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

Christopher Cornwell, Jason Rivera, and Ian M. Schmutte

Abstract

In Brazil, different employers report different racial classifications for the same worker. We use the variation in race across employers to estimate the relationship between race and wages. 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 racial wage gap remains after controlling for all individual characteristics that do not change across jobs. We formally test, and reject, the hypothesis that our results are driven by misclassification. We also evaluate several mechanisms that could explain our findings.

Christopher Cornwell is a professor of economics at the University of Georgia. Jason Rivera is Senior Manager and Econometrician at Inductis/EXL Service. Ian M. Schmutte is an assistant professor of economics at the University of Georgia. The authors are thankful for helpful comments from David Blau, Peter Brummond, Amanda Griffith, Barry Hirsch, Josh Kinsler, Daniel Kreisman, Kurt Lavetti, Kyle Mangum, Stephen L. Ross, Meghan Skira, Mike Strain, Ron Warren, and Doug Webber, along with audience members at the Southern Economic Association 2014 Meetings, LAMES 2014, the Econometric Society North American Winter Meetings 2014, Georgia State University, Georgia Tech, Temple University, University of South Carolina, and Wake Forest. Caio Canton graciously answered our questions about the hiring and data collection practices of Brazilian firms. The source data for this paper is restricted. The data are available by application to Brazil's Ministério do Trabalho e Emprego (MTE). The authors

will provide information upon request about the source files and the information required to go from raw data to our results as part of a replication archive.

I. Introduction

A longstanding question in studies of racial differences in wages is whether they can be explained by differences in skill, background, and preferences. The problem is that race is usually measured, and modeled, as an immutable characteristic of the individual. This restriction reflects the origins of the literature on labor market discrimination in the U.S., where racial identity is thought of as a fixed, inherited, trait. As a result, it has been impossible to separate the effect of race on wages from the effects of other characteristics that do not change over time. We address this question by exploiting a unique feature matched employer-employee data from Brazil's Relação Anual de Informações Sociais, or Annual Social Information Survey (RAIS). In RAIS, employers record the race of each new employee in an administrative register, along with other characteristics. It is not uncommon for different employers to assign a different racial classification to the same worker. A worker's measured race therefore can, and frequently does, change with each new job.

We use the observed changes in employer-reported race to control for the influence of fixed unobservables on wages. To be valid, the variation we observe in employers' reports of race must arise out of behavior associated with labor market outcomes. This is especially plausible in Brazil. Unlike the U.S., in Brazil, racial identity is at least as closely associated with skin color as with heredity. Consequently, different observers may have different perceptions of the race of the same individual. Racial classification is therefore subject to change over time and across contexts. Brazil also has a well-documented history of discriminatory treatment of non-white workers in hiring and promotion, as well as large racial disparities in labor market outcomes favoring whites. Thus, uniquely, the RAIS data are collected in a context where differences in the perception of a worker's race are quite plausible, and likely to be associated with wages.

We find that 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. We find that as much as 40 percent of the raw racial wage gap remains after controlling for all individual characteristics, observed and unobserved, that do not change across jobs.

While changes in reported race are especially plausible in the Brazilian context, it may nevertheless be that most of the changes observed in RAIS arise from measurement error. To explore this possibility, we specify and estimate a structural misclassification model. In the model, wages vary based on what we call 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 differences in wages paid to workers of different races. We formally reject the restrictions imposed by assuming changes in race arise from pure misclassification.

This paper is novel in using panel data methods to control for unobserved characteristics when estimating the racial wage gap rather than using performance on standardized tests to proxy for unobserved ability, as in Neal and Johnson (1996). In doing so, we draw on a key insight of the literature using field experiments to assess hiring discrimination (Rouse and Goldin 2000; Bertrand and Mullainathan 2004). The notion that race has an estimable effect on labor market outcomes is conceptually problematic if race is immutable. Field experiments work by manipulating something else: the employer's perception of an applicant's race. Our data offer a similar source of observed variation – employer-reported race – while implicitly addressing the drawback that correspondence studies measure the racial bias of the average, not the marginal, employer (Heckman 1998; Neumark 2012). Some recent research also exploits variation in the perception of racial identity to provide evidence of racial disparities in rules enforcement in professional sports (Price and Wolfers 2010; Parsons et al. 2011).

Our study also relates to the growing literature on the economics of identity. Akerlof and Kranton (2000) argue that the choice of identity is one of the most consequential economic decisions a person can make. Evidence that identity may be affected by economic incentives has emerged only recently. Antman and Duncan (2015) show for the U.S. that self-reported racial identity responds to changes in affirmative-action policies. Cassan (2015) presents evidence of identity manipulation in India in response to caste-based land policies. To date, very little empirical work has been done to understand how racial identity is related to labor market outcomes. An exception is research on the mechanisms of immigrant assimilation. Biavaschi, Giulietti and Siddique (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. In RAIS, 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.

Our results clearly show the persistence of a racial wage gap, conditional on unobserved worker characteristics. However, we cannot fully rule out the possibility that race change is endogenous to changes in the wage, and this necessarily limits how our results may be interpreted. Because we do not fully observe the job search process, our data are not sufficient to identify the precise mechanism linking wages and employer-reported race. Nevertheless, we can examine three highly plausible candidates. One concern is that our results are driven by plant-specific reporting behavior – for example, some plants simply classify all workers as white or non-white when race information is missing. A second possibility is that employers code workers' races differently based on how they move between jobs. For example, our results could arise if employers are more likely to report workers as non-white when they are hired from non-employment. Finally, our results may reflect reverse causality – that is, racial classification is a function of wages, rather than the other way around. This could arise either because employers, or workers, are more disposed to classify as white when starting higher-status jobs. We show that none of these candidate mechanisms can fully explain our findings.

II. 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).

II.A. 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 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 misperception and manipulation is not academic speculation. Telles (2002) finds survey enumerators and respondents disagree on racial classification in approximately 20 percent of cases. These disagreements cut both in the direction of "lightening" and "darkening", and are systematically associated with socio-economic status. Enumerators are more likely to perceive highly-educated and wealthier individuals as white when they self-report as 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] In a series of papers, Francis and Tannuri-Pianto (2012; 2013); Francis-Tan and Tannuri-Pianto (forthcoming) 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, workers may be willing and able to manipulate their race, as perceived by employers, to obtain better employment outcomes.

II.B. Racial Inequality and Discrimination in Brazil's Labor Market

While the Brazilian notion of race provides scope for the manipulation of racial identity, it does not constitute a motivation. 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.

II.B.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 reflected in other indicators such as development, literacy, and total wealth. Our own calculations, using PNAD data, 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 V, 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).[3]

II.B.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.[4]

II.B.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 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 in Rio de Janeiro identified work as the site of greatest racial tension. Furthermore, a majority of non-white respondents described experiences of discrimination in hiring and promotion, and that non-white workers struggle with advancement because of difficulty commanding the respect of their white subordinates.

II.C. 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 classification 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. This is also the case in our data. In unreported results, available from the authors, we find little evidence of racial disparities between workers classified as brown and black, nor across jobs for workers whose reported race changes from brown to black, or vice versa.

III. Data on Race and Job Mobility

We use data from the Relação Anual de Informações Sociais (RAIS), a matched employer-employee database. 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 managers who complete the survey on behalf of the employees. In smaller enterprises the data may be completed by 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 preceding calendar year. Completion of RAIS is mandatory, and 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.[5]

III.A. How Employers Collect and Report Information on Employee Race

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.[6] 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.[7]

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.[8] 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 employee, and subject to verification by the staff member responsible for hiring procedures.[9] 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 interpretation of that information may play a role.

III.B. 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 II.C, we restrict our analysis to jobs on which the race of the worker is reported as either white, brown, or black.[10]

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 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.[11] The remaining cases with missing race amount to approximately 5 percent of the sample and are evenly distributed across sectors, occupations, and basic demographic characteristics. We consider the implications of non-reporting in Section VI.B.

III.C. Descriptive Statistics

Table 1 reports sample averages of worker characteristics as reported by both the 'origin' 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 (3-6) 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 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.

III.C.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, they are also employed in plants with more employment volatility. These features of the data underscore the importance of controlling for plant-level characteristics in our earnings models.[12]

For interpretation, it is instructive to understand how we construct our sample of continuing workers from the raw source data. In Brazil, roughly 60 percent of employment in 2010 was in the formal sector, all of which is recorded in RAIS. The raw data include roughly 67 million observations, each of which corresponds to a single employment record (job-year). Of these, 57 million are full-time with valid person and plant identifiers and wage information. Restricting to continuing jobs with complete information for all covariates, and eliminating records associated with workers with many (more than 3) jobs during the year cuts the sample to 27 million. Our analysis must therefore be understood as representative of workers in stable, full-time, formal sector employment.

III.C.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.[13] 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.[14] 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 1 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.[15] 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 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 necessary for 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.

III.C.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 male, 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.

IV. 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 relationship between race and 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 for determining wages 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.[16]

IV.A. 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 1.

We begin by expressing wages as

$$(1) \qquad \ln \omega_{it} = a_t + b_t x_i + \delta r_{it}^* + \varepsilon_{it},$$

where $\ln \omega_{it}$ 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 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 *h*th possible race history, $h \in \{00, 01, 10, 11\}$. We assume that R_{ih}^* is strictly exogenous with respect to ε'_{it}, so that $E(R_{ih}^* \varepsilon'_{it}) = 0$ 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 unobserved market 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:

$$(2) \qquad \alpha_i = \phi_1 + \textstyle\sum_{h \neq 00} R_{ih}^* \phi_h + \lambda \mathrm{x}_i + \xi_i,$$

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

$$(3) \qquad \ln \omega_{i1} = a_1 + \phi_1 + (\beta_1 + \lambda)x_i + (\delta + \phi_{10})R_{i10}^* + \phi_{01}R_{i01}^* + (\phi_{11} + \delta)R_{i11}^* + \xi_i + \varepsilon'_{i1}$$

$$(4) \qquad \ln \omega_{i2} = a_2 + \phi_1 + (\beta_2 + \lambda)x_i + \phi_{10}R_{i10}^* + (\phi_{01} + \delta)R_{i01}^* + (\phi_{11} + \delta)R_{i11}^* + \xi_i + \varepsilon'_{i2}$$

The employer's report of race, r_{it}^{M}, 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 each possible race history R_{ih}^{*}, $h \in \{01, 10, 11\}$, onto R_i and x_i:

$$R_{ih}^{*} = \gamma_{0h} + \gamma_h R_i + \gamma_{xh} x_i + \eta_{ih}. \quad (5)$$

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:

$$\ln \omega_{i1} = a_1' + b_1 x_i + d_1 R_i + e_{i1} \quad (6)$$

$$\ln \omega_{i2} = a_2' + b_2 x_i + d_2 R_i + e_{i2} \quad (7)$$

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

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

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

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} (or R_{01}) between the wage equation for the first job and the wage equation for the second job. 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

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

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.

IV.B. 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,

$$P(r_{i1}, r_{i2} | r_{i1}^*, r_{i2}^*, \ x_i) = P(r_{i1} | r_{i1}^*). P(r_{i2} | r_{i2}^*). \quad (11)$$

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:

$$p = E(R_i) = E(R_i^* T) = \pi T. \quad (12)$$

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 $\zeta^{\mathrm{T}} = \Omega c^T$ 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

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

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

IV.C. 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, $(q_1, q_0, \pi_{11}, \pi_{10}, \pi_{01}, \phi_{11}, \phi_{10}, \phi_{01}, \delta)$. 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: $\pi_{10} = \pi_{01} = 0$.[17] 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, $\kappa \equiv (\delta + \phi_{11})$.[18] 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.

V. 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 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 report the reduced-form results for all years in Appendix 1.

V.A. Cross-Section Wage Gap

Table 2 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. Table A.1 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.

V.B. Reduced-Form Model for Wages

Table 3 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 destination 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 the effect of race on wages from differences in fixed unobservable worker-specific characteristics. If the observed race histories really correspond to differences in compensation – if there is no misclassification of race – 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 (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, while statistically different from zero, is an order of magnitude smaller at –.3 percent.[19] Table A.2 reports estimates of the benchmark specification from Column (3) for each year from 2003–2010, and show our findings are largely invariant to the sample year.

V.C. Tests of the Misclassification Model

The misclassification model predicts the estimated wage effect associated with race histories ‘10’ and ‘01’ in Column (3) of Table A.2 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 model in which race change is driven by misclassification that is random with respect to wages.[20] We now formally test the implications of these alternative models of the data-generating process.

Table 4 reports estimates of the misclassification model, along with tests of the restrictions discussed in Section IV. 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. Column (3), for completeness, reports the unrestricted model. Each model is fit to the reduced-form parameter estimates from Table 3 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 reported 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, $\kappa \equiv (\delta + \phi_{11})$. As discussed in Section IV, we test these restrictions using the statistic, $N \times (Q_r - Q_{nr})$. In this case, value of the test statistic is 1,588, 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 3. 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 Table 2.

VI. Possible Mechanisms and Alternative Specifications

The preceding results establish our key claim that workers are paid more when their employers report them to be white. We now consider possible mechanisms to explain why employer-reported race and wages change together in the way they do. Before doing so, we present reduced-form evidence on the correlates of race change. We find that a change in race is most strongly predicted by: (1) the share of white co-workers at the plant, and (2) the average log wage of the plant. Any explanatory mechanism must account for these facts.

Our results on wage determination in Table 3, along with the segregation exhibited in Figure 1, are consistent with the presence of employer discrimination. One possibility is that, during the hiring process, workers manipulate the way their race is perceived to obtain employment at better terms in a discriminatory labor market. As we have already discussed in Section II, such behavior is plausible in the Brazilian context (Francis and Tannuri-Pianto 2013).

Another possibility is that certain plants have a tendency to report all workers as white. Alternatively, our results could be driven by endogenous mobility, or by reverse causality – a tendency for employers to report workers as white when they are hired into higher-paying jobs. Here, we present evidence that these alternative mechanisms cannot fully explain our results.

VI.A. The Correlates of Race Change

Tables V and VI report estimates from the reduced-form linear probability models for the observed race histories, R_{11}, R_{10}, and R_{01} (Table 6 is a continuation of Table 5). These 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 5 and 6, all models include controls for industry and state of the origin and destination plant.

We will focus on several features of Tables 5 and 6. 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. 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 coworkers 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 between 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.

VI.B. Plant-Specific Reporting Behavior

Tables 1 and 6 show that race change is strongly associated with the share white in the plant. This is consistent with workers manipulating reported race to help obtain 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 III, 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. First we estimate the effect of plant-level non-reporting on race change (in either direction). 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 between plants that consistently report race. Table 7 presents the results from two linear probability models. The first captures the simple link between race change and the share of the origin and destination plant’s workers without a reported race (Column (1). The second adds the complete set of controls from Table 5 (Column 2). If anything, race change is between 1 and 3 percentage points less likely in moves across 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 8 provides the results of this exercise, carrying over the benchmark specification in first-differences from Table 3. 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 explained by plant-level reporting behavior that is correlated with missing race information, there may be other unobserved plant-level reporting practices, or plant heterogeneity more broadly, that may confound our analysis. To address this issue, we re-estimate the benchmark wage model incorporating plant effects for the destination

(Column 5) and origin plant (Column 6).[21] The estimated effect of race history '01' drops to 0.012 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 (or from) the same plant. In these specifications, which we regard as conservative, race change still accounts for 20 percent of the baseline cross-section wage gap. Our interpretation that the estimates in Columns (5) and (6) are conservative is supported by our analysis of an alternative identification strategy, which we consider next.

VI.C. Alternative Source of Variation

The reduced-form wage model restricts how individual heterogeneity enters the model. Table 9 reports results from estimating an alternative specification that controls for individual heterogeneity in the destination wage by directly controlling for the wage on the origin job:

$$w_{i2} = a + \zeta w_{i1} + bx_i + m \times OrigWhite_i + \theta_{10}R_{10} + \theta_{01}R_{01} + \psi_{J(i2)} + e_{i2}. \tag{16}$$

This specification relaxes the implied restriction of reduced-form model that $\zeta = 1$.[22] 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 $\psi_{J(i2)}$ (where $J(i2) = j$ indicates the plant j that 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'. 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.

VI.D. 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. An important empirical issue in such studies is that the decision to change jobs is typically based on new information about the current match, the new match, or both. Thus, 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.

In particular, one might be concerned that an employer may be more likely to report a worker as white under direct job-to-job moves and non-white when hired from non-employment. Table 10 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 3. Column (2) restricts the sample to workers who leave jobs in plants where there was a mass displacement event, defined as a plant that reduced its workforce by 30 percent or more over the course of the year. Column (3) limits the sample to workers who are not employed for three to four months between jobs, while Column (4) confines the analysis to workers who do not experience a spell of non-employment between jobs, and so can be regarded as having made direct job-to-job moves. The main results are remarkably consistent across these different groups of workers. We conclude that our findings are not driven by the manner in which workers move from one job to the next.[23]

VI.E. Reverse Causality: Does "Money Whiten"?

Another possibility is 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 respondents identify themselves as brown. Schwartzman (2007) finds parents of higher socio-economic status are more likely to classify their children as white.

Table 5 shows education is, if anything, negatively correlated with race change. However, if race change is associated with moves to higher-status jobs, we would expect the effect to be driven, in part, by changes in required education or occupation. Table 11 shows this is not likely to explain our results. Recall from Section III.C.2 that reported education falls for 18 percent of our sample as the move from one job to the next, which most likely signals a decrease in the level of education required for the job. Column (2) restricts the sample to workers whose education does not change, while Column (3) limits the sample to workers whose employer-reported education falls. Column (4) adds controls for the worker's occupation at the origin and destination jobs.

Confining the analysis to workers whose education does not change has essentially no effect on our results. Column (3), which focuses on workers moving into jobs where the reported level of education is lower, shows a modest attenuation of the race-change effect. Finally, Column (4), which controls for the occupation on the origin and destination, shows our results are not associated with occupational upgrading or downgrading, as would be expected if the process of reporting race were responsive to job status.

VII. Conclusion

We have shown that in Brazil's RAIS data, when a person changes jobs, their new employer may report a different race than their previous employer. This observed variation in race is systematically associated with variation in wages. We are able to show that the premium earned by workers reported to be white cannot be fully explained by fixed unobservable characteristics. Our analysis would typically be impossible since most socio-economic data measure race as an immutable individual characteristic.

Our results invite a more thorough investigation of the mechanisms governing the relationship between employer-reported race and wages. In addition to our central result, we find that workers whose employer-reported race changes from non-white to white are typically moving across highly segregated plants, and that they are moving from plants with higher average pay to plants to lower average pay. These findings could be explained by 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.

However, our data are not capable of clearly distinguishing the mechanism that drives changes in employer-reported race. Our results thus call for further theoretical research to jointly model the processes by which employers report race and wages. Extending the model in Lang, Manove and Dickens (2005) to incorporate race change is one promising approach. Their framework, along with and extensions described in Lang and Lehmann (2012) will be useful to address the facts outlined in our study. Any such model must also explain why some workers move into jobs where they are reported as non-white. There are many possibilities. Perhaps it is easier to find jobs at the non-white wage, or perhaps some workers prefer working with non-white workers. Alternately, in a job search framework, workers may sacrifice a wage premium associated with being reported white when they get an offer from a higher-paying employer in a non-white firm.

Regardless of the underlying mechanism, the results of this research are relevant beyond the Brazilian context. The difficulty of racial classification has a long history in the U.S. Using data from the 1880-1940 Decennial Censuses, Nix and Qian (2015) document that almost 20 percent of black men "passed" for white at some point, and that passing was associated with better social and economic outcomes. In the U.S., laws prohibiting interracial marriage were not fully eliminated until 1967, and the rate of interracial marriage 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, 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.

Our research highlights a connection between measured racial identity and economic outcomes that is echoed in recent demographic research in the U.S. The increasing difficulty of measuring race is the subject of a recent book by the former director of the Census Bureau

(Prewitt 2013). The Census has changed its procedure for collecting information on race in surveys and censuses 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, Rastogi, Fernandez, Noon, and Ennis (2014) document large changes in 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, and echo the similar churning we observe among Brazilians in employer-reported race. Saperstein and Penner (2012) also document changes over time in self-reported race in the 1979 National Longitudinal Study of Youth that are correlated with changes in socio-economic status. 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 the way we model, measure, and empirically assess the interaction between race and labor market outcomes.

References

Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. "High wage workers and high wage firms." *Econometrica* 67(2):251–333.

Abowd, John M., and Martha H. Stinson. 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, George A., and Rachel E. Kranton. 2000. "Economics and identity." *The Quarterly Journal of Economics* 115(3):715–753.

Antman, Francisca, and Brian Duncan. 2015. "Incentives to identify: Racial identity in the age of affirmative action." *The Review of Economics and Statistics* 97(3):710–713.

Bertrand, Marianne, and Sendhil Mullainathan. 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, Costanza, Corrado Giulietti, and Zahra Siddique. 2013. "The economic payoff of name Americanization." *IZA Discussion Papers 7725*, Institute for the Study of Labor (IZA).

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

Cassan, Guilhem. 2015. "Identity-based policies and identity manipulation: Evidence from colonial Punjab." *American Economic Journal: Economic Policy* 7(4):103–31.

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

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

Fiola, Jan. 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, Andrew M., and Maria Tannuri-Pianto. 2013. “Endogenous race in Brazil: Affirmative action and the construction of racial identity among young adults.” *Economic Development and Cultural Change* pp. 1–35.

Francis-Tan, Andrew, and Maria Tannuri-Pianto. (2015). “Inside the black box: affirmative action and the social construction of race in Brazil.” *Ethnic and Racial Studies*, 38(15):2771-2790…

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

Goldsmith, Arthur H., Darrick Hamilton, and William Darity Jr. 2006. “Shades of discrimination: Skin tone and wages.” *The American Economic Review*, 96(2):242–245.

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

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

Kreisman, Daniel, and Marcos A. Rangel. 2015. “Blurring the color line: Wages and employment for black males of different skin tones.” *Review of Economics and Statistics* 97(1):1–13.

Krishna, Pravin, Jennifer P. Poole, and Mine Zeynep Senses. 2014. “Wage effects of trade reform with endogenous worker mobility.” *Journal of International Economics* 93(2):239 – 252.

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

Lang, Kevin, Michael Manove, and William T. Dickens. 2005. “Racial discrimination in labor markets with posted wage offers.” *The American Economic Review* 95(4):1327–1340.

Liebler, Carolyn A., Sonya Rastogi, Leticia E. Fernandez, James M. Noon, and Sharon R. Ennis. 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, Rafael. 2013. “Firm wage differentials and labor market sorting: Reconciling theory and evidence.” Working Paper.

Menezes-Filho, Naércio A., Marc-Andreas Muendler, and Garey Ramey. 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, Marc-Andreas, James E. Rauch, and Oana Tocoian, O. 2012. “Employee spinoffs and other entrants: Stylized facts from Brazil.” *International Journal of Industrial Organization* 30(September):447-458.

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

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

Nix, Emily, and Nancy Qian. 2015. “The fluidity of race: Passing in the United States, 1880-1940.” *Working Paper 20828*, National Bureau of Economic Research.

Parsons, Christopher A., Johan Sulaeman, Michael C. Yates, and Daniel S. Hamermesh. 2011. “Strike three: Discrimination, incentives, and evaluation.” *American Economic Review* 101(June):1410–1435.

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

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

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

Racusen, S. 2009. “Affirmative action and identity.” In *Brazil’s New Racial Politics*., Lynne Rienner Publishers, pp. 89–122.

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

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

Schmutte, Ian M. 2015. “Job referral networks and the determination of earnings in local labor markets.” *Journal of Labor Economics* 33(1):1-32.

Schwarcz, Lilia Moritz. 2003. “Not black, not white: Just the opposite: Culture, race and national identity in Brazil.” *Centre for Brazilian Studies*, University of Oxford, Working Paper Number CBS-47-03.

Schwartzman, Luisa Farah. 2007. “Does money whiten? Intergenerational changes in racial classification in Brazil.” *American Sociological Review* 72(6):940–963.

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

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

Table 1
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)	'00' (6)
Race History						
'11': White/White	n/a	0.485	1	0	0	0
'10': White/Non-White	n/a	0.139	0	1	0	0
'01': Non-White/White	n/a	0.132	0	0	1	0
'00': Non-White/Non-White	n/a	0.244	0	0	0	1
White						
Orig. Job	0.644	0.624	1	1	0	0
Dest. Job	n/a	0.618	1	0	1	0
Log Wage						
Orig. Job	6.536	6.405	6.462	6.39	6.376	6.315
Dest. Job	n/a	6.460	6.517	6.452	6.431	6.368
Male						
Orig. Job	0.649	0.717	0.658	0.745	0.742	0.802
Dest. Job	n/a	0.717	0.659	0.745	0.744	0.802
Age						
Orig. Job	35.01	31.4	31.1	31.4	31.3	32.3
Dest. Job	n/a	31.4	31.1	31.4	31.2	32.1
Education						
LTHS	0.446	0.461	0.409	0.461	0.477	0.557
High School	0.421	0.436	0.451	0.451	0.443	0.393
Some College	0.041	0.040	0.052	0.035	0.033	0.023
Bachelor's (+)	0.092	0.063	0.088	0.053	0.047	0.027
Plant Mean Log Wage						
Orig. Job	6.528	6.460	6.503	6.446	6.449	6.389
Dest. Job	n/a	6.511	6.556	6.510	6.493	6.431
Plant White Share						
Orig. Job	0.626	0.615	0.822	0.749	0.363	0.268
Dest. Job	n/a	0.613	0.816	0.374	0.750	0.279
Plant Employment						
Orig. Job	755.4	662.5	551.5	549.6	703.1	921.9
Dest. Job	n/a	757.6	654.2	808.2	621.0	1004.4
Plant Separation Rate						
Orig. Job	0.633	1.150	1.139	1.197	1.121	1.161
Dest. Job	n/a	1.466	1.504	1.360	1.729	1.309
Num.Obs.	26, 512, 018	3, 000, 688	1, 443, 893	420, 759	397, 030	739, 006

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

Table 2
Cross-Section Racial Wage Gap Estimates: RAIS 2010

	All Workers		Job Changers	
			Orig. Wage	Dest. Wage
	(1)	(2)	(3)	(4)
White	0.132	0.078	0.065	0.048
	(0.000)	(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. 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 3
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. Equations (6) and (7), as 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 origin job. The dependent variable in column (2) is the log wage on the worker's destination job. The models estimated are the reduced-form. 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. 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 4
Summary of Structural Estimation: RAIS 2010

Parameter	No Race Change (1)	No Meas. Error (2)	Unrestricted (3)
Panel A: Structural Parameter Estimates			
$\kappa = (\delta + \phi_{11})$	0.283	0.071	––
	(0.0030)	(0.0001)	
δ	–	0.019	0.019
		($2.7e^{-5}$)	(0.0014)
ϕ_{11}	–	0.052	0.051
		($9.9e^{-5}$)	(0.0016)
ϕ_{10}	–	0.026	0.026
		($7.6e^{-5}$)	(0.0050)
ϕ_{01}	–	0.015	0.015
		($8.8e^{-5}$)	(0.0045)
q_1	0.884	––	1.000
	(0.0002)		(0.0140)
q_0	0.236	––	0.000
	(0.0002)		(0.0481)
π_{11}	0.583	0.481	0.481
	(0.0004)	(0.0003)	(0.0265)
π_{10}	–	0.141	0.140
		(0.0002)	(0.0005)
π_{01}	–	0.132	0.132
		(0.0002)	(0.0005)
Panel B: Implied Bias Parameters			
$\gamma(R^*_{11} \mid R_{11})$	0.244	1.000	1.000
Panel C: Model Fit			
Obj. Function 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 3) and their associated population shares (Table 1). Panel B reports the estimate of $\gamma(R^*_{11} \mid R_{11})$, 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 5
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 an observed race history. 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 6, and controls for the industry and state of the origin and destination plant. Heteroskedasticity-robust standard errors in parentheses.

Table 6
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 5, and controls for the industry and state of the origin and destination plant. Heteroskedasticity-robust standard errors in parentheses.

Table 7
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	0.001	0.071

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. 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 8
Race History and Wages: Plant Reporting Behavior – RAIS 2010

	Benchmark	Reporting Contols	Always Report	Not Always Report	Orig. Plant Effects	Dest. Plant Effects
	(1)	(2)	(3)	(4)	(5)	(6)
Race History						
‘11’: White/White	−0.003	−0.001	−0.002	0.009	−0.000	−0.001
	(0.0010)	(0.0010)	(0.0012)	(0.0031)	(0.0010)	(0.0010)
‘10’: White/Non-White	−0.021	−0.022	−0.021	−0.021	−0.009	−0.021
	(0.0010)	(0.0010)	(0.0013)	(0.0035)	(0.0010)	(0.0010)
‘01’: Non-White/White	0.017	0.020	0.016	0.032	0.012	0.012
	(0.0010)	(0.0010)	(0.0013)	(0.0036)	(0.0010)	(0.0010)
Orig. Plant Effects	N	N	N	N	Y	N
Dest. Plant Effects	N	N	N	N	N	Y
N	3,000,688	3,000,009	1,864,636	250,447	3,000,688	3,000,688
R^2	0.195	0.194	0.211	0.131	0.425	0.376

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 origin plant effects to the benchmark. Column (6) adds destination plant effects to the benchmark.

Table 9
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.

Table 10
Race History and Wages: Type of Job Change – RAIS 2010

	Benchmark	Mass Disp.	JUJ	J2J
	(1)	(2)	(3)	(4)
Race History				
‘11’: White/White	−0.003	0.005	−0.003	−0.000
	(0.0010)	(0.0037)	(0.0030)	(0.0016)
‘10’: White/Non-White	−0.021	−0.017	−0.018	−0.020
	(0.0010)	(0.0039)	(0.0032)	(0.0017)
‘01’: Non-White/White	0.017	0.025	0.018	0.019
	(0.0010)	(0.0038)	(0.0032)	(0.0017)
N	3,000,688	228,372	284,694	870,870
R^2	0.195	0.198	0.219	0.174

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 wages between the origin and 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. Column (2) restricts the sample to workers who separated from a plant that experienced a mass displacement, defined as a within-year reduction in total employment of more than 30 percent of the initial workforce. Column (3) restricts the sample to workers who were not observed to be employed for three to four months between the origin and destination job. Column (4) restricts the sample to workers who did not have a spell of non-employment between the origin and destination job.

Table 11
Race History and Wages: Type of Job Change – RAIS 2010

	Benchmark (1)	Education Same (2)	Education Down (3)	Occup. Controls (4)
Race History				
‘11’: White/White	−0.003	−0.002	−0.007	−0.004
	(0.0010)	(0.0013)	(0.0023)	(0.0010)
‘10’: White/Non-White	−0.021	−0.022	−0.019	−0.021
	(0.0010)	(0.0014)	(0.0024)	(0.0010)
‘01’: Non-White/White	0.017	0.017	0.013	0.017
	(0.0010)	(0.0014)	(0.0024)	(0.0010)
N	3,000,688	1,657,397	551,214	3,000,688
R^2	0.195	0.179	0.229	0.199

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 wages between the origin and 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.

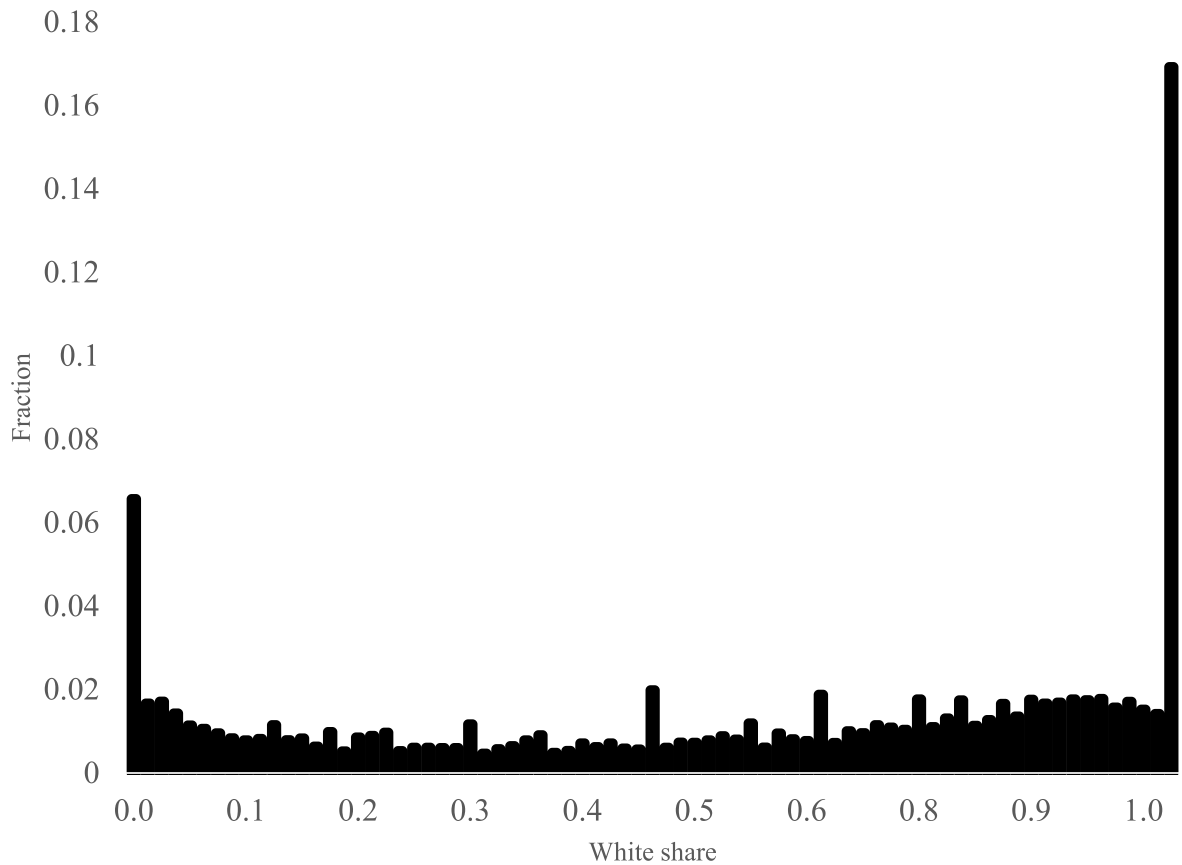

Figure 1: Share of White Workers, 2010 (Plant-level; Weighted by Plant Size).

Appendix 1: Model Details

App1.1. Reduced-Form Wage Equation

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

(A.1)
$$\omega_{i1} = a_1' + \{\beta_1 + \lambda + (\delta + \phi_{10})\gamma_{x10} + \phi_{01}\gamma_{x01} + (\delta + \phi_{11})\gamma_{x11}\}x_i \\ +\{(\delta + \phi_{10})\gamma_{10} + \phi_{01}\gamma_{01} + (\delta + \phi_{11})\gamma_{11})\}R_i + e_{i1}$$

(A.2)
$$\omega_{i2} = a_2' + \{\beta_2 + \lambda + \phi_{10}\gamma_{x10} + (\delta + \phi_{01})\gamma_{x01} + (\delta + \phi_{11})\gamma_{x11} +\}x_i \\ +\{\phi_{10}\gamma_{10} + (\delta + \phi_{01})\gamma_{01} + (\delta + \phi_{11})\gamma_{11}\}R_i + e_{i2}$$

where the composite error terms are

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

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

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.

App1.2. Derivation of Ω

To see this, the conditional expectation

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

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

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

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

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

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

App1.3. Derivation of γ_h

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

(A.9) $$\tilde{R}_{ih}^* = \tilde{x}_i c_h + \tilde{v}_{ih}$$

(A.10) $$\tilde{R}_{ih} = \tilde{x}_i \zeta_h + \tilde{v}_{ih}'$$

Applying the same transformation to (5) yields

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

The algebra of partitioned regression implies

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

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),

(A.13) $$\gamma_h = [(\tilde{R} - \tilde{x}\zeta)^{\mathrm{T}}(\tilde{R} - \tilde{x}\zeta)]^{-1}(\tilde{R} - \tilde{x}\zeta)^{\mathrm{T}}(\tilde{R}_h^* - \tilde{x} c_h)$$

(A.14) $$[\tilde{R}^{\mathrm{T}}\tilde{R} - \Omega c^{\mathrm{T}}\tilde{x}c\Omega^{\mathrm{T}}]^{-1}.\,(\tilde{R}^{\mathrm{T}}\tilde{R}_h^* - \Omega c^{\mathrm{T}}\tilde{x}^{\mathrm{T}}\tilde{x}c_h)$$

(A.15) $$[\mathrm{var}(R) - \Omega c^{\mathrm{T}}V_{xx}c\Omega^{\mathrm{T}}]^{-1}.\left[\mathrm{cov}\left(R, \tilde{R}_h^*\right) - \Omega c^{\mathrm{T}}V_{xx}c_h\right]$$

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 $\mathrm{cov}\left(R, \tilde{R}_h^*\right)$:

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

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: 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.2: 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.

Table A.3: Race History and Wages: Different Specifications and Samples: RAIS 2010

	Benchmark (1)	Male (3)	Prime-Age Male (4)
Race History			
'11': White/White	−0.003	−0.005	−0.004
	(0.0010)	(0.0012)	(0.0014)
'10': White/Non-White	−0.021	−0.023	−0.027
	(0.0010)	(0.0013)	(0.0035)
'01': Non-White/White	0.017	0.016	0.016
	(0.0010)	(0.0012)	(0.0036)
N	3,000,688	2,152,346	1,584,355
R2	0.195	0.2050	0.2049

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. Column (2) restricts the sample to male workers. Column (3) restricts the sample to prime-age workers (25–60 years of age).

Footnotes

---

[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.

[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.

[3] In the U.S., empirical evidence suggests that racial wage gaps are larger for darker skinned Black men and the intra-race disparity grows with labor-market experience (see Goldsmith et al. (2006) and Kreisman and Rangel (2015))

[4] 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.

[5] 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, Muendler and Ramey 2008), trade (Poole 2013; Krishna, Poole and Senses 2014), firm spin-offs (Muendler, Rauch and Tocoian 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.

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

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

---

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

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

[10] 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 V and the Appendix for details.

[11] 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".

[12] The mean separation rates reported in Table 1 do not imply most job changers are employed in plants that separate their entire workforce in a given year. Across all plants in RAIS, the average separation rate across plants is 0.51, but the median separation rate is 0.33. The median separation rate for continuing workers is 0.427. The median separation rate at the origin job in the job changer sample is 0.75, and at the destination job, it is 0.69.

[13] 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.

[14] 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.

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

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

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

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

[19] These findings are not sensitive to the inclusion of controls for the occupation of the origin and destination job (see Table 11 in Section VI.E), and hold when the sample is restricted to male workers only, and to prime-age male workers (see Table A.3).

[20] An exception to this statement occurs in the case where (1) reported race histories are uncorrelated with the history of observed covariates and (2) the probability of reporting a given racial classification is independent of the underlying market race. In this case, the reduced-form coefficients on all race history variables will be zero. We thank a referee for pointing out this exception.

[21] These specifications control for arbitrary plant and worker-specific heterogeneity in the spirit of Abowd et al. (1999).

[22] The estimated persistence across jobs 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.

[23] This analysis is inspired by Gibbons and Katz (1992), who focus on displaced workers to alleviate endogenous mobility bias in estimating the inter-industry wage premium.