Essays on the Formal Labor Market of Brazil

by

Jason M. Rivera

(Under the Direction of Christopher Cornwell)

Abstract

In this dissertation, I study the formal labor market of Brazil using restricted-access, employer-employee matched data from RAIS. In the first chapter, I show that estimates derived from administrative data smaller wage gaps, driven by racial bias; estimates from survey data yield much larger wage gaps, driven not just by racial bias, but also by differences in observable worker characteristics. In the second chapter, I identify wage discrimination using variation in the perception of the same worker's race by different employers and find that roughly 30 percent of the racial wage gap among workers changing jobs is associated with variation in the employer's perception of race and find that race change is strongly associated with movements of workers between racially segregated plants. In the final chapter, I estimate firm-specific starting pay and wage-tenure profiles and document features between these estimates and turnover that provide insight about firms personnel management practices.

Index words: Labor (General), Demography, Brazil, Econometrics

Essays on the Formal Labor Market of Brazil

by

Jason M Rivera

B.A., The University of Texas at Austin, 2004

M.A., Clemson University, 2009

A Dissertation Submitted to the Graduate Faculty

of The University of Georgia in Partial Fulfillment

of the

Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2014

Essays on the Formal Labor Market of Brazil

by

Jason M. Rivera

Approved:

Major Professor: Christopher M. Cornwell

Committee: Ian M. Schmutte
Ronald S. Warren, Jr.

Electronic Version Approved:

Julie Caulfield
Interim Dean of the Graduate School
The University of Georgia
August 2014

Essays on the Formal Labor Market of Brazil

Jason M. Rivera

July 21, 2014

Acknowledgments

Over the past few years I have received support, guidance and encouragement from many people. Dr. Christopher Cornwell's supervision of my research, from vague ideas to well-defined essays, has made this dissertation a thoughtful, challenging and ultimately rewarding process. Thank you to Dr. Ian Schmutte, who helped me learn the programming skills necessary to complete this project, and Dr. Ronald Warren, Jr. who provided me with sound advice and constructive criticism. All of my committee members have come to be trusted colleagues, invaluable teachers and friends.

I would like to thank Rosângela Jardim de Farias and Maria Emília Piccinini Veras of the *Ministério do Trabalho e Emprego*, the Brazilian labor ministry, for their assistance in providing the data used in this project. I would also like to thank Santanu Chatterjee, David B. Mustard and John Turner for their guidance throughout the graduate school process.

Finally, thank you to my family. To my parents, who supported my decision to pursue my doctorate and pushed me to get started. And of course, I would not have been able to complete this project without the unending support of my wife, Rebecca. Throughout my time as a graduate student she has supported me in my academic endeavors, tolerated countless late nights and proofread hundreds of pages of text. I am thrilled to tell her, "Let's get started with this life thing."

Contents

List of Figures

List of Tables

Chapter 1

Racial Classification and the Race-Wage Gap in Brazil: Evidence from Administrative and Survey Data

1.1 Introduction and Motivation

This paper serves two purposes. The first is to investigate the nature of racial inequality in Brazil by examining how a worker's race is reported in administrative and survey settings, how these differences affect measures of racial wage gaps, and what conventional earnings decompositions can tell us about reporting wage differences. The second purpose is a direct result of the first: this paper contributes to a literature of cautionary tales about how administrative incentives, or the lack thereof, may alter data and affect economic analysis.

Racial classification is particularly troublesome in settings where racial identity is fluid. In Brazil, most standard classifications classify Brazilians as either *branco* (White or light-skinned), *pardo* (Brown), *preto* (Black or dark-skinned), *amarela* (Yellow), or *Indigena* (Indigenous). These racial classifications exist on a spectrum of skin tones, and are not defined by ethnic background, which is how race is commonly defined in the U.S. and Europe. As a result, the racial classification of a worker depends on how he perceives his own race, how others perceive his race and whose perception prevails. Telles (2002) finds that newspaper survey enumerators and respondents disagree on racial classifications in about one-fifth of cases. Specifically, the third-party enumerators are more likely to perceive workers who are better educated and wealthier as White. Further, there may be economic gain for a worker if he can manipulate others' perception of his race, as shown in Cornwell et al. (2014). Additionally, such ambiguity is to be expected not just in Brazil, with its long history of miscegentation, but in western "post-racial" societies where miscegenation is strong and growing, like the U.S.

In addition to the opportunistic and sociological implications, fluid racial categories present special challenges and opportunities to researchers. Important descriptive research, such as the measurement of race-wage gaps becomes more difficult when race is no longer fixed. These studies are often used to make recommendations regarding policies such as affirmative action and equal employment legislation. However, fluidity in racial classification also provides researchers with the opportunity to experiment with racial classification and to study how varying racial assignments affect measures of economic outcomes.

I investigate racial inequality in Brazil using two sources: household survey data and administrative data. In the household survey, *Pesquisa Nacional por Amostra de Domicilios* (PNAD), race is reported by the individual; in the administrative data, *Relação Anual de Informações Sociais* (RAIS), race is reported by the employer. These programs serve different purposes, fall under the jurisdictions of different agencies and have different collection methods. However, much of the information collected by both PNAD and RAIS overlaps, specifically, both capture worker demographic information and labor market characteristics using equivalent measures and classifications. I use data from eight different waves of RAIS and PNAD to determine the existence, size and composition of a race-wage gap that persists between White and Non-white workers in the Brazil's formal labor market.

I first compare key demographic characteristics of Brazil's formal labor market as reported by RAIS and PNAD. I show that the different collection methods allow for large differences in racial composition and other areas where the information is either subjective or the respondent is ill-informed. Both data sources describe a work force that is mostly White and male, with shares of female and Non-white workers that are increasing over the sample period. However, RAIS and PNAD differ markedly in their depictions of the racial composition of the formal workforce.

I then decompose the race-wage gap for separately for RAIS and PNAD using the method prescribed in Oaxaca (1973) and Blinder (1973). Because of the above discrepancies in racial composition, I am curious to know what affect these differences may have on economic analysis. I find that RAIS data underestimate the size of the race-wage gap, relative to PNAD data, by between 7 and 30 percent, depending on the exact model specified. Further, the relative sizes of wage gap components also vary by data source and specification. Most RAIS specifications suggest very small race-wage gaps that are driven by racial biases or some other mechanism that the data cannot explain. PNAD results suggest larger gaps and that differences in worker characteristics are slightly more important than racial biases in determining the wage gap.

These results demonstrate that the size and composition of the race-wage gap varies systematically with perceived race, either the workplace administrator's or the survey enumerators's. RAIS data paint a somewhat rosier picture, because of racial distortion between the worker and the RAIS administrator and workers' strategic choice to change their race.

PNAD data paint a pessimistic picture, though this may result from the "Whitening" effect of higher incomes and education. Failure to consider one data source in favor of the other may lead to, at best, inaccurate research and, at worst, poor policy reccommendations in an important policy area.

This paper contributes to the ongoing discussion of Brazilian wage discrimination along various socio-economic criteria in general, and race in particular. As a result of RAIS's restricted access, most of the literature focuses on results from PNAD data. Loureiro et al. (2004) study the gender wage gap across the urban-rural wage gap using the 1992 and 1998 waves of PNAD data and find that in addition to substantial wage differentials between urban and rural workers, discrimination by worker gender and race persist. Cortez Reis and Vianna Crespo (2005) use PNAD data from 1987 through 2002 to note that the race gap in Brazil is decreasing over that five year period, most notably for younger workers. Marques Garcia et al. (2009) use the 1996 through 2006 waves to study gender and racial wage gaps in Brazil, and find the race-wage gap is more pronounced than the gender-wage gap and that both wage gaps are in decline over the sample period. Gradín (2009) uses the Oaxaca-Blinder decomposition to develop a causal analysis of poverty among "*preto*," or Black, workers. He finds that observed characteristics, as opposed to racial bias, explain about 90 percent of the non-black race-wage gap. Leite (2005) studies not just the race-wage gap, but the race-education gap as a determinant of the wage gap and finds that some part of the Oaxaca-Blinder component typically associated with racial discrimination is explained by educational inequalities between races. Madalozzo and Martins (2007) provide a dynamic study of the gender-wage gap using PNAD data. Using data from 1981, 1992 and 2004, they find that gender-wage gaps have decreased. Matos and Machado (2006) use the 1987 through 2001 waves to find persistent gender- and race-wage gaps, the latter gap originating primarily from poor education outcomes for black workers. Most recently, Gradín (2014) uses the Oaxaca-Blinder decomposition to study the race-wage gap in Brazil and compare it to the race-wage gaps of the United States and South Africa. He finds that the White-black earnings differential is smallest in the U.S. and largest in South Africa, with the Brazilian earnings differential slightly greater than that for the U.S.

Unlike previous papers, my research focuses on the formal labor market of Brazil. Formal labor markets have clearly defined worker rights and priveleges, including the ability to initiate legal action in the face of discriminatory behavior Gasparini and Tornarolli (2009a). Further, this paper is among the first to use RAIS to study the race-wage gap in Brazil and also to compare how RAIS and PNAD describe the formal labor market of Brazil.

1.2 Data

I first describe the the institutional details of each data source, including purpose, collection method and institutional processing. Then, I examine how each data source describes the racial and gender compositions of the formal labor force. Finally, I document the empirical differences in between the racial and gender compositions produced by RAIS and PNAD. While both data sources provide similar descriptions of formal labor in Brazil at an extremely broad level, they differ significantly when measuring racial composition. These quantitative differences drive important differences in estimating the race-wage gap.

1.2.1 RAIS

I use administrative data from the 2003 through 2010 waves of the *Relação Anual de Informações Sociais* (RAIS), or Annual Social Information Survey. RAIS is an annual administrative survey by the *Ministerio de Trabalho e Emprego* (MTE), Brazil's labor ministry. The MTE use RAIS to administer the *Abono Salarial*, or "Thirteenth Salary," a constitutionally-mandated bonus that is the equivalent to one month's earnings. Unique identifiers for each plant and worker allow the MTE to accurately record the earnings of each worker in every job they had over the year, and the worker to receive a prorated Thirteenth Salary from employers where he or she did not work the full year.

RAIS only covers formal employers and their employees; the Brazilian government defines formality in terms of the presence of a worker contract and contributions to one of two social security accounts, PIS for private sector workers and PASEP for public sector workers. RAIS covers the universe of formal labor, since a firm faces not only fines from the MTE, but also legal action from employees who have not received their bonus, if they do not participate. The actual survey is completed at the plant level by the employer, on behalf of the employee. In smaller plants and firms, the person completing the survey may be the owner or plant manager; in larger organizations, dedicated personnel submit information about the employee, a person they likely have never met. This disconnect between the subject of the survey – the employee – and the respondent of the survey – the plant's agent – may lead to errors in the reporting of particularly subjective information, such as race.

The survey is typically completed by March of the year following the survey year, and questions refer to the plant's and its employees' statuses as of December 31 in the survey year. Once the data are collected, the MTE cleans and prepares the data for their own use in administering the Thirteenth Salary, generating statistics and evaluating policy. Access to the microdata with employer and employee identifiers is restricted, and non-government researchers must obtain permission to use the data.

RAIS collects information on plant and firm characteristics such as industry, size and ownership type and worker characteristics such as age, education level, occupation, race and gender. However, my interest is restricted to those variables that both form the basis of a Mincer-type wage equation and have a counterpart in PNAD. To that end, I include earnings, age, gender, race, education, industry and occupation in my sample. RAIS records four measures of earnings: average monthly earnings, in reais and multiples of the minimum wage, and December earnings, in reais and multiples of the minimum wage. The minimum wage in Brazil is not an hourly pay rate, but the minimum monthly salary an employer must pay to a full-time employee. Interns, trainees and part-time workers may earn less than the minimum wage. For comparison with PNAD data, I use the reported monthly earnings in (nominal) reais.

Race is reported as one of six or seven categories in RAIS; five of these categories are the actual racial or color classifications, while the remainder consists of one or two non-reports, depending on the sample year. The five "identified" categories are Indigenous (*Indigena*), White (*Branco*), Black (*Preto*), Yellow (*Amarela*) and Brown (*Pardo*). The Indigenous category describes someone who is predominantly of Amerindian descent; the White category describes fairer-skinned workers; the Black category describes someone with darker skin; the Yellow category describes workers who are of predominantly East Asian descent and the Brown category describes workers with moderate skin tones.

RAIS allows for non-reports in racial classification. The first non-report category is Not Identified (*Não Identificado*), indicating that the race question was not answered for that worker. This category exists in every wave of RAIS in the sample. The second non-report category is Ignored (*Ignorado*), which indicates that the question was not merely skipped, but that the respondent actively chose to report the worker's race as not available or "known". From 2006 onward, when the Ignored option is available, the share of workers Not Identified is relatively smaller compared to 2003-2005.

1.2.2 PNAD

I also use data from the 2003 through 2009 and 2011 waves of the *Pesquisa Nacional por Amostra de Domicilios* (PNAD), or National Household Sample Survey. PNAD is a yearly survey of households and their occupants administered by the *Instituto Brasileiro de Geografia e Estatística* (IBGE), Brazil's national statistics institute. PNAD provides a statistical picture of Brazil in off-census years and reflects information gathered from households about their occupants and the physical space where they live. To get a complete picture of Brazilian life, PNAD records characteristics including education, labor, income, migration, fertility and health. The result is a pool of data that the IBGE can use to aid in the formation and evaluation of Brazilian policy.

The survey itself is completed by an enumerator/interviewer and is quite lengthy. For example, there are about 30 questions regarding the actual domicile and 190 questions regarding each person in the household. Since the 2007 wave, the interview has conducted the survey with a PDA or other electronic device. The questions refer to the status of the domicile and its residence during the survey week, which is typically the second or third week of September, depending on the year. IBGE workers interview some 300,000 people every year.

When the survey is complete, IBGE contructs additional variables, such as total household income, which it uses to produce descriptive statistics, assigns weights to each person and household and makes the microdata available for public use, usually by March of the following year.

In parallel with the RAIS sample, of the more than 200 variables present in PNAD, I am interested in those that might appear in a typical Mincer-style wage regression or define a person as a worker in the formal workforce. In addition to age, gender, race and earnings, I include education, industry and occupation. The way education levels are recorded changes over sample years, and some education variables require an intimate knowledge of Brazil's educational terminology. For consistency purposes, I use the number of years of schooling completed, which are reported in every sample year. I then aggregate these 13 values into seven categories so as to make them comparable with the categories from RAIS.

Since PNAD records information from all workers, formal or not, being able to distinguish formal workers from their informal counterparts is critical. PNAD records whether or not a worker contributes to the relevant social security organization, either PIS or PASEP. I use this variable as an indicator of the worker's attachment to the formal labor market. This definition of formality corresponds with the "legal definition" of formality from Gasparini and Tornarolli (2009a).

Unlike RAIS, PNAD does not allow for non-reports of a worker's race. The presence of an enumerator to aid in completion ensures that race question is almost always answered; non-reports are never more than 0.001 percent of the racial composition in PNAD.

1.2.3 Sample Selection

Since the subjects in RAIS should be represented by a subsample of PNAD, sample selection primarily involves constructing the subsample of PNAD which mirrors the sample taken from RAIS. Sample selection is based on five criteria: formality, age, full-time status, minimum wage and primary occupation.

Formal workforce participation in Brazil can be defined in one of two ways. The first is if the worker contributes to social security; the second is if the worker has a formal contract, the most common of which is the *Consolidação das Leis do Trabalho* (CLT)(Gasparini and Tornarolli 2009a); either definition would indicate a worker who is recorded in RAIS. PNAD indicates in every year whether or not a worker contributes to social security, and I use this as my indicator for formality.[1] Attachment to the formal workforce is the defining characteristic of workers in RAIS; all formal workers, and only formal workers, are included in the survey. Thus, no restrictions are necessary for RAIS in this instance.

I restrict my sample to workers between the ages of 20 and 65. These are the prime working years, when a worker is typically finished with his or her schooling and in the labor force full time. I also exclude workers who might not be completely attached to the formal labor sector. A worker who is employed part time may likely still be in school or working multiple jobs which may or may not be formal jobs themselves. For the same reason I include

[1] Alternative specifications, such as whether the worker has a formal contract, do not appear significantly different from this specification.

only workers earning the at least minimum wage. The minimum wage in Brazil is defined as the minimum *monthly* earnings a firm must pay its full-time employees.

Finally, I limit the sample to only one job for each worker, which I define as the primary job. The primary job is referred to explicitly in PNAD, and corresponds to the job in which the worker has the highest salary. I use this definition when selecting jobs from RAIS. When a worker has multiple jobs in a year, I select the highest-paying job as the primary job and include that job for my analysis.

To ensure comparability with RAIS, I weight individuals PNAD observations using weights provided by IBGE.

1.2.4 Descriptive Statistics and Comparison

RAIS

Table II provides basic descriptive statistics for the Brazilian formal labor force over the 2003-10 period generated from RAIS. The sample is restricted to full-time workers earning at least the minimum wage between the ages of 20 and 65. I focus here on race and gender.[2]

Race and Gender

The racial composition of the formal labor force, for any given year, is largely either White, Brown or Black and together comprise about 80 percent of the sample. White workers make up the largest segment of the formal labor force, although they are decreasing as a share of the labor force over the sample period, from about 60.8 percent in 2003 to 49.4 percent in 2010. The largest single decrease occurs between 2005 and 2006 (from 60.7 percent to 52.5 percent), when the "Ignored" category is included in the survey. Of the three major racial classifications, Black workers comprise the smallest share of the labor force. Throughout the sample period, the share of Black workers remains relatively stable, with a maximum share

[2] A description of table II statistics can be found in appendix B.1.

of 4.7 percent in 2005 and a minimum share of 4.4 percent in 2006. Brown workers make up the largest of the Non-white categories, between 20 and 25 percent of the formal workforce. Between 2003 and 2005, and again from 2006 through 2010, the share of Brown workers grows from 23.1 percent to 25.2 percent and then from 22.3 percent to 25.1 percent. The sharp decrease in the Brown share of the formal labor force between 2005 and 2006 results from a large increase in workers who do not report their race beginning in 2006. As a whole, the share of the three largest racial categories is decreasing over the sample period.

Workers without a reported race are a very large portion of the formal workforce in RAIS, as much as 20 percent. As mentioned earlier, non-reports are classified in one of two ways, Not Identified or Ignored. Not Identified workers are those whose race was not recorded at all in the RAIS survey. These workers make up about 10 percent of the workforce from 2003 through 2005.

The Ignored racial classification represents an active response by the RAIS respondent to not report the worker's race. The introduction of this category is associated with the largest changes in shares of other classifications.Starting in 2006, with the inclusion of the Ignored category, the share of workers classified as Not Identified becomes much smaller, about 2.5 percent. However, the share of workers who are Not Identified grows from 2006 through 2010 to about 4.6 percent of the workforce. The presence of the Ignored category coincides with declines in the shares of reported racial categories. From 2006 through 2010, the share of Ignored-race workers decreases from 17.3 percent to 15.7 percent. However, this decrease in Ignored workers does not correspond with an increase in reported (White, Black, Brown, etc.) workers, but an increase in the share of workers characterized as Not Identified. Overall, RAIS describes a formal workforce that is mostly White, with an increasing minority presence over the sample period. However, the presence of workers with non-reported race is large and significant.

The male share of the formal workforce decreases two percentage points over the sample period from about 61.7 percent in 2003 to 59.4 percent in 2010. This decrease in male share of the formal workforce reflects a broader trend of increased female participation in the formal labor market of Latin America as a whole, and Brazil in particular, over the sample period (Piras and Ripani 2005; Agüero and Marks 2008).

PNAD

Table III provides the basic descriptive statistics for the Brazilian formal labor force over the 2003-11 period generated from PNAD. As with the RAIS sample, this sample is restricted to full-time workers earning at least the minimum wage between the ages of 20 and 65. As with RAIS, I focus on race and gender.[3]

Race and Gender

Like RAIS, the racial composition of the formal labor force in PNAD is largely White, Brown or Black, and taken together comprise almost 90 percent of the sample. White workers form the largest racial group, though their share of the workforce is decreasing over the sample period, Brown workers are the largest Non-white group, and the share of workers who are Black is roughly one-fifth to one-fourth of the share of Brown workers. However, there is a shift in the composition from White workers to Brown workers.

However, there are some key differences between how RAIS and PNAD describe the formal labor markets racial composition. First, in PNAD, Indigenous workers' share of the labor force is increasing over the sample period; indeed, it more than doubles from 0.14 percent to 0.3 percent of the labor force. Second, Black workers also see their share of the labor force increasing over the sample period from 6.1 percent in 2003 to 8.8 percent in 2011. Third, Brown workers also increase as a share of the labor force from 31.6 percent in 2003 to

[3]A description of the table III statistics can be found in appendix B.1.

36.1 percent in 2011. These increasing trends in minority categories of PNAD are in sharp contrast to the decreases seen in RAIS.

A large reason, if not the single most important reason, for the discrepancy in racial composition trends between RAIS and PNAD is the lack of non-reported race in PNAD. In PNAD less than 0.01 percent of respondents do not report race and are classified as Not Identified. Further, the PNAD survey does not allow for an Ignored option in the race question.

As seen in RAIS, the gender composition of the labor force according to PNAD is mostly male, though the male share of the workforce is decreasing over time, reflecting the broader trends mentioned earlier. In PNAD, the male share decreases monotonically from 58.7 percent in 2003 to 57.3 percent in 2011. However, compared to RAIS, men comprise a smaller share of the formal workforce in every PNAD year. I defer discussion of these differences to the next sub-section.

Differences between RAIS and PNAD

Table IV provides the percentage-point differences in reported race, gender, industry and occupation compositions and the percent differences for mean age and income between RAIS and PNAD for each year, as well as the average differences. I again focus here on race and gender.[4]

Race and Gender

The key distinction between the RAIS and PNAD racial compositions is the inclusion of the Not Identified and Ignored categories. Prior to the introduction of the Ignored category (2003-5), RAIS is 8 to 9 percentage points "less Brown" and 1.5 to 2.2 percentage-points "less Black" than PNAD. White shares for this portion of the sample are roughly comparable at about 60 percent of the workforce. In these sample years the share of RAIS workers classified

[4] A full description of the statistics in table IV can be found in B.1.

as Not Identified is almost completely accounted for by the differences in the Black and Brown portions of the racial composition.

With the introduction of the Ignored category in 2006, the shares of White, Black and Brown workers in RAIS all fall relative to PNAD. From 2006 on, RAIS is also "less White" than PNAD. Starting in 2006, the differences in the Black, Brown and White shares, as a sum, are mostly offset entirely by the sum of the two non-report categories, 5.7 percentage-points for Not Identified and 10.3 percentage-points for Ignored. One reason these big differences may be the subjective nature of race in Brazil. The information available to RAIS respondents regarding worker race differs from the information available to PNAD respondents (third party versus first party information).

The difference between the male shares of the workforce varies between 2.9 and 1.9 percentage-points with a mean of 2.3 percentage-points, suggesting that RAIS is slightly more male than PNAD. Since gender is less likely to be subjective, or at least is much less subjective than race, it seems unlikely that the difference in gender composition is a result of differing information sets between respondents. Further, because the difference between RAIS and PNAD in male share always suggests that the male share of RAIS is about 2 percentage points greater than that of PNAD, it does not seem that this is a result of measurement error. This difference may be a result of systematic error in how either RAIS or PNAD records gender, or may be a result of female workers being omitted from the formal category of RAIS. Regardless, the consistency of the difference over time, and relatively small size – compared to the differences in race – suggests that the sample does represent an attempt to capture the portion of PNAD that most closely resembles RAIS. Further, these differences serve as a benchmark of differences between RAIS and PNAD for an objective variable.

Compared to this benchmark, the differences in racial composition between RAIS and PNAD indicate a discrepancy that goes beyond measurement error, and indicates something more systematic. Because these differences do not appear to be random, there remains the question of how measures of race-related wage inequality calculated from RAIS data may differ from those calculated from PNAD data. I set up the empirical framework to address this question in the following section.

1.3 Decomposing the Racial Earnings Gap

I now turn to the analysis of the racial earnings gap. I calculate the race-earnings gap produced by each data set and decompose these gaps into their characteristic and price components using the well-known method of Oaxaca (1973) and Blinder (1973). I will use the decomposition ot perform a counterfactual analysis of how much one group would earn if they were paid in the same manner as the other group, in this case.

In the current context, the Oaxaca-Blinder earnings gap can be derived as follows. First let Y_r be the monthly earnings for race group r, either White (W) or Non-white (NW). Then, express log earnings in a standard human-capital regression of the form

$$\ln Y_r = X_r \beta_r + \varepsilon_r, \tag{1.1}$$

where X_r contains worker characteristics, namely age, age squared, education, industry and occupation. Applying least-squares to (1.1) yields predicted earningsfor each race group.

$$\widehat{\ln Y}_r = X_r \hat{\beta}_r. \tag{1.2}$$

Define the race-wage gap, G as the difference in predicted earnings:

$$G = \widehat{\ln Y}_{NW} - \widehat{\ln Y}_W. \tag{1.3}$$

Substituting (1.2) into the right hand side of (1.3) yields

$$G = X_{NW}\hat{\beta}_{NW} - X_W\hat{\beta}_W. \tag{1.4}$$

Finally, adding and subtracting the product of the Non-white vector of characteristics with the White coefficient estimates produces the complete Oaxaca-Blinder decomposition.

$$G = \underbrace{X_{NW}(\hat{\beta}_{NW} - \beta_W)}_{\text{coefficient}} + \underbrace{(X_{NW} - X_W)\hat{\beta}_W}_{\text{characteristic}}. \tag{1.5}$$

The first term on the right-hand side of (1.5) is comprised of the difference between the estimated coefficients between Non-white and White workers evaluated at the characteristics of Non-white workers. The "coefficient component" captures how much Non-white workers in the sample would have earned if they had been paid the same for their market characteristics as their White counterparts, and is commonly interpreted as the racial bias component.

The second term in (1.5) consists of the between-group differences in characteristics weighted by White coefficient estimates. The "characteristic" component of the wage gap shows how much Non-white workers lose relative to White workers as a result of different education levels or working in different industries and occupations.

Of course, an alternative method for decomposing the race-wage gap would be to instead add and subtract the product of the White vector of characteristics with the Non-white coefficient estimates, producing

$$G = \underbrace{(X_{NW} - X_W)\hat{\beta}_{NW}}_{\text{characteristic 2}} + \underbrace{(\hat{\beta}_{NW} - \beta_W)X_W}_{\text{coefficient 2}}. \tag{1.6}$$

Here, the first term is an alternative characteristic component which is the difference between Non-white and White workers' observed characteristics evaluated at the Non-white price. The second term is an alternative price component capturing the difference between the estimated coefficients between Non-white and White workers evaluated at the characteristics of White workers.

Following Jones and Kelley (1984) and Daymont and Andrisani (1984) the usual Oaxaca-Blinder decomposition can be re-written as follows. Adding and subtracting the price component from (1.6), I re-write (1.5) as

$$G = \underbrace{X_{NW}(\hat{\beta}_{NW} - \beta_W)}_{\text{coefficient}} - \underbrace{(\hat{\beta}_{NW} - \beta_W)X_W}_{\text{coefficient 2}} + \underbrace{(X_{NW} - X_W)\hat{\beta}_W}_{\text{characteristic}} + \underbrace{(\hat{\beta}_{NW} - \beta_W)X_W}_{\text{coefficient 2}}.$$

Rearranging terms and factoring reveals the following three-part decomposition of the race-wage gap:

$$G = \underbrace{(X_{NW} - X_W)\hat{\beta}_W}_{\text{characteristic}} + \underbrace{(\hat{\beta}_{NW} - \beta_W)X_W}_{\text{coefficient 2}} \underbrace{(X_{NW} - X_W)(\widehat{\beta_{NW}} - \beta\hat{W})}_{\text{interaction}}. \tag{1.7}$$

In this three-fold decomposition the first term provides the difference in earnings resulting from differing characteristics of Non-white and White workers evaluated as if both groups were paid White prices. The second term captures the difference between Non-white and White prices evaluated with the characteristics of White workers. The third term is an interaction term. Algebraically, it is equivalent to the the price effect of (1.5) less the price effect of (1.6). More intuitively, the interaction term accounts for the simultaneous differences in characteristics and coefficients that persist between the two groups. Increased magnitudes in interaction effects indicate that the differences in characteristics and prices are closely related, and also suggest significant differences between the two counterfactuals of White workers paid as Non-White workers ($X_W\hat{\beta}_{NW}$) and Non-White workers paid as White workers ($X_{NW}\hat{\beta}_W$).

The relative sizes of the wage-gap components have important policy implications. A wage gap that is driven by its price component indicates discriminatory behavior on the part of the firm, and anti-discriminatory policies can be recommended. However, if the characteristic component appears to drive the wage gap, policy recommendations may focus more on continuted education and vocational training to improve the labor market characteristics of the disadvantaged group.

When comparing Oaxaca-Blinder results from RAIS and PNAD data, it is not just differences in the magnitude of the wage gap that matter, but also the relative magnitudes of each gap's price and characteristic components. Differences in total magnitude may suggest an overstatement or understatement of how Non-white workers are paid relative to White workers; differences in the gap components may result in different policy recommendations.

1.4 Results

For each data set, I estimate (1.1) twice, first using the Non-white observations, and then using the White observations. The regressions yield fitted values Non-white workers, $\widehat{\ln Y}_{NW}$, and for White workers, $\widehat{\ln Y}_{W}$ and as well as Non-white and White coefficient estimates, $\hat{\beta}_{NW}$ and $\hat{\beta}_{W}$. Using these results, I calculate and decompose the race-earnings gap according to (1.3) and (1.7).

For RAIS, the Not Identified and Ignored categories provide a unique opportunity to examine the race-wage gap. Thus, I consider three different cases. The first (A) includes all workers without a reported race as Non-white. The second (B) includes all workers without a reported race as White. In the third case (C), I use a worker's modal race; I exploit the panel nature of the data to identify the race of non-reported workers in their previous or future jobs. If a worker has more than one reported race, I use the worker's modal race.[5]

Overall, case A generates estimated gaps that are the smallest, and have the greatest contrast with the PNAD estimates. Model B generates estimated gaps that are closest to PNAD estimation, and case C – with imputed minority statuses – estimate gaps that are

[5]The estimation results for 2004-8 and 2010 can be found in appendix **??**.

slightly larger than those generated in case A, but smaller than those in case B, providing a rough middle ground. Further, these results are robust to relaxing the restrictions of full-time employment and earnings greater-than or equal-to the minimum wage. For any given year, RAIS estimates for the race-wage gap are smaller in every case specification than the PNAD estimate. Additionally, all RAIS gap estimates place much more emphasis on the unexplained component of the race-wage gap, suggesting that these smaller gaps are the result of racial discrimination. Even the case B gap decompositions - the case closest to PNAD in gap measurement - suggests that racial discrimination plays a greater role in the presence of the race-wage gap than worker characteristics do. PNAD estimates in any given year place greater emphasis on the characteristic component of the wage gap, though both components of the wage gap are rather large.

Tables V through XII report the results from the Oaxaca-Blinder decompositions for each sample year. In the discussion that follows, I distinguish the pre-2006 RAIS data – when non-reports of race are classified only as "Not Identified" – and the post-2006 data – when the "Ignored" category is added to the racial classification options. Recall that prior to the inclusion of the Ignored category, the share of workers classified as Not Identified roughly corresponded to the differences of Black and Brown workers between RAIS and PNAD. Beginning in 2006, the inclusion of the Ignored category coincides with a decrease in the shares of all three major racial categories, and is roughly equivalent to the differences of White, Black and Brown workers between the two data sources.

Consider first the results from 2003 presented in table V. For each specification, the table reports $\widehat{\ln Y}_{NW}$, the estimated monthly log earnings for Non-white workers, $\widehat{\ln Y}_{W}$, the estimated monthly earnings for White workers, the race-earnings gap, $\widehat{\ln Y}_{NW} - \widehat{\ln Y}_{W}$, and the components of the earnings gap. The decomposition consists of three components, the characteristic component, or portion of the gap resulting from differences in worker characteristics; the coefficient or price component, which is the portion of the gap left unexplained by worker characteristics; the interaction which is an augmentation of the effects, caused by the interaction of the characteristic and coefficient components of the wage gap.

The PNAD estimate of log earnings for Non-White workers in 2003 is about 6.2, or 492 R$, while estimated log earnings for White workers are 6.6, or roughly 735 R$. The estimated earnings gap is almost -0.4 log points, meaning Non-White workers earn about 40 percent less than their White counterparts. Decomposing this 40 percent gap into its 3 components reveals a large characteristics component of about -0.23 log points, meaning Non-White workers lose about 23 percent of their income relative to White workers as a result of less-favorable labor market characteristics. The coefficient component is about -0.18 log points, meaning that in addition to poor labor market characteristics, Non-White workers lose 18 percent of their earnings as a result of not paid how White workers are paid. The ratio between the characteristic and coefficient components is roughly 11:9, which suggests that worker characteristics are slightly more important in determining the race-wage gap. The interaction component is small, just 0.01 log points, suggesting that "changing" a Non-White worker's race and characteristics to whose of a White worker simultaneously does not affect the size of the earnings gap any more than the two components taken separately do.

RAIS estimates reflect the disconnect between respondent and subject in the administrative data. In case A, workers who are racially classified as Not Identified are included in the Non-White group, and the differences between case A and the PNAD case are drastic. In this case, estimated log earnings for Non-White workers are about 6.4 log points, or roughly 600 R$, which is significantly higher than the estimated Non-White earnings from PNAD. The case A estimated log earnings for White workers are roughly 6.47 log points, about 645 R$, which is slightly less than the estimated White earnings from PNAD. The estimated earnings gap in case A is therefore much smaller than in the PNAD case, about 8 compared to almost 40 percent in the survey data.

Decomposing the race-wage gap for case A reveals a very small characteristic component of just -0.015 and a larger coefficient component of about -0.1. The resulting characteristic-coefficient ratio is approximately 1:5, meaning that the much smaller wage gap of case A is largely driven by the coefficient – or racial – component. The interaction term for case A is about 0.03, and indicates a slightly stronger relationship between the way Non-White workers are paid and their characteristics.

Case B includes Not Identified workers as White workers and is the opposite of case A. In case B estimated log earnings for Non-White workers are lower than in A, about 6.25 log points or 515 R$, and estimated log earnings for White workers are greater than in case A at 6.5 log points or about 680 R$. The resulting estimated earnings gap is larger than that for case A, but still smaller than for PNAD, at about 27 percent. Decomposing the wage gap for case B reveals that the characteristic and coefficient components are almost equal at -0.13 and -0.16. The ratio between these components suggests that racial bias plays a slightly more important role in determining the race-wage gap than worker characteristics do. Again, the interaction component is relatively small, less than 0.02.

Finally, case C assigns to each Not Identified worker his or her modal race from other jobs where the worker's race was identified in RAIS. In general, case C is more like A than B, though the magnitudes of estimated earnings, earnings gap, gap components and component ratio are between those of cases A and B. Non-White earnings are estimated at about 6.37 log points, slightly less than those of case A, and estimated White earnings are estimated at 6.49 log points, slightly more than those of case A. The resulting wage gap estimate is naturally slightly more than that of case A, with Non-White workers earning about 12 percent less than White workers in case C, a 4 percentage-point increase over case A.

Decomposing the estimated gap from case C similarly reveals magnitudes that are slightly larger than those for case A, except for the interaction component. The estimated characteristic component is about -0.04, and the estimated coefficient component is about -0.11;

the resulting characteristic-coefficient ratio of 4:11 implies that racial bias is more important in determining the race-wage gap in case C, as it does in case A. The interaction term is about 0.03, again suggesting some relationship between the characteristic and coefficient components.

Tables VI and VII report the estimation results for 2004 and 2005. The qualitative and quantitative patterns produced from the 2003 data appear in 2004 and 2005 as well. In both years, PNAD provides the largest estimate of the race-wage gap, about 40 percent, and decomposing the gap suggests that characteristics are slightly more important than racial bias in determining the size of the gap. For RAIS estimates, case A provides the smallest estimated race-wage gap, case B the largest and case C provides estimates that are between those of A and B. All RAIS cases estimate race-wage gaps that are smaller than the PNAD estimated race-wage gaps, and place more emphasis on the coefficient component while the PNAD estimates place more emphasis on the characteristic component of the wage gap.

Starting in 2006, racial non-reports in RAIS may be categorized as either Not Identified or Ignored. Recall from the discussion of table IV that unlike the Not Identified category, which appears to consist mostly of Black and Brown workers, the Ignored category appears to include many White workers as well. The presence of the Ignored category and the effect it has on race reporting affects the calculation, and interpretation, of the race-wage gap.

Turn to the results from 2006 presented in table VIII. The estimates produced by PNAD are qualitatively and quantitatively similar to those from 2003-05. Non-White log earnings estimates for PNAD are about 6.5 log points while estimated log earning for White workers are approximately 6.8 log points; the resulting estimated race-wage gap is about -0.37, suggesting that Non-White workers earn about 37 percent less than White workers. Decomposing the race-wage gap reveals that Non-White workers lose about 22 percent of their earnings relative to White workers from the characteristic component, and about 16 percent from racial bias. And as in earlier years, the characteristic-coefficient component ratio, about

11:8, emphasizes the importance of worker characteristics over racial bias in determining the race-wage gap. The interaction component remains small, about 0.01, indicating that the combined effect of simultaneously changing a Non-White worker's characteristics to those of a White worker and paying him or her as a White person is roughly equal to the effects of each component.

Again, case A treats both Ignored and Not Identified workers as Non-White. For 2006, case A of the wage estimates provides the highest Non-White and lowest White earnings of all the specifications, as seen in earlier samples. However, here the estimated log earning of Non-White and White workers are the same at 6.7 log points, suggesting that there is no race-wage gap for this case in 2006. The decomposition of this "gap" reveals a very small, positive characteristic component of about 0.01 which suggests that Non-White workers actually have slightly better labor market characteristics. Surprisingly, the coefficient component is quite large and negative, about -0.10, which would suggest that Non-White workers earn considerably less than their White counterparts as a result of racial bias. However, the interaction term is important. It is large, almost 0.09 log points, and almost entirely negates the coefficient component. This large interaction term means that the chosen counterfactual – $X_W\hat{\beta}_{NW}$ or $X_{NW}\hat{\beta}_W$ – affects the measurement of the gap. It also means that paying a Non-White worker as a White worker and simultaneously endowing the Non-White worker with White characteristics has a joint effect above and beyond the combined sum of changing just one of those conditions.

Case B treats both Ignored and Not Identified workers as White, and more closely resembles the PNAD estimation results than case A does. Estimated earnings for Non-White workers are much smaller, about 21 percent, in case B than in case A. At the same time, estimated earnings for White workers in case B are about 8 percent greater than in case A. Together, the estimated earnings produce an estimated earnings gap of nearly 30 percent, which is much larger than the nearly non-existent gap of case A and approaches the 37 percent wage gap produced by PNAD. Decomposing the estimated wage gap reveals large characteristic and coefficient components, about -0.18 and -0.17 log points. The resulting characteristic-coefficient component ratio is almost 1:1 suggesting that worker characteristics both play large, almost equal roles in determining the race-wage gap. The interaction component for case B is larger than in the pre-2006 samples, about 0.06. This again suggests that the choice of counterfactual may affect component measurement, and that the coefficient and characteristic components react differently when changed simultaneously.

The final case, C, assigns Not Identified and Ignored workers their modal race from either past or future jobs where their race was reported. The general pattern established in the pre-2006 samples continues here; case C yields earnings and earnings gap estimates that lie between cases A and B in terms of magnitude, though they more closely resemble case A than B. This closeness to case A is first seen in the reported log earnings estimates for Non-White workers (6.65 log points) which are just 5 percent less than those of case A, and about 15 percent greater than those from case B. Log earnings estimates for White workers in case C (6.74 log points) are just 4 percent greater than those of case A and about 4 percent less than those from case B. The resulting race-wage gap estimate is about 9 percent, which about one-third the size of the case B gap.

The estimated wage gap components of case C resemble those from pre-2006. The characteristic effect is about -2 percent while the coefficient effect is about 5 times as large at -10 percent; this 1:5 characteristic-coefficient component ratio again heavily implicates racial bias as the driving force in the race-wage gap. The interaction term is smaller than both cases A and B at 0.03, but still larger than the characteristic component; changing characteristics and race simultaneously results in a change in wages that is smaller than the sum of the characteristic and coefficient components.

Tables IX through XII present the results of this exercise for the 2007-10 sample years. Overall, the results are qualitatively and quantitatively similar to those for 2006. PNAD estimates of the race-wage gap are the largest of all the specifications, around 35 percent. Decomposing the PNAD wage gaps reveal that characteristic components play a more important role in determining the gap than coefficient components do. Case A provides the smallest estimates of the race-wage gap, around 2 percent in the post-2006 samples. As in 2006, the characteristic component is very small while the coefficient component is quite large; however this large coefficient component is largely negated by a large interaction component. Case B most closely resemebles the PNAD case, though the estimated wage gap in case B is consistently smaller than the PNAD estimate by 7 to 8 percentage points. The ratio of characteristic and coefficient components for case B is close to 1:1. Lastly, case C represents a middle-ground between cases A and B, with an estimated race wage-gap near 9 percent and gap components that imply racial bias is more important in determining the race-wage gap than worker characteristics.

The administrative RAIS data, in which a disconnect exists between respondent and subject, overall provide smaller measures of racial inequality than the survey PNAD data. RAIS data do not report race for each worker in the formal sector; indeed they allow for respondents to actively decline to report a worker's race beginning in 2006. As a result, many workers are not assigned to a racial category. This ambiguity allows me the opportunity to assign these workers to different racial categories in 3 cases: one in which I classify them as Non-White (A), a second in which I classify them as White (B) and a final case in which I assign them to a category based upon their racial classifications in other jobs (C). Cases A and B produce the smallest and largest wage gaps from RAIS data, and do so in every sample year, indicating that the workers with no assigned race are not a random group of workers, but high earning workers in every year. Case C exists between the first two, though it more closely resembles A than B. Case B most closely resembles PNAD, in terms of wage gap magnitude and composition, but would imply drastic racial differences between the two data sources. Overall, this exercise provides a range of possible race-wage gap magnitudes and compositions while serving as a reminder of how administrative and survey data may differ in important ways despite many qualitative similarities.

1.5 Conclusion

In this paper, I use two different data sources – RAIS and PNAD – to explore the race-wage gap of Brazil's formal employment sector. Not only do I show that the race-wage gap in Brazil exists, I document that the size and composition of the wage gap vary greatly depending on the data source.

The differences in gap measurement stem from intrinsic properties of the RAIS and PNAD data and the subjective nature of race in Brazilian culture. RAIS is an administrative survey, and the subject of the survey and the respondent to the survey are almost always two different people. This disconnect between respondent and subject, combined with the subjective nature of race in Brazil, mean that the potential for mismeasurement in the racial variable is great. Additionally, RAIS allows for non-responses to the race question. In every year, the respondent has the option of skipping the race question, yielding a value of "Not Identified," and, beginning in 2006, the respondent may actively elect to classify the worker's race as "Ignored." In contrast, PNAD data originate from a survey where the subject and the respondent are one in the same; as a result the respondent has perfect information regarding how he classifies himself racially. Further, PNAD does not allow the respondent to omit race, either by passively skipping the question or actively answering the race question with an answer of "Ignored." Since PNAD is completed in an interview with an enumerator, an answer to the race question is ensured.

Using the Oaxaca-Blinder decomposition method, I show that the differences between in RAIS and PNAD in data collection methods and question enforcement drive large differences between RAIS and PNAD estimates of the race-wage gap for Brazil's formal labor market. While both data sources provide evidence of wage gaps in every sample year, RAIS estimates of the wage gap are as much as 30 percent less than PNAD estimates in the same year, depending on assumptions regarding workers without a reported race. Given the political and social importance of earnings gaps as indicators of socio-economic inequality, such a broad discrepancy in gap measurement provides for a very uncertain description of racial inequality among Brazil's formal labor force.

Decomposing the estimated race-wage gaps into characteristic and coefficient components reveals further differences between RAIS and PNAD estimates. RAIS estimates, again depending on the specification, suggest that the coefficient component, commonly interpreted as racial bias, are the driving force behind these smaller wage gaps. However, PNAD estimates suggest that the characteristic component, the portion of the gap resulting from unfavorable observable worker characteristics, are more important than racial bias in determining larger wage gaps. The policy implications derived from each data source are therefore very different: RAIS data imply that work needs to be done to eliminate the small amount of racial bias that exists; PNAD data suggest much more work needs to be done to eliminate racial bias and still more work to improve the job-market characteristics of Non-white formal sector workers.

While this paper serves as an example of the importance of data generation and survey construction and how these processes may affect analysis, it also contributes to a growing literature on the persistance of income differences between racial groups in Brazil. Instead of providing a concrete answer, however, I have provided a range of values which reflect the subjective nature of race in Brazil.

Chapter 2

Racial Identity and the Workplace: Evidence from Job Changers in Brazil

2.1 Introduction

In the US and Latin America, racial identity intersects every important dimension of life. Membership in certain racial categories may lead to greater opportunity; membership in others may be a handicap. Historically, Whites have benefited from greater access to education and fewer barriers in the labor market, while nonWhites have suffered from restricted access and greater barriers. Over the last thirty years, persistent racial disparities in educational attainment and inequality in labor-market outcomes have led to a variety of interventions (e.g. preferential admissions in education and affirmative action in hiring) designed to level the playing field. Despite these efforts, gaps between Whites and nonWhites in these areas remain.

Students in college may partially offset the intended effects of preferential admissions by becoming more steadfast in their racial identity. Workers may adjust their effort to satisfy the biases of their bosses. If it were possible, an alternative to these responses would be for an individual to change his racial identity altogether, at least as it is perceived by relevent decision-makers – employers or admissions committees. As Akerlof and Kranton (2000) argue, identity (racial or otherwise) affects both the payoffs of a person's own choices and the choices of others. In this paper, I present evidence that workers manipulate perceptions of their racial identity in response to labor market incentives. To do so, I exploit longitudinal, employer-employee matched data from Brazil in which demographic characteristics, including race, are reported by the employer rather than by the individual. The data come from the Annual Social Information Survey (RAIS), which is managed by the Brazilian labor ministry and completed at every registered place of employment. Brazilian employers are under an administrative requirement to complete the RAIS survey, including documenting the demographics of its workforce, but are not burdened with explicit mandates that accompany equal-opportunity laws or affirmative-action policies of the type common in the US.

I observe that approximately 30 percent of job changes are associated with a change in the employer's report of race. In the US, where racial categories are more strictly defined, either through heredity or ethnicity, there is little room for taking on a new racial identity.[1] In Brazil, however, the overwhelming majority of the population is of mixed African and European descent, tying racial identity more closely to a person's skin color than his inherited status or ethnic background. The reliance on skin color for classification creates more flexibility in racial identity and opens up the possibility of identity changes.[2]

In the Brazilian context, workers seemingly have the opportunity to manipulate employers' perceptions of race. As I show, they also have the incentive to do so. Despite the subjective nature of racial categories in Brazil, there is still a considerable amount of racial

[1] Saperstein and Penner (2012) show that there may be some scope for racial mobility in the U.S. Using data from the NLSY, they document non-random changes in race in which high-status people tend to reclassify as White and low-status people tend to reclassify as nonWhite.

[2] This flexibility is illustrated in the story of identical twin brothers, Alan and Alex Teixeira da Cunha, who applied to Universidade de Brasilia, seeking admission under newly instituted quotas for Black students. Separate admission counselors admitted one brother under affirmative action as Black; the other was denied admission, as his counselor considered him to be White (Marotto, 2007). Given a similar set of circumstances, Brazilian workers may provide employers with information about race that would increase their chances of employment and their income.

inequality in wages favoring White individuals. Taken together, these facts suggest that if workers change race to gain labor market advantage, I should observe that workers increase earnings when changing race to White, and lose earnings when changing race from White. Focusing on workers who change jobs in 2010, I find that among Brown or Black workers, those who change race to White enjoy a 2 percent increase in monthly earnings. Among White workers who change jobs, those who change race to Brown or Black experience a 3 percent decrease in monthly earnings.

I also find that race change is associated with movements to more segregated plants. Workers with higher levels of education and rates of earnings growth are more likely to change from 'Brown' to 'White' when changing jobs. However, individual characteristics account for no more than three percent of the variation in race change. In contrast, plant effects account for 60 percent. 'Whitening' is associated with moving to Whiter, smaller, and lower-paying plants.

2.2 Race in Brazil

My research question is whether individuals manipulate perceived racial identity to improve job-search outcomes. In Brazil, workers have both the means to manipulate perceived race and the motive to do so. The nature of racial identity in Brazil and the way race is measured is quite different from the US. According to the standard classification, individuals can be White (light skin tone), Brown (medium skin tone), Black (dark skin tone), Yellow (East Asian descent), or Indigenous (Amerindian descent). Brazilians perceive race as positions along a spectrum rather than as membership in a fixed set of biological groups. Individuals report membership in these categories in official population statistics. In RAIS, employers use the same categories to classify each of their employees.

Because race is strongly associated with skin tone, there can be considerable ambiguity regarding whether a given individual is light-skinned versus brown-skinned, and brown-skinned versus dark-skinned. That such ambiguity presents scope for mis-perception and manipulation is not academic speculation. Telles (2002) finds survey enumerators and respondents disagree on racial classification in approximately 20 percent of cases. These disagreements cut both in the direction of “lightening” and “darkening”, and are systematically associated with socio-economic status. Enumerators are more likely to perceive highly-educated and wealthier individuals as White when they self-report as Brown.

The notion of race embedded in these categories may be unfamiliar to those used to thinking about race and discrimination in the US context. Historically, the colonial powers in Brazil pursued policies of active miscegenation, quite different than the racial purity and “one drop” policies throughout the North American colonies. Brazil’s history of race relations was not dominated by a myth of white racial superiority in the same way as the US These historical differences resulted in a very different construction of the social meaning of race.

While the Brazilian notion of race provides the means for individuals to manipulate racial identity, it does not constitute a motivation. If there is no systematic racial discrimination, or if discrimination cannot be affected by manipulating perceived race, individuals have no incentive to do so. Despite Brazil’s history as a “racial democracy” there remains considerable discrimination in the labor market. Htun (2004), extensively documents racial inequality in both labor-market and educational outcomes. Leite (2005) finds that the race-education gap in Brazil is an important determinant of the race-wage gap. Gradín (2009) finds that poverty among Black workers is largely attributed to poor labor market characteristics rather than out-right racial bias.

The presence of a racial continuum combined with a history of discrimination results in individuals manipulating their perceived race, both as a matter of individual identity, and when dealing with educational institutions. Schwartzman (2007) finds that parents tend to classify their children as White when the parents are better educated and as non-White when the parents are not well educated. Francis and Tannuri-Pianto (2012) and Francis and Tannuri-Pianto (2013) find that the adoption of affirmative action policies leads both to students misrepresenting race to admissions offices, but also to students changing their self-perception of their own race.

In RAIS, information on race is collected and reported by employers, which are under strong incentives to complete the survey. Plants are penalized when data are late or not completed. Furthermore, RAIS includes information needed to administer a leave-loading program, which workers value. Scrutiny from the government and employees means formal employers are highly likely to comply with RAIS mandates. At the same time, employers are not under any particular pressure to manipulate race. Only recently have affirmative-action policies been introduced in admissions to state-funded universities and some public-sector agencies. To the best of my knowledge, there are no affirmative-action laws that bind on private-sector employers in Brazil.

My data also reveal evidence of racial inequality, despite the subjective nature of racial classifications. Table XIII presents estimates of a conventional Mincer-style wage regression using a full panel of data on jobs from RAIS between 2003 and 2010. The first column presents baseline regression with controls for Black and Brown race, education, experience and its square. The second and third columns control for heterogeneous plant and worker effects. Controls for experience and education have the expected signs and magnitudes. The baseline model suggests that being reported as Black or Brown is associated with 16 to 20 percent lower wages. After accounting for plant effects, about half of the Black/Brown discount remains. I still find that being Brown is associated with 2 percent lower wages after controlling for worker-specific heterogeneity.

Brazilian workplaces also exhibit segregation. Figure A.1 presents a histogram of the plant-size weighted distribution of the white share of all employees across all plants. This distribution strongly suggests the presence of segregation across plants. Fifteen percent of plants have no non-white workers and a further seven percent have no white workers. So, 22 percent of plants are completely homogeneous with respect to (reported) race. Across the support, the distribution exhibits no evidence of a mode near the white share of the formal sector workforce (roughly 60 percent). Brazilian workplaces are racially stratified relative to the overall population. These results hold up even when controlling for geographic variation in the racial composition.

2.3 Data

2.3.1 Source

I use the *Relação Anual de Informações Sociais*, or Annual Social Information Survey (RAIS) under an agreement with the Brazilian Ministry of Labor and Employment (MTE). The MTE collects the data in RAIS to administer a "Thirteenth Salary," or annual bonus, that is equivalent to one month's earnings to each worker, mandated by the Brazilian constitution. To administer the Thirteenth Salary, the MTE uses RAIS to collect the identity of each employee and their earnings in each plant where they worked during the year. RAIS also requires employers to provide detailed characteristics of every employee, including race, which MTE uses to produce labor market statistics for broader policy objectives.

RAIS collects data at the plant, not the firm, level. Plant management reports the data on behalf of the employees; in smaller enterprises the respondent may be the owner, while larger firms likely have an accountant, human resources manager or other administrator submitting the data. Uniquely among other employer-employee matched datasets, RAIS provides universal coverage of the formal labor market. For each plant, RAIS captures every worker in its employ during the survey year; for each worker, RAIS captures each registered employer he or she worked for over the same period.

2.3.2 A Statistical Description of Formal Labor Market of Brazil

My analysis is based on the 2010 wave of RAIS, the most recent wave available, which contains the over 65 million worker-job observations that define the Brazilian workforce. Restricting attention to full-time formal workers between 20 and 65 years of age in RAIS provides about 36.1 million unique workers. Almost 2.8 million plants employ these workers, producing about 42 million worker-plant observations.

Table XIV reports race, occupation classes and education levels in the 2010 RAIS. Race, as discussed earlier, is a complex issue in Brazil and, unlike in the US or Europe, is defined almost entirely on skin tone. This color-based definition leads to ambiguities that I do not observe in the US, where race is a matter of heredity and ethnicity. RAIS uses the following racial categories: Indigenous, White, Black, Yellow and Brown. With the obvious exception of the Indigenous category, these classifications reflect only a person's skin color.

RAIS depicts a formal workforce that is almost largely White and male. About 62 percent of the formal workforce is classified as White, about 6 percent are classified as Black and about 31 percent are classified as Brown. Workers classified as either Indigenous or Yellow combine to comprise just over 1 percent of the formal workforce.

Educational attainment in RAIS is reported in seven categories. Almost 45 percent of sampled workers have not completed high school. The largest single category is high school graduates, who comporise about 43 percent of the sample. Workers with education beyond high school combine to make up about 12 percent of the sample, with about 8 percent of the sample possessing at least a bachelor's degree.

Industry categories are aggregations of six-digit codes from RAIS as specified by the IBGE's National Registry of Economic Activity (CNAE 95). In 2010, almost 50 percent of sampled workers were employed in the trade and repair and production sectors. About 15 and 13 percent of workers are employed in the defense and real estate industries.

As with industry categories, occupational categories are aggregations of six-digit codes from RAIS specified by the IBGE's Brazilian Code of Occupations (CBO 2002). Service and production workers each make up more than a quarter of Brazil's workforce; combined these job classifications account for about 55 percent of the sampled workers. Agricultural workers, production, repair and maintenance workers, or the remaining blue-collar occupations, account for an additional 10.5 percent of the sample. The remaining White-collar and middle-skill occupations – public administrations, professionals, artists, scientists, mid-level technicians and administrative workers – make up between 32 percent of the sample. In 2010, sampled workers earned roughly R$1,354, roughly equivalent to US$800.[3]

2.4 Race Change

The matching feature of RAIS allows me to observe within-year job changes, which, in turn, reveal changes in racial status. In 2010, of the 1.8 million workers initially classified as Brown or Black who change jobs, over a third are classified as White in their next job. First, I describe the job and race changers in comparison to workers in general. Then, I attempt to sort out the contributions of individual and employer characteristics to changes in racial identity.

2.4.1 Race Changers and Their Employers

Table XV reports the racial, occupational and educational compositions of the entire workforce in 2010, and of job and race changers in that year. The mean age and monthly income of each group are also given. A job change occurs when I observe a worker associated with more than one plant in 2010; race change occurs when the reported race for a job changer differs between jobs. All figures are for workers in their first reported job in 2010 by hire

[3] These statistics are also available for RAIS 2003, and are similar in both sign and magnitude. Please contact the author for these statistics and results.

date. In 2010, almost five million workers, roughly 13.8 percent of the workforce, change jobs or take on a second job. Nearly 1.6 million workers, 31.2 percent of job changers, change race.

The overall workforce is about 62 percent White, 31 percent Brown and almost six percent Black. The racial composition of job changers is essentially the same. However, race changers look very different; only about 44 percent of workers who change race are White; 42 percent are Brown and 11 percent are Black. Table XVI describes the share of race changes between the White, Black and Brown classifications among all job changes and race changes. Changes between these three classifications are present among almost 30 percent of all job changes and account for over 95 percent of all race changes. Moreover, changes from White to Brown and from Brown to White make up over 73 percent of all race changes.

Workers only represent one side of the race-changing phenomenon; the plants where they work are the other. In 2010, RAIS records over two million distinct plants. Table XVII provides a description of these plants in terms of worker demographic and occupational compositions, industry affiliation and size.

On average, just over a third of a plant's workforce consists of service workers and vendors. An additional 20 percent are administrative and clerical workers. Other occupations make up between 2.5 and nine percent of a plant's workforce, except the military and police category which is 0.01 percent. The share of firms in a given size category decreases as the plant size decreases. Very small plants of one to four workers make up over 64 percent of all plants, and nearly 92 percent of plants have less than 20 workers.

Nearly 41 percent of plants are in the trade and repair industry, the largest sector by number of plants. Real estate, production and agriculture are the next largest sectors, making up between 9.5 and 12 percent of plants each. The typical labor force of a plant is roughly 68 percent White. The mean age of a plant's employees is almost about 35 years old and about 42 percent of a plant's employees are women.

2.4.2 Earnings Effects of Race Changes

Because racial classification in Brazil is fluid and racial categories there are strongly associated with socio-economic outcomes, workers have the scope and incentive to manipulate their perceived race to improve treatment in the labor market. If workers change race to gain labor market advantage, I should observe earnings should increase for individuals changing to White and decrease for those changing from White. To test this proposition, I estimate the effect of race change on earnings in the destination job, using the sample of workers jobs who changed jobs in 2010. My empirical model has the form

$$y_{i,j(i,1)} = \gamma \cdot racechange + \alpha y_{i,j(i,0)} + \mathbf{x}_i\beta + \mathbf{z}_{i,j(i,1)}\eta_{j(i,1)} + \mathbf{z}_{i,j(i,0)}\eta_{j(i,0)} + \psi_{j(i,1)} + \varepsilon_i \ , \quad (2.1)$$

where $j(i,0)$ is worker i's origin plant, $j(i,1)$ is the worker's destination, y_i^k is the worker's earnings ($k \in \{0,1\}$), $\mathbf{x}_i$ contains worker characteristics, $\mathbf{z}_{i,j(i,1)}$ contains plant and job characteristics, and $\psi_{j(i,1)}$ represents the plant effect for the destination job.

Table VI reports the results from estimating (2.1) for those workers who are Black or Brown in the origin job; Table VII presents the results those who are White in their job or origin. Each table distinguishes two cases, first omitting and then including the plant effects. In the full specification, with plant effects, I see evidence of the benefit of changing from Black or Brown to White and from White to Black or Brown. Changing race to White generates a 2 percent increase in monthly earnings, while changing race from White leads to a 3 percent decrease. Both effects are highly statistically significant.

2.4.3 An Empirical Model of Race Change

I take a two-step approach to characterize the features of the labor market correlated with change in race. First, I estimate linear probability models to predict whether workers change race conditional on changing jobs:

$$\Delta Race_{i,r} = \mathbf{x}_{i,r}\beta + \psi_{i,j,r} + \varepsilon_{i,r} \ , \tag{2.2}$$

where $\Delta Race_{i,r}$ is a race-change indicator for worker i with "race of origin" r, $\mathbf{x}_{i,r}$ is a vector of worker characteristics, $\psi_{i,j,r}$ is a fixed effect for destination plant j and $\varepsilon_{i,r}$ is a mean-zero error term. The set of worker characteristics includes origin and destination log earnings, origin and destination occupation, origin industry, age and education level. I carry out this step separately for the cases where the race of origin is White (W) and Black or Brown (B), using the 2010 job-changer sample. To gauge the relative importance of the employment destination on race changes, I estimate (2.2) both with and without plant effects. I compute standard errors that are clustered at the plant level. To foreshadow my results, almost all of the variation I are able to explain is explained by the plant controls.

Second, I link the estimated destination plant effects, $\hat{\psi}_{j,r}$ to the plant's racial composition and other observable characteristics:

$$\hat{\psi}_{j,r} = \delta White_j + \mathbf{z}_j\alpha + u_{j,r} \ , \tag{2.3}$$

where $White_j$ is the White share of plant employment at the beginning of the year and $\mathbf{z}_j$ contains the average worker age, average salary. Again, I estimate the White to Black/Brown and Black/Brown to White cases separately. For this step, I compute heteroskedasticity-robust standard errors.

2.5 Results

2.5.1 The Correlates of Race Change

Table XVIII reports my first-stage results from the estimation of (2.2). The first column presents a baseline regression wit controls for education and log wages in the worker's origin and destination jobs. The second column includes additional controls for occupation, industry and state and the third column adds controls for heterogeneous plant effects. The R^2 for columns (1) and (2) are 0.01 and 0.11. The inclusion of plant effects in column (3) results in an R^2 of 0.66. My discussion focuses first on characterizing the plant effects, and then individual characteristics. Note that all models in Table XVIII control for many other individual and plant-level characteristics, including age, gender, industry of origin plant, occupation of origin and destination job, and the legal form of ownership of the origin plant.

Plant Effects

The unit of analysis in Table XVIII is an individual worker who changes jobs. To interpret the plant effects, consider the results in columns (3) and (4) that condition on being reported Brown or Black in the initial job. The dependent variable is equal to 1 when the worker is reported as White in the destination job, and zero otherwise. In column (4), the plant effect picks up the share of Brown/Black workers that become White (net of worker-level observable characteristics).

Figure A.2 presents a plant-employment-weighted histogram of estimated plant effects $\hat{\psi}_{j,r}$ estimated in column (4). These are measures of the frequency with which workers classified as Black/Brown change to White within a plant.[4] Plant effects are highly bimodal. There is a negative mode around -0.3 and a positive mode with a center near 0.6. The

[4] Because some plants have very few new hires in a year – especially small plants – there is some concern of a limited mobility bias. Weighting by plant size attenuates this concern somewhat. Restricting my analysis to plants with at least 10 new hires yields comparable results.

baseline probability of race change is around 0.2. Therefore, the negative mode represents plants where workers almost never become White. The positive mode represents plants where around 80 percent of Brown or Black workers become White upon hire. I know from my estimation of equation (2.2) that plant effects explain a much larger amount of the variation in race changes than any worker characteristics do.

Education and Earnings

I continue to focus discussion on the case of Brown/Black workers who are classified as White by their destination employer. Workers with higher education levels are more likely to become White when changing jobs (and less likely to become Brown if previously White). Relative to workers in the with college-level educations, the likelihood of a Brown or Black worker decreases as education level decreases from 2.5 percent less for High School graduates to 8 percent less for workers with no schooling. This is an important result, in that it coalesces with previous research which suggests that education "Whitens" (Francis and Tannuri-Pianto 2013). However, where previous research has focused on how educated parents classify their children, my analysis suggests that education Whitens not just within families across generations, but for the same person across jobs. The estimated effect may reflects selection (more educated workers find it easier to manipulate their perceived race) or statistical discrimination (when plant managers have to predict race on the basis of other observables).

The pattern of earnings is less clear, though with regard to this variable I acknowledge that any interpretation is extremely tenuous given the endogenous determination between earnings, job mobility, and racial mobility. Column (2) shows workers who change to White tend also to be moving from higher-wage to lower-wage jobs (including controls for industry and occupation). Adding plant controls in column (3) eliminates, and even slightly reverses the effect – comparing workers who move to the same plant, and were both initially Brown or

Black, those who move to a higher-wage job are marginally (about 1 percent) more likely to be changing to White. The fact that the result from column (3) is eliminated by introduction of plant controls implies a negative correlation between plant-level wages and the propensity of the plant to report workers as White.

2.5.2 Why Do Some Plants "Whiten"

To help interpret my findings, I examine what it is about some plants that leads them to "Whiten" their incoming workers. Workplace discrimination can vary substantially across plants. Perhaps what I are seeing is evidence of workers with the opportunity to do so manipulating race to increase their chances of employment. If that is the case, then (1) plants should be highly segregated by race, and (2), the probability that a worker becomes White should be strongly associated with the share of the plants workforce that is White.

Figure A.1 illustrates the presence of segregation across plants, as discussed in Section 2. Suppose workers who can manipulate race do so to obtain employment with discriminatory firms. Then most of the plant-specific heterogeneity should be associated with the race share. Table XIX reports the coefficient on White share from projecting plant effects estimated in Table XVIII onto plant-level characteristics. The first column reports the White to Black/Brown case and the second column reports the Black/Brown to White case. The estimated coefficient is 0.78 for the plant effect on change to White. This single variable explains 40 percent of the variation in the plant effects. I also control for plant average age, plant average wages, plant average sex, the share of workers in each of 10 occupations, the industry, and the legal form of ownership of the plant. These other factors explain less than 5 percent of the total variation.

To investigate this result further, I assign to each plant its decile in the (size-weighted) distribution of the share of workers who are White. Table XX reports the results from regressing plant effects on a plant's location within the distribution of White share. Column

(1) shows the results of projecting onto the White share deciles, comlumn (2) includes controls for average earnings, gender composition and skill distribution, and column (3) adds additional controls for plant industry and location. All three specifications demonstrate that the relationship between White share and the propensity of a firm to Whiten non-White workers is monotonically increasing. Firms with very few White workers are highly unlikely to make workers White. Firms with 100 percent White workforce make workers White 85 percent of the time.

2.6 Conclusion

If it were possible, a rational response to racial discrimination would be to one's change racial identity. In the U.S., where racial categories are strictly defined through heredity or ethnicity, there is little room for taking on a new racial identity. In Brazil, racial categories are much more closely connected to skin color, creating more flexibility in racial identity and opening up the possibility for identity changes. Using employer-employee matched data from Brazil, I show that workers sometimes change race when changing jobs.

I then examine job changers in RAIS, linking observed race changes to worker and plant characteristics. I find that workers with higher levels of education and rates of earnings growth are more likely to change from 'Brown' to 'White'. However, individual characteristics account for no more than three percent of the variation in race change. In contrast, plant effects account for 60 percent. This is exactly what one would expect if workers manipulate race to obtain jobs with employers that have some kind of discriminatory hiring practice. Consistent with this explanation, I find that 'Whitening' is associated with moving to Whiter plants. These plants also tend to be smaller, and lower-paying.

An alternative explanation is that workers fail to report race in some fraction of cases, and employers "guess" based on their existing workforce. This seems highly unlikely. I find very little mobility of workers by gender, which is just as easy to omit from a form, but far less easy to manipulate in the hiring process. What little mobility there is is not correlated with racial mobility, and is not correlated with plant gender. Understanding mechanisms underlying the observed relationship between employment mobility and racial mobility is the subject of ongoing research.

Chapter 3

Pay Structure and Turnover: Evidence from Brazilian Employer-Employee Matched Data

3.1 Introduction

The persistence of heterogeneity in labor-market outcomes and its causes remain recurring themes in labor economics research. Countless studies have demonstrated the importance of worker characteristics – such as gender, race, education and experience – in wage determination. Many others have examined determinants of labor demand and how firm productivity, industry or ownership, may affect earnings. However, these studies do not explain why seemingly identical workers earn different wages when employed by different firms. Largely, this is a limitation of the data, as theoretical explanations, such as learning models, models of asymmetric information, firm-specific human capital models, all address this issue[1]. However, the work-horse data sources in labor economics – censuses, population surveys and

[1] See Lazear and Oyer (2013) for a comprehensive review.

industry surveys – rarely allow the researcher to simultaneously control for worker and firm characteristics in a meaningful way.

The limitations of conventional data sources have led modern economists to two solutions. The first, often referred to as "Insider Econometrics," typically involves an intimate knowledge of one or two firms, though some large-scale projects, such as the World Management Survey, have applied this method to hundreds of firms world-wide (Ichniowski and Shaw 2013; Bloom et al. 2012). The second solution is to use employer-employee matched data. The practice of using data that record identifiers of both workers and employers to study labor market outcomes is over 100 years old, but limitations in availability and computing capabilities have historically made the use of such data inconvenient, if not impossible (Moore 1911).

In this paper I study variation across firms in their pay structure and the relationship between pay structure and human resource management (HRM) outcomes, specifically employee turnover. I do so using the 2002-2010 waves of Brazil's *Relação Anual de Informações Sociais* (RAIS), which is effectively an annual census of all formal-sector jobs in Brazil. RAIS data are unique in that they provide highly detailed information for over 40 different firm and worker characteristics. They also distinguish between hires and transfers, and between employer-initiated and worker-initiated separations (layoffs and quits) retirements. The post-2002 waves include variables that were previously unavailable: race, age in years, date of hire, and nominal earnings denoted in reais, as opposed to multiples of the minimum wage. These variables are in addition to the rich set of firm characteristics which allow me to control for a firm's size, skill intensity, industry and ownership type. The ability to directly observe the date of hire is particularly important for this study, because it means that tenure is recorded exactly and not inferred from the date in which a match between worker and employer is first observed.

I focus on two features of a firm's pay structure: the starting pay premium and seniority profile. Using the RAIS data, I follow Abowd et al. (1999) (henceforth, AKM) and decompose log earnings into components associated with time-varying worker characteristics, worker fixed effects, firm fixed effects and firm-specific returns to seniority. The firm-specific components of the decomposition provide empirical measures of the firm's starting pay premium and seniority profile.

I then examine how these features of the firm's pay structure are related to its hiring, firing, and quit rates. I find that increased compensation through any channel is associated with a decrease in turnover. Furthermore, steeper and more linear tenure profiles are associated with a reduction in separations.

Consistent with profit-maximizing behavior, I find that starting pay structure and the slope of the seniority profile are negatively correlated. Firms with high wage premia have flatter tenure profiles, and firms with steep tenure profiles also have shorter periods over which earnings rise with tenure. This pattern characterizes private for-profit firms. Government firms, which are not as exposed to market discipline, exhibit little heterogeneity in pay structure, and what variation exists is largely uncorrelated with recruiting and retention.

My paper is closely related to the AKM paper, and their follow-up paper (Abowd et al. 2002) in which the authors use employer-employee matched data to describe the interaction of workers and firms in the French labor market. These papers have helped spawn a growing literature that uses matched data to characterize heterogeneity in labor market outcomes. Horny et al. (2009) use Portuguese data to examine job durations find that unobservable firm characteristics explain about 30 percent of the variation in job duration. Caliendo et al. (2012) use French employer-employee matched data to study firm hierarchies and how management structure affects pay grades to show that expanding firms which increase the number of levels in their hierarchies pay lower average wages in across all pre-existing levels. Zwick (2011) examines how German firms with differing responses to tenure seek different types of employees based on age and gender, and finds that firms that pay higher seniority wages keep older workers on staff, but hire fewer older workers as new employees. Mendes et al. (2010) use Portuguese matched data to explore the extent of assortative matching in the labor market, and present evidence that the most productive workers match with the most productive firms, especially among longer-established firms. Using hybrid matched data in Los Angeles county, Breau (2009) examines the relationship between export industries and immigration to show that Asian and Hispanic workers are relatively more dependent on export-related activity and that export-related activity is associated with increased wages for White and Black workers. Card et al. (2013) use West German data to explore differences in pay resulting from within-firm seniority as opposed to tenure on the job and show that incrasing plant-level heterogeneity and increasing assortative matching explain a large part of increased (West) German inequality from 1985 through 2009.

In Brazil, employer-employee matched data have previously been used to study a number of phenomena. Muendler et al. (2012) examine spinoffs in Brazil's manufacturing sector and show that spinoffs led by (former) employees have higher survival rates. Lopes de Melo (2013) studies assortative matching by skill level and Krishna et al. (2012) examine the relationship between trade liberalization and wage distributions. Among studies using Brazilian data, mine is most closely related to Menezes-Filho et al. (2008), who use two cross sections from RAIS in 1990 and 1997 to characterize changes in the Brazilian wage structure with comparisons to the U.S. and France. They find that plant-level wage premia explain very little of the cross-sectional variation in earnings relative to the U.S. Our results indicate that variation in starting pay and the shape of the seniority profile are empirically important. The discrepancy arises because I control separately for unobserved worker and firm-specific heterogeneity. Consistent with this literature, I find that in Brazil the correlation between worker and firm effects is large and positive.

3.2 Starting Pay, Returns to Seniority and Turnover

The literature suggests specific hypotheses regarding the relationship between a firm's starting pay and seniority profile, and its turnover outcomes. The first is that, hiring rates fall as starting salaries rise either because new hires are more costly to take on, or because higher paying firms have less need to replace their existing workers. Firms with steep seniority profiles tend to have lower starting pay, so there is some ambiguity as to how increased returns to seniority relate to hire rates, *ceteris paribus*. Other researchers have shown that firms with steep profiles seem to prefer hiring younger, less experienced workers who would be cheaper to employ (Lazear 1979; Adams and Heywood 2007; Daniel and Heywood 2007; Zwick 2011).

The second hypothesis concerns separations, and the literature has shown that increased starting pay and steeper seniority profiles are associated with reduced separations. The interaction between pay structure and separations works through two channels. The first is through fires and other firm-initiated separations. Increased pay, either through wage premia or the seniority profile, indicate a firm's commitment to its employees. A steeper seniority profile may be associated with firms that require higher amounts of firm-specific human capital, and a new worker is unlikely to be as productive as a currently employed worker. The second channel through which pay structure and separations interact is through quits. Here the theory is much more straightforward: increased pay increases the opportunity cost of leaving, meaning workers in firms with higher starting pay or steeper tenure profiles are less likely to leave the firm. Unlike many other data sources, my data allow me to distinguish between separation types and evaluate these hypotheses (Lazear 1979; Dostie 2005; Schönberg 2007).

A third hypothesis implied in the above research is that these relationships originate from the profit maximizing behavior of rational firms and their agents. As a result the above conclusions should be stronger for firms that are more susceptible to market forces. In sectors or industries where a firm's profitability and survival are tied closely to its ability to produce or serve efficiently, i.e. product demand is relatively elastic, the above predictions should be more apparent than in less elastic sectors or industries, such as the military or other state operations. Our data allow me to evaluate this hypothesis as well, because they distinguish firms by ownership type.

3.3 Empirical Framework

Our empirical strategy consists of two steps. In the first step, I estimate a Mincer-type wage equation to obtain firm-specific measures of starting pay and seniority profiles following the methods outlined in AKM and Abowd et al. (2002). In the second step, I model the relationship of starting pay and seniority profiles with a firm's hire, fire and quit rates using least-squares estimation.

3.3.1 Decomposition of Log Earnings

I decomposing earnings into worker and firm-specific components as follows:

$$y_{it} = \alpha + x_{it}\beta + \theta_i + \psi_{J(i,t)} + \gamma_{1J(i,t)}s_{it} + \gamma_{2J(i,t)}s_{it}^2 + \varepsilon_{it}. \tag{3.1}$$

y_{it} is the log monthly earnings of worker i at time t, x_{it} is a vector of K observed time-varying worker characteristics, s_{it} is years of tenure, and ε_{it} is a mean-zero error. The θ_i and $\psi_{J(i,t)}$ are unobserved worker and firm effects, and $J(i,t)$ indicates that firm j employs worker i at time t. Importantly, this empirical model allows for firm-specific returns to seniority. I let N and J designate the number of workers and firms, and N^* denote the number of worker-year observations.

The variables in x_{it} are the observable time-varying characteristics that are productive at all employers. RAIS provides information on educational attainment and labor-market experience, which I include as a quartic and interact with race and gender. Unobservable individual characteristics, such as ability and reliability, that pay off in all jobs are captured by the effect, θ_i. The firm-specific determinants of pay are captured by $\psi_{J(i,t)}$, which reflects starting pay, and $\gamma_{h,J(i,t)}$ $(h = 1, 2)$ which determine the slope of the firm's seniority profile.

Defining Y as a stacked vector of the N^* observations of log earnings, I rewrite equation (3.1) as

$$Y = X\beta + D\theta + F\psi + F_1\gamma_1 + F_2\gamma_2 + \varepsilon, \tag{3.2}$$

where X is $N^* \times K$ (including the intercept), D and F are $N^* \times N$ and $N^* \times J$ design matrices of worker and firm dummy variables, and F_1 and F_2 are the direct products of F with $N^* \times 1$ vectors with elements given by s_{it} and s_{it}^2. Estimation of (3.1) by ordinary least squares (OLS) is conceptually straightforward, but computationally burdensome because of the high-dimension matrix inversion involved in solution to the least-squares normal equations. In particular, the solution requires the inversion of a cross-product matrix with dimensions $(N + 3J + K) \times (N + 3J + K)$, which is infeasible for a sample as large as mine. Therefore, I rely on the iterative conjugate gradient algorithm developed by Abowd et al. (2002).[2] They show that the firm and worker effects, θ and ψ are separately identified within connected components of the "realized mobility network," $g = 1, \cdots, G$. Worker and firm effects are identified relative to the within-component mean, which I normalize to zero in the largest component.

[2]I separately identify firm and worker effects relative to the intercept within $G = 3$ different, mutually exclusive components of the realized mobility network. Abowd et al. (2002) show that identification requires G linear restrictions on worker and firm effects. I follow their example of setting the average person and group effect in the largest group to zero after identifying the grand mean of the regression. In the remaining two groups, I allocate $\lambda = \frac{1}{2}$ of the group mean to both the worker effect and the firm effect. An overwhelming majority of workers and firms belong to one very large group.

3.3.2 Relating Pay Structure Components to Turnover

Next, I relate the firm-specific components of worker pay to its hiring, firing (layoff) and quit rates. Together these measures of turnover represent important human resource management objectives which pay policies are presumably designed to influence. This exercise involves regressing each component of turnover on the estimated starting pay and seniority profile coefficients as follows:

$$h_{jk} = a_{k0}\hat{\psi}_j + b_{k1}\hat{\gamma}_{1j} + b_{k2}\hat{\gamma}_{2j} + Z_j\delta_k + \upsilon_{jk}, \tag{3.3}$$

where h_{jk} is either the hire, fire or quit rate at firm j (where k indexes the turnover variables), and Z_j contains measures of firm size, skill intensity, industry affiliation and ownership status.

First, I estimate simple, bivariate regressions of h_{jk} on each estimated pay structure coefficient, $\hat{\psi}_j$, $\hat{\gamma}_{1j}$ and $\hat{\gamma}_{2j}$ separately. These simple regressions serve as a benchmark for the results from (3.3), in which their effects are estimated jointly. Second, I estimate (3.3) with the variables in Z omitted, and then with the controls included. Finally, to gauge how the relationship between pay policies and turnover vary with market forces, I estimate (3.3) separately by ownership status: for-profit private sector, state-owned, non-profit and sole proprietorship.

3.4 The RAIS Data

3.4.1 The Formal Labor Market of Brazil and RAIS

The Brazilian economy, like most Latin American economies, is noteworthy for its large informal labor market. In Brazil a worker is formally employed if he or she has a registered identification number with one of two social security programs: the *Programa de Integração Social* (PIS), or Social Integration Program, or the *Programa de Formação do Patrimônio* *do Servidor Público* (PASEP), or Civil Servants Equity Formation Program, depending on if the worker is employed in the private sector or the public sector. PIS/PASEP numbers are consistent across workers and follow a worker for life. For firms, formal employment means that the employer contributes to a bank account administered by either *Caixa Econômica Federal*, if registered with PIS, or *Banco do Brasil*, for PASEP workers, covering all worker categories. Formal employers must also have employment contracts for all employees. The most common contract type is the *Consolidação das Leis de Trabalho* (CLT), or Labor Law Consolidation. Other contract types include internships, independent contractors, directorships and government contractors. The Brazilian government defines formal employment with these criteria, and this definition is consistent with definitions used by researchers when studying other Latin American economies (Gasparini and Tornarolli 2009b).

Formal employment entitles Brazilian workers to constitutionally mandated benefits such as 30 days paid vacation time, maternity leave and protections, disability insurance, transportation reimbursement, unemployment insurance and an "*Abono Salarial*." The *Abono Salarial*, also known as the 13th Salary or Christmas Bonus, is paid from employer contributions to a worker's PIS/PASEP account and is equivalent to one month's pay; workers with less than 12 months on the job receive a prorated amount. The *Ministerio do Trabalho e Emprego* (MTE), Brazil's labor ministry, administers the 13th Salary by collecting data from the *Relação Anual de Informações Sociais* (RAIS) or Annual Social Information Survey.

The information in RAIS is provided at the establishment level by a company administrator. In smaller firms and plants, this is likely the owner or plant manager; in larger establishments there may be dedicated personnel who submit the information. Coverage is universal, as employers who fail to complete the survey risk litigation from employees who have not received their 13th Salary. Since coverage is universal, RAIS serves as a census of formal labor in Brazil, which the MTE cleans and combines with other data to produce workforce statistics. Universal coverage gives me access to data from firms in all registered sectors of the Brazilian economy, including privately held firms, state enterprises, the military and sole proprietorships.

Through the sample period, formal employment grew steadily in Brazil, from nearly 42 million jobs in 2003 to over 65 million jobs in 2010. Unemployment is decreasing over this time period as well, from more than 11 percent in January 2003 to just over five percent in December 2010. Additionally, real wages grew consistently over this time period. Brazil's formal labor market proved particularly resilient during the global economic downturn, as real wages and formal employment grew in 2008, 2009 and 2010.[3]

Menezes-Filho et al. (2008) compare wage determinants for Brazil in 1990 and 1997 to the results from early 1990s France and the U.S. provided by Abowd et al. (2001). Brazil is comparable to both the U.S. and France in terms of occupation and gender wage differentials, however wage inequality is greater in Brazil. The authors also note that Brazil's wage structure resembles the French wage structure more so than the American wage structure, a likely result of France and Brazil sharing certain institutional similarities, specifically with regards to constitutionally mandated labor institutions dating to the Napoleonic Era.

3.4.2 Model Variables and Sample Construction

Estimation of (3.1) involves regressing log earnings on a vector of time-varying worker characteristics that consists of education, experience, and tenure. I also interact experience with worker race and gender, while I interact tenure with an indicator of the firm employing the worker in that period. To estimate (3.3), I include the firm's industry and ownership as explanatory variables, along with the estimated firm-specific starting pay and seniority profile coefficients. The dependent variable is either the firm's average hire rate, fire rate or quit rate, depending on the specification. Here, I provide more information about how RAIS reports the variables I use in my sample.

[3]Source: `www.tradingeconomics.com/brazil` | Instituto Brasileiro de Geografia e Estatística (IBGE).

Earnings

The dependent variable in (3.1) is log average monthly earnings, reported in 2003 Brazilian Reais. When a worker is employed for 12 months, average monthly earnings is simply his or her yearly earnings divided by 12. When a worker is employed fewer than 12 months, the total earnings paid for the year are divided by the number of months worked; for partial months, the earnings are pro-rated to reflect what the worker would have earned had he or she worked the entire month. All of these calculations are done by the MTE and included in the raw RAIS data. To express monthly earnings in 2003 reais, I deflate using the consumer price index.[4] The few observations reporting less than the monthly minimum wage are dropped as improperly recorded.

Individual Characteristics

Education is included as a time-varying characteristic because I do not restrict the sample to workers who have completed their schooling. RAIS provides detailed information regarding educational attainment, with elementary education decomposed into pre- and post-eighth grade, and categories for workers who are illiterate or never attended school; these categories are in addition to complete and incomplete elementary, high school, and university-level education. Later waves of RAIS expand educational attainment categories further, including separate categories for master's and doctorate/professional education levels. I aggregate these outcomes to five levels of education: less-than elementary school, elementary education, some high school, high school, some college, and bachelor's degree or better. Coverage for this variable is complete, as there is no "non-response" category.

RAIS does not explicitly report a worker's total experience in the workforce. However, RAIS does report if a new hire's job is the worker's first registered employment, allowing me to construct experience profiles for each worker. For those workers whose first employment

[4] Source: `www.tradingeconomics.com/brazil` | Instituto Brasileiro de Geografia e Estatística (IBGE).

I observe, experience is equal to one year, and in following years is the sum of all following tenure observations. When I do not observe a worker's first employment, I impute experience as the greater of either age less years of schooling less six or tenure in the first observed job. I adjust reported tenure and experience values, converting from months to years. I include experience through the quartic term, and also include experience interacted with non-white racial categories and gender through the quartic terms.

Racial identification is often difficult in Brazil. There are no affirmative action or equal employment opportunity programs in the labor market; as a result, the race of new hires and applicants is not closely monitored. In fact, for any given year, somewhere between eight and fifteen percent of observations do not have a reported race; I omit these from my analysis.

For each worker, RAIS records not only the months of tenure at the firm, but the actual date of hire, and the month and year of separation, where applicable. These dates are vital to my analysis, since RAIS top codes tenure at 60 months. With the month of both hire and separation, I calculate complete job durations for each worker in each firm.

Plant and Firm Characteristics

RAIS provides industrial codes at the five-digit level from the *Instituto Brasileiro de Geografia e Estatística*(IBGE), Brazil's statistical agency. I aggregate these codes into 16 different industries. The largest industries, by number of plants, are production/manufacturing and services/repairs while the smallest are international organizations and domestic non-profit organizations. For firms that change industries or operate plants in multiple industries, I designate the modal industry by number of observations as the industry of record.

RAIS also provides four-digit codes that specify the firm's "legal nature." These codes describe the firm's ownership and formation, and include designations such as limited liability corporation, publicly traded corporation, non-profit, state-owned business, or international non-profit. I aggregate these codes into four broad categories: private for-profit enterprises, public or state-owned enterprises, private non-profit enterprises and individually owned enterprises and sole proprietorships. Private for-profit firms include any and all incorporated entities that are neither governmental bodies nor non-profits. State-owned and government enterprises include all municipal, state and federal government entities. Private non-profit firms include religious organizations, political parties and labor unions, among other organizations. The last category, sole proprietorships, includes non-incorporated firms. The coverage of all government entities and non-incorporated firms in the data is a unique feature of RAIS, and critical to my analysis, as it allows me to relate patterns in starting pay and seniority profiles and verify that the behaviors described profit-maximizing objectives.

Another special feature of RAIS is the level of detail the data provide regarding all admission types – within firm transfers and new hires – and separations. New admissions to a firm are classified along several lines; I are concerned with identifying new hires to the firms. I calculate a firm's annual hire rate as the ratio of new hires during a calendar year, not including transfers within firm, to the total number of employees recorded in the firm on 31 December of that year. I then take the average of this ratio for all years a firm is observed as its "Hire Rate."

I repeat this procedure to compute fire and quit rates. Using the separation categories in RAIS, I define a quit as any employee-initiated separation, not including retirement or within-firm transfer. I define a fire as any employer-initiated separation, again not including mandatory retirements or within-firm transfers. I also exclude separations that result from the employee's death or disability from either calculation. As with hire rates, I first calculate annual fire and quit rates as the ratios of fires and quits to the total employed workforce at year's end, and then take the average of these ratios as the firm's "Fire Rate" and "Quit Rate." The difference between quits and fires is important: in Brazil, a worker who initiates a separation loses out on significant social security benefits, including a month of full pay before having to file for unemployment insurance. As a result, quit rates in Brazil are much lower than in the U.S.

3.4.3 Sample Selection

I carry out my analysis on a panel of workers and plants from the 2003-2010 waves of RAIS. Our sample is comprised of workers who are 20-65 years old, are contracted for at least 24 hours of work a week, have at least one month of tenure in the job and have a complete set of covariates, meaning I can construct the worker's experience and tenure and there are no missing values in race, gender or education. Workers are identified based upon their 16-digit PIS/PASEP numbers, which are consistent across employment type, as RAIS tracks workers in every job they hold. I restrict attention to the worker's highest paying job during a given year. These restrictions yield a sample with 329,963,533 unique worker-firm-year observations consisting of 72,384,247 unique workers and 3,934,097 unique firms. To estimate the firm-specific seniority effects with reasonable precision, I exclude firms with fewer than 200 observations over the sample period. This additional restriction reduces total observations to 143,561,805; total unique workers decreases to 36,590,870 and total unique firms decreases to just 18,556.

Table XXI reports summary statistics of the individual characteristics for each RAIS wave. Consistent with the growth of the formal workforce, the number of workers included in my sample grew from about 15.6 million to 19.9 million workers between 2003 and 2010. Real earnings rise over the sample period. Log average monthly earnings, deflated to 2003 reais, are 6.58 in 2003 and 6.81 in 2010, indicating an increase of more than 20 percent.

Between 2003 and 2010, educational attainment rose. The shares of workers in all three sub-high school education levels decrease over the sample period; workers with below elementary-level schooling decrease from about 28 to 17 percent of the sample between 2003 and 2010; the share of elementary-level educated workers decreases from about 14 percent to 11 percent and the share of workers with some high school education decreases slightly from about 7.2 percent to 6.5 percent. The largest group of observations by education level in all years is high school graduates, which account for about 31 percent of workers in 2003 and 42 percent in 2010. Over 16 percent of workers have a college degree or better in 2003, and this figure grows by almost three percentage points in 2010.

The average level of experience in 2003 is about 8.5 years, which declines slightly over the sample period as younger, less experienced workers enter the formal labor force.The average worker has about 6.7 years of tenure in 2003. Similar to experience levels, mean tenure falls slightly between 2003 and 2010.

In 2003, the racial composition of the sampled formal labor force in Brazil shows that the majority of workers, about 72 percent, are White. About four percent of workers are Black, over 23 percent of workers are Brown, and about one percent are either Yellow or Indigenous workers. The share of white workers decreases, with a sudden shift up to about 76 percent in 2006, after which the decreasing trend resumes. This jump in 2006 is a result in a change of how race is recorded beginning in 2006[5]. The male share of the sampled labor force decreases over the sample period, from about 61 percent in 2003 to about 57 percent in 2010.

The top panel of table XXII provides the distributions of firms across industry and ownership categories. For multi-industry firms or conglomerates, the reported industry is the modal industry by number of workers in that industry. The industries reported are 15 of the 16 aggregate categories designated by IBGE.[6] The single largest industry, accounting for nearly 26 percent of all firms, is Defense and Security, which includes all military, police, and fire-fighting organizations, as well as other public and private organization providing

[5]Until 2006, non-reports in race were identified by the question not being answered; in 2006 an option of explicitly ignoring the race variable was made available to respondents. Rivera (2014) for more on how this change in the survey and race in general may affect economic analysis, and **?** for more on racial subjectivity in Brazil.

[6]The sixteenth industry, domestic services, contained no observations that fulfilled the sample selection criteria.

similar, security-based functions. Production, or manufacturing, is the next largest sector, comprising 21 percent of all firms.

In terms of ownership status, 62 percent of sampled firms are classified as private, for-profit firms and a little more than a quarter are government and state-owned enterprises. Almost nine percent of firms are private non-profits and about two percent are individually owned enterprises. Since, as noted in Bloom et al. (2012), private, for-profit firms are the best managed and most efficient firms operating in the Brazilian economy, including these ownership categories serve as a way to verify that the behavior described by the estimation output is the behavior of profit-maximizing, rational firms.

The bottom panel of the table presents summary statistics for the constructed turnover outcomes. The average hire rate over the sample period is 24 percent, meaning that in the typical firm almost a quarter of the employees recorded on 31 December are new hires. The average fire rate is about 13 percent, while the average quit rate is only five percent. Thus, firings occur almost 2.5 times as often as quits. These differences may indicate something about the interchangeability, or lack thereof, between these separation types, since a worker who quits their job essentially "fires" the firm, forgoing many benefits provided to fired workers.

3.5 Results

To obtain the firm-specific estimates of starting pay and seniority profiles, I estimate (3.2) following Abowd et al. (2002). First, to demonstrate how much variation is absorbed by the individual worker and firm effects, I estimate (3.2), omitting D and F, and successively reinstate the model adding the worker and firm effects.

Descriptive Analysis

Table XXIII reports the estimated coefficients of the time-varying individual characteristics contained in X for the three model specifications. First, consider the results of model 1. The estimated coefficients on education increase monotonically with attainment. The estimated college-premium relative to high school education is 13 percent. The results for experience and its interactions with gender and non-white racial categories also follow the standard empirical Mincer-equation pattern; alternating signs, implying diminishing marginal returns. As with returns to education, the baseline specification produces a large estimated payoff to experience. The first year of experience is associated with a 12 percent increase in earnings for white male workers. The effects for female, black and brown workers are smaller by 11, 3 and 5 percentage points. These baseline regression results in model 1 complement the findings of Menezes-Filho et al. (2008).

Model 2 shows that including the unobserved worker effects accounts for most of the estimated returns to education and experience reported in model 1. For example, the estimated premium for completing college is now about 11 percent and the return to the first year of experience roughly 3 percent. Adding for the firm-specific terms in model 3 has relatively little impact on the estimated education coefficients, but predictably reduces the estimated experience coefficients. Since tenure effects are estimated at the firm level, I defer discussion of them until the composition exercise, which follows.

Earnings Decomposition

Next I present the results of a standard AKM-style earnings decomposition exercise, with emphasis on the firm-specific components. Table XXIV reports duration-weighted summaries of the components of log earnings: time-varying worker characteristics ($X\hat{\beta}$), worker effects ($\hat{\theta}$), starting pay ($\hat{\psi}$), linear and quadratic tenure terms ($\hat{\gamma}_1$ and $\hat{\gamma}_2$), and the sample residual ($\hat{\varepsilon}$).

Worker effects and starting pay both have mean values of zero; this result stems from the construction of the realized mobility networks in the estimation procedure. The linear seniority component is, on average, large and positive (0.22) while quadratic component is small and negative (less than -0.0001 in magnitude). These figures demonstrate the positive, but diminishing returns of seniority. The worker effect, starting pay and seniority effects all have large standard deviations and are highly correlated with log earnings.

Additionally, $\hat{\theta}$ is highly correlated with both $\hat{\psi}$ and $\hat{\gamma}$ (0.30 and 0.13). These results not only support recent findings of assortative matching in Brazil, as noted by Lopes de Melo (2013) and Menezes-Filho et al. (2008), but in Germany as well (Card et al. 2013).

Pay Structure and Turnover Outcomes

Now I turn to the analysis of the link between a firm's pay structure and its hiring and separation rates. First I present descriptive statistics on starting pay, returns to seniority and turnover. Then I report results from the estimation of (3.3), which empirically relates turnover to starting pay and the returns to seniority.

Descriptive Analysis

Figures A.3, A.4 and A.5 show the distributions of the estimated starting pay, linear tenure effects and quadratic tenure effects. The ranges of these figures have been trimmed to report the inner 95 percent of observations. Figure A.3 reveals a symmetric distribution for starting pay centered near zero with the majority of observations between -1 and 2; this lower value represents firms that offer starting pay that is roughly half of the average starting pay, while the higher value indicates firms that have starting pay that is almost 200 percent greater than the average starting pay. The distribution of linear tenure effects is depicted in figure A.4. Most observations lie between 0 – a completely flat seniority profile – and 0.2 – a linear increase in earnings of about 20 percent, year over year. In figure A.5 the overwhelming majority of firms have quadratic tenure effects between zero and -0.005, suggesting that tenure profiles are close to linear for the typical firm. These three figures taken together help to clarify the amount of variation actually contained within each pay characteristic; starting pay and linear tenure exhibit a great deal of variation, while the quadratic tenure term exhibits less.

Using the estimated tenure coefficients, $\hat{\gamma}_1$ and $\hat{\gamma}_2$, I compute the value of tenure corresponding to the maximum of the seniority profile as:

$$\hat{P}_j = \frac{\hat{\gamma}_{1j}}{2\hat{\gamma}_{2j}}.$$

The absolute value of $\hat{P}_j$ is the number of years that an employee will receive positive marginal to remaining with his or her current firm, j. Figure **??** plots the distribution of $\hat{P}$. Most values for $\hat{P}$ are small, indicating that workers in most firms receive increases in pay from increased seniority for only a few years. Further, the distribution of $\hat{P}$ implies a great deal of heterogeneity in the combination of how firms combine the linear and quadratic retursn to seniority to develop specific pay contracts.

Table XXV reports summary statistics and simple correlations of $\hat{\psi}$, $\hat{\gamma}_1$, $\hat{\gamma}_2$, and $\hat{P}$ and the correlations with turnover and firms' size and skill concentration. The mean values presented here are slightly different from those in table XXIV because the previous table reported how much each component contributes to mean, per worker-year, earnings, while table XXV reports the mean coefficient values per firm. The mean starting pay is about -0.05, meaning the typical firm pays about 5 percent less than the market wage to a new employee. However, the mean linear return to seniority is about 0.05, meaning the 5 percent loss in starting pay is made up for by the end of the first year and the mean quadratic returns to seniority are negative and very small, indicating again that returns to seniority in the typical firm are very close to linear. The mean value for $\hat{P}$ is about -13.4, meaning employees in the typical firm see increases in pay as a result of seniority for more than 13 years. These estimated firm-specific coefficients also have large standard deviations, which is important for the estimability of (3.3). The average firm in the sample employs over 1,100 workers and about 20 percent of its workers are high-skilled, meaning the worker has an education level beyond high school. Hire, fire and quit rates are as reported in table XXII.

Among the estimated coefficients, $\hat{\psi}$ and $\hat{\gamma}_1$ exhibit a large, negative correlation of about -0.23; this correlation indicates the trade-off between starting pay and returns to seniority present throughout the economic literature, notably Lazear (1979). The correlation between $\hat{\gamma}_1$ and $\hat{\gamma}_2$ indicates a similar tradeoff between increased linear returns to seniority and linearity in returns to seniority with a large correlation of -0.74. Among the other firm characteristics on the right-hand side of (3.3), the share of high skill workers in a firm has a strong positive correlation with starting pay at 0.28. The simple correlations between turnover rates and the estimated pay coefficients indicate that all turnover measures decrease with increases in starting pay, increase with increases in returns to seniority and decrease as returns to seniority become more linear.

Table XXVI replicates table XXV by ownership status; with the top panel describes private, for profit firms, the second panel describes public firms and government agencies, the third describes private, non-profit firms and the last panel describes firms with individual owners (sole proprietorships). Across the different ownership types the mean values of $\hat{\psi}$, $\hat{P}$ and turnover rates all vary considerably. Starting pay varies from zero for the private, for-profit firms to -0.13 for public firms to -0.2 for sole proprietorships. The amount of time before returns to seniority are maximized is close among the three types of private firms, between 7.5 and 8.5 years, but is over 27 years for government firms. Turnover measures are greatest for individually owned firms and lowest for government organizations by at least half for all measures. Government firms also standout from other types of firm in the correlations with $\hat{\psi}$; the correlation with $\hat{\gamma}_1$ (-0.11) is almost one-third the size of the correlation for private, for profit firms (-0.30). Correlations between $\hat{\psi}$ and the three turnover measures are very different for public firms than they are for other categories; magnitudes are much smaller (-0.09, 0.05 and 0.04) for the hire, fire and quit rate correlations, and the signs for fire and quit rates are different, compared to private, for profit firms (-0.33, -0.20 and -0.34). These differences help to illustrate the differences between the four firm types, and indicate that the relationship between the three estimated coefficients and the three turnover measures may vary across these ownership types.

Baseline Turnover Regressions

Table XXVII reports results from the estimation of (3.3) for the hire, fire and quite rates. To understand how the inclusion (or exclusion) of other variables may affect my analysis, I estimate three specifications of the model. First, I regress turnover outcomes on $\hat{\psi}$, $\hat{\gamma}_1$ and $\hat{\gamma}_2$ separately (columns 1, 4 and 7). However, because $\hat{\psi}$, $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are part of a package that is jointly determined, I then estimate (3.3) first excluding controls (columns 2, 5 and 8), and then adding industry, firm size and skill intensity covariates (columns 3, 6 and 9). Given the descriptive statistics in table XXV, I would expect negative coefficient estimates between turnover measures and starting pay and quadratic returns to seniority, and positive coefficient estimates between turnover measures and linear returns to seniority.

The results of the bivariate regressions listed in columns 1, 4 and 7 are characterized by a lack of precision and coherence. Though the signs of the coefficient estimate align with those of the simple correlations in Table XXVII, the lack of precision in the estimated coefficients $\hat{\gamma}_1$ and, especially, $\hat{\gamma}_2$ make inference difficult. Estimating the effect of starting pay and seniority profiles simultaneously on turnover rates, without controls, similarly results in estimated effects that lack precision and interpretive coherence, as seen in columns 2, 5 and 8. Indeed, model 2 only accounts for 3.6 percent of the variation within firm hire rates; the explanatory power of the model without controls is even lower for fire and quit rates.

Including controls for firm size, employee skill intensity and industry brings precision and coherence to the estimated results. Column 3 indicates a negative relationship between starting pay and hire rates of -0.08, meaning that a ten percent increase in starting pay would result in about a one percent decrease in hire rates. Coefficient estimates for seniority profile components are small and not statistically different from zero. However, this makes sense when considering that the marginal cost to the firm for a new hire is the starting pay, not the returns to seniority. Additionally, this specification explains 34 percent of the variation in hire rates across sampled firms, nearly ten times what was explained by the model without controls.

Column 6 reports regression results for hire rates on all three pay measures that are highly statistically significant and negative. The estimated coefficient for starting pay is about -0.02 and significant at the 0.1 percent level. Both seniority measures have estimated coefficients that are negative, about -0.01 and -0.004 for the linear and quadratic terms, and highly statistically significant. These terms indicate that as returns to seniority increase, both linearly and quadratically, fire rates decrease. These results are consistent with the theory that firms paying these larger premia may be better at securing high-quality matches when hiring, and thus do not need not initiate separations as often.

Column 9 reports the estimated coefficients of all three pay measures on quit rates, with controls. The estimated coefficients for starting pay and seniority profiles are all negative and highly significant, indicating that increased premia are related to decreased quit rates in sampled firms. In particular, starting pay has an estimated coefficient of about -0.04, meaning a ten percent increase in starting pay is associated with a 0.4 percent decrease in the firm's quit rate. Increasing the linear returns to seniority by ten percent would decrease the quit rate by about 0.1 percent. The intuition for these effects is straight forward: workers who are paid more are less likely to quit. While these relationships may seem quantitatively small, recall from table XXV that the mean firm has over 1,100 workers and hire and quit rates of about 13 and 5 percent. Taken together, these coefficients imply a ten percent change to starting pay may result in 3 fewer hires and 2 fewer separations per year in the average firm.

Turnover Regressions by Ownership Status

To verify that the results in the previous section are representative of the behavior of rational, profit-maximizing firms, I estimate (3.3) with all controls separately for firms in each ownership category. If the above results are indicative of economically rational behavior, the connection between pay structure and turnover should be strongest for private, for-profit firms.

Table XXVIII reports the results in four panels corresponding to ownership categories. Compared to the estimates for the whole sample, estimates for private, for profit firms are larger in magnitude and much more precise. The estimated coefficient for $\hat{\psi}$ in hire rates increases in magnitude from -0.08 in the whole sample to -0.13 for private for profit firms while the relationship between hire rates and both $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are small and statistically insignificant. This result provides further support for the interpretation above that increased starting pay, not returns to seniority, is the relevant cost borne by the firm in the hiring decision and increases in this cost result in decreased hire rates.

Compared to the whole sample, the estimated effect of starting pay on fire rates is about 3 percentage points greater (-0.05), the estimated linear seniority effect is about 1 percentage point greater (-0.02) and the estimated quadratic seniority effect is slightly larger at -0.005. For quit rates the coefficient estimate for $\hat{\psi}$ is -0.07, about 4 percentage points greater in magnitude than in the whole sample; the coefficient estimate for $\hat{\gamma}_1$ is -0.01, about 0.7 percentage points greater in magnitude and the coefficient estimate for $\hat{\gamma}_2$ is -0.003. As with hire rates, the increased estimated coefficient magnitudes in private, for-profit firms relative to those for the whole sample support the intuition that firms offering higher starting pay and steeper seniority profiles do so to reduce turnover. Securing better employees means firms have less need to initiate separations and higher pay means workers are less inclined to have a better outside option.

If the results of for-profit firms stem from their sensitivity to market forces, it follows that public and state-owned firms, including the military, state police and legislature should demonstrate weaker relationships between turnover measures and starting pay and returns to seniority, as a result of these firms' relative insulation from market pressures. The second panel reports the results from estimating (3.3) for public and state-owned entities, and the results are very different from those for private, for-profit firms. The only statistically significant coefficient estimates are for $\hat{\psi}$, and the magnitudes are no more than one-fifth the for-profit counterpart. In addition, the signs of estimated coefficients for $\hat{\psi}$ are positive in the separation regressions. For fire rates the estimated coefficient of starting pay is about 0.01 and for quit rates the estimated coefficient of starting pay is about 0.004.

These counterintuitive results may stem from the different objectives of these organizations. Many of the positions employed in government firms are appointed or elected positions and turnover decisions for these positions likely have little to do with starting pay or returns to tenure. Many, if not most, of these organizations are likely less affected by market pressures than for-profit firms, and may operate in such a manner as to wholly consume their allotted budget.

The estimated coefficients for private, non-profit firms, in the third panel, demonstrate further that the results of the whole sample reflect the behavior of rational profit-maximizing firms, but smaller in magnitude and, in the case of fire rates, not significant. In each case, non-profits have a large, highly significant estimated coefficient for the effect of quadratic returns to seniority, but insignificant estimated linear seniority effects. The estimated coefficients for starting pay are qualitatively similar to those for private, for-profit firms. Somewhat like government organizations, the objective function non-profits is unclear, and may vary greatly between firms.

The fourth panel of table XXVIII provides the estimation results for firms with individual owners. As with public organizations and non-profits, these firms do not closely resemble either the whole sample nor private for-profits in their estimation results. As noted earlier, these firms are among the worst managed in Brazil. As Bloom et al. (2012) point out, they are often motivated by familial obligations and outcomes other than profits. Further complicating inference is the small number of firms in this category.

The differences between these last two firm categories and private, for-profit firms only serves to highlight the findings in the baseline regressions as the result of rational economic agents. Private for-profit firms provide the textbook example of a rational economic agent and the resulting estimates from those firms closely resemble the results derived from the sample as a whole. Government organizations, non-profit firms and sole-proprietorships, however lack clear economic motivations and are often poorly managed, and as a result the estimates derived from these groups do not conform to the pattern produced by the whole sample.

3.6 Conclusion

In this paper I explore the relationship between the a firm's set of turnover outcomes – hire, fire and quit rates – and starting pay and seniority profiles. I characterize these relationships and compare the results to predictions from the literature. Our analysis begins by using employer-employee matched panel data from Brazil. These data provide universal coverage of the Brazilian formal labor market and describe in fine detail its firms and workers. These data are unique in that they allow me to accurately calculate job durations for every registered Brazilian worker and construct measures of turnover for each firm, and to control for firm size, skill composition, industry and ownership status.

I adapt the estimation method described in Abowd et al. (1999) to extract firm-specific coefficient estimates of starting pay, seniority and seniority squared from earnings regressions of Brazil's formal labor market. I then use these estimated firm-specific coefficients to define the elements of a firm's pay structure: starting pay, linear returns to tenure and quadratic returns to tenure. I estimate the relationship between the set of starting pay and seniority profile estimates and the set of turnover measures: hire, fire and quit rates.

The economic theory makes two specific claims regarding the relationship between turnover rates in general and both starting pay and seniority profiles. The first is that as starting pay increases, turnover measures should decrease. Increased starting pay implies that the firm is looking to hire high-quality employees and is more selective about job candidates. Simultaneously, increased pay indicates that the firm may have spent more resources to ensure higher match quality between the firm and its employees, reducing employer- and employee-initiated separations. The second is that as the slope of the tenure profile increases, fire rates and quit rates should decrease as well. This prediction follows the same logic as the first. Firms that are paying their workers more with seniority are seeing increased productivity – or other forms of better match quality – from longer tenured workers and are reluctant to fire them; workers who are paid more find that their outside option is not as attractive and are reluctant to quit.

In general, estimation results uphold the economic theory: as a firm's hire, fire and quit rates decrease as the individual elements of its pay structure increase. If a firm were to increase its starting pay, one could expect to see a decrease in its hiring rate; simultaneously one may expect a decrease in the firm's fire rate and a slightly larger decrease in its quit rate. However, the individual pay structure elements are themselves correlated; starting pay and returns to tenure, for example, exhibit a strong negative relationship. These findings are even stronger for firms which are more exposed to market shocks, such as private for-profit firms. For firms which are insulated from the market, such as governmental bodies, these results are weaker both statistically and economically. These results suggest that the estimated relationships are the actions of rational economic agents.

Bibliography

Abowd, J., Kramarz, F., Margolis, D. N. and Troske, K. R. (2001). *Journal of Japanese and International Economies* **15**: 419–436.

Abowd, J. M., Creecy, R. H. and Kramarz, F. (2002). Computing person and firm effects using linked longitudinal employer-employee data.

Abowd, J. M., Kramarz, F. and Margolis, D. N. (1999). High Wage Workers and High Wage Firms, *Econometrica* **67**(2): 251–333.

Adams, S. J. and Heywood, J. S. (2007). The Age of Hiring and Deferred Compensation: Evidence from Australia, *Economic Record* **83**(261): 174–190.

Agüero, J. M. and Marks, M. S. (2008). Motherhood and Female Labor Force Participation : Evidence from Infertility Shocks, *The American Economic Review* **98**(2): 500–504.

Akerlof, G. A. and Kranton, R. E. (2000). Economics and identity, *The Quarterly Journal of Economics* **115**(3): 715–753.

Blinder, A. S. (1973). Wage Discrimination: Reduced Form and Structural Estimates, *Journal of Human Resources* **8**(4): 436–456.

Bloom, N., Genakos, C., Sadun, R. and Van Reenen, J. (2012). Management Practices across Firms and Countries.

Breau, S. (2009). Exports and Local Labor Markets in the Greater Los Angeles Region, *Urban Geography* **30**(1): 40–62.

Caliendo, L., Monte, F. and Rossi-Hansberg, E. A. (2012). The Anatomy of French Production Hierarchies.

Card, D., Heining, J. and Kline, P. (2013). Workplace heterogeneity and the rise of west german wage inequality, *Quarterly Journal of Economics* **128**(3): 967–1015.

Cornwell, C., Rivera, J. M. and Schmutte, I. M. (2014). Racial Identity and the Workplace: Evidence from Job Changers in Brazil.

Cortez Reis, M. and Vianna Crespo, A. R. (2005). Race Discrimination in Brazil: An Analysis of the Age, Period and Cohort Effects.

Daniel, K. and Heywood, J. S. (2007). The Determinants of Hiring Older Workers: UK Evidence, *Labour Economics* **14**(1): 35–51.

Daymont, T. N. and Andrisani, P. J. (1984). Job Preferences, College Major, and the Gender Gap in Earnings, *Journal of Human Resources* **19**(3): 408–428.

Dostie, B. (2005). Job Turnover and the Returns to Seniority, *Journal of Business & Economic Statistics* **23**(2): 192–199.

Francis, A. M. and Tannuri-Pianto, M. (2012). Using Brazil's Racial Continuum to Examine the Short-Term Effects of Affirmative Action in Higher Education, *Journal of Human Resources* **47**(3): 754–784.

Francis, A. M. and Tannuri-Pianto, M. (2013). Endogenous Race in Brazil : Affirmative Action and the Construction of Racial Identity among Young Adults, *Economic Development and Cultural Change* (Forthcoming): 1–35.

Gasparini, L. and Tornarolli, L. (2009a). Labor Informality in Latin America and the Caribbean: Patterns and Trends from Household Survey Microdata, *Desarrollo Y Sociedad* **14**(63): 13–80.

Gasparini, L. and Tornarolli, L. (2009b). Labor Informality in Latin America and the Caribbean: Patterns and Trends from Household Survey Microdata*, *Desarrollo Y Sociedad* **63**(1): 13–80.

Gradín, C. (2009). Why is Poverty So High Among Afro-Brazilians? A Decomposition Analysis of the Racial Poverty Gap, *Journal of Development Studies* **45**(9): 1426–1452.
URL: *http://www.tandfonline.com/doi/abs/10.1080/00220380902890235*

Gradín, C. (2014). Race and Income Distribution: Evidence from the USA, Brazil and South Africa, *Review of Development Economics* **18**(1): 73–92.
URL: *http://doi.wiley.com/10.1111/rode.12070*

Horny, G., Mendes, R. and Berg, G. J. V. D. (2009). Job Durations with Worker and Firm Specific Effects: MCMC Estimation with Longitudinal Employer-Employee Data.

Htun, M. (2004). From "Racial Democracy" to Affirmative Action: Changing State Policy on Race in Brazil, *Latin American Research Review* **39**(1): 60–89.

Ichniowski, C. and Shaw, K. L. (2013). Insider Econometrics: Empirical Studies of How Management Matters, *in* R. Gibbons and J. Roberts (eds), *Handbook of Organizational Economics*, number December in *NBER Working Paper Series*, Princeton University Press, Princeton, NJ, chapter 7, pp. 263–314.

Jones, F. L. and Kelley, J. (1984). Decomposing Differences between Groups. A Caustionary Note on Measuring Discrimination, *Sociological Methods & Research* **12**(3): 323–343.

Krishna, P., Poole, J. P. and Senses, M. Z. (2012). Trade , Labor Market Frictions , and Residual Wage Inequality across Worker Groups, *American Economic Review* **102**(3): 417–423.

Lazear, E. P. (1979). Why Is There Mandatory Retirement ?, *Journal of Political Economy* **87**(6): 1261–1284.

Lazear, E. P. and Oyer, P. (2013). Personnel Economics, *in* R. Gibbons and J. Roberts (eds), *Handbook of Organizational Economics*, Princeton University Press, Princeton, NJ, chapter 12, pp. 479–519.

Leite, P. G. (2005). Race Discrimination or Inequality of Opportunities: The Brazilian Case.

Lopes de Melo, R. (2013). Firm Wage Differentials and Labor Market Sorting : Reconciling Theory and Evidence.

Loureiro, P. R., Carneiro, F. G. a. and Sachsida, A. (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.
URL: *http://www.emeraldinsight.com/10.1108/01443580410527114*

Madalozzo, R. and Martins, S. R. (2007). Gender Wage Gaps: Comparing the 80s, 90s and 00s in Brazil, *Revista de Economia e Administração* **6**(2): 141–156.

Marques Garcia, L., Ñopo, H. and Salardi, P. (2009). Gender and Racial Wage Gaps in Brazil 1996-2006: Evidence Using a Matching Comparisons Approach.

Matos, R. S. and Machado, A. F. (2006). Diferencial de Rendimento por Cor e Sexo no Brasil (1987-2001), *Econômica* **8**(1): 5–27.

Mendes, R., van den Berg, G. J. and Lindeboom, M. (2010). An empirical assessment of assortative matching in the labor market, *Labour Economics* **17**(6): 919–929.

Menezes-Filho, N. A., Muendler, M.-A. and Ramey, G. (2008). The Structure of Worker Compensation in Brazil, with a Comparison to France and the United States, *Review of Economics and Statistics* **90**(2): 324–346.

Moore, H. L. (1911). *Laws of Wages: An Essay in Statistical Economics*, Macmillan Company, New York.

Muendler, M.-A., Rauch, J. E. and Tocoian, O. (2012). Employee Spinoffs and Other Entrants: Stylized Facts from Brazil.

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets, *International Economic Review* **14**(3): 693–709.

Piras, C. and Ripani, L. (2005). The Effects of Motherhood on Wages and Labor Force Participation: Evidence from Bolivia, Brazil, Ecuador and Peru.

Rivera, J. M. (2014). Data Collection and Economic Analysis: Evidence from Brazilian Microdata.

Saperstein, A. and Penner, A. M. (2012). Racial fluidity and inequality in the United States, *American Journal of Sociology* **118**(3): 676–727.

Schönberg, U. (2007). Wage Growth Due to Human Capital Accumulation and Job Search: A Comparison between the United States and Germany, *Industrial and Labor Relations Review2* **60**(4): 562–586.

Schwartzman, L. F. (2007). Does Money Whiten? Intergenerational Changes in Racial Classification in Brazil, *American Sociological Review* **72**(December 2007): 940–963.

Telles, E. E. (2002). Racial ambiguity among the Brazilian population, *Ethnic and Racial Studies* **25**(3): 415–441.

Zwick, T. (2011). Consequences of Seniority Wages on the Employment Structure Consequences of Seniority Wages on the Employment Structure, *Industrial and Labor Relations Review* **65**(1): 108–125.

Appendix A

Tables and Figures

This section contains the tables and figures referred to throughout the paper, in the order of reference.

Table I: Industry and Occupation Categories and Descriptions

Variable	Description
Industry	
Ag/Fishing	Agriculture, fisheries, forestry; harvesting and exploration
Mining	Mining industries
Production	Production of intermediate and final goods
Utilities	Production and distribution of electricty, gas and water
Construction	Construction
Trade/Repair	Commercial trade/repair of automobiles, goods and households
Food/Lodging	Restaurants and hotels
Transp./Storage/Comm.	Transportation, storage and communications.
Finance/Banking	Financial intermediaries.
Real Estate	Real estate services, rentals, sales and property management
Defense/Security	Public defense and security
Education	Education, public and private
Health/Social Services	Health and social services
Other Social/Personal Services	Other collective, social and personal services
Domestic Services	Domestic services: housekeeping, cooking and childcare
Int'l Organization	Extraterritorial and international organizations
Occupation	
Military	Members of the armed forces, police and military firefighters
Public Administration/Mgmt	Senior members of public organizations and business managers
Professionals/Arts/Sci	Professionals in the sciences and arts
Mid-Level Techs	Technicians, middle-level
Admin/Clerical	Administrative service workers
Service/Vendors	Service employees, commercial vendors in stores and markets
Ag/Fish/Forestry	Farmers, forestry workers, hunters and fishermen
Production I	Production workers, industrial goods and services
Production II	Production workers, industrial goods and services
Repair/Maintenamce	Repair and maintenance workers

Sources: Comissão Nacional de Classificação, http://www.cnae.ibge.gov.br/estrutura.asp; Classificação Brasileira de Ocupações, http://www.mtecbo.gov.br/cbosite/pages/downloads

Table II: Sample Composition: RAIS 2003-2010, Full-time & Minimum Wage (+) Workers

	2003	2004	2005	2006	2007	2008	2009	2010
Race								
Indigenous	0.66	0.57	0.36	0.27	0.24	0.24	0.22	0.21
White	60.77	60.43	60.67	52.51	51.89	51.20	50.33	49.37
Black	4.51	4.57	4.71	4.44	4.50	4.58	4.53	4.49
Yellow	0.72	0.77	0.75	0.63	0.64	0.63	0.62	0.60
Brown	23.13	23.62	25.23	22.32	22.87	23.70	24.23	25.09
Not Identified	10.21	10.04	8.28	2.52	2.99	3.28	3.77	4.59
Ignored	0.00	0.00	0.00	17.31	16.87	16.37	16.31	15.70
Gender								
Male	61.65	61.38	61.04	60.67	60.42	60.23	59.78	59.43
Female	38.35	38.62	38.96	39.33	39.58	39.77	40.22	40.57
Education								
Not Reported	0.00	0.00	0.00	0.00	0.00	0.20	0.00	0.00
No Schooling	1.15	1.04	0.93	0.85	0.79	0.72	0.65	0.60
Some Elementary	29.16	27.15	25.23	23.45	21.97	20.66	19.15	17.93
Elementary	17.35	17.02	16.62	16.24	15.84	15.26	14.79	14.10
Some High School	8.25	8.31	8.25	8.22	8.10	7.94	7.76	7.62
High School	28.45	30.66	32.68	34.83	36.49	38.40	40.11	42.00
Some College	3.63	3.80	3.94	4.08	4.05	4.09	4.14	4.10
Bachelor's (+)	12.00	12.02	12.35	12.31	12.76	12.73	13.40	13.65
Industry								
Not Reported	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Ag/Fishing	5.69	5.89	5.77	5.56	5.32	5.07	4.77	4.43
Mining	0.42	0.45	0.44	0.50	0.48	0.51	0.49	0.46

Samples include all observations for workers ages 20-65, working at least 40 hours per week and earning at least the minimum wage.
Duplicate worker-plant observations are omitted.
Continued on next page.

Table II – Continued from Previous Page

	2003	2004	2005	2006	2007	2008	2009	2010
Production	18.37	18.70	18.78	18.92	19.02	18.94	18.18	17.75
Utilities	0.63	0.61	0.59	0.57	0.56	0.53	0.52	0.49
Construction	4.98	4.82	4.88	5.18	5.37	6.14	6.56	7.25
Trade/Repair	18.03	18.34	18.68	18.65	18.97	19.15	19.44	19.65
Food/Lodging	3.38	3.39	3.46	3.54	3.63	3.70	3.79	3.86
Transp./Storage/Comm.	5.05	5.02	5.06	5.06	5.09	5.10	5.20	5.25
Finance/Banking	1.80	1.74	1.70	1.74	1.74	1.73	1.66	1.63
Real Estate	12.10	12.07	12.07	12.03	12.23	12.24	12.35	12.75
Defense/Security	19.48	19.06	18.74	18.31	17.99	17.50	17.52	17.05
Education	2.54	2.54	2.50	2.86	2.66	2.70	2.72	2.72
Health/Social Services	3.49	3.40	3.44	3.35	3.28	3.21	3.33	3.31
Other Social/Personal Services	4.02	3.93	3.85	3.67	3.60	3.45	3.43	3.38
Domestic Services	0.03	0.04	0.03	0.03	0.04	0.03	0.03	0.01
Int'l Organization	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.02
Occupation								
Military	1.84	1.73	1.58	1.52	0.00	1.28	1.29	1.21
Public Administration/Mgmt	4.14	3.82	3.91	3.96	4.11	3.95	4.20	4.15
Professionals/Arts/Sci	8.05	8.02	8.05	8.14	8.06	8.13	8.23	8.25
Mid-level Techs	9.99	9.95	9.73	9.66	9.64	9.61	9.72	9.58
Admin	17.86	17.89	18.03	18.03	18.17	18.33	18.31	18.58
Service/Vendors	22.99	22.92	23.12	23.24	23.15	24.01	24.20	23.98
Ag/Fish/Forestry	6.28	6.46	6.26	6.07	5.90	5.60	5.23	4.79
Production I	21.17	21.42	21.36	21.48	21.75	22.62	22.46	23.23
Production II	3.87	3.94	4.04	3.97	3.95	3.93	3.83	3.71
Repair/Maintenance	3.81	3.86	3.93	3.94	3.98	2.54	2.54	2.52

Samples include all observations for workers ages 20-65, working at least 40 hours per week and earning at least the minimum wage.
Duplicate worker-plant observations are omitted.
Continued on next page.

Table II – Continued from Previous Page

	2003	2004	2005	2006	2007	2008	2009	2010
Not Reported	0.00	0.00	0.00	0.00	1.31	0.00	0.00	0.00
Age (years)	34.96	35.01	35.06	35.18	35.26	35.30	35.47	35.52
Income (Nominal R$)	927.56	1,008.47	1,075.23	1,158.34	1,231.24	1,336.68	1,448.94	1,558.52
Observations - Raw	34,075,265	35,828,595	38,000,728	40,218,116	42,751,207	45,885,972	47,583,825	50,789,219

Samples include all observations for workers ages 20-65, working at least 40 hours per week and earning at least the minimum wage. Duplicate worker-plant observations are omitted.

Table III: Sample Composition: PNAD 2003-2011, Full-time & Minimum Wage (+) Workers

	2003	2004	2005	2006	2007	2008	2009	2011
Race								
Indigenous	0.14	0.17	0.17	0.25	0.25	0.23	0.21	0.30
White	61.58	60.61	58.69	58.39	57.31	55.54	55.66	54.22
Black	6.12	6.24	6.89	7.31	7.97	7.48	7.44	8.75
Yellow	0.60	0.53	0.71	0.66	0.59	0.70	0.53	0.62
Brown	31.56	32.44	33.54	33.39	33.88	36.05	36.17	36.11
Not Identified	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Ignored	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Gender								
Male	58.74	58.71	58.69	58.25	58.41	58.30	57.90	57.27
Female	41.26	41.29	41.31	41.72	41.59	41.70	42.10	42.73
Education								
Not Reported	0.75	0.69	0.70	0.52	0.46	0.45	0.35	0.21
No Schooling	3.87	3.72	3.48	3.26	3.37	3.39	3.06	4.43
Some Elementary	28.64	27.24	26.40	25.36	24.35	22.50	21.95	18.56
Elementary	10.02	10.15	9.70	9.40	10.15	9.55	8.90	10.04
Some High School	5.51	5.52	5.48	5.55	5.23	5.68	5.50	5.10
High School	29.81	30.93	32.15	33.03	32.97	34.16	34.63	35.86
Some College	7.29	7.45	7.62	7.81	8.06	8.49	8.66	8.61
Bachelor's (+)	14.10	14.29	14.47	15.07	15.40	15.75	16.95	17.20
Industry								
Not Reported	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Ag/Fishing	4.87	5.17	5.16	5.39	5.62	5.33	5.26	4.87
Mining	0.53	0.60	0.53	0.60	0.63	0.61	0.65	0.58

Samples include all observations for workers ages 20-65, working at least 40 hours per week and earning at least the minimum wage, weighted as indicated.

Continued on next page.

Table III – Continued from Previous Page

	2003	2004	2005	2006	2007	2008	2009	2011
Production	18.34	19.13	18.82	18.38	18.82	18.64	17.63	16.01
Utilities	0.85	0.85	0.82	0.86	0.77	0.75	0.75	0.59
Construction	4.04	4.03	4.29	4.28	4.45	5.20	5.18	5.96
Trade/Repair	17.54	17.14	17.96	17.93	17.94	17.34	17.99	18.37
Food/Lodging	3.11	3.08	3.15	3.29	3.23	3.22	3.37	4.49
Transp./Storage/Comm.	6.06	6.06	6.00	5.87	6.07	6.32	5.99	6.70
Finance/Banking	2.32	2.16	2.05	2.06	2.13	2.09	1.94	1.97
Real Estate	8.40	8.32	8.53	8.62	8.40	8.71	8.83	9.79
Defense/Security	9.41	9.36	9.03	8.94	8.68	8.68	8.87	8.70
Education	10.11	10.11	9.79	9.67	9.59	9.53	9.54	8.53
Health/Social Services	6.28	6.02	5.97	6.05	5.96	6.11	6.00	5.81
Other Social/Personal Services	3.07	3.09	3.02	3.21	2.95	3.05	3.17	2.91
Domestic Services	5.01	4.80	4.81	4.82	4.65	4.34	4.80	4.63
Int'l Organization	0.07	0.09	0.06	0.03	0.12	0.08	0.04	0.08
Occupation								
Military	1.15	1.03	0.96	0.96	0.81	0.98	0.96	0.91
Public Administration/Mgmt	8.22	7.90	8.24	8.20	7.50	7.53	7.22	6.21
Professionals/Arts/Sci	10.00	9.86	9.82	10.20	10.18	10.41	11.43	11.86
Mid-level Techs	11.22	11.07	11.10	10.92	11.12	10.56	10.44	9.34
Admin	13.29	13.58	13.50	13.39	13.26	13.78	14.00	12.57
Service/Vendors	27.29	27.32	27.29	27.73	27.63	26.46	26.96	29.63
Ag/Fish/Forestry	4.74	4.92	5.00	5.18	5.50	4.93	5.04	4.62
Production I	18.85	19.30	18.92	18.58	18.87	19.72	18.73	19.31
Production II	2.76	2.94	3.00	2.83	2.99	3.23	2.97	2.74
Repair/Maintenance	2.05	2.09	2.17	2.01	2.15	2.39	2.25	2.83

Samples include all observations for workers ages 20-65, working at least 40 hours per week and earning at least the minimum wage, weighted as indicated.

Continued on next page.

Table III – Continued from Previous Page

	2003	2004	2005	2006	2007	2008	2009	2011
Not Reported	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Age (years)	36.62	36.68	36.72	37.01	37.12	37.19	37.31	37.53
Income (Nominal R$)	918.73	959.39	1,048.17	1,139.02	1,188.11	1,276.70	1,359.98	1,596.14
Observations - Raw	74,026	78,442	83,088	86,556	87,299	90,038	93,955	89,596
Observations - Weighted	34,607,990	36,508,684	38,401,812	40,274,874	42,230,929	44,831,420	46,108,013	49,829,595

Samples include all observations for workers ages 20-65, working at least 40 hours per week and earning at least the minimum wage, weighted as indicated.

Table IV: Sample Composition: RAIS-PNAD Differences 2003-2010/11, Full-time & Minimum Wage (+) Workers

	2003	2004	2005	2006	2007	2008	2009	2010/11	Avg
Race									
Indigenous	0.52	0.40	0.19	0.02	-0.01	0.01	0.01	-0.09	0.13
White	-0.81	-0.18	1.98	-5.88	-5.42	-4.34	-5.33	-4.85	-3.10
Black	-1.61	-1.67	-2.18	-2.87	-3.47	-2.90	-2.91	-4.26	-2.73
Yellow	0.12	0.24	0.04	-0.03	0.05	-0.07	0.09	-0.02	0.05
Brown	-8.43	-8.82	-8.31	-11.07	-11.01	-12.35	-11.94	-11.02	-10.37
Not Identified	10.21	10.03	8.28	2.52	2.99	3.28	3.77	4.59	5.71
Ignored	0.00	0.00	0.00	17.31	16.87	16.37	16.31	15.70	10.32
Gender									
Male	2.91	2.67	2.35	2.42	2.01	1.93	1.88	2.16	2.29
Female	-2.91	-2.67	-2.35	-2.39	-2.01	-1.93	-1.88	-2.16	-2.29
Education									
Not Reported	-0.75	-0.69	-0.70	-0.52	-0.46	-0.25	-0.35	-0.21	-0.49
No Schooling	-2.72	-2.68	-2.55	-2.41	-2.58	-2.67	-2.41	-3.83	-2.73
Some Elementary	0.52	-0.09	-1.17	-1.91	-2.38	-1.84	-2.80	-0.63	-1.29
Elementary	7.33	6.87	6.92	6.84	5.69	5.71	5.89	4.06	6.16
Some High School	2.74	2.79	2.77	2.67	2.87	2.26	2.26	2.52	2.61
High School	-1.36	-0.27	0.53	1.80	3.52	4.24	5.48	6.14	2.51
Some College	-3.66	-3.65	-3.68	-3.73	-4.01	-4.40	-4.52	-4.51	-4.02
Bachelor's (+)	-2.10	-2.27	-2.12	-2.76	-2.64	-3.02	-3.55	-3.55	-2.75
Industry									
Not Reported	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Ag/Fishing	0.82	0.72	0.61	0.17	-0.30	-0.26	-0.49	-0.44	0.10
Mining	-0.11	-0.15	-0.09	-0.10	-0.15	-0.10	-0.16	-0.12	-0.12
Production	0.03	-0.43	-0.04	0.54	0.20	0.30	0.55	1.74	0.36
Utilities	-0.22	-0.24	-0.23	-0.29	-0.21	-0.22	-0.23	-0.10	-0.22
Construction	0.94	0.79	0.59	0.90	0.92	0.94	1.38	1.29	0.97
Trade/Repair	0.49	1.20	0.72	0.72	1.03	1.81	1.45	1.28	1.09
Food/Lodging	0.27	0.31	0.31	0.25	0.40	0.48	0.42	-0.63	0.23
Transp./Storage/Comm.	-1.01	-1.04	-0.94	-0.81	-0.98	-1.22	-0.79	-1.45	-1.03
Finance/Banking	-0.52	-0.42	-0.35	-0.32	-0.39	-0.36	-0.28	-0.34	-0.37
Real Estate	3.70	3.75	3.54	3.41	3.83	3.53	3.52	2.96	3.53

Difference calculated as percentage point difference of RAIS less PNAD. Samples include all observations for workers ages 20-65, working at least 40 hours per week and earning at least the minimum wage. Duplicate worker-plant observations are omitted from RAIS. PNAD samples are weighted.

* Differences reported in percent, not percentage points.

Continued on next page.

Table IV – Continued from Previous Page

	2003	2004	2005	2006	2007	2008	2009	2010/11	Avg
Defense/Security	10.07	9.70	9.71	9.37	9.31	8.82	8.65	8.35	9.25
Education	-7.57	-7.57	-7.29	-6.81	-6.93	-6.83	-6.82	-5.81	-6.95
Health/Social Services	-2.79	-2.62	-2.53	-2.70	-2.68	-2.90	-2.67	-2.50	-2.67
Other Social/Personal Services	0.95	0.84	0.83	0.46	0.65	0.40	0.26	0.47	0.61
Domestic Services	-4.98	-4.76	-4.78	-4.79	-4.61	-4.31	-4.77	-4.62	-4.70
Int'l Organization	-0.06	-0.08	-0.05	-0.01	-0.11	-0.07	-0.03	-0.06	-0.06
Occupation									
Military	0.69	0.70	0.62	0.56	-0.81	0.30	0.33	0.30	0.34
Public Administration/Mgmt	-4.08	-4.08	-4.33	-4.24	-3.39	-3.58	-3.02	-2.06	-3.60
Professionals/Arts/Sci	-1.95	-1.84	-1.77	-2.06	-2.12	-2.28	-3.20	-3.61	-2.35
Mid-level Techs	-1.23	-1.12	-1.37	-1.26	-1.48	-0.95	-0.72	0.24	-0.99
Admin	4.57	4.31	4.53	4.64	4.91	4.55	4.31	6.01	4.73
Service/Vendors	-4.30	-4.40	-4.17	-4.49	-4.48	-2.45	-2.76	-5.65	-4.09
Ag/Fish/Forestry	1.54	1.54	1.26	0.89	0.40	0.67	0.19	0.17	0.83
Production I	2.32	2.12	2.44	2.90	2.88	2.90	3.73	3.92	2.90
Production II	1.11	1.00	1.04	1.14	0.96	0.70	0.86	0.97	0.97
Repair/Maintenance	1.76	1.77	1.76	1.93	1.83	0.15	0.29	-0.31	1.15
Not Reported	0.00	0.00	0.00	0.00	1.31	0.00	0.00	0.00	0.16
Age*	-4.54	-4.56	-4.50	-4.95	-5.02	-5.06	-4.91	-5.36	-4.86
Income (R$)*	0.96	5.12	2.58	1.70	3.63	4.70	6.54	-2.36	2.86

Difference calculated as percentage point difference of RAIS less PNAD. Samples include all observations for workers ages 20-65, working at least 40 hours per week and earning at least the minimum wage. Duplicate worker-plant observations are omitted from RAIS. PNAD samples are weighted.

* Differences reported in percent, not percentage points.

Table V: Oaxaca-Blinder Decompositions, RAIS and PNAD 2003: Full-time, Minimum Wage (+) Workers

	RAIS - 2003			PNAD
Estimated Earnings	A	B	C	2003
$\widehat{\ln Y}_{NW}$	6.390***	6.247***	6.367***	6.198***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
$\widehat{\ln Y}_{W}$	6.472***	6.521***	6.487***	6.596***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Earnings Gap	-0.082***	-0.274***	-0.119***	-0.398***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Decomposition				
Characteristic	-0.015***	-0.134***	-0.038***	-0.229***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Coefficient	-0.099***	-0.158***	-0.110***	-0.180***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Interaction	0.031***	0.018***	0.028***	0.011***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Observations	28,075,990	28,075,990	28,075,990	60,920

Standard errors in parentheses.

*p<0.05 **p<0.01 ***p<0.001

Full-time, minimum wage (+) earning workers, aged 20-65 years. All models include age, age^2, gender, education occupation and industry as control variables. Model A includes only non-reported race as non-white. Model B includes non-reported race as white. Model C uses workers' modal reported race.

Table VI: Oaxaca-Blinder Decompositions, RAIS and PNAD 2004: Full-time, Minimum Wage (+) Workers

	RAIS - 2004			PNAD
Estimated Earnings	A	B	C	2004
$\widehat{\ln Y}_{NW}$	6.429***	6.297***	6.407***	6.272***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
$\widehat{\ln Y}_{W}$	6.541***	6.585***	6.556***	6.646***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Earnings Gap	-0.112***	-0.288***	-0.148***	-0.374***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Decomposition				
Characteristic	-0.027***	-0.136***	-0.049***	-0.216***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Coefficient	-0.114***	-0.168***	-0.125***	-0.163***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Interaction	0.029***	0.016***	0.026***	0.005***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Observations	30,712,242	30,712,242	30,712,242	65,043

Standard errors in parentheses.

*p<0.05 **p<0.01 ***p<0.001

Full-time, minimum wage (+) earning workers, aged 20-65 years. All models include age, age^2, gender, education occupation and industry as control variables. Model A includes non-reported race as non-white. Model B includes non-reported race as white. Model C uses workers' modal reported race.

Table VII: Oaxaca-Blinder Decompositions, RAIS and PNAD 2005: Full-time, Minimum Wage (+) Workers

	RAIS - 2005			PNAD
Estimated Earnings	A	B	C	2005
$\widehat{\ln Y}_{NW}$	6.547***	6.429***	6.531***	6.375***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
$\widehat{\ln Y}_{W}$	6.648***	6.690***	6.659***	6.740***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Earnings Gap	-0.101***	-0.261***	-0.128***	-0.365***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Decomposition				
Characteristic	-0.022***	-0.130***	-0.043***	-0.220***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Coefficient	-0.104***	-0.147***	-0.111***	-0.157***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Interaction	0.025***	0.016***	0.026***	0.011***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Observations	31,336,795	31,336,795	31,336,795	68,825

Standard errors in parentheses.

*p<0.05 **p<0.01 ***p<0.001

Full-time, minimum wage (+) earning workers, aged 20-65 years. All models include age, age^2, gender, education occupation and industry as control variables. Model A includes non-reported race as non-white. Model B includes non-reported race as white. Model C uses workers' modal reported race.

Table VIII: Oaxaca-Blinder Decompositions, RAIS and PNAD 2006: Full-time, Minimum Wage (+) Workers

	RAIS - 2006			PNAD
Estimated Earnings	A	B	C	2006
$\widehat{\ln Y}_{NW}$	6.701***	6.494***	6.647***	6.467***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
$\widehat{\ln Y}_{W}$	6.701***	6.784***	6.738***	6.834***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Earnings Gap	-0.000	-0.291***	-0.091***	-0.368***
	(0.0003)	(0.0002)	(0.0003)	(0.0002)
Decomposition				
Characteristic	0.0145***	-0.176***	-0.0196***	-0.221***
	(0.0006)	(0.0002)	(0.0002)	(0.0002)
Coefficient	-0.103***	-0.171***	-0.104***	-0.160***
	(0.0002)	(0.0003)	(0.0002)	(0.0002)
Interaction	0.089***	0.056***	0.032***	0.014***
	(0.0005)	(0.0003)	(0.0001)	(0.0001)
Observations	32,901,938	32,901,938	32,901,938	71,246

Standard errors in parentheses.

*p<0.05 **p<0.01 ***p<0.001

Full-time, minimum wage (+) earning workers, aged 20-65 years. All models include age, age^2, gender, education occupation and industry as control variables. Model A includes non-reported race as non-white. Model B includes non-reported race as white. Model C uses workers' modal reported race.

Table IX: Oaxaca-Blinder Decompositions, RAIS and PNAD 2007: Full-time, Minimum Wage (+) Workers

	RAIS - 2007			PNAD
Estimated Earnings	A	B	C	2007
$\widehat{\ln Y}_{NW}$	6.753***	6.559***	6.704***	6.537***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
$\widehat{\ln Y}_{W}$	6.765***	6.844***	6.800***	6.899***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Earnings Gap	-0.012***	-0.286***	-0.097***	-0.362***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Decomposition				
Characteristic	-0.001*	-0.180***	-0.025***	-0.211***
	(0.0004)	(0.0002)	(0.0002)	(0.0002)
Coefficient	-0.098***	-0.156***	-0.101***	-0.167***
	(0.0002)	(0.0004)	(0.0002)	(0.0002)
Interaction	0.087***	0.050***	0.029***	0.017***
	(0.0004)	(0.0004)	(0.0001)	(0.0001)
Observations	35,371,174	35,371,174	35,371,174	72,401

Standard errors in parentheses.

*$p<0.05$ **$p<0.01$ ***$p<0.001$

Full-time, minimum wage (+) earning workers, aged 20-65 years. All models include age, age^2, gender, education occupation and industry as control variables. Model A includes non-reported race as non-white. Model B includes non-reported race as white. Model C uses workers' modal reported race.

Table X: Oaxaca-Blinder Decompositions, RAIS and PNAD 2008: Full-time, Minimum Wage (+) Workers

	RAIS - 2008			PNAD
Estimated Earnings	A	B	C	2008
$\widehat{\ln Y}_{NW}$	6.837***	6.651***	6.787***	6.638***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
$\widehat{\ln Y}_{W}$	6.849***	6.927***	6.885***	6.975***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Earnings Gap	-0.012***	-0.276***	-0.098***	-0.336***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Decomposition				
Characteristic	-0.012***	-0.163***	-0.013***	-0.198***
	(0.0004)	(0.0002)	(0.0002)	(0.0002)
Coefficient	-0.102***	-0.197***	-0.114***	-0.156***
	(0.0002)	(0.0004)	(0.0002)	(0.0002)
Interaction	0.102***	0.084***	0.029***	0.018***
	(0.0004)	(0.0004)	(0.0001)	(0.0001)
Observations	38,238,687	38,238,687	38,238,687	74,486

Standard errors in parentheses.

*p<0.05 **p<0.01 ***p<0.001

Full-time, minimum wage (+) earning workers, aged 20-65 years. All models include age, age^2, gender, education occupation and industry as control variables. Model A includes non-reported race as non-white. Model B includes non-reported race as white. Model C uses workers' modal reported race.

Table XI: Oaxaca-Blinder Decompositions, RAIS and PNAD 2009: Full-time, Minimum Wage (+) Workers

	RAIS - 2009			PNAD
Estimated Earnings	A	B	C	2009
$\widehat{\ln Y}_{NW}$	6.918***	6.744***	6.870***	6.719***
	(0.0002)	(0.0002)	(0.0002)	(0.0001)
$\widehat{\ln Y}_{W}$	6.933***	7.007***	6.970***	7.043***
	(0.0002)	(0.0001)	(0.0002)	(0.0002)
Earnings Gap	-0.015***	-0.263***	-0.100***	-0.324***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Decomposition				
Characteristic	-0.026***	-0.171***	-0.031***	-0.190***
	(0.0004)	(0.0002)	(0.0002)	(0.0002)
Coefficient	-0.082***	-0.163***	-0.093***	-0.149***
	(0.0002)	(0.0004)	(0.0002)	(0.0002)
Interaction	0.093***	0.071***	0.024***	0.015***
	(0.0004)	(0.0003)	(0.0001)	(0.0001)
Observations	39,628,675	39,628,675	39,628,675	78,126

Standard errors in parentheses.

*p<0.05 **p<0.01 ***p<0.001

Full-time, minimum wage (+) earning workers, aged 20-65 years. All models include age, age^2, gender, education occupation and industry as control variables. Model A includes non-reported race as non-white. Model B includes non-reported race as white. Model C uses workers' modal reported race.

Table XII: Oaxaca-Blinder Decompositions, RAIS and PNAD 2010/11: Full-time, Minimum Wage (+) Workers

Estimated Earnings	RAIS - 2010 A	B	C	PNAD 2011
$\widehat{\ln Y}_{NW}$	6.989***	6.823***	6.939***	7.088***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
$\widehat{\ln Y}_{W}$	7.011***	7.083***	7.050***	7.306***
	(0.0001)	(0.0001)	(0.0001)	(0.0002)
Earnings Gap	-0.022***	-0.260***	-0.111***	-0.218***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Decomposition				
Characteristic	-0.025***	-0.170***	-0.035***	-0.126***
	(0.0004)	(0.0001)	(0.0002)	(0.0002)
Coefficient	-0.082***	-0.163***	-0.096***	-0.112***
	(0.0002)	(0.0003)	(0.0002)	(0.0002)
Interaction	0.085***	0.072***	0.020***	0.019***
	(0.0003)	(0.0003)	(0.0001)	(0.0001)
Observations	43,106,617	43,106,617	43,106,617	51,521

Standard errors in parentheses.

*p<0.05 **p<0.01 ***p<0.001

Full-time, minimum wage (+) earning workers, aged 20-65 years. All models include age, age^2, gender, education occupation and industry as control variables. Model A includes non-reported race as non-white. Model B includes non-reported race as white. Model C uses workers' modal reported race. PNAD earnings reported in log 2010 reais and are adjusted for real wage growth.

Table XIII: Evidence of Racial Inequality in Earnings, 2003-10

Dependent Variable: Log Wage (2003 reais)

	Baseline	+ Plant Effects	+ Worker Effects
Black	-0.160^{***}	-0.098^{***}	-0.001
	(0.0003)	(0.0002)	(0.0004)
Brown	-0.208^{***}	-0.076^{***}	-0.017^{***}
	(0.0001)	(0.0001)	(0.0002)
Some Elementary	-0.334^{***}	-0.163^{***}	-0.008^{***}
	(0.0002)	(0.0001)	(0.0003)
Elementary	-0.178^{***}	-0.095^{***}	-0.000
	(0.0002)	(0.0002)	(0.0003)
Some HS	-0.156^{***}	-0.082^{***}	-0.002^{***}
	(0.0002)	(0.0002)	(0.0003)
Some College	0.403^{***}	0.237^{***}	0.052^{***}
	(0.0003)	(0.0003)	(0.0004)
Bachelors (+)	0.983^{***}	0.705^{***}	0.138^{***}
	(0.0003)	(0.0002)	(0.0004)
Experience	0.039^{***}	0.023^{***}	0.005^{***}
	(0.0000)	(0.0000)	(0.0000)
Experience^2	-0.068^{***}	-0.035^{***}	-0.008^{***}
	(0.0001)	(0.0000)	(0.0000)
Constant	6.183^{***}	6.295^{***}	6.384^{***}
	(0.0003)	(0.0002)	(0.0002)
R^2	0.3434	0.6764	0.9357
N	245, 210, 658	245, 210, 658	245, 210, 658

NOTE–The data are a panel of jobs from RAIS between 2003-2010 for workers age 16-65. The dependent variable is the natural logarithm of the average monthly wage in 2003 reais. I estimate the model separately for workers who are reported as *pardo* or *preto* on the origin job. I report controls for education, the wage on the origin job, sex. Column (2) adds plant effects, and column (3) absorbs worker effects.
Heteroskedasticity robust standard errors clustered by plant in parentheses. (*), (**), or (***) indicate the estimate is statistically different from zero at the 5, 1, and 0.1 percent level.

Table XIV: Characteristics of 2010 RAIS Sample

	RAIS 2010	
Variable	Freq.	Percent
Color or Race		
Indigenous	96,620	0.27
White	22,376,949	61.95
Black	2,021,984	5.60
Yellow 271,271	0.75	
Brown	11,352,471	31.43
Gender & Age		
Male	23,426,958	64.86
Age (years)†	34.3	-
Education		
No Schooling	227,058	0.63
Some Elementary	7,129,521	19.74
Elementary	5,611,783	15.54
Some High School	3,178,730	8.80
High School	15,476,901	42.85
Some University	1,432,472	3.97
Bachelor's (+)	3,062,830	8.48
Industry		
Ag/Fishing	1,945,128	5.39
Mining	164,576	0.46
Production	7,965,876	22.05
Utilities	182,644	0.51
Construction	3,193,594	8.84
Trade/Repair	9,052,152	25.06
Food/Lodging	1,729,897	4.79
Transp./Storage/Comm.	2,287,205	6.33
Finance/Banking	497,754	1.38
Real Estate	3,142,378	8.70
Defense/Soc.Sec.	5,395,321	14.94
Education	558,659	1.55
Health/Soc.Serv	1,061,254	2.94
Other Soc./Pers. Serv	1,306,458	3.62
Total	36,119,295	100

Source: RAIS 2010, MTE.

Full-time workers ages 20-65 with full-time employment. Occupation reclassified according to single digit CBO 2002, provided by the IBGE. †Monthly income and age reported reported in reais and years and difference is percentage difference.

Continued on next page.

Table XIV, Continued

	RAIS 2010	
Variable	Freq.	Percent
Domestic	5,578	0.02
Int'l Org	4,572	0.01
Occupation		
Military	3,819	0.01
Public Admin/Mgmt	1,204,695	3.34
Professionals/Arts/Sci	1,708,063	4.73
Mid-level Techs	2,642,037	7.31
Admin	6,213,993	17.2
Service/Vendors	9,419,447	26.08
Ag/Fish/Forestry	2,070,029	5.73
Production I	10,162,282	28.14
Production II	1,627,587	4.51
Repair/Mainten.	1,067,342	2.96
Monthly Income (R$)†	1,354	-
Total	36,119,295	100

Source: RAIS 2010, MTE.

Full-time workers ages 20-65 with full-time employment. Occupation reclassified according to single digit CBO 2002, provided by the IBGE. †Monthly income and age reported reported in reais and years and difference is percentage difference.

Table XV: Characteristics of Job-changers and Race-changers, 2010

	All Workers (%)	Job Changers (%)	Race Changers (%)
Race, color			
Indigenous	0.27	0.25	0.62
White	61.95	57.11	40.63
Black	5.60	5.26	8.86
Yellow	0.75	0.62	1.41
Brown	31.44	28.73	33.85
Education			
Some Elementary	19.75	19.78	19.97
Elementary	15.54	15.40	15.78
Some High School	8.80	9.74	9.66
High School	42.83	43.87	44.54
Some College	3.97	4.08	3.65
Bachelor's (+)	8.49	7.13	6.40
Age (years)	34.30	31.68	31.66
Observations	36,011,053	4,199,672	1,659.684

Source: RAIS 2010, MTE

Table XVI: Categories and Frequencies of Race Change, 2010

	White	Brown	Black	Total
White	1,724,113	437,910	56,612	2,218,635
	77.71	19.74	2.55	100.00
Brown	410,763	644,589	71,533	1,126,885
	36.45	57.20	6.35	100.00
Black	59,406	73,803	73,942	207,151
	28.68	35.63	35.69	100.00
Total	2,194,282	1,156,302	202,087	3,552,671
	61.76	32.55	5.69	100.00

Frequency of race changes among workers who change jobs in 2010. Table rows are disaggregated by race on the origin job. Columns indicate race on the destination job. The data are restricted to workers who start and end in one of three large race groups: White, Brown and Black.

Table XVII: Plant Occupation, Race, Age and Gender

Variable	Frequency	Variable	Frequency
Occupation		*Industry*	
Military	0.01	Ag/Fishing	9.61
Public Admin/Mgmt	4.89	Mining	0.25
Professionals/Arts/Sci	3.60	Production	10.18
Mid-level Techs	5.96	Utilities	0.15
Admin	19.91	Construction	3.37
Service/Vendors	33.70	Trade/Repair	40.90
Ag/Fish/Forestry	9.06	Food/Lodging	5.87
Production I	17.21	Transp./Storage/Comm.	4.63
Production II	2.66	Finance	1.41
Repair/Mainten.	2.97	Real Estate	11.76
Plant Size (workers employed)		Defense/Soc.Sec.	0.19
1 to 4	64.21	Education	1.95
5 to 9	17.93	Health/Soc.Serv	4.82
10 to 19	9.50	Other Soc./Pers. Serv.	4.79
20 to 49	5.21	Domestic	0.11
50 to 99	1.59	Int'l Org	0.01
100 to 249	0.95	Share White	68.45
250 to 499	0.35	Share Female	42.49
500 to 999	0.16	Mean Age (Yrs)	34.98
1,000 or more	0.10	Salary (R$)	1,064.48
Observations	2,772,993		

Source: RAIS 2010, MTE

All variables reported as percentages except age (years), salary (Reais) and total observations (plants).

Table XVIII: Explaining Change to White, 2010

	(1)	(2)	(3)
Education			
No Schooling	-0.166^{***}	-0.126^{***}	-0.083^{***}
	(0.0087)	(0.0056)	(0.0043)
< Elementary	-0.064^{***}	-0.080^{***}	-0.073^{***}
	(0.0077)	(0.0056)	(0.0043)
Elementary	-0.062^{***}	-0.079^{***}	-0.072^{***}
	(0.0078)	(0.0054)	(0.0042)
Some HS	-0.052^{***}	-0.063^{***}	-0.054^{***}
	(0.0068)	(0.0049)	(0.0040)
High School	-0.015^{**}	-0.035^{***}	-0.025^{***}
	(0.0067)	(0.0045)	(0.0043)
Origin log wage	0.016^{***}	-0.010^{***}	0.018^{***}
	(0.0047)	(0.0041)	(0.0025)
Destination log wage	0.011^{***}	0.005^{***}	0.004^{***}
	(0.0026)	(0.0019)	(0.0011)
Occupation, Industry, State Controls?	N	Y	Y
Plant Effects?	N	N	Y
Observations	1,427,530	1,427,530	1,427,530
R^2	0.0138	0.1101	0.6619

Estimates of a linear probability model for race change. The data are at the worker level. A worker is in the sample if he or she changed jobs during 2010. The dependent variable is an indicator = 1 if the worker changed race when he or she changed jobs. I estimate the model separately for workers who are reported as Brown or Black on the origin job. I report controls for education, the wage on the origin job, the wage on the destiation job. Column (2) adds controls for occupation, industry, and state of both origin and destination jobs. Column (3) adds plant effects.
Heteroskedasticity robust standard errors clustered by plant in parentheses. (*), (**), or (***) indicate the estimate is statistically different from zero at the 5, 1, and 0.1 percent level.

Table XIX: Plant Effects and the White Share of Employment, 2010

	(1) Plant Effect - To Black/Brown	(2) Plant Effect - To White
White share	-0.521^{***}	0.780^{***}
	(0.0014)	(0.0020)
Constant	0.281^{***}	-0.370
	(0.0295)	(0.0373)
Observations	548, 950	323, 115
R^2	0.642	0.556

Regression projecting plant effects estimated in in Table XVIII onto plant characteristics. I report the coefficient on the share white in the plant. The models also control for average wage, share male, and the share of workers with college, high school equivalence, and less than a high school education, as well as indicators for industry and municipality.
Heteroskedasticity robust standard errors clustered by plant in parentheses. (*), (**), or (***) indicate the estimate is statistically different from zero at the 5, 1, and 0.1 percent level.

Table XX: Relationship between Share White and Plant Effect, 2010

	(1)	(2)	(3)
Share White			
Decile 2	-0.087^{***}	-0.068^{***}	-0.033^{***}
	(0.0035)	(0.0036)	(0.0035)
Decile 3	0.073^{***}	0.051^{***}	0.041^{***}
	(0.0025)	(0.0024)	(0.0023)
Decile 4	0.157^{***}	0.133^{***}	0.138^{***}
	(0.0023)	(0.0023)	(0.0023)
Decile 5	0.247^{***}	0.232^{***}	0.245^{***}
	(0.0024)	(0.0024)	(0.0025)
Decile 6	0.326^{***}	0.323^{***}	0.371^{***}
	(0.0024)	(0.0024)	(0.0026)
Decile 7	0.422^{***}	0.430^{***}	0.493^{***}
	(0.0024)	(0.0024)	(0.0027)
Decile 8	0.515^{***}	0.530^{***}	0.615^{***}
	(0.0024)	(0.0024)	(0.0027)
Decile 9	0.742^{***}	0.735^{***}	0.812^{***}
	(0.0024)	(0.0024)	(0.0026)
Avg. Log Wage		-0.110^{***}	-0.047^{***}
		(0.0015)	(0.0017)
Male Share		0.036^{***}	0.023^{***}
		(0.0027)	(0.0035)
High School		0.072^{***}	0.154^{***}
		(0.0106)	(0.0102)
College		0.138^{***}	0.122^{***}
		(0.0115)	(0.0111)
Less Than		0.090^{***}	0.162^{***}
College		(0.0103)	(0.0101)
Constant	-0.344^{***}	0.285^{***}	-0.344^{***}
	(0.0018)	(0.0147)	(0.0371)
Worker Chars.?	N	Y	Y
Industry Controls?	N	N	Y
Municipality controls?	N	N	Y
Num. Obs.	323, 115	323, 115	323, 115
R^2	0.343	0.361	0.5641

Regression projecting plant effects estimated in in Table XVIII onto plant characteristics. The dependent variable is the plant effect in a linear probability model predicting whether a worker reported as *pardo/preto* in her first job is reported as *branco* in her subsequent job. Column (1) includes the share of workers in the plant who are white (not including the worker who changes). The white share is reported in deciles of the plant-size-weighted distribution, and these are included as indicators along with a constant. All observations are weighted by total employment in the plant at the start of the year. Column (2) adds other summaries of worker characteristics. Column (3) adds indicators that control for industry and the municipality where the plant is located.
Heteroskedasticity robust standard errors in parentheses. (*), (**), or (***) indicate the estimate is statistically different from zero at the 5, 1, and 0.1 percent level.

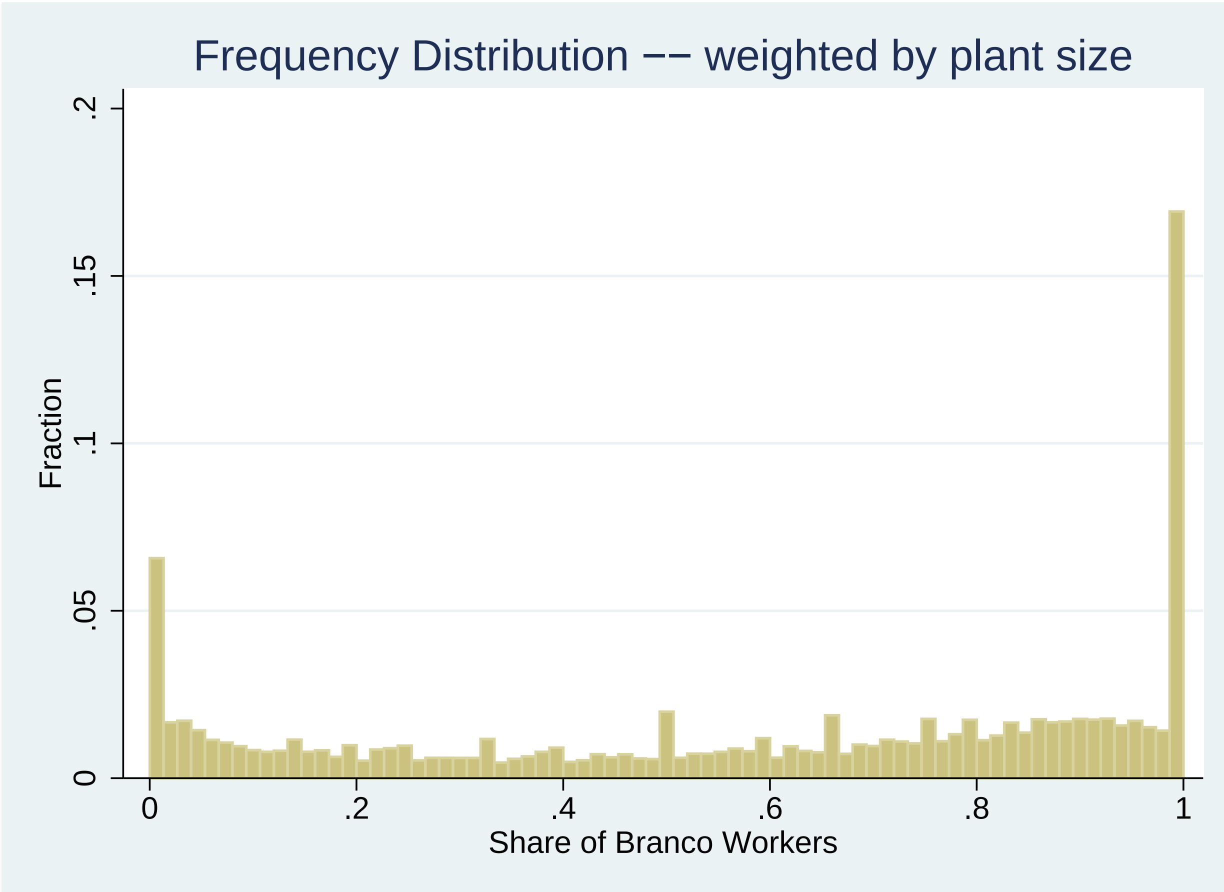

Figure A.1: Share of White-classified Workers, 2010 (Plant-level; Weighted by Plant Size).

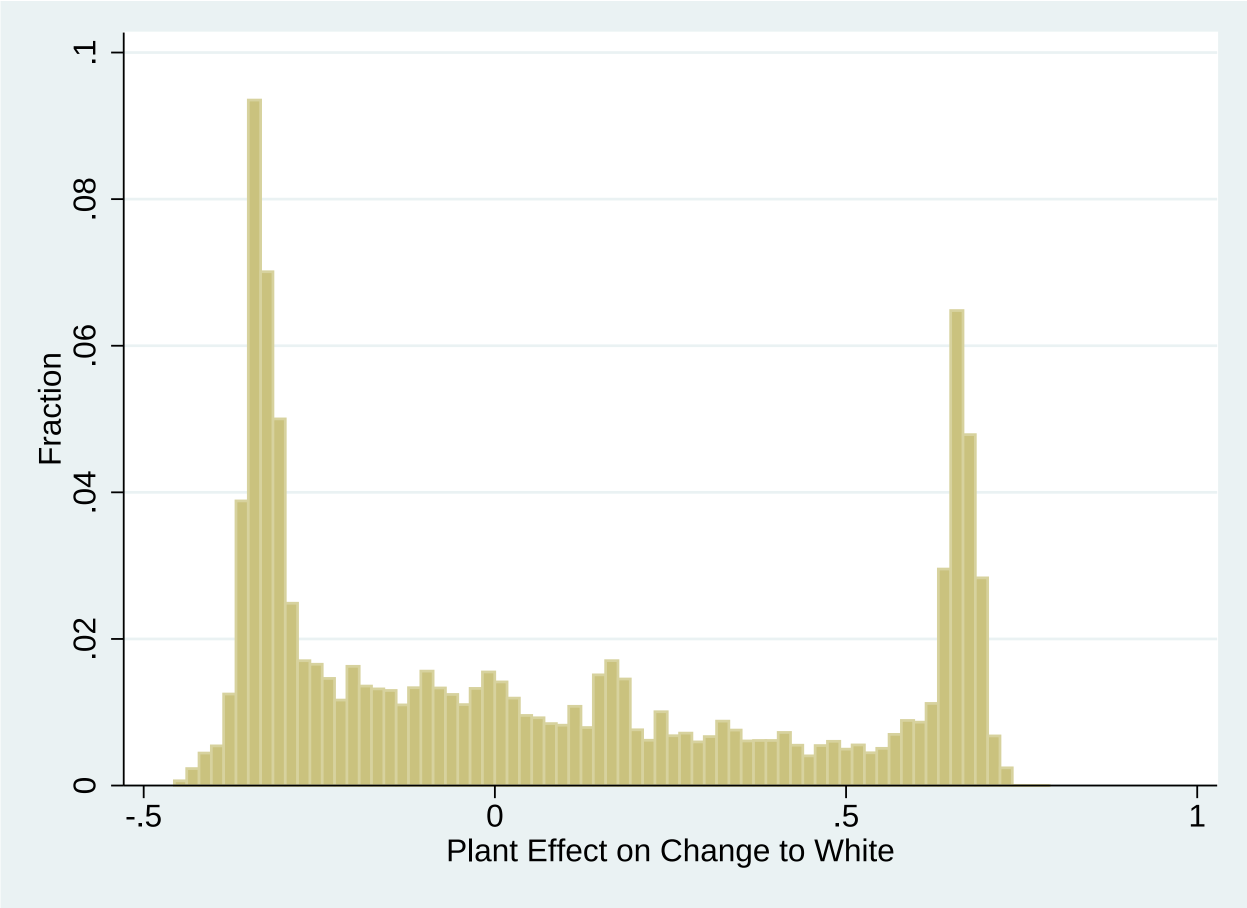

Figure A.2: Distribution of Plant Effects on to-White Race Change, 2010 (Plant-level; Weighted by Plant Size).

Table XXI: Summary Statistics of Unique Workers, 2003-2010

	Means and Frequencies							
Variable	2003	2004	2005	2006	2007	2008	2009	2010
Log Earnings	6.58	6.61	6.61	6.66	6.70	6.73	6.77	6.81
	(0.86)	(0.86)	(0.85)	(0.84)	(0.83)	(0.83)	(0.83)	(0.83)
Education								
< Elementary	27.62	25.65	23.83	22.11	20.91	19.75	18.42	17.24
Elementary	13.76	13.34	12.90	12.73	12.48	11.78	11.36	10.70
Some HS	7.23	7.27	7.00	7.03	7.03	6.87	6.66	6.48
High School	30.87	33.13	34.87	36.75	38.15	39.72	40.69	41.88
Some College	4.38	4.60	4.79	4.97	4.87	4.81	4.81	4.75
Bachelor's (+)	16.09	16.00	16.61	16.41	16.55	17.08	18.06	18.95
Experience (years)	8.45	8.37	8.10	7.81	7.54	7.48	7.40	7.30
	(8.98)	(8.61)	(8.63)	(8.50)	(8.42)	(8.52)	(8.57)	(8.56)
Tenure (years)	6.74	6.74	6.71	6.74	6.64	6.59	6.62	6.58
	(7.58)	(7.67)	(7.79)	(7.90)	(7.93)	(8.00)	(8.11)	(8.17)
Race								
White	71.99	71.25	69.04	75.58	74.84	73.67	73.09	72.25
Black	3.82	3.94	4.18	3.71	3.86	4.03	4.06	4.04
Brown	22.96	23.54	25.69	19.97	20.57	21.53	22.11	22.97
Other (Non-White)	1.23	1.27	1.09	0.74	0.73	0.77	0.74	0.75
Male Share	60.73	60.29	59.84	59.13	58.89	58.46	57.87	56.93
Workers (N)	15,623,463	16,573,158	17,158,485	17,721,892	18,287,340	19,103,691	19,243,030	19,850,746

Total observations, $N^* = 143,561,805$; Total unique workers, $N = 36,590,870$.

Standard deviations reported in parentheses.

Earnings reported in log 2003 Brazilian Reais.

Table XXII: Firm Characteristics, Summary Statistics, 2003-2010

Variable/Category	Percent/Mean
Industry	
Agriculture & Fishing	5.13
Mining	0.44
Production	20.93
Utilities	0.79
Construction	5.27
Trade & Repair	7.53
Food & Lodging	1.11
Transportation, Storage & Communication	7.03
Financial Intermediaries and Banks	1.38
Real Estate	12.09
Defense & Security	25.63
Education	3.66
Healthcare & Other Social Services	4.79
Other Services	4.23
International Organizations	0.01
Ownership Type	
Private, For Profit	62.11
Public/Government	26.74
Private, Non-Profit	8.81
Individual	2.35
Hire Rate	0.2417
	(0.1665)
Fire Rate	0.1316
	(0.1186)
Quit Rate	0.0501
	(0.0525)

Total observations: Firms, $J = 18,556$.

Standard deviations reported in parentheses.

Table XXIII: Estimated Coefficients of Worker Characteristics on Log Monthly Earnings, 2003-2010

Variable	$X\beta$ 1	$+\theta$ 2	Complete 3	Variable	$X\beta$ 1	θ 2	Complete 3
Education				Experience$^l \times$ Black			
< Elementary	-0.4135	-0.0190	-0.0175	$l = 1$	-0.0304	-0.0016	-0.0037
	(0.0015)	(0.0002)	(0.0000)		(0.0002)	(0.0001)	(0.0000)
Elementary	-0.2354	-0.0036	-0.0072	$l = 2/10^2$	0.1625	-0.0246	0.0272
	(0.0018)	(0.0002)	(0.0001)		(0.0018)	(0.0011)	(0.0005)
Some HS	-0.2308	-0.0403	-0.0298	$l = 3/10^3$	-0.0373	0.0121	-0.0073
	(0.0002)	(0.0002)	(0.0001)		(0.0006)	(0.0004)	(0.0002)
Some College	0.3606	0.0270	0.0148	$l = 4/10^4$	0.0029	-0.0014	0.0006
	(0.0027)	(0.0002)	(0.0001)		(0.0001)	(0.0000)	(0.0000)
Bachelor's (+)	0.8726	0.1096	0.0977	Experience$^l \times$ Brown			
	(0.0016)	(0.0002)	(0.0001)	$l = 1$	-0.0460	-0.0096	-0.0053
Experience^l					(0.0001)	(0.0000)	(0.0000)
$l = 1$	0.1194	0.0284	0.0058	$l = 2/10^2$	0.2654	0.0535	0.0363
	(0.0001)	(0.0000)	(0.0000)		(0.0007)	0.0004	(0.0002)
$l = 2/10^2$	-0.5377	-0.1686	-0.0313	$l = 3/10^3$	-0.0607	-0.0118	-0.0095
	(0.0005)	(0.0003)	(0.0002)		(0.0002)	(0.0001)	(0.0001)
$l = 3/10^3$	0.1068	0.0399	0.0090	$l = 4/10^4$	0.0046	0.0009	0.0008
	(0.0015)	(0.0001)	(0.0000)		(0.0000)	(0.0000)	(0.0000)
$l = 4/10^4$	-0.0079	-0.0031	-0.0009	Year			
	(0.0000)	(0.0000)	(0.0000)	2004	0.2731	0.1073	0.0712
Experience$^l \times$ Female					(0.0002)	(0.0001)	(0.0001)
$l = 1$	-0.1121	-0.0042	-0.0018	2005	0.2764	0.1430	0.0964
	(0.0001)	(0.0001)	(0.0000)		(0.0002)	(0.0001)	(0.0001)
$l = 2/10^2$	0.7047	0.0321	0.0160	2006	0.3190	0.2193	0.1640
	(0.0061)	(0.0005)	(0.0002)		(0.0002)	(0.0012)	(0.0001)
$l = 3/10^3$	-0.1657	-0.0078	-0.0046	2007	0.3562	0.2843	0.2175
	(0.0002)	(0.0001)	(0.0001)		(0.0002)	(0.0001)	(0.0001)
$l = 4/10^4$	0.0129	0.0006	0.0004	2008	0.3790	0.3422	0.2643
	(0.0000)	(0.0000)	(0.0000)		(0.0002)	(0.0001)	(0.0001)
Experience$^l \times$ Other				2009	0.4141	0.4093	0.3191
$l = 1$	-0.0011	-0.0113	-0.0072		(0.0024)	(0.0001)	(0.0001)
	(0.0003)	(0.0002)	(0.0001)	2010	0.4476	0.4811	0.3772
$l = 2/10^2$	-0.0030	0.0747	0.0552		(0.0002)	(0.0001)	(0.0001)
	(0.0031)	(0.0018)	(0.0010)	Observations			
$l = 3/10^3$	-0.0083	-0.0178	-0.0146	Groups (G)			3
	(0.0010)	(0.0006)	(0.0003)	Workers (N)			36,590,870
$l = 4/10^4$	0.0002	0.0014	0.0012	Firms (J)			18,556
	(0.0001)	(0.0001)	(0.0000)	Total (N^*)			143,561,805

N^* is the sum of all worker-year observations.

Standard errors in parentheses.

Column 1 estimates $Y = X\beta + \varepsilon$. Column 2 estimates $Y = X\beta + \theta + \varepsilon$. Column 3 estimates equation (3.2).

Table XXIV: Duration Weighted Summaries of Estimated Components of Log Earnings, 2003-2010

				Component Correlations						
Component	Label	Mean	Std. Dev.	Y	$X\hat{\beta}$	$\hat{\theta}$	$\hat{\psi}$	$\hat{\gamma}_1 s$	$\hat{\gamma}_2 s^2$	$\hat{\varepsilon}$
Y	Log earnings	6.6888	0.8440	1.0000						
$X\hat{\beta}$	Time varying characteristics†	0.0323	0.0479	0.5576	1.0000					
$\hat{\theta}$	Worker effect	0.0000	0.6061	0.8349	0.4987	1.0000				
$\hat{\psi}$	Firm effect/starting salary	0.0000	0.2962	0.5879	0.1950	0.2974	1.0000			
$\hat{\gamma}_1 s$	Linear seniority effect	0.2249	0.3887	0.3801	0.3134	0.1342	0.0846	1.0000		
$\hat{\gamma}_2 s^2$	Quadratic seniority effect	-0.0000	0.0003	-0.0880	-0.0364	-0.0682	-0.0018	-0.2771	1.0000	
$\hat{\varepsilon}$	Sample residual	0.0000	0.2033	0.2408	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

$N^* = 143,561,805$

†Time varying characteristics net of year effects.

Table XXV: Starting Pay, Returns to Seniority and Turnover: Summary Statistics and Simple Correlations

			Component Correlations								
Variable	Mean	Std. Dev.	$\hat{\psi}$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{P}$	Size	% High Skill	Hire Rate	Fire Rate	Quit Rate
$\hat{\psi}$	-0.0452	0.2811	1.0000								
$\hat{\gamma}_1$	0.0520	0.4496	-0.2260	1.0000							
$\hat{\gamma}_2$	-0.0006	0.9922	0.1004	-0.7361	1.0000						
$\hat{P}$	-13.3912	271.21	0.0105	0.0011	0.0001	1.0000					
Size	1,134.5	5,100.1	0.0408	-0.0044	0.0002	-0.0025	1.0000				
% High Skill	0.1947	0.2138	0.2772	-0.0114	0.0004	-0.0047	0.0850	1.0000			
Hire Rate	0.2417	0.1667	-0.1824	0.0618	-0.0032	0.0072	-0.0564	-0.3475	1.0000		
Fire Rate	0.1316	0.1186	-0.0153	0.0310	-0.0106	0.0145	-0.0684	-0.3613	0.4246	1.0000	
Quit Rate	0.0501	0.0525	-0.1635	0.0264	-0.0119	0.0117	-0.0301	-0.1783	0.4518	0.1774	1.0000

Total observations: $J = 18,556$.

Table XXVI: Starting Pay, Returns to Seniority and Turnover: Summary Stats and Correlations by Ownership Status

Private, For Profit							Component Correlations				
Variable	Mean	Std. Dev.	$\hat{\psi}$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{P}$	Size	% High Skill	Hire Rate	Fire Rate	Quit Rate
$\hat{\psi}$	0.0000	0.2517	1.0000								
$\hat{\gamma}_1$	0.0652	0.5552	-0.2971	1.0000							
$\hat{\gamma}_2$	0.0008	1.2585	0.1435	-0.7493	1.0000						
$\hat{P}$	-8.4980	100.44	-0.0019	0.0021	0.0030	1.0000					
Size	951.9	2,481.3	0.0696	-0.0068	0.0002	-0.0048	1.0000				
% High Skill	0.1358	0.1880	0.3581	0.0000	0.0004	0.0053	0.0702	1.0000			
Hire Rate	0.2767	0.1725	-0.3270	0.0563	-0.0115	0.0074	-0.0439	-0.3043	1.0000		
Fire Rate	0.1731	0.1144	-0.1971	0.0067	-0.0147	0.0196	-0.0838	-0.3045	0.3017	1.0000	
Quit Rate	0.0585	0.0532	-0.3378	0.0096	-0.0133	0.0068	0.0033	-0.1037	0.3894	0.0021	1.0000

Public/State-Owned							Component Correlations				
Variable	Mean	Std. Dev.	$\hat{\psi}$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\|\hat{P}\|$	Size	% High Skill	Hire Rate	Fire Rate	Quit Rate
$\hat{\psi}$	-0.1280	0.3183	1.0000								
$\hat{\gamma}_1$	0.0247	0.0681	-0.1121	1.0000							
$\hat{\gamma}_2$	0.0004	0.0610	-0.0435	-0.1954	1.0000						
$\hat{P}$	-27.3316	501.50	0.0094	-0.0076	0.0008	1.0000					
Size	1,730.8	9,125.5	0.0548	0.0082	-0.0022	0.0004	1.0000				
% High Skill	0.2971	0.1890	0.4918	0.0671	-0.0086	0.0094	0.1035	1.0000			
Hire Rate	0.1540	0.1095	-0.0854	0.0773	-0.0051	-0.0127	-0.0583	-0.0615	1.0000		
Fire Rate	0.0328	0.0535	0.0448	0.0094	-0.0137	-0.0229	-0.0313	0.0072	0.4128	1.0000	
Quit Rate	0.0246	0.0358	0.0394	0.0202	-0.0106	0.0031	-0.0245	0.0254	0.3708	0.0985	1.0000

Private, Non-Profit							Component Correlations				
Variable	Mean	Std. Dev.	$\hat{\psi}$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\|\hat{P}\|$	Size	% High Skill	Hire Rate	Fire Rate	Quit Rate
$\hat{\psi}$	-0.0780	0.2787	1.0000								
$\hat{\gamma}_1$	0.0318	0.1531	-0.2226	1.0000							
$\hat{\gamma}_2$	0.0045	0.2005	0.0004	-0.5906	1.0000						
$\hat{P}$	-8.0470	110.67	0.0268	-0.0085	0.0023	1.0000					
Size	837.6	1,296.9	0.1062	-0.0141	-0.0107	-0.0073	1.0000				
% High Skill	0.3508	0.2658	0.2775	0.0318	-0.0026	-0.0174	0.0847	1.0000			
Hire Rate	0.2068	0.1335	-0.3176	0.0381	0.1731	0.0032	-0.0392	-0.2611	1.0000		
Fire Rate	0.1082	0.0869	-0.0705	0.0141	-0.0330	0.0052	-0.1038	-0.0418	0.2210	1.0000	
Quit Rate	0.0482	0.0454	-0.1068	0.0573	-0.0394	-0.0050	0.0159	-0.0500	0.2284	-0.0051	1.0000

Individual Owners							Component Correlations				
Variable	Mean	Std. Dev.	$\hat{\psi}$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\|\hat{P}\|$	Size	% High Skill	Hire Rate	Fire Rate	Quit Rate
$\hat{\psi}$	-0.2002	0.2090	1.0000								
$\hat{\gamma}_1$	0.0488	0.2147	-0.3012	1.0000							
$\hat{\gamma}_2$	0.0070	0.2294	0.1996	-0.7821	1.0000						
$\hat{P}$	-7.4725	69.41	-0.0653	0.0080	0.0048	1.0000					
Size	353.7	301.5	0.2351	0.0465	-0.0152	-0.0506	1.0000				
% High Skill	0.0221	0.0520	0.1337	-0.0513	0.0055	-0.0293	0.1397	1.0000			
Hire Rate	0.4378	0.1626	-0.0859	0.0472	0.1978	-0.0705	-0.0491	-0.2800	1.0000		
Fire Rate	0.2142	0.1280	-0.0898	0.2045	-0.0871	-0.0298	0.0495	-0.1899	0.2516	1.0000	
Quit Rate	0.1131	0.0702	-0.1954	0.1105	-0.0674	0.0133	-0.0858	-0.1443	0.4954	0.0541	1.0000

Total observations: $J = 18,249$.

Table XXVII: Effects of Starting Pay and Returns to Seniority on Turnover, 2003-2010

	Hire Rate			Fire Rate			Quit Rate		
	Unconditional	Conditional	Conditional	Unconditional	Conditional	Conditional	Unconditional	Conditional	Conditional
Model	1	2	3	4	5	6	7	8	9
$\hat{\psi}$	-0.1080***	-0.1020***	-0.0808***	-0.0064*	-0.0029	-0.0220***	-0.0305***	-0.0311***	-0.0364***
	(0.0056)	(0.0052)	(0.0055)	(0.0032)	(0.0033)	(0.0031)	(0.0016)	(0.0016)	(0.0019)
$\hat{\gamma}_1$	0.0229*	0.0268	0.0145	0.0082**	0.0128*	-0.0109***	0.0031**	-0.0020	-0.0070***
	(0.0116)	(0.0159)	(0.0131)	(0.0026)	(0.0055)	(0.0029)	(0.0012)	(0.0016)	(0.0012)
$\hat{\gamma}_2$	-0.0005	0.0113	0.0063	-0.0013	0.0031	-0.0038***	-0.0006	-0.0004	-0.0021***
	(0.0042)	(0.0082)	(0.0065)	(0.0007)	(0.0026)	(0.0011)	(0.0004)	(0.0007)	(0.0004)
R^2	N/A	0.0360	0.3400	N/A	0.0010	0.4220	N/A	0.0270	0.2030
Controls	No	No	Yes	No	No	Yes	No	No	Yes

Total observations for all regressions: $J = 18,556$.
Standard errors in parentheses.
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.
Control variables include ownership status, industry, firm size and % high skill employees.

Table XXVIII: Effects of Starting Pay and Returns to Seniority on Turnover by Ownership Status, 2003-2010

	Private, For Profit		
	Hire Rate	Fire Rate	Quit Rate
$\hat{\psi}$	-0.1340***	-0.0530***	-0.0727***
	(0.0080)	(0.0053)	(0.0026)
$\hat{\gamma}_1$	0.0011	-0.0186***	-0.0142***
	(0.0121)	(0.0030)	(0.0014)
$\hat{\gamma}_2$	0.0028	-0.0054***	-0.0033***
	(0.0059)	(0.0011)	(0.0005)
R^2	0.3110	0.2640	0.1690
Obs	11,335		
	Public / State-Owned		
	Hire Rate	Fire Rate	Quit Rate
$\hat{\psi}$	-0.0148*	0.0092***	0.0044**
	(0.0068)	(0.0027)	(0.0017)
$\hat{\gamma}_1$	0.1280	0.0092	0.0106
	(0.0783)	(0.0085)	(0.0067)
$\hat{\gamma}_2$	0.0147	-0.0077	-0.0039
	(0.1070)	(0.0064)	(0.0039)
R^2	0.0260	0.0150	0.0110
Obs	4,879		
	Private, Non-Profit		
	Hire Rate	Fire Rate	Quit Rate
$\hat{\psi}$	-0.1070***	-0.0138	-0.0236***
	(0.0176)	(0.0091)	(0.0047)
$\hat{\gamma}_1$	0.1220	-0.0209	-0.0024
	(0.0851)	(0.0208)	(0.0095)
$\hat{\gamma}_2$	0.1700***	-0.0243**	-0.0076
	(0.0306)	(0.0080)	(0.0051)
R^2	0.2060	0.0660	0.1080
Obs	1,607		

Total observations for all regressions: $J = 18,249$. Additional control variables include firm size, % high skill employees and industry. Standard errors in parentheses. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table XXVIII, Continued

	Individual Owners		
	Hire Rate	Fire Rate	Quit Rate
$\hat{\psi}$	-0.0054	-0.0120	-0.0504**
	(0.0390)	(0.0386)	(0.0167)
$\hat{\gamma}_1$	0.3720***	0.1810	0.0321
	(0.1030)	(0.1410)	(0.0338)
$\hat{\gamma}_2$	0.4400***	0.0802	0.0187
	(0.0690)	(0.0909)	(0.0235)
R^2	0.2260	0.0950	0.0640
Obs	428		

Total observations for all regressions: $J = 18,249$.

Additional control variables include firm size, % high skill employees and industry. Standard errors in parentheses. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

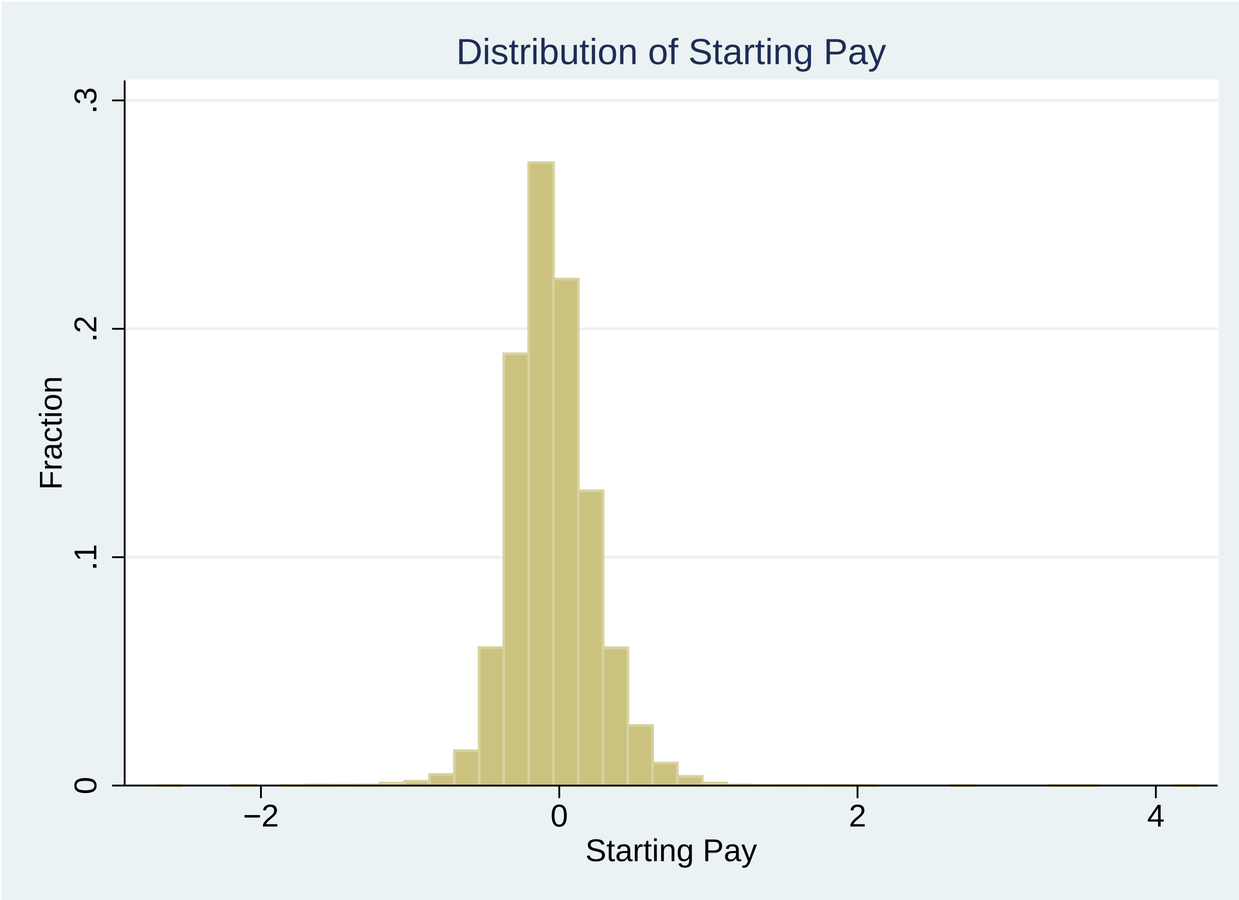

Figure A.3: Distribution of Firm-Specific Starting Pay, 2003-10.

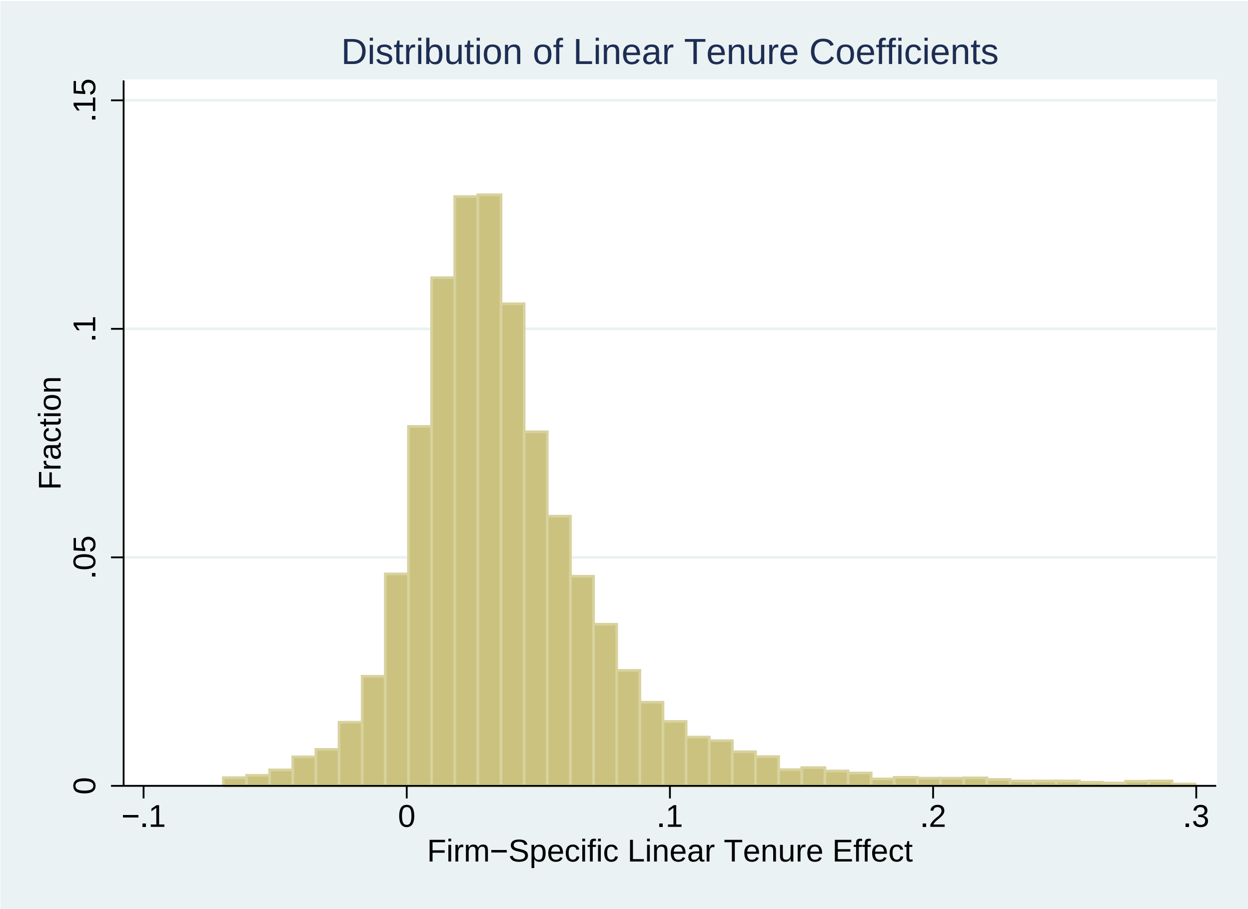

Figure A.4: Distribution of Firm-Specific Linear Returns to Seniority, 2003-10

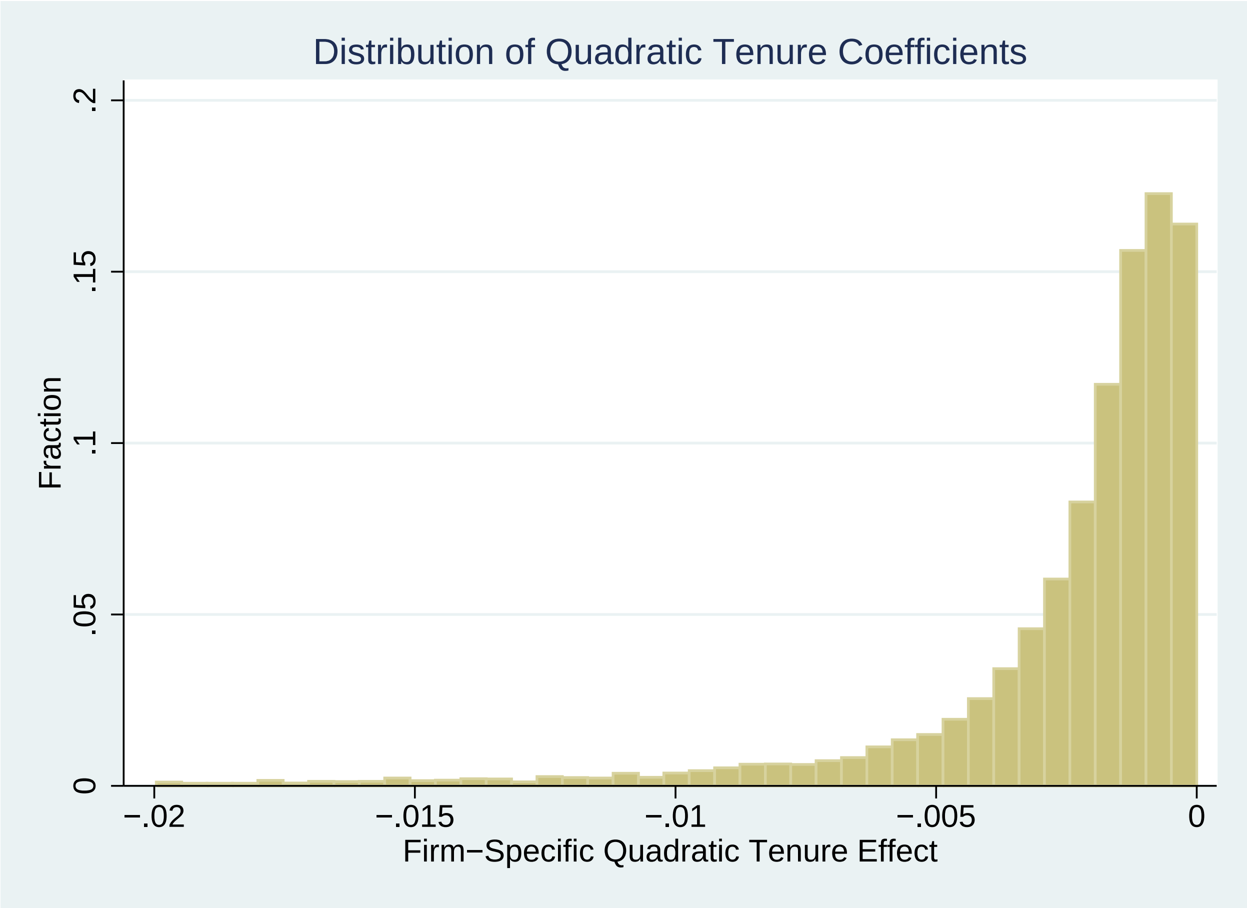

Figure A.5: Distribution of Firm-Specific Quadratic Returns to Seniority, 2003-10

Appendix B

Additional Materials for Racial Classification and the Race-Wage Gap in Brazil

B.1 Data Description

B.1.1 RAIS

Education is reported as one of seven categories: no schooling, some elementary, elementary, some high school, high school, some college and bachelor's degree or better. From 2005 onward, additional categories for workers with master's and doctoral/professional degrees are available; I opt to add these to the bachelor's category to remain consistent with the 2003 and 2004 waves.

RAIS records a plants industry using the *Classificação Nacional de Atividades Econômicas*, 1995 (CNAE 95) or the National Classification of Economic Activities, a five-digit system designed by the IBGE. I aggregate these codes according to the IBGE website to obtain the 16 broad industries listed in table I.

RAIS uses another IBGE coding system, the 2002 edition of the *Classificação Brasileiro de Ocupações* (CBO 2002). This six-digit code is very detailed and consists of hundreds of thousands of occupation types. I use the first digit to aggregate these very specific occupations into ten broadly defined occupations listed in table I.

B.1.2 PNAD

PNAD classifies a worker's industry according to the *Classificação Nacional de Atividades Econômicas – Domicilias* (CNAE-D), a less-detailed version of the CNAE-95 used in RAIS. The CNAE-D is roughly equivalent to the UN's ISIC, thrid edition. I aggregate industry codes from CNAE-D into the 16 different industries derived from the CNAE-95 using definitions listed on the IBGE website.

PNAD records occupation in several different ways. I use the *Classificação Brasileiro de Ocupações Domiciliar* (CBO-D). CBO-D regroups some occupational families into different sub-groups in order to eliminate difficulties in precisely capturing a worker's occupation during the interview. However, CBO-D is identical to CBO-2002 at the one digit level, allowing me to use the same ten occupational categories in PNAD as I do in RAIS.

B.1.3 Additional Descriptive Statistics–RAIS

Education

Education levels in Brazil are generally increasing over the sample period, as the lowest education levels – No Schooling, Some Elementary and Elementary – all decrease over the sample period. From 2003 to 2010, the share of the formal labor force with no schooling at all decreases from about 1.2 percent to 0.6 percent. The share of workers with only some elementary schooling, or less than the equivalent of an eighth grade education in the U.S., is large, but falls from 29.2 percent to 17.9 percent. Similarly, the share of workers with the equivalent of an eighth grade education – Elementary – decreases from 17.4 percent to 14.1

percent. Even the Some High School category decreases slightly from about 8.3 percent to 7.6 percent.

These decreases in the share of workers with lower education levels are necessarily offset by increases in the share of workers with higher levels of education. The largest increase occurs in the share of workers who have a complete High School education. The share of these workers in the formal labor market increases from 28.5 percent in 2003 to 42 percent in 2010. From 2004 onward, workers with a high school-level education make up the largest portion of the formal labor force. The share of workers with some college increases over the sample period as well, from 3.6 percent to 4.1 percent; the share of workers with a bachelor's degree (including workers with post-graduate and professional degrees) increases from 12 percent to 13.7 percent.

Industry

The IBGE provides an algorithm to aggregate both the CNAE 95 and CNAE-D codes into 16 industries. Table II provides the share of workers in each industry, not the number of firms or plants in each industry. Focusing on the larger industries, the production sector employs roughly 18 percent of all workers, according to RAIS, and this number varies slightly from as low as 17.8 percent in 2010 to as much as 19 percent in 2007. The construction industry grows over the sample period from 4.9 percent in 2003 to 7.3 percent in 2010. Similarly, the trade and repair sector grows over the sample period from 18 percent in 2003 to 19.7 percent in 2010. The share of workers in the real estate industry is relatively stable over the sample period, while the share of workers in the defense and security industry decreases over the sample period from about 19.5 percent to about 17 percent. The defense and security, trade and repair and production industries are the top three employing industries according to RAIS.

Occupation

Focusing on the larger occupations, professionals, artists and scientists comprise about eight percent of the formal labor force of Brazil. This share is very stable over the sample period, with a minimum of 8.0 percent in 2004 and a maximum of 8.3 percent in 2010. Mid-level technicians also form a relatively stable, if slightly decreasing, share of the formal labor force, with these workers comprising ten percent of formal workers in 2003 and 9.6 percent of workers in 2010.

The share of workers in administrative positions is generally large and increasing for the sample period, In 2003, 17.9 percent of workers are administrative and clerical workers while by 2010 this figure has increased to 18.6 percent. This trend appears for both the service and vendor occupation class and the production I occupation class. Service and vendor workers are the largest share of workers in any given year, and their share of the formal labor force is weakly increasing over the sample period from 22.9 percent to 24.2 percent. Production I workers are the second largest occupation class and their share of the workforce also weakly increases over time from 21.2 percent to 23.2 percent.

Mean Age and Earnings

Mean age and mean nominal earnings both increase over the sample period. Mean age increases from about 35 years about 6 months to 35.5 years in 2010. Mean nominal monthly earnings increase from about 928 R$ in 2003 to about 1,558 R$ in 2010.

Overall, RAIS provides a picture of Brazil's formal labor force that is decreasing in "whiteness," becoming more and more educated, growing in construction and trade industries, decreasing in defense and security industries, and increasing in administrative, service and production workers. Brazilian formal workers are also a getting a bit older, but earning much more money over the sample period.

B.1.4 Additional Descriptive Statistics–PNAD

Education

As with race, education as captured by PNAD is qualitatively similar to education as captured by RAIS. No Schooling is the smallest reported category, generally comprising about 3.5 percent of the workforce. This segment reaches a minimum of 3.1 percent of the labor force in 2009 and a maximum of 4.4 percent in 2011. Workers with an incomplete elementary education still make up a large share of all workers, though this portion of the workforce decreases monotonically over the sample period from 28.6 percent in 2003 to 18.6 percent in 2011.

Elementary (eighth grade) educated workers are a smaller portion of the labor force in both RAIS and PNAD, though in PNAD this portion of the labor force is neither increasing nor decreasing over the sample period. These workers have a maximum share of the workforce of 10.2 percent in 2004 and 2007 and a minimum share of 8.9 percent in 2009. As in RAIS, workers with an incomplete high school education represent a decreasing share of the formal workforce, decreasing slightly from 5.5 percent in 2003 and 2004 to 5.1 percent in 2011.

High school educated workers comprise the largest portion of formal workers in PNAD for any given year. Further, as in RAIS, workers with a high school-level education represent an increasing share of the labor force over the sample period, growing from 29.8 percent to 35.9 percent of the workforce from 2003 to 2011. The share of workers with some college increases as well from 7.3 percent to 8.6 percent. And the share of workers with at least a bachelor's degree monotonically increases from 14.1 percent to 17.2 percent. As with RAIS, the education composition of the formal workforce of Brazil in PNAD reveals a workforce that is becoming more educated and more highly skilled.

Industry

The industrial composition of the formal labor force from PNAD retains some similar characteristics to the RAIS composition. For example, production, construction, trade and repair, real estate and defense and security employ notably large portions of the workforce. Also, the share of workers employed in the defense and security industry is decreasing over time. However, according to PNAD, production tends to be the largest sector, except in 2009 and 2011. The production sector here employs about 18 percent of workersm though this varies from 19 percent in 2004 to 16 percent in 2011.

The construction sector is increasing over the sample period, as in RAIS, from about four percent to almost six percent. The trade and repair sector typically employs about 17 to 18 percent of workers, and, as in RAIS, does not exhibit any increasing or decreasing trends over the sample period; the minimum share is 17.1 percent in 2004 and the maximum is 18.4 percent in 2011. In PNAD, the trade and repair sector represents the second-largest employer of formal workers, except in 2009 and 2011, when it is the largest employer.

Whereas the real estate sector was relatively stable in RAIS, in PNAD the share of workers in real estate grows over the sample period from 8.4 percent in 2003 to 9.8 percent in 2011. As mentioned, the share of workers in the defense and security industry decreases over the sample period in PNAD, from 9.4 percent to 8.7 percent. Defense and security, however, employ a much smaller share of the formal labor force in PNAD than in RAIS. Also unlike in RAIS, education employs a larger portion of formal workers than defense and security does, as education's share of the labor force decreases from 10.1 percent to 8.5 percent over the sample period. Despite the many qualitative similarities between how RAIS and PNAD describe the industrial composition of the formal labor force, there are important differences that may affect analysis.

Occupation

The occupational composition from PNAD's description of the labor force retains some important similarities to that from RAIS. The occupations that make up the larger segments of the labor force in RAIS do so in PNAD. Service and vendor workers make up the largest share of workers, followed by production I-type workers. However, many of the trends observed over the sample in RAIS do not carry over in the PNAD statistics.

Professionals, artists and scientists increase as a share of the labor force from ten percent to 11.9 percent, though there is a trough of about 9.8 percent in 2005. In RAIS, the share of professionals, artists and scientists is relatively stable across the sample. However, mid-level technicians, which are decreasing as a share in RAIS, also decrease in PNAD from 11.2 percent of workers in 2003 to 9.3 percent in 2011.

The share of service workers and vendors, which was increasing in RAIS, is relatively stable in PNAD at slightly more than 27 percent of the workforce, with a minimum and maximum of 65.5 percent in 2008 and 29.6 percent in 2011. Service workers and vendors also make up the largest occupational class. Production I workers are also relatively stable in PNAD, when they were slightly increasing as a share of the workforce in RAIS. These production workers have a minimum share of 18.7 percent in 2009 and a maximum share of 19.3 percent in 2004 and 2011, and make up the second-largest occupational class.

Mean Age and Earnings

Mean age and mean nominal monthly earnings both increase over the sample period for PNAD, as they do in RAIS. However, the change in age is greater in PNAD, with average age increasing almost one year, as opposed to 6 months, from 36.6 years in 2003 to 37.5 years in 2011. Nominal mean monthly earnings increase from 919 R$ in 2003 to 1,596 R$ in 2011, a 677 R$ increase; nominal mean monthly earnings in PNAD increased by 670 R$[1].

[1] Average monthly earnings for 2011 are deflated to 2010 Reais.

Qualitatively, PNAD and RAIS are very close in very important ways. Both surveys convey a formal labor force that is majority white, but increasing in minority participation. Both surveys depict workforces that are majority male, but with increased female presence. They both convey a workforce that is mostly high school-educated, but that is becoming more educated and more highly skilled over time. These surveys also suggest a labor force that is largely employed by the production, construction, trade and repair, real estate and defense industries, and also suggest that most formal workers fall into the occupations of professionals, artists and scientists, mid-level technicians, administrative workers, service workers and vendors and production workers. However there are important differences in non-reported race, trends of changing occupational and industrial composition and the importance of the education sector as an employer. These qualitative differences may drive important differences in analysis outcomes.

B.1.5 Additional Descriptive Statistics–Differences

Education

Within the education composition, RAIS under-reports the extreme low-end of the education. RAIS under-reports the share of workers with no formal education by about 2.7 pp. This discrepancy improves over time, but only slightly, ranging from -2.7 pp in 2003 to -2.4 pp in 2010/11. RAIS also under-reports the share of workers with some elementary-level education by 1.3 pp, on average. The difference in elementary shares actual increases in magnitude from RAIS over-reporting by 0.52 pp to RAIS under-reporting by 2.8 pp.

RAIS tends to over-report the shares of workers with mid-level education. Differences in elementary education average 6.2 pp. These differences decrease in magnitude over the sample period from 7.3 pp in 2003 to 4.1 pp in 2010/11. The same thing happens for incomplete high school education; RAIS over-reports the share of the workforce with incomplete high school educations by about 2.6 pp. As with the share of workers with an elementary-level education, the differences in Some High School's share decrease over the sample period, albeit only slightly from 2.7 to 2.5 pp. RAIS also over-represents the share of workers with high school educations by about 2.5 pp on average. The discrepancy between RAIS and PNAD grows a lot over the sample period from -1.4 pp (under-reporting) to 6.1 pp (over-reporting).

For workers with higher levels of education, RAIS again under-reports their share of the workforce. The average discrepancy for the share of workers with some college-level education is -4 pp. The differences in this categor increase in absolute value over the sample period from -3.7 to -4.5. For those workers who have completed a bachelor's degree or have education beyond the undergraduate level, the average difference between RAIS and PNAD is -2.8 pp. As with the Some College category, differences for Bachelor's (+) are increasing in magnitude over the sample period from -2.1 pp to -3.6 pp.

In general, RAIS under-reports No Schooling, Some Elementary, Some College and Bachelor's (+), relative to PNAD. And while RAIS under-reports these more extreme categories, it also over-reports the middle categories: Elementary, Some High School and High School. Further, small gains in relative accuracy in some categories appear overshadowed by lost accuracy in others.

Industry

Of the sixteen different industries that employ workers in the formal labor market, there are exceptional differences between RAIS and PNAD for fois neither decreasing nor increasing over the sample period with a minimum difference of 3 pp in 2010/11 and a maximum difference of 3.8 pp in 2007; the average difference across all years is about 3.5 pp. For the share of workers employed in the defense and security industry, differences between RAIS and PNAD are decreasing over the sample period from 10.1 pp to 8.4 pp, with and average of about 9.3 pp. Similarly, the differences in the education industry's share of workers decrease in magnitude over time from -7.6 pp to -5.8 pp. On average, RAIS under-represents the size of the education sector's share of employee, with an average difference of -7.0 pp. Though it comprises a small share of the industrial composition of the formal workforce, there are large discrepancies in the share of employees working in the domestic services industry, with an average difference of -4.7 pp.

Other large sectors such as production, construction and trade and repairs all have demonstrate relatively small differences in their shares of the workforce, typically less than one percentage point. Further, none of these sectors exhibit and trend in the differences, neither increasing nor decreasing over the sample period.

In general, there are very large, opposing differences in just a few select industry categories, including the larger categories of real estate, defense and education. There are also large percentage-point differences in the share of workers employed in domestic services, a very small share of the overall workforce. These large differences are likely a result of how RAIS and PNAD treat defense and security workers. RAIS includes all workers that contribute to either PIS or PASEP within its survey; PNAD appears to capture civilian public sector employees, but does not necessarily include all military or defense sector workers within its survey.

Occupation

Turning to occupation shares, five of the ten occupation categories exhibit very large dicrepancies between their RAIS and PNAD shares. In public administration and management, RAIS under-represents workers in this occupation, with an average difference of about -3.6 pp. These differences decrease in magnitude over the sample period from -4.1 pp to -2.1 pp. RAIS also under-represent the share of professionals, artists and scientists in the labor force, with an average difference of about -2.4 pp. The difference in shares for this category increases in absolute value over the sample period from about -2 to -3.6 pp. RAIS tends to over-represent the share of workers in administrative positions by about 4.7 pp. With the exception of 2010/11, the difference in administrative worker shares is fairly consistent, and is neither increasing nor decreasing over the sample period. The share of workers who are service workers or vendors is under-represented in RAIS, with an average difference of -4.1 pp. These differences are mostly decreasing in magnitude for the sample period. Lastly, RAIS over-represents workers in the Production I category aby an average of 2.9 pp. This dscrepancy is increasing over time from 2.3 pp in 2003 to 3.9 pp in 2010.

In those occupations with smaller differences – military, mid-level technicians, agriculture/fishing/forestry, production II and repair/maintenance – the differences between RAIS and PNAD are decreasing in magnitude for the sample period.

Mean Age and Earnings

RAIS reports workers as roughly 1.7 years, or 4.8 percent, younger than they are in PNAD. This difference is fairly consistent over the sample period, and ranges from -4.5 percent in 2005 to -5.4 percent in 2010/11. The percentage difference in average monthly earnings is neither increasing nor decreasing, but varies greatly. On average, RAIS reports worker incomes as 2.9 percent greater. However, these numbers vary from -2.4 percent in 2010/11 to 6.5 percent in 2009.