Assortative Matching or Exclusionary Hiring? The Impact of Firm Policies on Racial Wage Differences in Brazil[*]

François Gerard
Columbia University and NBER

Lorenzo Lagos
Columbia University

Edson Severnini
Carnegie Mellon University and IZA

David Card
UC Berkeley and NBER

October 2018

Abstract

A growing body of research shows that firms' employment and wage-setting policies contribute to wage inequality and pay disparities between groups. We measure the effects of these policies on racial pay differences in Brazil. We find that nonwhites are less likely to work at establishments that pay more to all race groups, a pattern that explains about 20% of the white-nonwhite wage gap for both genders. The pay premiums offered by different employers are also compressed for nonwhites relative to whites, contributing another 5% of the overall gap. We then ask how much of the under-representation of nonwhites at higher-paying workplaces is due to the selective skill mix at these establishments. Using a counterfactual based on the observed skill distribution at each establishment and the nonwhite shares in different skill groups in the local labor market, we conclude that assortative matching accounts for about two-thirds of the under-representation gap for both men and women. The remainder reflects an unexplained preference for white workers at higher-paying establishments. The wage losses associated with unexplained sorting and differential wage setting are largest for nonwhites with the highest levels of general skills, suggesting that the allocative costs of race-based preferences may be relatively large in Brazil.

Race exerts a powerful influence on wage outcomes in many countries.[1] Though part of the earnings gap between different racial groups can sometimes be explained by differences in education or other observed factors, unexplained pay disparities are a major policy concern in the U.S., Brazil, and other nations. Most economic research on race-related wage differentials builds on the approach of Becker (1957), who assumed that each worker faces a *market-determined* wage.[2] A growing body of work on frictional labor markets, however, suggest that wages also incorporate *firm-specific* pay differences that can contribute to pay disparities.[3] When employers have wage-setting power, the racial pay gap will depend in part on the extent to which higher-paying firms differentially employ whites versus nonwhites – a between-firm *sorting effect* – and in part on the relative size of the pay premiums offered by a given firm to different race groups – a *relative wage-setting effect*.

Findings from three complementary strands of research suggest that these effects may be important. Randomized audit studies in many countries show that employer call-back rates for minority job applicants are lower than those for whites, implying that some employers set a higher bar for nonwhite candidates, or avoid hiring minorities altogether.[4] Observational studies show that minority managers are more likely to hire and retain minority applicants than their non-minority counterparts (e.g., Giuliano, Leonard, and Levine, 2009, 2011; Giuliano and Ransom, 2013; Aslund, Hensvik, and Skans, 2014), pointing to potential discrimination by some managers. And workplace-level studies of employee segregation show substantial segmentation by race (Hellerstein and Neumark, 2008; Hellerstein, Neumark, and McInerney, 2008) and ethnicity (e.g., Aslund and Skans, 2010; Glitz, 2014).

Nevertheless, it is unclear how much these patterns matter for white-nonwhite

---

[1]For overviews focusing on the U.S., see Altonji and Blank (1999), Fryer (2010), and Bayer and Charles (2018). For a summary of race-based differences in Latin America, see Nopo (2012) and Cano-Urbina and Maso (2016). For evidence on race differentials in the U.K. and Canada, see Blackaby, Leslie, and Murphy (2002) and Pendakur and Pendakur (1998, 2002), respectively.

[2]See Charles and Guryan (2008, 2011) for recent analyses that build directly on Becker's model, and Hirata and Soares (2016) for an application in Brazil.

[3]See Manning (2011) for a review of frictional and imperfect competition models of the labor market. Black (1995) presented an early search-based model of discriminatory hiring. Lang and Lehmann (2012) present a review of the discrimination literature emphasizing frictional labor market models. Card, Cardoso, and Kline (2016) study the effects of firm-specific wage setting on gender wage gaps in Portugal.

[4]Zschirnt and Ruedin (2016) summarize 36 studies in OECD countries: they find a median call-back rate for minorities relative to whites of 0.67, which is very close to the rate estimated in the seminal study by Bertrand and Mullainathan (2004). A recent audit study focusing on job openings for recent college graduates in Mexico City (Arceo-Gomez and Campos-Vasquez, 2014) finds a similar pattern for indigenous-looking female applicants.

earnings differences (see Lang and Lehmann, 2012, for a careful discussion). To quantify the impacts, we use rich administrative data on the formal sector in Brazil to estimate a series of two-way fixed effects (or "job ladder") models for wages, separately by race and gender, that include worker and establishment fixed effects. We then use the estimated wage premiums for white and non-white workers at each establishment, together with data on the distributions of workers across workplaces, to conduct a series of decompositions that identify the contributions of between-firm sorting and within-firm relative wage setting on racial pay gaps for each gender.

Racial differences in firm's employment and wage-setting policies are particularly relevant in our context. A steady stream of research since Silva (1978, 1980, 1985) and Oliveira, Porcaro, and Araújo (1981) has shown that the unexplained wage gaps between whites and nonwhites in Brazil are similar to those in the U.S., despite differences in the historical background and legal setting.[5] Several studies have shown that the gaps are especially large at the top of the distribution, pointing to the scarcity of nonwhites in high-paying industries and occupations (Soares, 2000; Henriques, 2001; Campante, Crespo, and Leite, 2004; Chadarevian, 2011; Mariano et al., 2018). Similar patterns have been documented in other Latin American countries (e.g., Nopo, 2012). Yet, employers often argue that this pattern is due to the relative shortage of qualified nonwhite workers, rather than to biases in their policies.

Our main analysis relies on the *Relação Anual de Informações Sociais* (RAIS), a longitudinal matched worker-firm database with information on workers' race, features that are key for our approach. RAIS covers essentially all formal employment in Brazil, but misses the informal sector. To address concerns over selectivity into the formal sector, we thus begin by using household survey data from the *Pesquisa Nacional por Amostra de Domicílios* (PNAD) to study wages and participation rates in formal and informal employment. We show that the raw shares of whites and nonwhites employed in the formal sector are comparable within each gender, and that there is no formality gap once we condition on age, location, and education. In contrast, the unexplained racial pay gaps are large (and similar) in the two sectors.

Two other important considerations in our context are the classification of race and the impacts of the minimum wage. Traditionally, nonwhites in Brazil are categorized into two groups: blacks and mixed race individuals, who comprise about 10% and 40% of the population, respectively. Consistent with previous studies, we find

---

[5]Andrews (1992) presents an historical comparison of racial differences in the U.S. and Brazil and concludes that differences were greater in the U.S. until the 1960s, but are now similar or even larger in Brazil. See also Cavalieri and Fernandes (1998), Arcard and d'Hombres (2004), Reis and Crespo (2015), Matos and Machado (2006), Garcia, Nopo, and Salardi (2009) and Bailey, Loveman, and Muniz (2013) for more recent studies of racial wage differences in Brazil.

that the wage and educational differences between these groups are small, though both have lower levels of education and earn substantially less than whites of the same gender. We therefore combine them into a single nonwhite category.[6] Regarding the minimum wage, simple histograms suggest that the federally legislated wage floor exerts strong upward pressure on wages in many regions of Brazil, reducing the effects of firm-specific wage setting and narrowing the wage gap between whites and nonwhites.[7] For our main analysis, we therefore focus on the Southeast region – the highest-income region of Brazil – though we also present results for the entire country. Additionally, we present separate results for workers with a high school education or more, whose pay is less impacted by the mininum wage.

With this background, we turn to our results. Consistent with findings from the U.S., Germany, and other countries, and with previous work by Lavetti and Schmutte (2016) and Alvarez et al. (2018) on Brazil, we find that differences in the wage premiums paid by different establishments explain an important share ($\approx 20\%$) of the variation in hourly wages for all four race-gender groups.[8] The average workplace premiums earned by whites are higher than those earned by nonwhites, accounting for about one-quarter of the racial pay gap for both men and women.[9] We also find a strong pattern of positive assortative matching *within* each race-gender group. Specifically, we estimate that establishments that pay 10% higher wage premiums have workers who would earn 5-8% more *at any workplace.* Given the education gaps, we would therefore expect to see some differential sorting of whites to higher-paying

---

[6]Cornwall, Rivera, and Schmutte (2017) present an interesting analysis of "endogeneity" in the measurement of race in RAIS. We thus conduct a number of checks to assess the impacts of potential measurement errors in employer-reported race in RAIS.

[7]Several recent papers, including Komastsu and Menezes-Filho (2016) and Alvarez et al. (2018), argue that the rise in the minimum wage after the mid-1990s contributed to lowering overall wage inequality in Brazil. Derenoncourt and Montialoux (2018) show that extensions in coverage of the minimum wage in the mid-1960s contributed to a narrowing of black-white wage gaps in the U.S.

[8]Abowd, Lengermann, and McKinney (2003) find that between-firm variation represents about 17% of the variance of U.S. wages. Card, Heining, and Kline (2013), Card, Cardoso, and Kline (2016), and Macis and Schivardi (2016) estimate a 15%-20% share for establishment effects in the case of German workers, Portuguese male workers, and Italian manufacturing workers, respectively.

[9]A key issue underlying this comparison is how to benchmark the estimated establishment effects for whites versus nonwhites. We use a normalization based on pay in the restaurant industry that allows us to standardize the establishment (and person) effects for the two race groups (by gender). We also evaluate the robustness of our conclusions to a range of assumptions on how much of the racial pay gap in that industry is attributable to wage premiums paid only to whites, and to normalizations based on other industries where wage-setting power is likely limited. To address the issue that whites are concentrated in geographic areas with a higher fraction of larger and more profitable firms, we also implement a simple reweighting procedure for all our decomposition results, which adjusts the geographic distribution of nonwhites to match the distribution of whites.

establishments even in the absence of any discriminatory employment policies.

To assess how much, we construct estimates of the distributions of white and non-white workers in different skill groups in each local labor market, based on age and the percentiles of their estimated person effects.[10] We then compare the actual employment shares of nonwhites at each workplace to the expected shares if establishments maintained the skill-age composition of their labor force but selected workers without regard to race from the available pool in their local labor market. This counterfactual suggests that about two-thirds of the overall *sorting effect* – accounting for about 12% of the overall racial wage gap – is explained by race-neutral assortative matching. The remainder, which incorporates discriminatory hiring and retention policies, accounts for another 6-8% of the overall racial wage gap.

Next, we use the estimated establishment-specific wage premiums to evaluate the within-firm *relative wage-setting effect*. We find that the wage premiums for nonwhites are compressed relative to those for whites – a pattern that is consistent with monopsonistic wage setting and lower elasticities of firm-specific supply ("price discrimination;" Barth and Dale-Olsen, 2009; Card et al., 2018) or lower bargaining power (Babcock and Laschever, 2003; Manning, 2011) for nonwhites than whites. These lower average premiums explain another 5-6% of the overall racial wage gap.

Finally, we show that the wage losses associated with unexplained sorting and differential wage-setting are largest for nonwhites with the highest levels of general skills (as captured by their person effects), suggesting that the allocative costs of race-based preferences may be relatively large in our setting.

Our work makes three main contributions. First, we advance the literature on discriminatory employment policies and workplace segregation, offering comprehensive estimates of the impacts of these practices on overall wage gaps for both males and females in Brazil, and showing how the effects vary across the skill distribution.[11] Second, we show how estimates from a two-way fixed effects model can be used to benchmark the employment patterns at a given workplace relative to its local labor market, while accounting in a flexible way for the skill composition at the workplace. This is particularly important in settings like Brazil where there is a large racial gap

---

[10]Our approach generalizes the method proposed by Aslund and Skans (2010), which accounts for observed skill characteristics of employees at a given workplace and in the surrounding labor market, by accounting for any unobserved but time-invariant skill characteristics.

[11]Previous work by Hellerstein and Neumark (2008) used U.S. data from a single cross-section and found that black workers were more likely than whites to work at higher-wage establishments - the opposite of the pattern in our data. An early study by Ashenfelter (1972) similarly found that black workers were more likely than whites to work at unionized jobs in the late 1960s. We are unaware of any work for the U.S. that has longitudinal data covering all (or most) establishments and includes racial status information.

in education levels. Third, we contribute to the literature on racial wage differences in Latin America, showing that firm-specific employment and pay-setting policies contribute a substantial share of these gaps, that assortative matching would exacerbate racial inequalities even in the absence of any discrimination, but that race-based preferences appear to play an important role at the top of the skill distribution.

The paper proceeds as follows. Section 1 begins with a descriptive analysis of racial differences in the Brazilian labor market, based on the PNAD data. Section 2 presents our econometric setup and decomposition methods. We then provide a brief descriptive analysis of our main RAIS samples in Section 3, focusing on the plausibility of the conditions for OLS to yield interpretable estimates of employer wage premiums. Sections 4 and 5 present our main estimation and decomposition results, using data from the Southeast region of Brazil. The latter section ends with a detailed analysis of how unexplained sorting and relative wage setting affect workers at different points in the skill distribution. We evaluate the robustness of our findings in Section 6, by replicating our analysis using other regions of the country, different measures of race, different sub-periods of our sample, industries with different degrees of interaction between employees and customers, and alternative normalizing assumptions regarding the wage premiums paid to whites and nonwhites in the lowest-wage sectors of the economy. Section 7 concludes.

1 Background and Data

We begin by providing some background information on our empirical setting and the data used in our analysis.

Legal Setting

Although African slave labor played a major role in colonial Brazil, a relatively fluid notion of racial identity emerged in the post-slavery era, manifested by the absence of de jure segregation and the acknowledgement of three main race groups: whites; mixed race individuals ("pardos," literally, brown people); and black/African race individuals.[12] Perhaps in part because of this fluidity, legal concerns over racial discrimination emerged relatively late (Skidmore, 1992). Indeed, it was only with the adoption of the 1988 Constitution and the passage of subsequent laws in 1989 and 1995 that racial discrimination in employment and pay setting became illegal

---

[12]See Skidmore (1974), Marx (1998), Telles (2004), and Andrews (1992) for detailed discussions.

in Brazil.[13] Nevertheless, a review of case law suggests that even in the early 2000's most claims of discrimination were dismissed (Equal Rights Trust, 2009).

Recently, however, the adoption of affirmative action policies in university admissions (Francis and Tannuri-Painto, 2013, 2015) has led to heightened awareness of racial issues in Brazil and a new law promoting racial parity (the Racial Equality Law) was passed in 2010. Therefore, it is possible that employer policies regarding employment and wage setting have evolved during the 13 years (2002-2014) included in our sample. We explore this possibility as part of our robustness analysis.

Initial Descriptive Analysis

We begin by studying data from the *Pesquisa Nacional por Amostra de Domicílios* (PNAD), a nationally-representative household survey. The PNAD covers both the formal and informal sectors, allowing us to assess the potential role of formality in mediating the size of the wage gap between different race groups. For consistency with our RAIS data set (see below), we pool the 2002-2014 PNAD surveys.[14] We also limit attention (here and throughout the paper) to men and women age 25 to 54 with at least one year of potential experience (using age and years of schooling).

Table 1 provides an overview of the characteristics of the working-age population in Brazil by gender (panels A and B) and race (columns 1-4), with a parallel analysis for the Southeast region in columns 5-8.[15] The Southeast region contains the states of Espírito Santo, Minas Gerais, Rio de Janeiro, and São Paulo, 42% of Brazil's population, and its most developed areas, including its three largest cities.[16]

The first row in each panel shows the fraction of the population in each race group. In Brazil as a whole, about 50% of the working-age population are white ("branco"), 42% are mixed race ("pardo"), and 8% are black ("preto"). Together, these three groups account for about 99% of the population, with Asians and indigenous groups making up the remainder. In the Southeast, the share of whites is higher (56% of men and 59% of women), while the share of mixed race individuals is lower (32-34%).

The second row shows the fractions of each group who were working as private-

---

[13]The Afonso Arinos Law, passed in 1951, was intended to deter racial discrimination, but is widely believed to have had at most a symbolic effect (Campos, 2015).

[14]Ferreira, Firpo, and Messina (2014) present an analysis of PNAD data from 1995 to 2012. They document trends over this period in returns to education, racial wage gaps, and overall wage inequality, with particularly large changes from the mid-1990s to the early 2000s. To assess the impacts of these underlying trends, we split our sample into earlier and later periods in Section 6.

[15]Similar descriptive statistics for the other regions are reported in Table D1 in the Appendix.

[16]A map showing Brazil's regions and micro-regions (our definition of local labor market; see below), as well as their racial composition, is presented in Figure C1 in the Appendix.

sector nonfarm employees at the PNAD survey date (in September). About 43% of men and 25% of women are in this class in Brazil; the remainder includes self-employed workers, farm laborers, domestic workers, public sector employees, and nonworkers. The private employment rate of men differs by only a few percentage points (ppts.) across the three main race groups, but is more variable for women. In the Southeast region, private employment rates of both genders are higher and less variable across race groups. Specifically, the rate ranges from 47% (for whites) to 51% (for blacks) for men and from 23% (for mixed race individuals) to 27% (for whites) for women. This similarity implies that there is relatively little room for differential selection biases among the subset of private-sector employees in each race group.

In the remaining rows, we show selected characteristics for nonfarm private-sector employees.[17] Among male private-sector employees, mean schooling is 8.4 years and 45% have completed high school. Both rates are higher among female employees, reflecting the fact that women tend to be better educated than men in Brazil. White men have about 1.6 years more schooling than mixed race or black men, and are 40% more likely to have finished high school. The corresponding gaps between white and mixed race or black women are smaller, but still notable. In the Southeast region, mean levels of schooling and the high school completion rate are both higher, but the gaps between the race groups are comparable to those in the country as a whole.

While the differences in education between whites and nonwhites in Brazil are larger than those that currently prevail in the U.S. (see, e.g., Bayer and Charles, 2018), they are broadly consistent with gaps in other Latin American countries. Esteve and López-Ruiz (2010), for example, document that the proportion of adults in Brazil who have not completed primary education is about 20 ppts. higher for nonwhites than whites. In Chile, the comparable gap between mapuches (the main indigenous group) and non-indigenous people is about 15 ppts.; in Ecuador the gap between blacks and whites is about 25 ppts.; and in Mexico the gap between indigenous Spanish speakers and non-indigenous people is about 23 ppts.

Looking next at the mean log hourly wage statistics presented in Table 1, two key patterns stand out. First, mixed race and black workers earn 30-35% less than whites. These gaps are similar for men and women, and in the Southeast region. Second, mean log wages of all groups are roughly 10% higher in the Southeast region.

The level of Brazil's minimum wage was relatively high during our sample period.[18] Indeed, 54% of male and 71% of female private-sector employees earned less

---

[17]All monetary values in the paper are deflated using the Consumer Price Index to a 2010 base.

[18]The ratio of the minimum wage to the median wage rose from 58% in 2002 to 70% in 2006, and was relatively constant afterwards (see Melo, 2014). By way of comparison, the ratio of the Federal minimum wage to the median wage was about 38% in the U.S. and 62% in France in 2012.

than 200% of the minimum wage. Consistent with the patterns for mean log wages, these fractions were higher for mixed race and black workers than for whites of either gender, and lower in the Southeast region for all groups.

The impact of the minimum wage is illustrated visually in Figure 1, where we show the distributions of log wages normalized relative to the minimum wage – i.e., $\log(w/min)$ – for white and non-white male workers in the Northeast region (the poorest region of the country, with the highest fraction of black workers) and the Southeast region (the richest region of the country).[19] We show the distributions for all male workers, for males with less than a high school education, and for males with a high school education or more in the top, middle, and bottom panels, respectively.

For all three education groups in the Northeast region, we see a large spike in the distribution at a relative wage of 1 (or a log relative wage of 0), coupled with a stark asymmetry between the upper and lower tails of the distribution. These graphs suggest that the minimum wage substantially attenuates any firm-specific component of pay in the Northeast region and equalizes pay across race groups. The graphs for workers in the Southeast region also show a notable spike at a relative wage of 1 and some asymmetry between the upper and lower tails. Nevertheless, there is a clear leftward shift in the distribution of wages for nonwhites relative to whites, suggesting that the impact of firm-specific wage setting may be detectable, though it might be attenuated for workers with less than a high school education.

To address concerns about the potential effects of the minimum wage we follow two strategies. First, we focus on the Southeast region of the country for our main analysis. Second, throughout our analysis, we present results for all education groups and for higher-educated (high school or more) workers separately. The impacts of the minimum wage appear to be relatively small for workers with at least a high school education in the Southeast region, so the findings for this group may give a clearer picture of what could be expected in the absence of a binding minimum wage.

Returning to Table 1, the last row of each panel shows the fraction of private-sector employees who have a valid working card ("carteira de trabalho") for their job and are thus in the formal sector. This rate is about 80% in Brazil and 83% in the Southeast region. Importantly, the formality rate is quite similar across race groups, suggesting that differences in formality between the groups are not a major concern for interpreting measured wage gaps in the formal sector (more on this below).

---

[19]Similar graphs for females are shown in Figure C2 in the Appendix.

An Overview of Racial Wage Gaps

Table 2 presents estimates from a series of simple regression models that use data from the PNAD to measure the size of the racial wage gaps for male and female private-sector employees in Brazil (including formal and informal employees). Again, we present parallel results for Brazil as a whole and for the Southeast region.[20] We also present separate results for all workers and the subset with a high school education or more. For each group we present two specifications: one that includes only state and year effects, and another that adds controls for education (a set of five dummies for incomplete elementary school, and complete elementary school, middle school, high school, or college) and a quadratic function of potential experience.

The estimated wage gaps between whites and the two main groups of nonwhites range from 27% to 33% when we control only for state and year effects. As has been found in numerous previous studies, including the seminal studies by Oliveira, Porcaro, and Araújo (1981) and Silva (1978, 1980, 1985), mixed race and black workers receive similar average wages that are both far below the average wages of whites. The racial wage gaps are quite similar for males and females, but are about 3 ppts. larger in the Southeast than in Brazil as a whole, perhaps reflecting the reduced impact of the minimum wage. Finally, the gaps are 3-5 ppts. larger for better educated men, but only slightly larger for better educated women.

The racial wage gaps are substantially reduced when we add controls for education and experience. For the country as a whole (column 2), the unexplained gaps fall to 11-13 ppts. for males and to 11 ppts. for females. The drop in magnitude compared to the gaps without controls (column 1) reflects the relatively large racial differences in educational achievement documented in Table 1, and the relatively high return to education in the Brazilian labor market (e.g., Psacharopoulos and Patrinos, 2002). Interestingly, the unexplained gaps are very similar in the Southeast region (column 6). The wage gaps among workers with a high school education or more also fall by nearly 50% when we add controls for education (in this case, a single dummy for having completed a bachelor's degree or more), though the gaps tend to be a little larger for this group (13%-21%, versus 11%-14% for all workers).

These simple models lead us to two main conclusions. First, we confirm the finding from other recent studies that mixed race and black workers face similar wage penalties relative to whites. In the remainder of the paper, we thus combine these two groups into a single non-white group. This has the advantage of creating a relatively large non-white group, which is useful when we turn to two-way fixed effects models. Second, we find that the *unexplained wage gaps* between whites

---

[20]Similar estimates for the other Brazilian regions are reported in Table D2 in the Appendix.

and nonwhites are fairly similar in the Southeast and in the overall Brazilian labor market. In the main analysis, we then focus on the Southeast region, but we present results for the whole country in the robustness checks.

Wage Gaps in RAIS

Our main analysis uses the *Relação Anual de Informações Sociais* (RAIS), a longitudinal data set that provides nearly universal coverage of formal jobs in Brazil.[21] Firms submit annual information to the Ministry of Labor on all employees who were on the payroll in the previous year, including their hiring and separation dates, average monthly earnings during the year, monthly earnings in December, contracted hours, age, gender, education, and race. Worker information is reported at the establishment level along with the industry and municipality of the workplace. Race is classified into the same categories used in PNAD, but is only available after 2002.[22] Hence, we use the RAIS files from 2002 to 2014, the last year available for this study.

To construct an hourly wage, we use information on contracted monthly hours and monthly earnings in December of each year, the month for which we measure earnings precisely, restricting attention to individuals who worked for their employer for the full month.[23] Conceptually, this wage is similar to that in PNAD, which also measures earnings and hours for a cross-section of jobs at one point in time in each year. Finally, we exclude farm workers and those outside the 25-54 age range from our RAIS samples (as in the PNAD samples), as well as workers on temporary contracts, those who are not paid on a monthly basis (the usual pay period in Brazil), and those with very low or very high wages (see details in Appendix A).

---

[21]Established in 1975, RAIS provides crucial information about the formal labor force in Brazil, including labor market indicators made available to public and private organizations. The data collected by RAIS are also used to administer a federal wage supplement to low-income formal employees ("Abono Salarial") and to monitor eligibility for various government programs, such as the Brazilian conditional cash transfer program ("Bolsa Familia"). Compliance with the mandatory reporting requirements is high because of large penalties when the data are late or incomplete.

[22]Cornwall, Rivera, and Schmutte (2017) describe in details the process by which employers record race. The requirement for employers to record workers' race was added as part of an effort to comply with the ILO Convention 111, which deals with discrimination in the workplace. Newspaper articles show that firms fought the new requirement at the time, but did not prevail.

[23]For the small fraction of workers who have more than one job in December, we first select the job with the highest contracted hours, breaking any ties by selecting the job with the highest hourly wage. In the few cases where hours and wages are identical on both jobs, we select one at random. RAIS does not report actual hours worked, but we find no racial gap in actual hours worked for formal-sector employees in PNAD (see Appendix Table D3). The PNAD data, which pertain to September, may not perfectly capture differences in hours worked in December.

A concern in RAIS is that race is sometimes recorded differently by different employers: among workers in the Southeast region whose modal race is white, their race is coded as mixed race or black about 10% of the time, while for those whose modal race is nonwhite, their race is coded as white about 15% of the time.[24] Similar anomalies occur, albeit less frequently, for the recording of education, gender, and birth year. To address these inconsistencies, we assign individuals their modal race, education, gender, and birth year across all their observations in the RAIS sample.

A second concern is that, although RAIS includes detailed longitudinal information for the formal sector, it excludes the entire informal sector. To assess how the exclusion of the informal sector affects the measured racial wage gaps, we present results from a series of models fit to the PNAD and RAIS samples in Table 3. The models have the same full set of controls as the models in Table 2, but combine black and mixed race individuals into a single non-white category.

The specifications in column 1 (for Brazil as a whole) and 5 (for the Southeast region) model the probability of being in a formal job, conditional on being a private-sector employee in PNAD. The coefficients fall in a narrow range from -0.01 to 0.01, suggesting that formality rates of whites and nonwhites of both genders are nearly identical when we control for education, experience, state of residence, and year.[25]

The remaining specifications in Table 3 model log hourly wages. The models in columns 2 and 6 are fit to samples that include all private-sector employees in PNAD, while those in columns 3 and 7 are fit to subsamples of formal-sector workers in PNAD. Importantly, the estimated wage gaps between white and non-white workers are very similar for both men and women regardless of whether informal sector workers are excluded or not. Together with the relatively high overall rates of formality among private-sector employees, this suggests that an analysis of racial wage gaps in the formal sector can provide useful insights for the entire labor market.

Finally, columns 4 and 8 present results for our main RAIS samples. In particular, we limit attention to employees at the largest connected sets of workplaces for both white and non-white workers (of a given gender), which we call the "dual connected" set of establishments (see the discussion in Section 3). Hourly wages are somewhat higher than in the corresponding PNAD samples,[26] while the racial wage gaps are a little smaller, particularly for men. One partial explanation for the smaller wage gaps

---

[24]As discussed by Cornwall, Rivera, and Schmutte (2017), the inconsistent reporting of race is not entirely due to random misclassification errors, since workers who move to a better paying job tend to be more likely to switch from nonwhite to white (and vice-versa).

[25]Similar estimates for the other Brazilian regions are presented in Table D4 in the Appendix.

[26]This might be due to the use of actual earnings (which can include overtime payments) but contracted hours. If overtime is more prevalent in December (when RAIS data are measured) than September (the data collection month for PNAD), hourly wages will be higher in RAIS.

in RAIS is that the mis-measurement of non-white racial status in the administrative data (even after assigning individuals their modal race) leads to an attenuation bias.[27] To assess this explanation, we re-estimated the models using only the subsample of "consistent race" individuals whose (binary) race classification is the same in every year they are observed in RAIS. This increases the magnitudes of the racial wage gaps, leading to gaps for females that are very similar to those in PNAD, and gaps for males that are only 2-3 ppts. smaller than in PNAD. In light of this finding, we present results for workers with consistent race histories in our robustness analysis, as well as for our full sample but classifying individuals by their first reported race in RAIS rather than their modal race.

2   Econometric Framework

In this section, we present our econometric framework for measuring the effects of firm-specific employment and pay-setting policies. We measure these effects by their net impact on the racial wage gap. Specifically, we estimate models that capture the wage premiums offered at each workplace, then perform simple counterfactual exercises to evaluate the effects of assuming that (1) each workplace offered the same wage premium (relative to other employers) to nonwhites and whites and (2) non-whites and whites had the same probabilites of employment at different workplaces. Our counterfactuals take the wages offered at each workplace as given and ignore the equilibrium effects emphasized by Becker (1957) that can arise from discriminatory preferences in the marketplace. Therefore, our procedure likely *under-estimates* the overall impact of discriminatory preferences on the racial wage gaps.

Job Ladder Model of Wages

Building on Abowd, Kramarz, and Margolis (1999) – henceforth AKM – we assume that the log of the hourly wage paid to worker i in race-gender group g in period t (y_{git}) is generated by a model of the form:

$$\ln y_{git} = \alpha_{gi} + \psi^g_{J(g,i,t)} + X'_{git}\beta_g + \varepsilon_{git} \tag{1}$$

where α_{gi} is a person effect that captures any time-invariant but fully portable components of earnings capacity, ψ^g_j represents a *wage premium* paid at establishment j to workers in group g, $J(g,i,t)$ is an index function indicating the workplace for

---

[27]In particular, the share of non-white workers among formal private-sector employees remains smaller in RAIS than in PNAD, even after assigning individuals their modal race.

worker i in group g in year t, X_{git} is a vector of time varying controls (e.g., year effects and controls for individual experience), and ε_{git} is a time-varying error capturing all other factors, including any person-specific job match effects.[28] The ψ_j^g terms capture workplace-specific pay premiums, but impose the assumption that the proportional premiums are the same for all workers in a given race-gender group.

One simple explanation for the presence of firm- or establishment-based premiums is monopsonist wage setting (see Card et al., 2018, for a review). Assuming that firms have some market power, they will set wages that are marked down relative to marginal revenue products, with a factor that depends on the elasticity of supply to the firm (Robinson, 1933; Boal and Ransom, 1997; Manning, 2003). Firms with a higher demand for labor (arising, e.g., from entrepreneurial skill) will set higher wages to attract a larger labor force. As discussed in Appendix B, under a set of simplifying assumptions about the choice model causing different workers to choose different employers and the substitutability between subgroups of workers, an optimal wage-setting policy will be characterized by a set of group-specific premiums:

$$\psi_j^g = \delta_g R_j \tag{2}$$

where R_j is a measure of latent productivity at establishment j and δ_g is a markdown factor that varies across subgroups of workers and is related to their elasticities of firm-specific supply (as in models of price discrimination; more on this below). An observationally equivalent model is that workers' wages incorporate a share of the surplus associated with their employment match, that the average match surplus is higher at more productive firms, and that different groups have different average bargaining power (Babcock and Laschever, 2003; Card, Cardoso, and Kline, 2016).

The empirical predictions of equation (1) depend on what is assumed about the error component ε_{git}. If the conditional expectation of ε_{git} is assumed to be independent of the job history of the worker – the "exogenous mobility" assumption required for OLS estimation of (1) to yield unbiased estimates of the establishment effects – then (1) implies that a worker in group g who moves from establishment k to establishment j will experience an average wage change of $\psi_j^g - \psi_k^g$, regardless of past or future mobility patterns. A worker who moves in the opposite direction (from j to k) will experience an equal and opposite expected wage change of $\psi_k^g - \psi_j^g$. This simple symmetry prediction contrasts with the predictions from models of mobility driven

---

[28]The contributions of the person effect and the time-varying covariates are not separately identified without a normalizing assumption, as discussed in Card et al. (2018). Following their work, we assume that in the baseline year $X'_{git}\beta_g = 0$ for 40-year old males and 35-year-old females, such that the person effects are measured as of age 40 for men and 35 for women, which correspond approximately to the peak of their experience profiles (see Figure C3 in the Appendix).

by job match effects, which imply that movers in *both directions* will experience positive average wage gains (see Eeckhout and Kircher, 2011, 2018, for a discussion of alternative theoretical models of mobility and sorting). A series of specification checks presented by Card, Heining, and Kline (2013) suggest that although the exogeous mobility assumption can be rejected by formal testing, the predictions of an AKM-type model with exogenous mobility are not too far off. In the next section we review these checks using RAIS data and confirm that this is also true in Brazil.

Estimates of AKM-style models for various countries, including the U.S., Germany, and Brazil, have found that workplace (or firm) effects in these models typically explain 15-25 percent of the overall variance of wages (see the review in Card et al., 2018). Workers with higher person effects also tend to work at firms that pay higher wage premiums. Such assortativeness has implications for wage inequality because the variance of log wages for workers in group g can be decomposed as:

$$Var\left(\ln y_{git}\right) = Var\left(\alpha_{gi}\right) + Var\left(\psi_{J(g,i,t)}^g\right) + Var\left(X_{git}'\beta_g\right) + Var\left(\varepsilon_{git}\right) \qquad (3)$$
$$+2Cov\left(\alpha_{gi}, \psi_{J(g,i,t)}^g\right) + 2Cov\left(\alpha_{gi}, X_{git}'\beta_g\right) + 2Cov\left(\psi_{J(g,i,t)}^g, X_{git}'\beta_g\right).$$

A positive covariance between worker and establishment effects will magnify the impacts of the person and firm components, contributing to higher overall inequality. As we discuss next, it also has important implications for the interpretation of differential employment patterns in higher- and lower-premium workplaces.

Impacts of Sorting and Relative Wage-Setting on Racial Wage Gaps

We now ask how the presence of establishment-specific wage premiums contributes to mean wage differences between groups. Suppose there are two groups, whites (W) and nonwhites (N), and let π_{Wj} and π_{Nj} represent the fractions of the groups employed at workplace j. Then the mean wages of the two groups can be written as:

$$E[\ln y_{Wit}] = E[\alpha_{Wi} + X_{Wit}'\beta_W] + \sum_j \psi_j^W \pi_{Wj}$$
$$E[\ln y_{Nit}] = E[\alpha_{Ni} + X_{Nit}'\beta_N] + \sum_j \psi_j^N \pi_{Nj}$$

Subtracting these equations leads to a simple expression for the mean log wage gap between whites and blacks:

$$E[\ln y_{Wit}] - E[\ln y_{Nit}] = \alpha_W - \alpha_N + \overline{X}_W'\beta_W - \overline{X}_N'\beta_N + \sum_j \psi_j^W \pi_{Wj} - \sum_j \psi_j^N \pi_{Nj} \quad (4)$$

where $\alpha_g = E[\alpha_{gi}]$ and $\overline{X}_g = E[X_{git}]$. For simplicity, assume that $\overline{X}'_W \beta_W = \overline{X}'_N \beta_N$, so we can ignore the time-varying person components.[29] Further rearrangement then leads to two alternative decompositions:

$$
\begin{aligned}
E[\ln y_{Wit}] - E[\ln y_{Nit}] &= \alpha_W - \alpha_N + \sum_j \psi_j^W (\pi_{Wj} - \pi_{Nj}) + \sum_j (\psi_j^W - \psi_j^N)\pi_{Nj} \quad (5)\\
&= \alpha_W - \alpha_N + \sum_j \psi_j^N (\pi_{Wj} - \pi_{Nj}) + \sum_j (\psi_j^W - \psi_j^N)\pi_{Wj} \quad (6)
\end{aligned}
$$

Following Oaxaca (1973), the mean wage gap can be decomposed into a difference in mean characteristics between the two groups, weighted by the coefficients for one of the two groups, and a difference in coefficients, weighted by the mean characteristics of the other group. In a job ladder model, the "characteristics" are simply person indicators and dummies for working at a given establishment, while the "coefficients" are the worker effects and the establishment pay premiums. The first term in equations (5) and (6) is just the difference in the mean person effects for the two groups - what might be called the "average skill gap" between the two groups. The other two terms measure the contribution of establishment pay and employment policies to the wage gap, with alternative choices for which group's wage premiums are used to weight the difference in employment shares, and which group's employment shares are used to weight the difference in establishment pay premiums.

In the analysis below we focus on the version of the decomposition specified by equation (5). In this variant, the difference in pay premiums received by whites versus nonwhites is weighted by the employment share of nonwhites, yielding an estimate of the effect of differential pay-setting *given the actual distribution of nonwhites across establishments* – a counterfactual that we believe is most natural. Likewise, the difference in employment shares of whites and nonwhites is weighted by the wage premium for white workers, yielding an estimate of the effect of differential sorting of the two race groups across workplaces *assuming that nonwhites were paid the same premiums as whites* – again, a counterfactual that we believe is natural.

Sorting Effect and Assortative Matching

The between workplace sorting term $\sum_j \psi_j^W (\pi_{Wj} - \pi_{Nj})$ in equation (5) will be zero if the two groups have the same distributions of employment across establishments (i.e., $\pi_{Wj} = \pi_{Nj}$ for all j) or if there are no establishment-specific pay premiums

---

[29]As discussed below, this assumption is roughly correct for males in our data. For females, however, there are some modest differences between whites and nonwhites.

15

(as in traditional discrimination models), but it will be positive if white workers are more likely to be employed at high-premium workplaces.

There are several reasons to suspect that this is true. One is that whites are concentrated in geographic areas with a higher fraction of larger and more profitable firms in Brazil. To address this issue, we implement a simple reweighting procedure that adjusts the geographic distribution of nonwhites to match the distribution of whites. Specifically, we form a weight for each nonwhite based on the relative fractions of whites and nonwhites in his or her micro-region.[30] We then construct a weighted average wage for nonwhites, and weighted fractions of nonwhites at each establishment. Since the decompositions in equations (4)-(6) remain valid using weighted means, we can account for the differing geographic distributions of whites and nonwhites in a simple non-parametric fashion.

A second explanation is suggested by the general finding of positive assortative matching between higher-skilled workers and higher-paying establishments – a pattern that is also true in Brazil (see below). Assuming that whites tend to have higher overall human capital than nonwhites, and that higher-paying establishments hire relatively more skilled workers, we would therefore expect to see more whites at these establishments, even in the absence of other factors.

To account for such skill-biased employment patterns, we classify individuals (by gender) into skill groups based on their age and the value of their estimated person effects. We then calculate the fractions of workers at each establishment in each skill group, and the share of nonwhites among all workers in each skill group in each local labor market. Next, we calculate counterfactual employment shares of whites and nonwhites, π^*_{Wj} and π^*_{Nj}, respectively, that would be expected if each establishment maintained the skill distribution of its labor force in each year but selected workers without regard to race from the available pool in its local labor market in that year. Using these shares we form the counterfactual *skill-based sorting effect*:

$$\sum_j \psi^W_j (\pi^*_{Wj} - \pi^*_{Nj}). \tag{7}$$

This gives the net effect of skill-based (race-neutral) employment probabilities on the racial wage gap, holding constant the skill distribution at each workplace, the wage premium paid to white workers, and the racial composition of local labor markets.

A third explanation for an under-representation of nonwhites at higher-paying establishments is discriminatory hiring and/or retention policies. We cannot directly

---

[30]A micro-region ("microrregião") is a legally defined geographic entity roughly equivalent to a county. It closely parallels the notion of local economies by grouping economically integrated contiguous municipalities with similar geographic and productive characteristics. The 557 micro-regions in Brazil (160 of them are in the Southeast region) are shown in Figure C1 in the Appendix.

16

test this explanation. We can, however, calculate the difference between the actual sorting effect and the counterfactual skill-based sorting effect:

$$D = \sum_j \psi_j^W (\pi_{Wj} - \pi_{Nj}) - \sum_j \psi_j^W (\pi_{Wj}^* - \pi_{Nj}^*) \tag{8}$$

To the extent that higher-premium establishments employ fewer nonwhites than would be expected given their skill distribution and the nonwhite shares in each skill group in their local labor market, the *residual sorting effect* D will be positive.

Relative Wage-Setting Effect

The relative pay-setting term $\sum_j (\psi_j^W - \psi_j^N) \pi_{Nj}$ in equation (5) will be zero if $\psi_j^W = \psi_j^N = 0$ for all j (as in traditional discrimination models), or if the pay premiums for whites and nonwhites at each establishment are equal. It will be positive, however, if whites tend to receive higher pay premiums than nonwhites at a given workplace.

As noted above, the simple monopsonistic pay-setting model developed by Card et al. (2018) predicts a set of group-specific pay premiums of the form $\psi_j^g = \delta_g R_j$ where R_j is a measure of relative productivity at establishment j and δ_g is a group-specific preference factor. This implies that

$$\psi_j^N = \gamma \psi_j^W, \tag{9}$$

where $\gamma = \delta_N / \delta_W$. Under these conditions, the relative wage-setting effect is:

$$\sum_j (\psi_j^W - \psi_j^N) \pi_{Nj} = \frac{1-\gamma}{\gamma} \sum_j \psi_j^N \pi_{Nj}. \tag{10}$$

If $\delta_N < \delta_W$ – so nonwhites' wage premiums at more productive employers are compressed relative to whites' – then $\gamma < 1$ and the pay-setting effect will be positive.

In the monopsonistic pay-setting model in Appendix B, δ_g depends on the relative value that individuals in a group place on the wage versus nonwage features of a job, and on the variation in the individual-specific valuations for a given job by individuals in the group. Groups with a higher value of δ_g have more elastic supplies to a given establishment, and therefore receive lower monopsonistic markdowns relative to their marginal revenue products. For example, if $\delta_N/\delta_W = 0.9$ (i.e., nonwhites receive pay premiums that are about 90% as large as whites) and the average pay premium earned by nonwhites is 10%, the pay-setting effect will be about 1.1 ppts.

17

Normalizing the Pay Premiums

An important feature of the sorting effect in equation (5) is that it depends on the *differences in establishment shares* of whites and nonwhites. Since these establishment shares sum to 1, the numerical value of the estimated sorting effect is invariant to any additive transformation of the estimated pay premiums. To see this, consider the transformation $\widetilde{\psi}_j^W = \psi_j^W + \tau$. Since $\sum_j \tau(\pi_{Nj} - \pi_{Wj}) = 0$ for any τ, the transformed pay premiums imply the same numerical value of the sorting effect. This is important because the pay premiums estimated in two-way fixed effects models are always normalized relative to the pay premium in some reference firm (τ). Substantively, this means that the overall sorting effect can be estimated without taking a stand on how one normalizes the estimated pay premiums for a given group.

In contrast, the relative pay-setting effect in equation (5) depends on the *difference in the estimated pay premiums* for whites and nonwhites, and therefore on the normalization of these premiums. Consider the same transformation as above: $\widetilde{\psi}_j^W = \psi_j^W + \tau$, which adds a positive constant τ to the premiums for whites (reflecting, for example, a premium paid by firms in the reference sector to whites). This will shift up the estimated pay-setting effect by the amount τ. Mechanically, it will also reduce all the person effects for whites by the same factor, potentially affecting the decomposition of the overall sorting effect into its *skill-based* and *residual* components (as the person effects are used to classify individuals into skill groups). Note that the key issue is how to normalize the establishment effects for whites relative to nonwhites: a renormalization that adds the same factor to the premiums for both race groups leaves these effects unchanged.

We address the normalization issue by assuming that in the restaurant sector, a large employment sector for both men and women that is comprised of many small firms that pay relatively low wages, the average pay premiums for whites and nonwhites of both genders are zero. In essence, we assume that firms with little or no rent to share pay zero premiums to all groups. It is possible to show that under the assumptions posed by Card et al. (2018) about the shape of workers' preferences for jobs at different workplaces, firms that require only a small workforce set wages so that their employees are close to indifferent between employment at the firm and a non-work alternative (i.e., just above their reservation wage). In this case, assuming that restaurants are "small" employers, our normalization assumption will be valid.

Interestingly, in both the PNAD and RAIS data, we find very small racial wage gaps in the restaurant sector (see Table D5 in the Appendix). Specifically, models like those in Table 3 that include state and year effects and controls for education and experience show a racial wage gap in our RAIS samples of about 2ppts. to 3ppts. Under the assumption that workers receive wages that are proportional to

18

their productivity in this sector, the implication is that the average skill gap between nonwhites and whites in the restaurant industry is about 2ppts. to 3ppts.

A concern with our normalizing assumption is that, even in the restaurant sector, there may be positive pay premiums for white workers that contribute to their higher wage (e.g., if guests prefer being served by whites). As a robustness check, we therefore evaluate the implications of choosing alternative normalizations such that differential establishment-specific wage premiums account for about 50% or 100% (i.e., 1.5 ppts. or 3 ppts.) of the observed wage gap between whites and nonwhites in the restaurant sector.[31] This amounts to adjusting all the establishment premiums for whites upward (and all the person effects for whites downward) by 1.5 ppts. or 3 ppts. We also replicate our results by using a normalization based on other sectors in which firms have likely little or no rent to share (e.g., auto repair services).

3   RAIS Samples and Specification Tests

We now describe our main RAIS samples and present specification tests supporting the plausibility of the "exogenous mobility" condition for OLS to yield interpretable estimates of establishment wage premiums, before we move to the estimation results.

RAIS Samples

We use longitudinal wage observations for private-sector employees in the RAIS data set to estimate our two-way fixed effects models. Columns 1-4 of Table 4 first show the characteristics of the four samples of workers in the Southeast region (one for each race-gender group) that meet the selection criteria laid out in Section 1 (without imposing any restriction related to connected sets). We have about 44 million person-year observations over our 13-year period for about 9 million white men, with samples about 70%, 45%, and 25% as big for white women, non-white men, and non-white women, respectively. The age distributions of the four groups are similar, while education varies more, with the highest levels of schooling among white women and the lowest among non-white men. Mean log wages are about 10% higher than in the PNAD samples described in Table 1, but the differences between groups are similar. Nearly all workers are employed full time, with about 185 hours per month ($\approx$ 43 hours per week) among women and just slightly more among men.[32]

The fourth subpanel shows the mean establishment size and the mean fractions of female and white employees in their workplace for the four groups. Weighted by the

---

[31]To be conservative, we use a gap of 3ppts. in the restaurant sector for both males and females.

[32]The most common contracts in Brazil specify a 44-hour or 40-hour workweek.

number of worker-year observations, mean establishment sizes are relatively large, and even larger for women than men and for nonwhites than whites.[33] The extent of segregation across establishments by race and gender is evident in the differences in exposure rates of the four groups to white and female colleagues. The mean fraction of white employees at a white worker's establishment is about 30 ppts. higher than at a non-white worker's (of the same gender), while the mean fraction of females at a female's workplace is about 40 ppts. higher than at a male's (of the same race).[34]

As was pointed out by AKM, the establishment effects in a two-way fixed effects model are only identifiable within "connected sets" of workplaces that are linked by worker mobility. Columns 5-8 in Table 4 present similar descriptive statistics for the subsamples of workers in each race-gender group who work in the largest connected set of establishments for that group. The largest connected set includes 97% of person-year observations for white men, 95% for non-white men, 95% for white women, and 90% for non-white women. Mean wages are 1-2% higher for observations in the largest connected sets, but other characteristics remain very similar.

The decomposition in equation (5) implicitly assumes that each establishment has both white and non-white workers, so that one can calculate race-specific pay premiums at each establishment. In reality there are many small establishments that hire only white (or less often, only non-white) workers, even in the largest connected set for each race-gender group. Columns 9-12 in Table 4 thus present the descriptive statistics for those workers employed at establishments in the *dual-connected* sets for their gender (i.e., in the connected sets for both white and non-white workers of the same gender). These are the samples used for column 8 in Table 3. Among males, the dual connected sets include about 91% of the person-year observations for nonwhites, but only 81% of the observations for whites, reflecting the higher share of all-white establishments.[35] Among females, the corresponding rates are 86% of the person-year observations for nonwhites and 71% of the observations for whites.

Narrowing the samples to workers at the dual-connected establishments has little impact on the average age or education of the workers in the sample, but it leads to an increase in average wages of about 5 ppts. for white men and women, and about 2 ppts. for non-white men and women. This differential effect arises because establishments that have only one race group tend to pay relatively low wages, and

---

[33]The finding that women work in larger establishments than men is also true in the U.S. (Papps, 2012) and the U.K. (Mumford and Smith, 2008).

[34]The difference in females' versus males' exposure to female coworkers in the RAIS data is slightly smaller than in Portugal (Card, Cardoso, and Kline, 2016), but similar to that in the U.K. (Mumford and Smith, 2008). Data for relatively large establishments in the U.S. show less segregation by race or gender than in RAIS (Hellerstein, Neumark, and McInerney, 2008).

[35]The shares of all-white establishments are higher because of the larger sample sizes for whites.

more of such establishments are present in the connected sets for white workers.

Some Specification Tests for Exogenous Mobility

A concern with any conclusion based on two-way fixed effects models is that OLS estimates of the firm wage premiums will be biased unless worker mobility is uncorrelated with the time-varying residual components of wages. Card, Heining, and Kline (2013) developed an event-study analysis of the wage changes experienced by workers moving between different groups of firms to assess the plausibility of this "exogenous mobility" assumption. Specifically, they proposed grouping establishments by the average pay of coworkers, and tracking the changes in wages for workers who move up and down the "job ladder" with rungs defined by quartiles of co-worker pay.

Figure 2 shows the results of this analysis using our four race-gender groups in RAIS. The samples are restricted to individuals who switch workplaces and are observed in two consecutive years at both the origin and destination establishments. Workplaces are grouped into coworker pay quartiles using wages of all coworkers (i.e., both races and both genders) in the year of hiring (for destination establishments) or separation (for origin establishments). For clarity, only the wage profiles of workers who move from jobs in quartile 1 and quartile 4 are shown in the figures.

The figures exhibit clear step-like patterns for all four race-gender groups: when workers move to higher-wage establishments, their wages tend to rise, but they tend to fall when workers move to lower-wage establishments. There is little evidence of differential trends before or after a move for workers who move up or down the job ladder, but there are clearly permanent differences in wages prior to a move that are correlated with the direction of the move. For example, workers who start at a 4th quartile establishment and move to another 4th quartile establishment have substantially higher wages in the two years prior to the move than those who start at a 4th quartile establishment and move down. Such differential mobility on the basis of the *permanent* component of wages is fully consistent with the exogenous mobility assumption, since the AKM model conditions on a worker fixed effect.

As discussed above, a stark prediction of an AKM model with exogenous mobility is that wage changes associated with movements up the job ladder should be equal and opposite to wage changes for corresponding movements down the ladder. Figure 2 suggests that this is the case in our data, but Figure 3 presents more systematic evidence in support of this symmetry prediction. We use the same sample of movers, but we group origin and destination firms in 20 quantiles of co-worker wages. For each of the 20×20 pairs of quantiles, we then plot the mean change in log wages for the movers in the year after vs. before the move against the mean difference in

log wages of co-workers at the destination vs. origin establishment.[36] For each of the four race-gender groups, the symmetry prediction seems to hold such that a linear fit captures the relationship between changes in own wages and co-worker wages closely. Interestingly, one can also see that the slope is flatter for nonwhites than for whites, indicating that nonwhites may benefit less from moving up the job ladder.

The results in Figures 2 and 3 suggest that a simple AKM model estimated by OLS will provide interpretable estimates of the wage premiums offered at different establishments for different race-gender groups. We discuss additional diagnostic evidence based on the residuals from the estimated models in the next section.

4   Estimation Results

In this section, we present the results from estimating the two-way fixed effects model in equation (1) by race-gender group, using the largest connected sets described in columns 5-8 in Table 4.[37] We also provide evidence of strong assortative matching between workers and establishments for each group, which is robust to the well-known bias in the covariance of person and establishment effects with AKM models.

Estimation Results and Model Fit

Table 5 summarizes the estimation results. We show the standard deviations of the person effects, of the establishment effects, and of the covariate index $X'_{git}\widehat{\beta}_g$, as well as the correlation of the worker and establishment effects, the adjusted R-squared of the models, and the implied variance decomposition based on equation (3). For reference, we also show the fit statistics for a more general "job match" model that includes a separate dummy for each worker-establishment match.

In general, the two-way fixed effects models fit well, with adjusted R-squared statistics of around 90%. Nevertheless, RMSE's (root-mean-squared-errors) of the models are around 15% higher than those of the corresponding job match model. A comparison of residual variances between these models allow us to calculate the variance of the job match effects, i.e., the common component of ε_{git} across all observations of a given worker at a given workplace. As shown in the table, the job match component is relatively small, accounting for just 3-4% of the overall variance

---

[36]Movers' wage changes are adjusted for trends based on coefficients from a regression estimated on the sample of stayers, workers who remain at the origin establishments. The model includes the same education dummies as in Table 2 and a quadratic in age fully interacted with these dummies.

[37]The covariates X_{git} include year dummies interacted with the same five education dummies as in Table 2, and quadratic and cubic terms in age interacted with the education dummies.

of wages. This small magnitude means there is only limited scope for job match effects to drive mobility patterns and invalidate the exogenous mobility assumption.

The variance shares show that fixed worker characteristics account for 50-60% of the variance of wages in the Southeast region, with a larger share among females than males, while the establishment effects account for 21-25% of the variance.[38] The covariance between these two sets of effects is positive and accounts for another 7-18% of the variance of wages. These variance shares are similar to those reported by Card, Heining, and Kline (2013) for Germany, and by Lavetti and Schmutte (2016) and Alvarez et al. (2018) for Brazil, based on broadly similar RAIS samples.

We present additional evidence on the goodness of fit of the AKM model for our four groups in Figure C5 in the Appendix. We show the mean residuals for each of 100 cells, formed by assigning workers and establishments into 10 equally-sized bins based on their corresponding estimated effects. The mean residuals in each cell are close to zero, with the exception of cells representing workers with low person effects employed at workplaces with low establishment effects, where the mean residuals are positive. This pattern is most pronounced for non-white females, and is consistent with upward pressure from the minimum wage that is particularly important for low-skilled workers at low-paying establishments. We evaluate the sensitivity of our decomposition results to these observations in our robustness checks in Section 6.

Assortative Matching

The results in Table 5 reveal that workers with higher earnings capacity at any establishment (as represented by their person effects) are more likely to work at establishments that pay higher wage premiums. This holds for all race-gender groups, but the correlations between the worker and establishment effects are smaller for non-whites than for whites, suggesting that there may be differences in the propensities of high-premium establishments to hire whites versus nonwhites.

As is well known in the literature, these estimated correlations have to be interpreted carefully because the sampling errors in the estimated worker and establishment effects are negatively correlated, leading to a downward bias (Maré and Hyslop, 2006; Andrews et al., 2008). The magnitude of the expected bias is larger for "thin networks" (Kline, Saggio, and Sølvsten, 2018), a problem that is likely more severe for nonwhites than whites in our samples. For example, there are 6.6 white males but only 5.1 non-white males per establishment in our largest connected sets, suggesting

---

[38]Table D6 in the Appendix presents the corresponding estimates using observations for the whole country. The results are quite similar: the correlation between the establishment effects estimated in each sample for Southeast establishments is close to 1 (see Figure C4 in the Appendix).

that there are likely to be fewer network links between establishments for nonwhites.

It is possible to derive a corrected correlation (Kline, Saggio, and Sølvsten, 2018), but it is more straightforward to use standard methods to correct the partial regression coefficient relating worker effects to establishment effects. Consider a descriptive regression fit over person-year observations in a given race-gender group g:

$$\widehat{\alpha}_{gi} = \lambda_{0g} + \lambda_{1g}\widehat{\psi}^g_{J(g,i,t)} + \xi_{git}. \tag{11}$$

The coefficient λ_{1g} gives the expected change in the person effect per unit increase in the estimated establishment effect, and provides a convenient metric for assessing the degree of assortative matching. Since the sampling errors in the person and establishment effects are negatively correlated, we expect OLS estimates of λ_{1g} to be negatively biased. The estimates of ψ^g_j for *other* race-gender groups are estimated on separate samples, however, and are therefore uncorrelated with the estimated person effects for a particular group. Assuming that establishment effects for different groups are correlated with each other, we can use the estimated establishment effects for another group as instrumental variables, yielding corrected estimates of λ_{1g}.

We implement this procedure in Table 6, restricting attention to person and establishment effects in the dual-connected sets for each gender, and using the establishment effects for the same gender but opposite race group as instruments.[39] For reference, the first row of the table shows the unadjusted correlations of the worker and establishment effects. Next, we present OLS and IV estimates of the λ_{1g} coefficients. We show estimates from two models: one with no other controls and one that controls for micro-region fixed effects (and therefore control for the availability of different subgroups of workers in each local labor market). Finally, we present the first stage coefficients, which are relatively large and show very strong correlations between the estimated establishment effects for whites and nonwhites of each gender, as would be expected if the wage-setting model given by equation (9) is correct.

Consistent with the patterns for the simple correlation coefficients, the OLS estimates of λ_{1g} are only about half as big for non-white men as white men and one-third as big for non-white women as white women. The IV estimates are uniformly larger but still show less assortative matching for nonwhites. Specifically, the IV estimate of λ_{1g} is about 20% lower for non-white men than white men, and 15% lower for non-white women than white women.

Figure 4 provides additional evidence on the degree of assortative matching. We show the fractions of workers employed at establishments in each quartile of the

---

[39]We do not re-estimate the AKM models, we simply estimate OLS and IV versions of equation (11) using the subsets of person-year observations in the dual-connected sets.

distribution of estimated pay premiums, separately for workers in five education categories: incomplete elementary school, and complete elementary school (including those with incomplete middle school), middle school (including those with incomplete high school), high school (including those with incomplete college), or college. For all race-gender groups, the college-educated subgroup is most likely to work at high-premium establishments: 47%-59% of college-educated workers are employed at quartile 4 establishments, compared to only 10%-19% of workers with only elementary schooling. Moreover, for both genders, there appears to be more assortative matching for whites than for nonwhites, consistent with the findings in Table 6.[40]

These results point to two main conclusions. First, there is strong positive assortative matching between workers and establishments for all four race-gender groups. On average, establishments that pay higher wage premiums hire workers with higher permanent components of wages and higher education. The IV estimates in Table 6 suggest that an establishment that pays a 10% higher wage premium has employees whose average earnings capacity is 5-7% higher. Second, the strength of the assortative matching appears to be *lower* for nonwhites than whites of either gender. This gap suggests that race matters in the determination of employment probabilities, even controlling for workers' skills.

5   Decomposition results

We now use the results from the estimated two-way fixed effects models summarized in Table 5 to measure the effects of employment and wage-setting policies on the racial wage gap. We begin by decomposing the racial wage gap into worker-specific components and a component attributable to establishment pay premiums. We also relate these components to results from a standard Mincerian model. We then decompose the contribution of establishments into a relative wage-setting effect and a sorting effect, and further decompose the latter into a skill-based sorting and a residual sorting effect. Finally, we investigate how the wage losses associated with residual sorting and differential wage setting vary across the skill distribution.

Decomposing the Racial Wage Gap into Person and Establishment Effects

As discussed in Section 2, an initial step is to normalize the establishment effects. This allows us to decompose the wages of any individual – or group – into a com-

---

[40]For instance, the difference in the share of workers with a college degree vs. with only elementary schooling employed in quartile 4 establishments reaches 46 ppts. and 41 ppts. for white men and women, respectively, but only 36 ppts. and 33 ppts. for non-white men and women, respectively.

ponent due to their person effect and time varying characteristics, and a component attributable to the premiums paid by their employer. As a baseline, we assume that establishments in the restaurant sector pay zero wage premiums to either race-gender group, on average. Under this assumption, the normalized employer effects represent wage differences relative to jobs where each worker is paid according to his or her productivity (i.e., with no monopsonistic markdown or rent premium). Figure 5 displays the distribution of implied average pay premiums by 3-digit sector for white workers. The estimated sector premiums for white males range from near zero – thus near the restaurant sector, which is the 9th in rank order – for sectors such as delivery services (-0.07), auto repair services (-0.03), and footwear manufacturing (-0.01) to around 0.9 for sectors such as auto manufacturing (0.85) and petroleum extraction (0.90). The ranking is similar for females; the rank correlation with male estimates is 0.93. Interestingly, the high- and low-premium sectors correspond fairly closely to the high- and low-wage sectors identified by Krueger and Summers (1988).[41]

Table 7 presents results from implementing the decomposition of the average racial pay gap based on equation (4) using individuals in the dual-connected set of each gender in the Southeast region, with and without reweighting to correct for differences in the geographic distribution of whites and nonwhites (see Section 2). We also present results for subsets of workers with vs. without a high school degree.[42]

Without reweighting, differences in the mean person effects account for about 70% of the white-nonwhite wage gap for both males and females, while differences in the establishment effects account for 30-35%. Differences in the covariate index account for a negligible share of the male wage gap, but actually widen the female gap slightly, particularly for the higher-educated subgroup. This arises because the experience profiles for higher-educated non-white women are somewhat flatter than the profiles for whites (this can be seen in Figure C3 in the Appendix). Adjusting women's wages to an age-35 basis, as we do, thus raises the white-nonwhite gap in the estimated person effects, with an offsetting negative gap in the covariate indexes.

Reweighting for the different locations of whites and nonwhites reduces the racial wage gaps in the first column of Table 7 by 3-4 ppts. A majority of the reduction comes from a reduction in the size of the establishment component, which is consistent with the idea that area-based wage differentials will be incorporated in the establishment premiums and that whites are more likely to live in high-wage areas. Nevertheless, the reweighting also leads to some reduction in the person-effects component (with the exception of females with completed high school), implying that

---

[41]Table D7 in the Appendix shows the rank correlation between our ranking of sectors and theirs.

[42]We do not re-estimate the AKM models, we simply implement the decomposition in equation (4) in subsets of the dual-connected sets restricted by education levels.

there are relatively more high-earning whites than nonwhites in local labor markets with more high-premium firms (this is shown in Figure C6 in the Appendix).[43] Overall, the wage gaps remain sizable, however, particularly for higher-educated workers for whom the wage gaps are 5 to 6 times larger than for lower-educated workers.

Focusing on the location-adjusted results, our estimates imply that the 18.6 ppt. average racial wage gap for males is attributable to a 14.4 ppt. gap in permanent earnings capacity and a 4.4 ppt. gap in average pay premiums. Similar comparisons for females show that the 26.1 ppt. overall racial wage gap arises from a 21.6 ppt. gap in permanent earnings capacity, a 7.1 ppt. gap in average pay premiums, and a -2.6 ppt. gap in experience-related factors. Thus, the white-nonwhite difference in mean establishment premiums accounts for about one quarter of the overall wage gap for both genders (the same holds true for the higher-educated subgroup).

This decomposition relies on the assumption that workers of both race groups are paid their true productivity in the restaurant industry. If one assumes instead that the 3 ppt. wage gap in that sector (see Section 2) is entirely due to wage premiums for whites, then the component attributed to differences in earnings capacity falls mechanically to 11.4 ppts. for men and 18.6 ppts. for women, and the component attributed to differences in average employer pay premiums rises to 7.4 ppts. for men and 10.1 ppts. for women, or about 40% of the overall racial wage gap.

Relating our Decomposition Results to a Standard Mincerian Model

A question that naturally arises with the decomposition in Table 7 is how it relates to the usual Mincerian decomposition of the racial wage gap into an "explained" component (due to education and experience) and an "unexplained" component. We address this in Appendix Table D8, which presents the joint distribution of the Mincerian components of the wage gap and the AKM-based components (due to person effects, establishment effects, and covariates).

This exercise points to three main conclusions. First, a majority of the unexplained wage gap from a Mincer specification is attributable to differences in average person effects between whites and nonwhites (about 80% for males and 65% for females). In other words, differences in person effects explain a fairly large share of the unexplained residual from a simple cross-sectional wage model, though 20-35% is attributed to differences in firm effects. Second, about 60% of the racial gap in person effects from an AKM decomposition is due to differences in education and experience. The remaining 40% may be attributed to factors like school quality or

---

[43]The figure shows the correlation between the fractions of high-skilled white workers and high-paying jobs across micro-regions. Figure C7 displays the distributions of estimated person effects.

parental inputs that are *unobserved* in a conventional Mincer model.

Third, and most interestingly, differences in education and experience explain an important share of the racial gap in the AKM establishments effects (around 60% of the gap for workers of all education levels, and 45% for those with a high-school degree). Such a pattern is expected given the degree of sorting of higher-educated workers to higher-paying establishments shown in Figure 4. Indeed, if employers paid similar premiums to whites and nonwhites, and education and age were the only factors that affected the probability of being employed at a higher-wage premium workplace, then education and experience would explain the entire difference in the average establishment premiums between whites and nonwhites. The fact that there is a residual component for the establishment effects motivates our next analyses.

Decomposing the Effect of Employment and Wage-Setting Policies

We decompose the gap in average establishment premiums into a between-firm sorting effect and a within-firm relative wage-setting effect in Table 8, using the framework of equation (5). We show the overall wage gap (column 1), the mean establishment effects for whites and nonwhites (columns 2 and 3, respectively), the mean difference in establishment effects (column 4), and the two components of equation (5): the sorting and relative pay-setting effects (columns 5 and 6, respectively).

As noted in the discussion of Table 7, the average gap in establishment effects between whites and nonwhites accounts for about one quarter of the overall wage gap for both men and women (see column 4). The entries in column 5 imply that most of the establishment effect is attributable to the under-representation of nonwhites in higher-premium workplaces. Evaluated using the wage premiums earned by whites at different workplaces, the differential sorting of white and non-white workers accounts for about 20% of the overall wage gap for all groups but lower-educated males, for whom the average gap in establishment effects is small to begin with (1 ppt.).

Relative to the size of the sorting effects, the wage-setting effects in column 6 are modest in size, on the order of 1 ppt. for men and 1.5 ppts. for women. Some insight into this finding is provided by equation (10) and the pattern in Figure 6, which shows simple bin-scatters of the relationship between the estimated pay premiums for whites and nonwhites (separately by gender). For both gender groups, we find that nonwhite pay premiums are strongly correlated with white pay premiums, and that an empirical relationship of the form $\psi_j^N = \gamma \psi_j^W$ is highly plausible. To estimate the slope parameter γ while accounting for estimation errors in the white premiums, we use the premiums for white women as instruments for the premiums for white men (and vice versa), using the fact that establishment effects for different genders are

correlated with each other. This approach leads to estimates of $\gamma = 0.94$ for males and $\gamma = 0.9$ for females. Given the magnitudes of the average premiums earned by non-white men (0.175) and women (0.085), equation (10) predicts pay-setting effects that are close to the estimates in Table 8 (particularly for men), providing some empirical support for the monopsonistic pay-setting model in Card et al. (2018).

As noted earlier, the value of the pay-setting effects – but not of the sorting effects – depends on the normalization of the establishment effects. If, for example, we assume that the 3 ppt. white-nonwhite gap in the restaurant sector is attributable to a difference in pay premiums received by whites rather than a difference in productivity, we would then increase the relative pay-setting effect by 3 ppts., leading to an effect of about the same size as the sorting effect (and implying that differences in establishment premiums explain about 40% of the racial wage gap for both genders).

Decomposing the Sorting Effect into Skill-Based and Residual Sorting

Our final step is to decompose the sorting effect into a skill-based component – due to assortative race-neutral matching – and a residual component, using equations (7) and (8), respectively. As discussed in Section 3, we form a counterfactual racial composition for each establishment by calculating the expected fraction of nonwhites if the establishment selected randomly in the pool of suitable workers in their local labor market. Specifically, we divide workers of each gender into 16 bins defined by four age categories (25-27, 28-36, 37-45, and 46-54) and four quartiles of the overall distribution of person effects (combining whites and nonwhites).[44] Next, we calculate the fraction of employees at each establishment in each skill bin, and the nonwhite share of each skill bin in its local labor market (micro-region). We then combine these to calculate the expected fractions of whites and nonwhites at the establishment, which we then use to calculate the counterfactual establishment shares π_{Wj}^* and π_{Nj}^*, and the counterfactual skill-based sorting effect given by equation (7).[45]

Figure 7 presents our results graphically (given the negligible contribution of establishments for lower-educated workers shown in Tables 7 and 8, we simplify our graphs – here and in subsequent figures – by showing results for all workers and those with at least a high school education). In each panel, the black line displays the actual share of nonwhites by decile of the establishment-effect distribution. The red and blue lines in each panel represent counterfactual shares under two scenarios.

---

[44]We obtain similar results using finer skill categories, e.g., using octiles rather than quartiles of the distribution of estimated person effects (see Table D9 in the Appendix).

[45]Define N_{kj} the number of workers at establishment j in skill bin k, and p_{Wk} the share of white workers in skill bin k in the local labor market. The counterfactual numbers of white and non-white workers at establishment j are simply: $\sum_k N_{kj} \cdot p_{Wk}$ and $\sum_k N_{kj} \cdot (1 - p_{Wk})$, respectively.

First, we assume that each establishment maintains its existing age structure but selects workers at random within age categories (i.e., without regard for race or skill) from its local labor market. This gives the "naive" counterfactual shown by the red line. Second, we assume that each establishment maintains its joint distribution of age and skill but selects workers at random within age-skill categories (thus, without regard for race), yielding the "full" counterfactual shown by the blue line.

The top panels combine workers of all education levels. We see that the actual shares of non-white workers are relatively constant across the lower deciles of the establishment effect distribution, but then decrease sharply, falling from about 0.35 to 0.25 for men and a little more for women. In contrast, the share of nonwhites predicted by our naive benchmark remains nearly constant, as we move across the firm effect deciles, implying that age differences are unimportant in understanding the differential sorting of whites and nonwhites to high- and low-premium establishments.

Unlike the pattern for the naive counterfactual, the predicted non-white shares under our full counterfactual exhibit a downward-sloping pattern across the deciles, reflecting the racial gap in person effects and the tendency for higher-premium work-places to hire workers with higher person effects. The blue lines fall between the red and black lines, suggesting that skill-based employment policies explain some, but not all, of the under-representation of nonwhites at higher-premium workplaces.

The bottom panels of Figure 7 show results for men and women with at least a high school education. For these workers, the share of nonwhites at establishments in the top two deciles of the premium distribution is about 6 ppts. and 8 ppts. lower than would be expected under the naive counterfactual for men and women, respectively. Moreover, skill-based sorting explains only about half of that difference.

These results are summarized quantitatively in columns 7 and 8 of Table 8. Skill-based sorting accounts for about two-thirds of the sorting effect for men, and a similar share for the subgroup with high school or more. These shares are slightly lower for women (about 55%-60%). The residual sorting effect – which includes any effect of discriminatory employment policies by higher-premium establishments – accounts for about 6% of the overall racial wage gap for men and 8% for women.

Variation in the Residual Sorting and Wage-Setting Effects by Skill

The results in Table 8 suggest that the impacts of race-based preferences in firms' pay-setting and employment policies are larger for higher-educated nonwhites. To give a fuller picture, we show how the magnitudes of the residual sorting and relative wage-setting effects vary across skill levels (as captured by the person effects).

Specifically, we first divide workers of both races into 10 decile groups based on

their estimated person effects (the deciles are defined pooling whites and nonwhites together, separately by gender). For workers in each decile group, we then calculate the mean value of the estimated white establishment effects at workplaces of white workers (which corresponds to $\sum_j \widehat{\psi}_j^W \pi_{Wj}$ using employment shares for whites in that decile group) and at workplaces of non-white workers (which corresponds to $\sum_j \widehat{\psi}_j^W \pi_{Nj}$). The gap between the mean white establishment effects of the two groups represents the sorting effect for workers in that person-effects decile. Since we are conditioning on the values of the person effect, however, the difference has the interpretation of a *residual* sorting effect (it differs very slightly from the concept used in Table 8 because we do not also condition on age). Figure 8 shows the means for each decile group for all male and female workers (upper panels) and for those with at least a high school education (lower panels).

A concern with this simple exercise is that the estimation errors in the white establishment effects are negatively correlated with the errors in the estimated white person effects. This correlation will be magnified in the bottom and top deciles: whites classified as having the lowest person effects (resp. highest person effects) will on average have negative estimation errors (resp. positive estimation errors) in their person effects, implying that the associated average firm effects will be "too high" (resp. "too low"). This is clearly visible in the "dip" in the value of the mean white establishment effect for *white workers* between the first and second deciles. As expected, there is no such pattern in the mean white establishment effects for *non-white workers*, since there is no correlation between the estimation errors in the establishment effects for whites and the person effects for nonwhites.

With that caveat in mind, these is a clear tendency for higher-skilled workers of both race groups to work at establishments with higher pay premiums. There is also a systematic divergence between the mean establishment effects for whites and non-whites, implying that higher-skilled nonwhites are systematically under-represented at the best-paying workplaces, even conditional on their ability. The difference in the top decile reaches 4.3 ppts. for men and 4.9 ppts. for women, and given the sampling error issue, these are likely *under-estimates* of the true residual sorting effect.

Figure 9 presents a parallel analysis for the relative pay-setting effect. Here we plot the mean of the estimated white establishment effects and the mean of the estimated nonwhite establishment effects at workplaces of non-white workers (i.e., $\sum_j \widehat{\psi}_j^W \pi_{Nj}$ and $\sum_j \widehat{\psi}_j^N \pi_{Nj}$, respectively) in each decile. The gap between these two means represents the relative pay-setting effect for workers in that person-effects decile. Apart from the dip from the first to second deciles in the non-white establishment effects, which is likely due to the sampling error issue discussed above (accordingly, there is little or no dip in the means of the white establishment effects),

there is a strong tendency for establishment effects to rise across the skill distribution. Moreover, the difference between the mean white and nonwhite establishment effects at workplaces of non-white workers widens across the deciles. For example, the gap amounts to a 4.7 ppt. and 6.4 ppt. wage penalty for male workers in the 8th and 9th skill deciles, respectively. The gaps are even larger in the 10th decile, but these are likely *over-estimated* by a negative correlation between the estimation error in the nonwhite establishment effects and the error in the estimated nonwhite person effects among nonwhites in the top decile.

To show that our main conclusions are robust to the estimation error issue, we reproduce Figures 8 and 9 in the Appendix (see Figures C8 and C9, respectively) but using the estimated establishment effects *of the other gender.* For instance, in the top-left panel of Figure C8, we plot the mean value of the estimated establishment effects for *white females* at workplaces of *white males* and at workplaces of *non-white males* for each decile of the *male* person-effects distribution. The establishment effects for different genders are correlated with each other (a fact that we used for our IV strategy in Figure 6), but there is no correlation between the estimation error in the female establishment effects and the error in the estimated male person effects (and vice-versa). Accordingly, the differential "dips" from the first to the second deciles in Figures 8 and 9 disappear. As in Figure 8 and 9, in contrast, the residual sorting and relative pay-setting effects (now evaluated using establishment effects of the other gender) still widen systematically across the person-effects deciles.[46]

In summary, we infer that the costs of unequal employment and pay setting policies is disproportionately born by highly-skilled nonwhites. These results corroborate the work of scholars who highlight the "elitist" nature of racial discrimination in the Brazilian labor market (e.g., Campante, Crespo, and Leite, 2004), and suggest that the allocative costs of race-based preferences may be relatively large in Brazil.

6   Robustness

The results presented so far are based on a series of choices about sample and specification. We focus on workers in the Southeast region; use the modal race for people with a changing racial classification over time; impose a specific normalization for the establishment effects; and pool all the data for the available sample period. In this section, we show how our decomposition results vary as we vary these choices.

For ease of comparison, we summarize our findings graphically in Figure 10,

---

[46]The specific point estimates in Figures C8 and C9 have to be interpreted with caution because the establishment effects for different genders are unlikely to be perfectly correlated with each other.

with separate panels for workers of all education levels and for the higher-educated subgroups – for whom all the effects that we estimated are systematically larger.[47] For each alternative sample or specification choice, the height of the stacked column represents the mean white-nonwhite wage gap and its four components represent the gaps attributed to skill-based sorting, residual sorting, relative wage-setting, and differences in person effects and covariates, respectively. For reference, the first column in each panel reproduces our baseline results from Table 8.

We begin by extending our analysis to the whole country. To do this, we re-estimated our AKM models pooling RAIS data from all Brazilian regions, and then repeated our decompositions with the new set of worker and establishment effects.[48] As shown in the second column of each panel, the overall wage gap is about 3 ppts. smaller for males and 4 ppts. smaller for females in the national sample. The relative shares of the gap attributed to the various components, however, are typically within 1 ppt. of the corresponding shares in the Southeast sample.

We also fitted a separate set of AKM models, and repeated our decompositions, using only observations from the Northeast region. In this case, the wage gaps are smaller, and the sorting and relative pay-setting effects are scaled down accordingly. We find that skill-based sorting is relatively less important in the Northeast than in our baseline sample (or the national sample), while residual sorting is somewhat more important. Overall, however, the relative share of the wage gap accounted for by establishment-specific pay premiums is remarkably stable across regions.

As noted in Section 1, some individuals in RAIS are classified as white in some years and as nonwhite (black or mixed race) in others. For our baseline results, we resolved this issue by assigning each individual his or her modal race group. As a simple alternative, we classified each individual using the first race group they were ever assigned to in RAIS. We then re-estimated our AKM models and performed a new set of decompositions. In so doing, the overall racial wage gap increases a bit for men but remains very similar for women. The relative shares of the various components are again comparable to those in our baseline results, although the pay-setting effects are slightly smaller and the residual sorting effects slightly larger.

A more aggressive approach is to drop any worker classified in more than one (binary) race group in RAIS, which reduces our sample sizes by about 37%, with a relatively larger loss of nonwhites than whites.[49] Importantly, it removes an even

---

[47]All the values displayed in those figures are reported in Tables D10 and D11 in the Appendix.

[48]As mentioned earlier, the estimated effects based on this sample for the establishments located in the Southeast are very close to those obtained using observations for the Southeast region only.

[49]The percentage of person-year observations lost when imposing this restriction is 31% for white males, 48% for non-white males, 24% for white females, and 44% for non-white females.

larger share of between-establishment movers, since nearly all changes in race occur with a change in employer. The average wage gaps in this "consistent race" sample are about 10 ppts. larger than in our baseline sample, reflecting a slight increase in the mean wage of nonwhites and a larger rise in the mean wage of whites. Nevertheless, the decomposition shares are not too different from those in our baseline results. Most notably, the share of the wage gap attributed to the pay-setting effect is larger with this sample, rising to 9% and 7% of the overall wage gap – vs. 5% and 6% in our baseline results – for men and women, respectively.

In discussing the fit of our two-way fixed effects models, we noted that the models tend to underestimate wages of workers with low estimated person effects who are employed at establishments with low wage premiums. To evaluate the sensitivity of our results to these observations, we took our baseline samples and excluded observations for workers with estimated person effects in the bottom decile of the person effect distribution, as well as for workers at establishments in the bottom decile of the estimated wage premium distribution. We then repeated our decompositions finding that these exclusions have only negligible effects on our baseline results.

Next, we show the impact of attributing 50% or 100% of the racial wage gap in the restaurant sector to higher premiums earned by whites, rather than to a difference in productivity, as in our baseline normalization. As mentioned above, this raises the wage premiums for whites mechanically by 1.5 ppts. or 3 ppts., and the relative wage-setting effects by the same amounts. In contrast, by construction, the sorting effect remains identical. Its composition is affected, however, because these alternative normalizations also lower the person effects of whites mechanically by 1.5 ppts. or 3 ppts. Yet, the impact on the skill distribution is limited, such that the reduction in the share of the sorting effect attributed to skill-based sorting is minimal.

We also show the impact of using other sectors, in which firms have likely little or no rent to share, for the normalization of the establishment effects. Specifically, we assumed that wage premiums of both races were zero in two alternative sectors with low average establishment pay premiums, namely the auto repair sector for males and the clothing and apparel sector (i.e., firms that sew clothing) for females. These choices have the advantage of sidestepping concerns that some restaurant customers may prefer to be served by whites, possibly leading to differential wage premiums in that sector. Importantly, the "unexplained" racial wage gap in those sectors is nearly identical to the gap in the restaurant sector, so these choices lead to similar normalizations and only neglibible changes relative to our baseline results.

Finally, we investigate how our results vary with possible variation in the incentive of employers to engage in discriminatory employment or wage-setting policies. First, as mentioned earlier, legal or social sanctions against discrimination in the Brazilian

labor market may have evolved over time. To evaluate this possibility, we divided our sample into two (overlapping) 8-year periods, from 2002 to 2009 and from 2007 to 2014.[50] We then re-estimated our AKM models separately by sub-period and repeated our decompositions. Interestingly, we see a small decline in the overall racial wage gap between the periods, with a reduction in the magnitudes of the residual sorting and relative pay-setting effects, the two components that are most likely to reflect discriminatory practices. Second, employers may have less incentives to discriminate against nonwhites in industries where interactions between employees and customers are less frequent. Accordingly, we see that the residual sorting and relative pay-setting effects are larger in high "face-time" sectors (commerce, hospitality, financial services and insurance, real estate) than in low "face-time" sectors (extractive and transformation industries, utilities, construction).[51]

7   Conclusions

This paper measured the contribution of firms' employment and wage-setting policies to the white-nonwhite pay gap in Brazil. It showed that firms exacerbate racial inequalities in general skills in three ways. First, the strong assortativeness in the matching between workers and establishments means that nonwhites are less likely to be employed in high-premium workplaces, even in the absence of any discriminatory practices, an effect that accounts for about half of the contribution of firms to the racial wage gap. Yet, non-white workers also tend to be sorted into lower-premium establishments, and tend to receive lower pay premiums in the establishments that they are sorted into, compared to white workers of similar skill levels. The associated wage losses are particularly severe for nonwhites at the top of the skill distribution.

The results of this paper relate to active policy debates that are taking place across Latin American countries, where racial differences in education levels are persistent and nonwhites remain under-represented in high-paying industries and occupations. Plaintiffs have used the disproportionately high share of white employees in several industries as evidence of (potentially "unconscious") discrimination, as in the high profile case filed in 2009 by the Union of Bank Employees of Brasilia against Itau Unibanco. In this case, Brazil's highest labor court agreed that the disparities were alarming, but ruled against the plaintiff as it found no direct evidence of irregularities in the hiring and promotion practices of the bank. Interestingly, we show in Tables D14 and D15 in the Appendix that our estimates of the "unexplained"

---

[50]Tables D12 and D13 in the Appendix present the statistics in Table 4 for the two sub-periods.
[51]We simply implement our decompositions separately for the two groups of sectors.

under-representation of nonwhites is largest in the "banking and finance" sector.

The findings of this paper naturally raise the question of how policies and management practices could help narrow the racial pay gap. The large white-nonwhite skill gap, combined with the strong assortative matching, highlights the importance of investments towards narrowing the educational gap. Yet, it is important to underline that the skill gap that we estimate is not necessarily determined prior to workers entering the labor force. Differential mentoring and on-the-job training opportunities could also lower the skills that workers bring to any job, or even undermine the impact of educational investments. An interesting agenda for future research is to go beyond the "static" decomposition in this paper and examine the dynamic process through which workers end up with higher person and firm effects.

In that respect, it would be interesting to study the impact of affirmative action policies that exist in several countries. Despite the recent adoption of racial quotas for public-sector jobs, Brazil does not require private-sector employers to take race into account in their recruitment process. For instance, in contrast to the US, Brazil does not require government contractors to make efforts to employ disadvantaged groups at rates proportional to their shares in the (qualified) local labor market. Miller (2017) shows that such policies can have powerful and lasting impacts on the racial composition of "treated" firms. Nevertheless, it is yet to be seen what the effects of similar policies would imply for the overall quality of the matching of nonwhites in the labor market, and whether these effects would generalize to a Latin American context. There is also a "softer" but increasing social pressure in many countries on large companies to improve their standing on racial equality. For instance, the Ethos Institute in cooperation with the Inter-American Development Bank releases periodically the so-called *Social, Racial, and Gender Profile of the 500 Largest Brazilian Companies*, which analyzes the workforce of these companies to reveal possible ethno-racial inequalities (among others) and reports on best employment practices and affirmative action programs already in place in those corporations. It remains unclear, however, to which extent this growing movement actually affects firms' employment and wage-setting policies towards non-white workers.

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., Paul Lengermann, and Kevin L. McKinney. 2003. "The Measurement of Human Capital in the U.S. Economy." Tech. Rep. Technical Paper No. TP-2002-09, Washington DC: US Census Bureau, Center for Economic Studies.

Altonji, Joseph and Rebecca Blank. 1999. "Race and Gender in the Labor Market." In *Handbook of Labor Economics, Vol. IIIc*, edited by Orley Ashenfelter and David Card. Amsterdam: Elsevier.

Alvarez, Jorge, Felipe Benguria, Niklas Engbom, and Christian Moser. 2018. "Firms and the decline in earnings inequality in Brazil." *American Economic Journal: Macroeconomics* 10 (1):149–89.

Andrews, George Reid. 1992. "Racial Inequality in Brazil and the United States: A Statistical Comparison." *Journal of Social History* 26(2):229–263.

Andrews, Martyn J., Leonard Gill, Thorsten Schank, and Richard Upward. 2008. "High Wage Workers and Low Wage Firms: Negative Assortative Matching or Limited Mobility Bias?" *Journal of the Royal Statistical Society: Series A* 171(3):673–697.

Arcard, Jean-Louis and Beatrice d'Hombres. 2004. "Racial Discrimination in the Brazilian Labour Market: Wage, Employment and Segregation Effects." *Journal of International Development* 16:1053–1066.

Arceo-Gomez, Eva and Raymundo Campos-Vasquez. 2014. "Race and Marriage in the Labor Market: A discrimination correspondence audit study in a developing country." *American Economic Review* 104(5):376–380.

Ashenfelter, Orley. 1972. "Racial Discrimination and Trade Unionism." *Journal of Political Economy* 8(3):435–64.

Aslund, Olof, Lena Hensvik, and Oskar Nordstrom Skans. 2014. "Seeking Similarity: How Immigrants and Natives Manage in the Labor Market." *Industrial and Labor Relations Review* 32(3):405–441.

Aslund, Olof and Oskar Nordstrom Skans. 2010. "Will I See You at Work? Ethnic Workplace Segregation in Sweden." *Industrial and Labor Relations Review* 63(3):471–493.

Babcock, Linda and Sarah Laschever. 2003. *Women Don't Ask: Negotiation and the Gender Divide.* Princeton, NJ: Princeton University Press.

Bailey, Stanley R., Mara Loveman, and Jeronimo O. Muniz. 2013. "Measures of "Race" and the Analysis of Racial Inequality in Brazil." *Social Science Research* 42(1):106–119.

Barth, Erling and Harald Dale-Olsen. 2009. "Monopsonistic Discrimination, Worker Turnover, and the Gender Wage Gap." *Labour Economics* 16(5):589–597.

Bayer, Patrick and Kerwin Kofi Charles. 2018. "Divergent Paths: A New Perspective on Earnings Differences Between Black and White Men Since 1940." *Quarterly Journal of Economics* 133 (3):1459–1501.

Becker, Gary. 1957. *The Economics of Discrimination (second edition 1971).* Chicago: University of Chicago Press.

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.

Black, Dan A. 1995. "Discrimination in an Equilibrium Search Model." *Journal of Labor Economics* 13(2):309–334.

Blackaby, D.H., D.G. Leslie, and P.D. Murphy. 2002. "White-Ethnic Minority Earnings and Employment Differentials in Britain: Evidence from the LFS." *Oxford Economic Papers* 54(2):270–297.

Boal, William M. and Michael R. Ransom. 1997. "Monopsony in the the Labor Market." *Journal of Economic Literature* 35(1):86–112.

Campante, Filipe R., Anna R. V. Crespo, and Phillippe G. P. G. Leite. 2004. "Desigualdade Salarial entre Raças no Mercado de Trabalho Urbano Brasileiro: Aspectos Regionais." *Revista Brasileira de Economia* 58(2):185–210.

Campos, Walter de Oliveira. 2015. "Expectativas em torno da Lei Afonso Arinos (1951): A 'Nova Abolição' ou 'Lei Para Americano Ver'?" *Revista Latino-Americana de História* 4(13):257–278.

Cano-Urbina, Javier and Patrick L. Maso. 2016. "Acculturation and the Labor Market in Mexico." *IZA Journal of Labor Policy* (2016):5–21.

Card, David, Ana Rute Cardoso, Jorg Heining, and Patrick M. Kline. 2018. "Firms and Labor Market Inequality: Evidence and Some Theory." *Journal of Labor Economics* 36(Supplement):S13–S70.

Card, David, Ana Rute Cardoso, and Patrick M. Kline. 2016. "Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women." *Quarterly Journal of Economics* 131(2):633–686.

Card, David, Jorg Heining, and Patrick M. Kline. 2013. "Workplace Heterogeneity and the Rise of West German Wage Inequality." *Quarterly Journal of Economics* 128(3):967–1015.

Cavalieri, Claudia and Reynaldo Fernandes. 1998. "Diferenciais de Salários por Gênero e por Cor: Uma Comparação entre as Regiões Metropolitanas Brasileiras." *Revista de Economia Política* 18(1):158–175.

Chadarevian, Pedro C. 2011. "Para Medir as Desigualdades Raciais no Mercado de Trabalho." *Revista de Economia Política* 31(2):283–304.

Charles, Kerwin K. and Jonathan Guryan. 2008. "Prejudice and Wages: An Empirical Assessment of Becker's The Economics of Discrimination." *Journal of Political Economy* 116(5):773–809.

———. 2011. "Studying Discrimination: Fundamental Challenges and Recent Progress." *Annual Review of Economics* 3(1):479–511.

Cornwall, Christopher, Jason Rivera, and Ian M. Schmutte. 2017. "Wage Discrimination When Identity is Subjective: Evidence from Changes in Employer-Reported Race." *Journal of Human Resources* 52(3):719–755.

Derenoncourt, Ellora and Claire Montialoux. 2018. "Minimum Wages and Racial Inequality." *University of Michigan CREST Working Paper* .

Eeckhout, Jan and Philipp Kircher. 2011. "Identifying Sorting - In Theory." *Review of Economic Studies* 78(3):872–906.

———. 2018. "Assortative Matching with Large Firms." *Econometrica* 86(1):85–132.

Equal Rights Trust. 2009. *Brazil: Step-by-Step Guide: Race Discrimination in the Field of Employment.* London: Equal Rights Trust.

Esteve, Albert and Luis Ángel López-Ruiz. 2010. "Union Formation Implications of Race and Gender Gaps in Educational Attainment: The Case of Latin America." *Population Research and Policy Review* 29(5):609–637.

Ferreira, Francisco H.G., Sergio P. Firpo, and Julian Messina. 2014. "A More Level Playing Field? Explaining the Decline in Earnings Inequality in Brazil, 1995-2012." *International Research Initiative on Brazil and Africa (IRIBA) Working Paper* .

Francis, Andrew and Maria Tannuri-Painto. 2013. "Endogenous Race in Brazil: Affirmative Action and the Construction of Racial Identity Among Young Adults." *Economic Development and Cultural Change* 61(4):731–753.

———. 2015. "Inside the Black Box: Affirmative Action and the Social Construction of Race in Brazil." *Ethnic and Racial Studies* 38(15):2771–2790.

Fryer, Roland G. 2010. "Racial Inequality in the 21st Century: The Declining Significance of Race." In *Handbook of Labor Economics, Vol. IV*, edited by Orley Ashenfelter and David Card. Amsterdam: Elsevier, 885–971.

Garcia, Luana Marques, Hugo Nopo, and Paola Salardi. 2009. "Gender and Racial Wage Gaps in Brazil 1996-2006: Evidence Using a Matching Comparisons Approach." *Inter-American Development Bank Research Department Working Paper* 681.

Giuliano, Laura, Jonathan Leonard, and David I. Levine. 2009. "Manager Race and the Race of New Hires." *Journal of Labor Economics* 27(4):589–631.

———. 2011. "Racial Bias in the Manager-Employee Relationship: An Analysis of Quits, Dismissals, and Promotions at a Large Retail Firm." *Journal of Human Resources* 46(1):26–52.

Giuliano, Laura and Michael R. Ransom. 2013. "Manager Ethnicity and Employment Segregation." *Industrial and Labor Relations Review* 66(2):346–379.

Glitz, Albrecht. 2014. "Ethnic Segregation in Germany." *Labour Economics* 29(4):28–40.

Hellerstein, Judith and David Neumark. 2008. "Workplace Segregation in the United States: Race, Ethnicity, and Skill." *Review of Economics and Statistics* 90(3):459–477.

Hellerstein, Judith, David Neumark, and Melissa McInerney. 2008. "Changes in Workplace Segregation in the United States Between 1990 and 2000: Evidence from Matched Employer-Employee Data." In *The Analysis of Firms and Employees: Quantitative and Qualitative Approaches*, edited by Stefan Bender, Julia Lane, Kathryn Shaw, Fredrik Andersson, and Till von Wachter. Chicago: University of Chicago Press.

Henriques, Ricardo. 2001. "Desigualdade Racial no Brasil: Evolução das Condições de Vida na Década de 90." *Institute for Applied Economic Research (IPEA) Discussion Paper* 807.

Hirata, Guilherme and Rodrigo R. Soares. 2016. "Competition and the Racial Wage Gap: Testing Becker's Model of Employer Discrimination." *IZA Working Paper* 9764.

Kline, Patrick M., Raffaele Saggio, and Mikkel Sølvsten. 2018. "Leave-Out Estimation of Variance Components." *UC Berkeley Working Paper* .

Komastsu, Bruno Kawaoka and Naercio A. Menezes-Filho. 2016. "Does the Rise of the Minimum Wage Explain the Fall in Wage Inequality in Brazil?" *Instituto de Ensino e Pesquisa (INSPER) Working Paper* .

Krueger, Alan B. and Lawrence H. Summers. 1988. "Efficiency Wages and the Inter-Industry Wage Structure." *Econometrica* 56(2):259–293.

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.

Lavetti, Kurt and Ian M. Schmutte. 2016. "Estimating Compensating Wage Differentials with Endogenous Job Mobility." *Ohio State University Working Paper* .

Macis, Mario and Fabiano Schivardi. 2016. "Exports and Wages: Rent Sharing, Workforce Composition, or Returns to Skills?" *Journal of Labor Economics* 34(4):945–978.

Manning, Alan. 2003. *Monopsony in Motion: Imperfect Competition in Labor Markets*. Princeton, NJ: Princeton University Press.

———. 2011. "Imperfect Competition in the Labor Market." In *Handbook of Labor Economics, Vol. IV*, edited by Orley Ashenfelter and David Card. Amsterdam: Elsevier, 973–1041.

Maré, David C. and Dean R. Hyslop. 2006. "Worker-Firm Heterogeneity and Matching: An Analysis Using Worker and Firm Fixed Effects Estimated from LEED." *Statistics New Zealand LEED Research Paper* .

Mariano, Francisca Z., Edward M. Costa, Guimar aes Daniel B., and Daniel T. de Sousa. 2018. "Diferenciais de Rendimentos entre Raças e Gêneros, nas Regiões Metropolitanas, por Níveis Ocupacionais: Uma Análise Através do Pareamento de Nopo." *Estudos Econômicos* 48(1):137–173.

Marx, Anthony W. 1998. *Making Race and Nation: A Comparison of South Africa, the United States, and Brazil.* Cambridge: Cambridge University Press.

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

Melo, Frederico L. B. 2014. "The Minimum Wage Campaign in Brazil and the Fight Against Inequality." *Global Labor Journal* 6(3):283–301.

Miller, Conrad. 2017. "The Persistent Effect of Temporary Affirmative Action." *American Economic Journal: Applied Economics* 9(3):152–190.

Mumford, Karen and Peter N. Smith. 2008. "What Determines the Part-Time and Gender Earnings Gaps in Britain: Evidence from the Workplace." *Oxford Economic Papers* 61(Supplement):i56–i75.

Nopo, Hugo. 2012. *New Century, Old Disparities: Gender and Ethnic Earning Gaps in Latin America and the Caribbean.* Washington DC: The Inter-American Development Bank and The World Bank.

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

Oliveira, Lucia Elena Garcia, Rosa Maria Porcaro, and Tereza Cristina N. Araújo. 1981. *O Lugar do Negro na Força de Trabalho.* Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística (IBGE).

Papps, Kerry L. 2012. "Spillovers and Wage Determination Within Firms." *University of Bath Working Paper* .

Pendakur, Krishna and Ravi Pendakur. 1998. "The Colour of Money: Earnings Differentials Among Ethnic Groups in Canada." *Canadian Journal of Economics* 31(3):518–548.

———. 2002. "Colour my World: Have Earnings Gaps for Canadian-Born Ethnic Minorities Changed Over Time?" *Canadian Public Policy* 28(4):489–512.

Psacharopoulos, George and Harry A. Patrinos. 2002. "Returns to Investment in Education: A Further Update." *World Bank Policy Research Working Paper* 2881.

Reis, Mauricio Cortez and Ana Risi Vianna Crespo. 2015. "Race Discrimination in Brazil: An Analysis of the Age, Period and Cohort Effects." *Institute for Applied Economic Research (IPEA) Discussion Paper* 1114.

Robinson, Joan. 1933. *The Economics of Imperfect Competition*. London: Macmillan.

Silva, Nelson do V. 1978. *Black-White Income Differentials in Brazil*. University of Michigan PhD Dissertation. Ann Arbor: University of Michigan.

———. 1980. "O Preço da Cor: Diferenciais Raciais na Distribuição da Renda no Brasil." *Pesquisa e Planejamento Econômico* 10(1):21–44.

———. 1985. "Updating the Cost of Not Being White in Brazil." In *Races, Classes and Power in Brazil*, edited by P.M. Fontaine. Los Angeles: UCLA Press, 42–55.

Skidmore, Thomas E. 1974. *Black into White: Race and Nationality in Brazilian Thought*. Oxford: Oxford University Press.

———. 1992. "Fact and Myth: Discovering a Racial Problem in Brazil." *Kellogg Institute Working Paper* 173.

Soares, Sergei S. 2000. "O Perfil da Discriminação no Mercado de Trabalho – Homens Negros, Mulheres Brancas e Mulheres Negras." *Institute for Applied Economic Research (IPEA) Discussion Paper* 769.

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

Zschirnt, Eva and Didier Ruedin. 2016. "Ethnic Discrimination in Hiring Decisions: A Meta-Analysis of Correspondence Tests 1990-2015." *Journal of Ethnic and Migration Studies* January:1–19.

Figure 1: Log Hourly Wage Distributions Among Private-Sector Employees (Males)
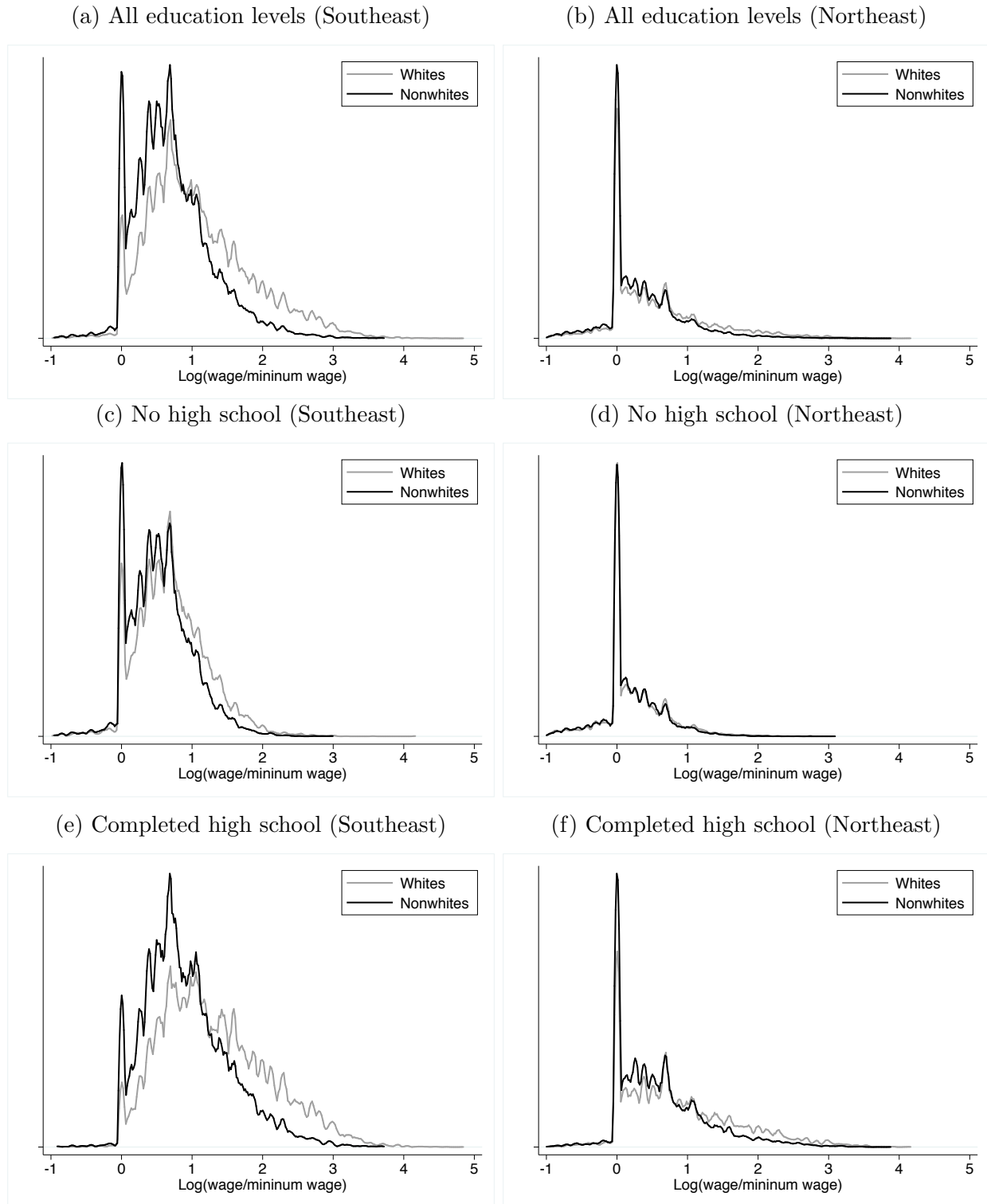
(a) All education levels (Southeast)

(b) All education levels (Northeast)

(c) No high school (Southeast)

(d) No high school (Northeast)

(e) Completed high school (Southeast)

(f) Completed high school (Northeast)

Notes: The figure displays kernel densities (Epanechnikov kernel with a 0.025 half-width) of the log wage-to-minimum-wage ratio for whites and nonwhites, based on PNAD 2002-2014 and constructed using survey weights (PNAD was not conducted in 2010). The samples include male non-farm private-sector employees (either formal or informal), age 25 to 54, with potential labor market experience of at least 1 year, and non-missing data on race, gender, education, wage, and hours worked. The left and right panels restrict the samples to the Southeast and the Northeast regions, respectively. The top panel pools all education levels together; the middle and bottom panels are restricted to workers with no high school degree and with at least a high school degree, respectively.

Figure 2: Event Studies Around Job Moves

(a) White males

(b) Non-white males

(c) White females

(d) Non-white females
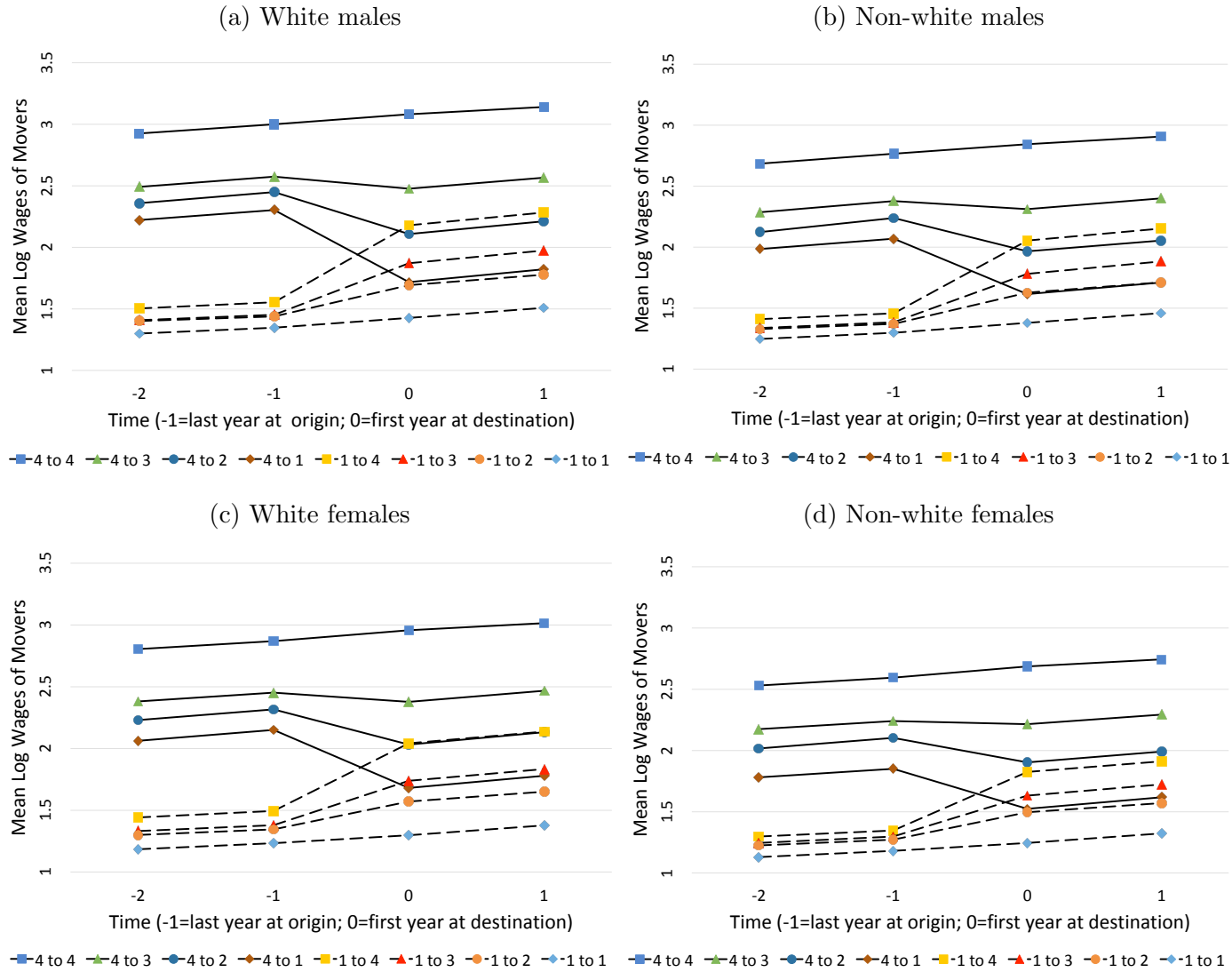
Notes: The figure displays the evolution of wages for workers in the Southeast region who moved from origin establishments in the top and bottom quartile groups to destination establishments in any of the other quartile groups. We use the samples described in columns (1)-(4) in Table 4; the movers are defined as workers at establishments employing at least one worker of each race-gender group, who separated from the origin establishment in 2003-2012, were reemployed in the destination establishment the next or the following year, and were employed at the origin and destination establishments for 2+ consecutive years. Origin/destination groups are based on quartiles of co-worker wages during the calendar year of separation/hiring.

Figure 3: Wage Changes of Movers vs. Changes of Co-Worker Wages

(a) White males

(b) Non-white males

(c) White females

(d) Non-white females
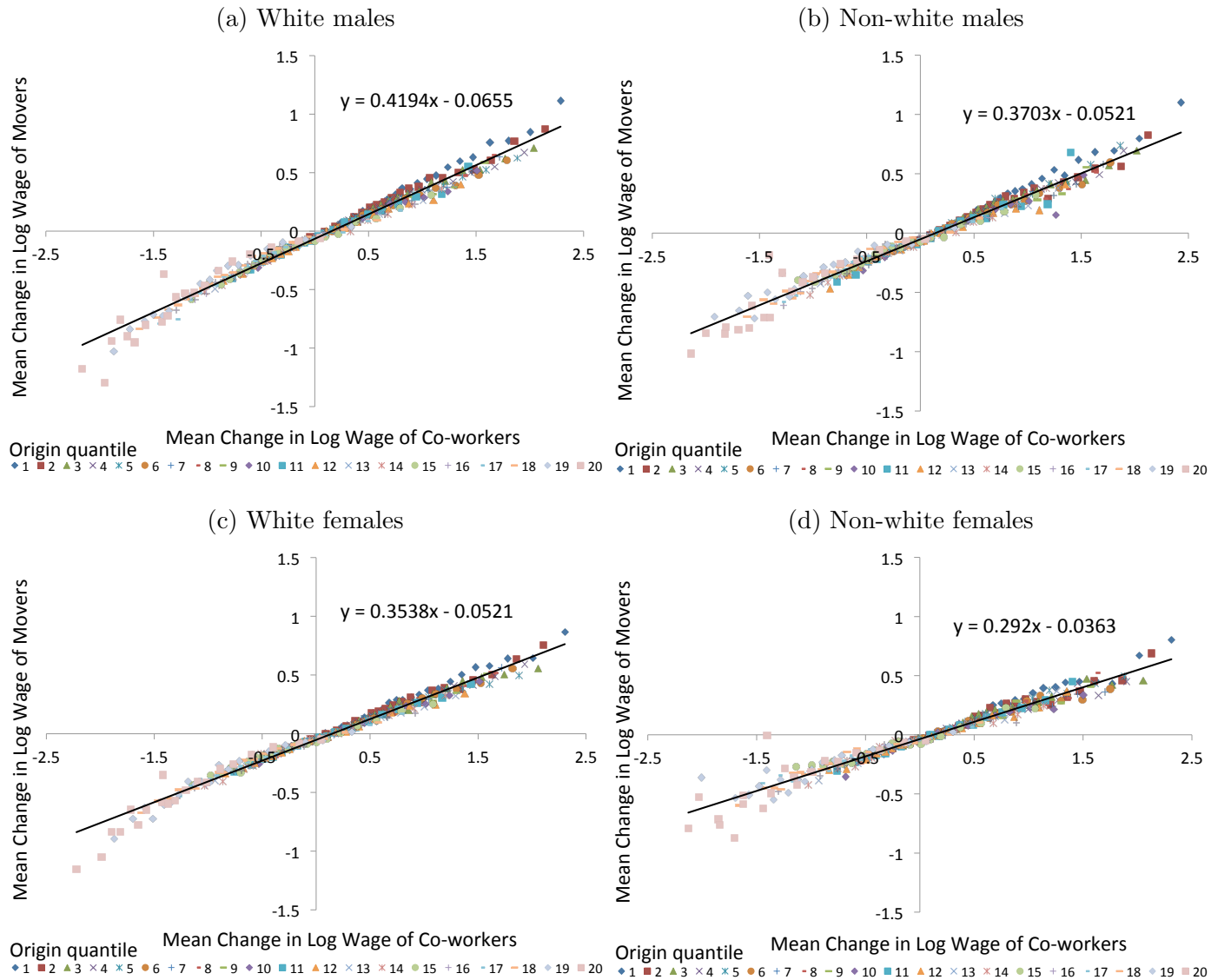
Notes: The figure plots the mean change in movers' wages between the years prior to separation and after hiring against the mean change in co-worker wages between origin and destination establishments. Origin and destination establishments are grouped by 20 quantiles of co-worker wages; each of the 20x20 dots corresponds to movers from/to given origin/destination quantiles. Movers are defined as in Figure 2; their wage changes are adjusted for trends based on coefficients from a regression estimated on the sample of stayers, workers who remain at the same origin establishments over the years around a move. The model includes the same education dummies as in Table 2 and a quadratic in age fully interacted with these dummies.

Figure 4: Evidence of Assortativeness in the Matching Between Workers and Establishments

(a) White males (b) Non-white males

(c) White females (d) Non-white females
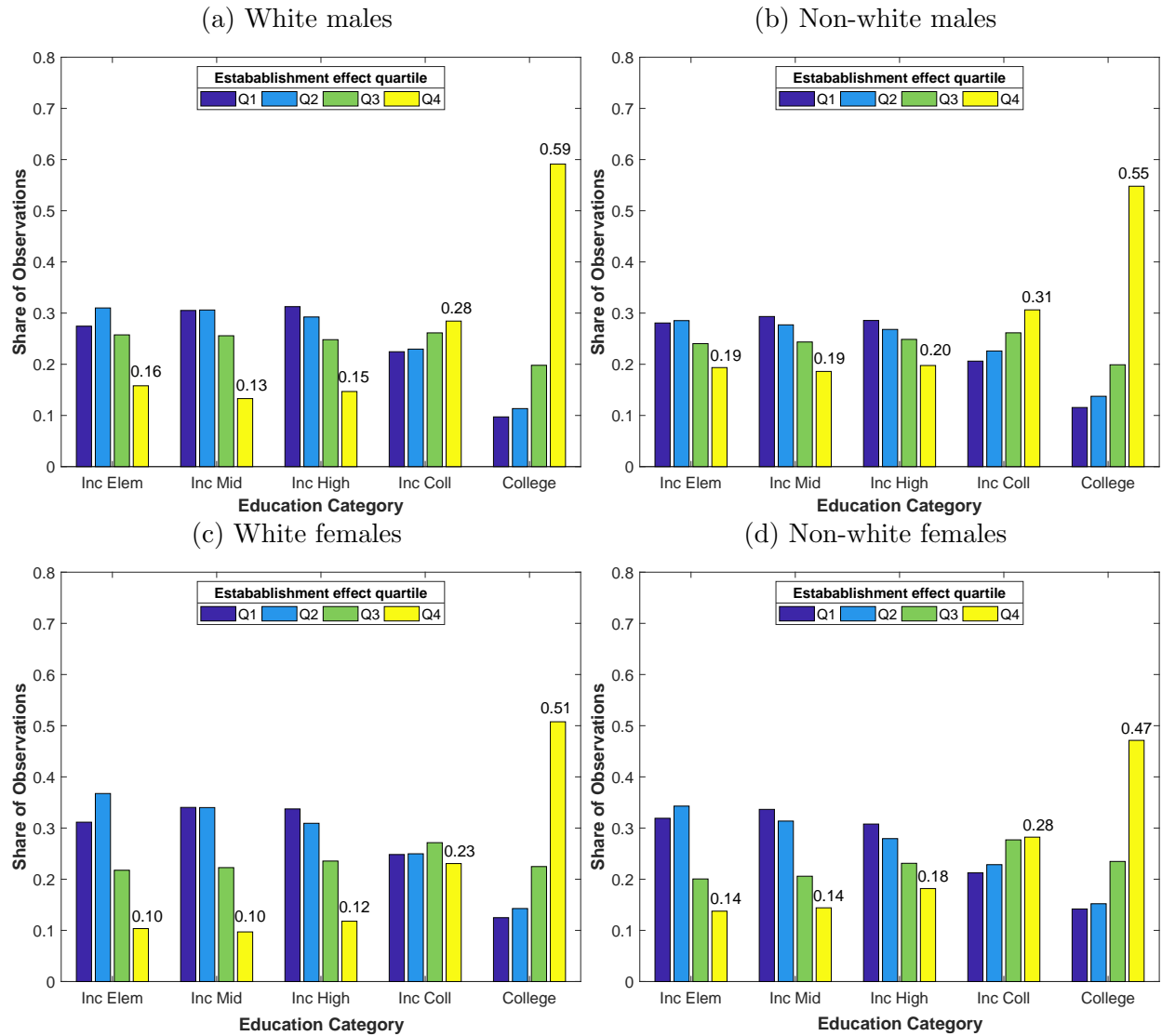
Notes: The figure displays the share of person-year observations across establishment effect quartiles by education categories, which correspond to the same education dummies as in Table 2: incomplete elementary school (*IncElem*), and complete elementary school (including those with incomplete middle school; *IncMid*), middle school (including those with incomplete high school; *IncHigh*), high school (including those with incomplete college; *IncColl*), or college (*College*). The samples are restricted to the largest connected set for each race-gender group in the Southeast region, described in columns (5)-(8) in Table 4.

Figure 5: Cumulative Distribution of Normalized Establishment Effect by Sector

(a) Males

(b) Females
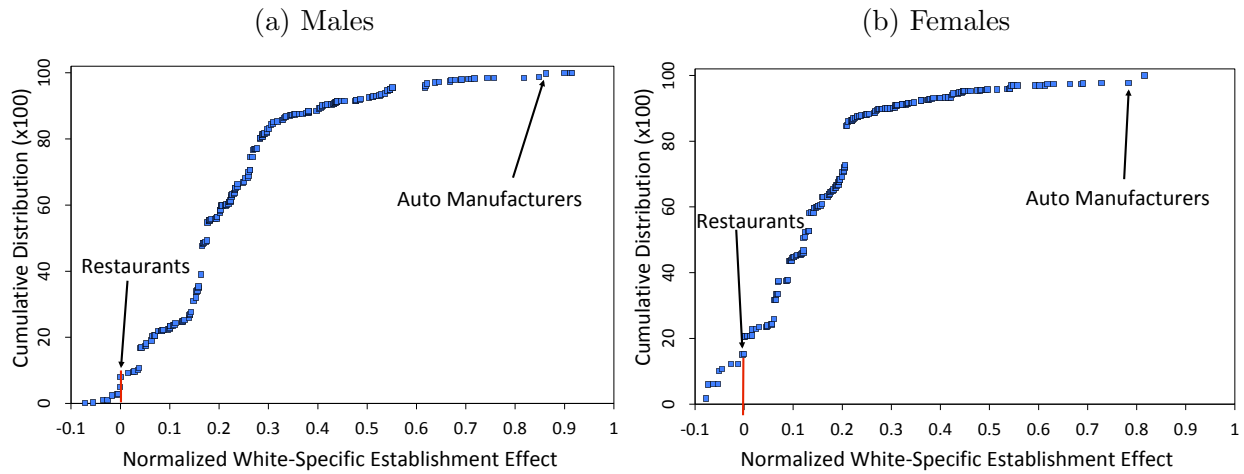
Notes: The figure displays the cumulative distribution of average white-specific establishment effects for three-digit industry sectors (CNAE), after normalizing these with respect to the average in the restaurant sector (all establishments effects are normalized in the same way in subsequent figures and tables). The averages are computed over person-year observations in the dual-connected set of each gender in the Southeast region, i.e., the set of establishments in the largest connected set for both whites and nonwhites of that gender, which is described in columns (9)-(12) in Table 4. The rank correlation between these sector-level averages for males and females is 0.93.

Figure 6: Correlation of Establishment Effects for Whites and Nonwhites

(a) Males

(b) Females
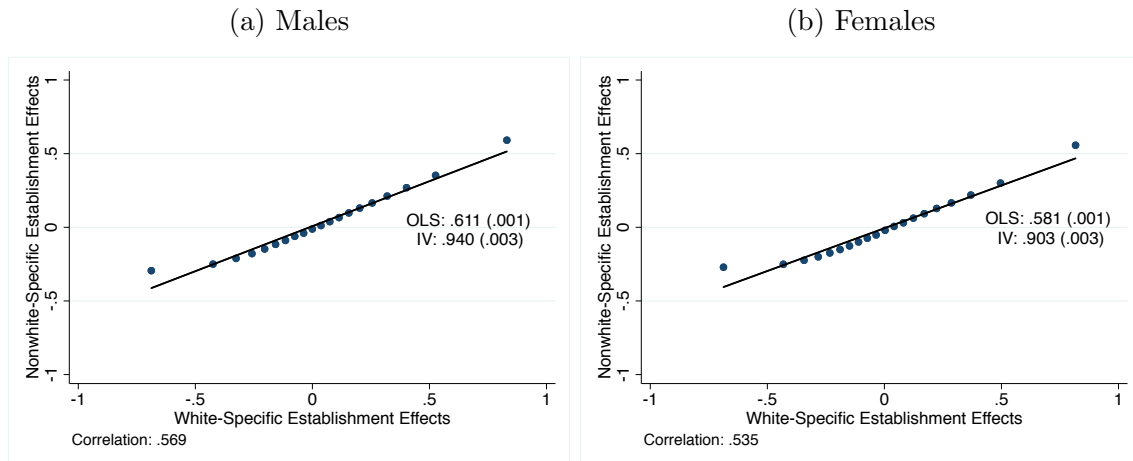
Notes: The figure displays binned scatterplots of white-specific and nonwhite-specific establishment effects from the Southeast region's dual connected set of each gender. Observations are unique establishments. Scatterplots use 20 equal-sized bins and plot the within-bin means. The slope of the OLS fit line (with its standard error in parentheses) is reported in the graph (the correlation coefficient of the variables are reported under the graph), as well as an IV estimate that accounts for estimation errors in the white premiums by using the premiums for white women as instruments for the premiums for white men (and vice versa; in that case, we use establishments in the tetra-connected set, i.e., the intersection of the dual-connected sets of both genders).

Figure 7: Skill-Based Assortative Matching by Establishment Effect Deciles

(a) Males (all education)

(b) Females (all education)

(c) Males (completed high school)

(d) Females (completed high school)
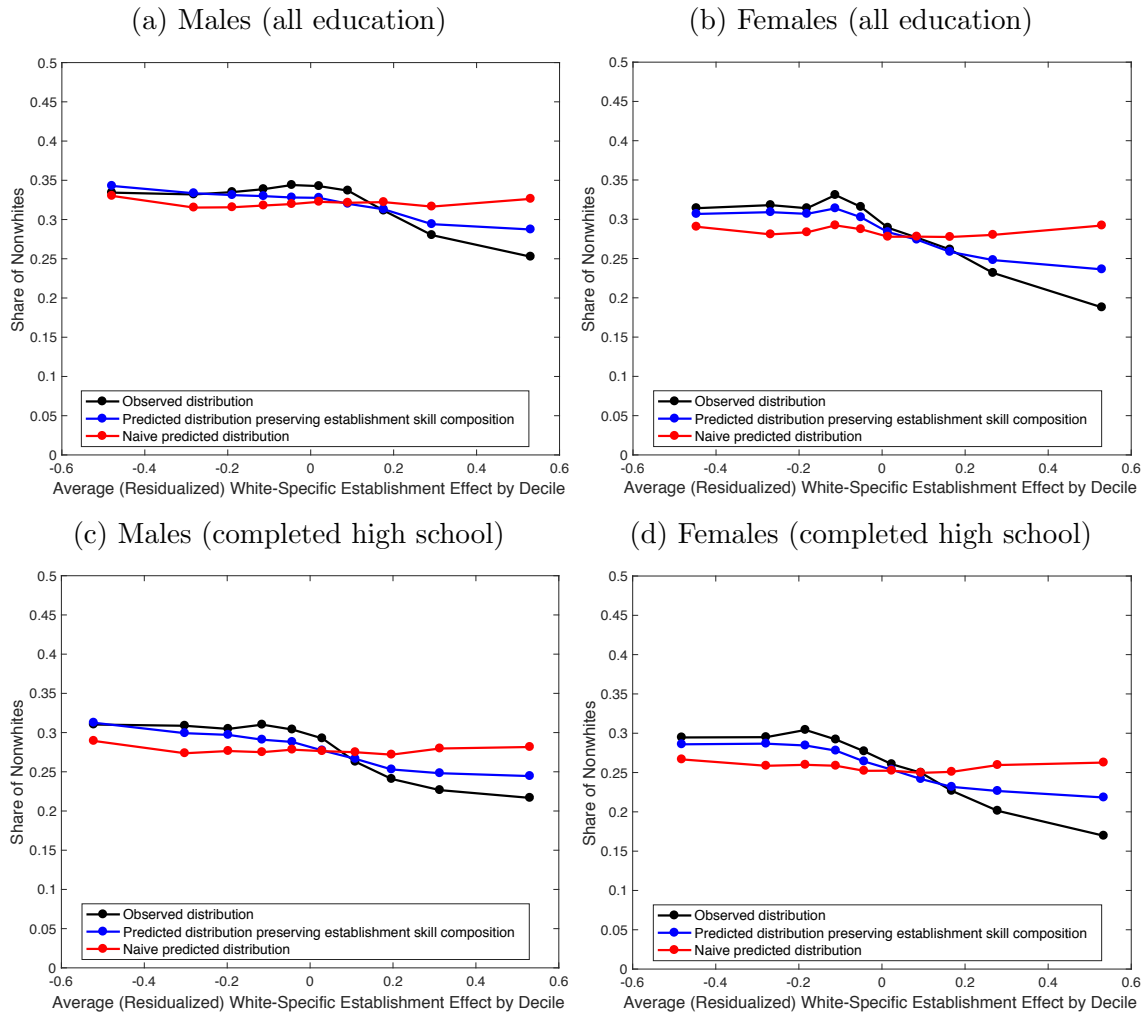
Notes: The figure displays the share of non-white person-year observations by deciles of the white-specific establishment effects (residualized of micro-region fixed effects). The naive and predicted distributions reshuffle person-year observations – regardless of race – across establishments within a micro-region. The naive distribution maintains the age distribution of establishments for each year. The predicted distribution maintains the age and skill distribution of establishments for each year. Samples are restricted to the dual-connected set of each gender (and education) group in the Southeast region. Nonwhites are reweighted so as to have the same distribution across micro-regions as whites (of the same gender).

Figure 8: Residual Sorting Effect by Person Effect Deciles
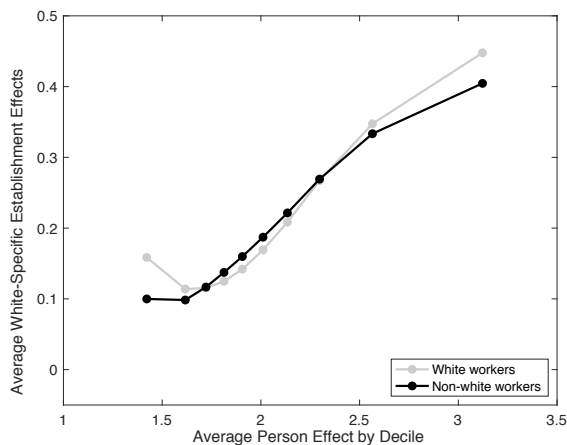
(a) Males (all education)

(b) Females (all education)
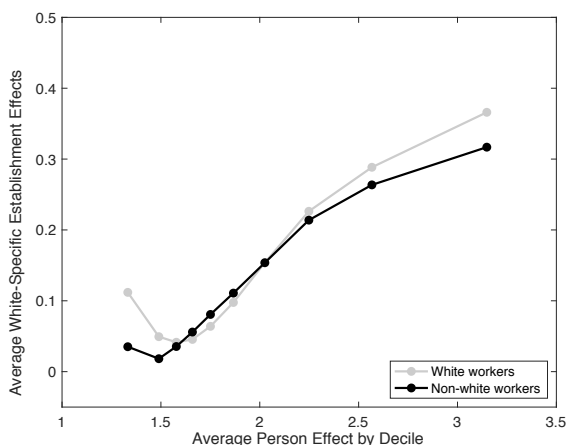
(c) Males (completed high school)
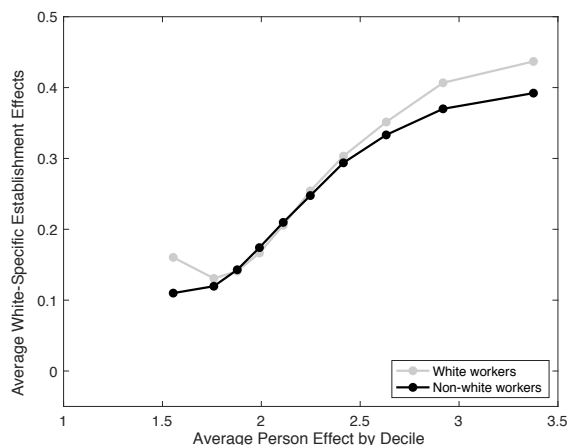
(d) Females (completed high school)
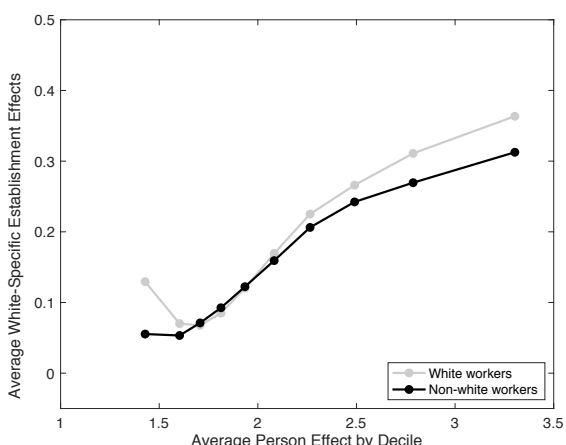
Notes: The figure displays the means of the estimated white-specific establishment effects at workplaces of whites and non-white workers in each decile of the person effect distribution. The gap between the mean white-specific establishment effects of the two groups represents the residual sorting effect for workers in that decile of person effects. We use person-year observations in the dual-connected set of each gender in the Southeast region (the person effect deciles are defined pooling whites and nonwhites together). Nonwhites are reweighted so as to have the same distribution across micro-regions as whites (of the same gender).

Figure 9: Relative Wage-Setting Effect by Person Effect Deciles
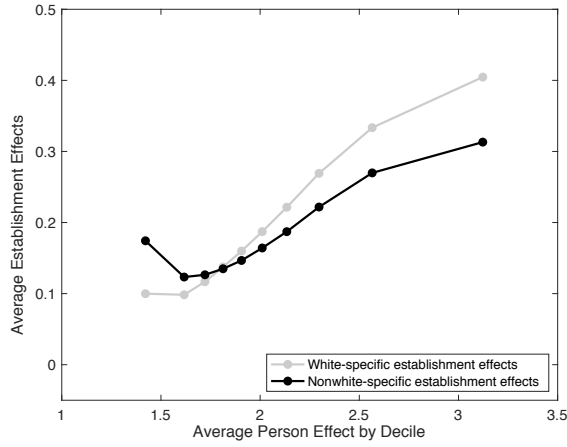
(a) Males (all education)

(b) Females (all education)
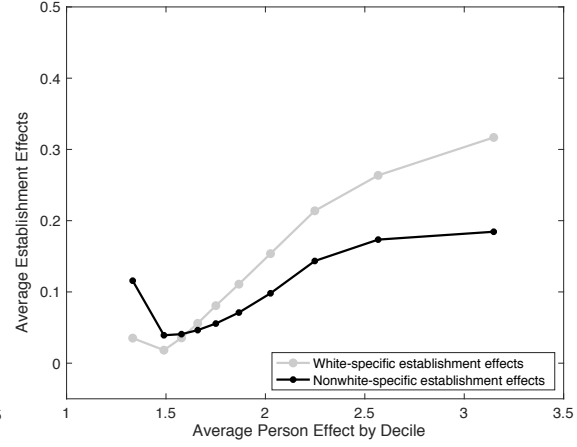
(c) Males (completed high school)
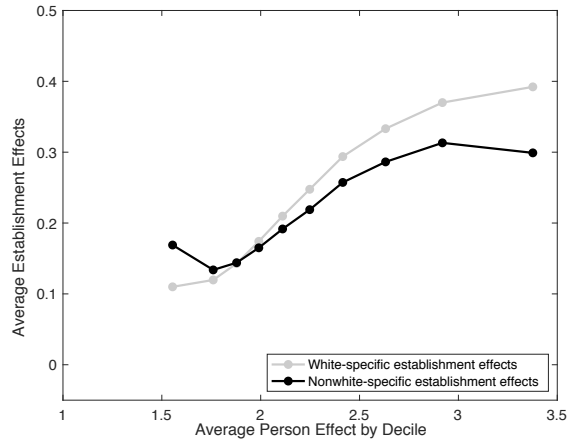
(d) Females (completed high school)
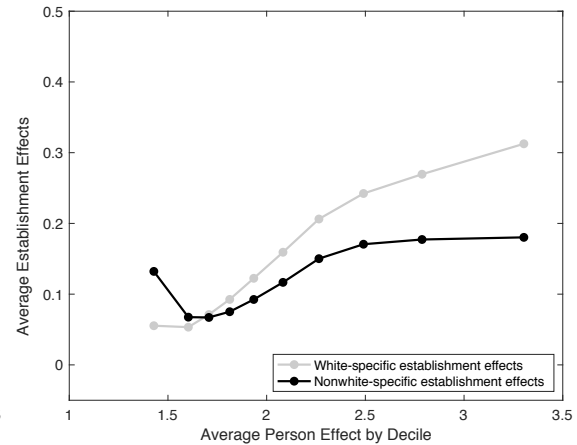
Notes: The figure displays the means of the estimated white-specific and nonwhite-specific establishment effects at workplaces of non-white workers in each decile of the person effect distribution. The gap between the mean white-specific and nonwhite-specific establishment effects represents the relative wage-setting effect for workers in that decile of person effects. We use person-year observations in the dual-connected set of each gender in the Southeast region (the person effect deciles are defined pooling whites and nonwhites together). Nonwhites are reweighted so as to have the same distribution across micro-regions as whites (of the same gender).

Figure 10: Robustness and Heterogeneity

(a) Males (all education)

■ Skill-based sorting ■ Residual sorting ■ Relative wage-setting ■ Person effect and covariates

(b) Females (all education)

■ Skill-based sorting ■ Residual sorting ■ Relative wage-setting ■ Person effect and covariates

(c) Males (completed high school)

■ Skill-based sorting ■ Residual sorting ■ Relative wage-setting ■ Person effect and covariates

(d) Females (completed high school)

■ Skill-based sorting ■ Residual sorting ■ Relative wage-setting ■ Person effect and covariates

Notes: The figure displays similar decomposition results as in Table 9 for a series of robustness checks and heterogeneity analyses (see text for details). In each case, the height of the stacked column represents the mean white-nonwhite wage gap and its four components represent the gap attributed to skill-based sorting, to residual sorting, to relative wage-setting, and to differences in person effects and covariates, respectively. For reference, the first column in each panel reproduces our baseline results from Table 9. All samples are restricted to the dual-connected set of each gender (and education) group. Nonwhites are reweighted so as to have the same distribution across micro-regions as whites (of the same gender).

Table 1: Characteristics of Private-Sector Employees by Race Group

	Brazil				Southeast region			
	All	White	Mixed race	Black	All	White	Mixed race	Black
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Males								
Share of sample in column race group	1.00	0.48	0.42	0.08	1.00	0.56	0.34	0.09
Share of column race group in private employment	0.41	0.43	0.39	0.45	0.48	0.47	0.48	0.51
Characteristics of private-sector employees in column race group:								
Mean years of education	8.44	9.24	7.59	7.76	8.90	9.60	7.99	8.00
Fraction with high school or more	0.45	0.53	0.37	0.38	0.49	0.56	0.40	0.39
Mean log hourly wage (R$)	1.61	1.78	1.42	1.46	1.75	1.89	1.57	1.55
Share with wage ≤ 2 minimum wages	0.54	0.44	0.66	0.66	0.46	0.38	0.56	0.59
Share in formal sector employment	0.79	0.82	0.75	0.78	0.83	0.84	0.81	0.82
B. Females								
Share of sample in column race group	1.00	0.51	0.41	0.08	1.00	0.59	0.32	0.08
Share of column race group in private employment	0.21	0.25	0.17	0.20	0.25	0.27	0.23	0.24
Characteristics of private-sector employees in column race group:								
Mean years of education	10.15	10.66	9.41	9.39	10.36	10.94	9.39	9.28
Fraction with high school or more	0.65	0.69	0.58	0.59	0.67	0.72	0.57	0.57
Mean log hourly wage (R$)	1.48	1.62	1.28	1.31	1.58	1.70	1.38	1.37
Share with wage ≤ 2 minimum wages	0.71	0.63	0.82	0.82	0.66	0.58	0.78	0.80
Share in formal sector employment	0.80	0.82	0.76	0.80	0.82	0.83	0.81	0.83

Notes: The table displays statistics based on PNAD 2002-2014 and constructed using survey weights (PNAD was not conducted in 2010). The samples include males (panel A) and females (panel B), age 25 to 54, with potential labor market experience of at least 1 year, and non-missing data on race, gender, and education. Private employment status includes nonfarm formal and informal private-sector employees, excluding those with missing wages or hours. Columns (1)-(4) present statistics for the whole country; columns (5)-(8) restrict the sample to the Southeast region only. All monetary values in the paper are deflated using the CPI to a 2010 base. In each case, the first column pools all race groups together; the other columns report statistics for each of the three main race groups, separately (the race groups in PNAD are white, mixed race, black, asian, and indigenous).

Table 2: Racial Differences in Log Wages in PNAD (Private-Sector Employees)

	Brazil				Southeast region			
	All		Completed high school		All		Completed high school	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Males								
Dummy if mixed race	-0.27	-0.11	-0.30	-0.16	-0.30	-0.11	-0.33	-0.16
(std err.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Dummy if black	-0.29	-0.13	-0.34	-0.19	-0.33	-0.14	-0.37	-0.21
(std err.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Year and state fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Education and experience	no	yes	no	yes	no	yes	no	yes
B. Females								
Dummy if mixed race	-0.28	-0.11	-0.28	-0.13	-0.31	-0.11	-0.31	-0.13
(std err.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Dummy if black	-0.30	-0.11	-0.32	-0.16	-0.33	-0.11	-0.35	-0.16
(std err.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Year and state fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Education and experience	no	yes	no	yes	no	yes	no	yes

Notes: The table displays the results of regressing log hourly wages on a series of race group dummies, using data from PNAD 2002-2014 (PNAD was not conducted in 2010). The samples include male (panel A) and female (panel B) private-sector nonfarm employees (either formal or informal), age 25 to 54, with potential labor market experience of at least 1 year, and non-missing data on race, gender, education, wage, and hours worked. All specifications include year and state fixed effects and use survey weights. Education and experience controls include five education dummies (incomplete elementary school, and complete elementary school, middle school, high school, or college) and a quadratic in potential experience. The omitted race group is white. Other race dummies not reported are indigenous and Asian. The samples in columns (1)-(4) use data for the whole country; columns (5)-(8) restrict the samples to the Southeast region only. In each case, the first two columns pool workers of all education levels; the other columns restrict the sample to workers with completed high school.

Table 3: Racial Differences in Formality and Log Wages in PNAD and RAIS (Private-Sector Employees)

	Brazil				Southeast region			
	Formality	Log hourly wage			Formality	Log hourly wage		
	PNAD	PNAD	PNAD-formal	RAIS	PNAD	PNAD	PNAD-formal	RAIS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Males - All education								
Dummy if nonwhite	0.00	-0.11	-0.11	-0.06	0.00	-0.12	-0.11	-0.07
(std err.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Mean for whites	0.82	1.78	1.85	1.94	0.84	1.89	1.94	2.03
B. Males - Completed high school								
Dummy if nonwhite	0.01	-0.16	-0.16	-0.11	0.01	-0.17	-0.16	-0.12
(std err.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Mean for whites	0.86	2.10	2.13	2.19	0.87	2.19	2.21	2.30
C. Females - All education								
Dummy if nonwhite	-0.01	-0.11	-0.10	-0.09	0.00	-0.11	-0.11	-0.10
(std err.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Mean for whites	0.82	1.62	1.67	1.80	0.83	1.70	1.75	1.88
D. Females - Completed high school								
Dummy if nonwhite	-0.01	-0.14	-0.13	-0.11	0.00	-0.14	-0.13	-0.13
(std err.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Mean for whites	0.85	1.82	1.85	1.99	0.86	1.89	1.92	2.09

Notes: The table displays the results of regressing the outcome on top of each column on a series of race group dummies, using data from PNAD 2002-2014 (PNAD was not conducted in 2010) or RAIS 2002-2014. The samples include male (panels A and B) and female (panels C and D) private-sector employees, age 25 to 54, with potential labor market experience of at least 1 year, tenure of at least 1 month, and non-missing data on race, gender, education, wage, and hours worked. All specifications include year and state effects, the same education dummies as in Table 2, and quadratic in potential experience (we use survey weights with the PNAD data). The omitted race category is white; nonwhite includes both black and mixed race. Other race dummies not reported are indigenous and Asian. The samples in columns (1)-(4) use data for the whole country; columns (5)-(8) restrict the samples to the Southeast region only. In each case, the first three columns use the PNAD samples. The outcome in the first column is a dummy for being formally employed. The specification in the second column is the same as in columns (2), (4), (6), and (8) in Table 2. The third column restricts the PNAD sample to formal employees. The specification in the fourth column is identical but uses the RAIS samples that we use for our decomposition results in the rest of the paper, namely the dual-connected set of each gender, that is described in columns (9)-(12) in Table 4.

Table 4: Descriptive Statistics Comparing the Analysis Samples, the Largest Connected Sets, and the Dual-Connected Sets

	Analysis samples (all valid observations)				Largest connected sets (by race-gender group)				Dual-connected sets (by gender)			
	White male	Non-white male	White female	Non-white female	White male	Non-white male	White female	Non-white female	White male	Non-white male	White female	Non-white female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Age												
Mean age	36.4	36.2	35.5	35.6	36.4	36.2	35.4	35.5	36.4	36.2	35.5	35.5
Share ≤ 30 years old	0.30	0.30	0.34	0.33	0.30	0.30	0.34	0.33	0.30	0.30	0.34	0.33
Share ≥ 50 years old	0.08	0.08	0.06	0.06	0.08	0.08	0.06	0.06	0.08	0.08	0.06	0.06
Education												
Mean years of schooling	9.4	8.5	10.7	9.7	9.5	8.5	10.7	9.7	9.5	8.5	10.8	9.7
Share completed high school	0.52	0.42	0.70	0.60	0.53	0.42	0.70	0.61	0.53	0.42	0.71	0.61
Share completed college	0.11	0.04	0.19	0.08	0.11	0.04	0.20	0.09	0.11	0.04	0.21	0.09
Wages and Hours												
Mean log hourly wage (R$)	1.96	1.77	1.80	1.56	1.98	1.79	1.82	1.57	2.03	1.80	1.88	1.59
(standard dev.)	(0.70)	(0.60)	(0.71)	(0.56)	(0.70)	(0.60)	(0.72)	(0.57)	(0.71)	(0.60)	(0.74)	(0.57)
Mean monthly hours	188.1	188.4	184.2	185.2	188.0	188.3	183.9	184.8	187.8	188.3	182.7	184.6
Share full-time	0.99	0.99	0.96	0.97	0.99	0.99	0.96	0.97	0.99	0.99	0.96	0.97
Establishment Characteristics												
Mean establishment size	478	601	498	708	494	634	524	783	586	657	688	820
Share females at establishment	0.24	0.22	0.65	0.64	0.24	0.22	0.64	0.64	0.24	0.22	0.63	0.63
Share whites at establishment	0.80	0.50	0.82	0.50	0.80	0.51	0.82	0.52	0.77	0.52	0.78	0.53
Sample Sizes												
No. person-year obs.	43,960,676	18,442,544	31,072,137	10,214,292	42,504,499	17,464,447	29,467,955	9,213,854	35,700,872	16,858,515	22,181,982	8,795,728
No. persons	9,066,023	4,064,366	6,943,395	2,547,634	8,643,741	3,783,424	6,443,519	2,224,382	7,677,443	3,717,140	5,258,580	2,169,911
No. establishments	1,640,702	971,550	1,575,138	788,035	1,314,220	736,179	1,186,607	520,940	597,178	597,178	413,653	413,653

Notes: The table displays descriptive statistics by race-gender group in three samples based on RAIS 2002-2014. The analysis samples in columns (1)-(4) include nonfarm private-sector formal employees in the Southeast region, age 25 to 54, with potential labor market experience of at least 1 year, tenure of at least 1 month, and non-missing data on race, gender, education, wage, and hours worked, who are employed on December 31st of each year with an open-ended contract in which they are paid on a monthly basis. The entire history of an individual is dropped if that individual reports earning a wage below the minimum wage or above the 99th percentile in a given year, as well as in the case of extreme wage changes between consecutive years. The analysis sample is restricted to the largest connected set for each race-gender group in columns (5)-(8), i.e., the largest set of establishments that are linked by worker mobility. The analysis sample is further restricted to the dual-connected set of each gender in columns (9)-(12), i.e., the set of establishments in the largest connected set for both whites and nonwhites of that gender. All statistics are calculated across person-year observations.

Table 5: Summary of Estimated Two-Way Fixed Effects Models by Race-Gender Group

	White male	Non-white male	White female	Non-white female
	(1)	(2)	(3)	(4)
Standard deviation of log wages	0.705	0.598	0.716	0.571
Summary of Parameter Estimates:				
Std. dev. of person effects (across person-yr obs.)	0.499	0.419	0.541	0.446
Std. dev. of estab. effects (across person-yr obs.)	0.331	0.297	0.331	0.283
Std. dev. of Xb (across person-yr obs.)	0.178	0.182	0.185	0.187
Correlation of person/estab. effects	0.273	0.153	0.255	0.086
RMSE of model	0.231	0.222	0.218	0.194
Adjusted R-squared of model	0.893	0.862	0.907	0.885
Comparison job-match effects model:				
RMSE of match-effects model	0.197	0.194	0.187	0.173
Adjusted R-squared of match-effects model	0.922	0.895	0.932	0.909
Variance of job match effect	0.017	0.014	0.014	0.009
Percent of variance of wages due to job match	3.4	4.0	2.7	2.7
Variance decomposition:				
Percent of variance of log wages due to:				
person effects	50.1	49.1	57.0	61.0
establishment effects	22.1	24.8	21.3	24.6
covariance of person and estab. effects	18.2	10.7	17.7	6.7
Xb and associated covariances	1.5	5.2	-2.9	-0.3
residual (including job match and time-varying)	8.2	10.2	6.9	8.1

Notes: The table summarizes the results from estimating two-way fixed effects models for the log hourly wage using person-year observations in the largest connected set for each race-gender group in the Southeast region. The samples are described in columns (5)-(8) in Table 4. The models include dummies for individual workers and individual establishments, year dummies interacted with the same five education dummies as in Table 2, and quadratic and cubic terms in age interacted with the education dummies. The "comparison" job-match effects models include dummies for each job match as well as the other covariates in the basic models.

Table 6: Evidence of Attenuation Bias in Correlation Between Worker and Establishment Effects

	White male	Non-white male	White female	Non-white female
	(1)	(2)	(3)	(4)
Correlation of person/establishment effects	0.273	0.153	0.255	0.086
Regression coefficient of person effect on establishment effect:				
OLS estimate	0.549	0.249	0.620	0.181
OLS estimate including micro-region fixed effects	0.521	0.214	0.598	0.154
IV estimate	0.672	0.523	0.746	0.638
IV estimate including micro-region fixed effects	0.660	0.520	0.756	0.678
Addendum: first stage coefficients, using establishment effect for opposite race (same gender) as instrument				
First stage coefficient	0.763	0.812	0.706	0.737
First stage coefficient including micro-region fixed effects	0.731	0.796	0.656	0.711

Notes: The table displays regression results providing evidence of attenuation bias in the correlation between the estimated worker and establishment effects in our two-way fixed effects models. The samples are restricted to person-year observations in the dual-connected set of each gender in the Southeast region, described in columns (9)-(12) in Table 4. The first row shows the unadjusted correlations of the worker and establishment effects, as in Table 5, but for the dual-connected sets. Next, we present OLS estimates from regressing the person effect on the establishment effect, and IV estimates in which we use the establishment effect for the opposite race (but same gender) as instrument. The bottom rows present the first stage regression coefficients. In each case, we show results from two models: one with no other controls and one that controls for micro-region fixed effects (and therefore control for the availability of different subgroups of workers in each establishment's local labor market).

Table 7: Decomposition of the Racial Wage Gap

	Overall racial wage gap	Component attributable to:		
		Person effects	Covariate index	Establishment effects
	(1)	(2)	(3)	(4)
A. Males				
Unadjusted for locational differences:				
All education	0.227	0.160 (0.70)	-0.001 (0.00)	0.068 (0.30)
No high school	0.089	0.054 (0.61)	0.006 (0.07)	0.029 (0.33)
Completed high school	0.290	0.210 (0.72)	-0.006 -(0.02)	0.086 (0.30)
Adjusted for locational differences (nonwhites reweighted to white distribution):				
All education	0.186	0.144 (0.78)	-0.003 -(0.01)	0.044 (0.24)
No high school	0.052	0.038 (0.73)	0.004 (0.08)	0.010 (0.19)
Completed high school	0.262	0.203 (0.78)	-0.007 -(0.03)	0.065 (0.25)
B. Females				
Unadjusted for locational differences:				
All education	0.295	0.217 (0.73)	-0.022 -(0.07)	0.100 (0.34)
No high school	0.081	0.051 (0.63)	-0.009 -(0.11)	0.039 (0.48)
Completed high school	0.322	0.239 (0.74)	-0.027 -(0.08)	0.110 (0.34)
Adjusted for locational differences (nonwhites reweighted to white distribution):				
All education	0.261	0.216 (0.83)	-0.026 -(0.10)	0.071 (0.27)
No high school	0.046	0.041 (0.90)	-0.013 -(0.29)	0.018 (0.40)
Completed high school	0.291	0.243 (0.84)	-0.032 -(0.11)	0.079 (0.27)

Notes: The table displays the results from implementing the decomposition of the average white-nonwhite pay gap based on equation (4) using person-year observations in the dual-connected set of each gender in the Southeast region, without and with reweighting so that nonwhites have the same distribution across micro-regions as whites (of the same gender). See text for a detailed explanation of the decomposition. Entries in parentheses represent the percentage of the overall racial wage gap in the first column that is explained by the factor represented in the column.

Table 8: Decomposition of the Racial Gap in Establishment Effects

	Overall racial wage gap	Means of establishment effect		Gap in mean establishment effect	Decomposition of gap in establishment effects		Decomposition of sorting effect	
		Whites	Nonwhites		Sorting	Relative wage-setting	Skill-based sorting	Residual sorting
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Males								
All education	0.186	0.220	0.175	0.044	0.035	0.009	0.023	0.012
				(0.24)	(0.19)	(0.05)	(0.13)	(0.06)
No high school	0.052	0.150	0.140	0.010	0.004	0.006	0.003	0.000
				(0.19)	(0.07)	(0.12)	(0.06)	(0.01)
Completed high school	0.262	0.267	0.202	0.065	0.054	0.011	0.035	0.019
				(0.25)	(0.21)	(0.04)	(0.13)	(0.07)
B. Females								
All education	0.261	0.156	0.085	0.071	0.056	0.015	0.034	0.022
				(0.27)	(0.21)	(0.06)	(0.13)	(0.09)
No high school	0.046	0.043	0.025	0.018	0.010	0.008	0.003	0.007
				(0.40)	(0.22)	(0.17)	(0.06)	(0.16)
Completed high school	0.291	0.191	0.112	0.079	0.062	0.017	0.035	0.027
				(0.27)	(0.21)	(0.06)	(0.12)	(0.09)

Notes: The table displays the results from implementing the decomposition of the average white-nonwhite gap in establishment effects into a relative wage setting effect and a sorting effect based on equation (5), as we well as the decomposition of the sorting effect into a skill-based sorting and a residual sorting effect based on equations (7) and (8). The samples include all person-year observations in the dual-connected set of each gender in the Southeast region. Nonwhite observations are reweighted so that they have the same distribution across micro-regions as whites (of the same gender). Column (1) reproduces the overall racial wage gap from column (1) of Table 7; column (2) and (3) report the average establishment effect for white and nonwhites, respectively, and column (4) calculates the difference, which corresponds to the racial gap in mean establishment effects in column (4) of Tables 7. Column (5) and (6) decompose the gap in column (4) into a sorting and relative wage-setting effect, respectively. Columns (7) and (8) decompose the sorting effect into a skill-based and a residual sorting effect. Entries in parentheses represent the share of the overall racial wage gap (in column 1) that is explained by the source in the column heading.

Online Appendix

Appendix A: Data Construction

The raw RAIS data are provided in state-year files. The variable names, labels, types, formats, and value labels are standardized across years. For each state-year file, we keep workers employed on December 31st whose tenure is greater than one month to ensure employment throughout December—the month at which wages are calculated. Workers with invalid information for individual identifiers, establishment identifiers, and December wages are dropped. Log hourly wages are constructed by taking the natural logarithm of the real value of December wages (using Brazil's CPI for that month) divided by the monthly contracted hours (using weekly contracted hours multiplied by 4.348). When there is more than one December job for a given person-year pair, we keep the observation with the highest contracted hours. If tied in contracted hours, we keep the observation with the highest log hourly wage. If tied in contracted hours and log hourly wages, we randomly selected one observation. This ensures that person-year observations are unique within each state.

The selected unique person-year observations for each state are then stacked across 2002-2014 into a single state file. Each establishment is assigned its modal legal classification, municipality, and industry code. Each worker is assigned its modal gender, race, date of birth, and education (we record the original value of the race variable for each observation for our robustness checks).[1] We then keep observations belonging to the private sector based on the legal classification of each establishment (we remove observations with the Central Bank industry code as well as those with invalid industry codes), and workers who are hired on open-ended (i.e., not temporary) non-farm contracts and are paid on a monthly basis.

The remaining observations in each state file are then stacked across states into a single master file. The entire employment history of an individual is removed when one of the following four conditions is satisfied. First, the worker has a repeated person-year observation across states. Second, the nominal value of the reported December wage is below the federal minimum wage for that month. Third, the log hourly wage is in the 99th percentile of the wage distribution in the state. Fourth, the log hourly wage changes by more than 100 log points in adjacent years. Based on these person-year observations, the modal assignments from the previous paragraph are applied again, allowing us to categorize workers into mutually exclusive and exhaustive race-gender groups (the establishment size in Table 4 is based

---

[1]Date of birth is reported for 2002-2010 and age is reported for 2012-2014. We can thus calculate a worker's age for all years except 2011. We use 2010 and 2012 observations to calculate the age of 2011 observations; workers only observed in 2011 have a missing value for age, and are ultimately dropped from our samples. These workers only appear in one year and would not help identify establishment effects.

on the count of workers per establishment in this sample).

Finally, the above sample is restricted to the desired race-gender group and region. The education variable is used to calculate years of *schooling*. We calculate the years of potential labor market experience as $age - schooling - 6$. The remaining person-year observations, age 25 to 54 and with at least one year of potential labor market experience, constitute the analysis samples described in columns 1-4 in Table 4.[2]

---

[2]Workers only observed in 2011 are dropped as their potential labor market experience is missing.

Appendix B: A Simple Model of Monopsonistic Wage Setting

In this appendix we summarize implications of the monopsonistic wage setting model proposed by Card et al. (2018). In the model, a large number of firms (or establishments) compete over workers who have idiosyncratic tastes for different jobs. Building on standard monopsony wage setting models, we assume that firms cannot negotiate individually with workers, but instead post group-specific wages and are willing to hire any worker in a given group who is willing to work at that wage.

Worker Preferences

There are J firms (or establishments) in a local labor market and two groups of workers denoted by 1 and 2. Each firm j posts a pair of group-specific wages (w_{1j}, w_{2j}) that workers costlessly observe. Assume that the indirect utility of a job at firm j for worker i in group $g \in \{1, 2\}$, is:

$$u_{igj} = \delta_g^0 \ln(w_{gj} - b_g) + a_{gj}^0 + v_{igj}, \tag{1}$$

where b_g is a reference wage level (arising for example from the value of non-employment), a_{gj}^0 is a firm-specific amenity common to all workers in group g, v_{igj} is a worker-specific component of the value of a job at firm j, and $\delta_g^0 > 0$ is a factor expressing the relative valuation of the excess wage offered by the firm versus its non-pecuniary amenities.

Assume that $v_{igj} = \tau_g \epsilon_{igj}$ where ϵ_{igj} is an EV-1 error that is independent across workers, and τ_g is a scale factor reflecting the dispersion of idiosyncratic preferences within group g. Under this assumption the fraction of workers in group g who would choose to work at firm j is:

$$p_{gj} \equiv P(u_{igj} = \underset{k \in \{1,..,J\}}{\arg\max} \{u_{igk}\}) = \frac{\exp(\delta_g(\ln(w_{gj} - b_g) + a_{gj})}{\sum_{k=1}^{J} \exp(\delta_g \ln(w_{gk} - b_g) + a_{gk})}, \tag{2}$$

where $\delta_g = \delta_g^0 / \tau_g$ and $a_{gj} = a_{gj}^0 / \tau_g$. Note that the differences between groups in δ_g reflect both differences in the relative valuation placed on the excess wage versus the nonwwage amenity, and differences in the dispersion of idiosyncratic values for different firms.

To abstract from strategic interactions in wage-setting, assume that the number of firms J is large, in which case the logit probabilities in equation 2 are closely approximated by exponential probabilities:

$$p_{gj} \approx D_g \exp(\delta_g \ln(w_{gj} - b_g) + a_{gj}),$$

where D_g is a group-specific constant common to all firms in the market. In this case, the

firm-specific supplies of workers in the two groups, N_{1j} and N_{2j}, are:

$$\ln N_{1j}(w_{1j}) = d_1 + \delta_1 \ln(w_{1j} - b_1) + a_{1j} \tag{3}$$

$$\ln N_{2j}(w_{2j}) = d_2 + \delta_2 \ln(w_{2j} - b_2) + a_{2j}, \tag{4}$$

where d_1 and d_2 are market specific constants.

Firm Optimization

Firms have production functions of the form:

$$Y_j = T_j f(N_{1j}, N_{2j}), \tag{5}$$

where T_j is a firm-specific productivity shifter. The firm's problem is to post a pair of group-specific wages that minimize the cost of labor services given knowledge of the supply functions (3) and (4). These choices solve the cost-minimization problem:

$$\min_{w_{1j}, w_{2j}} w_{1j} N_{1j}(w_{1j}) + w_{2j} N_{2j}(w_{2j}) \text{ s.t. } T_j f(N_{1j}(w_{1j}), N_{2j}(w_{2j})) \geq Y.$$

The associated first order conditions can be written:

$$w_{1j} = \frac{e_{1j}}{1 + e_{1j}} T_j f_1 \mu_j \tag{6}$$

$$w_{2j} = \frac{e_{2j}}{1 + e_{2j}} T_j f_2 \mu_j \tag{7}$$

where e_{1j} and e_{2j} represent the elasticities of supply of group 1 and 2 workers at the optimal choice of wages, and μ_j represents the marginal cost of production, which the firm will equate to marginal revenue at an optimal choice for Y. Thus the terms $T_j f_1 \mu_j$ and $T_j f_2 \mu_j$ on the right hand sides of equations (6) and (7) represent the marginal revenue products of the two types of labor. These equations express the traditional "markdown" condition that the firm sets the wage for a given group equal to a fraction of its marginal revenue product, where the fraction is just $e_{gj}/(1 + e_{gj})$. If, for example the elasticity of supply is around 5 then the wage is about 15% less than marginal revenue product.

Using equations (3) and (4), the elasticities of supply are:

$$e_{1j} = \frac{\delta_1 w_{1j}}{w_{1j} - b_1}$$

$$e_{2j} = \frac{\delta_2 w_{2j}}{w_{2j} - b_2}.$$

4

Note that when $b_g = 0$ the firm's labor supply elasticity for group g is just $e_{gj} = \delta_g$, which is constant across firms and independent of the wage. Otherwise, when $b_g > 0$, the elasticity becomes large as $N_{gj} \to 0$, and falls in magnitude as N_{gj} becomes larger.

Using these expressions, the firm's first order conditions can be re-written as:

$$w_{1j} = \frac{1}{1+\delta_1}b_1 + \frac{\delta_1}{1+\delta_1}T_j f_1 \mu_j \tag{8}$$

$$w_{2j} = \frac{1}{1+\delta_2}b_2 + \frac{\delta_2}{1+\delta_2}T_j f_2 \mu_j. \tag{9}$$

The optimal wage choice for group g is a weighted average of the reference wage b_g and the group's marginal revenue product.

A Simple Benchmark: Linear Production and Fixed Output Price

To proceed we need to specify the production function and the firm's marginal revenue function. To keep things as simple as possible, we assume a linear technology – so the two groups are perfect substitutes in production – and we assume that the firm is a price-taker in its output market. Specifically, suppose that

$$f(N_{1j}, N_{2j}) = N_j \equiv \theta_1 N_{1j} + \theta_2 N_{2j}$$

where θ_g gives the efficiency units of each worker in group g and N_j represents the total efficiency units of labor at firm j. Suppose in addition that the firm's output price is P_j^0. Then the first order conditions (8) and (9) evaluate to:

$$w_{1j} = \frac{1}{1+\delta_1}b_1 + \frac{\delta_1}{1+\delta_1}T_j P_j^0 \theta_1$$

$$w_{2j} = \frac{1}{1+\delta_2}b_2 + \frac{\delta_2}{1+\delta_2}T_j P_j^0 \theta_2.$$

To understand the implications of this model for the wage structure, suppose that the reference wages of the two groups are proportional to their relative productivities, so that

$$b_1 = \theta_1 b, \; b_2 = \theta_2 b.$$

Now the first order conditions can be re-written:

$$\ln w_{1j} = \ln \frac{\theta_1 b}{1 + \delta_1} + \ln(1 + \delta_1 R_j) \tag{10}$$

$$\ln w_{2j} = \ln \frac{\theta_2 b}{1 + \delta_2} + \ln(1 + \delta_2 R_j) \tag{11}$$

where $R_j \equiv \frac{T_j P_j^0}{b}$ gives the ratio of the marginal revenue product of labor at firm j to the reference wage. Wages of both groups contain a firm-specific component that depends on R_j and the group-specific supply parameter δ_g. To interpret these expressions, note that value added per standardized unit of labor is $\lambda_j \equiv P_j^0 Y_j / N_j = P_j^0 T_j$, so $R_j = \lambda_j / b$ is the ratio of value added per standardized unit of labor to reference wage for a worker with 1 efficiency unit of labor.

An important implication of these expressions is that for firms with $R_j \approx 1$ – i.e., "marginally efficient" firms whose value added per worker is approximately equal to the outside option available to workers – the wage of each group is approximately equal to its marginal productivity:

$$\ln w_{1j} \approx \ln(\theta_1 T_j P_j^0)$$
$$\ln w_{2j} \approx \ln(\theta_2 T_j P_j^0)$$

These "marginal" firms have essentially no market power (since the elasticity of labor supply tends to infinity as the wage falls to the reference wage level), so their offered wages reveal the productivities of the two groups.

Implications for AKM-Style Wage Models

To illustrate the implications of equations 10 and 11, suppose that $\delta_g R_j$ is relatively small. In this case:

$$\ln w_{1j} \approx \ln \frac{(1 - \theta)b}{1 + \delta_1} + \delta_1 R_j$$
$$\ln w_{2j} \approx \ln \frac{\theta b}{1 + \delta_2} + \delta_2 R_j.$$

These equations imply that the wages of workers at different firms can be written in the form:

$$\ln w_{gj} = \alpha_g + \psi_j^g, \tag{12}$$

where $\psi_j^g = \delta_g R_j$ is a group-specific firm component of wages. Note that groups with a higher relative valuation of wages versus non-wage amenities (i.e., larger values of δ_g^0) and groups with less dispersion in the firm-specific valuations of individual workers (i.e., smaller values of τ_g) would be expected to have higher values of δ_g. These groups will have "larger steps" in the job ladder across firms.

References

Card, David, Ana Rute Cardoso, Jorg Heining, and Patrick M. Kline. 2018. "Firms and Labor Market Inequality: Evidence and Some Theory." *Journal of Labor Economics* 36(Supplement):S13–S70.

Figure C1: Brazil's Regions and Racial Composition of its Micro-Regions

Share of nonwhites
- 7%-20%
- 20%-30%
- 30%-40%
- 40%-50%
- 50%-60%
- 60%-70%
- 70%-87%

North
15.8 M

Northeast
53.1 M

Midwest
14.1 M

Brasilia
3.6 M

Southeast
80.3 M

Rio de Janeiro
12.5 M

Sao Paulo
21.1 M
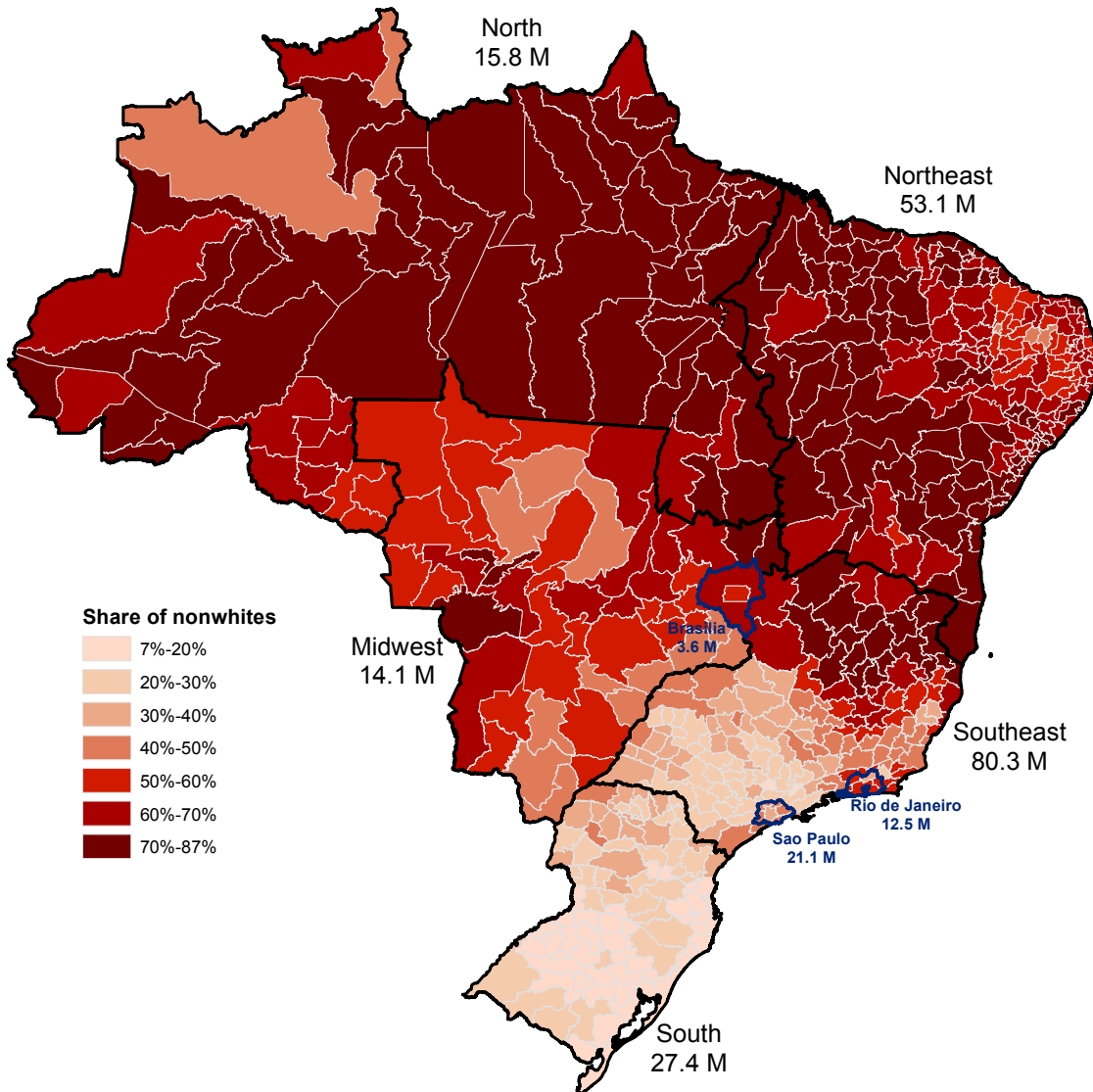
South
27.4 M

Notes: The figure displays a map of Brazil; the black lines correspond to the borders of Brazil's five regions (North, Northeast, Southeast, South, and Midwest; a region's population according to the 2010 census is reported under its name); the white lines correspond to the borders of Brazil's 557 micro-regions; and blue lines identify Brazil's two largest cities (São Paulo and Rio de Janeiro) and capital (Brasilia). The coloring provides information on the share of nonwhites in a micro-region's population according to the 2010 census.

Figure C2: Log Hourly Wage Distributions Among Private-Sector Employees (Females)
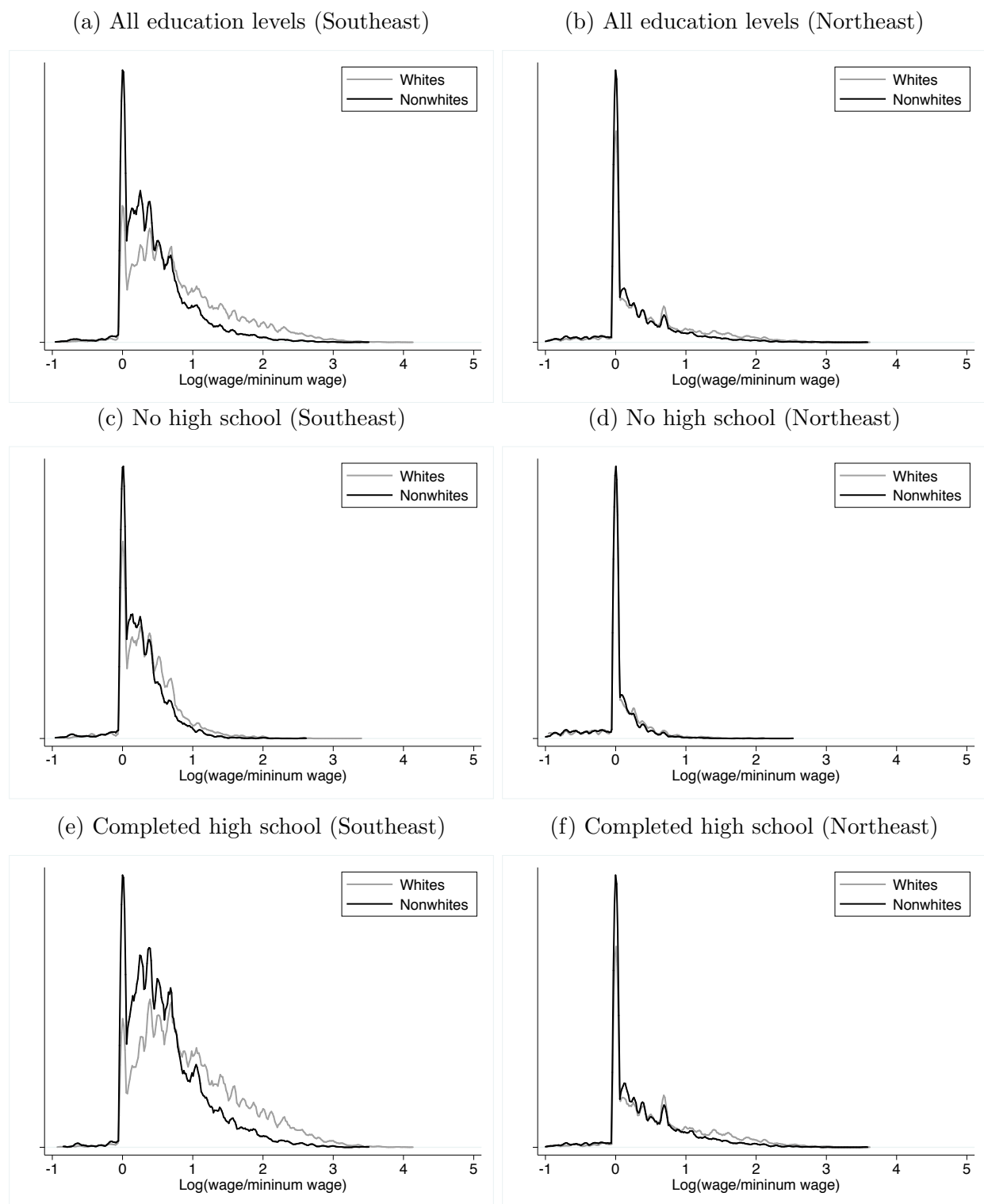
(a) All education levels (Southeast)

(b) All education levels (Northeast)

(c) No high school (Southeast)

(d) No high school (Northeast)

(e) Completed high school (Southeast)

(f) Completed high school (Northeast)

Notes: The figure displays kernel densities (Epanechnikov kernel with a 0.025 half-width) of the log wage-to-minimum-wage ratio for whites and nonwhites, based on PNAD 2002-2014 and constructed using survey weights (PNAD was not conducted in 2010). The samples include female non-farm private-sector employees (either formal or informal), age 25 to 54, with potential labor market experience of at least 1 year, and non-missing data on race, gender, education, wage, and hours worked. The left and right panels restrict the samples to the Southeast and the Northeast regions, respectively. The top panel pools all education levels together; the middle and bottom panels are restricted to workers with no high school degree and with at least a high school degree, respectively.

Figure C3: Age-Wage Profiles by Cohort

(a) White males

(b) Non-white males

(c) White females

(d) Non-white females
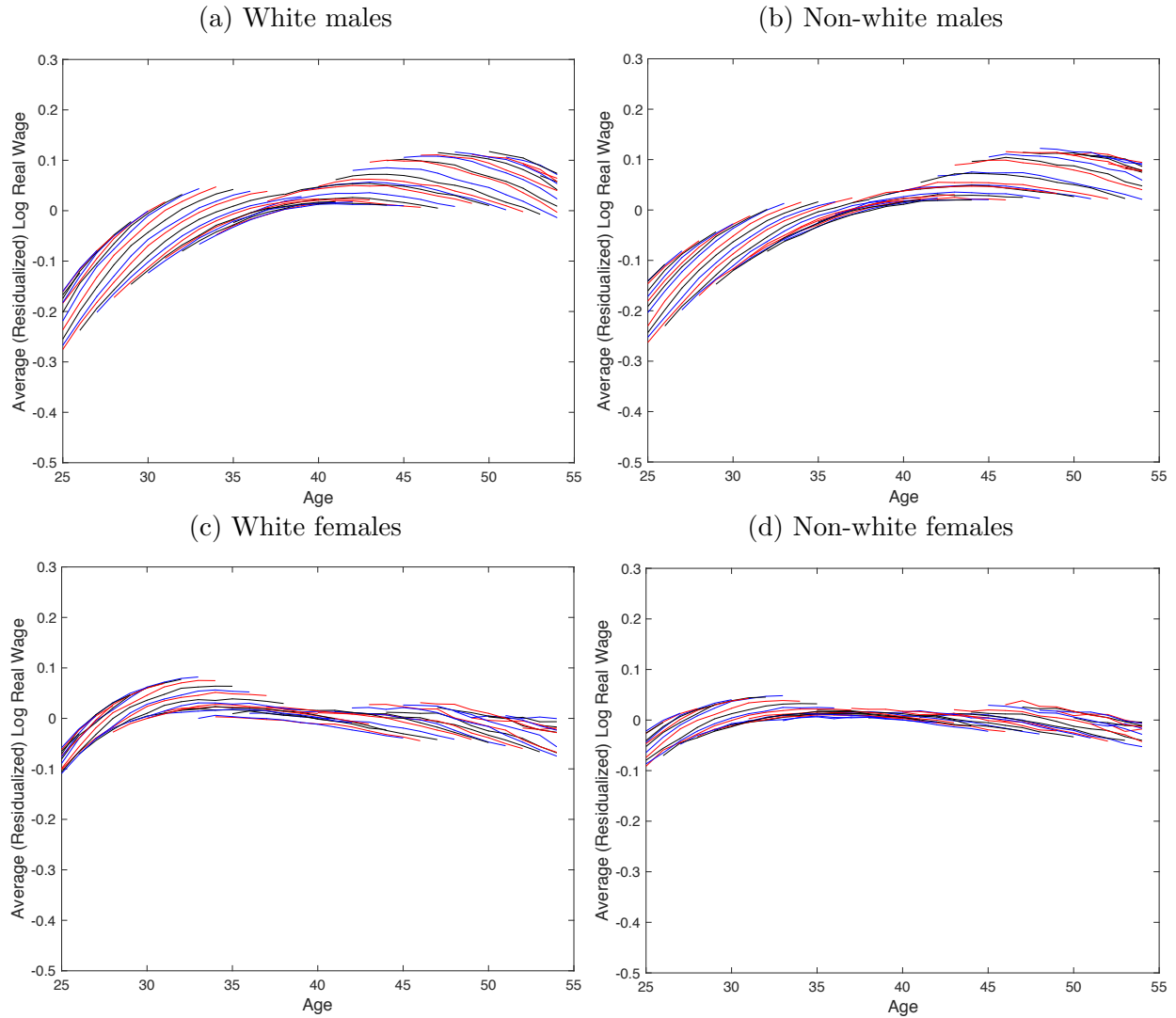
Notes: The figure displays the age-wage profile by cohort, separately for each race-gender group. Specifically, each panel shows the residuals of a regression of mean log wages by age and cohort on year fixed effects; the lines correspond to age-wage profiles for different cohorts. The mean log wages by age and cohort are constructed from the person-year observations in the largest connected set of each race-gender group, described in columns (5)-(8) in Table 4. The age-wage profiles tend to peak around age 40 for white and non-white males and around age 35 for white and non-white females.

Figure C4: Correlation of Establishment Effects in AKM Models for Southeast vs. Brazil

(a) White males
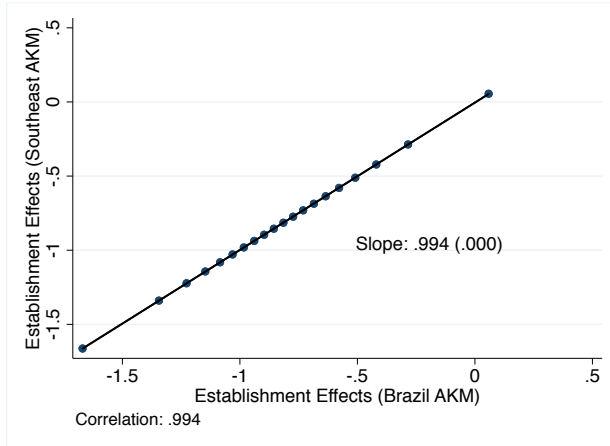
Slope: .994 (.000)

Correlation: .994

(b) Non-white males

Slope: .987 (.000)

Correlation: .990

(c) White females

Slope: .994 (.000)

Correlation: .995

(d) Non-white females

Slope: .989 (.000)

Correlation: .991
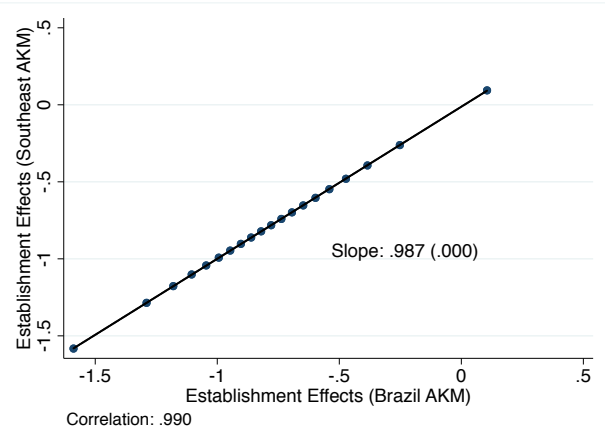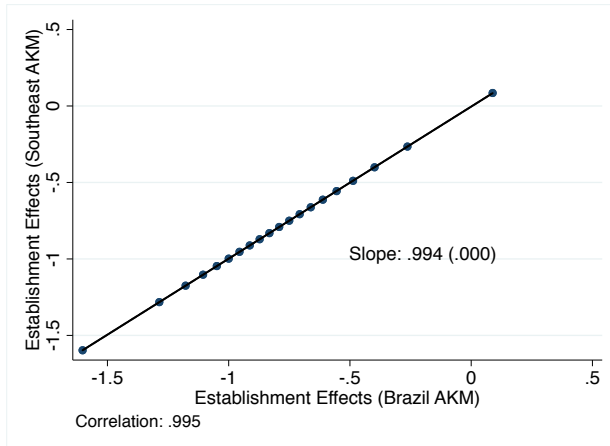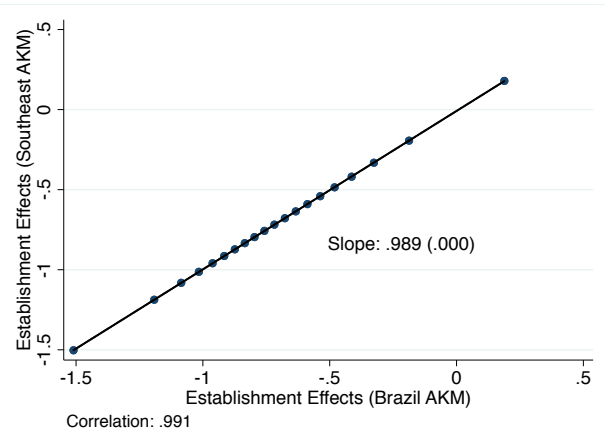
Notes: The figure displays binned scatterplots of establishment effects from AKM models estimated on the dual connected set of each race-gender group in the Southeast region versus in all of Brazil. Observations are unique establishments that appear in both the Southeast and Brazil samples. Scatterplots use 20 equal-sized bins and plot the within-bin means. The slope of the OLS fit line with its standard error (in parentheses) are reported in each graph, as well as the correlation coefficient of the variables (under each graph).

Figure C5: Mean AKM Residuals by Person Effect and Establishment Effect Deciles

(a) White males

(b) Non-white males

(c) White females

(d) Non-white females
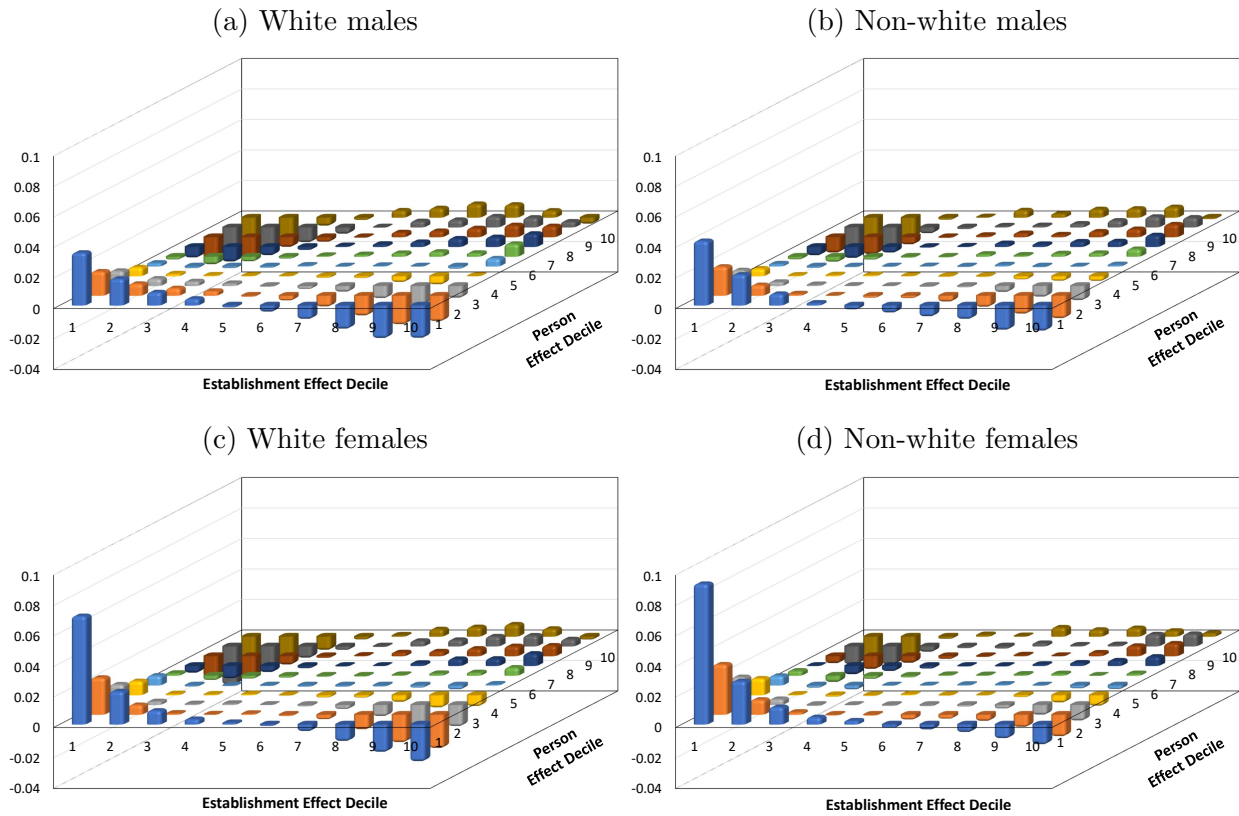
Notes: The figure displays mean residuals from AKM models estimated on the largest connected set of each race-gender group in the Southeast region, for 100 cells defined by deciles of estimated establishment effects interacted with deciles of estimated person effects. The mean residuals in each cell are close to zero, with the exception of cells representing workers with low person effects employed at workplaces with low establishment effects, where the mean residuals are systematically positive. This pattern is most pronounced for non-white females, and is consistent with upward pressure from the minimum wage that is particularly important for low-skilled workers employed at low-paying establishments. We evaluate the sensitivity of our results to these observations in Section 6.

Figure C6: Shares of Jobs in High-Premium Establishments and of High-Skilled White Workers Across Micro-Regions

(a) Males   (b) Females
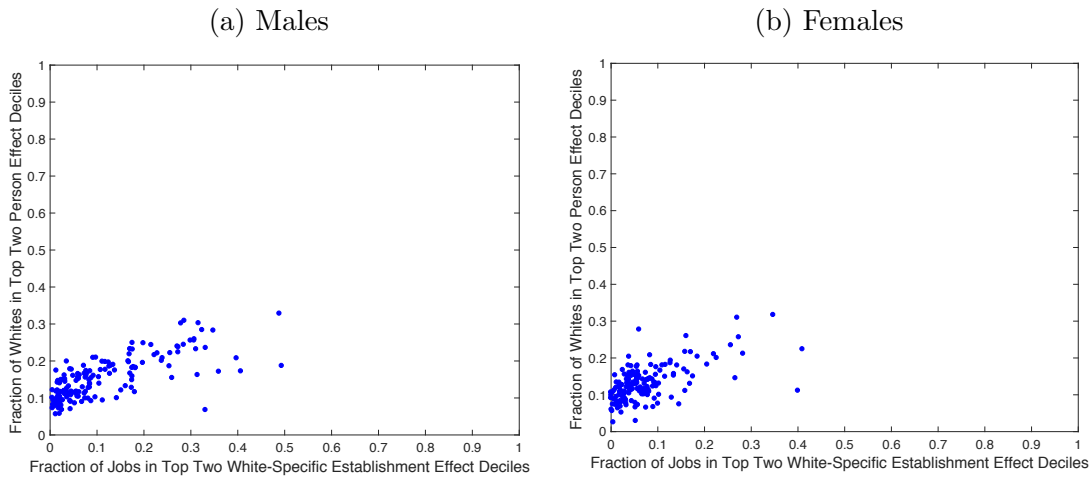
Notes: The figure displays scatterplots of the relationship between the share of jobs in high-premium establishments and the share of high-skilled white workers across micro-regions, based on person-year observations in the dual-connected set of each gender in the Southeast region. Each dot corresponds to a micro-region. The value on the x-axis is the share of person-year observations in establishments found in the top two deciles of the white-specific establishment effect distribution. The value on the y-axis is the share of whites found in the top two deciles of the person effect distribution (deciles are defined based on the full distribution in the dual-connected set, thus pooling white and non-white workers). The positive relationship implies that micro-regions with relatively more jobs in high-premium establishments have a relatively more skilled population of white workers.

Figure C7: Distribution of Estimated Person Effects for White and Non-White Workers

(a) Males (all education)

(b) Females (all education)

(c) Males (no high school)

(d) Females (no high school)

(e) Males (high school)

(f) Females (high school)
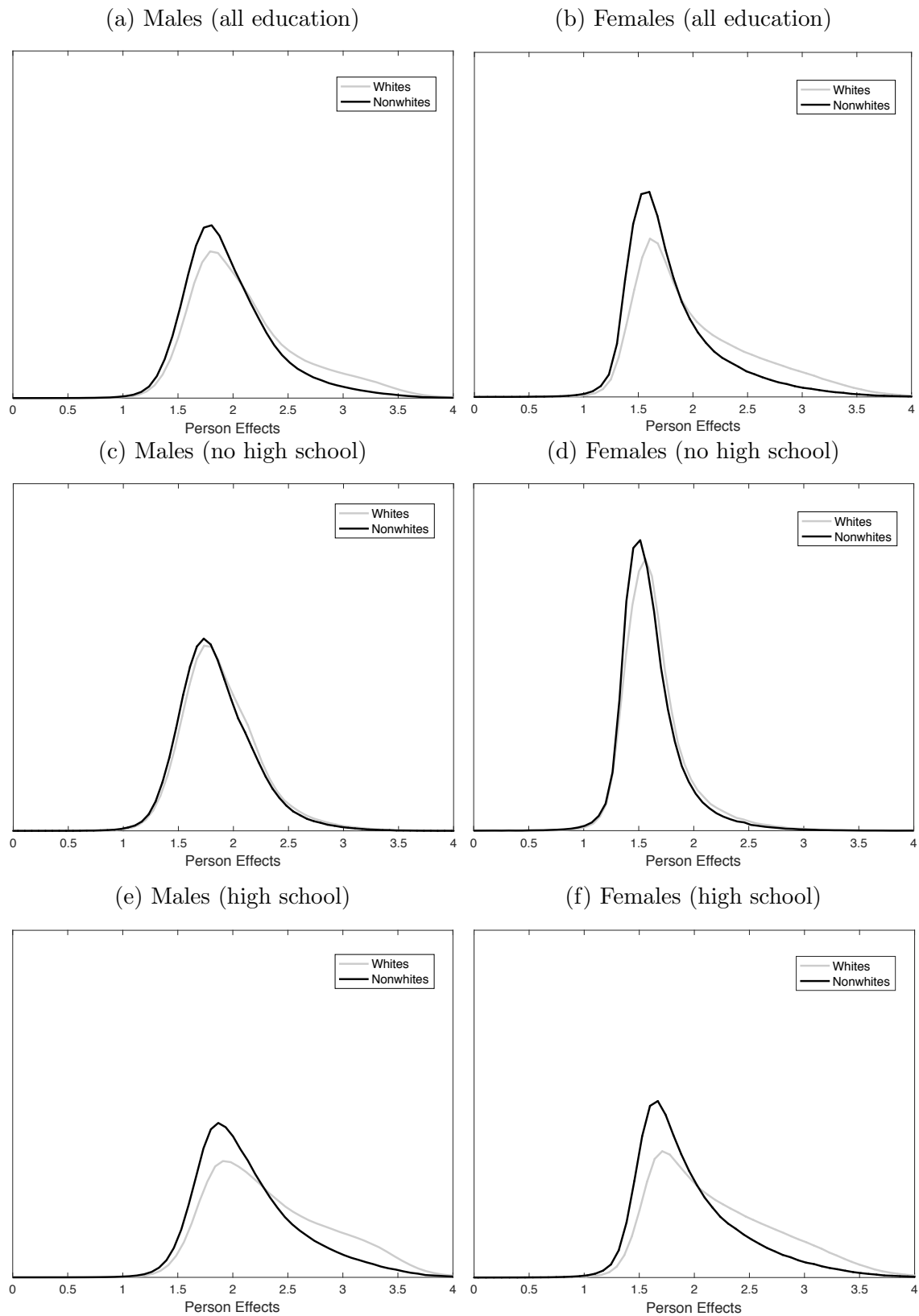
Notes: The figure displays kernel densities (normal kernel with optimal bandwidth) of the estimated person effects, using person-year observations in the dual-connected set of each gender (and education) group in the Southeast region. Nonwhites are reweighted so as to have the same distribution across micro-regions as whites (of the same gender).

Figure C8: Residual Sorting Effect by Person Effect Deciles (Using Other Gender)

(a) Males (all education)

(b) Females (all education)

(c) Males (completed high school)

(d) Females (completed high school)
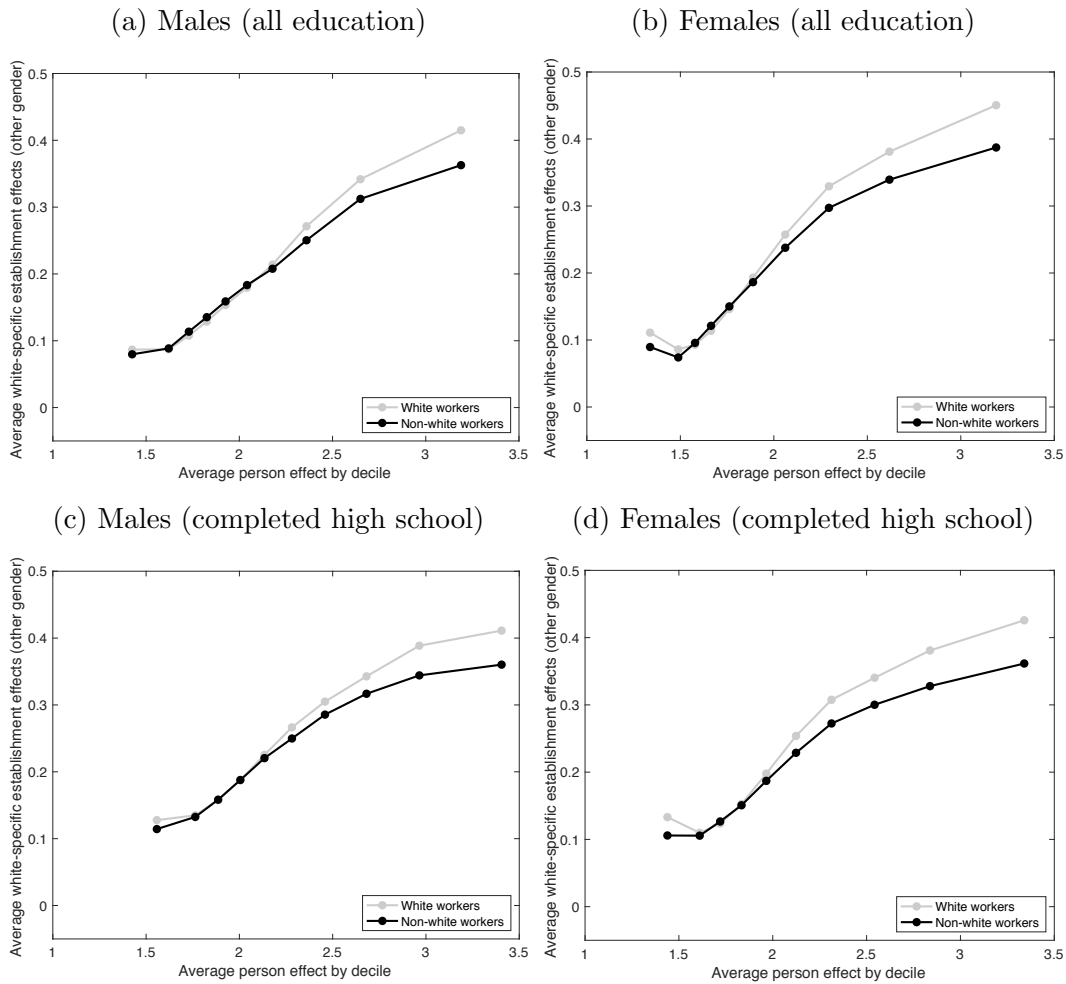
Notes: The figure is constructed similarly as Figure 8, but uses the establishment effects of the other gender to show that the main patterns in Figure 8 are robust to the sampling error issue discussed in the text. The figure displays the means of the estimated white-specific establishment effects for the other gender at workplaces of whites and non-white workers in each decile of the person effect distribution. The gap between the mean white-specific establishment effects of the two groups represents a residual sorting effect for workers in that decile of person effects, evaluated using the estimated firm effects for the other gender of the same race group. We use person-year observations in the tetra-connected set (the intersection of the dual-connected sets of both genders) in the Southeast region (the person effect deciles are defined pooling whites and nonwhites together). Nonwhites are reweighted so as to have the same distribution across micro-regions as whites (of the same gender). The dip from the first to the second decile, which was more severe for white workers in Figure 8, is no much smaller and similar between the race groups because there is no more sampling error issue at the bottom of the skill distribution (the small remaining dip for women is likely due to the upward pressure on wages from the minimum wage). As in Figure 8, in contrast, there is still a strong tendency for higher-skilled workers to work at establishments with higher pay premiums. There is also still a clear divergence between the mean establishment effects for whites and nonwhites, implying that nonwhites are systematically under-represented at the best-paying workplaces, even conditional on their ability. In this case, however, there is no more sampling error issue at the top of the skill distribution.

Figure C9: Relative Wage-Setting Effect by Person Effect Deciles (Other Gender)

(a) Males (all education)

(b) Females (all education)

(c) Males (completed high school)

(d) Females (completed high school)
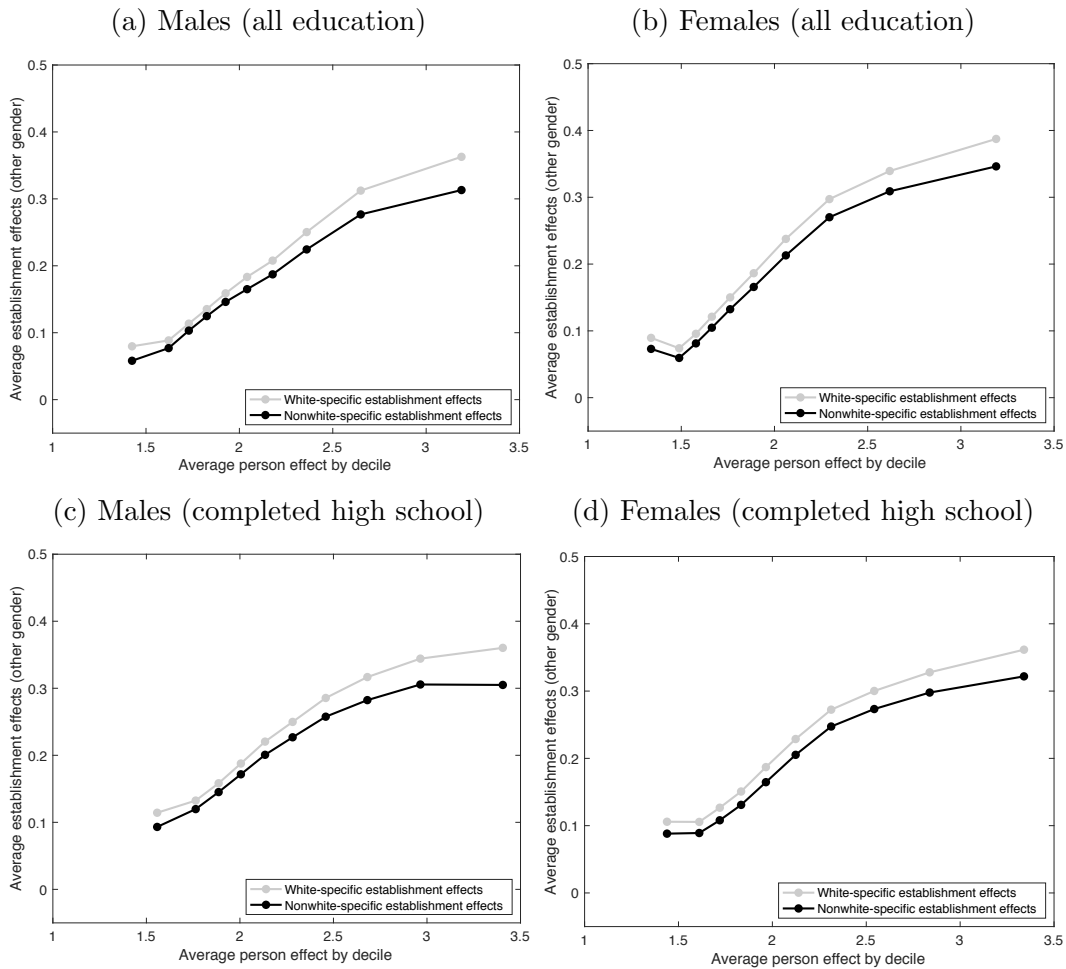
Notes: The figure is constructed similarly as Figure 9, but uses the establishment effects of the other gender to show that the main patterns in Figure 9 are robust to the sampling error issue discussed in the text. The figure displays the means of the estimated white-specific and nonwhite-specific establishment effects for the other gender at workplaces of non-white workers in each decile of the person effect distribution. The gap between the mean white-specific and nonwhite-specific establishment effects represents a relative wage-setting effect for workers in that decile of person effects, evaluated using the estimated firm effects for the other gender of the same race group. We use person-year observations in the tetra-connected set (the intersection of the dual-connected sets of both genders) in the Southeast region (the person effect deciles are defined pooling whites and nonwhites together). Nonwhites are reweighted so as to have the same distribution across micro-regions as whites (of the same gender). The dip from the first to the second decile, which was more severe for the nonwhite-specific establishment effect in Figure 9, is no much smaller and similar for the mean white-specific and nonwhite-specific establishment effects because there is no more sampling error issue at the bottom of the skill distribution (the small remaining dip for women is likely due to the upward pressure on wages from the minimum wage). As in Figure 9, in contrast, the gap between the mean white-specific and nonwhite-specific establishment effects at the workplaces of nonwhites also widens across the deciles. In this case, however, there is no more sampling error issue at the top of the skill distribution.

Table D1: Characteristics of Private-Sector Employees by Race Group (Other Regions)

	North region				Northeast region				South region				Midwest region			
	All	White	Mixed race	Black	All	White	Mixed race	Black	All	White	Mixed race	Black	All	White	Mixed race	Black
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
A. Males																
Share of sample in column race group	1.00	0.23	0.69	0.08	1.00	0.27	0.62	0.10	1.00	0.78	0.17	0.04	1.00	0.41	0.51	0.08
Share of column race group in private employment	0.33	0.34	0.32	0.34	0.31	0.31	0.30	0.37	0.44	0.43	0.47	0.52	0.39	0.37	0.40	0.44
Characteristics of priv. employees in column race group:																
Mean years of education	7.79	8.60	7.58	7.27	7.43	8.29	7.03	7.48	8.58	8.92	7.41	7.65	8.14	9.01	7.62	7.43
Fraction with high school or more	0.40	0.48	0.38	0.35	0.38	0.47	0.34	0.39	0.44	0.48	0.32	0.34	0.41	0.50	0.35	0.36
Mean log hourly wage (R$)	1.42	1.55	1.38	1.39	1.22	1.35	1.16	1.24	1.71	1.76	1.54	1.51	1.61	1.75	1.51	1.53
Share with wage ≤ 2 minimum wages	0.68	0.59	0.71	0.72	0.79	0.72	0.82	0.81	0.46	0.42	0.58	0.61	0.55	0.47	0.61	0.61
Share in formal sector employment	0.68	0.71	0.68	0.67	0.67	0.69	0.66	0.72	0.84	0.85	0.81	0.83	0.76	0.77	0.75	0.77
B. Females																
Share of sample in column race group	1.00	0.25	0.68	0.06	1.00	0.29	0.61	0.09	1.00	0.80	0.15	0.04	1.00	0.43	0.49	0.07
Share of column race group in private employment	0.14	0.17	0.13	0.13	0.14	0.16	0.12	0.15	0.27	0.28	0.24	0.28	0.21	0.22	0.19	0.21
Characteristics of priv. employees in column race group:																
Mean years of education	9.95	10.65	9.68	9.27	10.01	10.69	9.60	9.98	9.73	10.00	8.37	8.67	10.14	10.86	9.51	9.53
Fraction with high school or more	0.67	0.73	0.64	0.61	0.67	0.73	0.63	0.69	0.58	0.61	0.44	0.47	0.64	0.71	0.58	0.59
Mean log hourly wage (R$)	1.33	1.47	1.27	1.27	1.17	1.31	1.09	1.16	1.51	1.55	1.33	1.32	1.48	1.62	1.35	1.36
Share with wage ≤ 2 minimum wages	0.80	0.71	0.84	0.83	0.84	0.77	0.88	0.89	0.70	0.67	0.84	0.84	0.72	0.63	0.79	0.82
Share in formal sector employment	0.70	0.75	0.68	0.67	0.69	0.72	0.67	0.71	0.85	0.86	0.84	0.85	0.79	0.79	0.78	0.81

Notes: The table displays statistics based on PNAD 2002-2014 and constructed using survey weights (PNAD was not conducted in 2010). The samples include males (panel A) and females (panel B), age 25 to 54, with potential labor market experience of at least 1 year, and non-missing data on race, gender, and education. Private employment status includes formal and informal employees, but excludes those with missing wages or hours. Columns (1)-(4), (5)-(8), (9)-(12), and (13)-(16) restrict the sample to the North region, the Northeast region, the South region, and the Midwest region, respectively. In each case, the first column pool all race groups together; the other columns report statistics for each of the three main race groups (the race groups in PNAD are white, mixed race, black, asian, and indigenous).

Table D2: Racial Differences in Log Wages in PNAD (Private-Sector Employees, Other Regions)

	North region				Northeast region				South region				Midwest region			
	All		Completed high school		All		Completed high school		All		Completed high school		All		Completed high school	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
A. Males																
Dummy if mixed race	-0.18	-0.09	-0.20	-0.12	-0.21	-0.08	-0.25	-0.14	-0.26	-0.11	-0.27	-0.16	-0.26	-0.12	-0.28	-0.16
(std err.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Dummy if black	-0.18	-0.08	-0.19	-0.12	-0.21	-0.09	-0.29	-0.16	-0.25	-0.13	-0.29	-0.17	-0.28	-0.12	-0.33	-0.19
(std err.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Year and state fixed effects	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Education and experience	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes
B. Females																
Dummy if mixed race	-0.21	-0.09	-0.21	-0.11	-0.24	-0.10	-0.23	-0.12	-0.26	-0.09	-0.27	-0.14	-0.29	-0.12	-0.29	-0.15
(std err.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Dummy if black	-0.21	-0.07	-0.18	-0.08	-0.25	-0.11	-0.28	-0.15	-0.24	-0.11	-0.27	-0.14	-0.32	-0.14	-0.32	-0.17
(std err.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Year and state fixed effects	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Education and experience	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes

Notes: The table displays the results of regressing the log(hourly wage) on a series of race group dummies, using data from PNAD 2002-2014 (PNAD was not conducted in 2010). The samples include male (panel A) and female (panel B) private-sector employees (either formal or informal), age 25 to 54, with potential labor market experience of at least 1 year, and non-missing data on race, gender, education, wage, and hours worked. All specifications include year and state fixed effects and use survey weights. Education and experience controls include five education dummies (incomplete elementary school, and complete elementary school, middle school, high school or college) and a quadratic in potential experience. The omitted race group is white. Other race dummies not reported are indigenous and asian. Columns(1)-(4), (5)-(8), (9)-(12), and (13)-(16) restrict the samples to the North, Northeast, South, and Midwest regions, respectively. In each case, the first two columns pool workers of all education levels; the other columns restrict the sample to workers with completed high school.

Table D3: Racial Differences in Log Work Hours in PNAD (Formal Private-Sector Employees)

	Brazil				Southeast region			
	All		Completed high school		All		Completed high school	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Males								
Dummy if mixed race	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00
(std err.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Dummy if black	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00
(std err.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Year and state fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Education and experience	no	yes	no	yes	no	yes	no	yes
B. Females								
Dummy if mixed race	0.01	0.00	0.02	0.00	0.02	0.00	0.02	0.00
(std err.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Dummy if black	0.02	0.00	0.02	0.00	0.02	0.00	0.03	0.01
(std err.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Year and state fixed effects	yes	yes	yes	yes	yes	yes	yes	yes
Education and experience	no	yes	no	yes	no	yes	no	yes

Notes: The table displays the results of regressing log monthly work hours on a series of race group dummies, using data from PNAD 2002-2014 (PNAD was not conducted in 2010). The samples include male (panel A) and female (panel B) private-sector non-farm employees (formal employees only), age 25 to 54, with potential labor market experience of at least 1 year, and non-missing data on race, gender, education, wage, and hours worked. All specifications include year and state fixed effects and use survey weights. Education and experience controls include five education dummies (incomplete elementary school, and complete elementary school, middle school, high school, or college) and a quadratic in potential experience. The omitted race group is white. Other race dummies not reported are indigenous and Asian. The samples in columns (1)-(4) use data for the whole country; columns (5)-(8) restrict the samples to the Souteast region only. In each case, the first two columns pool workers of all education levels; the other columns restrict the sample to workers with completed high school.

Table D4: Racial Differences in Formality and Log Wages in PNAD (Private-Sector Employees)

	North region			Northeast region			South region			Midwest region		
	Formality	Log hourly wage		Formality	Log hourly wage		Formality	Log hourly wage		Formality	Log hourly wage	
	PNAD	PNAD	PNAD-formal	PNAD	PNAD	PNAD-formal	PNAD	PNAD	PNAD-formal	PNAD	PNAD	PNAD-formal
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
A. Males - All education												
Dummy if nonwhite	0.00	-0.09	-0.09	0.01	-0.08	-0.08	-0.02	-0.12	-0.11	0.00	-0.12	-0.11
(std err.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Mean for whites	0.71	1.56	1.67	0.70	1.36	1.50	0.85	1.76	1.80	0.77	1.75	1.80
B. Males - Completed high school												
Dummy if nonwhite	0.01	-0.12	-0.12	0.01	-0.14	-0.14	0.00	-0.17	-0.16	0.02	-0.17	-0.16
(std err.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Mean for whites	0.80	1.85	1.91	0.80	1.73	1.78	0.88	2.06	2.07	0.82	2.08	2.09
C. Females - All education												
Dummy if nonwhite	-0.03	-0.09	-0.08	-0.02	-0.10	-0.10	-0.01	-0.10	-0.09	0.00	-0.12	-0.12
(std err.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Mean for whites	0.75	1.47	1.56	0.72	1.32	1.46	0.86	1.55	1.58	0.79	1.62	1.66
D. Females - Completed high school												
Dummy if nonwhite	-0.03	-0.11	-0.10	-0.02	-0.12	-0.12	-0.01	-0.14	-0.14	0.01	-0.16	-0.15
(std err.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Mean for whites	0.80	1.64	1.68	0.79	1.51	1.58	0.88	1.77	1.78	0.82	1.83	1.84

Notes: The table displays the results of regressing the outcome on top of each column on a series of race group dummies, using data from PNAD 2002-2014 (PNAD was not conducted in 2010). The samples include male (panels A and B) and female (panels C and D) private-sector employees, age 25 to 54, with potential labor market experience of at least 1 year, tenure of at least 1 month, and non-missing data on race, gender, education, wage, and hours worked. All specifications include year and state effects, the same education dummies as in Table 2, and quadratic in potential experience (we use survey weights with the PNAD data). The omitted race category is white; nonwhite includes both black and mixed race. Other race dummies not reported are indigenous and Asian. The samples in columns (1)-(3), (4)-(6), (7)-(9), and (10)-(12) use data for the North, Northeast, South and Midwest regions only, respectively. In each case, the outcome in the first column is a dummy for being formally employed. The specification in the second column is the same as in columns (2), (4), (6), and (8) in Table 2. The third column restricts the PNAD sample to formal employees.

Table D5: Wage Differentials in the Normalizing Sector and Other Sectors

	PNAD		RAIS	
	Restaurant sector	All other sectors	Restaurant sector	All other sectors
	(1)	(2)	(3)	(4)
Dummy if nonwhite	-0.034	-0.120	-0.024	-0.071
Dummy if female	-0.261	-0.326	-0.187	-0.302
Nonwhite *x* female	-0.002	0.016	-0.004	-0.038

Notes: The table displays the results of regressing log hourly wages on a nonwhite dummy fully interacted with a female dummy using data from PNAD 2002-2014 (PNAD was not conducted in 2010) from the Southeast region as in Table 2 columns (5)-(8) and the pooled RAIS dual-connected sets for males and females from the Southeast region shown in Table 4 columns (9)-(12). All specifications include year and state effects, the same education dummies as in Table 2, and quadratic in potential experience (we use survey weights with the PNAD data). The omitted race category is white; nonwhite includes both black and mixed race. Other race dummies not reported are indigenous and Asian. All coefficients are statistically significant.

Table D6: Summary of Estimated Two-Way Fixed Effects Models by Race-Gender Group (Brazil)

	White male	Non-white male	White female	Non-white female
	(1)	(2)	(3)	(4)
Standard deviation of log wages	0.674	0.582	0.682	0.558
Summary of Parameter Estimates:				
Std. dev. of person effects (across person-yr obs.)	0.473	0.400	0.514	0.437
Std. dev. of estab. effects (across person-yr obs.)	0.328	0.297	0.323	0.283
Std. dev. of Xb (across person-yr obs.)	0.185	0.191	0.194	0.197
Correlation of person/estab. effects	0.229	0.141	0.203	0.042
RMSE of model	0.230	0.221	0.216	0.196
Adjusted R-squared of model	0.884	0.856	0.900	0.876
Comparison job-match effects model:				
RMSE of match-effects model	0.196	0.192	0.186	0.175
Adjusted R-squared of match-effects model	0.916	0.891	0.926	0.902
Variance of job match effect	0.017	0.015	0.013	0.009
Percent of variance of wages due to job match	3.7	4.4	2.8	2.9
Variance decomposition:				
Percent of variance of log wages due to:				
person effects	49.3	47.1	56.9	61.3
establishment effects	23.6	26.1	22.4	25.8
covariance of person and estab. effects	15.6	9.9	14.5	3.3
Xb and associated covariances	2.7	6.2	-1.2	0.8
residual (including job match and time-varying)	8.9	10.7	7.4	8.7

Notes: The table summarizes the results from estimating two-way fixed effects models for the log hourly wage using person-year observations in the largest connected set for each race-gender group in Brazil as a whole. The models include dummies for individual workers and individual establishments, year dummies interacted with the same five education dummies as in Table 2, and quadratic and cubic terms in age interacted with the education dummies. The "comparison" job-match effects models include dummies for each job match as well as the other covariates in the basic models.

Table D7: Rank Correlation of Sector Premiums

	Krueger-Summers: CPS	White male: RAIS	White female: RAIS
	(1)	(2)	(3)
Krueger-Summers: CPS	1.00		
	(243)		
	--		
White male: RAIS	0.54	1.00	
	(247)	(247)	
	[0.00]	--	
White female: RAIS	0.60	0.93	1.00
	(237)	(236)	(237)
	[0.00]	[0.00]	--

Notes: The table reports Spearman's rank correlation coefficients of sector premiums (with number of observations in parentheses and p-values in brackets). Krueger-Summers sector premiums are the estimated wage differentials for three-digit census industries (CIC) reported in Krueger and Summer (1988) Table A1. These estimates are obtained from a cross-section regression of log wage on industry dummies with human capital and demographic controls on the 1984 CPS. The average white-specific establishment effects by three-digit industry sectors (CNAE) in RAIS are the same as those reported in Figure 5, i.e., computed over person-year observations in the dual-connected set of each gender in the Southeast region. The CIC were matched to CNAE using a cross-walk. Since this matching is not one-to-one we use an outer join that creates repeated values for both CIC- and CNAE-based sector premiums.

Table D8: Relating Decomposition Results to Standard Mincerian Model

	Overall racial wage gap	Component attributable to:		
		Person Effects	Covariate Index	Establishment Effects
	(1)	(2)	(3)	(4)
A. Males				
(i) All education				
Wage gap	0.186	0.144	-0.003	0.044
Explained by educ./exp.	0.118	0.088	0.000	0.029
Unexplained component	0.069	0.056	-0.003	0.015
(Share explained by educ./exp.)	(0.63)	(0.61)	-(0.05)	(0.65)
(ii) No high school				
Wage gap	0.052	0.038	0.004	0.010
Explained by educ./exp.	0.012	0.007	0.004	0.001
Unexplained component	0.039	0.030	0.001	0.009
(Share explained by educ./exp.)	(0.24)	(0.20)	(0.88)	(0.12)
(iii) Completed high school				
Wage gap	0.262	0.203	-0.007	0.065
Explained by educ./exp.	0.137	0.110	-0.002	0.029
Unexplained component	0.124	0.093	-0.006	0.036
(Share explained by educ./exp.)	(0.53)	(0.54)	(0.23)	(0.45)
B. Females				
(i) All education				
Wage gap	0.261	0.216	-0.026	0.071
Explained by educ./exp.	0.169	0.139	-0.008	0.039
Unexplained component	0.091	0.077	-0.018	0.032
(Share explained by educ./exp.)	(0.65)	(0.64)	(0.32)	(0.55)
(ii) No high school				
Wage gap	0.046	0.041	-0.013	0.018
Explained by educ./exp.	0.008	0.007	0.000	0.001
Unexplained component	0.037	0.034	-0.013	0.017
(Share explained by educ./exp.)	(0.18)	(0.16)	-(0.02)	(0.06)
(iii) Completed high school				
Wage gap	0.291	0.243	-0.032	0.079
Explained by educ./exp.	0.163	0.140	-0.012	0.034
Unexplained component	0.128	0.103	-0.020	0.045
(Share explained by educ./exp.)	(0.56)	(0.58)	(0.37)	(0.43)

Notes: The table displays the results from decomposing the overall racial wage gap in the dual-connected set of each gender in the Southeast region, and the gaps attributed to each of the three components specified by equation (4), into a part that is explained by education and experience and a part that remained unexplained. In each panel, the first row reproduces the racial gaps reported in Table 7 (with reweighting). The third row is the component of the gap that is not explained by education and experience ("unexplained component"). It is estimated by the coefficient on a white dummy in a regression model for the log hourly wage (column 1), the person effect (column 2), the establishment effect (column 3), or the covariate index (column 4) that controls for education and experience as in Table 3, and is fitted by WLS to accomodate the reweighting adjustment (so that nonwhites have the same distribution across micro-regions as whites of the same gender). The second row is the component of the gap that is explained by education and experience, the difference between the overall gap and the unexplained component. The share of the overall gap that is explained by education and experience is reported in parenthesis.

Table D9: Decomposition of the Racial Gap in Establishment Effects (Using Octiles of Person Effect Distributions)

	Overall racial wage gap	Means of establishment effect		Gap in mean establishment effect	Decomposition of gap in establishment effects		Decomposition of sorting effect (using octiles of person effects)	
		Whites	Nonwhites		Sorting	Relative wage-setting	Skill-based sorting	Residual sorting
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Males								
All education	0.186	0.220	0.175	0.044 (0.24)	0.035 (0.19)	0.009 (0.05)	0.025 (0.14)	0.010 (0.05)
No high school	0.052	0.150	0.140	0.010 (0.19)	0.004 (0.07)	0.006 (0.12)	0.003 (0.06)	0.001 (0.01)
Completed high school	0.262	0.267	0.202	0.065 (0.25)	0.054 (0.21)	0.011 (0.04)	0.036 (0.14)	0.018 (0.07)
B. Females								
All education	0.261	0.156	0.085	0.071 (0.27)	0.056 (0.21)	0.015 (0.06)	0.035 (0.14)	0.020 (0.08)
No high school	0.046	0.043	0.025	0.018 (0.40)	0.010 (0.22)	0.008 (0.17)	0.003 (0.07)	0.007 (0.15)
Completed high school	0.291	0.191	0.112	0.079 (0.27)	0.062 (0.21)	0.017 (0.06)	0.036 (0.12)	0.026 (0.09)

effect and a sorting effect based on equation (5), as we well as the decomposition of the sorting effect into a skill-based sorting and a residual sorting effect based on equations (7) and (8). As in Table 9, the samples include all person-year observations in the dual-connected set of each gender in the Southeast region. Nonwhite observations are reweighted so that they have the same distribution across micro-regions as whites (of the same gender). Column (1) reproduces the overall racial wage gap from column (1) of Table 7; column (2) and (3) report the average establishment effect for white and nonwhites, respectively, and column (4) calculates the difference, which corresponds to the racial gap in mean establishment effects in column (4) of Tables 7. Column (5) and (6) decompose the gap in column (4) into a sorting and relative wage-setting effect, respectively. Columns (7) and (8) decompose the sorting effect into a skill-based and a residual sorting effect. The only difference with Table 9 is that we use octiles instead of quartiles of the person effect distribution to construct the predicted distribution used for the decomposition in columns (7) and (8), which maintains the age and skill distribution of establishments for each year. Entries in parentheses represent the share of the overall racial wage gap (in column 1) that is explained by the source in the column heading.

Table D10: Robustness of Decompositions to Changes in Samples and Assumptions (All Education)

	By Region:			Treatment of Inconsistent Race		Remove 1st decile workers and estab's.	Adjust Normalization		Alternative normalizing sector: other sector	By period:		By importance of customer interactions	
	South-east	Brazil	North-east	Use first obs. race	Consistent subsample		50% of race gap = white premium	100% of race gap = white premium		2002-2009	2007-2014	Low face-time	High face-time
A. Males													
Racial wage gap	0.186	0.154	0.077	0.196	0.279	0.189	0.186	0.186	0.186	0.205	0.195	0.217	0.192
Sorting	0.035	0.027	0.018	0.043	0.057	0.036	0.035	0.035	0.035	0.041	0.030	0.048	0.039
Skill-based sorting	0.023	0.017	0.007	0.024	0.035	0.023	0.022	0.020	0.024	0.024	0.024	0.036	0.018
Residual sorting	0.012	0.010	0.011	0.018	0.022	0.013	0.014	0.016	0.012	0.017	0.005	0.012	0.020
Relative wage-setting	0.009	0.009	0.004	0.006	0.025	0.009	0.023	0.037	0.007	0.015	0.005	0.006	0.009
Human capital gap	0.142	0.118	0.056	0.148	0.198	0.143	0.128	0.114	0.144	0.149	0.160	0.163	0.144
Share of wage gap													
Sorting	*0.190*	*0.177*	*0.230*	*0.217*	*0.204*	*0.193*	*0.190*	*0.190*	*0.190*	*0.199*	*0.152*	*0.221*	*0.201*
Skill-based sorting	*0.126*	*0.113*	*0.084*	*0.124*	*0.124*	*0.124*	*0.116*	*0.105*	*0.127*	*0.118*	*0.124*	*0.165*	*0.095*
Residual sorting	*0.065*	*0.063*	*0.146*	*0.093*	*0.080*	*0.069*	*0.075*	*0.085*	*0.064*	*0.081*	*0.028*	*0.056*	*0.107*
Relative wage-setting	*0.048*	*0.061*	*0.051*	*0.031*	*0.088*	*0.047*	*0.123*	*0.199*	*0.039*	*0.071*	*0.028*	*0.027*	*0.049*
Human capital gap	*0.761*	*0.763*	*0.719*	*0.752*	*0.708*	*0.760*	*0.686*	*0.611*	*0.771*	*0.730*	*0.820*	*0.752*	*0.750*
B. Females													
Racial wage gap	0.261	0.220	0.106	0.261	0.362	0.261	0.261	0.261	0.261	0.285	0.275	0.259	0.283
Sorting	0.056	0.046	0.022	0.058	0.080	0.056	0.056	0.056	0.056	0.061	0.050	0.059	0.074
Skill-based sorting	0.034	0.027	0.009	0.033	0.049	0.034	0.032	0.031	0.033	0.032	0.032	0.047	0.043
Residual sorting	0.022	0.019	0.014	0.025	0.031	0.023	0.024	0.025	0.022	0.029	0.019	0.012	0.031
Relative wage-setting	0.015	0.013	0.002	0.010	0.025	0.015	0.029	0.043	0.016	0.023	0.010	0.005	0.019
Human capital gap	0.190	0.161	0.082	0.193	0.257	0.190	0.176	0.162	0.189	0.201	0.214	0.196	0.191
Share of wage gap													
Sorting	*0.214*	*0.210*	*0.213*	*0.222*	*0.222*	*0.215*	*0.214*	*0.214*	*0.214*	*0.215*	*0.183*	*0.227*	*0.261*
Skill-based sorting	*0.129*	*0.122*	*0.084*	*0.126*	*0.137*	*0.129*	*0.123*	*0.118*	*0.128*	*0.114*	*0.116*	*0.183*	*0.151*
Residual sorting	*0.086*	*0.088*	*0.129*	*0.096*	*0.085*	*0.086*	*0.091*	*0.097*	*0.086*	*0.102*	*0.067*	*0.044*	*0.109*
Relative wage-setting	*0.058*	*0.058*	*0.015*	*0.039*	*0.068*	*0.058*	*0.112*	*0.165*	*0.062*	*0.080*	*0.037*	*0.018*	*0.067*
Human capital gap	*0.728*	*0.732*	*0.773*	*0.739*	*0.710*	*0.727*	*0.674*	*0.620*	*0.724*	*0.705*	*0.779*	*0.755*	*0.673*

Notes: The table displays the racial wage gap decomposition for the robustness checks and heterogeneity analyses presented in Figure 10. Samples include workers in the dual connected set for each gender. All estimates are formed using weights that reweight nonwhites so as to have the same distribution across micro regions as whites (of the same gender).

Table D11: Robustness of Decompositions to Changes in Samples and Assumptions (Completed High School)

	By Region:			Treatment of Inconsistent Race		Remove 1st decile workers and estab's.	Adjust Normalization		Alternative normalizing sector: other sector	By period:		By importance of customer interactions	
	South-east	Brazil	North-east	Use first obs. race	Consistent subsample		50% of race gap = white premium	100% of race gap = white premium		2002-2009	2007-2014	Low face-time	High face-time
A. Males													
Racial wage gap	0.262	0.220	0.115	0.271	0.348	0.262	0.262	0.262	0.262	0.287	0.262	0.271	0.271
Sorting	0.054	0.043	0.026	0.062	0.070	0.054	0.054	0.054	0.054	0.060	0.043	0.054	0.059
Skill-based sorting	0.035	0.026	0.011	0.036	0.045	0.035	0.033	0.031	0.035	0.036	0.033	0.039	0.033
Residual sorting	0.019	0.017	0.015	0.026	0.025	0.019	0.021	0.023	0.019	0.024	0.010	0.015	0.026
Relative wage-setting	0.011	0.015	0.009	0.010	0.019	0.011	0.025	0.039	0.010	0.013	0.007	0.008	0.012
Human capital gap	0.196	0.162	0.081	0.199	0.259	0.197	0.182	0.168	0.198	0.215	0.212	0.209	0.199
Share of wage gap													
Sorting	*0.206*	*0.197*	*0.223*	*0.228*	*0.202*	*0.207*	*0.206*	*0.206*	*0.206*	*0.208*	*0.165*	*0.198*	*0.218*
Skill-based sorting	*0.133*	*0.120*	*0.093*	*0.132*	*0.129*	*0.133*	*0.126*	*0.118*	*0.134*	*0.126*	*0.127*	*0.144*	*0.122*
Residual sorting	*0.073*	*0.078*	*0.129*	*0.096*	*0.073*	*0.074*	*0.081*	*0.088*	*0.072*	*0.083*	*0.037*	*0.054*	*0.096*
Relative wage-setting	*0.044*	*0.066*	*0.075*	*0.037*	*0.054*	*0.043*	*0.097*	*0.151*	*0.037*	*0.045*	*0.026*	*0.031*	*0.046*
Human capital gap	*0.750*	*0.736*	*0.702*	*0.734*	*0.745*	*0.750*	*0.697*	*0.643*	*0.757*	*0.747*	*0.809*	*0.771*	*0.736*
B. Females													
Racial wage gap	0.291	0.250	0.119	0.292	0.389	0.291	0.291	0.291	0.291	0.313	0.302	0.339	0.306
Sorting	0.062	0.053	0.029	0.064	0.082	0.062	0.062	0.062	0.062	0.067	0.053	0.076	0.078
Skill-based sorting	0.035	0.028	0.011	0.035	0.048	0.035	0.033	0.032	0.035	0.032	0.033	0.056	0.043
Residual sorting	0.027	0.024	0.018	0.029	0.034	0.027	0.028	0.030	0.027	0.034	0.020	0.020	0.035
Relative wage-setting	0.017	0.015	0.002	0.011	0.026	0.017	0.031	0.045	0.019	0.025	0.012	0.009	0.019
Human capital gap	0.212	0.182	0.088	0.217	0.280	0.212	0.198	0.184	0.210	0.221	0.236	0.254	0.208
Share of wage gap													
Sorting	*0.212*	*0.211*	*0.244*	*0.221*	*0.211*	*0.212*	*0.212*	*0.212*	*0.212*	*0.213*	*0.175*	*0.225*	*0.256*
Skill-based sorting	*0.120*	*0.114*	*0.090*	*0.120*	*0.123*	*0.120*	*0.115*	*0.109*	*0.120*	*0.103*	*0.108*	*0.166*	*0.141*
Residual sorting	*0.092*	*0.097*	*0.154*	*0.101*	*0.088*	*0.092*	*0.098*	*0.103*	*0.093*	*0.110*	*0.067*	*0.059*	*0.115*
Relative wage-setting	*0.059*	*0.060*	*0.018*	*0.037*	*0.068*	*0.059*	*0.107*	*0.155*	*0.064*	*0.080*	*0.041*	*0.027*	*0.063*
Human capital gap	*0.729*	*0.730*	*0.738*	*0.743*	*0.721*	*0.729*	*0.681*	*0.633*	*0.724*	*0.707*	*0.783*	*0.748*	*0.681*

Notes: The table displays the racial wage gap decomposition for the robustness checks and heterogeneity analyses presented in Figure 10. Samples include workers in the dual connected set for each gender. All estimates are formed using weights that reweight nonwhites so as to have the same distribution across micro regions as whites (of the same gender).

Table D12: Descriptive Statistics Comparing the Analysis Samples, the Largest Connected Sets, and the Dual-Connected Sets (2002-2009)

	Analysis samples (all valid observations)				Largest connected sets (by race-gender group)				Dual-connected sets (by gender)			
	White male	Non-white male	White female	Non-white female	White male	Non-white male	White female	Non-white female	White male	Non-white male	White female	Non-white female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Age												
Mean age	36.1	35.9	35.2	35.4	36.1	35.9	35.1	35.3	36.1	35.9	35.2	35.4
Share ≤ 30 years old	0.31	0.32	0.35	0.34	0.31	0.32	0.36	0.34	0.31	0.32	0.35	0.34
Share ≥ 50 years old	0.07	0.07	0.06	0.06	0.07	0.07	0.05	0.05	0.07	0.07	0.05	0.06
Education												
Mean years of schooling	9.2	8.2	10.5	9.5	9.3	8.3	10.6	9.5	9.3	8.3	10.7	9.6
Share completed high school	0.49	0.37	0.67	0.55	0.50	0.38	0.68	0.57	0.51	0.39	0.69	0.57
Share completed college	0.10	0.04	0.20	0.09	0.11	0.04	0.21	0.09	0.11	0.04	0.23	0.10
Wages and Hours												
Mean log hourly wage (R$)	1.86	1.65	1.70	1.43	1.89	1.68	1.75	1.47	1.95	1.69	1.82	1.49
(standard dev.)	(0.71)	(0.60)	(0.73)	(0.58)	(0.71)	(0.60)	(0.74)	(0.59)	(0.73)	(0.60)	(0.76)	(0.60)
Mean monthly hours	188.1	188.5	184.0	185.3	188.0	188.3	183.4	184.6	187.7	188.3	181.9	184.3
(standard dev.)	(11.90)	(11.12)	(19.23)	(16.63)	(11.94)	(11.19)	(19.68)	(16.92)	(12.14)	(11.22)	(20.23)	(17.11)
Share full-time	0.99	0.99	0.96	0.97	0.99	0.99	0.96	0.97	0.99	0.99	0.95	0.97
Establishment Characteristics												
Mean establishment size	422	514	421	573	454	578	473	718	577	608	703	764
Share females at establishment	0.23	0.21	0.64	0.63	0.23	0.21	0.63	0.62	0.23	0.21	0.61	0.62
(leave-out mean)	0.26	0.23	0.58	0.59	0.25	0.22	0.60	0.60	0.24	0.22	0.60	0.60
Share whites at establishment	0.81	0.50	0.83	0.51	0.80	0.51	0.82	0.53	0.77	0.53	0.78	0.55
(leave-out mean)	0.79	0.56	0.81	0.58	0.79	0.55	0.81	0.57	0.75	0.56	0.77	0.58
Sample Sizes												
No. person-year obs.	24,075,008	9,611,684	16,040,182	4,788,756	22,369,983	8,532,716	14,245,257	3,793,803	17,406,539	8,109,848	9,311,910	3,559,887
No. persons	6,571,244	2,748,752	4,619,317	1,468,558	6,019,751	2,404,030	3,997,264	1,122,407	4,959,677	2,332,390	2,824,707	1,078,389
No. establishments	1,134,130	645,649	1,046,157	480,913	774,065	392,415	630,365	225,672	304,838	304,838	170,989	170,989

Notes: The table displays descriptive statistics by race-gender group in three samples based on RAIS 2002-2009. The analysis samples in columns (1)-(4) include nonfarm private-sector formal employees in the Southeast region, age 25 to 54, with potential labor market experience of at least 1 year, tenure of at least 1 month, and non-missing data on race, gender, education, wage, and hours worked, who are employed on December 31st of each year with an open-ended contract in which they are paid on a monthly basis. The entire history of an individual is dropped if that individual reports earning a wage below the minimum wage or above the 99th percentile in a given year, as well as in the case of extreme wage changes between consecutive years. The analysis sample is restricted to the largest connected set for each race-gender group in columns (5)-(8), i.e., the largest set of establishments that are linked by worker mobility. The analysis sample is further restricted to the dual-connected set of each gender in columns (9)-(12), i.e., the set of establishments in the largest connected set for both whites and nonwhites of that gender. All statistics are calculated across person-year observations.

Table D13: Descriptive Statistics Comparing the Analysis Samples, the Largest Connected Sets, and the Dual-Connected Sets (2007-2014)

	Analysis samples (all valid observations)				Largest connected sets (by race-gender group)				Dual-connected sets (by gender)			
	White male	Non-white male	White female	Non-white female	White male	Non-white male	White female	Non-white female	White male	Non-white male	White female	Non-white female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Age												
Mean age	36.6	36.4	35.7	35.6	36.6	36.3	35.5	35.5	36.6	36.3	35.5	35.5
Share ≤ 30 years old	0.29	0.30	0.33	0.33	0.29	0.30	0.33	0.33	0.29	0.30	0.33	0.33
Share ≥ 50 years old	0.09	0.08	0.07	0.06	0.09	0.08	0.06	0.06	0.09	0.08	0.06	0.06
Education												
Mean years of schooling	9.6	8.7	10.8	9.8	9.6	8.7	10.8	9.9	9.7	8.7	10.9	9.9
Share completed high school	0.55	0.44	0.72	0.63	0.56	0.45	0.73	0.63	0.57	0.45	0.74	0.64
Share completed college	0.11	0.04	0.19	0.08	0.11	0.04	0.20	0.09	0.11	0.04	0.22	0.09
Wages and Hours												
Mean log hourly wage (R$)	2.03	1.85	1.85	1.62	2.05	1.87	1.88	1.64	2.11	1.88	1.95	1.65
(standard dev.)	(0.68)	(0.57)	(0.69)	(0.54)	(0.69)	(0.58)	(0.70)	(0.55)	(0.70)	(0.58)	(0.72)	(0.55)
Mean monthly hours	188.1	188.4	184.4	185.2	188.0	188.3	184.0	184.7	187.7	188.2	182.6	184.4
(standard dev.)	(11.36)	(10.57)	(17.87)	(15.95)	(11.38)	(10.61)	(18.18)	(16.28)	(11.57)	(10.63)	(19.01)	(16.48)
Share full-time	0.99	0.99	0.96	0.98	0.99	0.99	0.96	0.97	0.99	0.99	0.96	0.97
Establishment Characteristics												
Mean establishment size	523	665	548	782	553	723	594	899	677	755	809	948
Share females at establishment	0.25	0.23	0.65	0.65	0.25	0.23	0.64	0.64	0.25	0.23	0.63	0.63
(leave-out mean)	0.28	0.25	0.60	0.61	0.27	0.24	0.61	0.61	0.26	0.24	0.61	0.62
Share whites at establishment	0.79	0.50	0.81	0.50	0.79	0.51	0.81	0.51	0.75	0.52	0.77	0.53
(leave-out mean)	0.77	0.55	0.79	0.57	0.77	0.54	0.79	0.55	0.74	0.55	0.75	0.56
Sample Sizes												
No. person-year obs.	30,187,495	13,112,908	22,154,540	7,662,857	28,514,275	12,056,128	20,416,320	6,648,507	23,184,341	11,542,387	14,785,049	6,299,182
No. persons	7,847,841	3,589,664	6,078,282	2,288,125	7,320,848	3,257,489	5,495,282	1,933,699	6,293,717	3,179,095	4,301,305	1,873,283
No. establishments	1,369,430	801,301	1,331,083	664,801	996,665	542,371	908,737	389,053	427,946	427,946	301,629	301,629

Notes: The table displays descriptive statistics by race-gender group in three samples based on RAIS 2007-2014. The analysis samples in columns (1)-(4) include nonfarm private-sector formal employees in the Southeast region, age 25 to 54, with potential labor market experience of at least 1 year, tenure of at least 1 month, and non-missing data on race, gender, education, wage, and hours worked, who are employed on December 31st of each year with an open-ended contract in which they are paid on a monthly basis. The entire history of an individual is dropped if that individual reports earning a wage below the minimum wage or above the 99th percentile in a given year, as well as in the case of extreme wage changes between consecutive years. The analysis sample is restricted to the largest connected set for each race-gender group in columns (5)-(8), i.e., the largest set of establishments that are linked by worker mobility. The analysis sample is further restricted to the dual-connected set of each gender in columns (9)-(12), i.e., the set of establishments in the largest connected set for both whites and nonwhites of that gender. All statistics are calculated across person-year observations.

Table D14: Decomposition Statistics by Sector Group (All Education)

	Overall racial wage gap	White establishment effects	Observed share non-white	Predicted share non-white	Unexplained under-representation of nonwhites	Relative wage-setting
	(1)	(2)	(3)	(4)	(5)	(6)
A. Males						
Education	0.263	0.212	0.254	0.293	0.040	0.037
Banking and finance	0.161	0.697	0.166	0.251	0.085	0.021
Other services and organizations	0.155	0.229	0.299	0.297	-0.002	0.028
Electricity/gas/utilities	0.137	0.582	0.216	0.246	0.030	0.043
Manufacturing	0.133	0.307	0.240	0.249	0.009	0.016
Farming and fishing	0.100	0.133	0.313	0.273	-0.041	0.010
Trade	0.093	0.105	0.296	0.310	0.014	0.015
Real estate	0.085	0.190	0.335	0.322	-0.013	0.024
Construction	0.081	0.158	0.392	0.330	-0.062	0.015
Extractive industries	0.072	0.391	0.378	0.382	0.003	0.008
Transportation and communication	0.064	0.208	0.307	0.307	0.000	0.020
Accomodation and food	0.056	0.015	0.304	0.353	0.049	0.001
B. Females						
Education	0.261	0.091	0.182	0.229	0.047	0.050
Farming and fishing	0.204	0.120	0.241	0.221	-0.019	0.049
Electricity/gas/utilities	0.189	0.461	0.148	0.198	0.050	0.056
Manufacturing	0.165	0.246	0.199	0.213	0.014	0.022
Other services and organizations	0.161	0.167	0.238	0.227	-0.010	0.027
Banking and finance	0.156	0.575	0.146	0.202	0.056	0.041
Construction	0.154	0.201	0.264	0.248	-0.016	0.045
Transportation and communication	0.144	0.238	0.217	0.243	0.026	0.055
Real estate	0.137	0.109	0.274	0.257	-0.016	0.043
Accomodation and food	0.128	0.016	0.305	0.289	-0.016	-0.001
Extractive industries	0.115	0.501	0.282	0.293	0.011	0.044
Trade	0.095	0.108	0.288	0.280	-0.008	0.021

Notes: The table displays key decomposition statistics by broad sector of activities. Columns (1), (2), and (6) are analogous to those same columns in Table 9 but for a specific sector. Column (3) is the average observed share of nonwhites in the sector, similar to the black line in Figure 7. Column (4) is the average predicted share of nonwhites in the sector group preserving establishment skill composition, akin to the blue line in Figure 7. The difference between the predicted and the observed shares shown in column (5) provides a measure of the under-representation of nonwhites in the sector. The samples include all person-year observations in the dual-connected set of each gender in the Southeast region. Non-white observations in each sector are reweighted so as to have the same distribution across micro-regions as whites (of the same gender) in the same sector.

Table D15: Decomposition Statistics by Sector Group (Completed High School)

	Overall racial wage gap	White establishment effects	Observed share nonwhite	Predicted share nonwhite	Unexplained under-representation	Relative wage-setting
	(1)	(2)	(3)	(4)	(5)	(6)
A. Males						
Education	0.263	0.190	0.220	0.255	0.035	0.034
Other services and organizations	0.184	0.236	0.258	0.258	0.000	0.028
Manufacturing	0.173	0.375	0.204	0.212	0.008	0.019
Farming and fishing	0.163	0.207	0.240	0.222	-0.019	0.018
Banking and finance	0.158	0.667	0.161	0.224	0.063	0.016
Electricity/gas/utilities	0.145	0.559	0.203	0.212	0.008	0.038
Construction	0.137	0.179	0.333	0.284	-0.049	0.019
Trade	0.123	0.127	0.277	0.280	0.003	0.021
Real estate	0.116	0.182	0.291	0.281	-0.010	0.028
Extractive industries	0.095	0.540	0.391	0.387	-0.004	0.012
Accomodation and food	0.094	0.028	0.308	0.321	0.013	0.002
Transportation and communication	0.093	0.229	0.284	0.282	-0.002	0.020
B. Females						
Education	0.261	0.091	0.182	0.229	0.047	0.050
Farming and fishing	0.204	0.120	0.241	0.221	-0.019	0.049
Electricity/gas/utilities	0.189	0.461	0.148	0.198	0.050	0.056
Manufacturing	0.165	0.246	0.199	0.213	0.014	0.022
Other services and organizations	0.161	0.167	0.238	0.227	-0.010	0.027
Banking and finance	0.156	0.575	0.146	0.202	0.056	0.041
Construction	0.154	0.201	0.264	0.248	-0.016	0.045
Transportation and communication	0.144	0.238	0.217	0.243	0.026	0.055
Real estate	0.137	0.109	0.274	0.257	-0.016	0.043
Accomodation and food	0.128	0.016	0.305	0.289	-0.016	-0.001
Extractive industries	0.115	0.501	0.282	0.293	0.011	0.044
Trade	0.095	0.108	0.288	0.280	-0.008	0.021

Notes: The table displays key decomposition statistics by broad sector of activities. Columns (1), (2), and (6) are analogous to those same columns in Table 9 but for a specific sector. Column (3) is the average observed share of nonwhites in the sector, similar to the black line in Figure 7. Column (4) is the average predicted share of nonwhites in the sector group preserving establishment skill composition, akin to the blue line in Figure 7. The difference between the predicted and the observed shares shown in column (5) provides a measure of the under-representation of nonwhites in the sector. The samples include all person-year observations in the dual-connected set of each gender in the Southeast region with at least a high school education. Non-white observations in each sector are reweighted so as to have the same distribution across micro-regions as whites (of the same gender) in the same sector.