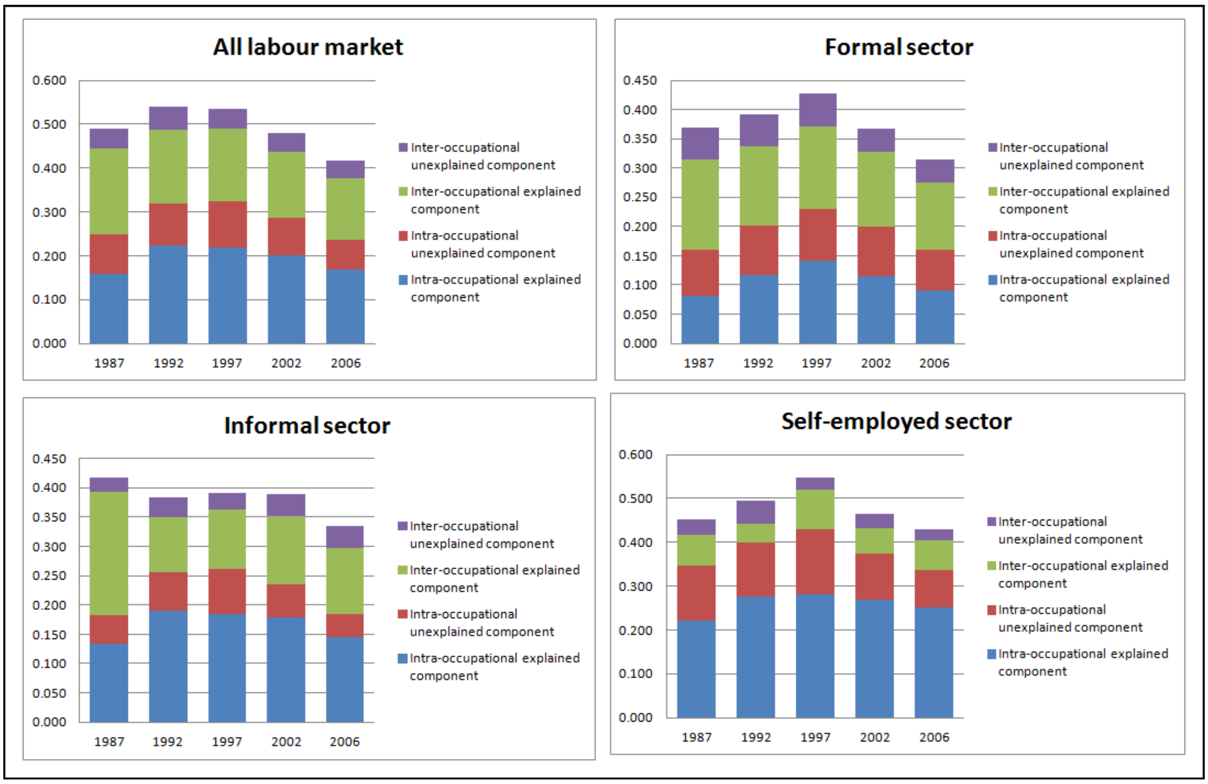
In summary, gender wage gaps are primarily explained by the treatment component. The BMZ decomposition results further reveal that this treatment component is driven by the intra-occupational component, with gender wage gaps thus primarily explained by the presence of vertical segregation. Differences in endowments favour women and are primarily inter-occupational differences. By contrast, racial wage gaps are primarily explained by the endowment component, while the BMZ decomposition results reveal that it is the inter-occupational component of the endowment component that has been declining somewhat over time. Both vertical and horizontal segregation are comparatively modest but still account for about 25% of the total gap, on average, and, more importantly, have remained stable over time.

Figure 3: Brown, Moon and Zoloth (1980) decomposition over time

PANEL A – Gender wage gaps



PANEL B – Racial wage gaps



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

6.5.4 Sensitivity checks

In addition to the core results we now perform several sensitivity checks in relation to the standard OB decomposition in order to investigate whether this decomposition is sensitive to certain choices that we have made in terms of equation specification and sample. The results are presented in table 6 and in figures 4 and 5.

We first explore the sensitivity of the results to the use of the log of hourly wages as the dependent variable. We find that if we adopt monthly wages as the dependent variable the gender wage gap increases by roughly 0.22-0.23 log points across all years, while the treatment component disappears in earlier years. By contrast, the racial wage gap remains unchanged when using the alternative dependent variable. We then explore what happens if we retain only workers who are employed full-time.[84] Again the gender wage gap increases, this time by 0.04-0.10 log points, while the racial wage gaps increases by a modest 0.02 log points. In both cases a likely explanation for increased gender wage gaps is the distinction between full-time and part-time workers, with women working full-time more likely to be low-wage earners.

Finally, we explore the impact of adding a work experience variable. As noted earlier, the PNADs began to collect work experience data only in the 1990s, and as such this analysis can only be conducted for those years. We consider two different variables for work experience: the number of years in the current job and the number of years since the individual began to work. In both cases we add the work experience variable and its square to the original specification, while also retaining the age and age squared variables.[85] While the age variable may capture general labour market experience over a lifetime, the job tenure variable in particular captures potentially more relevant firm specific job experience. Ultimately, the results, reported in table 6 and in panel B of figures 4 and 5, are essentially unchanged, implying that the exclusion of this variable does not alter our results.

---

[84]Following the OECD definition (http://stats.oecd.org/glossary/detail.asp?ID=3046), a part-time employee is someone who works less than 30-35 hours per week in their primary job. According to ILO data from 2001 to 2009 the incidence of part-time employment has been equal to 17%, on average, for the entire workforce, but higher for females, at roughly 28%. Among part-time workers females are the large majority: the gender ratio for the part-time workforce is 0.68, on average (ILO (2011), The Key Indicator of the Labour Market (KILM) dataset available at http://kilm.ilo.org/kilmnet/). If we explore this feature in our PNAD datasets, we find that the large majority of part-time workers are not in the formal sector and they tend to report being self-employed (in 2002, for example, 37% of part-time workers were self-employed). In Brazil part-time employment is not generally associated with a deliberate choice, and as such this analysis needs to be interpreted with caution, particularly in relation to drawing policy implication.

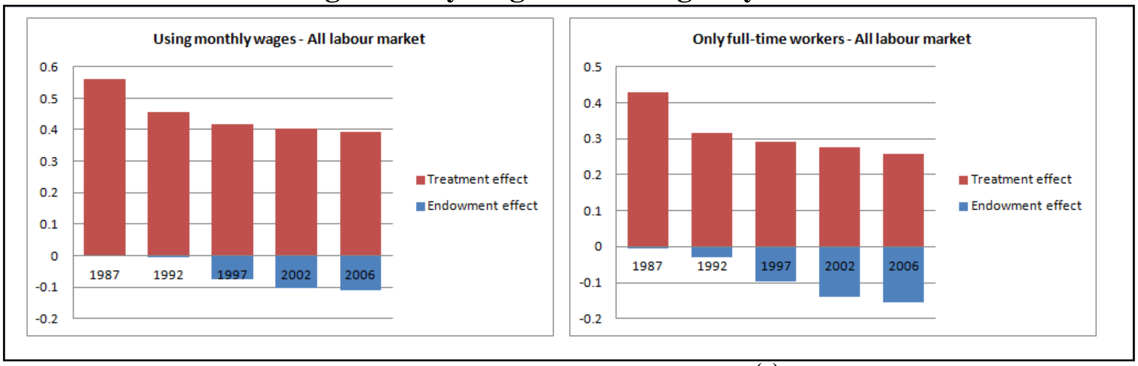
[85] By substituting age and age squared with the experience variable we obtain similar results as well.

To conclude, these sensitivity checks leave the main findings about the magnitude and trend of gender and racial pay gaps, drawn from the implementation of the OB and BMZ decomposition techniques, largely unchanged. Of greatest note, the results remain almost entirely unchanged when we include the experience variable for the years in which it is available. The only difference of note is in relation to gender wage gaps, where we find somewhat larger wage gaps when using monthly wages, or when we restrict the sample to full-time workers. A possible explanation for these results is that the average hourly wage for the excluded 'part-time' female workers is observed to be higher, on average, than for 'full-time' female workers. This may initially appear contradictory, given that part-time employment is generally perceived to be precarious in Brazil. However, we in fact notice a bimodal distribution of wages among 'part time' female workers, while the fact that one third of 'part-time' females report being self-employed is also consistent with this being a highly heterogeneous group. Ultimately, it appears that a significant portion of 'part-time' female workers are involved in highly qualified professions with high hourly pay, which explains how 'part-time' work may have higher average wages than 'full time' work, while also being home to larger numbers of women whose employment is comparatively precarious.

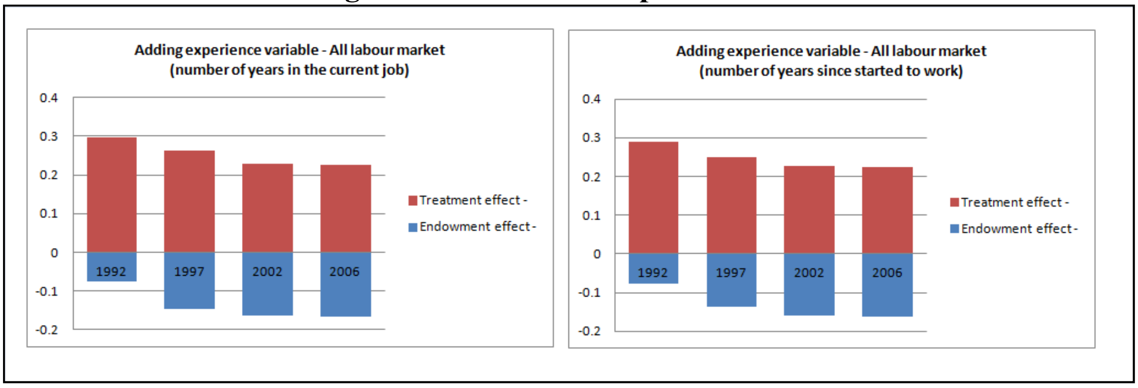
This is an interesting pattern, and quite different from the results of other studies that have looked at wage gaps between part-time and full-time workers. For example, Manning and Petrongolo (2008) have investigated the wage penalties for 'part time' female workers in Britain, finding that women working parti-time have average hourly earnings that are 22% lower than full-time female workers. They, in turn, attribute this wage penalty primarily to differences in occupational attachments and to limited occupational mobility, as women switching to part-time employment were forced to make a downward occupational move.

However, while these messages are of obvious importance in their own right, the key message for us is that even when we focus exclusively on full-time workers the two main conclusions from our analysis are entirely unchanged. First, gender wage gaps are still declining faster than racial wage gaps. Second, the decline in gender wage gaps is still driven by a declining treatment component, while the more modest decline in racial wage gaps continues to be attributable to the decline in the endowments component, with the treatment component remaining stable over time.

Figure 4: OB decomposition for gender wage gaps – Sensitivity checks
PANEL A – Check using monthly wages and using only full-time workers

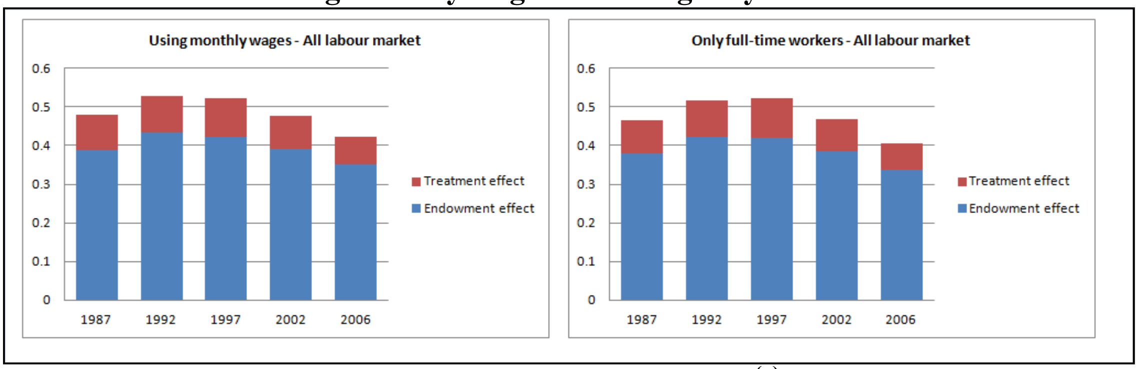


PANEL B – Check adding variable on work experience[(a)]

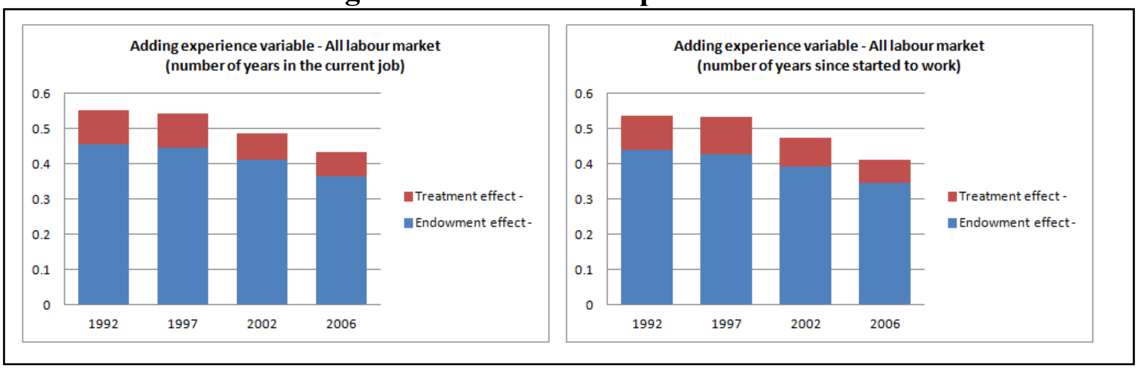


Source: Author's computations using PNAD 1987 – 1992 – 1997 – 2002 – 2006.
Note: (a) variables on experience are only available from the 1990s.

Figure 5: OB decomposition for racial wage gaps – Sensitivity checks
PANEL A – Check using monthly wages and using only full-time workers



PANEL B – Check adding variable on work experience[(a)]



Source: Author's computations using PNAD 1987 – 1992 – 1997 – 2002 – 2006.
Note: (a) variables on experience are only available from the 1990s.

Table 6: Oaxaca-Blinder decomposition – Sensitivity checks

PANEL A – GENDER WAGE GAPS

	1987	t-test	1992	t-test	1997	t-test	2002	t-test	2006	t-test
CHECK 1: using monthly wages										
Endowment effect	0.0033		-0.0031		-0.0735	***	-0.1033	***	-0.1106	***
s.e.	0.0082		0.0079		0.0069		0.0055		0.0049	
Treatment effect	0.5561	***	0.4567	***	0.4169	***	0.4022	***	0.393	***
s.e.	0.0084		0.008		0.0067		0.0055		0.0049	
Total gap	0.5595	***	0.4536	***	0.3434	***	0.299	***	0.2824	***
s.e.	0.0069		0.0066		0.0058		0.0052		0.0048	
CHECK 2: keeping only full-time workers										
Endowment effect	-0.0062		-0.0302	***	-0.0963	***	-0.1411	***	-0.1534	***
s.e.	0.0083		0.0082		0.0072		0.0059		0.0051	
Treatment effect	0.4276	***	0.3153	***	0.2924	***	0.2745	***	0.259	***
s.e.	0.008		0.0079		0.0068		0.0056		0.0049	
Total gap	0.4215	***	0.2851	***	0.1961	***	0.1334	***	0.1056	***
s.e.	0.0076		0.0074		0.0065		0.0057		0.005	
CHECK 3a: adding the variable on experience (number of years in the current job)(a)										
Endowment effect	-		-0.0743	***	-0.1474	***	-0.1634	***	-0.1652	***
s.e.	-		0.0095		0.0083		0.0066		0.0059	
Treatment effect	-		0.2954	***	0.2637	***	0.2285	***	0.2249	***
s.e.	-		0.0093		0.0079		0.0064		0.0058	
diff	-		0.2211	***	0.1163	***	0.0652	***	0.0597	***
s.e.	-		0.0078		0.007		0.0061		0.0054	
CHECK 3b: adding the variable on experience (number of years since started to work)(a)										
Endowment effect	-		-0.0773	***	-0.1367	***	-0.1589	***	-0.1623	***
s.e.	-		0.0082		0.0071		0.0057		0.005	
Treatment effect	-		0.2888	***	0.2505	***	0.2281	***	0.2228	***
s.e.	-		0.0081		0.0069		0.0056		0.0051	
Total gap	-		0.2115	***	0.1138	***	0.0692	***	0.0605	***
s.e.	-		0.0066		0.0059		0.0052		0.0046	

PANEL B – RACIAL WAGE GAPS

	1987	t-test	1992	t-test	1997	t-test	2002	t-test	2006	t-test
CHECK 1: using monthly wages										
Endowment effect	0.3895	***	0.4341	***	0.4231	***	0.3908	***	0.3505	***
s.e.	0.0056		0.0056		0.005		0.0044		0.0041	
Treatment effect	0.0906	***	0.0929	***	0.0994	***	0.0871	***	0.0719	***
s.e.	0.0055		0.0058		0.0049		0.0044		0.0041	
Total gap	0.4801	***	0.527	***	0.5225	***	0.478	***	0.4224	***
s.e.	0.0065		0.0062		0.0056		0.005		0.0046	
CHECK 2: keeping only full-time workers										
Endowment effect	0.3797	***	0.4226	***	0.4202	***	0.3867	***	0.3375	***
s.e.	0.0059		0.0059		0.0054		0.0047		0.0042	
Treatment effect	0.0854	***	0.0944	***	0.1017	***	0.0806	***	0.0676	***
s.e.	0.0056		0.006		0.0051		0.0046		0.0041	
Total gap	0.4651	***	0.517	***	0.5219	***	0.4673	***	0.4051	***
s.e.	0.0067		0.0065		0.0059		0.0052		0.0047	
CHECK 3a: adding the variable on experience (number of years in the current job)(a)										
Endowment effect	-		0.4534	***	0.4452	***	0.4112	***	0.3627	***
s.e.	-		0.0065		0.0059		0.0051		0.0046	
Treatment effect	-		0.0976	***	0.0964	***	0.0732	***	0.0686	***
s.e.	-		0.0067		0.0057		0.0051		0.0047	
Total gap	-		0.5511	***	0.5416	***	0.4844	***	0.4313	***
s.e.	-		0.0072		0.0065		0.0058		0.0052	
CHECK 3b: adding the variable on experience (number of years since started to work)(a)										
Endowment effect	-		0.4399	***	0.4274	***	0.3912	***	0.344	***
s.e.	-		0.0055		0.005		0.0043		0.0039	
Treatment effect	-		0.0972	***	0.1043	***	0.0823	***	0.0684	***
s.e.	-		0.0057		0.005		0.0045		0.0041	
Total gap	-		0.5371	***	0.5317	***	0.4736	***	0.4124	***
s.e.	-		0.0062		0.0056		0.0049		0.0045	

Source: Author's computations using PNAD 1987 – 1992 – 1997 – 2002 – 2006.
Note: (a) variables on experience are only available from the 1990s.

6.6 Results when correcting for selection

We now explore the impact on the results of correcting for potential selectivity bias. We explore two different parametric corrections, the Heckman (1979) and the Lee (1983) procedures, for both the OB and BMZ decomposition methods. While the Heckman procedure conceptualizes the participation process as a binary outcome between participation and non-participation, the Lee procedure permits the disaggregation of the participation decision into the formal and non-formal sectors. What follows begins with a description of the first stage estimation, the selection equation, after which we present our findings when the selection corrected decomposition methodologies are applied. We conclude with some additional robustness checks.

6.6.1 The selection equations

The specification of the selection equation is an important first step, as we need to identify instruments that affect the likelihood of participation but not the determination of wages. In general, an instrument is valid when i) it is relevant (the instrument needs to be related to the likelihood of participating), and ii) it is exogenous (the instrument must be related to wages only via participation). The failure of either condition can have serious implications for the empirical analysis. At the same time, it can be difficult to satisfy both conditions. In selecting an instrument it is further important to understand the particular features of the selection process within individual countries, and for this reason we look to existing studies that have estimated the selection process for the Brazilian labour market in order to choose appropriate instruments. The key identifiers that we select are a set of variables that describe the households, including number of children younger than 1 or 6 years, the presence of elderly individuals in the household, living with parents, being married, having a housekeeper and earning non-labour related income at the household level. [86] These

[86] In order to accurately choose these identifiers we have also reviewed several studies that have applied selection correction to the Mincerian equation for the Brazilian market. Birdsall and Berhman (1991) used the following instruments: spouse income, other household income, dummies for residing with parents, household head, spouse, presence of children less than 6, interaction of resides with parents and other household income and interaction of if has child aged 6 or less and number of dependents over 14. Stelcner et al (1992) used household size, number of children in the family, wife's income, husband's total earnings, number of rooms in the house, dummies for whether the husband is an employee, whether the wife works, whether the house is owned, presence of assets and transferred income. Tiefenthaler (1992) used number of children, number of sons over 13 and number of daughter over 13, dummies for

variables are employed in the Brazilian literature as instruments for the selection process because they capture demands on women's time within the household, which may prevent labour market participation (i.e., having children), and factors that may reduce the extent of household responsibilities and thus free women to participate in the labour market (i.e., having additional adult household members, domestic workers or additional income).

In the appendix, tables A4 and A5 report the results for the Heckman procedure. For female workers the estimated coefficient for the inverse of the Mills' ratio is found to be positive in 1987 but contracts over time and is statistically insignificant in 2006. It is important to briefly explain the meaning of these terms and their signs. The estimated coefficient for the selection term is the product of the standard deviation of the wage equation errors and the correlation coefficient between the unobservables determining participation status and those determining the hourly wage. As such, a positive coefficient implies a positive correlation between the unobservables that determine the participation decision and those that determine the wage levels. That is, women that are more likely to participate are also those that are more likely to earn higher wages. Alternatively, positive selection implies that lower-wage earners are disproportionately likely to be out of the labour market. In the case of positive selection the estimated wage offer is then the difference between the conditional wage value[87] and the sample selection effect. Put simply, when the selection effect is positive, the wage offer will be smaller than the observed wage. As such, if we control for the fact that "more able" women participate disproportionately, we would conclude that the average female wage in the absence of selection would be lower, resulting potentially in a higher estimate of the wage gap if selection effects were not relevant for men.

With this interpretation in mind, over time we have found a decrease in positive female selection effects. The most likely interpretation of this finding is that the increase in female labour market participation has involved proportionately more unskilled/semi-skilled female workers. Table A9 lends support to this interpretation, as

---

whether the house is owned and presence of unearned income. Loureiro, Carneiro and Sachsida (2004) used years of education and experience for the household head, dummies for head and spouse, for presence of children by age and gender and for presence of other family income. Carvalho, Neri and Silva (2006) used the number of children younger than five years old and dummies for being the household head or being the son/daughter of the household.

[87] The conditional value of wages is the expected value of the dependent variable conditional on the dependent variable being observed, namely the uncorrected observed wage (Dolton and Makepeace, 1987).

we see that female participation has increased primarily among women with moderate levels of education (i.e., between 1 and 10 years of education).

Turning to male workers we find a negative and significant selection coefficient in 1987 but the coefficient is no longer statistically significant by 2006, indicating the absence of any selection bias. This is a not surprising given the very high rate of male participation in the labour market. By contrast, the negative coefficient in 1987 is quite surprising, as the literature suggests that negative selection is rare among male workers. Looking more closely, the results in table A5 reveal similarly negative selection among non-white workers in 2006, despite the existence of positive selection among women. These two results together indicate that negative selection is driven primarily by non-white male workers, and this result is consistent across alternative specifications of the selection equation. Thus, the overall negative selection term for males in 1987 reflects negative selection among non-white males' and a negligible selection effect among white males. By 2006 this effect has disappeared due to a decline in negative selection for non-white males and a slight increase in positive selection among white males. Similarly, the negative selection term for non-whites in 2006, which did not appear in 1987, reflects continued negative selection among non-white males and declining positive selection among non-white females.

Negative selection implies that more able or higher-earners among non-white men are more likely to be out of the labour force. There could be several plausible explanations for this result. First, we can think about the link between education and ability. If education (which is an observed characteristic) is highly correlated with the ability of a worker (unobserved), we can observe negative selection if there is a lower participation rate among the relatively more educated compared to those with zero education. Second, we can think about the link between education and reservation wages. While those with very little education may choose to work even when wages are low, those workers with slightly more education – that is, those with an intermediate level of education – might have a reservation wage that is higher than the wage offered in the market and thus opt out of the labour market. Such a pattern may further be facilitated by the rapid entry of moderately educated women into the workforce (as seen in table A9), as men who do not enter the labour market may be supported by a working spouse.

When we look at the educational achievement and labour market participation of specific population sub-groups it offers support for this possibility. Figure B1 in the appendix presents kernel density distributions by years of education for both participants and not participants, disaggregated into white and non-white females and white and non-white males. The pattern for females is consistent with positive selection, as women with at least 11 years of education comprise a significantly larger proportion of participants than non-participants, while the reverse is true of less educated women. In contrast, we do not find the same monotonic pattern among men, particularly among non-white men. We first see that non-white men with low levels of education (zero or less than five years) participate disproportionately in the labour market, representing a much larger share of participants than non-participants. As we would expect, the same is true among the highly educated (11 years or more) men. The more surprising result is that non-white men with between five and 10 years of education comprise a much larger proportion of non-participants than participants. They are disproportionately outside of the labour market. It is this combination of disproportionately high participation among the least educated, and disproportionately low participation among the moderately educated, which appears to lie behind the overall pattern of negative selection among non-white men. While somewhat similar patterns are apparent for white males, the effects are more muted, as we see a much more disproportionate participation among the highly educated, and less high participation among the less educated. The former pattern among highly educated workers, particularly in 2006, would account for increasingly positive selection among white males.

Tables A6 and A7 present the results for the Lee procedure. In interpreting the results, it is important to note that, owing to the construction of the Lee selection correction terms, in these tables a negative sign on the coefficient implies positive selection and thus a positive relationship between the unobservables determining the participation status and those determining the wage equation (see Gyourko and Tracy, 1988). We immediately observe quite considerable differences between the selection estimates using the Lee procedure and the Heckman procedure. These differences can be explained by differences in the way that the different selection terms are computed, as the IMR estimated via the Lee procedure is derived from a broader selection mechanism into the formal, informal and self-employed sectors. To this end we investigate the selection process by looking at each sector (i.e., formal, informal or self-employed individually) in turn. We note that alongside the major findings from the Heckman procedure - positive selection for females declining over time and negative selection for non-white males - the Lee procedure highlights differences across the different sectors. In particular, the selection terms from the Lee procedure differ because the Lee procedure puts more weight on the positive selection processes into the non-formal sectors for females and into the formal sector for males (which, of course, affect the results for non-whites and whites as well).

We can account to some extent for the differences between the two methods. However, these complications suggest a difficulty in accurately capturing selection processes given that they are unobserved and that the validity of the analysis is highly dependent on the appropriateness of the identifying instruments. Perhaps these discrepancies are a signal of a more severe problem associated with the existence of unobserved heterogeneity in the selection process (Machado, 2011). While recognizing the difficulties in detecting these unobserved and heterogeneous selection processes, there remains value in such analysis, and we continue with implementing the selection-corrected decomposition analysis to assess how our results may vary when selection is taken into account.

6.6.2 Selectivity-corrected decomposition results

Table 7 presents the OB and BMZ decomposition results after correcting for selectivity bias using the Heckman procedure. We report the estimated wage offer gaps, which are the differences between the estimated observed wage gaps and the selection term differentials. As described in the methodological section, we adopt the decomposition specification that keeps the difference in selection terms separate from the endowment and wage structure components (Neuman and Oaxaca, 2003). Therefore, the selection term differential corresponds to the last term of equation (9) for the BMZ decomposition and to the last term of equation (8) for the standard OB decomposition technique.

We first analyse the BMZ gender pay gap decomposition results corrected for selectivity reported in the upper part of panel A in table 7. The selection term differential is negative, as a result of greater positive selection among female workers. This positive selection among females results in a smaller female wage offer and, consequently, in a larger gender wage offer gap. In other words, gender differentials are larger after correcting for selection, as this process accounts for the fact that less able women are disproportionately out of the labour force. These findings are consisted with findings elsewhere in the literature on gender wage gaps. Olivetti and Petrongolo (2008) have, for instance, observed corrected wage gaps than are higher than observed wage gaps, as women are more positively selected into work than men. They also stress the prominent role of sample selection in countries with high employment differentials between gender groups.

The difference in the selection terms decreases over time, reflecting declining positive selection among female workers, as a greater proportion of less able or low-wage women have entered the labour market relative to more skilled women. When we turn to the selection-corrected results for the OB decomposition, reported at the bottom of the panel, we find broadly similar results.

In panel B, we report the corresponding results for the racial wage gaps. The selection term differential is generally small and positive, reflecting the existence of negative selection for non-white workers, and non-white males in particular. This negative selection implies that the estimated non-white wage offers are greater than the observed wages and the wage offer gaps are smaller than the observed racial pay gaps, although the difference is relatively modest.

Table 8 reports a similar analysis of changes in the results, this time when employing the Lee correction. The first important difference between the Heckman and Lee corrections is that while the selection term differential is large when employing both methods, they are significantly larger when the Lee correction is used. This is true for both gender and racial pay gaps, but is particularly sizeable in the case of gender wage gaps. For example, when applying the Heckman correction to the OB decomposition in 1987 the selection differentials component represents 45% of the overall gender gap, though this decreases to 5.4% in 2006, while the corresponding figures using the BMZ decomposition are 37% in 1987 and 1.3% in 2006. However, when we look at the magnitude of the selection differentials when using the Lee procedure the selection components are even larger. Neuman and Oaxaca (2003, 2004) have claimed that selection correction introduces ambiguities in decomposition analysis. Manski (1989) has also warned of a potential lack of robustness given that "seemingly small misspecifications might generate large biases in estimates" (Manski, 1989, p. 256).

A second important difference between the Heckman and the Lee corrections lies in the evolution of the selection term differentials over time. For gender wage gaps, we find using the Heckman correction that the selection differentials component is negative and increases over time, reflecting decreasing positive selection for female workers. By contrast, when using the Lee procedure we estimate a large negative selection differential, but this differential remains persistently negative over time despite some decline in its magnitude. In the case of racial wage gaps, we find that the selection term differentials computed using the Heckman procedure are generally positive while using the Lee procedure the selection term differentials are negative and remain negative. The patterns for race are more mixed because the estimated selection term differentials are small and highly sensitive to the methodology used.

These discrepancies between the Heckman and Lee procedures can be traced to the likely presence of unobserved heterogeneity in the selection process, as was noted earlier when we offered a possible interpretation of certain aspects of this heterogeneity in the selection process. More importantly, despite the difficulties in interpreting and accounting for the selection process, our selection-corrected decomposition results are reassuring. The main findings are unaltered, both in term of the evolution of gender and wage gaps and in terms of the contribution of each component in determining the gaps. We still find that vertical segregation (unexplained intra-occupational component) is the main determinant of gender wage gaps and of their decline over time. By contrast, racial wage gaps are primarily explained by differences in endowments, with the inter-occupational component declining slightly over time. Meanwhile vertical and horizontal segregation continue to account for about 25% of racial wage gaps and have remained stable over time. In other words, the narrative that emerges from the uncorrected decomposition analysis survives the selectivity correction procedures. Nevertheless, while we are able to reconcile the different results, and the selection-corrected decomposition results convey the same message as the uncorrected decomposition results, we remain suspicious of such a large explanatory role for the differences in selection terms components, particularly for the gender wage gaps reported here. For this reason we opt to implement further sensitivity checks.

Table 7: Brown, Moon and Zoloth (1980) and Oaxaca-Blinder (1973) decomposition with Heckman correction

PANEL A – GENDER WAGE GAPS

	1987	t-test	1992	t-test	1997	t-test	2002	t-test	2006	t-test
Brown, Moon and Zoloth (1980) with Heckman correction										
Explained intra-occupational component	0.027	***	0.043	***	-0.019	***	-0.019	***	-0.016	***
s.e.	0.0103		0.0087		0.0060		0.0064		0.0056	
Unexplained intra-occupational component	0.487	***	0.420	***	0.320	***	0.270	***	0.198	**
s.e.	0.1719		0.1708		0.1344		0.1238		0.1186	
Explained inter-occupational component	-0.156	***	-0.161	***	-0.177	***	-0.139	***	-0.135	***
s.e.	0.0011		0.0012		0.0011		0.0008		0.0008	
Unexplained inter-occupational component	0.087	***	0.033	***	0.061	***	0.003		0.000	
s.e.	0.0058		0.0056		0.0045		0.0030		0.0026	
Wage offer gap	0.444	2.580	0.334	1.954	0.186	1.378	0.115	0.925	0.047	0.391
s.e.	0.1723		0.1711		0.1346		0.1240		0.1188	
Selection term differential	-0.120	***	-0.125	***	-0.074	***	-0.053	***	0.007	
s.e.	0.0224		0.0221		0.0165		0.0157		0.0143	
Oaxaca-Blinder decomposition with Heckman correction										
Endowment effect	-0.072	***	-0.069	***	-0.131	***	-0.154	***	-0.156	***
s.e.	0.0083		0.0080		0.0071		0.0056		0.0050	
Treatment effect	0.541	***	0.442	***	0.336	***	0.277	***	0.213	***
s.e.	0.0170		0.0178		0.0150		0.0143		0.0136	
Wage offer gap	0.468	***	0.373	***	0.205	***	0.123	***	0.057	***
s.e.	0.0163		0.0170		0.0145		0.0142		0.0135	
Selection term differential	-0.146	***	-0.161	***	-0.092	***	-0.053	***	0.003	
s.e.	0.0146		0.0156		0.0132		0.0132		0.0127	
Total gap	0.324	***	0.208	***	0.111	***	0.064	***	0.055	***
s.e.	0.0069		0.0066		0.0059		0.0052		0.0046	

PANEL B – RACIAL WAGE GAPS

	1987	t-test	1992	t-test	1997	t-test	2002	t-test	2006	t-test
Brown, Moon and Zoloth (1980) with Heckman correction										
Explained intra-occupational component	0.160	***	0.227	***	0.224	***	0.201	***	0.168	***
s.e.	0.0063		0.0079		0.0065		0.0057		0.0054	
Unexplained intra-occupational component	0.074		0.038		0.053		0.030		0.032	
s.e.	0.1334		0.1301		0.1110		0.1114		0.1075	
Explained inter-occupational component	0.198	***	0.169	***	0.165	***	0.154	***	0.143	***
s.e.	0.0012		0.0011		0.0010		0.0010		0.0009	
Unexplained inter-occupational component	0.043	***	0.053	***	0.044	***	0.043	***	0.040	***
s.e.	0.0005		0.0006		0.0006		0.0005		0.0004	
Wage offer gap	0.475	3.556	0.487	3.737	0.486	4.373	0.428	3.835	0.384	3.565
s.e.	0.1336		0.1304		0.1112		0.1115		0.1076	
Selection term differential	0.014		0.052	***	0.048	***	0.054	***	0.036	***
s.e.	0.0165		0.0171		0.0134		0.0133		0.0121	
Oaxaca-Blinder decomposition with Heckman correction										
Endowment effect	0.407	***	0.448	***	0.436	***	0.395	***	0.344	***
s.e.	0.0057		0.0056		0.0051		0.0044		0.0039	
Treatment effect	0.090	***	0.066	***	0.072	***	0.028	***	0.028	***
s.e.	0.0141		0.0155		0.0132		0.0127		0.0117	
Wage offer gap	0.497	***	0.514	***	0.508	***	0.423	***	0.372	***
s.e.	0.0141		0.0150		0.0128		0.0124		0.0113	
Selection term differential	-0.008		0.023	**	0.025	***	0.051	***	0.041	***
s.e.	0.0125		0.0137		0.0115		0.0113		0.0104	
Total gap	0.489	***	0.541	***	0.535	***	0.484	***	0.422	***
s.e.	0.0064		0.0061		0.0055		0.0049		0.0045	

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

Table 8: Oaxaca-Blinder (1973) and Brown et al (1980) decomposition with Lee correction

PANEL A – GENDER WAGE GAPS

	1987	t-test	1992	t-test	1997	t-test	2002	t-test	2006	t-test
Brown, Moon and Zoloth (1980) with Lee correction										
Explained intra-occupational component	0.012	***	-0.013	***	-0.041	***	-0.040	***	-0.040	***
s.e.	0.006		0.005		0.004		0.004		0.004	
Unexplained intra-occupational component	0.616	***	0.608	***	0.529	***	0.404	***	0.388	***
s.e.	0.110		0.127		0.107		0.095		0.088	
Explained inter-occupational component	-0.156	***	-0.161	***	-0.177	***	-0.139	***	-0.135	***
s.e.	0.001		0.001		0.001		0.001		0.001	
Unexplained inter-occupational component	0.087	***	0.033	***	0.061	***	0.003		0.000	
s.e.	0.006		0.006		0.005		0.003		0.003	
Wage offer gap	0.559	***	0.467	***	0.372	***	0.228	***	0.212	***
s.e.	0.110		0.127		0.107		0.095		0.088	
Selection term differential	-0.235	***	-0.257	***	-0.260	***	-0.166	***	-0.159	***
s.e.	0.021		0.027		0.024		0.021		0.017	
Oaxaca-Blinder decomposition with Lee correction										
Endowment effect	-0.086	***	-0.107	***	-0.148	***	-0.165	***	-0.173	***
s.e.	0.011		0.012		0.010		0.008		0.007	
Treatment effect	0.662	***	0.567	***	0.438	***	0.338	***	0.290	***
s.e.	0.097		0.100		0.086		0.083		0.079	
Wage offer gap	0.576	***	0.460	***	0.290	***	0.173	***	0.116	***
s.e.	0.097		0.100		0.087		0.083		0.080	
Selection term differential	-0.254	***	-0.249	***	-0.177	***	-0.104	***	-0.056	***
s.e.	0.015		0.017		0.016		0.016		0.014	
Total gap	0.322	***	0.212	***	0.114	***	0.069	19.022	0.060	***
s.e.	0.005		0.005		0.004		0.004		0.003	

PANEL B – RACIAL WAGE GAPS

	1987	t-test	1992	t-test	1997	t-test	2002	t-test	2006	t-test
Brown, Moon and Zoloth (1980) with Lee correction										
Explained intra-occupational component	0.154	***	0.215	***	0.216	***	0.197	***	0.165	***
s.e.	0.006		0.007		0.006		0.005		0.004	
Unexplained intra-occupational component	0.161	**	0.188	***	0.166	*	0.172		0.155	
s.e.	0.079		0.086		0.077		0.076		0.073	
Explained inter-occupational component	0.198	***	0.169	***	0.165	***	0.154	***	0.143	***
s.e.	0.001		0.001		0.001		0.001		0.001	
Unexplained inter-occupational component	0.043	***	0.053	***	0.044	***	0.043	***	0.040	***
s.e.	0.001		0.001		0.001		0.000		0.000	
Wage offer gap	0.556	***	0.625	***	0.592	***	0.566	***	0.504	***
s.e.	0.079		0.086		0.077		0.076		0.016	
Selection term differential	-0.067	***	-0.085	***	-0.058		-0.084		-0.083	*
s.e.	0.015		0.020		0.019		0.018		0.0156	
Oaxaca-Blinder decomposition with Lee correction										
Endowment effect	0.398	***	0.427	***	0.426	***	0.386	***	0.337	***
s.e.	0.006		0.008		0.007		0.005		0.005	
Treatment effect	0.156	*	0.161	***	0.156	**	0.150	**	0.141	***
s.e.	0.099		0.111		0.096		0.088		0.084	
Wage offer gap	0.554	***	0.588	***	0.581	***	0.536	***	0.478	***
s.e.	0.099		0.110		0.096		0.088		0.084	
Selection term differential	-0.065	***	-0.051	***	-0.049	***	-0.062	***	-0.065	***
s.e.	0.013		0.017		0.016		0.015		0.013	
Total gap	0.489	***	0.537	***	0.532	***	0.474	***	0.413	***
s.e.	0.005		0.004		0.004		0.003		0.003	

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

6.7 Sensitivity analysis

The parametric selection correction models employed in the previous section have been criticized for their restrictive distributional assumptions and robustness (Manski, 1989) and for the additional ambiguity that they introduce in the interpretation of the wage gap decomposition results (Neuman and Oaxaca, 2003). Given these concerns, the large estimated differences in selection terms, particularly for gender wage gaps, provide grounds for some suspicion, although it is important to reiterate that the findings from the uncorrected decomposition analysis appear invariant to the selectivity correction procedures. Given the potential limitations of the preceding analysis, we implement several sensitivity checks in this section in order to further verify and confirm our core results. We first re-estimate the results using the same methodology but employing an alternative sample. We then implement two alternative correction methods: the non-parametric imputation method by Olivetti and Petrongolo (2008) and the local wage gap estimation by Machado (2011).

6.7.1 Sensitivity checks using alternative versions of the sample

A possible explanation for the large selection differential components lies in the high levels of heterogeneity in the sample resulting from exploring the entire labour market simultaneously. We thus re-estimate both the OB decomposition and the BMZ decomposition with both corrections (Heckman and Lee procedures) employing more restricted samples. We consider three different, progressively more restricted, samples: full-time workers only, full-time workers excluding the self-employed, and full-time workers excluding the self-employed and agricultural workers. We present these robustness checks in tables 9 and 10.

Perhaps the most interesting finding relates to the magnitude of female selection effect when we restrict the sample to full-time workers only. We find that the extent of positive female selection into the labour market declines significantly within the restricted sample, such that positive selection in 1987 declines by roughly 20% relative to the full sample, while by 2006 we observe significant negative selection, with less able women entering disproportionately into the labour market. The explanation for this pattern appears to lie in significant differences between women who are employed part-time and full-time, as women who are employed part-time exhibit significant positive selection that has, if anything, marginally increased over time (i.e. they are disproportionately more able), while those who are employed full-time exhibit negative selection by 2006. This is consistent with the results reported earlier about the sensitivity of the OB decomposition results to employing monthly wages as the dependent variable and to restricting the sample to full-time workers. In both cases the changes led to an increase in the wage gap, consistent with the notion that part-time female workers are on average more able and better paid.

When we progressively exclude the self-employed and agricultural workers it further alters the magnitude of the results somewhat, but has no effect on the broad pattern of our results. The exclusion of self-employed workers leads to somewhat smaller selection differentials components, particularly when employing the Lee procedure to correct for selection in estimating gender wage gaps. This may reflect the fact that the exclusion of the self-employed cleans the estimates of some of the gender heterogeneity in the self-employed sector. The further exclusion of agricultural workers has virtually no additional impact on the selection differentials component. The only notable change in this last case is a relatively large increase in the estimated gender wage offer gap, but this change is almost entirely restricted to the inter-occupational components. Most importantly, while we thus see some changes in the magnitudes of the wage gaps or the differences in selection terms, the main findings from the uncorrected OB and BMZ decompositions are again unaltered.

Table 9: Heckman correction – Sensitivity checks

PANEL A – GENDER WAGE GAPS

	only full-time				**only full-time and no self-empl.**				**only full-time, no self-empl. and no agric.**			
	1987		**2006**		**1987**		**2006**		**1987**		**2006**	
Brown, Moon and Zoloth (1980) with Heckman correction												
Explained intra-occupational component	0.043	***	-0.025	***	0.062	***	-0.009	*	0.074	***	-0.004	
s.e.	0.013		0.005		0.018		0.005		0.021		0.005	
Unexplained intra-occupational component	0.498	***	0.175		0.414	***	0.182		0.427	*	0.184	
s.e.	0.153		0.117		0.191		0.117		0.225		0.120	
Explained inter-occupational component	-0.145	***	-0.143	***	-0.116	***	-0.110	***	-0.055	***	-0.063	***
s.e.	0.001		0.001		0.001		0.001		0.001		0.001	
Unexplained inter-occupational component	0.119	***	0.016	***	0.119	***	0.010	***	0.171	***	0.031	***
s.e.	0.005		0.003		0.005		0.003		0.005		0.003	
Wage offer gap	0.515	3.350	0.022	0.189	0.479	2.500	0.073	0.626	0.617	2.729	0.148	1.233
s.e.	0.154		0.117		0.192		0.117		0.226		0.120	
Selection term differential	-0.092	***	0.077	***	-0.066	***	0.055	***	-0.073	***	0.055	***
s.e.	0.020		0.013		0.024		0.013		0.027		0.013	
Oaxaca-Blinder decomposition with Heckman correction												
Endowment effect	-0.011		-0.153	***	0.055	***	-0.087	***	0.179	***	-0.022	***
s.e.	0.008		0.005		0.009		0.006		0.009		0.006	
Treatment effect	0.539	***	0.189	***	0.418	***	0.161	***	0.416	***	0.164	***
s.e.	0.017		0.013		0.018		0.012		0.019		0.012	
Wage offer gap	0.528	***	0.036	***	0.473	***	0.074	***	0.595	***	0.142	***
s.e.	0.017		0.013		0.017		0.012		0.019		0.013	
Selection term differential	-0.106	***	0.069	***	-0.063	***	0.057	***	-0.053	***	0.062	***
s.e.	0.015		0.012		0.015		0.011		0.017		0.011	
Total gap	0.423	***	0.100	***	0.412	***	0.130	***	0.543	***	0.205	***
s.e.	0.008		0.005		0.008		0.005		0.009		0.005	

PANEL B – RACIAL WAGE GAPS

	only full-time				**only full-time and no self-empl.**				**only full-time, no self-empl. and no agric.**			
	1987		**2006**		**1987**		**2006**		**1987**		**2006**	
Brown, Moon and Zoloth (1980) with Heckman correction												
Explained intra-occupational component	0.151	***	0.161	***	0.131	***	0.145	***	0.128	***	0.146	***
s.e.	0.007		0.005		0.006		0.004		0.006		0.004	
Unexplained intra-occupational component	0.088		0.026		0.076		0.015		0.083		0.014	
s.e.	0.133		0.107		0.127		0.101		0.144		0.106	
Explained inter-occupational component	0.180	***	0.136	***	0.222	***	0.145	***	0.196	***	0.122	***
s.e.	0.001		0.001		0.001		0.001		0.001		0.001	
Unexplained inter-occupational component	0.046	***	0.042	***	0.050	***	0.048	***	0.059	***	0.046	***
s.e.	0.001		0.000		0.001		0.001		0.001		0.001	
Wage offer gap	0.464	3.495	0.365	3.423	0.480	3.759	0.352	3.487	0.466	3.220	0.327	3.069
s.e.	0.133		0.107		0.128		0.101		0.145		0.107	
Selection term differential	0.001	0.092	0.046	***	-0.009	-0.562	0.045	***	-0.001	-0.068	0.053	***
s.e.	0.015		0.011		0.016		0.011		0.018		0.012	
Oaxaca-Blinder decomposition with Heckman correction												
Endowment effect	0.379	***	0.328	***	0.413	***	0.334	***	0.389	***	0.310	***
s.e.	0.006		0.004		0.007		0.005		0.007		0.005	
Treatment effect	0.096	***	0.022	**	0.080	***	0.006		0.094	***	0.005	
s.e.	0.013		0.011		0.014		0.011		0.015		0.012	
Wage offer gap	0.475	***	0.351	***	0.493	***	0.340	***	0.483	***	0.315	***
s.e.	0.013		0.011		0.014		0.011		0.015		0.012	
Selection term differential	-0.010		0.054	***	-0.023	*	0.052	***	-0.019		0.060	***
s.e.	0.011		0.009		0.012		0.010		0.013		0.011	
Total gap	0.465	***	0.413	***	0.470	***	0.398	***	0.464	***	0.381	***
s.e.	0.007		0.005		0.007		0.005		0.008		0.005	

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

Table 10: Lee correction – Robustness checks

PANEL A – GENDER WAGE GAPS

	only full-time				only full-time and no self-empl.				only full-time, no self-empl. and no agric.			
	1987		**2006**		**1987**		**2006**		**1987**		**2006**	
Brown, Moon and Zoloth (1980) with Lee correction												
Explained intra-occupational component	0.040	***	-0.028	***	0.059	***	-0.010	**	0.072	***	-0.004	
s.e.	0.009		0.004		0.010		0.004		0.012		0.004	
Unexplained intra-occupational component	0.648	***	0.317	***	0.433	***	0.270	**	0.432	***	0.257	**
s.e.	0.112		0.099		0.126		0.108		0.139		0.112	
Explained inter-occupational component	-0.145	***	-0.143	***	-0.116	***	-0.110	***	-0.055	***	-0.063	***
s.e.	0.001		0.001		0.001		0.001		0.001		0.001	
Unexplained inter-occupational component	0.119	***	0.016	***	0.119	***	0.010	***	0.171	***	0.031	***
s.e.	0.005		0.003		0.005		0.003		0.005		0.003	
Wage offer gap	0.662	5.862	0.162	1.632	0.495	3.922	0.160	1.478	0.620	4.442	0.220	1.958
s.e.	0.113		0.099		0.126		0.108		0.140		0.112	
Selection term differential	-0.239	***	-0.063	***	-0.082	***	-0.032	-1.895	-0.076	***	-0.017	
s.e.	0.022		0.020		0.020		0.017		0.022		0.017	
Oaxaca-Blinder decomposition with Lee correction												
Endowment effect	-0.130	***	-0.195	***	-0.123	***	-0.164	***	-0.018	***	-0.105	***
s.e.	0.004		0.002		0.004		0.001		0.005		0.001	
Treatment effect	0.754	***	0.142	*	0.706	***	0.178	**	0.731	***	0.190	**
s.e.	0.081		0.077		0.086		0.085		0.098		0.089	
Wage offer gap	0.624	7.686	-0.053	-0.684	0.583	6.752	0.013	0.159	0.713	7.246	0.085	0.956
s.e.	0.081		0.077		0.086		0.085		0.098		0.089	
Selection term differential	-0.203	***	0.159	***	-0.173	***	0.117	***	-0.171	***	0.118	***
s.e.	0.016		0.014		0.015		0.014		0.016		0.014	
Total gap	0.422	***	0.106	***	0.411	***	0.131	***	0.542	***	0.204	***
s.e.	0.005		0.003		0.006		0.004		0.006		0.004	

PANEL B – RACIAL WAGE GAPS

	only full-time				only full-time and no self-empl.				only full-time, no self-empl. and no agric.			
	1987		**2006**		**1987**		**2006**		**1987**		**2006**	
Brown, Moon and Zoloth (1980) with Lee correction												
Explained intra-occupational component	0.150	***	0.166	***	0.132	***	0.147	***	0.129	***	0.148	***
s.e.	0.006		0.004		0.006		0.004		0.005		0.004	
Unexplained intra-occupational component	0.121		0.071		0.066		0.046		0.094		0.053	
s.e.	0.083		0.078		0.090		0.085		0.106		0.093	
Explained inter-occupational component	0.180	***	0.136	***	0.222	***	0.145	***	0.196	***	0.122	***
s.e.	0.001		0.001		0.001		0.001		0.001		0.001	
Unexplained inter-occupational component	0.046	***	0.042	***	0.050	***	0.048	***	0.059	***	0.046	***
s.e.	0.001		0.000		0.001		0.001		0.001		0.001	
Wage offer gap	0.496	5.926	0.415	5.277	0.470	5.192	0.386	4.513	0.478	4.491	0.368	3.965
s.e.	0.084		0.079		0.091		0.085		0.106		0.093	
Selection term differential	-0.031	*	-0.003		0.001		0.012		-0.014		0.012	
s.e.	0.016		0.017		0.015		0.015		0.017		0.016	
Oaxaca-Blinder decomposition with Lee correction												
Endowment effect	0.350	***	0.295	***	0.366	***	0.269	***	0.334	***	0.243	***
s.e.	0.005		0.004		0.005		0.004		0.005		0.004	
Treatment effect	0.212	***	0.069		0.239	***	0.097		0.245	**	0.069	
s.e.	0.076		0.071		0.082		0.081		0.096		0.087	
Wage offer gap	0.563	7.421	0.364	5.096	0.605	7.340	0.367	4.521	0.579	6.008	0.312	3.579
s.e.	0.076		0.071		0.082		0.081		0.096		0.087	
Selection term differential	-0.098	***	0.041	***	-0.135	***	0.025	2.022	-0.115	***	0.064	***
s.e.	0.013		0.013		0.013		0.012		0.014		0.013	
Total gap	0.465	***	0.405	***	0.470	***	0.392	***	0.464	***	0.375	***
s.e.	0.005		0.003		0.005		0.003		0.006		0.004	

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

6.7.2 Using alternative methods of selection correction

In order to assess the reliability of the estimated wage offer gaps we now consider the results when employing two alternative methods for selection correction: the non-parametric imputation method, as illustrated by Olivetti and Petrongolo (2008), and the estimation of the local wage gap, proposed by Machado (2011). The next sub-section briefly introduces these methodologies, after which we present the core results.

6.7.2.1 Olivetti and Petrongolo (2008) imputation methodology and Machado (2011) local wage gap

The imputation method for selection correction is based on the notion of recovering information on wages for the non-employed by adopting several imputation techniques. This allows the median wage gap to be computed on the sample of employed and the additional non-employed individuals whose wages are imputed. Because the wage gap is computed via median regressions, we only need information on the position with respect to the median of the imputed wages and not the actual level of missing wages. This is possible because the median regressions are only affected by the position of wage observations. As long as the imputed position with respect to the median is correct the median regression will generate unbiased estimates.

The imputation of missing wages is undertaken by applying different sets of progressively more inclusive sample inclusion rules. Because of the cross-sectional nature of our data, we are only able to implement the missing wages imputations based on the observable characteristics of those who are not employed. The first imputation method is to simply assign the minimum level of observed wages to those that are unemployed. The second imputation method is to enlarge the sample by exploiting information on education levels. We assign the minimum level of observed wages to non-employed (unemployed and out of the labour force) individuals with zero years of education, while we assign the maximum level of observed wages to non-employed individuals with more than 11 years of education. The third imputation method further expands the sample by employing a probability model in which the likelihood of earning less than the median wage is estimated with a probit model and the estimated scores are used as sampling weights to construct the imputed sample.

The most important limitation of this imputation methodology lies in the reliability of the observed median wage. Given the selection problem, the observed median wage might be very different to the latent and unbiased median. For this reason, Olivetti and Petrongolo (2008) recommend the use of this methodology only when the participation rate is above 50%. This does not pose any problem in the case of racial wage gaps, as participation rates for both non-white and white workers are significantly above 50% across the entire time period. However, this poses a significant problem in the case of gender wages gaps, as female participation moves from only roughly 41% in 1987 to 48% in 2006. For this reason, it is important to take this specific exercise as suggestive rather than compelling. Finally, it is important to stress that we adopt this imputation methodology to estimate median wage gaps, as this only requires information on the position and not the level of the missing wages. For this reason, we cannot take this method too far in decomposing wage gaps because reliable levels of missing wages are needed in order to properly estimate the returns to observables in the Mincerian equations.

The second non-parametric selection correction is the estimation of the "local wage gap" as proposed by Machado (2011). The existence of unobserved heterogeneity in the selection process is often overlooked in the use of both parametric and non-parametric methods, which impose structure on the selection rules. In practice different selection rules might co-exist, in which case the sign of an estimated selection rule would simply capture the average of this unobserved heterogeneity. In order to address this concern Machado (2011) method builds on the concept of "always takers" in the treatment effect models developed by Angrist, Imbens and Rubin (1996). In the selection process there are individuals for whom the identifying instrument decreases or increases the likelihood of participation (i.e. "switchers" and "defiers"), while there are some individuals that will always be employed or never employed. By focusing exclusively on a sample of the "always employed", for whom the employment decision does not change, we are able to recover an unbiased local measure of the wage gap. In order to implement this method we need an instrument to identify the "always employed", and we opt to consider those individuals that work regardless of whether they have children younger than six years old. By adopting this method we are able to consider two comparable groups in the computation of the wage gaps: women and men, or non-white and white individuals, that show a higher attachment to work.

6.7.2.2 Comparing estimated gaps

In tables 11 and 12 we report the comparison of gender and racial wage gaps estimated using several alternative selection correction methodologies. The first five rows in Table 11 report the actual observed wage gap and the estimated wage offer gaps when employing the OB and Brown et al (1980) decomposition techniques with, in turn, the Heckman and Lee parametric correction models. In the three rows that follow we report the estimated median wage gaps using three different methods of imputation, based on the three different sets of sample inclusion rules described above. In the final row we report the estimated local wage gap, following Machado (2011), where the "always employed" individuals are identified as those individuals that work irrespective of whether they have children younger than six years old.

When reviewing the various results, we first note that the selection corrected gender wage gaps (or estimated wage offer gaps) are generally greater than the observed wage gaps. Thus, both parametric and non-parametric methods for selection correction support the existence of positive selection for women. This is also clearly visible in panel A of figure 6 where the plot of the observed wage gap is lower than all of the other plots. We further find that in most specifications the gender wage gap decreases more over time when we correct for selection, though by 2006 the selection corrected results still lead in most cases to a higher wage gap than that observed in the raw data. As such, the faster decline in the gender wage gap largely reflects the higher starting point in the selection corrected decomposition results. One possible explanation lies in the changing pattern of labour participation. Female positive selection is decreasing over time, which implies that more unskilled ("less able") women are entering the labour market relatively more than more able women. In other words, the gender wage gap has decreased over time both because of less discrimination and because of different patterns of labour market participation, with less qualified women disproportionately entering the labour market. This is also evident if we look at the evolution of the Machado (2011) local wage gaps. The local wage gaps show the smallest decrease over time, consistent with the idea that the decline in the pay gap is smaller if we focus attention on more comparable women and men.

Table 11: Summary of gender wage gaps – with alternative methods to correct for selectivity

PANEL A – Comparison of wage gaps estimated using different correction methods for selectivity

	1987	1992	1997	2002	2006
observed wage gap	0.322	0.209	0.111	0.064	0.061
s.e.	0.007	0.007	0.006	0.005	0.005
wage offer gap - Heckman & OB	0.468	0.373	0.205	0.123	0.057
s.e.	0.016	0.017	0.015	0.014	0.013
wage offer gap - Heckman & Brown et al (1980)	0.444	0.334	0.186	0.115	0.047
s.e.	0.172	0.171	0.135	0.124	0.119
wage offer gap - Lee & OB	0.576	0.460	0.290	0.173	0.116
s.e.	0.097	0.100	0.087	0.083	0.080
wage offer gap - Lee & Brown et al (1980)	0.559	0.467	0.372	0.228	0.212
s.e.	0.110	0.127	0.107	0.095	0.088
median wage gap - Olivetti and Petrongolo (2008) 1st imputation method	0.316	0.219	0.201	0.110	0.090
s.e.	0.000	0.000	0.009	0.011	0.000
median wage gap - Olivetti and Petrongolo (2008) 2nd imputation method	0.511	0.289	0.223	0.095	0.056
s.e.	0.000	0.012	0.001	0.001	0.005
median wage gap - Olivetti and Petrongolo (2008) 3rd imputation method	0.640	0.396	0.433	0.316	0.174
s.e.	0.000	0.008	0.008	0.003	0.001
local gender gap - (Machado, 2011)	0.414	0.335	0.219	0.212	0.196
s.e.	0.010	0.010	0.009	0.008	0.008

PANEL B – Percentage of adult population in sample for the imputation methodology by Olivetti and Petrongolo (2008)

	1987		1992		1997		2002		2006	
	F	M	F	M	F	M	F	M	F	M
No of obs.	86859	75519	91993	79991	104385	91832	120657	108034	130745	117532
Employed	40.7%	83.3%	41.9%	78.9%	42.5%	75.7%	45.7%	74.3%	48.4%	75.2%
Imputation method 1	42.7%	86.7%	47.3%	85.3%	49.1%	82.5%	53.8%	81.9%	56.4%	81.8%
Imputation method 2	54.2%	86.6%	54.3%	83.0%	53.5%	80.1%	55.5%	78.9%	57.9%	79.7%
Imputation method 3	66.1%	100.0%	66.8%	100.0%	68.7%	100.0%	71.3%	100.0%	73.4%	100.0%

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

Table 12: Summary of racial wage gaps – with alternative methods to correct for selectivity

PANEL A – Comparison of wage gaps estimated using different correction methods for selectivity

	1987	1992	1997	2002	2006
observed wage gap	0.489	0.541	0.535	0.484	0.422
s.e.	0.006	0.006	0.006	0.005	0.005
wage offer gap - Heckman & OB	0.497	0.514	0.508	0.423	0.372
s.e.	0.014	0.015	0.013	0.012	0.011
wage offer gap - Heckman & Brown et al (1980)	0.475	0.487	0.486	0.428	0.384
s.e.	0.134	0.130	0.111	0.112	0.108
wage offer gap - Lee & OB	0.554	0.588	0.581	0.536	0.478
s.e.	0.099	0.110	0.096	0.088	0.084
wage offer gap - Lee & Brown et al (1980)	0.556	0.625	0.592	0.566	0.504
s.e.	0.079	0.086	0.077	0.076	0.073
median wage gap - Olivetti and Petrongolo (2008) 1st imputation method	0.462	0.462	0.511	0.470	0.357
s.e.	0.010	0.007	0.002	0.001	0.001
median wage gap - Olivetti and Petrongolo (2008) 2nd imputation method	0.550	0.522	0.600	0.514	0.417
s.e.	0.001	0.000	0.008	0.008	0.007
median wage gap - Olivetti and Petrongolo (2008) 3rd imputation method	0.377	0.362	0.482	0.406	0.270
s.e.	0.001	0.000	0.001	0.001	0.004
local gender gap - (Machado, 2011)	0.494	0.589	0.581	0.561	0.514
s.e.	0.008	0.009	0.008	0.007	0.006

PANEL B – Percentage of adult population in sample for the imputation methodology by Olivetti and Petrongolo (2008)

	1987		1992		1997		2002		2006	
	NW	W	NW	W	NW	W	NW	W	NW	W
No of obs.	74847	87531	81585	90398	94566	101651	116027	112664	134829	113448
Employed	61.4%	59.7%	59.0%	59.3%	57.6%	58.5%	58.3%	60.1%	60.0%	62.4%
Imputation method 1	64.3%	62.2%	65.6%	64.4%	65.0%	64.5%	67.1%	67.1%	68.1%	68.8%
Imputation method 2	71.6%	67.3%	69.0%	66.5%	66.2%	65.7%	65.8%	67.4%	66.9%	69.8%
Imputation method 3	89.0%	89.6%	81.9%	86.5%	82.2%	86.5%	82.2%	88.4%	84.1%	92.4%

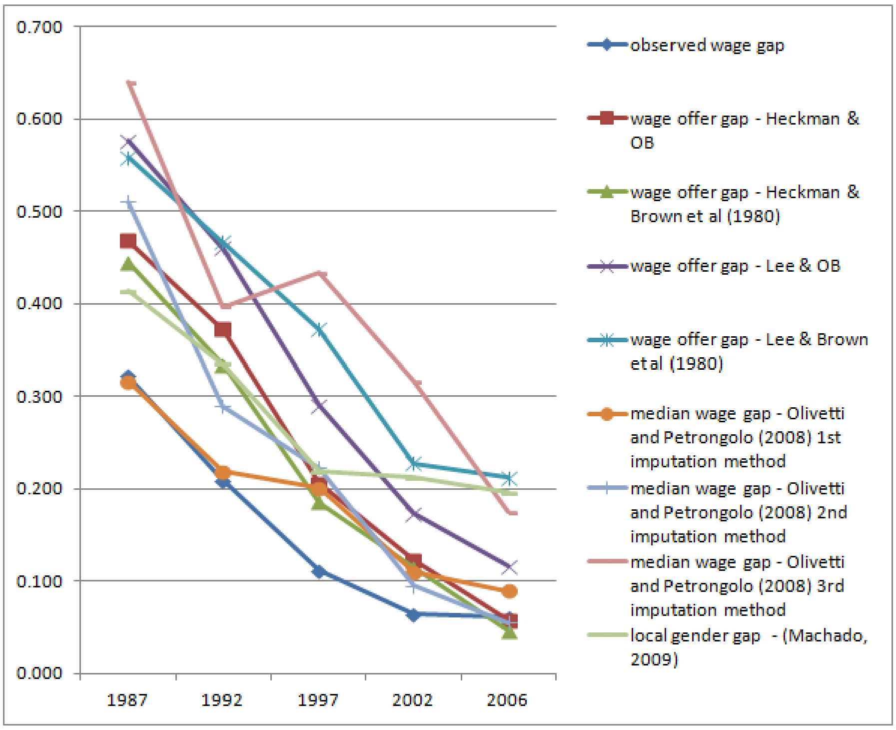
Source: Author's computations using PNAD 1987 – 1992 – 1997 – 2002 – 2006.

Looking at the selection corrected racial wage gaps in table 12, we observe mixed results when using alternative methods of correction. Among the parametric methods, employing the Lee correction yields wage offer gaps that are generally greater than the observed gaps, owing to the negative difference in selection terms. On the other hand, when we employ the imputation methodology using three alternative sets of sample inclusion rules we find estimated wage offer gaps that are smaller than the observed gaps in two cases, and larger in the third case. Finally, the local wage gap is considerably greater than the observed wage gap, and, more strikingly, shows an increase over time that is significant at the 10% level. In order to explain this last result it is necessary to recall that the observed racial wage gap is driven primarily by differences in observed characteristics, while the decline in the racial wage gap over time is driven almost entirely by a relative improvement in non-white endowments. What the slight increase in the local wage gap thus suggests is that within the "always employed" group either this narrowing of the endowments gap among whites and non-whites has not occurred or, alternatively, has been offset by an increased treatment effect that is not apparent for the whole sample. Put differently, any improvements in wage gaps appear to be concentrated among those who fall outside of the sample used for calculating the local wage gap. Analysing why this might be would require investigation that is beyond the scope of this thesis.

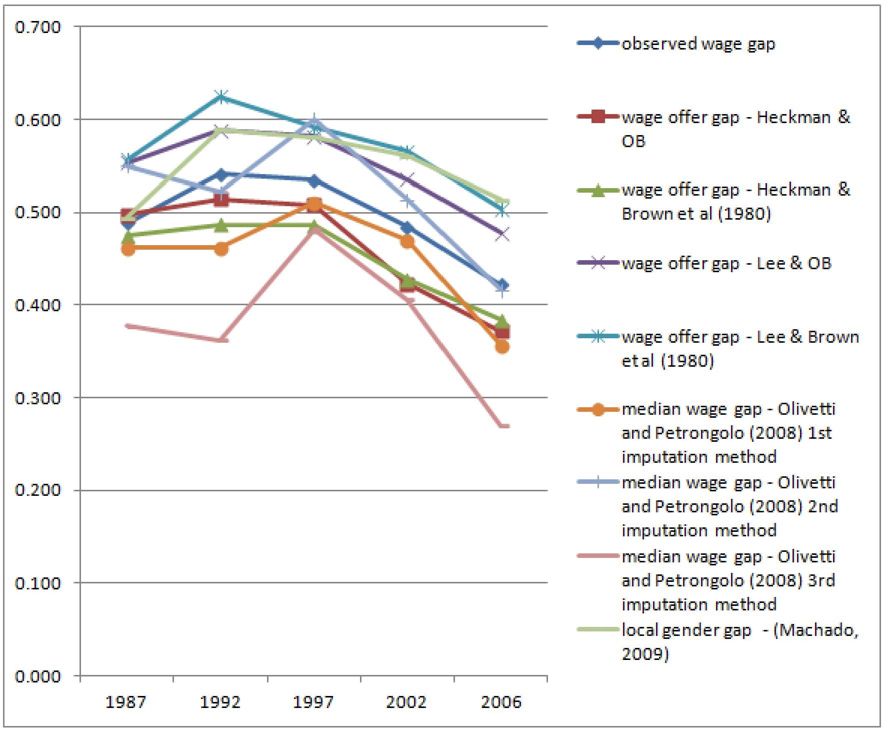
While it would be possible to devote more attention to understanding these discrepancies between the different methods, this is not our primary concern, as our focus has been on exploiting multiple alternative corrections for selection in order to verify the reliability of our core results. On this count the results are, again, reassuring, as the overall trends in gender and racial pay gaps remain essentially unchanged across all of the different methodologies. This is particularly true for gender wage gaps, though with the caveat that the results when applying the Olivetti and Petrongolo (2008) imputation method should be treated as suggestive given that female labour market participation remains below 50%. There is somewhat more variability in the racial wage gap results, but the modest overall increase in the wage gap over time when using the local wage gap is the only genuinely conflicting result, and even there the results over five year intervals follow the trend of other results very closely. On balance these sensitivity checks reinforce confidence in our results, while emphasizing the difficulties inherent in efforts to capture the selection process.

Figure 6: Comparison of wage gaps estimated using different correction methods for selectivity

PANEL A – Gender wage gaps



PANEL B – Racial wage gaps



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

6.8 Conclusions

In this chapter we investigated the evolution of gender and racial pay gaps in the Brazilian labour market over time by applying and comparing two different wage gap decomposition methods: the standard Oaxaca-Blinder (1973) decomposition technique and the Brown, Moon and Zoloth (1980) decomposition method, which decomposes wage gaps while explicitly accounting for occupational segregation. We further enriched our approach by accounting for the impact of selectivity on the results. We have considered two alternative parametric corrections for selection, treating participation both as a simple binary decision (the Heckman procedure) and as a more complex decision that incorporates a distinction between the formal and non-formal sectors (the Lee procedure). We have also performed additional sensitivity checks by applying the non-parametric imputation method by Olivetti and Petrongolo and the local wage gap estimation by Machado (2011).

The standard OB decomposition revealed that gender wage gaps are explained primarily by the treatment effect (i.e., differences in the wage structure), while racial wage gaps are primarily explained by the endowment effect (i.e., differences in observed characteristics). Both gender and racial differentials were higher at the end of the 1980s, however only gender gaps have steadly declined over time. The rapid decrease in gender wage gaps is primarily attributable to changes in wage structure (and, among other unobserved factors, discriminatory behaviour) while the small decrease in racial wage gaps is almost entirely due to improved characteristics among non-white individuals, such as increased educational attainment.

Applying the Brown, Moon and Zoloth (1980) decomposition adds to our findings further. It reveals that the decrease over time in gender wage gaps is explained primarily by reduced vertical segregation, by which we mean differences in wage structures within occupations (or unexplained intra-occupational component). It also reveals that differences in observed characteristics exist primarily across occupations, with female workers employed disproportionately in more skilled professions, which would justify higher female wages on average. Finally, we find that horizontal segregation represents a minor issue, which modestly favours women over time.

In the case of racial wage gaps, we find that they are primarily driven by differences in observed characteristics, both within and across occupations, and that small improvements over time are entirely explained by declining differences across occupations, while within occupations white workers continue to have significantly better endowments. Both horizontal and vertical segregation are comparatively small in magnitude when compared to gender wage gaps, as they account for 25% of the total gap, but are persistent over time.

Having arrived at a core set of results we then explored various methods to account for selectivity bias. In doing so, we find different selection processes by gender and race. Selection corrected gender wage gaps (or gender wage offer gaps) are greater than observed wage gaps. This reflects the existence of positive female selection into the labour market, which inflates average female wages above what they would be in the absence of selection. However, this positive female selection has decreased markedly over time, and is only marginally positive by 2006. The inference is that more unskilled women, with no more than compulsory education, are entering the labour market in comparatively larger numbers. That said, these results appear to demand some caution, as the differences in selection terms component of the gender decompostions results is generally very large.

In contrast, accounting for selectivity bias yields less clear-cut trends for racial wage gaps. The Heckman procedure yields selection corrected racial wage gaps are smaller than observed wage gaps, owing primarily to negative selection among non-white workers, and particularly among non-white men. Among the latter, moderately skilled workers are disproportionately outside of the labour market, while less educated men are more likely to be employed. However, this trend is reversed in using the Lee correction, which finds positive selection among non-white workers by virtue of giving a greater weight to positive selection into the formal sector.

While the different methodologies thus yield different estimates of the effects of selection, this is not a major concern. Previous studies have highlighted the potential problems associated with accounting for selectivity bias in decomposition results, and particularly the sensitivity of results to methodological choices (Manski, 1989; Neuman and Oaxaca 2003). This reflects, among other things, the likely heterogeneity of underlying selection processes. Owing to these difficulties related to the correction for selection bias (i.e. validity of the instruments, ambiguities related to the introduction of selection processes within the decomposition framework and unobservable heterogeneity of the selection process) we have also implemented additional sensitivity checks by adopting a non-parametric imputation method developed by Olivetti and Petrongolo (2008) and the local wage gap estimation by Machado (2011).

The goal has been to verify the reliability of our core results by exploring their sensitivity to employing multiple alternative corrections for selection. The core decomposition results are largely unchanged across a wide variety of methods that account for selectivity bias, and across various efforts to test the sensitivity of the results to changes in the sample. Given that our focus has been on testing the robustness of our core results, the consistency of these broad trends across methods is very encouraging, and suggests that we have accurately captured the broad evolution of wage discrimination by gender and race over time.

Appendix to Chapter 6

Table A1: Summary statistics for the main covariates

PANEL A – Sample of year 1987

	female		male			non white		white			total	
	mean	s.d.	mean	s.d.	t-test	mean	s.d.	mean	s.d.	t-test	mean	s.d.
male						0.476	0.499	0.456	0.498	***	0.465	0.499
white	0.548	0.498	0.529	0.499	***						0.539	0.498
age	33.190	13.495	33.266	13.438	n.s.	32.655	13.406	33.713	13.502	***	33.225	13.468
edu	5.254	4.148	5.038	4.125	***	4.123	3.685	6.034	4.298	***	5.153	4.138
urban	0.820	0.384	0.785	0.411	***	0.769	0.421	0.834	0.373	***	0.804	0.397
formal	0.430	0.495	0.467	0.499	***	0.395	0.489	0.506	0.500	***	0.454	0.498
focc3	0.653	0.270	0.207	0.214	***	0.362	0.324	0.373	0.313	***	0.367	0.318
nwocc3	0.456	0.114	0.471	0.115	***	0.494	0.102	0.441	0.119	***	0.466	0.115

PANEL B – Sample of year 2006

	female		male			non white		white			total	
	mean	s.d.	mean	s.d.	t-test	mean	s.d.	mean	s.d.	t-test	mean	s.d.
male						0.486	0.500	0.459	0.498	***	0.473	0.499
white	0.470	0.499	0.443	0.497	***						0.457	0.498
age	35.229	13.693	34.701	13.564	***	34.260	13.430	35.834	13.825	***	34.979	13.635
edu	7.889	4.235	7.252	4.215	***	6.833	4.072	8.476	4.256	***	7.587	4.238
urban	0.885	0.319	0.848	0.359	***	0.845	0.362	0.895	0.307	***	0.868	0.339
formal	0.477	0.499	0.460	0.498	***	0.418	0.493	0.523	0.499	***	0.467	0.499
focc3	0.646	0.222	0.277	0.247	***	0.426	0.304	0.445	0.293	***	0.435	0.299
nwocc3	0.517	0.110	0.540	0.114	***	0.554	0.101	0.502	0.119	***	0.530	0.113

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

Table A2: Wage equations for the OB decomposition for female and male workers - all labour market

PANEL A - year 1987

	(1) Females b/se	(2) Females b/se	(3) Females b/se	(4) Females b/se	(5) Males b/se	(6) Males b/se	(7) Males b/se	(8) Males b/se
white	0.144***	0.108***	0.134***	0.107***	0.127***	0.111***	0.130***	0.110***
	(0.009)	(0.008)	(0.009)	(0.008)	(0.006)	(0.006)	(0.006)	(0.006)
age	0.092***	0.087***	0.092***	0.086***	0.101***	0.089***	0.099***	0.089***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
agesq	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
edu	0.137***	0.089***	0.136***	0.089***	0.120***	0.089***	0.123***	0.089***
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
urban	0.243***	0.292***	0.297***	0.293***	0.298***	0.153***	0.300***	0.151***
	(0.013)	(0.014)	(0.013)	(0.014)	(0.008)	(0.008)	(0.008)	(0.008)
formal	0.194***	0.172***	0.151***	0.149***	0.122***	0.071***	0.124***	0.073***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.006)	(0.006)	(0.006)	(0.006)
focc3			-0.473***	-0.401***			-0.262***	0.095***
			(0.015)	(0.025)			(0.013)	(0.019)
Regions FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupations FE	No	Yes	No	Yes	No	Yes	No	Yes
N	35077	35077	35077	35077	62602	62602	62602	62602
r2	0.536	0.579	0.550	0.582	0.480	0.528	0.483	0.528

* p<0.10, ** p<0.05, *** p<0.01

PANEL B - year 2006

	(1) Females b/se	(2) Females b/se	(3) Females b/se	(4) Females b/se	(5) Males b/se	(6) Males b/se	(7) Males b/se	(8) Males b/se
white	0.117***	0.087***	0.114***	0.087***	0.133***	0.104***	0.134***	0.104***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
age	0.058***	0.053***	0.059***	0.053***	0.069***	0.062***	0.069***	0.062***
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
agesq	-0.001***	-0.000***	-0.001***	-0.000***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
edu	0.105***	0.068***	0.104***	0.068***	0.096***	0.065***	0.096***	0.065***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
urban	0.111***	0.100***	0.125***	0.101***	0.166***	0.057***	0.165***	0.057***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.007)	(0.007)	(0.007)	(0.007)
formal	0.224***	0.201***	0.216***	0.194***	0.214***	0.212***	0.214***	0.212***
	(0.006)	(0.005)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
focc3			-0.179***	-0.153***			-0.023**	0.018
			(0.013)	(0.018)			(0.009)	(0.013)
Regions FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupations FE	No	Yes	No	Yes	No	Yes	No	Yes
N	62202	62202	62202	62202	86758	86758	86758	86758
r2	0.404	0.460	0.406	0.461	0.429	0.488	0.429	0.488

* p<0.10, ** p<0.05, *** p<0.01

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

Table A3: Wage equations for the OB decomposition for non-white and white workers - all labour market

PANEL A - year 1987

	(1) Non-whites b/se	(2) Non-whites b/se	(3) Non-whites b/se	(4) Non-whites b/se	(5) Whites b/se	(6) Whites b/se	(7) Whites b/se	(8) Whites b/se
male	0.523***	0.387***	0.516***	0.387***	0.454***	0.332***	0.444***	0.332***
	(0.007)	(0.008)	(0.007)	(0.008)	(0.007)	(0.008)	(0.007)	(0.008)
age	0.091***	0.081***	0.087***	0.081***	0.104***	0.094***	0.100***	0.094***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
agesq	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
edu	0.113***	0.078***	0.088***	0.078***	0.134***	0.096***	0.102***	0.095***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
urban	0.267***	0.195***	0.253***	0.195***	0.287***	0.201***	0.250***	0.200***
	(0.009)	(0.010)	(0.009)	(0.010)	(0.010)	(0.011)	(0.010)	(0.011)
formal	0.186***	0.166***	0.147***	0.166***	0.133***	0.086***	0.078***	0.085***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
nwocc3			-1.629***	-0.224***			-1.899***	-0.691***
			(0.044)	(0.077)			(0.038)	(0.070)
Regions FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupations FE	No	Yes	No	Yes	No	Yes	No	Yes
N	45695	45695	45695	45695	51984	51984	51984	51984
r2	0.435	0.481	0.454	0.481	0.522	0.565	0.547	0.565

* p<0.10, ** p<0.05, *** p<0.01

PANEL B - year 2006

	(1) Non-whites b/se	(2) Non-whites b/se	(3) Non-whites b/se	(4) Non-whites b/se	(5) Whites b/se	(6) Whites b/se	(7) Whites b/se	(8) Whites b/se
male	0.214***	0.197***	0.228***	0.202***	0.272***	0.230***	0.275***	0.235***
	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)
age	0.061***	0.057***	0.061***	0.056***	0.067***	0.061***	0.068***	0.061***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
agesq	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
edu	0.083***	0.056***	0.065***	0.055***	0.116***	0.077***	0.085***	0.075***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
urban	0.151***	0.066***	0.139***	0.069***	0.153***	0.082***	0.136***	0.090***
	(0.007)	(0.008)	(0.007)	(0.008)	(0.010)	(0.010)	(0.009)	(0.010)
formal	0.272***	0.250***	0.242***	0.252***	0.175***	0.163***	0.145***	0.166***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
nwocc3			-1.476***	-0.529***			-1.923***	-1.105***
			(0.028)	(0.054)			(0.028)	(0.055)
Regions FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupations FE	No	Yes	No	Yes	No	Yes	No	Yes
N	79094	79094	79094	79094	69866	69866	69866	69866
r2	0.350	0.404	0.374	0.405	0.429	0.489	0.469	0.492

* p<0.10, ** p<0.05, *** p<0.01
Source: Author's computations using PNAD 1987 and 2006.

Table A4: Wage equations with Heckman correction for female and male sample – 1987 and 2006 (with first step) – FOURTH SPECIFICATION (WITH FOCC3 AND OCC FE)

	1987 Females b/se	1987 Males b/se	2006 Females b/se	2006 Males b/se
main				
white	0.092***	0.112***	0.087***	0.104***
	(0.008)	(0.006)	(0.006)	(0.005)
age	0.094***	0.080***	0.053***	0.062***
	(0.002)	(0.003)	(0.002)	(0.002)
agesq	-0.001***	-0.001***	-0.000***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
edu	0.093***	0.089***	0.067***	0.065***
	(0.001)	(0.001)	(0.001)	(0.001)
urban	0.315***	0.168***	0.100***	0.057***
	(0.013)	(0.009)	(0.011)	(0.008)
formal	0.159***	0.076***	0.193***	0.212***
	(0.009)	(0.007)	(0.006)	(0.005)
focc3	-0.392***	0.094***	-0.153***	0.018
	(0.026)	(0.020)	(0.018)	(0.013)
Regions FE	Yes	Yes	Yes	Yes
Occupations FE	Yes	Yes	Yes	Yes
part				
white	-0.171***	-0.031**	-0.051***	0.018*
	(0.010)	(0.013)	(0.008)	(0.010)
age	0.151***	0.221***	0.189***	0.216***
	(0.002)	(0.003)	(0.002)	(0.002)
agesq	-0.002***	-0.003***	-0.002***	-0.003***
	(0.000)	(0.000)	(0.000)	(0.000)
edu	0.044***	0.004**	0.060***	0.032***
	(0.001)	(0.002)	(0.001)	(0.001)
urban	0.107***	-0.482***	0.114***	-0.478***
	(0.013)	(0.018)	(0.012)	(0.014)
hijos	-0.174***	-0.143***	-0.200***	-0.240***
	(0.014)	(0.018)	(0.011)	(0.013)
relacion_ci==Conyuge	-0.626***	-0.600***	-0.390***	-0.089***
	(0.011)	(0.110)	(0.008)	(0.021)
nmenor6_ch	-0.067***	0.215***	-0.028***	0.161***
	(0.007)	(0.010)	(0.006)	(0.008)
nmenor1_ch	-0.162***	0.177***	-0.208***	0.074***
	(0.018)	(0.030)	(0.017)	(0.022)
nmayor65_ch	-0.018	-0.207***	-0.167***	-0.302***
	(0.013)	(0.016)	(0.010)	(0.010)
domwork	0.799***	-0.100**	0.805***	0.014
	(0.023)	(0.040)	(0.043)	(0.069)
other	0.331***	-0.496***	-0.324***	-1.131***
	(0.013)	(0.016)	(0.009)	(0.014)
mills				
lambda	0.140***	-0.119***	-0.006	-0.004
	(0.015)	(0.029)	(0.017)	(0.013)
N	86467	75168	128316	115552
r2				

* p<0.10, ** p<0.05, *** p<0.01
Source: Author’s computations using PNAD 1987 and 2006.

Table A5: Wage equations with Heckman correction for non-white and white sample – 1987 and 2006 (with first step) FOURTH SPECIFICATION (WITH NWOCC3 AND OCC FE)

	1987 Non-whites b/se	1987 Whites b/se	2006 Non-whites b/se	2006 Whites b/se
main				
male	0.475***	0.406***	0.164***	0.232***
	(0.014)	(0.014)	(0.008)	(0.008)
age	0.090***	0.103***	0.048***	0.060***
	(0.002)	(0.002)	(0.002)	(0.002)
agesq	-0.001***	-0.001***	-0.000***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
edu	0.080***	0.097***	0.053***	0.074***
	(0.001)	(0.001)	(0.001)	(0.001)
urban	0.190***	0.196***	0.074***	0.090***
	(0.010)	(0.010)	(0.008)	(0.010)
formal	0.166***	0.088***	0.250***	0.166***
	(0.008)	(0.007)	(0.005)	(0.005)
nwocc3	-0.223***	-0.688***	-0.529***	-1.105***
	(0.075)	(0.071)	(0.049)	(0.052)
Regions FE	Yes	Yes	Yes	Yes
Occupations FE	Yes	Yes	Yes	Yes
part				
male	0.943***	0.873***	0.623***	0.564***
	(0.013)	(0.012)	(0.009)	(0.010)
age	0.179***	0.188***	0.205***	0.213***
	(0.002)	(0.002)	(0.002)	(0.002)
agesq	-0.002***	-0.002***	-0.002***	-0.003***
	(0.000)	(0.000)	(0.000)	(0.000)
edu	0.027***	0.033***	0.041***	0.050***
	(0.002)	(0.001)	(0.001)	(0.001)
urban	-0.134***	-0.114***	-0.147***	-0.177***
	(0.014)	(0.014)	(0.011)	(0.014)
hijos	-0.156***	-0.211***	-0.203***	-0.228***
	(0.016)	(0.014)	(0.011)	(0.012)
relacion_ci==Conyuge	-0.824***	-0.845***	-0.416***	-0.483***
	(0.015)	(0.014)	(0.010)	(0.011)
nmenor6_ch	0.022***	0.055***	0.058***	0.072***
	(0.007)	(0.008)	(0.006)	(0.009)
nmenor1_ch	-0.021	-0.087***	-0.104***	-0.114***
	(0.020)	(0.021)	(0.016)	(0.021)
nmayor65_ch	-0.114***	-0.108***	-0.244***	-0.228***
	(0.015)	(0.014)	(0.009)	(0.010)
domwork	0.932***	0.369***	0.781***	0.361***
	(0.033)	(0.026)	(0.050)	(0.053)
other	-0.076***	0.041***	-0.582***	-0.655***
	(0.015)	(0.013)	(0.010)	(0.011)
mills				
lambda	0.144***	0.122***	-0.087***	-0.008
	(0.018)	(0.019)	(0.014)	(0.015)
N	74449	87186	131894	111974
r2				

* p<0.10, ** p<0.05, *** p<0.01

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

Table A6: Wage equations with Lee correction for female and male sample

	(2) 1987 Females b/se	(4) 1987 Males b/se	(2) 2006 Females b/se	(4) 2006 Males b/se
white	0.097***	0.112***	0.088***	0.101***
	(0.008)	(0.006)	(0.006)	(0.005)
age	0.092***	0.084***	0.056***	0.064***
	(0.002)	(0.001)	(0.002)	(0.001)
agesq	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
edu	0.092***	0.089***	0.070***	0.067***
	(0.002)	(0.001)	(0.001)	(0.001)
urban	0.320***	0.159***	0.106***	0.076***
	(0.014)	(0.008)	(0.012)	(0.007)
focc3	-0.463***	0.098***	-0.241***	0.009
	(0.024)	(0.019)	(0.018)	(0.013)
Regions FE	Yes	Yes	Yes	Yes
Occupations FE	Yes	Yes	Yes	Yes
IMR	0.110***	-0.103***	-0.034***	-0.092***
	(0.010)	(0.007)	(0.009)	(0.007)
N	35077	62602	62202	86758
r2	0.581	0.529	0.452	0.478

* $p<0.10$, ** $p<0.05$, *** $p<0.01$
Source: Author's own computations using PNAD 1987 and 2006.

Table A7: Wage equations with Lee correction for non-white and white sample

	(2) 1987 Non-whites b/se	(4) 1987 Whites b/se	(2) 2006 Non-whites b/se	(4) 2006 Whites b/se
male	0.417***	0.319***	0.223***	0.222***
	(0.009)	(0.008)	(0.006)	(0.006)
age	0.085***	0.092***	0.062***	0.059***
	(0.002)	(0.002)	(0.001)	(0.001)
agesq	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
edu	0.081***	0.095***	0.060***	0.075***
	(0.001)	(0.001)	(0.001)	(0.001)
urban	0.204***	0.206***	0.082***	0.099***
	(0.010)	(0.011)	(0.008)	(0.010)
nwocc3	-0.247***	-0.690***	-0.450***	-1.031***
	(0.077)	(0.070)	(0.055)	(0.055)
Regions FE	Yes	Yes	Yes	Yes
Occupations FE	Yes	Yes	Yes	Yes
IMR	0.009	-0.048***	-0.034***	-0.091***
	(0.008)	(0.008)	(0.008)	(0.008)
N	45695	51984	79094	69866
r2	0.476	0.565	0.387	0.486

Table A8: Local gap estimates using Machado (2011) methodology

PANEL A – Gender local wage gaps across years

	(1) 1987 b/se	(1) 1992 b/se	(1) 1997 b/se	(1) 2002 b/se	(1) 2006 b/se
Female	-0.273***	-0.152***	-0.070***	-0.001	-0.005
	(0.008)	(0.007)	(0.007)	(0.006)	(0.005)
Female#kid6	-0.141***	-0.183***	-0.149***	-0.211***	-0.190***
	(0.011)	(0.011)	(0.010)	(0.009)	(0.008)
N	98094	101652	113826	135257	151581
r2	0.024	0.013	0.005	0.005	0.005
Local gap	-0.414	-0.335	-0.219	-0.212	-0.196
s.e.	0.010	0.010	0.009	0.008	0.008

PANEL B – Racial local wage gaps across years

	(1) 1987 b/se	(1) 1992 b/se	(1) 1997 b/se	(1) 2002 b/se	(1) 2006 b/se
Non-white	-0.484***	-0.514***	-0.513***	-0.442***	-0.381***
	(0.008)	(0.007)	(0.006)	(0.006)	(0.005)
Non-white #kid6	-0.010	-0.075***	-0.069***	-0.120***	-0.133***
	(0.009)	(0.009)	(0.009)	(0.007)	(0.007)
N	98094	101652	113826	135257	151581
r2	0.056	0.072	0.076	0.069	0.059
Local gap	-0.494	-0.589	-0.581	-0.561	-0.514
s.e.	0.008	0.009	0.008	0.007	0.006

* p<0.10, ** p<0.05, *** p<0.01

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

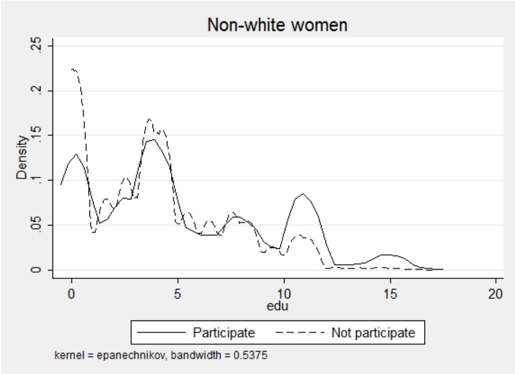
Table A9: Differentials in participation rates between 1987 and 2006 by different sub-groups of population (and by levels of education)

	All sample	Zero years of edu	1 - 11 years of edu	11+ years of edu
Females	0.077	-0.031	0.063	-0.038
Males	-0.081	-0.14	-0.077	-0.06
Non-whites	-0.014	-0.065	-0.017	-0.082
Whites	0.027	-0.057	0.021	-0.043
NW F	0.049	-0.045	0.033	-0.1
W F	0.107	-0.012	0.091	-0.013
NW M	-0.091	-0.14	-0.081	-0.051
W M	-0.07	-0.151	-0.071	-0.061

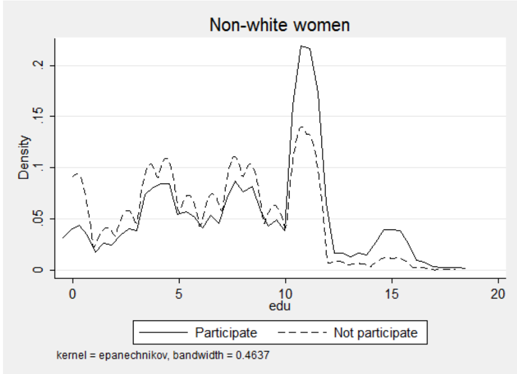
Source: Author's computations using PNAD 1987 and 2006.

Figure B1: Density distribution of years of education by participants and not participants – 1987 and 2006

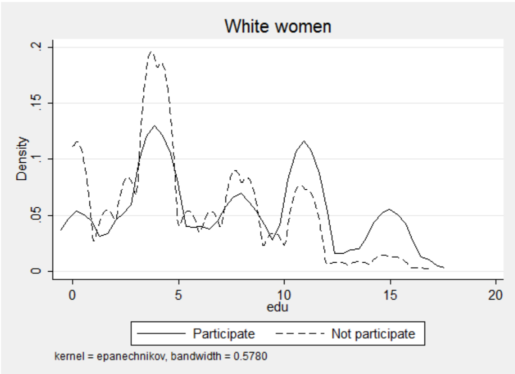
1987



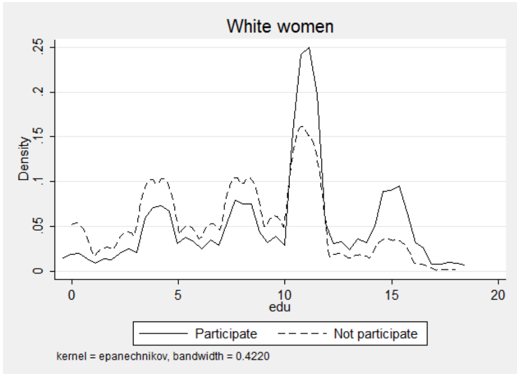
2006



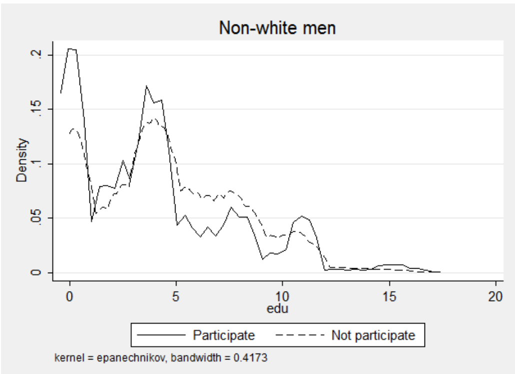
1987



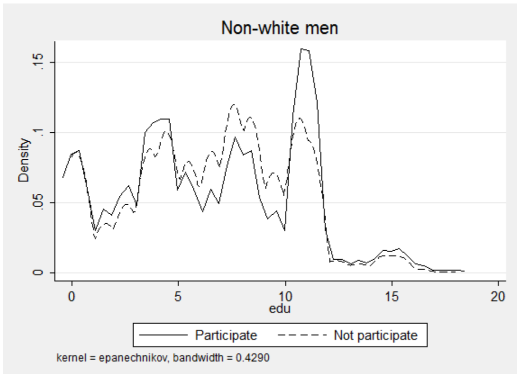
2006



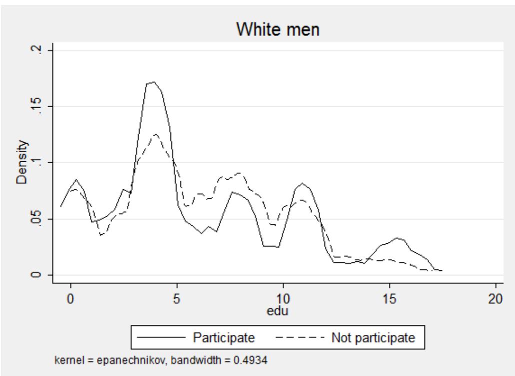
1987



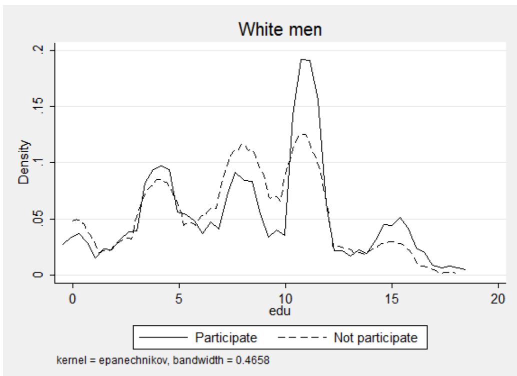
2006



1987



2006



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

Chapter 7

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

7.1 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.

In the previous chapter we addressed the second limitation by decomposing wage gaps using the Brown, Moon and Zoloth (1980) decomposition technique, which accounts for the impact of occupational segregation. In this chapter we go further by addressing both of these limitations. This chapter 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. This focus on occupational intensity is made possible by the use of the harmonized occupational codes explained in chapter 4.

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.

Complementing the analysis undertaken in previous chapters, the empirical analysis presented here also makes three 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). Finally, we not only analyse the entire labour market but also disaggregate the analysis between the formal and non-formal sectors, as has been the case in previous chapters, although we consign this part of the investigation to an appendix.

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 generally remained stable over time, while the effect of the female occupational intensity variable has, on average, declined 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 chapter 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.

7.2 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 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). [88] 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[89] 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 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 chapter 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.

---

[88] The Juhn, Murphy and Pierce (1991) methodology has been subject to several criticisms, sumamrized by Yun (2009). 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).

[89] 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.

In analyzing gender and racial wage gaps in Brazil, this study builds on a large number of existing studies (see the review in chapter 6). 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, and chapter 6 of this study). 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 chapter 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. Third, 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. Finally, though not the primary focus of the paper, it modestly contributes to the analysis of informality within the Brazilian labour market by presenting results in the appendix disaggregated between formal and non-formal sectors.

7.3 Data and overview of wage gaps

The analysis in this chapter employs the same data used in earlier chapters. 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).

It is important to again highlight that, unlike many studies of Brazil, we initially consider the entire labour market by including civil servants, domestic workers and individuals involved in agricultural activities across all five regions of Brazil, in both urban and rural areas. In previous chapters we have considered the entire labour market as well as grouping the labour force into three main sectors: formal, informal and self-employed. However, due to space limitations, in this chapter we focus the analysis on the entire labour market only, while consigning the disaggregated analysis to a comparatively brief section in an appendix.

The analysis presented here is, again, crucially dependent on the use of our newly constructed 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 and described in chapter 4 and employed previously in the empirical analysis in both chapters 5 and 6. 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 as already pointed out in the previous chapter. For this reason, we employ an austere wage equation specification, which has nonetheless proven to have high explanatory power in the earlier chapter.

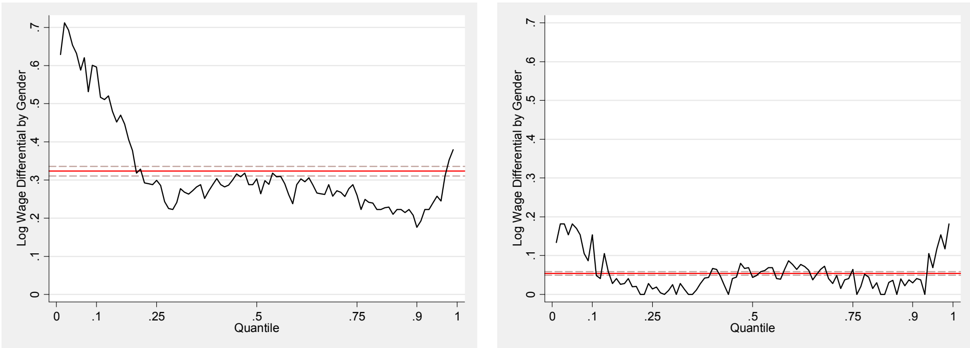
Having reviewed the main features of the data employed in this chapter, we now report some preliminary descriptive analysis. Figure 1 illustrates the distribution of wage gaps across the wage distribution by both gender and race (the plots to the left side are for the first year, 1987, and the plots to the right side are for the last year, 2006), presenting data at selected points of the wage distribution (0.1, 0.25, 0.5, 0.75 and 0.9). 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.[90]

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.

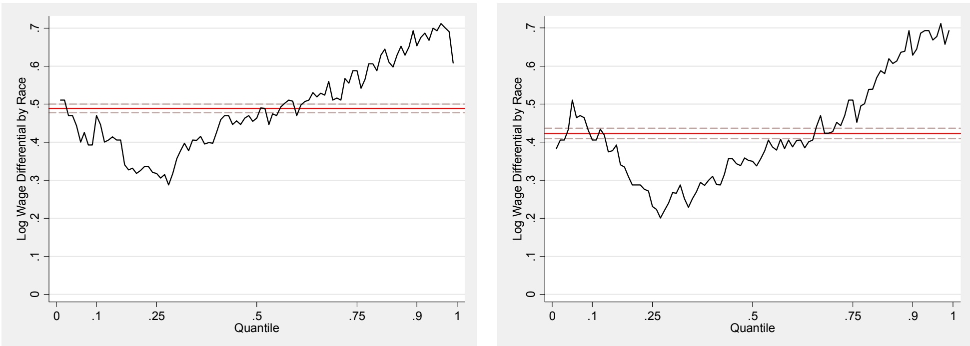
[90] The concepts of 'glass ceilings' and 'sticky floors' are, in fact, closely related. Glass ceilings are invisible but concrete barriers that prevent career advancement and restrict minorities from reaching the best paying and most prestigious occupations, despite their characteristics. Sticky floors 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).

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.

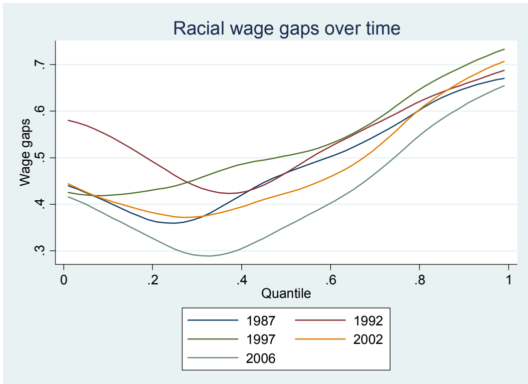
Figure 2 provides a more general portrait of both gender and racial wage gaps for five years spanning the entire period (1987, 1992, 1997, 2002 and 2006). These plots reaffirm the key findings from figure 1. First, we again see that gender wage gaps are wider at the bottom of the wage distribution, while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

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.[91]

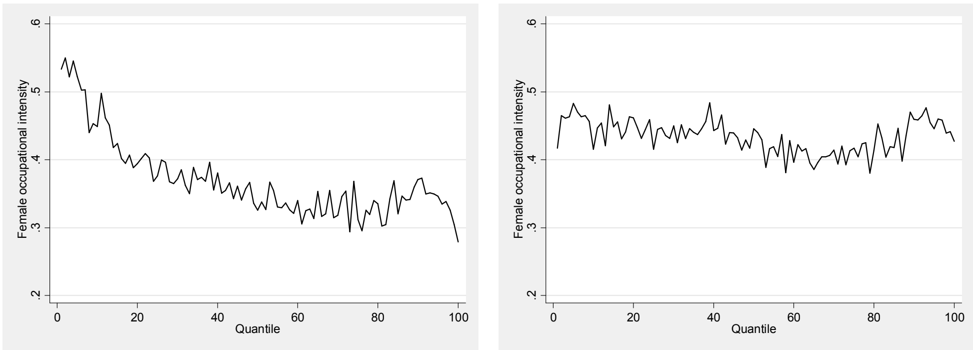
Figure 3 provides additional insights into how female and non-white occupational intensity vary across wage quantiles. The values of the female and non-white occupational intensity variables at different quantiles of the wage distribution are derived using a variation of the Machado and Mata (2005) approach, 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 60th and the 80th 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.

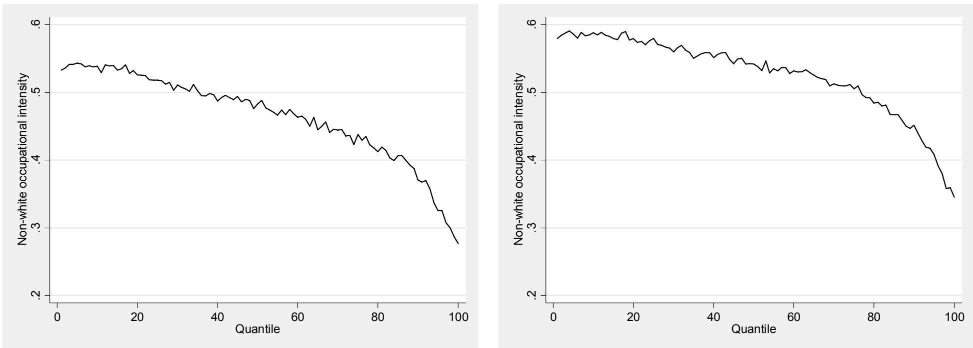
[91] Both female and non-white occupational intensity have, on average, increased over time (by 7 and 6 percentage points, respectively). However, female occupational intensity has increased more homogenously across occupations than non-white occupational intensity. These patterns are consistent with the findings about occupational segregation presented in chapter 5, where we found a sizeable decline in gender segregation but only a small contraction in racial segregation.

Figure 3: Occupational intensity over wage quantiles

Panel A- Female occupational intensity, 1987 and 2006



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



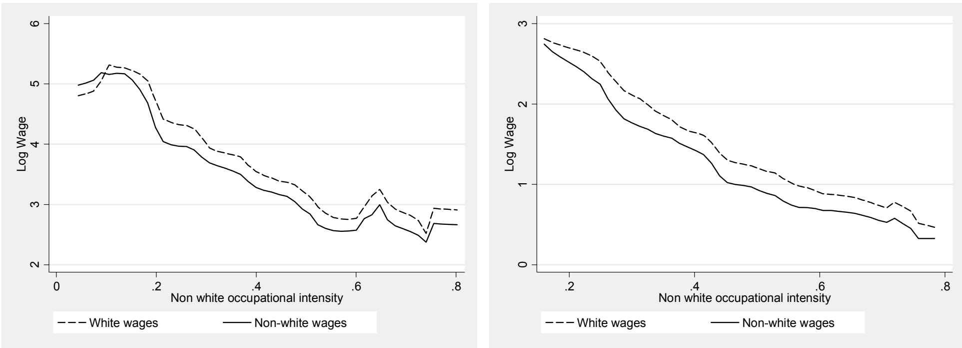
Source: Author's computations using PNAD 1987 and 2006.

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 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: 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.

7.4 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).

7.4.1 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).[92]

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

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

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:

$$E[W_B] - E[W_A] = \{E[W_B|D_B = 1] - E[W_A|D_B = 1]\} + \{E[W_A|D_B = 1] - E[W_A|D_A = 1]\} \tag{2}$$

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:

$$ATT = E[W_B|D_B = 1] - E[W_A|D_B = 1] \tag{3}$$

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:

$$E[W_B] - E[W_A] = \{E[W_B|D_B = 1] - E[W_B|D_A = 1]\} + \{E[W_B|D_A = 1] - E[W_A|D_A = 1]\} \tag{4}$$

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:

$$ATNT = E[W_B|D_A = 1] - E[W_A|D_A = 1] \tag{5}$$

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).

7.4.2. 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), Pham and Reilly (2007) and Melly (2006). 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.[93]

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).[94]

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 decomposition using simulation methods, but with the drawback that it is

[93] 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.

[94] 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.

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.

7.4.3. Estimation of counterfactual distributions using RIF-regression

An important limitation of the Melly (2006) decomposition technique is that it does 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.[95] 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.[96] 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.

---

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

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

The primary limitation of this methodology lies in the linear approximation of a non-linear distributional function. This decomposition procedure provides only a first-order approximation of the composition effects and this approximation is not precise and 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).[97]

7.4.4 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.

[97] 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.

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 unconfoundedness (or ignorability) assumption does not hold, and the decomposition terms are not identified correctly. Machado (2011) invokes three different self-selection cases (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 Buchinsky (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 Buchinsky (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; Buchinsky , 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 chapter 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 chapter is already dense.

7.5 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.

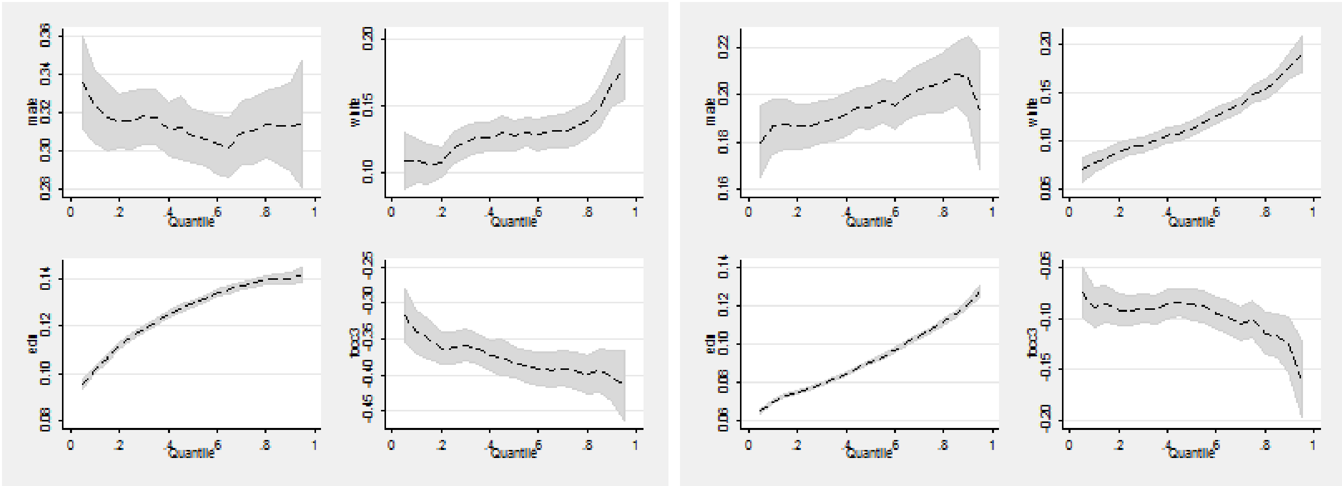
7.5.1 Quantile regression estimates: the effect of female and non-white occupational intensity

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, following the strategy employed in the analysis of mean wage regressions in chapter 6. 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.

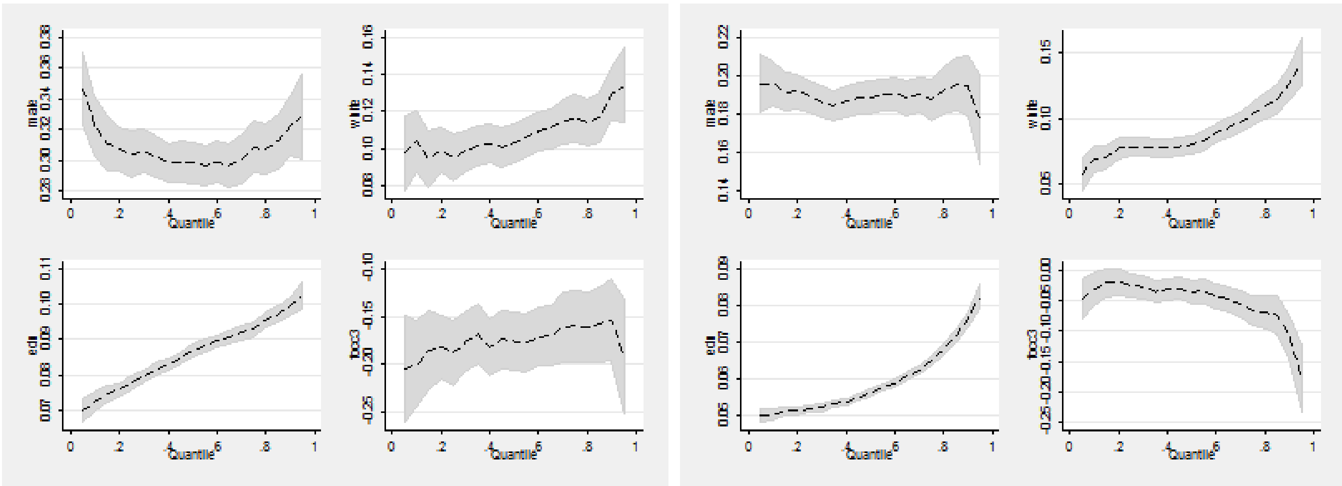
To conserve space we report the pooled quantile regressions only for the third and fourth specifications; that is, those that include female or non-white occupational intensity (focc3 or nwocc3), with and without dummies for occupations (i.e., occupation effects). These regressions are presented in tables A1 and A2 of the appendix for the years 1987 and 2006 respectively. In these tables we consider both the mean regressions and quantile regressions at 0.01, 0.10, 0.25, 0.50, 0.75, 0.90 and 0.99 quantiles. Standard errors are bootstrapped using 200 replications. Inter-quantile regression estimates are also reported in table A3 of the appendix in order to test the statistical significance of differences across the main quantiles (i.e., the 10^{th}, the 50^{th} and the 90^{th} quantiles). While these detailed results are confined to an appendix, 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: Main covariates' effect (including focc3) from pooled quantile regressions

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



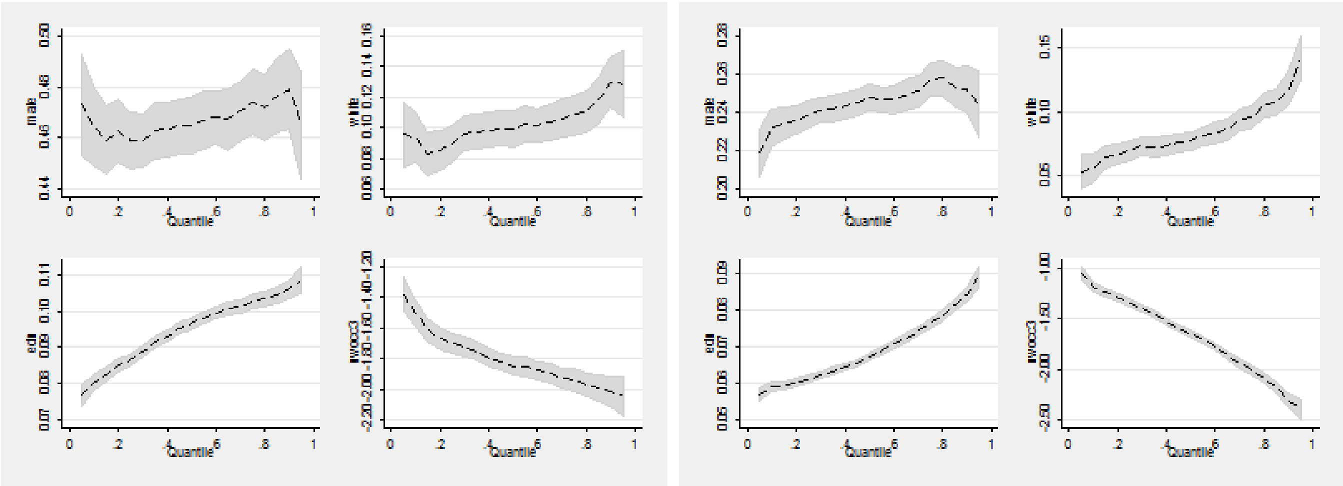
Panel B – with occupations (using the 4th specification), 1987 and 2006



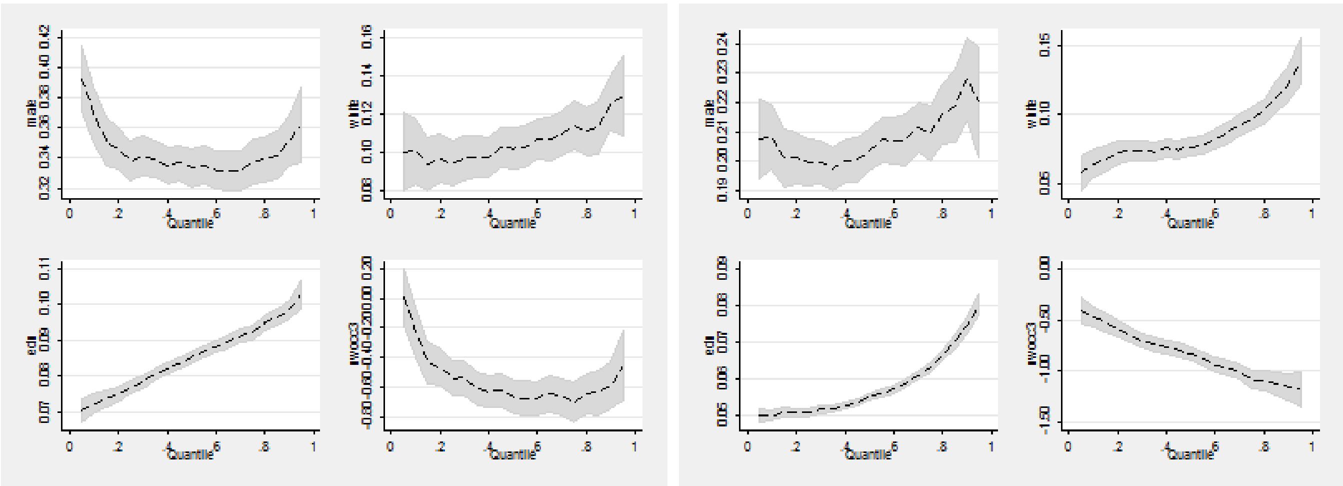
Source: Author's computations using PNAD 1987 and 2006. Note: Bootstrapped standard errors using 200 replications.

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.

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.[98] 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 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.

---

[98] 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 Buchinsky (1998)).

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 the current chapter. 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: 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.

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: 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.

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.

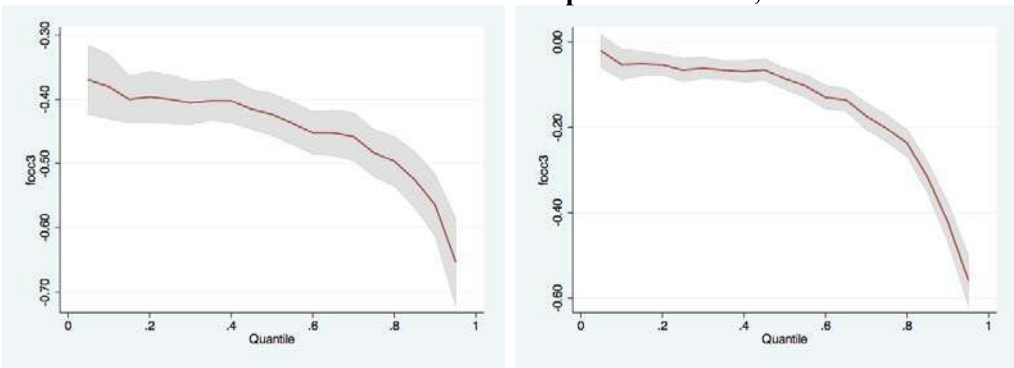
Turning to differences by race, employment in non-white dominated occupations reduces wages for both non-white and white workers, though the effect is slightly more pronounced for white workers. This negative effect increases, in absolute terms, as we move up the distribution, independent of whether 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 the effect of the female occupational intensity variable has, on average, declined over time.

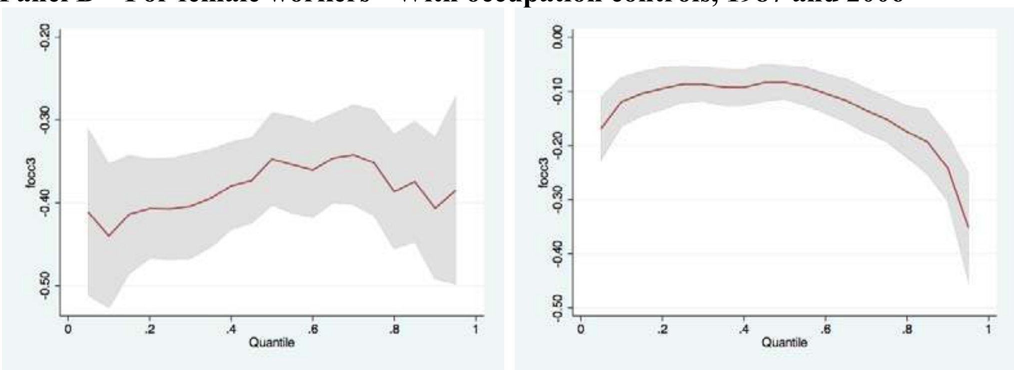
It is important to re-emphasize that many of these estimates represent a novel contribution to existing research, not only for Brazil but internationally, and thus highlight important, but previously overlooked, aspects of wage determination. These new insights fall into three broad categories. First, 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: The role of female occupational intensity

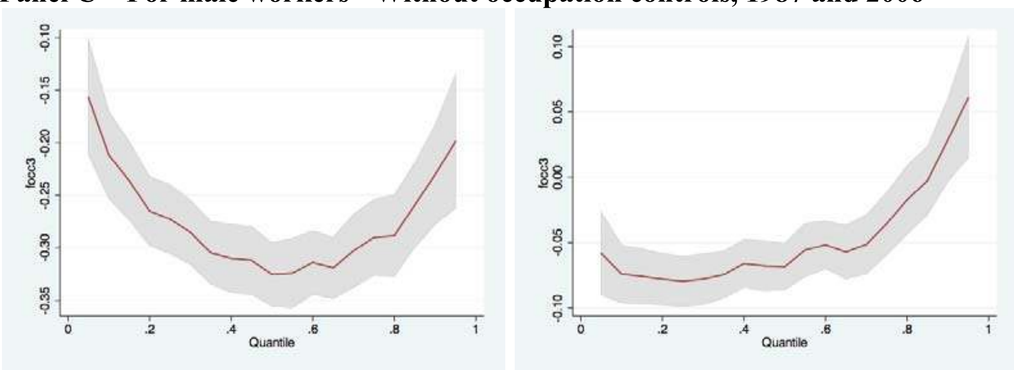
Panel A – For female workers - Without occupation controls, 1987 and 2006



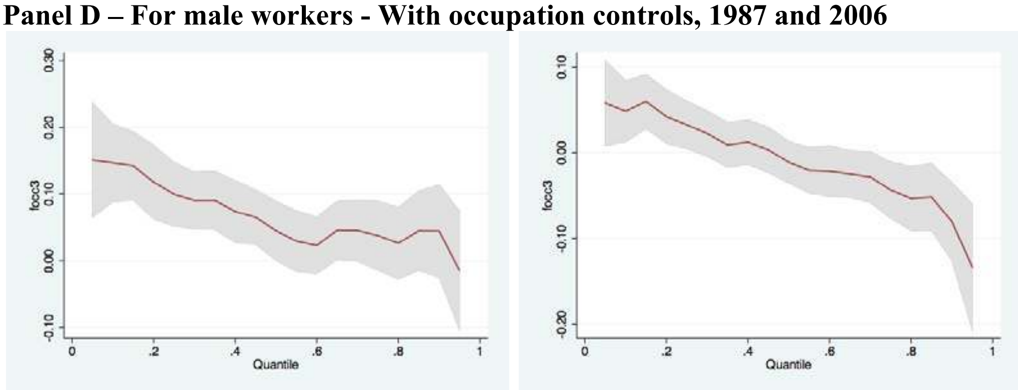
Panel B – For female workers - With occupation controls, 1987 and 2006



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



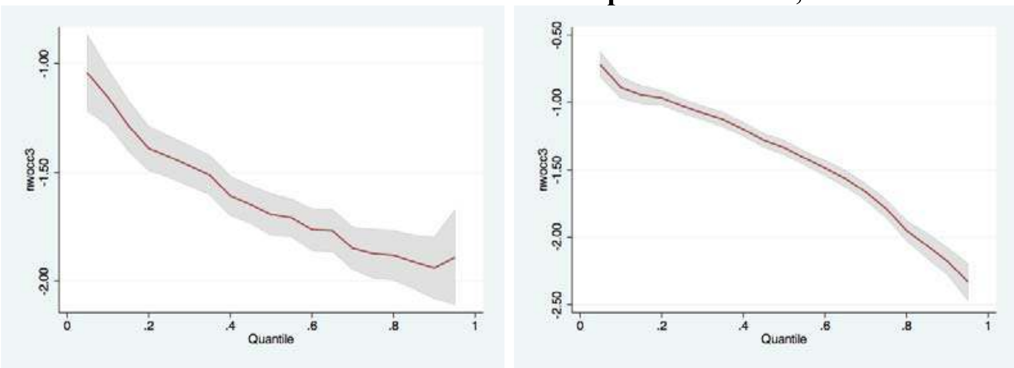
Panel D – For male workers - With occupation controls, 1987 and 2006



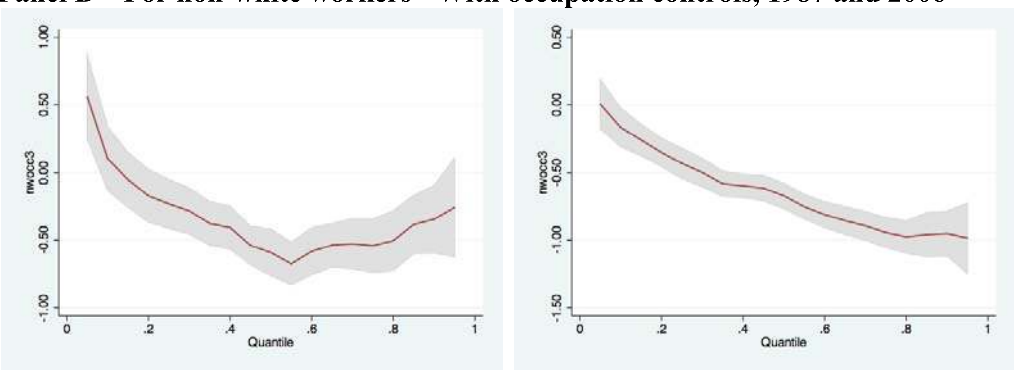
Source: Author's computations using PNAD 1987 and 2006.
Note: Panels A and C correspond to the 3rd specification of the wage equation, while panels B and D correspond to the 4th specification.

Figure 8: The role of non-white occupational intensity

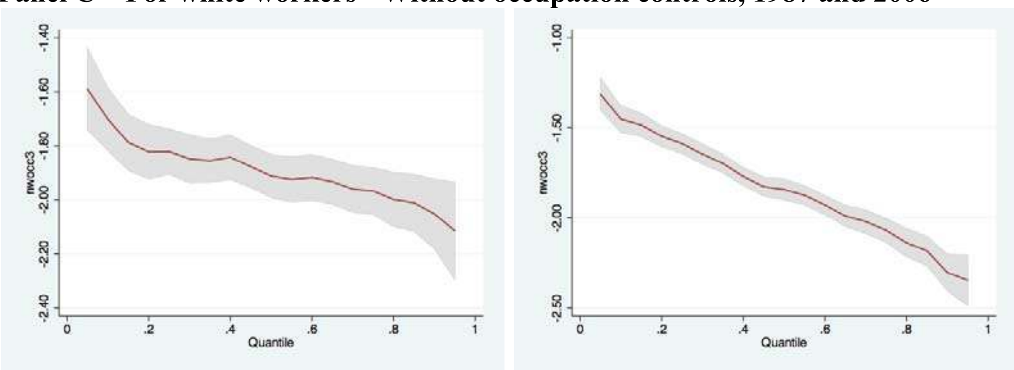
Panel A – For non-white workers - Without occupation controls, 1987 and 2006



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.

7.5.2 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.[99] 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.[100]

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.[101] In the upper panels of tables 3 and 4 we

[99] 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).

[100] 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'.

[101] 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.

report the quantile decomposition results using the most complete wage equation specification (the 4th specification).[102] The first three panels report the Machado and Mata (2005) and Melly (2006) 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.

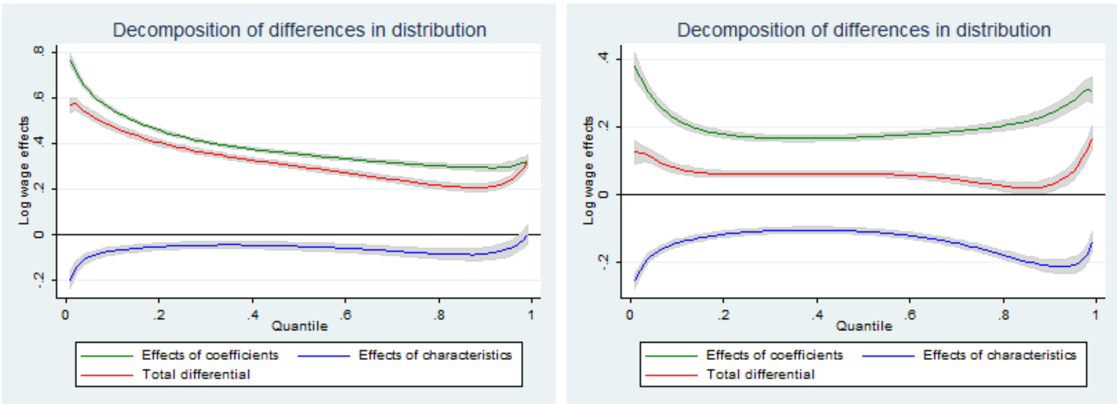
When we turn to the results for 2006 we see that the size of wage gaps has contracted over time, while differences across the wage distribution have also declined. Wage gaps have fallen most rapidly at the bottom end of the wage distribution, with this reduction explained primarily by a decline in the 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 10th, 50th and 90th 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

[102] 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 while consigning to the appendix the tables for the other specifications (tables A4 to A9).

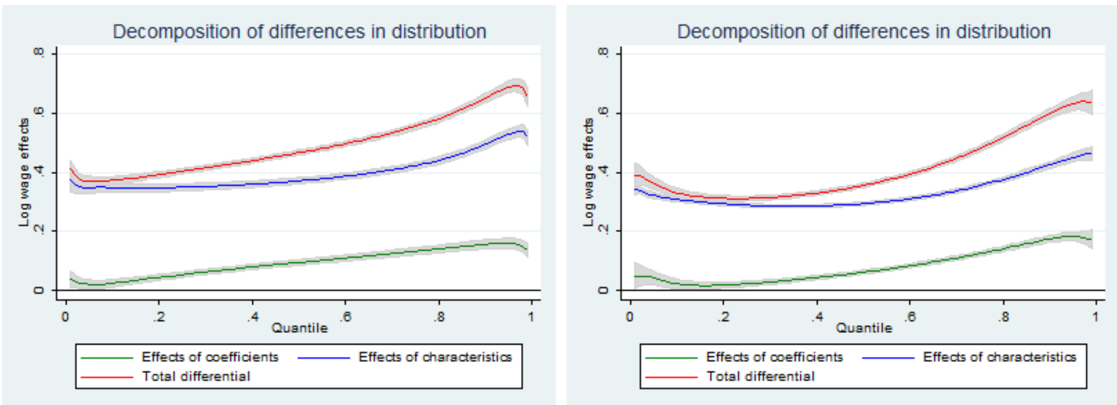
(2009) similarly detect a U-shaped pattern of unequal treatment in computing gender wage gaps in Brazil using an alternative non-parametric matching decomposition methodology (see also in Ñopo, 2012: 171). In addition, this U-shaped pattern has been similarly found in other South American countries, such as Chile and Colombia (Ñopo, 2012).

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.

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: 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.

7.5.3 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 were described at length in chapter 6, and in tables 5a and 5b we report the OB decomposition results at average values while adding the decomposition results. This allows us to compare the detailed decomposition results from the RIF-OLS procedure to these mean results using the OB technique.[103] 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

[103] When implementing detailed decomposition analysis, we encounter the problem of choosing a base group, as explained in the methodological section of the previous chapter. 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.

occupational controls is more muted in 2006 amidst the broader decline in gender 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: 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.

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.

7.5.4 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.

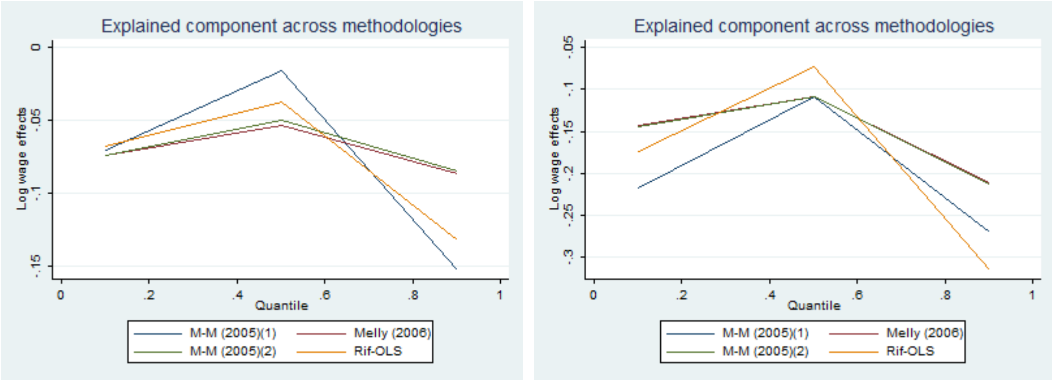
Figures 10 and 11 plot the decomposition results across these methodologies for gender and racial wage gaps, respectively, looking separately at the explained component, the unexplained component and the total gap. These figures are based on the results reported in tables 3 and 4, which are computed 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.

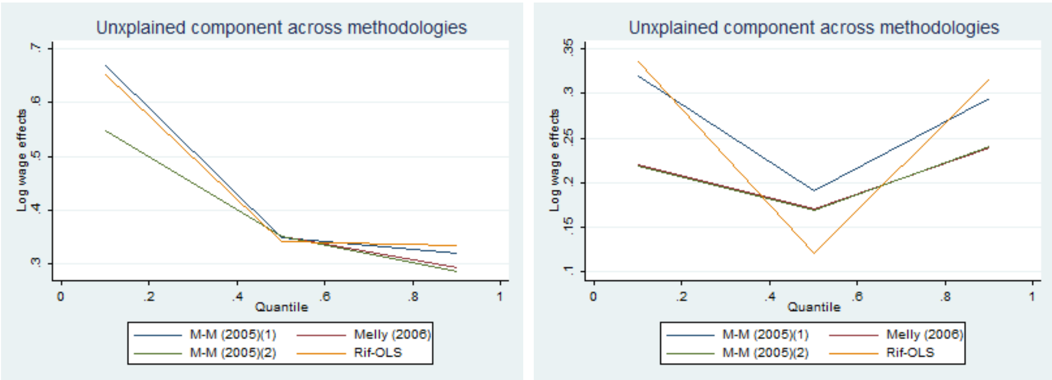
Where we notice differences between the methods, these tend to occur at the extremes of the wage distribution, and most notably at the 10^{th} quantile for gender and the 90^{th} quantile for race. By contrast, the median decomposition results are less likely to differ across methods. Overall, the similarity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

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

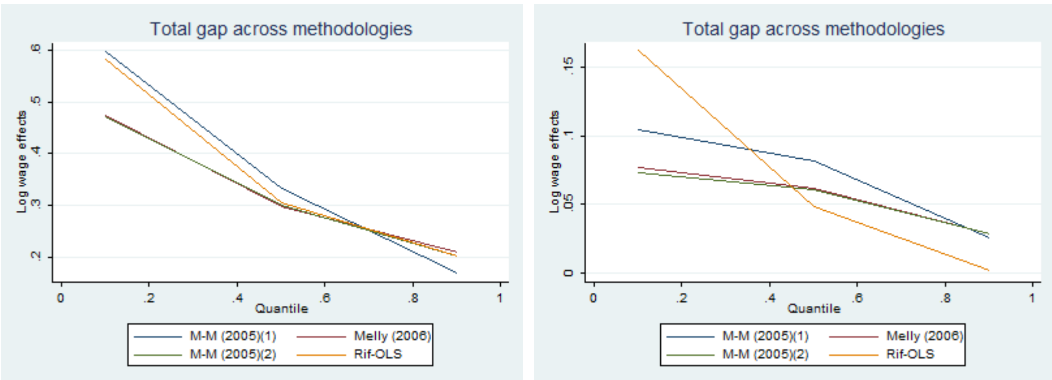
Panel A – Explained component (effect of characteristics)



Panel B – Unexplained component (effect of coefficients)



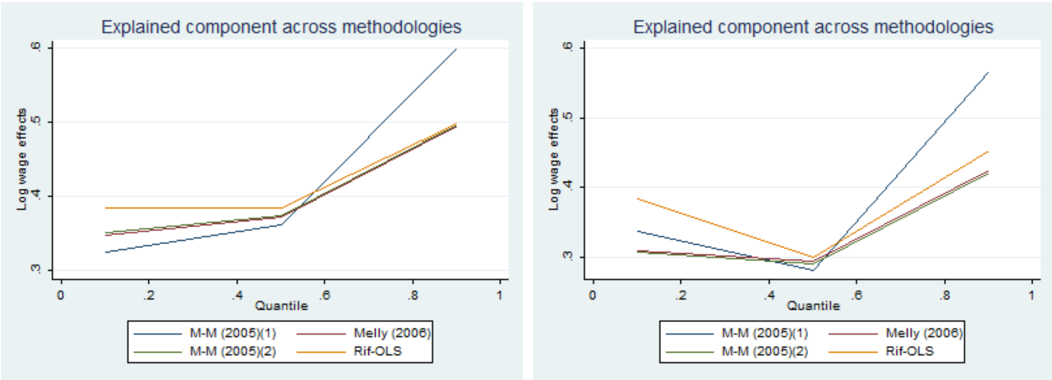
Panel C- Total gap



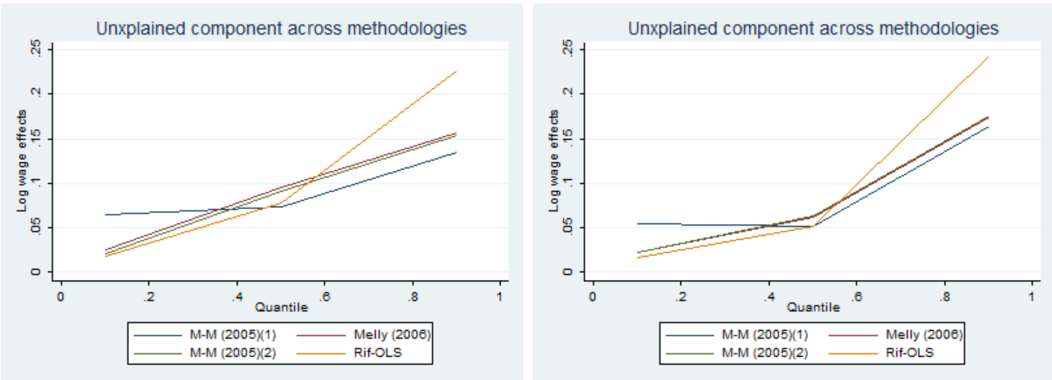
Source: Author's computations using PNAD 1987 and 2006.
Note: M-M (2005) (1) refers to Machado and Mata (2005) as Albrecht, Bjorklund and Vroman (2003); M-M (2005) (2) refers to Machado and Mata (2005) as Albrecht, van Vuuren and Vroman (2009).

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).

7.6 Conclusions

The chapter 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. This builds on the analysis in previous chapters, which has highlighted the significance of occupational segregation in the Brazilian labour market, with female occupational segregation high but declining and non-white occupational segregation less pronounced, but more persistent.

The chapter 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 generally remained stable over time, while the effect of the female occupational intensity variable has, on average, declined 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 chapter, 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. While not the primary focus of this paper, additional insights into this pattern emerge when we disaggregate the impact of the wage structure component between the formal and non-formal labour markets (reported in Appendix B). The wage structure effect is greater at higher quantiles in the formal market, while in the non-formal sectors the effect of coefficients is considerably greater at the bottom end of the distribution. This suggests the existence of a *sticky floor* phenomenon for women working in non-formal sectors, while also revealing the existence of persistent *glass ceilings* in the formal sector where, despite higher levels of endowment than men, women continue to receive lower wages.

Turning to racial wage differentials a single key message emerges across the formal and non-formal sectors: wage differentials tend to widen at higher wage quantiles due to both larger differences in characteristics in favour of white workers and higher returns to those characteristics, 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 formal jobs and in the low paid informal occupations. These trends reinforce the apparent existence of *sticky floors* in non-formal occupations and of *glass ceilings* in formal activities. Looking at racial wage gaps, occupational intensity again plays an important role, though in the opposite direction, as non-white workers receive higher wages in non white-dominated occupations, particularly 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 (the latter of which was explored in much more detail in the previous chapter). 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.

Appendices to Chapter 7

Appendix A

Table A1: Pooled quantile regressions, 1987

Panel A - Mean and quantile regressions for all sample (using focc3)

	(1) mean	(2) 1	(3) 10	(4) 25	(5) 50	(6) 75	(7) 90	(8) 99
male	0.326***	0.337***	0.324***	0.316***	0.308***	0.311***	0.313***	0.329***
	(0.007)	(0.028)	(0.009)	(0.008)	(0.007)	(0.008)	(0.011)	(0.024)
white	0.132***	0.091***	0.109***	0.119***	0.128***	0.134***	0.166***	0.146***
	(0.005)	(0.020)	(0.007)	(0.006)	(0.006)	(0.008)	(0.009)	(0.023)
age	0.095***	0.069***	0.082***	0.089***	0.095***	0.102***	0.105***	0.107***
	(0.001)	(0.005)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.005)
agesq	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
edu	0.129***	0.088***	0.101***	0.116***	0.130***	0.138***	0.140***	0.140***
	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
urban	0.289***	0.584***	0.323***	0.271***	0.248***	0.254***	0.248***	0.263***
	(0.007)	(0.041)	(0.011)	(0.009)	(0.007)	(0.008)	(0.011)	(0.030)
biformal	0.137***	0.689***	0.360***	0.226***	0.127***	0.044***	-0.046***	-0.180***
	(0.005)	(0.020)	(0.008)	(0.006)	(0.006)	(0.006)	(0.010)	(0.022)
focc3	-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)
Constant	0.264***	-1.073***	-0.305***	0.004	0.310***	0.619***	1.017***	1.771***
	(0.022)	(0.101)	(0.035)	(0.026)	(0.026)	(0.027)	(0.035)	(0.098)
Region effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	97679	97679	97679	97679	97679	97679	97679	97679
R2								
pr2	0.518	0.312	0.287	0.278	0.311	0.339	0.351	0.318

Note: bootstrapped s.e. 200 replications, * p<0.10, ** p<0.05, *** p<0.01.
Source: Author's computations using PNAD 1987.

Panel B - Mean and quantile regressions for all sample (using focc3 and occupations)

	(1) mean	(2) 1	(3) 10	(4) 25	(5) 50	(6) 75	(7) 90	(8) 99
male	0.324***	0.348***	0.323***	0.304***	0.298***	0.309***	0.323***	0.357***
	(0.006)	(0.032)	(0.011)	(0.008)	(0.006)	(0.008)	(0.011)	(0.026)
white	0.110***	0.099***	0.104***	0.095***	0.103***	0.116***	0.129***	0.109***
	(0.005)	(0.022)	(0.008)	(0.006)	(0.005)	(0.007)	(0.008)	(0.025)
age	0.087***	0.068***	0.076***	0.080***	0.086***	0.092***	0.094***	0.092***
	(0.001)	(0.006)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.005)
agesq	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
edu	0.090***	0.071***	0.072***	0.078***	0.087***	0.093***	0.099***	0.107***
	(0.001)	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)
urban	0.198***	0.380***	0.258***	0.212***	0.172***	0.158***	0.139***	0.115***
	(0.007)	(0.044)	(0.015)	(0.009)	(0.009)	(0.009)	(0.011)	(0.030)
biformal	0.108***	0.590***	0.334***	0.213***	0.104***	0.013*	-0.078***	-0.235***
	(0.005)	(0.025)	(0.009)	(0.006)	(0.006)	(0.007)	(0.009)	(0.029)
focc3	-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)
Constant	1.541***	0.260*	1.006***	1.385***	1.680***	1.916***	2.159***	2.954***
	(0.036)	(0.140)	(0.061)	(0.048)	(0.041)	(0.046)	(0.053)	(0.167)
Region effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occup. effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	97679	97679	97679	97679	97679	97679	97679	97679
r2	0.554							
pr2		0.329	0.312	0.308	0.342	0.368	0.379	0.348

Note: bootstrapped s.e. 200 replications, * p<0.10, ** p<0.05, *** p<0.01.
Source: Author's computations using PNAD 1987.

Panel C - Mean and quantile regressions for all sample (using nwocc3)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	mean	1	10	25	50	75	90	99
male	0.481***	0.446***	0.465***	0.459***	0.465***	0.474***	0.479***	0.460***
	(0.005)	(0.023)	(0.008)	(0.006)	(0.006)	(0.006)	(0.009)	(0.024)
white	0.104***	0.074***	0.094***	0.089***	0.099***	0.108***	0.130***	0.128***
	(0.005)	(0.019)	(0.008)	(0.007)	(0.006)	(0.007)	(0.009)	(0.027)
age	0.094***	0.071***	0.084***	0.088***	0.095***	0.098***	0.102***	0.103***
	(0.001)	(0.005)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.005)
agesq	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
edu	0.098***	0.075***	0.080***	0.086***	0.096***	0.103***	0.106***	0.118***
	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)
urban	0.246***	0.551***	0.285***	0.238***	0.209***	0.213***	0.197***	0.179***
	(0.007)	(0.042)	(0.011)	(0.008)	(0.007)	(0.008)	(0.012)	(0.034)
biformal	0.102***	0.663***	0.334***	0.203***	0.089***	-0.001	-0.090***	-0.251***
	(0.005)	(0.027)	(0.009)	(0.006)	(0.006)	(0.007)	(0.008)	(0.028)
nwocc3	-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)
Constant	1.153***	-0.563***	0.356***	0.814***	1.214***	1.621***	2.024***	2.717***
	(0.027)	(0.100)	(0.046)	(0.034)	(0.033)	(0.033)	(0.043)	(0.135)
Region effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	97679	97679	97679	97679	97679	97679	97679	97679
r2	0.532							
pr2		0.315	0.295	0.288	0.322	0.351	0.364	0.329

Note: bootstrapped s.e. 200 replications, * p<0.10, ** p<0.05, *** p<0.01.
Source: Author's computations using PNAD 1987.

Panel D - Mean and quantile regressions for all sample (using nwocc3 and occupations)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	mean	1	10	25	50	75	90	99
male	0.360***	0.378***	0.369***	0.338***	0.333***	0.338***	0.351***	0.379***
	(0.006)	(0.028)	(0.009)	(0.007)	(0.006)	(0.007)	(0.011)	(0.024)
white	0.108***	0.088***	0.100***	0.094***	0.102***	0.114***	0.125***	0.115***
	(0.005)	(0.021)	(0.008)	(0.006)	(0.006)	(0.007)	(0.009)	(0.022)
age	0.088***	0.070***	0.077***	0.081***	0.086***	0.092***	0.095***	0.093***
	(0.001)	(0.005)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.005)
agesq	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
edu	0.089***	0.071***	0.072***	0.077***	0.085***	0.092***	0.098***	0.107***
	(0.001)	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)
urban	0.196***	0.358***	0.259***	0.211***	0.170***	0.156***	0.141***	0.114***
	(0.007)	(0.051)	(0.016)	(0.008)	(0.009)	(0.009)	(0.011)	(0.035)
biformal	0.118***	0.598***	0.346***	0.223***	0.108***	0.019***	-0.074***	-0.216***
	(0.005)	(0.029)	(0.009)	(0.006)	(0.006)	(0.006)	(0.008)	(0.024)
nwocc3	-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)
Constant	1.579***	0.100	0.981***	1.427***	1.777***	2.004***	2.246***	2.869***
	(0.037)	(0.135)	(0.064)	(0.045)	(0.047)	(0.044)	(0.064)	(0.161)
Region effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occup. effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	97679	97679	97679	97679	97679	97679	97679	97679
r2	0.553							
pr2_1_6_1		0.329	0.312	0.308	0.342	0.368	0.379	0.348

Note: bootstrapped s.e. 200 replications, * p<0.10, ** p<0.05, *** p<0.01.
Source: Author's computations using PNAD 1987.

Table A2: Pooled quantile regressions, 2006

Panel A - Mean and quantile regressions for all sample (using focc3)

	(1) mean	(2) 1	(3) 10	(4) 25	(5) 50	(6) 75	(7) 90	(8) 99
male	0.211***	0.167***	0.187***	0.187***	0.195***	0.204***	0.208***	0.151***
	(0.005)	(0.012)	(0.006)	(0.005)	(0.005)	(0.007)	(0.009)	(0.029)
white	0.127***	0.061***	0.077***	0.095***	0.113***	0.149***	0.177***	0.200***
	(0.004)	(0.011)	(0.005)	(0.004)	(0.004)	(0.005)	(0.008)	(0.028)
age	0.064***	0.042***	0.054***	0.054***	0.061***	0.069***	0.073***	0.079***
	(0.001)	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.006)
agesq	-0.001***	-0.000***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
edu	0.100***	0.060***	0.070***	0.077***	0.090***	0.107***	0.121***	0.133***
	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
urban	0.146***	0.330***	0.233***	0.151***	0.104***	0.080***	0.084***	0.104***
	(0.006)	(0.031)	(0.013)	(0.007)	(0.006)	(0.007)	(0.009)	(0.036)
biformal	0.218***	1.170***	0.520***	0.315***	0.182***	0.080***	-0.019***	-0.271***
	(0.004)	(0.018)	(0.006)	(0.004)	(0.004)	(0.005)	(0.007)	(0.028)
focc3	-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)
Constant	-1.521***	-2.561***	-1.910***	-1.521***	-1.368***	-1.237***	-1.008***	-0.131
	(0.019)	(0.067)	(0.028)	(0.022)	(0.021)	(0.023)	(0.031)	(0.119)
Region effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	148960	148960	148960	148960	148960	148960	148960	148960
r2	0.419							
pr2		0.332	0.273	0.220	0.228	0.263	0.289	0.233

Note: bootstrapped s.e. 200 replications, * p<0.10, ** p<0.05, *** p<0.01.
Source: Author's computations using PNAD 2006.

Panel B - Mean and quantile regressions for all sample (using focc3 and occupations)

	(1) mean	(2) 1	(3) 10	(4) 25	(5) 50	(6) 75	(7) 90	(8) 99
male	0.208***	0.175***	0.196***	0.189***	0.189***	0.187***	0.195***	0.126***
	(0.005)	(0.014)	(0.006)	(0.005)	(0.004)	(0.006)	(0.008)	(0.028)
white	0.097***	0.059***	0.069***	0.078***	0.080***	0.103***	0.127***	0.147***
	(0.004)	(0.010)	(0.005)	(0.004)	(0.004)	(0.004)	(0.007)	(0.025)
age	0.059***	0.046***	0.053***	0.053***	0.054***	0.058***	0.064***	0.065***
	(0.001)	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.006)
agesq	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
edu	0.066***	0.044***	0.050***	0.052***	0.056***	0.065***	0.076***	0.088***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
urban	0.072***	0.095***	0.106***	0.085***	0.060***	0.042***	0.033***	0.020
	(0.006)	(0.031)	(0.009)	(0.006)	(0.005)	(0.007)	(0.009)	(0.038)
biformal	0.203***	1.096***	0.501***	0.309***	0.175***	0.070***	-0.042***	-0.287***
	(0.004)	(0.022)	(0.007)	(0.004)	(0.004)	(0.004)	(0.006)	(0.026)
focc3	-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)
Constant	-0.175***	-1.716***	-0.819***	-0.331***	0.141***	0.382***	0.488***	1.152***
	(0.028)	(0.091)	(0.042)	(0.033)	(0.031)	(0.038)	(0.047)	(0.128)
Region effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occup. effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	148960	148960	148960	148960	148960	148960	148960	148960
r2	0.475							
pr2		0.353	0.301	0.257	0.275	0.312	0.335	0.276

Note: bootstrapped s.e. 200 replications, * p<0.10, ** p<0.05, *** p<0.01.
Source: Author's computations using PNAD 2006.

Panel C - Mean and quantile regressions for all sample (using nwocc3)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	mean	1	10	25	50	75	90	99
male	0.255***	0.182***	0.231***	0.238***	0.248***	0.257***	0.252***	0.188***
	(0.004)	(0.012)	(0.005)	(0.004)	(0.004)	(0.004)	(0.006)	(0.027)
white	0.088***	0.050***	0.056***	0.070***	0.077***	0.096***	0.117***	0.128***
	(0.004)	(0.010)	(0.006)	(0.004)	(0.004)	(0.005)	(0.007)	(0.026)
age	0.065***	0.046***	0.058***	0.059***	0.063***	0.066***	0.068***	0.067***
	(0.001)	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.005)
agesq	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
edu	0.075***	0.054***	0.059***	0.061***	0.067***	0.076***	0.084***	0.096***
	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)
urban	0.132***	0.335***	0.226***	0.137***	0.090***	0.065***	0.055***	0.031
	(0.006)	(0.033)	(0.011)	(0.008)	(0.005)	(0.007)	(0.009)	(0.033)
biformal	0.188***	1.150***	0.500***	0.292***	0.156***	0.044***	-0.060***	-0.323***
	(0.003)	(0.021)	(0.006)	(0.004)	(0.004)	(0.005)	(0.006)	(0.022)
nwocc3	-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)
Constant	-0.391***	-2.227***	-1.278***	-0.774***	-0.329***	0.136***	0.605***	1.678***
	(0.022)	(0.073)	(0.030)	(0.024)	(0.022)	(0.029)	(0.045)	(0.124)
Region effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	148960	148960	148960	148960	148960	148960	148960	148960
r2	0.451							
pr2		0.335	0.284	0.236	0.251	0.290	0.318	0.256

Note: bootstrapped s.e. 200 replications, * p<0.10, ** p<0.05, *** p<0.01.
Source: Author's computations using PNAD 2006.

Panel D - Mean and quantile regressions for all sample (using nwocc3 and occupations)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	mean	1	10	25	50	75	90	99
male	0.222***	0.177***	0.208***	0.199***	0.204***	0.209***	0.228***	0.195***
	(0.004)	(0.013)	(0.006)	(0.004)	(0.004)	(0.005)	(0.008)	(0.027)
white	0.092***	0.059***	0.064***	0.074***	0.077***	0.097***	0.121***	0.153***
	(0.004)	(0.011)	(0.005)	(0.004)	(0.003)	(0.005)	(0.007)	(0.025)
age	0.058***	0.047***	0.053***	0.053***	0.053***	0.057***	0.063***	0.066***
	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.005)
agesq	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
edu	0.065***	0.044***	0.050***	0.051***	0.055***	0.063***	0.074***	0.089***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)
urban	0.077***	0.092***	0.109***	0.090***	0.061***	0.047***	0.038***	0.021
	(0.006)	(0.033)	(0.008)	(0.007)	(0.005)	(0.008)	(0.009)	(0.040)
biformal	0.207***	1.095***	0.502***	0.312***	0.179***	0.076***	-0.033***	-0.264***
	(0.004)	(0.020)	(0.006)	(0.004)	(0.003)	(0.004)	(0.007)	(0.030)
nwocc3	-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)
Constant	0.011	-1.644***	-0.744***	-0.180***	0.338***	0.627***	0.743***	1.163***
	(0.030)	(0.094)	(0.050)	(0.033)	(0.033)	(0.042)	(0.048)	(0.133)
Region effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occup. effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	148960	148960	148960	148960	148960	148960	148960	148960
r2	0.477							
pr2		0.353	0.302	0.258	0.276	0.314	0.337	0.275

Note: bootstrapped s.e. 200 replications, * p<0.10, ** p<0.05, *** p<0.01.
Source: Author's computations using PNAD 2006.

Table A3: Inter-quantile regressions for the pooled sample, 1987 and 2006

Panel A - Inter-quantile regressions for all sample (using focc3)

	(1) 1987 10-50	(2) 1987 50-90	(3) 1987 10-90	(1) 2006 10-50	(2) 2006 50-90	(3) 2006 10-90
male	-0.016	0.005	-0.011	0.008	0.013*	0.021**
	(0.010)	(0.010)	(0.014)	(0.007)	(0.008)	(0.010)
white	0.019**	0.038***	0.057***	0.035***	0.065***	0.100***
	(0.008)	(0.009)	(0.011)	(0.005)	(0.008)	(0.008)
age	0.013***	0.009***	0.023***	0.007***	0.012***	0.020***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
agesq	-0.000***	-0.000	-0.000***	-0.000	-0.000**	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
edu	0.028***	0.011***	0.039***	0.020***	0.031***	0.051***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
urban	-0.075***	0.001	-0.075***	-0.128***	-0.020**	-0.149***
	(0.012)	(0.011)	(0.015)	(0.011)	(0.010)	(0.014)
biformal	-0.234***	-0.172***	-0.406***	-0.338***	-0.201***	-0.539***
	(0.009)	(0.008)	(0.011)	(0.006)	(0.007)	(0.009)
focc3	-0.042***	-0.019	-0.061***	0.002	-0.039***	-0.037**
	(0.016)	(0.016)	(0.020)	(0.011)	(0.013)	(0.017)
Constant	0.615***	0.707***	1.321***	0.543***	0.359***	0.902***
	(0.033)	(0.037)	(0.044)	(0.030)	(0.035)	(0.038)
Region effects	Yes	Yes	Yes	Yes	Yes	Yes
N	97679	97679	97679	148960	148960	148960

Note: bootstrapped s.e. 200 replications, * p<0.10, ** p<0.05, *** p<0.01.
Source: Author's computations using PNAD 1987 and 2006.

Panel B - Inter-quantile regressions for all sample (using focc3 and occupations)

	(1) 1987 10-50	(2) 1987 50-90	(3) 1987 10-90	(1) 2006 10-50	(2) 2006 50-90	(3) 2006 10-90
male	-0.025**	0.025**	-0.000	-0.007	0.006	-0.001
	(0.011)	(0.011)	(0.014)	(0.006)	(0.008)	(0.010)
white	-0.001	0.026***	0.025**	0.011**	0.047***	0.058***
	(0.008)	(0.007)	(0.011)	(0.005)	(0.007)	(0.008)
age	0.010***	0.008***	0.018***	0.001	0.010***	0.011***
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
agesq	-0.000***	-0.000	-0.000***	0.000**	-0.000*	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
edu	0.015***	0.013***	0.027***	0.006***	0.020***	0.026***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
urban	-0.086***	-0.033***	-0.119***	-0.046***	-0.027***	-0.073***
	(0.014)	(0.012)	(0.019)	(0.009)	(0.010)	(0.013)
biformal	-0.231***	-0.182***	-0.413***	-0.326***	-0.217***	-0.543***
	(0.010)	(0.009)	(0.011)	(0.007)	(0.006)	(0.010)
focc3	0.024	0.023	0.046	-0.004	-0.071***	-0.076***
	(0.021)	(0.023)	(0.028)	(0.014)	(0.016)	(0.020)
Constant	0.674***	0.479***	1.153***	0.960***	0.347***	1.307***
	(0.067)	(0.062)	(0.081)	(0.052)	(0.047)	(0.067)
Region effects	Yes	Yes	Yes	Yes	Yes	Yes
Occup. effects	Yes	Yes	Yes	Yes	Yes	Yes
N	97679	97679	97679	148960	148960	148960

Note: bootstrapped s.e. 200 replications, * p<0.10, ** p<0.05, *** p<0.01.
Source: Author's computations using PNAD 1987 and 2006.

Panel C - Inter-quantile regressions for all sample (using nwocc3)

	(1) 1987 10-50	(2) 1987 50-90	(3) 1987 10-90	(1) 2006 10-50	(2) 2006 50-90	(3) 2006 10-90
male	0.001	0.014*	0.014	0.016***	0.005	0.021***
	(0.008)	(0.008)	(0.011)	(0.005)	(0.006)	(0.008)
white	0.006	0.031***	0.036***	0.021***	0.040***	0.062***
	(0.008)	(0.008)	(0.012)	(0.005)	(0.008)	(0.009)
age	0.011***	0.007***	0.018***	0.005***	0.005***	0.011***
	(0.002)	(0.002)	(0.003)	(0.001)	(0.002)	(0.002)
agesq	-0.000***	-0.000	-0.000***	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
edu	0.016***	0.010***	0.026***	0.008***	0.017***	0.025***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
urban	-0.077***	-0.012	-0.088***	-0.136***	-0.035***	-0.171***
	(0.011)	(0.011)	(0.014)	(0.010)	(0.009)	(0.015)
biformal	-0.246***	-0.179***	-0.424***	-0.345***	-0.215***	-0.560***
	(0.009)	(0.009)	(0.013)	(0.006)	(0.007)	(0.009)
nwocc3	-0.343***	-0.166***	-0.509***	-0.462***	-0.655***	-1.117***
	(0.042)	(0.043)	(0.058)	(0.028)	(0.037)	(0.042)
Constant	0.858***	0.810***	1.668***	0.948***	0.935***	1.883***
	(0.044)	(0.048)	(0.060)	(0.034)	(0.041)	(0.052)
Region effects	Yes	Yes	Yes	Yes	Yes	Yes
N	97679	97679	97679	148960	148960	148960

Note: bootstrapped s.e. 200 replications, * p<0.10, ** p<0.05, *** p<0.01.
Source: Author's computations using PNAD 1987 and 2006.

Panel D - Inter-quantile regressions for all sample (using nwocc3 and occupations)

	(1) 1987 10-50	(2) 1987 50-90	(3) 1987 10-90	(1) 2006 10-50	(2) 2006 50-90	(3) 2006 10-90
male	-0.036***	0.018*	-0.018	-0.005	0.024***	0.020**
	(0.009)	(0.010)	(0.013)	(0.006)	(0.007)	(0.010)
white	0.001	0.024***	0.025**	0.012**	0.045***	0.057***
	(0.008)	(0.008)	(0.011)	(0.006)	(0.007)	(0.009)
age	0.010***	0.009***	0.019***	-0.000	0.009***	0.009***
	(0.002)	(0.002)	(0.003)	(0.001)	(0.002)	(0.002)
agesq	-0.000***	-0.000*	-0.000***	0.000***	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
edu	0.013***	0.013***	0.026***	0.005***	0.020***	0.024***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
urban	-0.089***	-0.029**	-0.118***	-0.048***	-0.023**	-0.072***
	(0.016)	(0.013)	(0.019)	(0.009)	(0.010)	(0.013)
biformal	-0.238***	-0.182***	-0.420***	-0.323***	-0.212***	-0.535***
	(0.009)	(0.010)	(0.012)	(0.006)	(0.007)	(0.009)
nwocc3	-0.441***	0.078	-0.364***	-0.360***	-0.319***	-0.679***
	(0.091)	(0.084)	(0.116)	(0.058)	(0.069)	(0.082)
Constant	0.796***	0.469***	1.265***	1.082***	0.405***	1.487***
	(0.063)	(0.065)	(0.083)	(0.047)	(0.054)	(0.062)
Region effects	Yes	Yes	Yes	Yes	Yes	Yes
Occup. effects	Yes	Yes	Yes	Yes	Yes	Yes
N	97679	97679	97679	148960	148960	148960

Note: bootstrapped s.e. 200 replications, * p<0.10, ** p<0.05, *** p<0.01.
Source: Author's computations using PNAD 1987 and 2006.

Table A4: Quantile decomposition results for gender wage gaps (using the 1st 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	-0.068	-0.147	-0.306	-0.196	-0.156	-0.223
s.e.	0.003	0.001	0.004	0.002	0.001	0.003
Unexplained	0.700	0.474	0.467	0.324	0.230	0.270
s.e.	0.011	0.006	0.015	0.007	0.004	0.012
Total gap (conditional wages)	0.362	0.314	0.309	0.094	0.057	0.043
s.e.	0.005	0.005	0.005	0.003	0.003	0.003
Residual	0.270	0.014	-0.147	0.034	0.016	0.004
Total gap (predicted wages)	0.632	0.327	0.162	0.128	0.074	0.048
s.e.	0.012	0.006	0.015	0.007	0.003	0.012
Decomposition method: Melly (2006)						
Explained	-0.120	-0.158	-0.199	-0.173	-0.160	-0.181
s.e.	0.006	0.005	0.009	0.004	0.003	0.003
Unexplained	0.596	0.446	0.419	0.267	0.204	0.233
s.e.	0.005	0.004	0.007	0.006	0.004	0.008
Total gap	0.476	0.288	0.221	0.094	0.045	0.052
s.e.	0.008	0.006	0.009	0.008	0.004	0.009
Decomposition method: Machado & Mata (2005) as Albrecht, van Vuuren and Vroman (2009)						
Explained	-0.118	-0.155	-0.195	-0.172	-0.159	-0.181
s.e.	0.030	0.024	0.038	0.027	0.019	0.033
Unexplained	0.594	0.443	0.415	0.263	0.204	0.233
s.e.	0.033	0.025	0.036	0.030	0.018	0.036
Total gap	0.476	0.288	0.220	0.091	0.044	0.052
s.e.	0.033	0.024	0.037	0.029	0.019	0.033
Decomposition method: RIF-OLS regressions (Firpo, Fortin and Lemieux, 2009)						
Explained	-0.069	-0.150	-0.308	-0.132	-0.147	-0.337
s.e.	0.004	0.005	0.009	0.004	0.003	0.006
Unexplained	0.653	0.455	0.511	0.294	0.195	0.339
s.e.	0.011	0.007	0.014	0.008	0.004	0.012
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	0.007	0.021	0.045	-0.008	-0.005	0.005
s.e.	0.002	0.002	0.003	0.001	0.001	0.003
Expl: edu	-0.041	-0.130	-0.350	-0.068	-0.114	-0.340
s.e.	0.002	0.003	0.009	0.002	0.002	0.006
Unexp: age	-0.680	0.539	0.326	-0.515	0.546	0.265
s.e.	0.100	0.053	0.080	0.094	0.036	0.080
Unexp: edu	-0.221	-0.151	0.227	-0.300	-0.090	0.083
s.e.	0.014	0.010	0.028	0.019	0.009	0.030

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

Table A5: Quantile decomposition results for gender wage gaps (using the 2nd 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	-0.024	-0.004	-0.147	-0.205	-0.111	-0.283
s.e.	0.009	0.006	0.018	0.005	0.002	0.008
Unexplained	0.620	0.340	0.312	0.306	0.194	0.305
s.e.	0.015	0.008	0.022	0.008	0.004	0.013
Total gap (conditional wages)	0.362	0.311	0.311	0.090	0.057	0.046
s.e.	0.005	0.005	0.005	0.003	0.003	0.003
Residual	0.233	0.025	-0.147	0.011	0.026	-0.024
Total gap (predicted wages)	0.595	0.336	0.165	0.101	0.083	0.022
s.e.	0.017	0.009	0.028	0.009	0.004	0.014
Decomposition method: Melly (2006)						
Explained	-0.039	-0.040	-0.079	-0.134	-0.112	-0.222
s.e.	0.006	0.008	0.012	0.004	0.004	0.009
Unexplained	0.513	0.339	0.288	0.211	0.174	0.252
s.e.	0.005	0.004	0.007	0.007	0.004	0.009
Total gap	0.474	0.299	0.209	0.077	0.062	0.030
s.e.	0.008	0.007	0.010	0.007	0.004	0.008
Decomposition method: Machado & Mata (2005) as Albrecht, van Vuuren and Vroman (2009)						
Explained	-0.037	-0.038	-0.082	-0.133	-0.113	-0.221
s.e.	0.027	0.023	0.036	0.025	0.018	0.043
Unexplained	0.510	0.337	0.281	0.206	0.173	0.248
s.e.	0.028	0.026	0.041	0.029	0.019	0.040
Total gap	0.472	0.299	0.199	0.073	0.060	0.027
s.e.	0.032	0.024	0.042	0.029	0.018	0.036
Decomposition method: RIF-OLS regressions (Firpo, Fortin and Lemieux, 2009)						
Explained	-0.066	-0.025	-0.085	-0.130	-0.107	-0.289
s.e.	0.007	0.009	0.019	0.006	0.004	0.011
Unexplained	0.649	0.331	0.288	0.292	0.155	0.291
s.e.	0.012	0.010	0.021	0.009	0.005	0.014
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	0.034	-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	-0.084	-0.205
s.e.	0.002	0.003	0.007	0.002	0.002	0.005
Expl: occ	-0.035	0.074	0.137	-0.039	-0.001	-0.074
s.e.	0.007	0.008	0.017	0.005	0.003	0.010
Unexp: age	-0.725	0.418	0.199	-0.685	0.513	0.343
s.e.	0.100	0.052	0.079	0.096	0.035	0.078
Unexp: edu	-0.119	0.032	0.169	-0.284	0.037	0.042
s.e.	0.022	0.015	0.033	0.027	0.012	0.030
Unexp:occ	0.047	0.212	0.123	0.028	0.170	-0.510
s.e.	0.062	0.039	0.317	0.042	0.022	0.223

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

Table A6: Quantile decomposition results for gender wage gaps (using the 3rd 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	0.034	0.002	-0.221	-0.166	-0.130	-0.232
s.e.	0.011	0.007	0.009	0.005	0.003	0.006
Unexplained	0.591	0.330	0.375	0.293	0.203	0.285
s.e.	0.016	0.009	0.016	0.009	0.005	0.013
Total gap (conditional wages)	0.360	0.309	0.309	0.094	0.056	0.040
s.e.	0.005	0.005	0.005	0.003	0.003	0.003
Residual	0.265	0.023	-0.155	0.033	0.016	0.013
Total gap (predicted wages)	0.625	0.332	0.154	0.127	0.073	0.053
s.e.	0.019	0.011	0.018	0.009	0.006	0.014
Decomposition method: Melly (2006)						
Explained	-0.002	-0.029	-0.106	-0.145	-0.139	-0.187
s.e.	0.009	0.007	0.010	0.005	0.004	0.005
Unexplained	0.477	0.322	0.321	0.239	0.184	0.236
s.e.	0.005	0.004	0.007	0.004	0.003	0.008
Total gap	0.475	0.293	0.214	0.095	0.045	0.049
s.e.	0.009	0.005	0.008	0.006	0.003	0.009
Decomposition method: Machado & Mata (2005) as Albrecht, van Vuuren and Vroman (2009)						
Explained	-0.001	-0.030	-0.105	-0.144	-0.138	-0.185
s.e.	0.028	0.025	0.040	0.028	0.019	0.033
Unexplained	0.481	0.324	0.314	0.235	0.183	0.234
s.e.	0.035	0.024	0.036	0.031	0.018	0.035
Total gap	0.596	0.303	0.176	0.092	0.045	0.048
s.e.	0.033	0.024	0.041	0.029	0.019	0.033
Decomposition method: RIF-OLS regressions (Firpo, Fortin and Lemieux, 2009)						
Explained	0.026	-0.002	-0.208	-0.103	-0.083	-0.462
s.e.	0.010	0.010	0.018	0.008	0.005	0.012
Unexplained	0.558	0.307	0.411	0.264	0.131	0.464
s.e.	0.014	0.011	0.021	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	0.007	0.021	0.045	-0.008	-0.004	0.004
s.e.	0.001	0.002	0.003	0.001	0.001	0.003
Expl: edu	-0.044	-0.135	-0.354	-0.069	-0.117	-0.334
s.e.	0.002	0.003	0.009	0.002	0.002	0.006
Expl: focc3	0.098	0.153	0.104	0.031	0.067	-0.129
s.e.	0.009	0.009	0.016	0.007	0.004	0.010
Unexp: age	-0.702	0.492	0.305	-0.542	0.505	0.292
s.e.	0.099	0.053	0.080	0.094	0.036	0.081
Unexp: edu	-0.201	-0.124	0.249	-0.288	-0.070	0.064
s.e.	0.014	0.010	0.028	0.020	0.009	0.030
Unexp: focc3	0.205	0.012	0.264	0.029	-0.067	0.595
s.e.	0.026	0.019	0.034	0.025	0.012	0.030

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

Table A7: Quantile decomposition results for racial wage gaps (using the 1st 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	0.337	0.346	0.547	0.342	0.275	0.443
s.e.	0.004	0.004	0.007	0.003	0.002	0.005
Unexplained	0.037	0.090	0.192	0.051	0.073	0.265
s.e.	0.012	0.007	0.011	0.009	0.004	0.008
Total gap (conditional wages)	0.465	0.493	0.507	0.379	0.398	0.476
s.e.	0.004	0.004	0.005	0.003	0.003	0.003
Residual	-0.091	-0.056	0.232	0.014	-0.050	0.232
Total gap (predicted wages)	0.374	0.436	0.739	0.393	0.348	0.708
s.e.	0.013	0.007	0.013	0.009	0.004	0.009
Decomposition method: Melly (2006)						
Explained	0.362	0.363	0.429	0.320	0.291	0.318
s.e.	0.008	0.006	0.008	0.005	0.004	0.005
Unexplained	0.003	0.111	0.209	-0.005	0.084	0.258
s.e.	0.006	0.005	0.008	0.008	0.003	0.006
Total gap	0.365	0.473	0.638	0.316	0.375	0.576
s.e.	0.007	0.005	0.008	0.008	0.004	0.008
Decomposition method: Machado & Mata (2005) as Albrecht, van Vuuren and Vroman (2009)						
Explained	0.365	0.365	0.431	0.319	0.287	0.315
s.e.	0.034	0.024	0.039	0.029	0.021	0.032
Unexplained	0.002	0.106	0.206	-0.006	0.085	0.258
s.e.	0.032	0.024	0.033	0.028	0.019	0.030
Total gap	0.367	0.472	0.637	0.314	0.372	0.573
s.e.	0.032	0.024	0.038	0.031	0.019	0.033
Decomposition method: RIF-OLS regressions (Firpo, Fortin and Lemieux, 2009)						
Explained	0.361	0.359	0.501	0.371	0.276	0.411
s.e.	0.008	0.006	0.011	0.007	0.004	0.008
Unexplained	0.041	0.103	0.225	0.029	0.075	0.285
s.e.	0.014	0.008	0.012	0.011	0.005	0.010
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	0.021	0.022	0.010	0.019	0.043
s.e.	0.002	0.002	0.003	0.001	0.001	0.003
Expl: edu	0.134	0.275	0.531	0.119	0.182	0.410
s.e.	0.003	0.004	0.009	0.003	0.003	0.007
Unexp: age	0.199	0.360	0.154	-0.624	0.552	0.187
s.e.	0.092	0.048	0.077	0.085	0.034	0.075
Unexp: edu	0.033	0.129	0.098	0.061	0.309	0.333
s.e.	0.010	0.007	0.018	0.014	0.007	0.022

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

Table A8: Quantile decomposition results for racial wage gaps (using the 2nd 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	0.323	0.359	0.599	0.333	0.275	0.566
s.e.	0.004	0.003	0.007	0.003	0.002	0.005
Unexplained	0.068	0.078	0.139	0.063	0.058	0.171
s.e.	0.013	0.006	0.011	0.009	0.004	0.008
Total gap (conditional wages)	0.490	0.487	0.505	0.400	0.392	0.462
s.e.	0.004	0.005	0.005	0.003	0.003	0.003
Residual	-0.099	-0.050	0.232	-0.003	-0.059	0.275
Total gap (predicted wages)	0.392	0.436	0.737	0.396	0.333	0.737
s.e.	0.013	0.007	0.013	0.009	0.004	0.009
Decomposition method: Melly (2006)						
Explained	0.346	0.369	0.494	0.303	0.289	0.418
s.e.	0.008	0.005	0.007	0.005	0.003	0.006
Unexplained	0.026	0.099	0.159	0.027	0.067	0.181
s.e.	0.005	0.005	0.007	0.008	0.004	0.007
Total gap	0.372	0.467	0.652	0.330	0.356	0.600
s.e.	0.007	0.005	0.007	0.007	0.004	0.009
Decomposition method: Machado & Mata (2005) as Albrecht, van Vuuren and Vroman (2009)						
Explained	0.349	0.371	0.495	0.302	0.285	0.416
s.e.	0.033	0.022	0.041	0.029	0.019	0.033
Unexplained	0.022	0.094	0.156	0.027	0.068	0.182
s.e.	0.030	0.021	0.033	0.028	0.017	0.031
Total gap	0.371	0.465	0.651	0.329	0.354	0.598
s.e.	0.032	0.022	0.039	0.029	0.018	0.038
Decomposition method: RIF-OLS regressions (Firpo, Fortin and Lemieux, 2009)						
Explained	0.383	0.383	0.491	0.383	0.293	0.439
s.e.	0.008	0.006	0.011	0.007	0.004	0.008
Unexplained	0.019	0.079	0.235	0.017	0.058	0.257
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	0.035
s.e.	0.002	0.002	0.003	0.001	0.001	0.002
Expl: edu	0.112	0.186	0.377	0.098	0.120	0.258
s.e.	0.005	0.004	0.009	0.004	0.002	0.006
Expl: occ	0.053	0.125	0.138	0.053	0.080	0.161
s.e.	0.004	0.004	0.006	0.003	0.002	0.005
Unexp: age	0.255	0.358	0.001	-0.539	0.510	0.039
s.e.	0.092	0.048	0.075	0.086	0.034	0.073
Unexp: edu	0.053	0.094	0.081	0.061	0.188	0.221
s.e.	0.014	0.010	0.021	0.019	0.009	0.023
Unexp:occ	-0.066	-0.161	-0.821	-0.140	-0.201	0.173
s.e.	0.052	0.043	0.266	0.041	0.024	0.188

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

Table A9: Quantile decomposition results for racial wage gaps (using the 3rd 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	0.336	0.370	0.598	0.345	0.305	0.520
s.e.	0.004	0.003	0.007	0.003	0.003	0.005
Unexplained	0.040	0.063	0.151	0.037	0.038	0.186
s.e.	0.012	0.006	0.011	0.009	0.005	0.009
Total gap (conditional wages)	0.486	0.487	0.503	0.395	0.397	0.456
s.e.	0.004	0.004	0.005	0.003	0.003	0.003
Residual	-0.111	-0.053	0.246	-0.013	-0.054	0.250
Total gap (predicted wages)	0.375	0.434	0.749	0.382	0.343	0.707
s.e.	0.012	0.007	0.012	0.009	0.005	0.010
Decomposition method: Melly (2006)						
Explained	0.369	0.386	0.474	0.330	0.320	0.383
s.e.	0.009	0.005	0.007	0.007	0.004	0.006
Unexplained	-0.001	0.086	0.165	-0.014	0.052	0.196
s.e.	0.005	0.005	0.007	0.004	0.003	0.006
Total gap	0.368	0.472	0.639	0.317	0.372	0.579
s.e.	0.008	0.004	0.007	0.005	0.002	0.008
Decomposition method: Machado & Mata (2005) as Albrecht, van Vuuren and Vroman (2009)						
Explained	0.371	0.387	0.474	0.328	0.316	0.381
s.e.	0.035	0.022	0.040	0.030	0.021	0.032
Unexplained	-0.003	0.081	0.163	-0.015	0.054	0.195
s.e.	0.032	0.023	0.033	0.028	0.018	0.030
Total gap	0.368	0.468	0.637	0.313	0.369	0.575
s.e.	0.032	0.023	0.038	0.030	0.019	0.033
Decomposition method: RIF-OLS regressions (Firpo, Fortin and Lemieux, 2009)						
Explained	0.371	0.384	0.544	0.378	0.304	0.489
s.e.	0.008	0.006	0.011	0.007	0.004	0.009
Unexplained	0.031	0.078	0.182	0.021	0.047	0.208
s.e.	0.014	0.007	0.012	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	0.020	0.021	0.009	0.018	0.039
s.e.	0.002	0.002	0.003	0.001	0.001	0.003
Expl: edu	0.106	0.202	0.409	0.106	0.134	0.277
s.e.	0.004	0.004	0.008	0.003	0.002	0.005
Expl: nwocc3	0.041	0.107	0.179	0.022	0.082	0.225
s.e.	0.004	0.003	0.006	0.003	0.002	0.005
Unexp: age	0.188	0.347	0.140	-0.630	0.549	0.200
s.e.	0.092	0.048	0.076	0.085	0.033	0.074
Unexp: edu	0.012	0.095	0.052	0.008	0.210	0.155
s.e.	0.013	0.009	0.020	0.018	0.008	0.022
Unexp: nwocc3	-0.121	-0.149	-0.168	-0.257	-0.318	-0.263
s.e.	0.051	0.038	0.076	0.044	0.022	0.066

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

Appendix B: Performing the aggregate Melly (2006) decomposition by disaggregating formal and non-formal markets

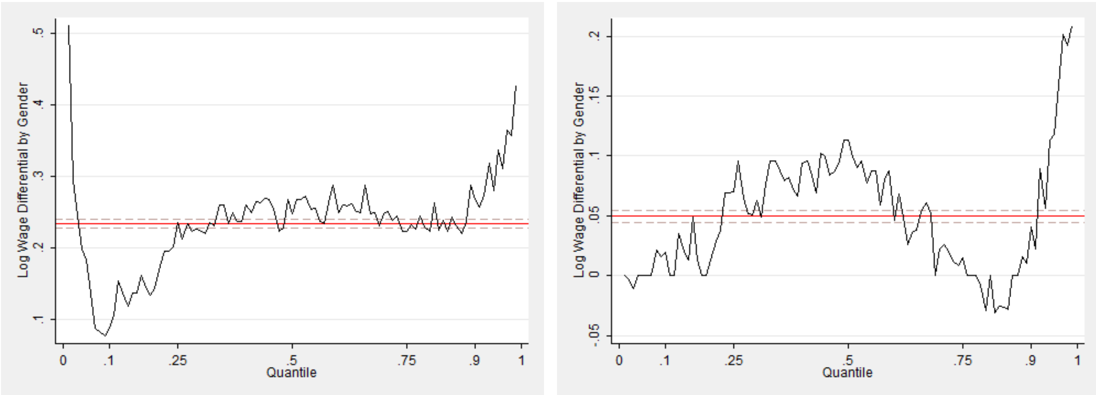
In chapter 3, we presented an extensive discussion of the evolution of wage gaps over time by gender and race, disaggregated into the formal and non-formal sectors. In this appendix we extend the analysis of wage gaps along the entire wage distribution by looking separately at the formal and non-formal sectors.

Figures B1 and B2 plot the gender and racial wage gaps across wage quantiles, disaggregated into the formal and non-formal sectors. We see that gender wage differentials are only more pronounced at the bottom of the wage distribution in the non-formal sectors. By contrast, within the formal sector the gender pay gap seems to increase as we move towards the top of the wage distribution, with particularly large wage gaps at the very top of the distribution. Interestingly, and particularly for the informal sector, we record negative wage gaps in the upper half of the wage distribution, before observing large wage gaps at the very top of the distribution (see Figure B1). Thus, the U-shape that we notice when looking at gender wage gaps over quantiles for the entire labour market disguises different patterns in the formal and non formal sectors: greater gender gaps within low-paid occupations occur primarily in the non-formal sector, while greater gender gaps within top occupations are a more prominent feature of formal sector activities. Turning to racial wage gaps, we do not see large differences in patterns across sectors, as racial wage gaps tend to increase as we move toward the top of the wage distribution in both the formal and non-formal sectors (see Figure B2).

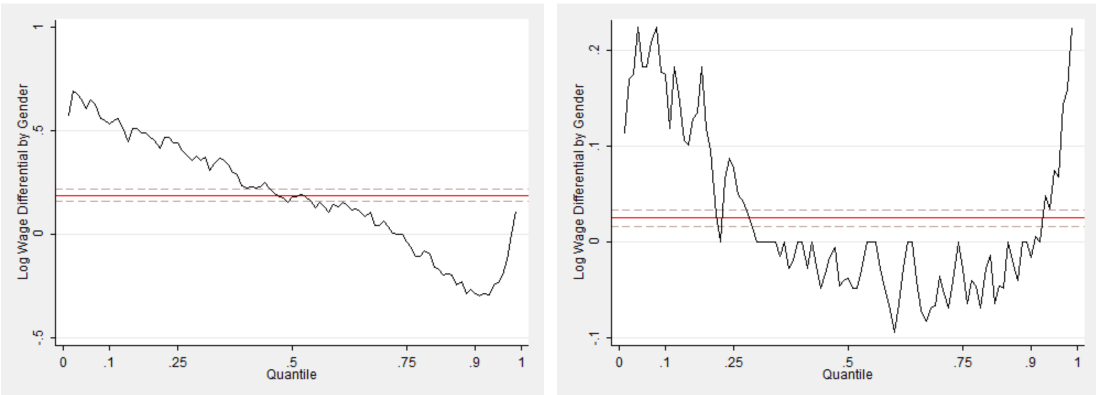
In sum, our analysis at the aggregate level revealed that women suffer from more severe pay gaps at the extremes of the wage distribution. Disaggregating athe analysis into the formal and non-formal labour markets further reveals higher wage gaps within low paid non-formal occupations and within the very top paid formal jobs. On the other hand, non-white workers seem to suffer from higher wage gaps among higher wage quantiles within all segments of the labour market, independent to the degree of informality.

Figure B1: Wage differentials over wage quantiles by gender and disaggregated by formal and non-formal sectors

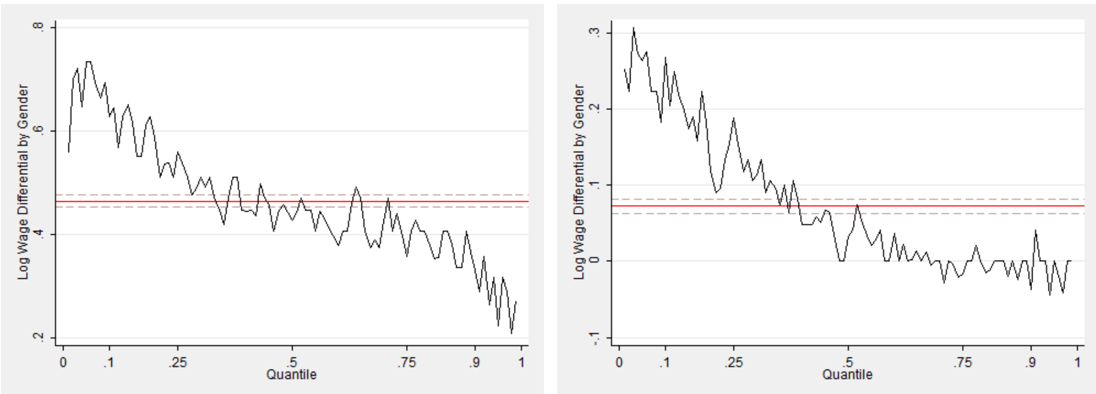
Panel A – Formal sector – 1987 and 2006



Panel B - Informal sector – 1987 and 2006



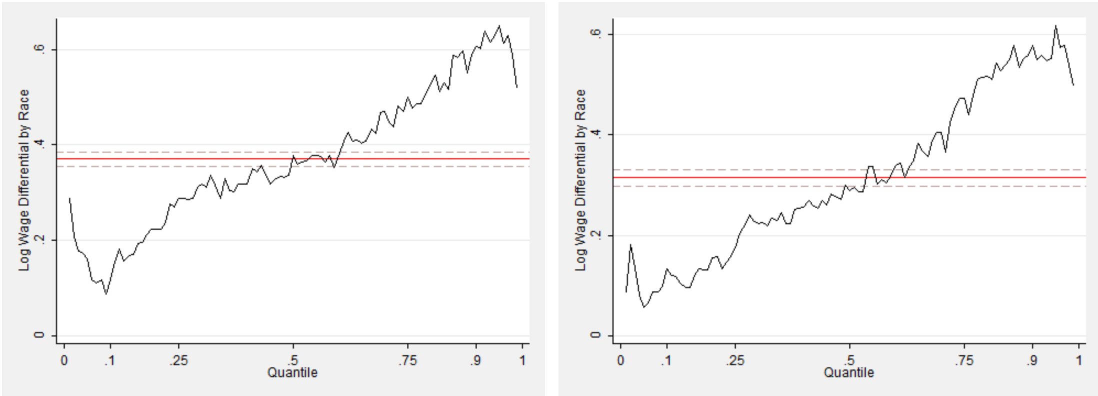
Panel C – Self-employed sector – 1987 and 2006



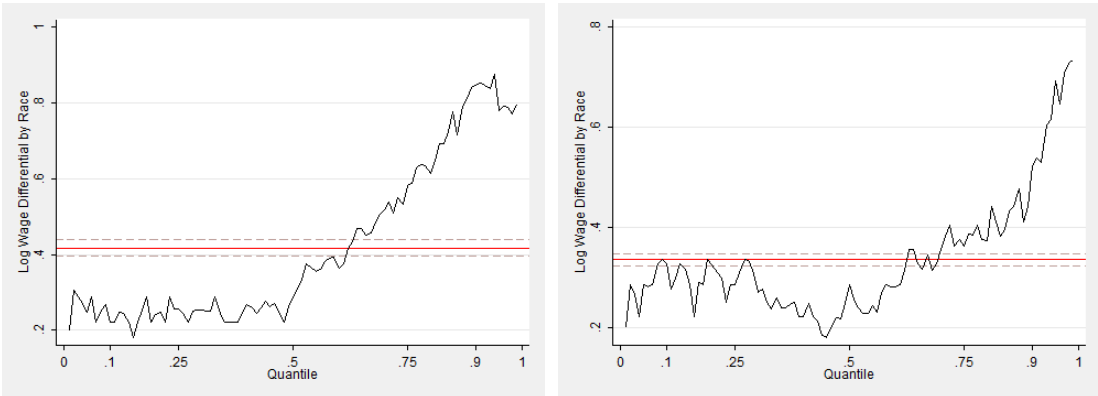
Source: Author's computations using PNAD 1987 and 2006.
Note: the horizontal lines represent the mean values for wage gaps. The wage differentials are the difference of the value of wages for each percentile computed separately for each sub-group.

Figure B2: Wage differentials over wage quantiles by race and disaggregated by formal and non-formal sectors

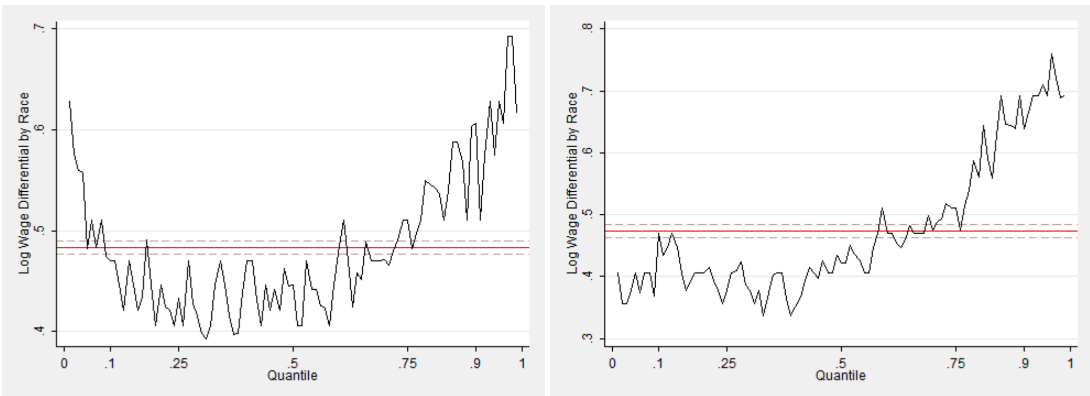
Panel A – Formal sector – 1987 and 2006



Panel B - Informal sector – 1987 and 2006



Panel C – Self-employed sector – 1987 and 2006



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

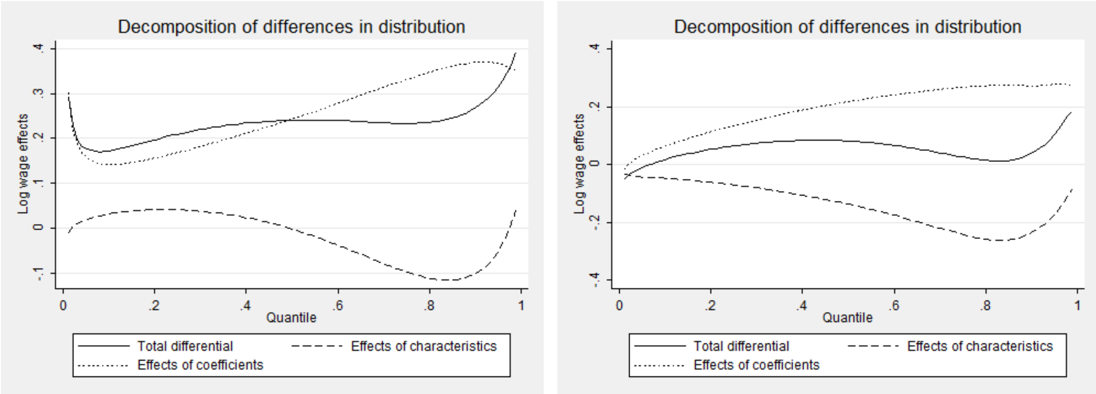
In order to further investigate these pay differentials, we compute the aggregate Melly (2005) decomposition across the formal and non-formal labour markets for the first and the last years of the data. We begin with gender wage gaps, with the results reported in figure B3. We find that the wage structure component (or coefficients' effect) acts differently between the formal and non-formal labour markets. While it is higher at higher quantiles in the formal market, in the non-formal sectors the effect of the coefficients (or wage structure effect) is considerably greater at the bottom end of the wage distribution.

In the case of racial wage gaps, wage differentials widen at the top end of the wage distribution both because of greater characteristics for whites and higher returns to these characteristics. The disaggregation of the analysis reveals quite similar patterns within the formal, informal and self-employed sectors, as reported in figure B4. This suggests less acute differences in labour conditions across the three sectors for non-white and white workers.

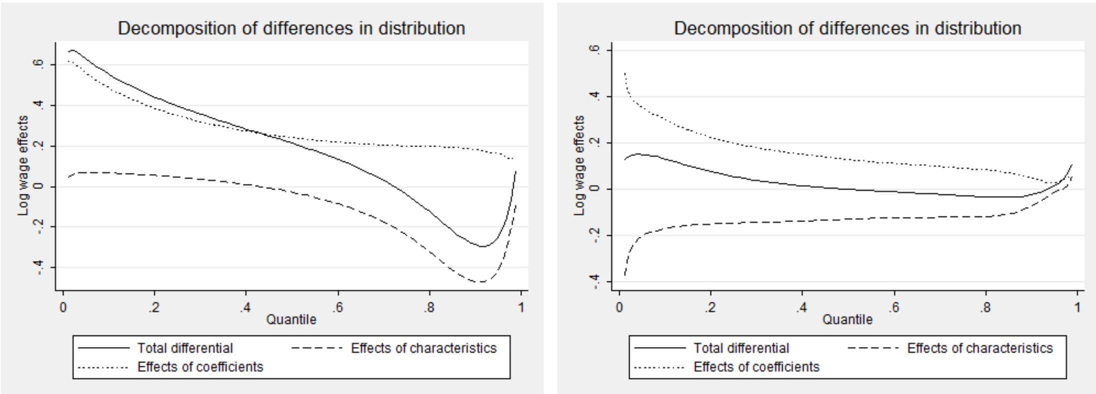
In summary, there is a *sticky floor* phenomenon for women, but it is largely confined to the non-formal sectors, while there continues to be a *glass ceilings* phenomenon within the highest echelons of the formal sector. On the other hand, non-white workers suffer from the existence of glass ceilings and this racial discrimination appears to be a fairly persistent feature of all segment of the Brazilian labour market.

Figure B3: Melly (2006) quantile decomposition results of gender wage gaps, disaggregated by formal and non-formal sectors

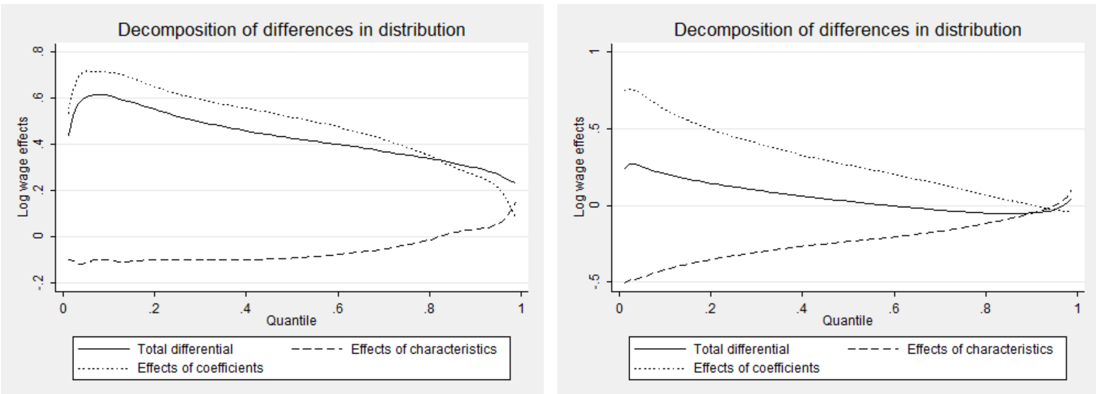
Panel A – Formal sector – 1987 and 2006



Panel B – Informal sector – 1987 and 2006



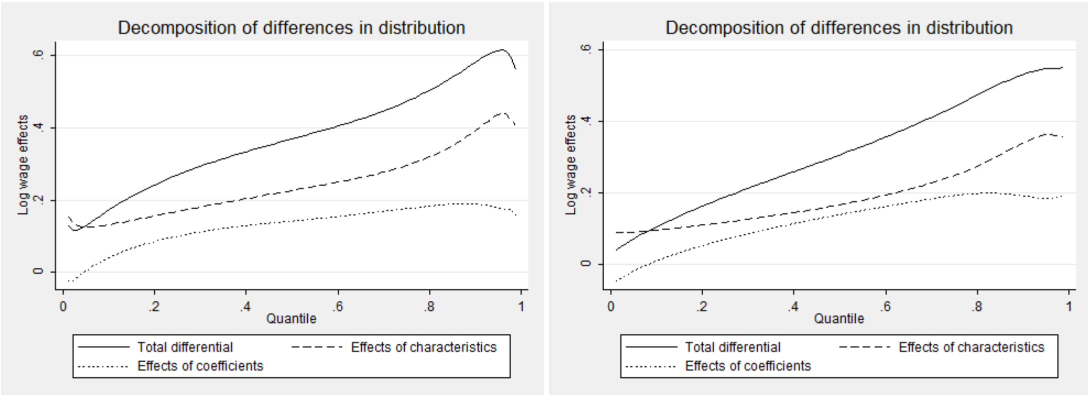
Panel C – Self-employed sector – 1987 and 2006



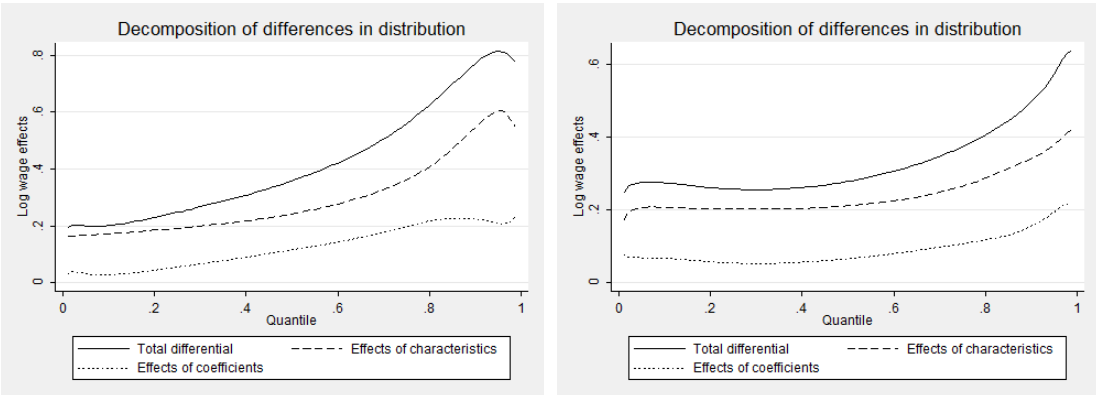
Source: Author's computations using PNAD 1987 and 2006.

Figure B4: Melly (2006) quantile decomposition results of racial wage gaps, disaggregated by formal and non-formal sectors

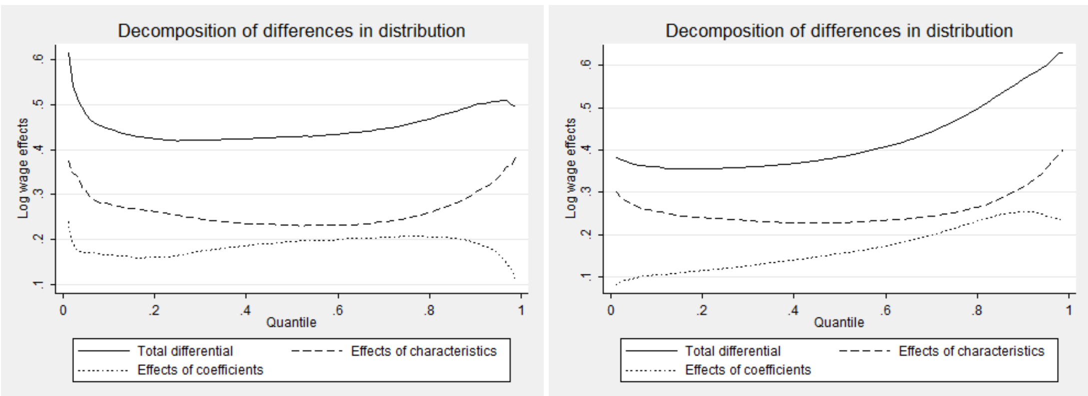
Panel A – Formal sector – 1987 and 2006



Panel B – Informal sector – 1987 and 2006



Panel C – Self-employed sector – 1987 and 2006



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

Chapter 8

Conclusions

This thesis presented a comprehensive portrait of the evolution of occupational segregation and wage discrimination over time in Brazil. These topics are central to the broader field of labour economics and we believe some novel contributions have been made to the understanding of these topics in Brazil. This contribution began with the development of a new harmonized re-classification of occupational codes for the PNAD surveys from 1987 to 2006. This has made it possible to study labour market trends over a more protracted period than previously possible, while also allowing us to compare our results to those found for other countries. The availability of these data has allowed us to present a systematic investigation of patterns of occupational segregation over time, and to link them to a deeper understanding of wage discrimination. In doing so, we have provided a more detailed account of the evolution of these labour market outcomes over time while also appreciating the nature of divergent outcomes by gender and race.

In addition, we have provided a particularly detailed and robust set of results by virtue of having conducted a highly disaggregated analysis, along three dimensions. First, we have conducted the analysis disaggregated by both gender and race, allowing us to compare and contrast experiences within both population sub-groups. Second, we have conducted analysis disaggregated into the formal, informal and self-employed sectors, thus illuminating important differences between these different segments of the labour market. Finally, throughout the analysis we have employed a variety of different methods – including alternative segregation indexes, decomposition methods and corrections for selection – in order to ensure the robustness of the results, and highlighting the important differences across these methods.

The narrative that emerges is of a continuing difference in labour market outcomes across sub-groups, though with significant changes over time and significant differences between groups. For women, we observe very high levels of occupational segregation and initially high levels of wage disadvantage, though both have declined significantly over time. However, despite gains over time, differences in female outcomes in both areas continue to be driven overwhelmingly by unexplained factors. While it is acknowledged that very high levels of occupational segregation may in part reflect differences in tastes and preferences, wage differentials are less easily explained, suggesting the continued existence of unequal treatment in both domains, particularly at the top and bottom of the wage distribution. It is reassuring that the estimated treatment differentials appear to have diminished over time, but this has occurred primarily within low-paid jobs while we see more persistent unequal treatment within the high-paid occupations.

In focusing on differences by race we observe lower levels of occupational segregation, but significantly higher levels of pay differentials. More importantly, we see that both patterns have only improved relatively gradually over time. The divergent outcomes for white and non-white workers are primarily explained by differences in endowments, particularly differences in educational attainment, suggesting an important pre-market barrier to equal opportunity. This pattern is particularly severe within the highly-paid and highly-skilled occupations. The unexplained components of wage discrimination, though comparatively small by race, are extremely persistent having decreased only negligibly over two decades despite an increasing official awareness and concern about racial discrimination.

Finally, it is important to highlight that we uncover significant connections between occupational structure, earnings and wage discrimination. Wages in jobs dominated by both women and non-whites are systematically lower than we would expect. This effect is larger for non-white workers, as non-white dominated occupations are disproportionately low-paid and unskilled. Working in female-dominated occupations constricts female wages while actually increasing male wages, though the latter only applies in low-paid jobs. The unequal treatment affecting female workers is primarily driven by intra-occupational differences, or what we refer to here 'vertical segregation'. In contrast, the endowment differentials, which explain racial pay gaps, reflect differences both between and within occupations. Our analysis revealed that 'vertical' and 'horizontal' segregation explained only one-quarter of the observed racial pay gaps, but these unexplained treatments appear persistent over time.

These patterns point to the complexity and persistence of discrimination, and to the importance of systematically linking the study of occupational segregation and wage discrimination. A key contribution of this thesis thus lies in the fact that the analysis spans two decades during which Brazil has transitioned from a period of economic and political uncertainty to becoming an emerging global power. This transitional period has included important institutional and macroeconomic changes, and it is through the labour market that these broad economic changes impact actual individuals and their level of welfare.

Our analysis suggests clear policy implications, but ones which differ between gender and racial concerns. While the problem of gender inequality remains highly relevant we observe a persistent decline in unequal treatment, which reflect positive changes in many areas. First, Brazilian female workers now have better endowments than men, and higher educational attainment on average. Second, we observed increased and more homogenous participation of women across occupations. Third, despite remaining the main determinant of gender pay disparities, the unexplained treatment effect has declined steadily over time and at a considerable pace. At a broader level, while gender disparities remain a significant concern, it is appears evident that Brazilian economic progress, coupled perhaps with explicit anti-discrimination legislation (though this was not investigated explicitly in this thesis), has helped enhance gender equality. The direction of policy is thus encouraging, though this thesis has pointed towards important areas of concern, including the continued concentration of women in low-paid jobs in the informal sector and strikingly persistent wage gaps at either extremes of the wage distribution.

The portrait of racial inequalities is more worrying. Although occupational segregation by race is lower than that for gender, non-white dominated occupations are predominantly low-skilled and low-paid while improvements over time have been negligible. Racial pay gaps are similarly high and persistent, while both occupational segregation and pay gaps are most pronounced in highly-skilled and high-paid jobs. Unlike the gender case, the main determinant of pay differentials is differences in observed characteristics, and particularly in educational attainment levels. This endowment differential has not declined at the pace hoped. At the same time, unexplained wage differences, while comparatively modest, remain significant and have been equally persistent over time.

The implication is that policies need to focus particularly on addressing pre-market barriers to opportunity. Differential labour market outcomes are explained primarily at earlier stages, residing in pre-market factors related to residential segregation, the quality of schools, and the lack of opportunities for young individuals before entering the job market. There remain important unexplained differences in outcomes, and particularly acute wage gaps at the top of wage distribution, and the persistence of these differences suggests that despite new legislation anti-discrimination efforts need to be made more effective. However, the overarching message remains that addressing racial inequality in the labour market demands a multifaceted approach, with a focus on pre-market inequalities of opportunity serving a key role.

Although this thesis has provided a detailed account of the magnitude and evolution of occupational segregation and wage gaps in Brazil there remain important challenges and opportunities for future research. The remainder of this concluding chapter highlights the limitations of the research undertaken in the thesis and potential directions for future research.

The construction of the re-classification of occupational codes employed in this thesis inevitably encountered cases in which the occupational classifications employed in the PNAD surveys could not be easily matched to corresponding international classifications as discussed in chapter 4. Although we have undertaken extensive checks in order to ensure the overall consistency of the data over time, it is nonetheless important to highlight that we encountered constraints in reconciling occupational codes in particular for specific top managerial positions, within agricultural occupations.

It is important to re-emphasize that while some of these limitations are specific to this re-classification exercise, and the data contained in the PNAD surveys, some imperfections are likely to be inherent in the re-classification exercises undertaken. One reason is that the intrinsic meaning of different occupational labels is likely to change over time, driven by changing technologies and skills requirements. The 'label' for an occupation may remain the same but the actual job may change, as can be easily seen, for example, in the changing skills required of mechanics or machine operators. Alongside the changing nature of occupations lies the problem of how to deal with the creation and destruction of particular occupations when looking at an individual classification over time. At a sufficiently fine level of disaggregation some professions will disappear over time, while other will be created, as economies and technologies change. By adopting a relatively high level of aggregation we seek to avoid these challenges.

The thesis provides a detailed account of the scale of unequal treatment in Brazil across the wage and occupational dimension. However, it did not explicitly assess the role of anti-discrimination legislation and this remains an obvious topic for further investigation. For example, exploiting sub-national variation may present a useful strategy in seeking to capture the impact of this type of public policy on trends in segregation and discrimination over time. This is of special interest during the period under study in Brazil, as the government has introduced a wide range of new legislation since the promulgation of the new constitution in 1988, though with little systematic evaluation of these measures. However, isolating the impact of ADL is extremely difficult, and this thesis has only approached this question only imperfectly by comparing trends in the formal and informal sectors, the former of which is expected to be more affected by public policy measures. Consistent with an impact of ADL, the past two decades have witnessed a more rapid decline in segregation in the formal than in the non-formal sector, while highly segregated occupations have expanded primarily in the informal labour market. However, while these findings are consistent with an impact of ADL on segregation, they fall far short of establishing clear causation.

This remains an avenue for future research. For instance, it may be possible to adopt an econometric strategy that exploits differences over time and across states in order to conduct an *ad hoc* quasi experiment looking at the role of state level differences in either the passage of ADL or the enforcement of federal anti-discrimination laws. This follows research conducted across U.S. States by Neumark and Stock (2006). However, such research requires significant additional data on variation in laws, enforcement and governance over time and across states, and such data are currently unavailable, to the best of the author's knowledge. This research would require significant investment of time and effort in compiling relevant data.

We noted the importance of controlling for potential selectivity bias but also the existence of significant challenges. The validity of the selection correction relies on the validity of the chosen instruments. However, even where we are confident of the validity of our instruments previous research has highlighted a potential lack of robustness (Manski, 1989) and the potential for ambiguities in the interpretation of decomposition results (Neuman and Oaxaca, 2003, 2004). Further challenges may derive from the presence of unobservable heterogeneous selection processes (see Machado, 2011).

These challenges are apparent in the analysis presented in chapter 6, as we find significant differences in the estimated selection effects when employing alternative selection methods. Our solution has been to employ a variety of methods in order to ensure the robustness of our results to alternative methods, and the uncorrected decomposition findings appear invariant to the selection correction procedures. This analysis could be strengthened by examining more deeply the causes of the different outcomes across the different selection processes. This might include disaggregating the workforce further in order to consider non-white females, white female, non-white males and white males separately to attempt to deal with underlying heterogeneity in the selection process.

The selection correction procedure should also be applied for the more complicated quantile decomposition analysis presented in chapter 7. However, this represents a very complex set of challenges. There is, at present, little consensus regarding the correct procedure for applying selection corrections to quantile decomposition analysis. Aside from the potential for lack of robustness and ambiguities which applies to any decomposition technique, selection correction within a quantile framework suffers from a number of drawbacks. These include the choice of the correct estimation method for the first stage (the non-parametric single index model is currently generally preferred) and the problem of the identification of the constant. While selection correction within decomposition techniques is acknowledged to be problematic, its application within a quantile framework is more complex. However, this represents an interesting area for future research and one that would be useful to undertake verify (or otherwise) the key findinsg reported in this chapter.

At the centre of this thesis, and of most comparable studies, is a desire to understand whether observed wage differentials can be explained by differences in the characteristics of different workers, or reflect unexplained differences in outcomes, attributable to unobserved factors. The latter are of special interest, because it is with this ‘unobserved category’ that we may detect the existence of explicit discrimination against particular population groups. Throughout this thesis we have drawn special attention to such unexplained differences, as they appear likely to be of particular concern. However, the methods adopted here do not explicitly distinguish between discrimination *per se* and other unobservable factors that may explain contrasting outcomes.

A potential method for overcoming this limitation is the analysis of gender and race based wage differentials in the Brazilian formal labour market by employing employer-employee matching data, broadly following the methodology employed by Hellerstein and Neumark (2005). It is possible using such data to estimate both production functions and wage equations in order to compare productivity differentials with wage differentials, as this comparison is likely to control for a range of unobservable factors such as innate ability. Such an approach represents a stronger test for discrimination at the establishment-level.

The relevant data for this type of exercise are potentially available from from Brazilian administrative files - *the Relação Anual de Informações Sociais* (RAIS)- which are maintained by the Brazilian Ministry of Employment and Labour (*Ministério do Trabalho e Emprego*). In Brazil, all registered, tax-paying, establishments must send the Ministry information on all employees who worked anytime during the reference year. The RAIS data are essentially a matched employer-employee longitudinal database, and provide information on workers (such as gender, age and schooling), as well as on firms (such as location, industry, activity, input and output indicators, namely costs, revenues, number of employers, profit). To our knowledge the only study to have used these data in exploring gender wage differentials is that of Foguel (2006), who studied the relationship between female segregation across establishments and the wages of male and female workers. However, a drawback of these data is that they only cover the formal labour market in Brazil.

Brazil presents an ideal contextual setting for a further investigation of discrimination, particularly in regard to race. This thesis has made a significant contribution in highlighting the extent and evolution of segregation and discrimination by both gender and race in Brazil, and provides a useful benchmark for researchers in this field. The trends observed will be central to shaping future developments in Brazil, and particularly in shaping the nature and magnitude of inequality in the country. Thus, future research aimed at further understanding the determinants of segregation and discrimination, and explicitly incorporating the role public policy interventions, is likely to be of significant policy benefit.

References

Abramo, L. (2004). Desigualdades e discriminacao de genero e raca no Mercado de trabalho brasileiro e suas implicacoes para a formulacao de uma politica de emprego. Testo elaborado para o Seminario Nacional: Politica peral de emprego: Necesidades, opcoes, prioridades, OIT, Brasilia, 9 e 10 de dezembro de 2004.

Akelof, G. A. (1997). Social distance and social decisions. Econometrica 65 (5), 1005-1027.

Akerlof, G. A. and R. E. Kranton (2000). Economics and Identity. *The Quarterly Journal of Economics*, 115(3), 715-753.

Albelda, R. P. (1986). Occupational segregation by race and gender, 1958-1981. *Industrial and Labor Relations Review* 39 (3), 404-411.

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

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

Alexim, J. C., P. Cappellin and C. Lettieri, (2005). A experiencia dos nucleos de promocao da igualdade de oportunidades e combate a discriminacao no emprego e na ocupacao. Brasilia:ILO.

Almeida, R. and P. Carneiro (2007). Inequality and employment in a dual economy: Enforcement of labor regulation in Brazil. IZA Discussion Papers 3094, Institute for the Study of Labor (IZA).

Alonso-Villar, O., C. Del Rio, and C. Gradin (2010). The extent of occupational segregation in the U.S.: Differences by race, ethnicity, and gender. Working Paper 2010-180, ECINEQ.

Altonji, J. G. and R. M. Blank (1999). Race and gender in the labor market. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3 of Handbook of Labor Economics, Chapter 48, pp. 3143-3259. Elsevier.

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

Angrist, J. D., G. D. Imbens, and D. B. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434), 444-454.

Anker, R. (1997). Theories of occupational segregation by sex: An overview. *International Labour Review* 136 (3).

Anker, R. (1998). Gender and Jobs: Sex Segregation of Occupations in the World. Geneva: International Labor Organization.

Anker, R., H. Melkas, and A. Korten (2003). Gender-based occupational segregation in the 1990's. Working Paper 16, ILO Declaration.

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

Arabsheibani, G. R., F. G. Carneiro, and A. Henley (2003). Gender wage differentials in Brazil: Trends over a turbulent era. Policy Research Working Paper 3148, World Bank.

Arbache, J. S. (2001). Wage differentials in Brazil: theory and evidence. *The Journal of Development Studies* 38 (2), 109-130.

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.

Arias, O., G. Yamada, and L. Tejerina (2004). Education, family background and racial earnings inequality in Brazil. *International Journal of Manpower* 25 (3/4), 355-374.

Arriagada, A.-M. And A. Ziderman (1992). Vocational secondary schooling, occupational choice, and earnings in Brazil. Policy Research Working Paper Series 1037, The World Bank.

Arrow, K. (1973). The theory of discrimination. In O. Ashenfelter and A. Rees (Eds.), Discrimination in Labor Markets, pp. 143-164. New Jersey: Princeton University Press.

Arrow, K. J. (1998). What has economics to say about racial discrimination? *The Journal of Economic Perspectives* 12 (2), 91-100.

Ashenfelter, O. (1970). Changes in labor market discrimination over time. *Journal of Human Resources* 5, 403-429.

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.

Azzoni, C. R. and L. M. S. Servo (2002). Education, cost of living and regional wage inequality in Brazil. *Papers in Regional Science* 81 (2).

Baer, W. (2008). The Brazilian Economy: Growth and Development. Westport, Connecticut: Praeger.

Bailey, S. R. (2002). The race construct and public opinion: Understanding Brazilian beliefs about racial inequality and their determinants. *American Journal of Sociology* 108 (2), 406-39.

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

Baldwin, M. L., R. J. Butler, and W. G. Johnson (2001). A hierarchical theory of occupational segregation and wage discrimination. *Economic Inquiry* 39 (1), 94-110.

Barros, R. P. d. and C. H. Corseuil (2001). The impact of regulations on Brazilian labor market performance. Mimeo.

Barros, R. P. d., A. F. Machado, and R. S. P. d. Mendonca (1997). A desigualdade da pobreza: estrategias ocupacionais e diferenciais por genero. Texto para discussao 453, IPEA, Rio de Janeiro.

Barros, R. P. d., M. de Carvalho, S. Franco, and R. Mendonca (2006). Una analise de principais causas da queda recente na disegualdade de renda brasileira. Texto para Discussao 1203, IPEA, Rio de Janeiro.

Barros, R. P. d., R. Mendonça, P.Pereira Deliberalli and C. Lopes (2000). Impacto da Distribuição da Terra sobre a Eficiência Agricola e a pobreza no Nordeste. In Henriques, R. (ed) Desigualdade e Pobreza no Brasil, Rio de Janeiro: IPEA.

Baunach, D. M. (2002). Trends in occupational sex segregation and inequality, 1950 to 1990. *Social Science Research* 31, 77-98.

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

Becker, G. (1971). The Economics of Discrimination (2nd ed.). Chicago: Chicago University Press.

Behrman, J. R. and N. Birdsall (1983). The quality of schooling: Quantity alone is misleading. *American Economic Review* 73 (5), 928-46.

Beltrao, K. I. (2003). Alfabetizacao por sexo e raca no Brasil: Un modelo linear generalizado para explicar a evolucao no periodo 1940-2000. Texto para Discussao 1003, IPEA, Rio de Janeiro.

Beltrao, K. I. and M. de Poli Teixeira (2004). O vermelho e o negro: Raca e genero na universidade brasileira. uma analise da seletividade das carreiras a partir dos censos demograficos de 1960 a 2000. Texto para Discussao 1052, IPEA, Rio de Janeiro.

Bertrand, M. (2011). New Perspectives on Gender, Volume 4 of Handbook of Labor Economics, Chapter 17, pp. 1543-1590. Elsevier.

Binelli, C. (2008). Returns to education and increasing wage inequality in Latin America. Institute for Fiscal Studies, mimeo.

Binelli, C., C. Meghir, and N. Menezes-Filho (2009). Education and wages in Brazil. Mimeo.

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.

Blackaby, D. H., D. G. Leslie, P. D. Murphy, and N. C. O'Leary (1998). The ethnic wage gap and employment differentials in the 1990s: Evidence for Britain. *Economics Letters*, 97-103.

Blackburn, R. M. and J. Jarman (2005). Segregation and inequality. Working Paper 3, GeNet.

Blackburn, R. M., B. Brooks, and J. Jarman (2001). The vertical dimension of occupational segregation. *Work Employment Society* 15, 511-538.

Blackburn, R. M., J. Jarman, and J. Siltanen (1993). The Analysis of Occupational Gender Segregation over Time and Place: Considerations of Measurement and Some New Evidence. *Work, Employment and Society* 7.

Blau, F. D. and L. M. Kahn (1996, Suppl.). Wage structure and gender earnings differentials: An international comparison. *Economica* 63 (250), S29-62.

Blau, F. D. and W. E. Hendricks (1979). Occupational segregation by sex: Trends and prospects. *Journal of Human Resources* 14 (2), 197-210.

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

Boisso, D., K. Hayes, J. Hirschberg, and J. Silber (1994). Occupational segregation in the multidimensional case: Decomposition and tests of statistical significance. *Journal of Econometrics* (61), 161-171.

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

Bosch, M., E. Goni, and W. Maloney (2007). The determinants of rising informality in Brazil: Evidence from gross worker flows. Working Paper No. 4375, *World Bank Policy Research*.

Bourguignon, F., F.H.G. Ferreira (2005). Decomposing Changes in the Distribution of Household Incomes: Methodological Aspects. In The Microeconomics of Income Distribution Dynamics in East Asia and Latin America. F. Bourguignon, F.H.G. Ferreira and N. Lustig (eds.) World Bank. pp. 17-46.

Bourguignon, F., F.H.G. Ferreira, and P. G. Leite (2008). Beyond Oaxaca. Blinder: Accounting for differences in household income distributions. *Journal of Economic Inequality* 6: 117-148.

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

Brito, F. and J. A. M. de Carvalho (2006). As migracoes internas no brasil: As novidades sugeridas pelos censos demograficos de 1991 e 2000 e pela pnads recentes. Departamento de Demografia e do Centro de Desenvolvimento e Planejamento Regional (CEDEPLAR), UFMG, mimeo.

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.

Butler, R. J. (1987). New indices of segregation. *Economics Letters* 24, 359-362.

Cacciamali, M. C. (1982). Um estudo sobre o setor informal urbano e formas de participacao na producao. Ph. D. thesis, Faculdade de Economia, Adminsitracao e Contabilidade da Universidade de Sao Paulo.

Cacciamali, M. C. (1988). Mudancas estruturais no produto e emprego no Brasil: 1950-80. FEA/USP, mimeo.

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.

Cain, G. G. (1987). The economic analysis of labor market discrimination: A survey. In O. Ashenfelter and R. Layard (Eds.), *Handbook of Labor Economics* (1 ed.), Volume 1, Chapter 13, pp. 693-781. Elsevier.

Calonico, S. and H. Ñopo (2008). Gender segregation in the workplace and wage gaps: Evidence from urban Mexico 1994-2004. RES Working Papers 4579, Inter-American Development Bank, Research Department.

Campante, F. R., A. R. V. Crespo, and P. G. P. G. Leite (2004). Desigualdade salarial entre racas no mercado de trabalho urbano brasileiro: Aspectos regionais. *Revista Brasileira de Economia* 58 (2), 185-210.

Card, D. (1999). The causal effect of education on earnings. Handbook of Labor Economics, in: O. Ashenfelter & D. Card (ed.), Handbook of Labor Economics, edition 1, volume 3, chapter 30, pages 1801-1863 Elsevier.

Card, D. and J.E. DiNardo (2002). Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles. *Journal of Labor Economics*, 20(4), 733-783.

Carneiro, F. (1997). The changing informal labour market in brazil: Cyclicality versus excessive intervention. *Labour* 11 (1), 3-22.

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.

Cavalieri, C. H. and R. Fernandes (1998). Diferenciais de salarios por genero e por cor: Uma comparacao entre as regioes metropolitanas brasileiras. *Revista de Economia Politica* 18 (1), 158-175.

Charles, M. and D. B. Grusky (1995). Models for describing the underlying structure of sex segregation. *The American Journal of Sociology* 100 (4), 931-971.

Chernozhukov, V., I. Fernandez-Val, and B. Melly (2009). 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). 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.

Curi, A. Z. And N. A. Menezes Filho (2007). As Causas Da Estagnação Educacional Recente No Brasil. Anais do XXXV Encontro Nacional de Economia. Proceedings of the 35th Brazilian Economics Meeting. 127, ANPEC - Associação Nacional dos Centros de Pósgraduação em Economia. Brazilian Association of Graduate Programs in Economics.

Darity, W. A. and P. L. Mason (1998). Evidence on discrimination in employment: Codes of color, codes of gender. *Journal of Economic Perspectives* 12 (2), 63-90.

de la Rica, S., J. J. Dolado, and V. Llorens (2005). 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.

Deutsch, J. and J. Silber (2005). Comparing segregation by gender in the labour force across ten European countries in the late 1990s: An analysis based on the use of normative segregation indices. *International Journal of Manpower* 26 (3), 237-264.

Deutsch, J., Y. Fluckinger, and J. Silber (1994). Measuring occupational segregation. Summary statistics and the impact of classification errors and aggregation. *Journal of Econometrics* 61, 133-146.

Deutsch, J., Y. Flueckiger, and J. Silber (2009). Analysing changes in occupational segregation: the case of Switzerland 1970-2000. In J. Silber, Y. Fluckiger, and S. F. Reardon (Eds.), Occupational and Residential Segregation Research on Economic Inequality. Emerald Group Publishing Limited.

Deutsch, R., A. Morrison, C. Piras, and H. Ñopo (2005). Working within confines: Occupational segregation by gender in three Latin American countries. The Icfai Journal of Applied Economics IV (3), 50-59.

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. and G. H. Makepeace (1985). The statistical measurement of discrimination. *Economic Letters*, 391-395.

Dolton, P. J. and G. H. Makepeace (1987). Interpreting sample selection effects. *Economics Letters* 24 (4), 373-379.

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 (2000). 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.

Duncan, O. D. and B. Duncan (1955). A methodological analysis of segregation indexes. *American Sociological Review* 20, 210-17.

Efron B. (1979). Bootstrap methods: another look at the Jackknife. *Annals of Statistics* 7, 1-26.

Efron, B. and R. J. Tibshirani (1991). Statistical data analysis in the computer age. *Science* 253, 390-395.

Efron, B. and R. J. Tibshirani (1993). An Introduction to the Bootstrap. New York: Chapman & Hall.

Ehrenberg, R. G. and R. S. Smith (2009). Modern Labor Economics: Theory and Public Policy. Pearson/Addison Wesley.

Ferreira, F. H. G. and R. P. D. Barros (1999). The slippery slope: Explaining the increase in extreme poverty in urban Brazil 1976-1996. Textos para discussao 404, Department of Economics PUC-Rio (Brazil).

Ferreira, F. H. G., P. G. Leite, and M. Wai-Poi (2007). Trade liberalization, employment flows and wage inequality in Brazil. Policy Research Working Paper 4108, World Bank.

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

Fiess, N. M., M. Fugazza, and W. F.Maloney (2008). Informality and macroeconomic fluctuations. Working Paper 3519, IZA.

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

Firpo, S., N. M. Fortin, and T. Lemieux (2011). Occupational tasks and changes in the wage structure. IZA Discussion Papers 5542, Institute for the Study of Labor (IZA).

Flueckiger, Y. and J. Silber (1999). The Measurement of Segregation in the Labor Force, Heidelberg. New York: Physica-Verlag.

Fonseca, M. and G. Rayp (2011). A view of trade liberalization on the agricultural labor market in Brazil. Ghent University, mimeo.

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). Decomposition Methods in Economics, Volume 4 of Handbook of Labor Economics, Chapter 1, pp. 1-102. Elsevier.

Freguglia, R., G. Spricigo, G. Johnes, and A. Aggarwal (2011). Education and labour market outcomes: Evidence from Brazil. Working paper, Lancaster University Management School.

Freyre, G. (1933). Casa Grande e Senzala. Rio de Janeiro: Jose Olimpio.

Fryer Jr, R. G. (2010). The importance of segregation, discrimination, peer dynamics, and identity in explaining trends in the racial achievement gap. Working Paper 16257, NBER.

Gaddis, I. and J. Pieters (2012). Trade liberalization and female labor force participation: Evidence from Brazil. Discussion Papers 6809, IZA.

Gasparini, L. and L. Tornarolli (2007). Labor informality in Latin America and the Caribbean: Pattern and trends from household survey microdata. CEDLAS, mimeo.

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

Goldberg, P. K. and N. Pavcnik (2003). The response of the informal sector to trade liberalization. *Journal of Development Economics* 72 (2), 463-496.

Goldthorpe, J.H. (2000). Globalisation and social class. *West European Politics*, 25: 1-28.

Gomes, B. F. (2006). Migracao interna e urbanizacao no Brasil contemporaneo: Um estudo da rede de localidades centrais do Brasil (1980/2000). Universidade Salgado de Oliveira, mimeo.

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

Gradin, C. (2007). Why is poverty so high among Afro-Brazilians? A decomposition analysis of the racial poverty gap. Discussion Paper 2809, IZA.

Gradin, C. (2010). Conditional occupational segregation of minorities in the US. Working Paper 185, ECINEQ.

Gradin, C., C. del Rio, and O. Alonso-Villar (2011). Occupational segregation by race and ethnicity in the US: Differences across states. ECINEQ WP 2011 – 190.

Green, F., A. Dickerson, and J. S. Arbache (2001). A picture of wage inequality and the allocation of labor through a period of trade liberalization: The case of Brazil. *World Development* 29 (11), 1923-1939.

Gronau, R. (1974). Wage Comparisons-A Selectivity Bias. *Journal of Political Economy* 82(6), 1119-43.

Guimares, R. d. O. (2006). Desigualdade salarial entre nogros e brancos no Brasil: discriminacao ou exclusao? *Economica* 8 (2), 227-251.

Gyourko, J. and J. Tracy (1988). An analysis of public- and private-sector wages allowing for endogenous choices of both government and union status. *Journal of Labor Economics* 6 (2), 229-53.

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.

Hakim, C. (1992). Explaining trends in occupational segregation: The measurement, causes, and consequences of the sexual division of labour. *European Sociological Review* 8 (2), 127-152.

Hakim, C. (1993). Segregated and integrated occupations: A new approach to analysing social change. *European Sociological Review* 9 (3), 289-314.

Hart, K. (1971). Informal Income Opportunities and Urban Employment in Ghana. ILO, mimeo.

Hart, K. (1972). Employment, income and inequality: A strategy for increasing productive employment in Kenya. Geneva: ILO.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica* 47 (1), 153-161.

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). 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. and D. Neumark (2008). Workplace segregation in the United States: Race, ethnicity, and skill. *The Review of Economics and Statistics* 90 (3), 459-477.

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

Henley, A., R. G. Arabheibani, and F. G. Carneiro (2009). On defining and measuring the informal sector: Evidence from Brazil. *World Development* 37 (5), 992-1003.

Henriques, R. (2001). Desigualdade racial no Brasil: Evolucao das condicoes de vida na decada de 90. Texto para Discussao 807, IPEA, Rio de Janeiro.

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

Hirsch, B. T. and D. A. Macpherson (2003). Wages, sorting on skill, and the racial composition of jobs. IZA Discussion Papers 741, Institute for the Study of Labor (IZA).

Hoffman, R. (1989). Evolucao da Distribuicao da Renda no Brasil, Entre Pessoas e Entre Familias, 1979/86. In Sedlacek, G.L. and R. Paes de Barros (eds.) Mercado de Trabalho e Distribuicao de Renda: Uma Coletanea (Rio de Janeiro: Instituto de Pesquisa Economica Aplicada (IPEA) Serie Monografica No. 35).

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.

Hutchens, R. (1991). Segregation curves, Lorenz curves, and inequality in the distribution of people across occupations. *Mathematical Social Sciences* 21, 31-51.

Hutchens, R. (2004). One measure of segregation. *International Economic Review* 45 (2), 555-578.

IACHR (1997). Report on the Situation of Human Rights in Brazil 1997. Inter-American Commission on Human Rights Organization of American States.

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

ILO (2010). Trabalho domestico no Brasil: rumo ao reconhecimento institutional. Technical report, International Labour Organization (ILO).

Isaza-Castro, J. G. and B. Reilly (2011). Occupational segregation and gender wage differences: evidence from urban Colombia. Mimeo, University of Sussex, Brighton, U.K.

James, D. R. and K. E. Taeuber (1985). Measures of segregation. *Sociological Methodology* 15, 1-32.

Jatoba, G. (1987). Urban poverty, labor markets and regional development: the case of Brazil. Article presented in Brazilian Meeting on Labor Markets and Income Distribution. Rio de Janeiro, mimeo.

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

Jones, F. L. (1983). On Decomposing the Wage Gap: A Critical Comment on Blinder's Method. *Journal of Human Resources* 18, 126.130.

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.

Kakwani, N. C. (1994). Segregation by sex: Measurement and hypothesis testing. In S. Neuman and J. Silber (Eds.), Inequality in Labor Markets: The Economics of Labor Market Segregation and Discrimination, pp. 1-26. Greenwich, CT and London, England: Research on Economic Inequality. JAI Press.

Karmel, T. and M. Maclachlan (1988). Occupational sex segregation - increasing or decreasing? *The Economic Record* 64, 187-95.

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

King, M. C. (1992). Occupational segregation by race and sex, 1940-1988. *Monthly Labor Review* 115 (4), 30-37.

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

Koenker, R. W. and G. J. Bassett, (1978). 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).

Lago, L. C. d. (2006). Trabalho, moradia e (i)mobilidade especial na metropole do Rio de Janeiro. Written for the XVI World Congress of the International Sociological Association.

Laing, D. (2011). Labor Economics: Introduction to Classic and the New Labor Economics. W. W. Norton & Co.

Langoni, C. G. (1973). Distribucao da Renda e desenvolvimento Economico do Brasil. Rio de Janeiro: Editora Expressao e cultura.

Lee, L.-F. (1983). Generalized econometric model with selectivity. *Econometrica* 51 (2), 507-512.

Lemieux, T. (2006). Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill? *American Economic Review* 96 (3), 461-498.

Lewis, D. E. (1982). The measurement of the occupational and industrial segregation of women. *Journal of Industrial Relations* 24 (3), 406-423.

Litchfield, J. A. (2001). Poverty, Inequality and Social Welfare in Brazil, 1981-1995. PhD Thesis. London School of Economics and Political Science.

Liu, P.-W., J. Zhang, and S.-C. Chong (2004). Occupational segregation and wage differentials between natives and immigrants: evidence from Hong Kong. *Journal of Development Economics* 73 (1), 395-413.

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.

Lovell, P. A. (1994). Race, gender and development in Brazil. *Latin American Research Review* 29 (3), 1-35.

Lovell, P. A. (2000). Race, gender and regional labour market inequalities in Brazil. *Review of Social Economy* 58 (3), 277-293.

Lovell, P. A. (2006). Race, gender, and work in Sao Paulo, Brazil, 1960-2000. *Latin American Research Review* 41 (3).

Lovell, P. A. and C. H. Wood (1998). Skin colour, racial identity and life chances in Brazil. *Latin American Perspectives*, Issue 25 (3), 90-109.

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

Machado, A. F., A. M. H. C. d. Olivera, and N. F. Carvalho (2003). Tipologia de qualificacao da forma de trabalho: Uma proposta a partir da nocao de impompatibilidade entre ocupacao e escolaridade. Technical report, Universidade Federal de Minas Gerais.

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. (2010). Occupational segregation and the gender wage gap in brazil: An empirical analysis. *Economia aplicada* 14 (2), 147-168.

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.

Maloney, W. F. (1999). Does informality imply segmentation in urban labor markets? Evidence from sectoral transitions in Mexico. *The World Bank Economic Review* 13 (2), 275-302.

Maloney, W. F. (2004). Informality revisited. *World Development* 32 (7), 1159-1178.

Manning, A. (1996). The equal pay act as an experiment to test theories of the labour market. *Economica* 63, 191-212.

Manning, A. (2003). The real thin theory: monopsony in modern labour markets. *Labour Economics*. 10(2), 105-131.

Manning, A. and B. Petrongolo (2008). The part-time pay penalty for women in Britain. *Economic Journal* 118 (526), 28-51.

Manski, C. F. (1989). Anatomy of the selection problem. *Journal of Human Resources* 24 (3), 343-360.

Marquez Garcia, L.,H. Ñopo, and P. Salardi (2009). 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.

Marshall, R. (1974). The economics of racial discrimination: A survey. *Journal of Economic Literature* 12 (3), 849-871.

Maume, D. J. (1999). Glass ceilings and glass escalators. *Work and Occupations* 26 (4), 483-509.

Melkas, H. and R. Anker (1997). Occupational segregation by sex in Nordic countries: An empirical investigation. *International Labour Review* 136 (341).

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.

Meng, X. and Miller, P. W. (1995). Occupational Segregation and Its Impact on Gender Wage Discrimination in China's Rural Industrial Sector. *Oxford Economic Papers*, 47(1), 136-55.

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

Miller, P. W. and P. A. Volker (1985). On the determination of occupational attainment and mobility. *The Journal of Human Resources* 20 (2), 197-213.

Mincer, J. A. (1974). The Human Capital Earnings Function. NBER Chapters, in: Schooling, Experience, and Earnings, pages 83-96 National Bureau of Economic Research, Inc.

Moir, H. and J. S. Smith (1979). Industrial segregation in the Australian labour market. *Journal of Industrial Relations* 21 (3), 281-291.

Mora, R. and J. Ruiz-Castillo (2003). Additively decomposable segregation indexes. the case of gender segregation by occupations and human capital levels in Spain. *Journal of Economic Inequality* 1, 147-179.

Muendler, M.-A., J. Poole, G. Ramey, and T. Wajnberg (2004). Job concordances for Brazil: Mapping the classicacao brasileira de ocupacoes(CBO) to the international standard classification of occupations (isco-88). University of California, San Diego, mimeo.

Neuman, S. (1994). Ethnic occupational segregation in israel. In S. Neuman and J. Silber (Eds.), Inequality in Labor Markets: The Economics of Labor Market Segregation and Discrimination. Greenwich, CT and London, England: Research on Economic Inequality. JAI Press.

Neuman, S. (1998). Occupational segregation in the Israeli labour market: The gender-ethnicity interaction. *International Journal of Manpower* 19 (8).

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 (2003). Estimating Labor Market Discrimination with Selectivity-Corrected Wage Equations: Methodological Considerations and An Illustration from Israel. Discussion Paper No. 2-2003. The Pinhas Sapir Center for Development Tel-Aviv University.

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.

Neumark, D. and W. A. Stock (2006). The labor market effects of sex and race discrimination laws. *Economic Inquiry* 44 (3), 385-419.

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

Nicodemo, C. (2009). 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 (1988). Searching for the Effect of Unionism on the Wages of Union and Nonunion Workers. *Journal of Labor Research* 9, 139-148.

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

Oaxaca, R. L. and M. R. Ransom (1999). Identification in Detailed Wage Decompositions. *Review of Economics and Statistics* 81, 154.157.

Oliveira, A. M. H. C. d. (2001). Occupational gender segregation and effect on wages in Brazil. XXIV General Population Conference.

Olivetti, C. and B. Petrongolo (2008). Unequal Pay or Unequal Employment? A Cross-Country Analysis of Gender Gaps. *Journal of Labor Economics*, 26, 621-654.

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.

Oppenheimer, V. K. (1969). The Female Labor Force in the United States: Demographic and Economic Factors governing its growth and changing composition. Westport: Greenwood Press Publishers.

Osorio, R. G. (2008). Is all socioeconomic inequality among racial groups in Brazil caused by racial discrimination? Working Papers 43, International Policy Centre for Inclusive Growth.

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

Phelps, E. S. (1972). The statistical theory of racism and sexism. *The American Economic Review* 62 (4), 659-661.

Piore M. (1972). Notes for a theory of labor market stratification (1972), MIT Working Paper No. 95, Dept. of Econ., Mass. Instit. of Tech. Oct. 1972.

Pochmann, M. (2007). Nova geoeconomia do emprego no brasil: um balanco de 15 anos nos estados da federacao. UNICAMP, mimeo.

Ramos, L. and V. Ferreira (2005). Padroes easpacial e setorial da evolucao da informalidade no Brazil - 1991-2003. Texto para Discussao 1099, IPEA, Rio de Janeiro.

Reardon, S. F. and G. Firebaugh (2002). Response: Segregation and social distance - a generalized approach to segregation measurement. *Sociological Methodology* 32, 85-101.

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

Reimers, C. (1983). Labor market discrimination against Hispanic and black men. *The Review of Economics and Statistics* 65 (4), 570-79.

Ribeiro, M. (2008). Las mujeres negras en la lucha por sus derechos. *Nueva Sociedad* 218, 131-147.

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.

Santos, S. A. d. and N. O. I. d. Silva (2006). Brazilian indifference to racial inequality in the labour market. *Latin American Perspective* 149(33)(4), 13-29.

Sastre, M. and A. Trannoy (2002). Shapley inequality decomposition by factor components: Some methodological issues. *Journal of Economics* 9 (1), 51-89.

Semyonov, M. and F. L. Jones (1999). Dimensions of gender occupational differentiation in segregation and inequality: A cross-national analysis. *Social Indicators Research* 46, 225-247.

Shorrocks, A. F. (1999). Decomposition procedures for distributional analysis: A unified framework based on the Shapley value. University of Essex, mimeo.

Silber, J. (1989). On the measurement of employment segregation. *Economics Letters* 30, 237-43.

Siltanen, J., J. Jarman, and R. M. Blackburn (1995). Gender inequality in the labour market - Occuaptional concentration and segregation. Geneva: ILO press.

Silva Jr., H. (1998). Anti-Racismo – Coletanea de Leis Brasileiras (Federais, Estaduais, Municipais), Sao Paulo: Editora Oliveira Mendes.

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).

Skidmore, T. E. (1985). Race and class in Brazil: Historical perspectives. In P. Fontaine (Ed.), Race, Class and Power in Brazil. Los Angeles: CAAS.

Skidmore, T. E. (1993). Bi-racial U.S. versus multi-racial Brazil: Is the contrast still valid? *Journal of Latin American Studies* 25 (2), 373-86.

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

Soares, F. (2004). Some stylized facts of the informal sector in Brazil in the 1980's and 1990's. Texto para Discussao 1020, IPEA.

Soares, S. S. D. (2000). O perfil da discriminacao no mercado de trabalho – homens negros, mulheres brancas e mulheres negras. Texto para Discussao 769, IPEA Brasilia.

Soares, S. S. D. (2006). Distribuicao de renda no Brasil de 1976 a 2004 com enfase no periodo entre 2001 e 2004. (1166).

Soares, S. S. D. and R. S. Izaki (2002). A participacao feminina no mercado de trabalho. Texto Para Discussao 023, IPEA, Rio de Janeiro.

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.

Strauss, J. and T. Duncan (1991). Wages, Schooling and Background: Investments in Men and Women in Urban Brazil. In Opportunity Foregone: Education in Brazil edited by Nancy Birdsall, and Richard Sabot.

Telles, E. E. (1992a). Residential segregation by skin colour in Brazil. *American Sociological Review* 57 (2), 186-197.

Telles, E. E. (1992b). Who gets formal sector jobs? Determinants of formal-informal participation in Brazilian metropolitan areas. *Works and Occupations* 19 (2), 108-

Telles, E. E. (1994). Industrialization and racial inequality in employment: The Brazilian example. American Sociological Review 59 (1), 46-63.

Telles, E. E. (2006). Race in Another America: The Significance of Skin Colour in Brazil. Princeton Press.

Telles, E. E. and N. Lim (1998). Does it matter who answers the race question? Racial classification and income inequality in Brazil. *Demography* 35 (4), 465-474.

Theil, H. (1967). Economics and Information Theory. Amsterdam: North Holland.

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.

Tzannatos, Z. (1999). Women and labor market changes in the global economy: Growth helps, inequalities hurt and public policy matters. *World Development* 27, 551-569.

Ulyssea, G. (2005). Informalidade no mercado de trabalho brasileiro: uma resenha da literatura. Texto para Discussao 1070, IPEA, Rio de Janeiro.

Ulyssea, G. (2010). Regulation of entry labor market institutions and the informal sector. *Journal of Development Economics* 91, 8799.

UNIFEM (2006). O Progreso das Mulheres no Brasil. Brasilia: UNIFEM and Ford Fundation.

Urani, A. (1996). Ajuste macroeconomico e exibilidade do mercado de trabalho no brasil - 1981/95. Texto para discussao 380, IPEA, Rio de Janeiro.

Wajnman, S. and E. R. Neto (2000a). Quanta serao as mulheres: Cenario para a atividade feminina. In M. I. B. Rocha (Ed.), Trabalho e Genero - Mudancas, Permanencias e Desafios, pp. 59-84. Sao Paulo: Editora 34.

Wajnman, S. and E. R. Neto (2000b). Women's participation in the labor market in Brazil: elements for projecting levels and trends. *Brazilian Journal of Population Studies* 2 (1999).

Wajnman, S., A. M. H. C. d. Oliveira, and E. L. d. Oliveira (2004). Os idosos no mercado de trabalho: tendencias e consequencias. In A. A. Camarano (Ed.), Os novos idosos brasileiros Muito alem dos 60?, pp. 453-79. Brasilia: IPEA.

Watts, M. (1998). Occupational gender segregation: Index measurement and econometric modeling. *Demography* 35 (4), 489-496.

Wood, C. (1991). Categorias censitarias e classificacoes subjectivas de raca no Brasil. In L. P. A. (Ed.), Desigualdade racial no Brasil Contemporaneo, pp. 93-113. Belo Horizonte: MGSP Editores.

World Bank (2001). Brazil Country Brief. Washington D.C.: World Bank.

World Bank (2002a). Brazil Jobs report. Volume I: Policy briefing. Report, World Bank.

World Bank (2002b). Brazil Jobs report. Volume II: Background papers. Report, World Bank.

World Bank (2011). World Development Indicator 2011. Washington D.C.: 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.

Zabalza, A. and Z. Tzannatos (1985). The effect of Britain's anti-discriminatory legislation on relative pay and employment. *The Economic Journal* 95 (379), 679- 699.