WAGE DISCRIMINATION WHEN IDENTITY IS SUBJECTIVE: EVIDENCE FROM CHANGES IN EMPLOYER-REPORTED RACE

Christopher Cornwell
Department of Economics
Terry College of Business
University of Georgia
cornwl@uga.edu

Jason Rivera
Inductis/EXL Service
jas.rivera@gmail.com

Ian M. Schmutte
Department of Economics
Terry College of Business
University of Georgia
schmutte@uga.edu

December 2, 2014

Abstract

In Brazil, different employers often report different racial classifications for the same worker. We use this variation in employer-reported race to identify wage discrimination. Workers whose reported race changes from non-white to white receive a wage increase; those who change from white to non-white realize a symmetric wage decrease. As much as 40 percent of the raw racial wage gap is explained by the employer's report of race, after controlling for all individual characteristics that do not change across jobs. The results are consistent with workers manipulating perceived race in an environment where racial classification is subjective, but discrimination persists.

JEL Codes: J15, J31, J62, J71, D83, C2.
Keywords: Discrimination, Wage Differential, Job Mobility.

1 Introduction

We identify the effect of race on wages from variation in the report of the same worker's race by different employers. No prior study has been able to separately identify the effect of racial classification from arbitrary unobserved characteristics of the worker in a panel setting. Our ability to do so comes from the combination of our data and the setting in which it is collected. Using employer-employee matched data from Brazil, we observe that different employers often report different racial classifications for the same worker. Among workers who change jobs, those reported as non-white by their original employer and white by their destination employer realize a wage increase. Workers reported as white by the original employer and non-white by their destination employer experience a symmetric wage decrease. As much as 40 percent of the raw racial wage gap is explained by the employer's report of race, after controlling for all individual characteristics that do not change across jobs.

The Brazilian context is important because employers report race in an environment where racial categories are highly subjective. In Brazil, one's racial identity is closely associated with skin color. Racial classification is therefore subjective – different people may have different perceptions of the race of the same person. As a result, racial identity is subject to change over time and across contexts. Despite this subjectivity, there are large racial disparities in labor market outcomes favoring whites. Most economic analyses of racial discrimination implicitly assume workers and employers agree about the worker's race. However, theories of labor market discrimination imply that it is the employer's perception of race that matters. So, uniquely, we observe a context in which workers have the both the scope and incentive to manipulate their perceived race for an advantage in the labor market.

The central empirical challenge is, therefore, to demonstrate that differences in employer-reported race correspond to actual differences in labor market outcomes. While changes in racial classification are especially plausible in the Brazilian context, it may nevertheless be that most of the differences in reported race we observe arise from classification errors. To address this possibility, we specify and estimate a structural misclassification model. In the model, wages vary based on the 'market' race, which affects the distribution from which wages are drawn. The market race is unobserved, and may be imperfectly correlated with employer-reported race, which we do observe. Pure misclassification corresponds to a setting in which market race does not change over time. In that setting, observed changes in employer-reported race are not informative about market discrimination. We estimate the structural misclassification model and formally reject the hypothesis that observed changes in employer-reported race are represent pure misclassification error, unassociated with wages.

We turn to a consideration of different mechanisms that could drive our finding that racial wage gaps persist after controlling for arbitrary unobserved characteristics of the individual. Most changes in racial classification are associated with workers obtaining employment in segregated plants. The observed patterns are consistent with a model in which workers manipulate the way employers perceive race to obtain favorable treatment in a discriminatory labor market. Such "passing" behavior is rational in a context where race is subjective and affects wages. We consider two alternative mechanisms. First, our results may reflect reverse causality – higher pay induces employers to classify workers as white. Second, our results could be generated by plant-specific reporting behavior – some plants simply classify all workers are white or non-white. We show that neither mechanism is a likely explanation of our results.

We contribute most directly to the large literature on racial discrimination in the labor market. We use panel data methods to directly control for unobserved characteristics, rather than proxy for unobserved ability using performance on standardized tests, as in Neal and Johnson (1996). In doing so, we draw on a key insight of analyses of discrimination using field experiments (Rouse and Goldin 2000; Bertrand and Mullainathan 2004). Identifying the causal effect of race on labor market outcomes is conceptually problematic if race is understood as an immutable feature of an individual's identity. Field experiments work by manipulating something else: the employers perception of an applicant's race. We exploit a plausibly similar source of observed variation

– employer-reported race, while implicitly addressing an essential drawback of correspondence studies, which is they measure the racial bias of the average employer. As Heckman (1998) and Neumark (2012) argue, it is the racial bias of the marginal employer that determines whether workers experience discriminatory treatment in the labor market. Some recent research on discrimination also exploits variation in implied racial perceptions in non-experimental settings (Price and Wolfers 2010; Parsons et al. 2011). Both of these papers use data from major league sports. Ours is the first such paper to use nationally representative labor market data.

Our research also relates to the growing literature on the economics of identity. Akerlof and Kranton (2000) argue that the choice of identity, racial or otherwise, is one of the most consequential economic decisions a person can make. Nevertheless, very little empirical work has been done to understand how racial identity responds to economic incentives. An exception is research on the mechanisms of immigrant assimilation. Biavaschi et al. (2013) show that immigrants to the U.S. who adopted more 'American' names experienced large wage gains. Duncan and Trejo (2011) find self-reports of Hispanic-origin decline with economic status. Both lead to a downward bias in measures of immigrant achievement. Our results offer the first evidence, to our knowledge, of a causal link from labor market outcomes to racial identity.

2 Race in Brazil

Here, we describe aspects of race relations in Brazil most relevant to our study: the subjectivity and malleability of racial categories, on the one hand, and persistent racial inequality on the other. It is beyond the scope of this paper to provide a comprehensive survey of these topics. We refer the interested reader to Telles (2004).

2.1 Racial Classifications

In Brazil, race is generally characterized in terms of skin tone rather than in categories fixed by heredity. In 1976, Brazil's national household survey, the

Pesquisa Nacional por Amostra de Domicilios (PNAD), asked for an open-ended answer to a question about race. The responses yielded 136 different descriptions of skin color (Racusen 2009; Schwarcz 2003). Official statistics in Brazil, including the RAIS data we use, employ a standardized system of racial categorization that reflects an emphasis on skin tone. A person's race can be recorded as *branco* (white, or light-skinned), *pardo* (brown-skinned), *preto* (dark-skinned), *amarela* (yellow), or *Indigena* (Indigenous).[1] In the PNAD data, individual survey responders choose their race category; in RAIS, employers classify the race of their employees.

The notion of race embedded in these categories is unfamiliar to those used to thinking about race and discrimination in the U.S. context. As in the U.S., Brazil's history of race relations involves a narrative of white racial superiority. A key difference is that in the U.S., racial domination was supported through explicit laws against racial intermarriage and segregation. In Brazil, miscegenation was encouraged, leading, by the beginning of the twentieth century, to a large multi-racial population (Daniel 2010). The absence of a clear color line and lack of discriminatory laws coalesced in a national perception of Brazil as a "racial democracy", in which any racial inequality was mild, unintentional, and ultimately transitory (Fiola 1990). Statistical evidence of persistent racial disparities has challenged the "racial democracy" narrative. Nevertheless, there is still no affirmative action or equal opportunity legislation that binds on private Brazilian employers. Hence, there is no legal incentive for employers to alter the reported race of their workers.

Because race is defined by skin tone, there can be considerable ambiguity regarding whether an individual is light-skinned versus brown-skinned, or 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-

[1] The *amarela* and *Indigena* groups are very small and geographically concentrated. We omit them from our analysis. Their inclusion has no effect on our results.

educated and wealthier individuals as white when they self-report as non-white.

There is also evidence Brazilians manipulate perceived race for social and economic advantage. Since 2004, Brazilian universities have adopted aggressive affirmative action policies.[2] Francis and Tannuri-Pianto (2012) and Francis and Tannuri-Pianto (2013) show that the adoption of affirmative action policies led to students misrepresenting race to admissions offices. Policy makers are aware of this problem, which is a direct consequence of the *criterion of self-determination* – you are the race you report yourself to be – that characterizes racial identity in Brazil (Racusen 2009; Telles 2004). If students are willing and able to manipulate their race, as perceived by university admissions committees, to obtain better admissions outcomes, we speculate workers may be willing and able to manipulate their race, as perceived by employers, to obtain better employment outcomes.

2.2 Racial Inequality and Discrimination in Brazil's Labor Market

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, individuals have no incentive to manipulate their perceived race. The results surveyed in this section document a considerable degree of racial inequality in the labor market, as well as the prevalence of labor market discrimination through access to jobs and opportunities for advancement.

2.2.1 Racial Disparities in Labor Market Earnings

Data from the PNAD and Brazilian census indicate non-white men earn roughly 50-60 percent as much as white men. These discrepancies persist, though are more muted, when conditioning on industry, occupation, and region, and are

[2] Affirmative action policies have been introduced in university admissions in part because in the public higher-education system, slots are rationed to begin with. There is no equivalent affirmative action law that binds on private sector employers, though some state government agencies have adopted preferential hiring policies.

reflected in other indicators such as development, literacy, and total wealth. Our own calculations, using PNAD, indicate that from 2003-2010 non-white workers earned 20 percent less than white workers, after controlling for education, work experience, region, and industry. In Section 5, we report a racial wage gap in RAIS of 8 percent after controlling for both worker and employer characteristics. Interestingly, racial inequality in social and labor market outcomes is primarily between white and non-white workers. While there are differences in outcomes between brown and black workers, they are relatively negligible (Telles 2004).

2.2.2 Workplace Segregation

The RAIS data allow us to contribute new descriptive evidence on workplace segregation in Brazil. Brazilian formal-sector workplaces are highly racially stratified relative to the overall population. Figure 1 presents a histogram of the plant-size weighted distribution of the white share of all employees across all plants. Fifteen percent of plants have no non-white workers, and a further seven percent have no white workers. Thus, 22 percent of plants are completely homogeneous with respect to employer-reported race. There is no evidence of a mode near the white share of the formal sector workforce, which is 62 percent.[3]

2.2.3 Discrimination in the Workplace

There is considerable qualitative evidence of discrimination in recruiting and hiring. Through the 1950s, classified advertisements would explicitly exclude non-white applicants. Once this exclusion became socially unacceptable, explicitly racial terms were replaced by coded terms ("good appearance") that remained in use until at least the 1980s (Telles 2004). Telles (2004, p.161) describes an attempt to conduct an audit study in Brazil that failed because the low-skilled jobs he planned to test were always filled through word-of-mouth and employed only white applicants. Extrapolating from this example, one mechanism through which workers might manipulate perceived race is by

[3]The racial composition of the workforce varies considerably across different regions of Brazil. The results on racial stratification across plants are the same if we condition on region. All subsequent analysis will control for the geographic variation.

obtaining access to influential social networks. After hiring, non-white workers may experience discriminatory attitudes and practices in the workplace. When surveyed, 54 percent of people Rio de Janeiro identified work as the place of greatest racial tension. Furthermore, a majority of non-white respondents described experiences of discrimination in hiring and promotion. Bento (2000) reports that non-white workers struggle with advancement because of difficulty commanding the respect of their white subordinates.

2.3 A Note on Terminology and Racial Categories

In this paper, we focus on what happens to workers when they are classified by their employer as white (*branco*), versus when they are classified as either brown (*pardo*) or black (*preto*). Following Telles (2004), we group the brown and black categories together, and refer to them as “non-white”. This may seem an odd choice given the rich and complex nature of racial categories in Brazil. In particular, it may appear that we are incorrectly applying a U.S.-centric concept of race to the Brazilian context. However, the data show the white/non-white margin to be the most salient racial divide for labor market outcomes. The differences between brown and black workers are much smaller.

3 Data on Race and Job Mobility

The data in RAIS are collected to administer a constitutionally-mandated annual wage supplement (the *Abono Salarial*, or 13th salary), and to produce national statistics. RAIS data are collected at the plant level by plant management officials who complete the survey on behalf of the employees. In smaller enterprises this official may be the owner, while larger firms likely have an accountant, human resources manager or other administrator submitting the data. RAIS provides universal coverage of the formal labor market. For each registered plant, RAIS records information for every worker in its employ during the survey year. Completion of RAIS is mandatory, and plant compliance is very high. Plant owners are subject to large penalties when the data are late or are not completed. These penalties together with scrutiny from employees give employers strong incentives to comply with RAIS mandates.[4]

3.1 How Employers Collect and Report Information on Employee Race

All worker characteristics – race in particular – are reported by employers. To understand the race data in RAIS, we describe the process by which employers obtain and record information on worker characteristics. At the date of hire, the employee is required to produce a large number of official documents. Those documents include a "Worker Record Booklet" (*Carteira de Trabalho e Previdência Social*, CTPS). The CTPS includes basic information, including the worker's name, date of birth, gender, and place of residence as well as an identification number, but not race.[5] The worker is also required to provide the employer with a photograph and proof of education required for the position. The CTPS looks like a passport, and includes some of the same information.[6]

Upon hiring a new worker, the employer is required to make an entry in an "Employee Registration Book" (*Livro de Registro dos Empregados*, LRE), which is maintained by the plant. Information from the LRE is used to comply with several mandatory reporting requirements, including RAIS. In contrast to the CTPS, the LRE commonly includes a field for race (*COR*, literally "color"). The LRE also includes space for a photograph of the employee. The law requires employers collect each worker's name, date of birth, date of hire, and identification number, along with several other fields related to the job.[7] Employers are not required to collect information on race and gender, but they are, nevertheless, routinely reported.

In general, all information entered into the LRE is completed by the em-

[4]RAIS data have been little-used by labor economists. Existing economic applications of RAIS data include the study the role of firms in wage determination (Menezes-Filho et al. 2008), trade (Poole 2013; Krishna et al. 2014), firm spin-offs (Muendler et al. 2012), and labor market sorting (Lopes de Melo 2013). As far as we are aware, ours is the first study to use the unique features of RAIS to examine the role of race in wage determination.

[5]See http://www.planalto.gov.br/ccivil_03/LEIS/L8260.htm for requirements of the CTPS.

[6]Further information on the CTPS with visual examples is available from http://portal.mte.gov.br/ctps/tipos-de-ctps.htm.

[7]See http://www.alcon-sc.com.br/registro_de_empregados.htm for details

ployee, and subject to verification by the staff member responsible for hiring procedures.[8] Some of the information collected in RAIS, such as age and gender, is less ambiguous than race, and is generally reported consistently for the same worker across jobs. Other information, such as educational attainment, can be verified with other documents that employees are required to produce at hire.

The social convention regarding race in Brazil is that "you are what you say you are" (Telles 2004). No affirmative-action or equal opportunity laws bind on private-sector employers in Brazil that might induce them to manipulate the racial composition of their workforce (Telles 2002; Racusen 2009). Furthermore, the race information reported by employers does not appear to be subject to any systematic audit. Thus, our data on race emerges from a process that is primarily based on information provided by the worker, but where the employer's iterpretation of that information may play a role.

3.2 Data Preparation and Sample Construction

Our analysis is based on a sample of workers from the 2010 RAIS who change employers during the year. To construct that sample, we begin with the complete set of all jobs. We restrict attention to full-time jobs in which the employee is contracted to work 40 hours per week. For the reasons outlined in Section 2.3, we restrict our analysis to jobs on which the race of the worker is reported as either white, brown, or black.[9]

From this set of all full-time jobs in the formal sector, we locate all workers employed on what we will call a 'continuing' job. These are workers observed in a full-time job that started prior to the beginning of 2010. All workers with continuing jobs are at risk to enter our analysis sample. They enter if and only if we observe them starting exactly one other job during 2010. The number of workers with multiple new full-time jobs during the calendar year

[8]This information was provided in an e-mail exchange with a Brazilian human resource management consultant, Caio Canton.

[9]Through our agreement with MTE, we have access to RAIS data for 2003–2010. Carrying out our analysis on each of the previous years, using the same sample construction, produces very similar quantitative findings and the same basic conclusions. See Section 5 and the Appendix for details.

is small. We exclude them to focus on workers whose employment histories are more stable. The final analysis sample is constructed by taking the set of continuing workers, finding those with exactly one new job, and assembling all of the employer-reported information for both jobs.

Focusing on workers whose race is reported as white, brown, or black has two consequences. The first is that we eliminate workers in the very small (less than 2 percent of the population) and geographically concentrated 'amarela' and 'Indigena' categories. The second is that we exclude workers for whom race is not reported. In 2010, approximately 17 percent of workers do not have a race reported by their employer. There are two non-response categories: 'Not Identified' (4.76 percent) and 'Ignored' (12.16 percent). Among workers in the 'Ignored' category, almost all (93 percent) are public employees in the 'Defense and Social Security' sector.[10] The remaining cases with missing race amount to approximately 5 percent of the sample and are evenly distributed across sectors, occupations, and other demographic characteristics. We consider the implications of non-reporting in Section 6.2.

3.3 Descriptive Statistics

Table I reports sample averages of worker characteristics as reported by both the 'origin' job and 'destination' employer, the wage paid on each job, and several characteristics of each employing plant. Our key independent variables are indicators for each possible 'race history'. There are four possible cases: the worker is reported white by both employers (race history '11'); white by the origin employer and non-white by the destination employer (race history '10'); non-white by the origin employer and white by the destination employer (race history '01'); non-white by both employers (race history '00'). Columns (3-5) of the table report descriptive statistics for workers with each race history.

We calculate several plant-level summaries and merge them to our primary analysis sample. In calculating plant-level summaries, we use data from all RAIS workers – not just the job changers we otherwise focus on. For each

[10] The sector of employment corresponds to the United Nations' International Standard Industrial Classification (ISIC) rev.3 code 75: "Public administration and defense; compulsory social security".

plant we find all workers who were employed on January 1, 2010, and measure their average log wage, the share that are reported white, and the total number of such workers. These become our measures of the mean log wage, share white, and employment, respectively. We repeat these measurements for each plant, using instead workers employed on December 31, 2010. To compute the separation rate, we count the total number of jobs in the plant that were reported to have ended for any reason, and divide by the simple average of beginning-of-year and end-of-year employment. For the origin job, we use the beginning-of-year plant characteristics. For the destination job, we use the end-of-year plant characteristics.

3.3.1 Sample Selection

Column (1) reports summary statistics for continuing workers – all workers at risk for inclusion in our analysis sample of job changers. Column (2) reports summary statistics for the analysis sample. For the sample of continuing workers, most do not have a second job, so we report descriptive statistics just on the origin job. There are 26,512,018 continuing workers, of whom 3,000,688 (11 percent) are in the job-change sample. Relative to this population, workers who change jobs are slightly less white, more likely to be male, and slightly less educated. Job changers are younger, with an average age of 31 versus 35 among all continuing workers. Job changers have slightly lower average wages and are employed in smaller plants. Workers changing jobs are drawn from plants with much higher levels of turnover. The average plant-level separation rate among continuing workers is 0.633. Among job changers, the average plant-level separation rate is 1.15 – nearly twice as large. While one might expect workers who change jobs to be quite different from workers who do not, the largest observable difference is in the kinds of plants that employ them.

3.3.2 Race Histories and other Individual Characteristics

The white share of the workforce is 62 percent, whether we measure the race as reported by the origin or the destination employer. The stability of this stock measure masks rather large flows of workers between racial classifications. Job mobility is associated with a large rate of 'racial churn'. Among the sample of job changers, 27.1 percent are reported with a different race by their origin and destination employer. Of these, 14 percent are classified as white by the original employer, and as non-white by the destination employer. A slightly smaller flow, 13 percent of workers, make the reverse transition – classified as non-white by the origin employer and classified as white by the destination employer.[11] Among workers whose race is consistently reported by both employers, 48.5 percent are reported to be white by both, and 24.4 percent are reported to be non-white by both.

Our primary dependent variable is the natural logarithm of the monthly wage in 2003 Brazilian Reais.[12] Contracts that specify the wage rate by month rather than by hour are common in Brazil. Wages increase, on average, among our sample of job changers. The average log monthly wage is 6.404 (604 2003 Brazilian Reais) at the origin job, and 6.460 (639 2003 Brazilian Reais) at the destination job.

Employers also report gender, age, and educational attainment. Table I shows the share male (71.7 percent) and average age (31.4 years) are the same when reported by origin or destination employer. Age and gender are reported with great, but not perfect, consistency by different employers. There is no difference in age, on average, as reported by different employers, though we do find cases of disagreement. Across our sample, approximately 2 percent of workers are reported with a different gender by their destination employer. The greatest inconsistency is in reported education.[13] Forty-four percent of workers have different levels of education reported by the origin and destination employers. Furthermore, 18 percent of workers are reported with less education

[11] If these estimates represent stable flow rates, then over time, the workforce should become less white. That is indeed what we observe when measuring the white share as reported by origin and destination employer (62.4 versus 61.8 percent white). The share of workers reported as white is also decreasing across years in RAIS.

[12] We refer to this measure as a monthly wage, though technically the variable is reported as the average monthly earnings. When the worker separates mid-month, his earnings are adjusted so the average monthly earnings reflect what the worker would have earned had he stayed the full month. This is done so the average monthly earnings may be accurately compared with the monthly minimum wage for calculating the value of wage supplements.

[13] To save space, we only show education as reported by the destination employer.

by the destination employer than by the origin employer.

The greatest consistency in employer reports of individual characteristics are on variables about which there is little uncertainty. The worker's date of birth is recorded on the CTPS, which is provided to all employers. Gender is not on the CTPS, but is arguably subject to much less ambiguity than skin tone. Education is verifiable in some cases, but employers only require verification of the level of education required by the job. Therefore, employer-reported education may proxy for both the skill demand of the job as well as the general human capital accumulated by the worker. In our analysis, we control for employer-reported education on both the origin and destination job. We also address the possibility that race change is correlated with changes in reported education.

3.3.3 Race Change and Plant Characteristics

We focus next on the contrast in Columns (3)-(5) between workers who are consistently reported as white by both employers and those whose employer-reported race changes. Workers with race histories '10' and '01' have lower average wages than workers with race history '11'. They are also around ten percentage points more likely to be white, are slightly older, and have slightly less education. Among workers who change race, those who move from white to non-white ('10') are demographically nearly identical to those who move from non-white to white ('01').

By contrast, there is a clear association between race change and plant characteristics. Among workers with race history '11', the average share of white workers is 82 percent in both the origin and destination plant. Among workers whose reported race changes, those with race history '01' on average move from plants that are 36 percent white to plants that are 75 percent white. They also move to slightly smaller plants. Those with race history '10' move from plants that are 75 percent white to plants that are 37 percent white, and also to larger plants.

4 Modeling Racial Classification and Wages

We observe race as it is reported by an individual's employer. Because we follow the same individuals across two jobs, we observe how their wages change and how their employers' reports of race change. In principle, the variation over time in reported race provides variation in racial classification that is separate from fixed unobservable worker attributes. We would like to exploit this variation to measure the manner in which race affects labor market earnings, holding individual ability constant. An obstacle to implementing this strategy is that the observed variation in racial identity might reflect measurement error rather than true variation in the process determining individual wages. The following model develops a formal test of the measurement error hypothesis.

We posit three different notions of race:

- The 'market race' that determines the data generating process from which wages are drawn (r^*).
- The 'employer race' that is reported by an individual's employer at the date of hire (r^M).
- The 'self-race'; a worker's self-reported race, or what she would report to a survey enumerator (r^S).

A worker's wage is drawn from a distribution that depends on observable characteristics, unobservable stationary characteristics, and the 'market race'. It is common in studies based on household survey data to assume that market race is immutable and equal to self-reported race ($r^* = r^S$). In principle, though, the employer's perception of race should matter more if discrimination is driven by the employer's tastes or beliefs. When race is subjective, as is the case in Brazil, the employer and the individual may perceive, and report, race differently.

A difficulty in applying a misclassification model to our setting is that there is no ground truth behind racial categories. Race is whatever people decide it is in a particular setting. Defining race as we have avoids taking a stand on the meaning of racial categories. The race that determines which wage equation a worker draws from is potentially different both from the race that is reported by the employer and from the race that the worker would report in a survey. These are both potentially noisy measures of the racial characteristic that affects the data-generating process.[14]

4.1 A Proposed Test of Pure Misclassification

Our purpose is to exploit variation in the employer's report of race, r^M, to help identify the effect of race on wages. This approach is based on the assumption that the variation in reported race is associated with variation in the data-generating process determining wages. An alternative possibility is that race really is an immutable characteristic as far as wage determination is concerned. In that case, observed variation in the employer's report of race is pure measurement error. If so, we cannot use that variation to identify the effect of race. At best, we can use the observed variation to find bounds on the attenuation bias in the measured relationship between race and wages.

We develop a test of the assumption that variation in racial classification is pure measurement error. Our approach closely follows Card (1996), who estimates the effect of union status on wages using longitudinal data in a setting where union status may be misclassified. Detailed derivations are removed to Appendix A.

We begin by expressing wages as

$$\ln w_{it} = a_t + \beta_t x_i + \delta r^*_{it} + \varepsilon_{it} \tag{1}$$

where $\ln w_{it}$ is the log monthly wage reported by worker i in period $t \in \{1, 2\}$ and x_i is a vector containing the history of time-varying worker and plant characteristics. Here, a 'period' coincides with an employer, so the elements of x_i correspond to origin and destination employer values. Our goal is to test whether the market race is constant within individuals; that is, whether the data are best explained by a model in which each worker always draws from the same wage distribution. We allow wages and race to be correlated with an

[14] Abowd and Stinson (2013) make the related point that earnings are measured with error in both survey and administrative data sources.

additive, unobserved person-specific effect (α_i), which implies that the error in (1) can be written as $\varepsilon_{it} = \alpha_i + \varepsilon'_{it}$.

We consider two racial categories, white (1) and non-white (0). Let R^*_{ih} be an indicator for the hth possible race history, $h \in \{00, 01, 10, 11\}$. We assume that R^*_{ih} is strictly exogenous with respect to ε'_{it}, so that $\mathrm{E}(R^*_{ih}\varepsilon'_{it}) = 0$ for all h and t. If workers are always paid according to the same wage-generating process – their market race does not change over time – we get the testable restriction that the set of possible race histories is limited to $\{00, 11\}$.

In the spirit of Chamberlain (1982), we take α_i to be a linear function of the race-history indicators and observable worker and plant characteristics:

$$\alpha_i = \phi_1 + \sum_{h \neq 00} R^*_{ih}\phi_h + \lambda x_i + \xi_i, \tag{2}$$

where $\mathrm{E}[(R^*_{ih},\, x_i)\xi_i] = 0$. Thus, the complete two-period (employer) model of wages is given by

$$\ln w_{i1} = a_1 + \phi_1 + (\beta_1 + \lambda)x_i + (\delta + \phi_{10})R^*_{i10} + \phi_{01}R^*_{01} + (\phi_{11} + \delta)R^*_{i11} + \xi_i + \varepsilon'_{i1} \tag{3}$$

$$\ln w_{i2} = a_2 + \phi_1 + (\beta_2 + \lambda)x_i + \phi_{10}R^*_{i10} + (\phi_{01} + \delta)R^*_{01} + (\phi_{11} + \delta)R^*_{i11} + \xi_i + \varepsilon'_{i2}. \tag{4}$$

The employer's report of race, r^M_{it}, which we observe, may not accurately measure the market race, r^*_{it}. Let R_i be a vector of observed race-history indicators, $R_i = (R_{i01}, R_{i10}, R_{i11})$, where R_{i00} is the baseline category. Then, consider the system of equations projecting the each possible race history R^*_{ih}, $h \in \{01, 10, 11\}$, onto R_i and x_i:

$$R^*_{ih} = \gamma_{0h} + \gamma_h R_i + \gamma_{xh}x_i + \eta_{ih}. \tag{5}$$

The elements of γ_h capture the conditional correlation between each true history h and the observed race histories $k = 01, 10, 11$. If there is no misclassification, $\gamma_{h,k} = 0$ for all $k \neq h$ and $\gamma_{h,h} = 1$ for all h.

Substitution of (5) into the structural wage equations, (3) and (4), leads to the reduced-form model for wages in terms of worker and plant characteristics and observed race histories:

$$\ln w_{i1} = a_1' + b_1 x_i + d_1 R_i + e_{i1} \tag{6}$$

$$\ln w_{i2} = a_2' + b_2 x_i + d_2 R_i + e_{i2}. \tag{7}$$

Our interest is in the parameters measuring the conditional correlation between wages and observed race histories:

$$d_1 = (\delta + \phi_{10})\gamma_{10} + \phi_{01}\gamma_{01} + (\delta + \phi_{11})\gamma_{11} \tag{8}$$

$$d_2 = \phi_{10}\gamma_{10} + (\delta + \phi_{01})\gamma_{01} + (\delta + \phi_{11})\gamma_{11}. \tag{9}$$

By construction, the composite errors, e_{i1} and e_{i2} are uncorrelated with x_i and R_i. Consistent estimates of d_1 and d_2 can therefore be obtained by applying OLS to (6) and (7).

In the absence of measurement error, the discrimination coefficient, δ, is identified by differencing the parameters associated with R_{10} and R_{01}. However, measurement error will lead to bias and cannot be resolved without further information on the misclassification process. For example, the difference in parameters associated with observed history R_h is

$$d_{2,h} - d_{1,h} = \delta(\gamma_{01,h} - \gamma_{10,h}). \tag{10}$$

Under additional assumptions about the misclassification process, we can estimate the bias parameters ($\gamma_{k,h}$) and then test whether the data could have been generated by a model in which market race never changes within person.

4.2 The Misclassification Process

If misclassification is independent of observables, conditional variation in employer-reported race is informative about the underlying distribution of market race histories, R_i^*. We assume misclassification is constant across workers and independent across employers. Formally,

$$P(r_{i1}, r_{i2}|r^*_{i1}, r^*_{12}, x_i) = P(r_{i1}|r^*_{i1}) \cdot P(r_{i2}|r^*_{i2}). \tag{11}$$

Define $P(r_{it} = 1|r^*_{it} = 1) = q_1$ and $P(r_{it} = 1|r^*_{it} = 0) = q_0$. Hence, q_0 is the probability of a false positive, and $1 - q_1$ is the probability of a false negative.

Define π as a vector of population shares of workers with $R^*_{ih} = 1$ and p as a vector of population shares of workers with $R_{ih} = 1$, $h \in \{00, 01, 10, 11\}$. Let T be the 4×4 matrix whose (j, k) element is the misclassification probability $\tau_{jk} = P(R_{ij} = 1|R^*_{ik} = 1)$. Then, true and observed race histories are related as follows:

$$p = \mathrm{E}(R_i) = \mathrm{E}(R^*_i T) = \pi T. \tag{12}$$

Because p is observable, with assumptions on the misclassification probabilities, q_1 and q_0, we can recover the bias parameters in γ. Consider the projections of true and observed race histories onto worker and plant characteristics, transformed into deviations from means so that the constant terms represent the relevant population shares:

$$R^*_{ih} = \pi_h + (x_i - \bar{x})c_h + \nu_{ih} \tag{13}$$

$$R_{ih} = p_h + (x_i - \bar{x})\zeta_h + \nu'_{ih} \tag{14}$$

It is then straightforward to show $\zeta^{\mathrm{T}} = \Omega c^{\mathrm{T}}$ where Ω is a matrix whose j, k entry is $\tau_{jk} - \tau_{j00}$.

Finally, using (13) and (14), we write γ_h in (5) as

$$\gamma_h = \left[\mathrm{var}(R) - \Omega c^{\mathrm{T}} V_{xx} c \Omega^{\mathrm{T}}\right]^{-1} \cdot \left\{\mathrm{cov}(R, R^*_h) - \Omega c^{\mathrm{T}} V_{xx} c_h\right\}, \tag{15}$$

where V_{xx} is the covariance matrix of x_i. Intuitively, γ_h is identified from between-group variation in the white share along with modeling assumptions on the misclassification probabilities embodied in Ω.

4.3 Estimation and Testable Restrictions

The model is estimated in two stages. First, we estimate the reduced-form models for wages and observed race histories from (6), (7), and (14). Second, we use a minimum distance estimator to fit nine unrestricted sample moments, (d_{11}, d_{12}, d_{13}, d_{21}, d_{22}, d_{23}, p_{11}, p_{10}, p_{01}), to nine parameters, (q_1, q_0, π_{11}, π_{10}, π_{01}, ϕ_{11}, ϕ_{10}, ϕ_{01}, δ). The estimating equations are those relating the structural parameters to the reduced-form parameters on observed race histories, (8) and (9), and the equations defining the misclassification model, (12).

We test two models that are nested within the unrestricted model. In the first, market race does not change across employers, implying the observed variation in employer-reported race is uninformative. This model imposes the testable restrictions: $\pi_{10} = \pi_{01} = 0$.[15] Furthermore, if there is no variation in market race, we cannot separately identify the discrimination parameter, δ, from the part of the person effect correlated with race, ϕ_{11}. Instead, we identify the combined effect, $\kappa \equiv (\delta + \phi_{11})$.[16] In the second, market race is the employer-reported race. If correct, there is no measurement error, which implies the parameter restrictions $q_1 = 1$ and $q_0 = 0$.

We test both models comparing the values of the minimized objective function with (Q_r) and without (Q_{nr}) the restrictions imposed. The test statistic, $N \times (Q_r - Q_{nr})$, is asymptotically χ^2 under the null with degrees of freedom equal to the number of restrictions.

5 Results

We present our main results as estimates of the reduced-form relationship between wages and observed race histories (6) and (7). We first establish benchmark estimates of cross-sectional wage gap between white and non-white workers. Next, we report the estimated reduced-form effect of race change on wages. We then formally test, and reject, the hypothesis that the data are

[15] The model also imposes the restrictions $\phi_{10} = \phi_{01} = 0$, but technically these parameters are not identified.

[16] This is the classic problem, that in fixed effects estimation it is not possible to separately identify the parameters associated with fixed observable characteristics

generated by a model in which market race does not vary across jobs. We also are unable to reject a model in which the market race is identical to employer-reported race.

While we focus our attention on the findings produced from our 2010 RAIS sample, we also carry out the same analysis on all available years (2003–2010) with comparable results. We present the reduced-form results for all years in Appendix A.

5.1 Cross-Section White Wage Gap

Table II reports the estimated cross-sectional log wage gap between white and non-white workers. Columns (1) and (2) estimate the gap for all continuing workers, regardless of whether they enter the sample of job changers. In a model that controls for gender, education, a quadratic in age, along with controls for industry and state of employment, the estimated wage gap is 0.132 (Column (1)), but adding plant characteristics erases about 40 percent of it (Column (2)). Columns (3) and (4) restrict attention to workers who change jobs, and present the estimated wage differences on the origin and destination jobs, using the same specification as in Column (2). Whites earn about 6.5 percent more at the origin job and 4.8 percent more at the destination job. Tables A.2 reports estimates of cross-section white/non-white wage gap and reduced-form wage model for each year from 2003-2010, showing they are quite consistent over the period.

5.2 Reduced-Form Model for Wages

Table III presents estimates of the observed race-history (R_i) coefficients in the reduced form wage equations, (6) and (7). The results are conditional on a set of covariates (x_i), which include a worker's gender, education, age (as a quadratic), industry and state, as reported by their origin and destination employers, and the mean log wage, share white, employment, and separation rate of the origin and destination plants. In Column (1), the dependent variable is the log wage on the worker's origin job. In Column (2), the dependent variable is the log wage on the workers destination job. In Column (3), the dependent variable is the difference between the log wage on the origin and destintion job. The specification in Column (3) represents a benchmark against which we compare subsequent estimates.

Surveying the results in Columns (1) and (2), we find that workers who are reported as white by a given employer earn more from that employer than workers who are reported as non-white. Not surprisingly, the largest premium – on the order of 7 percent – accrues to those who are reported as white on both jobs (race history '11'). Workers reported as white in the origin job (race history '10') earn a premium of 4.6 percent, while those reported as white on the destination job (race history '01') earn a 3.3 percent premium. In contrast, starting out and ending up non-white carry smaller estimated wage effects of 1.6 and 2.5 percent.

A goal of our analysis is to separate wage discrimination from differences in unobservable, but fixed, worker-specific characteristics. If the observed race histories really correspond to differences in compensation – if there is no measurement error – then the effect of race on wages is identified by the wage changes of workers who also change race. The estimates in Column (3) measure the difference between reduced-form parameters, $\hat{d}_2 - \hat{d}_1$, as described in Equation 10. Workers who are reported as non-white on the origin job and then white on the destination job experience an average wage gain of 1.7 percent. In contrast, workers who make the racial-status transition in the other direction realize a loss in wages of 2.1 percent, on average. Finally, the estimated residual wage change for those workers who are reported white by both employers is almost an order of magnitude smaller at –.03 percent. All results are statistically distinct from zero, though the effect associated with R_{11} is measured with much less precision. Table A.3 reports estimates of the benchmark specification from Column (3) for each year from 2003–2010. The estimates suggest our findings are largely invariant to the sample year.

5.3 Tests of the Misclassification Model

The misclassification model predicts the estimated wage effect associated with race histories '10' and '01' should be zero. They are not, indicating that the variation in employer-reported race is systematically correlated with the earnings process. Further, the symmetry of the wage changes associated with changing race, along with the relatively small estimated effect associated with race history '11' are inconsistent with a measurement-error story. We now formally test the implications of these alternative models of the data-generating process.

Table IV reports tests of two restricted versions of the misclassification model of Section 4. Column (1) reports the 'No Race Change' model, in which the market race of each worker is immutable, and does not change from job to job. Column (2) reports the 'No Measurement Error' model, in which the market race is identical to the observed employer-reported race. Each model is fit to the reduced-form parameter estimates from Table III and the corresponding population shares from Table I. The structural model also involves estimation of a reduced-form linear probability model for each of the observed race histories (14), the details of which are given in Table VI.

The test of the parameter restrictions in the 'No Race Change' model is the key result. If market race is immutable, then only four model parameters are identified: the share of workers who are always white, π_{11}, the true-positive and false-positive parameters, q_1 and q_0, and the composite parameter, $\kappa = (\delta + \phi_{11})$. As discussed in Section 4, we test these restrictions using the statistic $N \times (Q_r - Q_{nr})$. In this case, value of the test statistic is 1588, so the null that observed race changes are not associated with wage changes is soundly rejected.

The alternative version of the model is that employer-reported race always corresponds to the way workers are paid, so that there is no measurement error. In this case, the restrictions, $q_1 = 1 = (1 - q_0)$, are supported by the data. The value of the test statistic is only 0.531. Unsurprisingly, with no measurement error in race, the effect of race on wages is very similar to the reduced-form differences in race history coefficient estimates for race changers reported in Table III. The true coefficient of wage discrimination is $\hat{\delta} = 0.019$, which is approximately 40 percent of the estimated cross-section wage gap of 0.048 reported in II. We report the complete set of structural parameter estimates in Table A.1.

6 Possible Mechanisms and Alternative Specifications

We now consider possible behavioral mechanisms that cause wages to change with employer-reported race. We begin by summarizing predictors of race change. Tables V and VI report estimates from the reduced-form linear probability models for each observed race histories, R_{11}, R_{10}, and R_{01}. These models are estimated as part of the misclassification model, and include the same control variables as the reduced-form wage equations. In addition to the variables reported in Tables V and VI, all models include controls for industry and state of the origin and destination plant.

We will focus on several features of Tables V and VI. First, race change is weakly associated with worker characteristics, but strongly associated with plant characteristics on the origin and destination job. Plant characteristics provide almost all of the explanatory power; individual characteristics explain very little. More specifically, race change is most strongly predicted by two plant characteristics: (1) the share of white co-workers at the plant, and (2) the average log wage of the plant. Workers are more likely to be reported white when a large share of their co-workers are white. A worker is more likely to be reported as non-white by the origin employer and white by the destination employer (race history '01') when the share of white co-workers at the destination employer is high and the share of white co-workers at the origin plant is low. Workers are more likely to move from white to non-white when they are moving into plants with a *higher* average wage. They are more likely to move from non-white to white when moving to a plant with a lower average wage. Finally, there is a strong symmetry the coefficient estimates on the share white and the plant average log wage at the origin and destination plants for race histories 10 and 01.

Our results on wage determination in Table III along with the segregation exhibited in Figure 1 are consistent with the presence of employer discrimination. The correlates of race change support a model in which workers manipulate the way race is perceived to obtain employment in a discriminatory labor market. Such behavior is plausible in the Brazilian context, and has a coherent economic foundation, which we discuss at length toward the end of this section. Before we do, we consider two alternative mechanisms.

6.1 Reverse Causality: Does "Money Whiten"?

It is possible that changes in earnings lead to changes in the way race is reported; that is, that "money whitens" (Schwartzman 2007). Perhaps workers are more likely to report themselves, or to be classified by company representatives, as white when they enter a high-status, high-paying job. These concerns are not mere speculation: there is evidence that racial classification in Brazil is affected by socio-economic status. Using a 1995 survey, Telles (2002) shows interviewers classify respondents with high levels of education as white, even when the repsondents identify themselves as brown. Schwartzman (2007) finds parents of higher socio-economic status are more likely to classify their children as white.

A formal test of reverse causality is difficult to construct without imposing more structure on the analysis. If race change is associated with moves to higher-status jobs, we would expect the effect to be driven, in part, by changes in occupation or changes in required education. Table V shows education is, if anything, negatively correlated with race change. Workers with at least some college are generally less likely to change race in either direction. Second, as we discuss later, the wage effect of race change is not attenuated when we restrict our sample to jobs on which education does not change. Finally, the wage effect of race change is also not diminished when we add controls for the occupation of the origin and destination job.

6.2 Alternative Mechanism: Plant-Specific Reporting Behavior

Tables I and VI show that race change is strongly associated with the share white in the plant. We argue this is consistent with workers changing reported race as part of obtaining employment in segregated plants. An alternative non-economic explanation is that some employers systematically misreport race. This might happen if, for instance, plants with poor human resource management systems simply classify workers as either white or non-white 'by default' when race information is missing.

As discussed in Section 3, a worker's race may be missing because it is either 'Not Identified' or 'Ignored' by the employer. We now leverage the plant-level variation in missing-race information to examine whether our results could be explained by plant reporting behavior. If race change is driven by certain plants using default 'imputation' of racial classifications, then we should observe workers changing race more often in moves to plants that consistently report race. Furthermore, controlling for the extent of non-reporting should attenuate the estimated wage effect of changes in employer-reported race.

So, first we estimate the effect of plant-level non-reporting on race change (in either direction). Table VII presents the results from two linear probability models. The first captures the simple link between race change and the share of the destination plant's workers without a reported race (Column (1)). The second adds the complete set of controls from Table V (Column (2)). If anything, race change is between 1 and 3 percentage points *less* likely in moves to plants that consistently report race. The opposite would be true if these plants systematically assigned a particular race to every worker with missing data.

Next, we explore whether plant-level non-reporting can account for the estimated effect of race change on wages. Table VIII provides the results of this exercise, carrying over the benchmark specification in first-differences from Table III. In Column (2), we include controls for the share of workers in both the origin and destination plants without a reported race. Compared with the benchmark estimates, the payoff to becoming white is larger and closer in magnitude to the penalty associated with movement in the opposite direction. In addition, the estimated effect of being reported white at both jobs falls sharply and becomes statistically insignificant. We then restrict the analysis to workers whose origin and destination employer always report race (Column (3)) and have at least some non-reporting workers (Column(4)). While there is some variation in the point estimates associated with the white/white and non-white/white race histories, the pattern of results remains consistent with the benchmark model.

Although our findings cannot be accounted for by plant-level reporting behavior correlated with missing race information, there may be other (unobserved) plant-level reporting policies that may confound our analysis. To address this issue, we re-estimate the benchmark wage model controlling for arbitrary destination plant heterogeneity (Column (5)).[17] The estimated wage effect drops to 0.010 from 0.017, but the pattern of results remains the same. Indeed, some attenuation should be expected, because now the wage effect is identified solely from workers with different race histories who move to the same plant. In this specification, which we view as conservative, race change still accounts for 20 percent of the baseline cross-section wage gap.

6.3 Robustness

6.3.1 Alternative Sources of Variation

The reduced-form wage model restricts how individual heterogeneity enters the model. Table IX introduces an alternative specification that controls for individual heterogeneity in the destination wage by directly controlling for the wage on the origin job:

$$w_{i2} = a + \zeta w_{i2} + bx_i + m \times OrigWhite_i + \theta_{10}R_{10} + \theta_{01}R_{01} + \psi_{J(i2)} + e_{2i}. \quad (16)$$

This specification relaxes the implied restriction of reduced-form model that $\zeta = 1$.[18] The covariate vector, x_i, still includes all worker and plant characteristics from the origin and destination job. The model also controls for arbitrary plant heterogeneity on the destination job through the plant effect, $\psi_{J(2i)}$ (where $J(i2)$ indicates that plant j employs worker i in period 2). For clarity of presentation, we change the set of race controls in the model, including an indicator for whether the worker is reported white on the origin job, $OrigWhite$, along with dummies for race history '10' and race history '01'.

[17] This specification controls for arbitrary plant and worker-specific heterogeneity in the spirit of Abowd et al. (1999).

[18] The estimated persistence in wages, 0.307, is in line with other estimates of wage changes or earnings volatility associated with job change (Hospido 2010; Schmutte 2015). Estimates of wage persistence based on within-job variation are typically much higher.

Therefore, we interpret the coefficient on R_{10} as the wage gap for a worker who is reported as white on the origin job and non-white on the destination job relative to a worker who was reported as white on both jobs. The coefficient on R_{01} has an analogous interpretation as the wage gap for a worker who is reported as non-white on the origin job and white on the destination job relative to a worker reported non-white on both jobs.

Using this alternative source of identifying variation, which controls for all wage-relevant characteristics of the worker as well as arbitrary plant-level characteristics, we obtain results that are quite similar to the benchmark model. Workers whose race changes from white to non-white earn -0.034 less than workers who remain white on both jobs. Workers whose race changes to white from non-white earn 0.022 more than workers who are non-white on both jobs.

6.3.2 Endogenous Mobility

Our empirical model is motivated by Card (1996) and related research using longitudinal data to estimate the effect of employer characteristics, such as industry, on wages. A concern in such studies is that the decision to change jobs is based on new information about the current match, the new match, or both. In that case, the estimated effect of race change may not be attributable to employer-reported race, *per se*, but reflects a correlation between match characteristics, wages, and the way the employer reports race.

We address two specific forms of endogenous mobility. The first is that the employer may be more likely to report a worker is white when the worker makes a direct job-to-job move and to report him or her as non-white when the worker is hired from non-employment (or vice-versa). The second is that the employer may be more likely to report a worker is white when the job has high education requirements. Either case could explain the observed pattern of results. Before we present our analysis, we note that in our context, any correlation between employer-reported race and wage outcomes is informative about how race is related to, and determined by, labor market phenomena.

Table X reports estimates of the reduced-form wage model, expressed in terms of the difference between (6) and (7), restricted to particular types of job change. Column (1) repeats the benchmark specification from Table

III. Column (2) restricts the sample to workers whose job change involves a spell of unemployment lasting at least one full month (job-unemployment-job; JUJ). This specification is inspired by Gibbons and Katz (1992), who focus on displaced workers to alleviate endogenous mobility bias in estimating the inter-industry wage premium. The estimated coefficients for workers whose reported race changes are not affected at all. There is a slight increase in the magnitude of the estimated effect for workers with race history '11'.

Columns (3) and (4) address the endogeneity of education changes by restricting attention to workers whose education does not change (Column 3) and workers whose employer-reported education decreases when they change jobs. In the latter case, we see a modest attenuation of the estimated effect of changing to white (from 0.017 to 0.013). It is possible that some of the apparent effect of being reported as 'white' comes from workers who also move into jobs where they also have a higher reported level of education. The possibility remains that workers and employers manipulate both perceived race and education, or that employer perception of race is affected by the type of job a worker obtains. These relationships suggest many possible extensions of our research.

6.4 Discussion

We conclude this section with a discussion of what may be driving observed changes in racial classification. Workers whose employer-reported race changes from non-white to white are typically moving from non-white to white majority plants, and they are moving from plants with higher average pay to plants to lower average pay. These patterns are consistent with job search in which workers apply for many different types of jobs, and can attempt to modify potential employers' perceptions of race. This kind of behavior could be an equilibrium outcome in extensions of a directed search model with wage posting of the type developed by Lang et al. (2005).

We have excluded the possibility that the effect of changing employer-reported race on wages is driven by employer misreporting. An alternative possibility is that workers manipulate their employer's perception of race when searching for jobs in a discriminatory environment. If observed workplace segregation reflects discrimination, then workers would manipulate race precisely to obtain employment with those discriminating firms. This is related to the economic mechanism Heckman (1998) and Neumark (2012) use in their criticism of correspondence studies: discriminating firms do not hire non-white workers. Discriminating firms may occasionally hire non-white workers, but are much more likely to do so if they believe they are white.

We also observe that workers are more likely to change race when moving into a plant with lower average pay, even though the race change itself is associated with a wage premium. This, too, is consistent with a labor market search equilibrium in which workers with scope to manipulate perceived race should be indifferent in applying to and accepting jobs across different types of plants. In an equilibrium, discriminating plants know there is some scope for manipulation. The plants that tolerate manipulation should be lower paying.

This sort of argument can accommodate a counter-intuitive feature of the data as well. We observe workers changing racial classification from white to non-white and drawing a wage penalty when they do. In an equilibrium where workers and employers know there is scope for manipulation, workers who can manipulate perceived race should actually be indifferent between applying for a job where they will be reported as white and one where they are reported as non-white. In our setting, the observed wage premium for being perceived as white may be offset by a reduction in the probability of being hired. This would be consistent with the fact that more workers change from white to non-white than from non-white to white.

While the theoretical framework presented in this discussion helps resolve our reduced-form evidence, other frameworks are possible, and the details of such a model remain to be worked out. Furthermore, if we are correct that changes in reported race are the product of strategic behavior on the part of workers, it may ultimately be necessary to systematically address the endogeneity between reported race and wages. The problem of endogenous mobility in notoriously difficult in estimating wage differentials in other settings (compensating differentials, inter-industry differentials, etc.). Here, our goal has been to document variation in employer-reported race and show that we can use that variation to identify the presence of labor market discrimination in wages. In doing so, we invite new theoretical and empirical research to clarify and applies this source of variation to better the nature of racial discrimination.

7 Conclusion

If it were possible, a rational response to racial discrimination would be to 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 subjectivity in racial identity and opportunities to manipulate the way other people see one's race. Using employer-employee matched data from Brazil, we observe what may be the outcome of this process.

We show that when workers change jobs in Brazil, sometimes their new employer will report a different race than their previous employer. The results of our structural misclassification model imply that the observed variation in race is associated with variation in wages. Under the model, we separately identify the part of wages due to changes in the employer's report of race from other unobservable, non-varying observable characteristics that could affect the wage. This task would typically be considered impossible since almost all previous economic research adopts the perspective of race an immutable individual characteristic.

It is more difficult to pin down the precise mechanism that leads one employer to report a worker's race as white and another employer to report her race as non-white. The economics of the situation implicate workers. In an environment where hiring, coworker relations, performance evaluation and advancement are characterized by discrimination, and where there is scope to manipulate how others perceive race, some workers can and will manipulate race. This perspective suggests an avenue for further theoretical research: if race becomes completely subjective and malleable, it will be impossible to support discrimination since anyone could change their race. Precisely how subjective can race be and still support discrimination? Extending Lang et al. (2005) to incorporate race change is one avenue we have pursued to formally model this phenomenon. Their framework, along with and extensions described in Lang and Lehmann (2012) will be useful to address the facts outlined in our study.

The results of this research, and the perspective on analyzing racial discrimination we advance, are relevant beyond the Brazilian context. In the U.S., laws prohibiting interracial marriage were not fully eliminated until 1967. The rate of interracial marriage in the U.S. increased from 6.7 percent of new marriages in 1980 to 15.1 percent in 2010. These trends together with the election of the nation's first African-American president in 2008 have prompted a public discussion over whether the U.S. is becoming a 'post-racial' society. Brazil's experience suggests that a high rate of inter-racial socialization can co-exist with persistent racial inequality and discrimination, while the measurement of racial categories, and hence discrimination, becomes more complex.

The issues of racial subjectivity and the manipulation of perceived race we address are echoed in recent demographic research in the U.S. The difficulty of measuring racial identity in the U.S. is the subject of a recent book by a former director of the U.S. Census Bureau (Prewitt 2013). The Census recently changed its procedure for collecting information on race to allow for more detailed responses, shedding new light onto the complexity with which individuals perceive their own race. In a widely publicized paper, Liebler et al. (2014) document large changes in the self-reported race across the 2000 and 2010 Decennial Censuses for the U.S. Their results suggest extensive 'racial churn' of individuals moving back and forth between racial categories. Their results echo the similar churning we observe among Brazilians in employer-reported race. Saperstein and Penner (2012) also document changes in self-reported race in the 1997 National Longitudinal Study of Youth. As our evidence and the Brazilian context suggest, these trends do not imply discrimination will disappear, but that economists will need to become more sophisticated in our treatment of race.

References

Abowd, J. M., Kramarz, F. and Margolis, D. N. (1999). High wage workers and high wage firms, *Econometrica* **67**(2): 251–333.

Abowd, J. M. and Stinson, M. H. (2013). Estimating measurement error in annual job earnings: A comparison of survey and administrative data, *The Review of Economics and Statistics* **95**(5): 1451–1467.

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

Bertrand, M. and Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination, *American Economic Review* **94**(4): 991–1013.

Biavaschi, C., Giulietti, C. and Siddique, Z. (2013). The economic payoff of name Americanization, *IZA Discussion Papers 7725*, Institute for the Study of Labor (IZA).

Card, D. (1996). The effect of unions on the structure of wages: A longitudinal analysis, *Econometrica* **64**(4): 957–79.

Daniel, G. (2010). *Race and Multiraciality in Brazil and the United States: Converging Paths?*, Pennsylvania State University Press.

Duncan, B. and Trejo, S. J. (2011). Intermarriage and the intergenerational transmission of ethnic identity and human capital for Mexican Americans, *Journal of Labor Economics* **29**(2): pp. 195–227.

Fiola, J. (1990). *Race Relations in Brazil: A Reassessment of the "Racial Democracy" Thesis*, Occasional papers series, Latin American Studies Program. University of Massachusetts at Amherst.

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* pp. 1–35.

Gibbons, R. and Katz, L. F. (1992). Does unmeasured ability explain inter-industry wage differentials?, *Review of Economic Studies* **59**(3): 515–35.

Heckman, J. J. (1998). Detecting discrimination, *The Journal of Economic Perspectives* **12**(2): pp. 101–116.

Hospido, L. (2010). Job changes and individual-job specific wage dynamics, *IZA Discussion Papers 5088*, Institute for the Study of Labor (IZA).

Krishna, P., Poole, J. P. and Senses, M. Z. (2014). Wage effects of trade reform with endogenous worker mobility, *Journal of International Economics* **93**(2): 239 – 252.

Lang, K. and Lehmann, J.-Y. K. (2012). Racial discrimination in the labor market: Theory and empirics, *Journal of Economic Literature* **50**(4): 959–1006.

Lang, K., Manove, M. and Dickens, W. T. (2005). Racial discrimination in labor markets with posted wage offers, *The American Economic Review* **95**(4): pp. 1327–1340.

Liebler, C., Rastogi, S., Fernandez, L., Noon, J. and Ennis, S. (2014). Americas churning races: Race and ethnic response changes between Census 2000 and the 2010 Census, *Technical Report 2014-09*, U.S. Census Bureau, Center for Administrative Records Research and Applications Working Paper.

Lopes de Melo, R. (2013). Firm wage differentials and labor market sorting: Reconciling theory and evidence. Working Paper.

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.

Muendler, M.-A., Rauch, J. E. and Tocoian, O. (2012). Employee spinoffs and other entrants: Stylized facts from Brazil.

Neal, D. A. and Johnson, W. R. (1996). The role of premarket factors in black-white wage differences, *Journal of Political Economy* **104**(5): 869–95.

Neumark, D. (2012). Detecting discrimination in audit and correspondence studies, *Journal of Human Resources* **47**(4): 1128–1157.

Parsons, C. A., Sulaeman, J., Yates, M. C. and Hamermesh, D. S. (2011). Strike three: Discrimination, incentives, and evaluation, *American Economic Review* **101**(June): 1410–1435.

Poole, J. P. (2013). Knowledge transfers from multinational to domestic firms: Evidence from worker mobility, *The Review of Economics and Statistics* **95**(2): 393–406.

Prewitt, K. (2013). *What is Your Race? The Census and Our Flawed Efforts to Classify Americans*, Princeton University Press.

Price, J. and Wolfers, J. (2010). Racial discrimination among NBA referees, *Quarterly Journal of Economics* **125**(4): 1859–1887.

Racusen, S. (2009). Affirmative action and identity, *Brazil's New Racial Politics.*, Lynne Rienner Publishers, pp. 89–122.

Rouse, C. and Goldin, C. (2000). Orchestrating impartiality: The impact of "blind" auditions on female musicians, *American Economic Review* **90**(4): 715–741.

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

Schmutte, I. M. (2015). Job referral networks and the determination of earnings in local labor markets, *Journal of Labor Economics* **33**(1): 0.

Schwarcz, L. M. (2003). Not black, not white: Just the opposite: Culture, race and national identity in Brazil, *Technical report*.

Schwartzman, L. F. (2007). Does money whiten? intergenerational changes in racial classification in Brazil, *American Sociological Review* **72**(6): 940–963.

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

Telles, E. E. (2004). *Race in Another America: The Significance of Skin Color in Brazil*, Princeton paperbacks, Princeton University Press.

Table I: Descriptive Statistics for Job Changers by Employer-Reported Race: RAIS 2010

	Continuing Workers (1)	Job Changers (2)	By Race History '11' (3)	'10' (4)	'01' (5)
Race History					
'11': White/White	n/a	0.485	1	0	0
'10': White/Non-White	n/a	0.139	0	1	0
'01': Non-White/White	n/a	0.132	0	0	1
'00': Non-White/Non-White	n/a	0.244	0	0	0
White					
Orig. Job	0.644	0.624	1	1	0
Dest. Job	n/a	0.618	1	0	1
Log Wage					
Orig. Job	6.536	6.404	6.462	6.390	6.376
Dest. Job	n/a	6.460	6.517	6.452	6.431
Male					
Orig. Job	0.649	0.717	0.658	0.745	0.742
Dest. Job	n/a	0.717	0.659	0.745	0.743
Age					
Orig. Job	35.010	31.4	31.1	31.4	31.3
Dest. Job	n/a	31.4	31.1	31.4	31.2
Education					
LTHS	0.446	0.461	0.409	0.461	0.477
High School	0.421	0.436	0.451	0.451	0.443
Some College	0.041	0.040	0.052	0.035	0.033
Bachelor's (+)	0.092	0.063	0.088	0.053	0.047
Plant Mean Log Wage					
Orig. Job	6.528	6.459	6.503	6.445	6.449
Dest. Job	n/a	6.510	6.556	6.510	6.493
Plant White Share					
Orig. Job	0.626	0.614	0.822	0.749	0.363
Dest. Job	n/a	0.613	0.816	0.374	0.750
Plant Employment					
Orig. Job	755.4	662.5	551.5	549.6	703.1
Dest. Job	n/a	757.6	654.2	800.2	621.0
Plant Separation Rate					
Orig. Job	0.633	1.150	1.139	1.197	1.121
Dest. Job	n/a	1.466	1.503	1.360	1.693
Num.Obs.	26,512,018	3,000,688	1,443,893	420,759	397,030

NOTE–Column (1) reports summaries for all workers who start 2010 in a continuing job. Column (2) reports summaries for our analysis sample of job changers. The remaining columns ('By Race History') disaggregate by the way race is reported by the different employers at the origin and destination job. Workers with race history '11' are reported as white by both the origin and destination employer. Workers with race history '01' are reported as non-white on the origin job and white on the destination job. In Column (1) we report characteristics as measured at the origin job. Since most continuing workers do not have a destination job, those entries are marked 'n/a'.

Table II: Cross-Section Racial Wage Gap Estimates: RAIS 2010

	All Workers		Job Changers	
			Orig. Job Wage	Dest. Job Wage
	(1)	(2)	(3)	(4)
White	0.132	0.078	0.065	0.048
	(0.0002)	(0.001)	(0.001)	(0.001)
Plant Characteristics?	N	Y	Y	Y
N	$26,512,018$	$26,512,018$	$3,000,688$	$3,000,688$
R^2	0.362	0.680	0.552	0.528

NOTE-Heteroskedasticity-robust standard errors in parentheses. Each column reports the estimated coefficient on an indicator for whether a worker is reported 'white' by their employer. Columns (1) and (2) are estimated for all workers in 2010 at risk to enter our analysis sample. The models in Columns (3) and (4) are estimated on the sample of workers who change employers. The dependent variable in column (3) is the log wage on the origin job. The dependent variable in column (4) is the log wage on the destination job. In addition to the White indicator, all models control for gender, education, and a quadratic in age. The models in columns (2), (3), and (4) also control for the following plant characteristics: industry, state, employment, white share, average log wage, and separation rate.

Table III: Reduced-Form Relationship Between Race History and Wages: RAIS 2010

	Orig. Job Wage (1)	Dest. Job Wage (2)	ΔLog Wage (3)
Race History			
'11': White/White	0.072	0.069	-0.003
	(0.001)	(0.001)	(0.001)
'10': White/Non-White	0.046	0.025	-0.021
	(0.001)	(0.001)	(0.001)
'01': Non-White/White	0.016	0.033	0.017
	(0.001)	(0.001)	(0.001)
N	$3,000,688$	$3,000,688$	$3,000,688$
R^2	0.565	0.599	0.195

NOTE-Heteroskedasticity-robust standard errors in parentheses. Estimated on a sample of workers from the 2010 RAIS observed to change primary employer during the year. The dependent variable in column (1) is the log wage on the worker's original job. The dependent variable in column (2) is the log wage on the worker's destination job. The models estimated are the reduced-form equations (6) and (7) from the misclassification model. They include a full set of indicators for the history of employer-reported race. The dependent variable in Column (3) is the difference between the log wage on the destination and origin job. The estimates are equivalent to the difference in estimates from Columns (1) and (2). All models control for gender, educational attainment, industry, state of employment, as well as the share white, employment, separation rate, and average wage at both the origin and destination plant.

Table IV: Tests of Structural Misclassification Model: RAIS 2010

	Model	
	No Race Change (1)	No Meas. Error (2)
Obj. Fcn Value	0.0005	$1.049e^{-5}$
Test Statistic	$1,588$	0.5313

NOTE–Tests of the misclassification model of Section 4.1. The table reports the value of the distance function at the solution and test statistics for each of the restricted models. Under the null hypothesis that the parameter restrictions are valid, the test statistics are distributed χ^2_d with degrees of freedom equal to the number of restrictions. The complete set of structural parameter estimates is reported in Table A.1.

Table V: Reduced-Form Observed Race History Models, Worker Characteristics: RAIS 2010

	(1) Always White '11'	(2) From White '10'	(3) To White '01'
Male			
Orig. Job	−0.010	0.006	0.001
	(0.0015)	(0.0012)	(0.0012)
Dest. Job	−0.011	0.002	0.007
	(0.0015)	(0.0012)	(0.0012)
Age			
Orig. Job	−0.005	0.003	0.003
	(0.0007)	(0.0006)	(0.0006)
Dest. Job	0.001	−0.001	−0.002
	(0.0007)	(0.0006)	(0.0006)
Age Sq.			
Orig. Job	0.000	−0.000	−0.000
	(0.0000)	(0.0000)	(0.0000)
Dest. Job	0.000	0.000	0.000
	(0.0000)	(0.0000)	(0.0000)
Education (Orig. Job)			
LTHS	0.013	0.004	−0.002
	(0.0009)	(0.0007)	(0.0007)
High School	0.025	0.006	−0.005
	(0.0007)	(0.0006)	(0.0006)
Some College	0.063	−0.003	−0.019
	(0.0014)	(0.0011)	(0.0011)
Bachelor's (+)	0.088	−0.009	−0.024
	(0.0014)	(0.0012)	(0.0012)
Education (Dest. Job)			
LTHS	0.006	0.003	0.010
	(0.0009)	(0.0008)	(0.0008)
High School	0.021	0.001	0.012
	(0.0007)	(0.0006)	(0.0006)
Some College	0.067	−0.018	0.003
	(0.0013)	(0.0011)	(0.0011)
Bachelor's (+)	0.091	−0.025	0.000
	(0.0013)	(0.0011)	(0.0011)

NOTE- This table reports estimated coefficients of worker-specific controls from the reduced-form model for observed race histories. The dependent variable is an indicator for whether the worker was reported as white by the plant at their origin and at their destination job. For example, '10' indicates the worker was reported as white by the origin plant and as non-white by the destination plant. In addition to the reported controls, the estimated model includes plant-specific characteristics, as reported in Table VI, and controls for the industry and state of the origin and destination plant. Heteroskedasticity-robust standard errors in parentheses.

Table VI: Continued – Reduced-Form Observed Race History Models, Plant Characteristics: RAIS 2010

	(1) Always White '11'	(2) From White '10'	(3) To White '01'
Plant Share White			
Orig. Job	0.528	0.435	−0.479
	(0.0008)	(0.0007)	(0.0007)
Dest. Job	0.513	−0.488	0.422
	(0.0008)	(0.0007)	(0.0007)
Plant Mean Log Wage			
Orig. Job	−0.005	−0.022	0.019
	(0.0006)	(0.0005)	(0.0005)
Dest. Job	−0.019	0.033	−0.024
	(0.0006)	(0.0005)	(0.0005)
Plant Employment			
Orig. Job	0.000	0.000	−0.000
	(0.0000)	(0.0000)	(0.0000)
Dest. Job	0.000	−0.000	0.000
	(0.0000)	(0.0000)	(0.0000)
Plant Separation Rate			
Orig. Job	0.001	−0.001	0.001
	(0.0001)	(0.0001)	(0.0001)
Dest. Job	0.000	−0.000	0.000
	(0.0000)	(0.0000)	(0.0000)
N	3,000,688	3,000,688	3,000,688
R^2	0.459	0.220	0.220

NOTE- This table reports estimated coefficients from plant-specific controls in the reduced-form model for observed race histories. The dependent variable is an indicator for whether the worker was reported as white by the plant at their origin and at their destination job. For example, '10' indicates the worker was reported as white by the origin plant and as non-white by the destination plant. In addition to the reported controls, the estimated model includes worker-specific characteristics, as reported in Table V, and controls for the industry and state of the origin and destination plant. Heteroskedasticity-robust standard errors in parentheses.

Table VII: Probability of Race Change and Plant Reporting Behavior – RAIS 2010

	No Controls (1)	Full Contols (2)
Non-reporting share = 0	-0.031	-0.012
(Always report)	(0.0006)	(0.0007)
Non-reporting share	-0.163	0.012
	(0.0031)	(0.0037)
N	3,000,009	3,000,009
R^2	0.0010	0.0709

NOTE-Heteroskedasticity-robust standard errors in parentheses. Estimated on a sample of workers from the 2010 RAIS observed to change primary employer during the year. The dependent variable is an indicator equal to 1 if the employer-reported race is different on the origin and destination job. Coefficient estimates reported for an indicator of whether the non-reporting share at the destination plant is equal to zero, along with the total non-reporting share. Column (1) includes no additional controls. Column (2) controls for gender, educational attainment, industry, state of employment, the share white, employment, separation rate, and average wage at the origin and destination plant.

Table VIII: Race History and Wages: Plant Reporting Behavior – RAIS 2010

	Benchmark (1)	Reporting Contols (2)	Always Report (3)	Not Always Report (4)	Plant Effects (5)
Race History					
'11': White/White	-0.003	-0.001	-0.002	0.009	0.001
	(0.0010)	(0.0010)	(0.0012)	(0.0031)	(0.001)
'10': White/Non-White	-0.021	-0.022	-0.021	-0.021	-0.010
	(0.0010)	(0.0010)	(0.0013)	(0.0035)	(0.001)
'01': Non-White/White	0.017	0.020	0.016	0.032	0.010
	(0.0010)	(0.0010)	(0.0013)	(0.0036)	(0.001)
Plant Effects	N	N	N	N	Y
N	$3,000,688$	$3,000,009$	$1,864,636$	$250,447$	$3,000,688$
R^2	0.195	0.1938	0.2111	0.1313	0.378

NOTE-Heteroskedasticity-robust standard errors in parentheses. Estimated on a sample of workers from the 2010 RAIS observed to change primary employer during the year. The dependent variable in all models is the change in log wage between origin and destination job. All models control for gender, educational attainment, industry, state of employment, the share white, employment, separation rate, and average wage at the origin and destination plant. The model in Column (2) adds controls for the share of workers for whom no race is reported at the origin and destination plant. Column (3) restricts the sample to workers who move between plants for which the share of workers with no reported race is zero. Column (4) restricts the sample to workers who move between plants for which the share of workers with no reported race is positive. Column (5) adds plant effects to the benchmark.

Table IX: Alternative Model Specification – RAIS 2010

	Dest. Wage
Race History	
'10': White/Non-White	-0.034
	(0.001)
'01': Non-White/White	0.022
	(0.001)
Log Wage (Origin Job)	0.307
	(0.001)
White (Origin Job)	0.043
	(0.001)
Plant Effects	Y
N	$3,000,688$
R^2	0.745

NOTE-Heteroskedasticity-robust standard errors in parentheses. Estimated on a sample of workers from the 2010 RAIS observed to change primary employer during the year. The dependent variable is the log wage on the destination job. The model includes plant effects along with controls for gender, educational attainment, and the industry, state of employment, share white, employment, separation rate, and average wage at the origin plant. See the text for details.

Table X: Race History and Wages: Type of Job Change – RAIS 2010

	Benchmark (1)	JUJ (2)	Education Same (3)	Education Down (4)
Race History				
'11': White/White	-0.003	-0.007	-0.002	-0.007
	(0.0010)	(0.0022)	(0.0013)	(0.0023)
'10': White/Non-White	-0.021	-0.021	-0.022	-0.019
	(0.0010)	(0.0024)	(0.0014)	(0.0024)
'01': Non-White/White	0.017	0.019	0.017	0.013
	(0.0010)	(0.0024)	(0.0014)	(0.0024)
N	$3,000,688$	$513,335$	$1,657,397$	$551,214$
R^2	0.195	0.254	0.179	0.229

NOTE-Heteroskedasticity-robust standard errors in parentheses. Estimated on a sample of workers from the 2010 RAIS observed to change primary employer during the year. The dependent variable in all models is the wage on the destination job. All models control for gender, educational attainment, industry, state of employment, and, where relevant, the share white, employment, separation rate, and average wage at the origin and destination plant.

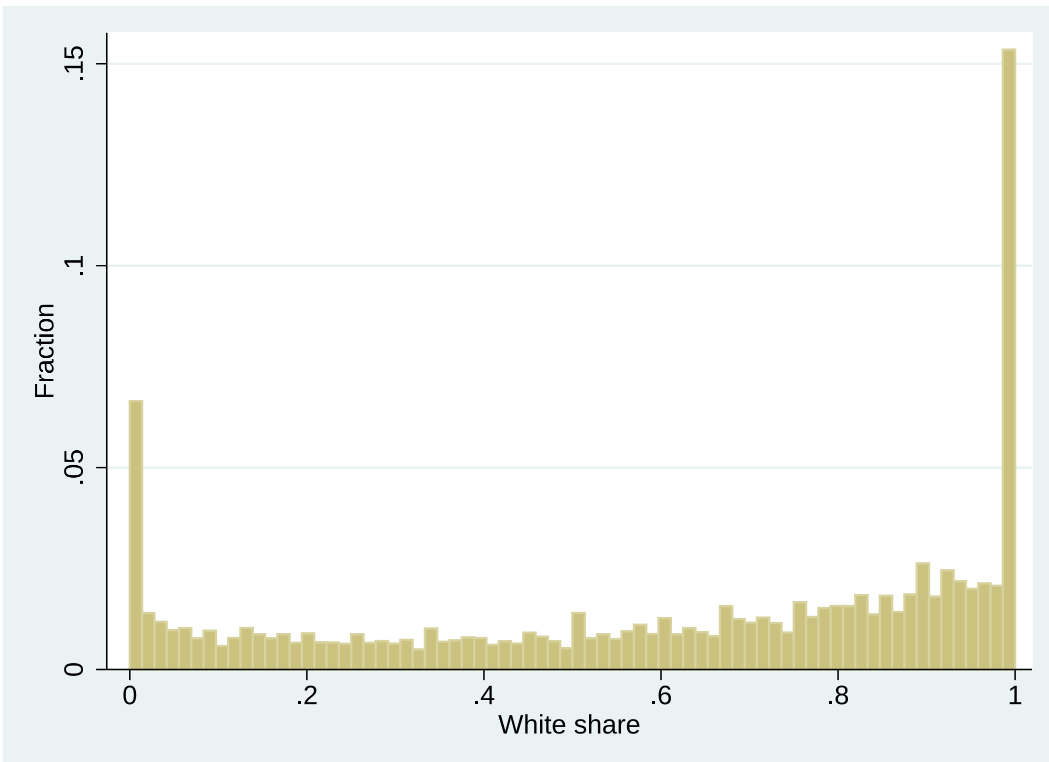

Figure 1: Share of White Workers, 2010 (Plant-level; Weighted by Plant Size).

A Appendix: Model Details

A.1 Reduced-Form Wage Equation

Substitution of (5) into the structural wage equations, (3) and (4) gives

$$\begin{aligned} w_{i1} =& a_1' + \{\beta_1 + \lambda + (\delta + \phi_{10})\gamma_{x10} + \phi_{01}\gamma_{x01} + (\delta + \phi_{11})\gamma_{x11}\}\, x_i \qquad (A.1)\\ &+ \{(\delta + \phi_{10})\gamma_{10} + \phi_{01}\gamma_{01} + (\delta + \phi_{11})\gamma_{11})\}\, R_i + e_{i1}\\ w_{i2} =& a_2' + \{\beta_2 + \lambda + \phi_{10}\gamma_{x10} + (\delta + \phi_{01})\gamma_{x01} + (\delta + \phi_{11})\gamma_{x11}\}\, x_i \qquad (A.2)\\ &+ \{\phi_{10}\gamma_{10} + (\delta + \phi_{01})\gamma_{01} + (\delta + \phi_{11})\gamma_{11})\}\, R_i + e_{i2} \end{aligned}$$

where the composite error terms are

$$e_{i1} = (\delta + \phi_{10})\eta_{10} + \phi_{01}\eta_{01} + (\delta + \phi_{11})\eta_{11} + \xi_i + \varepsilon_{i1}' \qquad (A.3)$$

$$e_{i2} = \phi_{10}\eta_{10} + (\delta + \phi_{01})\eta_{01} + (\delta + \phi_{11})\eta_{11} + \xi_i + \varepsilon_{i2}'. \qquad (A.4)$$

By construction, the composite errors are uncorrelated with the observables: x_i and R_i. Consistent estimates can therefore be obtained by OLS regression of observed wages in each period onto observables.

A.2 Derivation of Ω

To see this, the conditional expectation

$$\mathrm{E}(R_{ij}|x_i) = P(R_{ij} = 1|x_i) = \sum_h P(R_{ij} = 1|R_{ih}^*, x_i) \cdot P(R_{ih}^*|x_i) \qquad (A.5)$$

Our assumptions on the misclassification process give us $P(R_{ij} = 1|R_{ih}^*, x_i)$ in terms of τ. So the equation is

$$\mathrm{E}(R_{ij}|x_i) = \tau_{j|00}\left[1 - \sum_{h\neq 00}(\pi_h + (x_i - \bar{x})c_h)\right] + \sum_{h\neq 00}\tau_{j|h}\left[\pi_h + (x_i - \bar{x})c_h\right] \qquad (A.6)$$

$$= \sum_h \tau_{j|h}\pi_h + (x_i - \bar{x})\sum_{h\neq 00}(\tau_{j|h} - \tau_{j|00})c_h \qquad (A.7)$$

$$= T_j\pi + (x_i - \bar{x}) \cdot c \cdot \Omega_j^{\mathrm{T}}. \qquad (A.8)$$

This clearly implies that $\zeta_j = \Omega_j^{\mathrm{T}}$ where Ω_j is the jth row of Ω. It follows that $\zeta^{\mathrm{T}} = \Omega c^{\mathrm{T}}$. Note ζ and c are $K \times 3$ matrices of covariate parameters.

A.3 Derivation of γ_h

Let a tilde designate variables that have been transformed into mean deviations (e.g., $\tilde{y}_i = y_i - \bar{y}$), so that (13) and (14) become

$$\tilde{R}^*_{ih} = \tilde{x}_i c_h + \tilde{\nu}_{ih} \tag{A.9}$$
$$\tilde{R}_{ih} = \tilde{x}_i \zeta_h + \tilde{\nu}'_{ih}. \tag{A.10}$$

Applying the same transformation to (5) yields

$$\tilde{R^*}_{ih} = \tilde{R}_i \gamma_h + \tilde{x}_i \gamma_{xh} + \tilde{\eta}_{ih}. \tag{A.11}$$

The algebra of partitioned regression implies

$$\gamma_h = \left(\tilde{R} M_{\tilde{x}} \tilde{R}\right)^{-1} \tilde{R} M_{\tilde{x}} \tilde{R}^*_h, \tag{A.12}$$

where $M_{\tilde{x}} = I - P_{\tilde{x}}$ is the idempotent "residual maker" matrix that projects onto the column null space of $\tilde{x}$.

Then, using (A.9) and (A.10),

$$\gamma_h = \left[(\tilde{R} - \tilde{x}\zeta)^{\mathrm{T}}(\tilde{R} - \tilde{x}\zeta)\right]^{-1} (\tilde{R} - \tilde{x}\zeta)^{\mathrm{T}}(\tilde{R}^*_h - \tilde{x}c_h) \tag{A.13}$$
$$= \left[\tilde{R}^{\mathrm{T}}\tilde{R} - \Omega c^{\mathrm{T}}\tilde{x}^{\mathrm{T}}\tilde{x}c\Omega^{\mathrm{T}}\right]^{-1} \cdot (\tilde{R}^{\mathrm{T}}\tilde{R}^*_h - \Omega c^{\mathrm{T}}\tilde{x}^{\mathrm{T}}\tilde{x}c_h) \tag{A.14}$$
$$= \left[\mathrm{var}(R) - \Omega c^{\mathrm{T}} V_{xx} c\Omega^{\mathrm{T}}\right]^{-1} \cdot \left[\mathrm{cov}(R, R^*_h) - \Omega c^{\mathrm{T}} V_{xx} c_h\right], \tag{A.15}$$

where V_{xx} is the covariance matrix of x_i.

We can use these expressions to compute γ_h, as long as we have sufficient structure in Ω to recover c_h from our estimate of ζ. We also use the misclassification model to calculate $\mathrm{cov}(R, R^*_h)$:

$$\mathrm{cov}(R_j, R^*_k) = (\tau_{j,k} - p_j)\pi_k. \tag{A.16}$$

Therefore, γ_h is a function of observed data (V_{xx}, $\mathrm{var}\,(R)$, and p), prior information on misclassification probabilities (τ and Ω), and model parameters, π.

Table A.1: Summary of Structural Estimation: RAIS 2010

Panel A: Structural Parameter Estimates		
Parameter	Model	
	No Race Change (1)	No Meas. Error (2)
$\kappa = (\delta + \phi_{11})$	0.283	0.071
	(0.0030)	(0.0001)
δ	–	0.019
		$(2.7e^{-5})$
ϕ_{11}	–	0.052
		$(9.9e^{-5})$
ϕ_{10}	–	0.026
		$(7.6e^{-5})$
ϕ_{01}	–	0.015
		$(8.8e^{-5})$
q_1	0.884	−−
	(0.0002)	
q_0	0.236	−−
	(0.0002)	
π_{11}	0.583	0.481
	(0.0004)	(0.0003)
π_{10}	–	0.141
		(0.0002)
π_{01}	–	0.132
		(0.0002)
Panel B: Implied Bias Parameters		
$\gamma(R^*_{11} \| R_{11})$	0.244	1.000
Panel C: Model Fit		
Obj. Fcn Value	0.0005	$1.049e^{-5}$
Test Statistic	$1,588$	0.5313

NOTE–Standard errors in parentheses. Parameters are estimated by minimum distance, fitting the reduced-form coefficients for employer-reported race histories (Table III) and their associated population shares (Table I). Panel B reports the estimate of $\gamma(R^*_{11}|R_{11})$, which is the parameter on an indicator for observed race history '11' in a linear probability model for true race history '11'. Panel C reports the value of the distance function at the solution and test statistics for each of the restricted models. Under the null hypothesis that the parameter restrictions are valid, the test statistics in Panel C are distributed χ^2_d with degrees of freedom equal to the number of restrictions.

Table A.2: Cross-Section Wage Gap for Workers Who Change Employers: RAIS 2003–2010

	2010	2009	2008	2007	2006	2005	2004	2003
White	0.048	0.049	0.048	0.054	0.047	0.050	0.050	0.046
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
N	3,000,688	2,575,019	2,621,915	2,210,629	1,922,121	1,865,234	1,569,839	1,419,995
R^2	0.528	0.519	0.548	0.543	0.539	0.532	0.536	0.517

Heteroskedasticity-robust standard errors in parentheses. The dependent variable is the log wage on the job in which a worker is employed at the end of the year; his or her destination job. Each column reports the estimated coefficient on an indicator for whether a worker is reported 'white' by their employer. The models are estimated on the sample of workers who change employers during the indicated year. In addition to the White indicator, all models control for gender, education, a quadratic in age, and for the following plant characteristics: industry, state, employment, white share, average log wage, and separation rate.

Table A.3: Benchmark Specification First-Difference Models: RAIS 2003–2010

	2010	2009	2008	2007	2006	2005	2004	2003
Race History								
'11': White/White	−0.003	−0.004	−0.004	−0.003	−0.005	0.000	−0.011	−0.009
	(0.0009)	(0.0011)	(0.0011)	(0.0012)	(0.0013)	(0.0015)	(0.0017)	(0.0018)
'10': White/Non-White	−0.021	−0.025	−0.020	−0.023	−0.026	−0.018	−0.035	−0.025
	(0.0010)	(0.0011)	(0.0011)	(0.0013)	(0.0014)	(0.0016)	(0.0018)	(0.0020)
'01': Non-White/White	0.017	0.018	0.018	0.016	0.012	0.022	0.017	0.008
	(0.0010)	(0.0011)	(0.0011)	(0.0013)	(0.0014)	(0.0016)	(0.0018)	(0.0020)
N	3,000,688	2,575,019	2,621,915	2,210,629	1,922,121	1,865,234	1,569,839	1,419,995
R^2	0.1948	0.2160	0.2024	0.2077	0.2232	0.2304	0.2542	0.2414

NOTE-Heteroskedasticity-robust standard errors in parentheses. Estimated on a sample of workers in each year, 2003–2010, from RAIS, observed to change primary employer during the year. The dependent variable is the difference between the log wage on the destination job and log wage on the origin job. The model includes plant effects along with controls for gender, educational attainment, and the industry, state of employment, share white, employment, separation rate, and average wage at the origin and destination plant.