Assortative Matching or Exclusionary Hiring? The Impact of Employment and Pay Policies on Racial Wage Differences in Brazil

By François Gerard, Lorenzo Lagos,
Edson Severnini and David Card*

We measure the effects of firm policies on racial pay differences in Brazil. Nonwhites are less likely to be hired by high-wage firms, explaining about 20% of the racial wage gap for both genders. Firm-specific pay premiums for nonwhites are also compressed relative to whites, contributing another 5% for that gap. A counterfactual analysis reveals that about two-thirds of the under-representation of nonwhites at higher-wage firms is explained by race-neutral skill-based sorting. Non-skill-based sorting and differential wage setting are largest for college-educated workers, suggesting that the allocative costs of discriminatory hiring and pay policies may be relatively large in Brazil.

In many countries around the world nonwhites earn less than whites.[1] Traditionally, economic studies of racial pay disparities have built on the framework of Becker (1957), who assumed that each worker has a *market-determined* wage that is independent of the choices of any single employer.[2] A growing body of work, however, suggests that firm-specific hiring and pay policies also contribute to between-group wage differentials.[3] When employers have wage-setting power, the racial pay gap will depend partly on whether higher-paying firms differentially hire whites versus nonwhites—a between-firm *sorting effect*—and partly on the pay premiums that firms offer to different race groups—a *relative wage-setting*

* Gerard: Queen Mary University of London and CEPR, f.gerard@qmul.ac.uk. Lagos: Princeton University and IZA, llagos@princeton.edu. Severnini: Carnegie Mellon University and IZA, edsons@andrew.cmu.edu. Card: UC Berkeley and NBER, card@berkeley.edu. We are grateful to three referees, seminar participants at Bocconi University, Carnegie Mellon University, Central European University, Columbia University, Cornell University, New York University, Stockholm University, University of Sao Paulo, University of Zurich, Yale University, and conference participants at the NBER Labor Studies Spring Meetings, and the Annual Meetings of the Brazilian Econometric Society for many helpful comments and suggestions. We also thank Raffaele Saggio for his generous help with the KSS codes, and Dario Fonseca and Samira Noronha for their research assistance and help with the case law on labor market discrimination in Brazil. This article is part of a project that obtained Institutional Review Board (IRB) approval from Columbia University (IRB-AAAQ8871).

[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., Canada, and France see Blackaby, Leslie and Murphy (2002), Pendakur and Pendakur (1998, 2002), and Longhi (2017), respectively.

[2]See Charles and Guryan (2008, 2011) for recent analyses that build directly on Becker's model, and Hirata and Soares (2020) for an application in Brazil.

[3]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.

effect.

Findings from four complementary strands of literature suggest that both mechanisms may be important. Randomized audit studies in several countries show that employer callback rates are lower for minority job applicants, implying that some firms set a higher bar for nonwhite candidates, or avoid hiring minorities altogether.[4] Observational studies show that white managers are less likely to hire and retain minority applicants (e.g., Giuliano, Leonard and Levine, 2009, 2011; Giuliano and Ransom, 2013; Aslund, Hensvik and Skans, 2014), and that workplaces are substantially segregated by race and ethnicity (Hellerstein and Neumark, 2008; Hellerstein, Neumark and McInerney, 2008; Aslund and Skans, 2010; Glitz, 2014). Finally, studies of occupational stratification (e.g., Penner, 2008) suggest that employers assign nonwhite workers to lower-paid occupations, accounting for some of the racial pay gap within firms.

Nevertheless, it is unclear how much these patterns contribute to average wage gaps between whites and nonwhites (see Lang and Lehmann, 2012). To quantify the impacts, we use linked employer-employee data from 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.

Concerns over differential hiring and pay setting are particularly relevant in Brazil, where close to one-half of all workers identify as nonwhite. 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 are as large as those in the U.S., despite differences in the historical background and legal setting in the two countries.[5] Several studies have shown that racial wage gaps are particularly large for higher skilled workers, and pointed to the scarcity of nonwhites in high-paying industries and occupations as evidence of exclusionary hiring policies (Soares, 2000; Henriques, 2001; Campante, Crespo and Leite, 2004; Chadarevian, 2011; Mariano et al., 2018).

In addition to the salience of race in the Brazilian labor market, an advantage of our setting is the availability of rich administrative data—covering the entire formal sector—that include information on education, hours of work, occupation,

[4]Zschirnt and Ruedin (2016) summarize 36 studies in OECD countries: they find a median callback 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). An 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.

[5]One difference is that nonwhites in Brazil are traditionally classified into blacks (roughly 10% of the population) and mixed race individuals (40%), though previous studies find that wage differences between these groups are very small. Confirming this pattern, we follow the literature in distinguishing only two groups: whites and nonwhites. Prior studies of wage differences include Cavalieri and Fernandes (1998), Arcard and d'Hombres (2004), Matos and Machado (2006), Garcia, Nopo and Salardi (2009), Bailey, Loveman and Muniz (2013), and Reis and Crespo (2015).

and industry. These data allow us to address concerns over differences in education and hours between white and nonwhites, to document the role of occupation-specific skills in observed workplace segregation, and to explore heterogeneity by industry. To address concerns over potential selectivity into the formal sector, we use national survey data that include information on both formal and informal jobs. Reassuringly, we find that formality rates are quite similar for whites and nonwhites of each gender, and virtually identical once we condition on location, age, and education. Unexplained racial pay gaps are also nearly identical—and large—in the two sectors.

An important consideration for the structure of wages in Brazil is the minimum wage. Simple comparisons across states suggest that the federally-legislated wage floor exerts strong upward pressure on wages that narrows the gap between whites and nonwhites, particularly in poorer regions of the country.[6] Our main analysis therefore focuses on the Southeast region of the country, where the ratio of the minimum wage to the median wage is comparable to other developing and developed countries (see, for example, Kristensen and Cunningham, 2006; Dube, 2019). We also present results separately for lower-educated workers, who are the most impacted by the minimum wage, and higher-educated workers, who are less impacted.

Consistent with findings from the U.S., Germany, and other countries, and with previous work on Brazil by Lavetti and Schmutte (2016) and Alvarez et al. (2018), we find that firms play a large role in explaining the variation in wages for all race-gender groups. The differences in the wage premiums paid by different establishments, together with a strong pattern of positive assortative matching between workers and establishments *within* each race-gender group, explains about 30%-40% of the variation in wages for all groups (adjusting for sampling errors in the estimated worker and firm effects). Importantly, since nonwhites in Brazil tend to have lower education and lower average earnings irrespective of where they work, such positive assortative matching would lead to a lower share of non-white workers at higher-paying establishments, even in the absence of discriminatory employment policies.

To assess how much lower, 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.[7] 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

[6]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 (Forthcoming) and Bailey, DiNardo and Stuart (Forthcoming) 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.

[7]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 measured by the person effects in our two-way models.

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 12%-13% of the white-nonwhite wage gap, is explained by race-neutral "skill-based" sorting. The remainder, incorporating discriminatory hiring and retention policies, accounts for about 6%-7% 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 whites—a pattern that is consistent with monopsonistic wage setting and lower elasticities of firm-specific supply (Barth and Dale-Olsen, 2009; Card et al., 2018), or with lower bargaining power (Babcock and Laschever, 2003; Manning, 2011), for nonwhites. These lower average premiums explain another 4-5% of the overall racial wage gap.

Our main findings are summarized in Figure 1. It shows the mean white-nonwhite wage gaps by gender and education group, and the shares of these gaps attributable to skill-based sorting, unexplained or residual sorting, and differential wage setting. The wage losses associated with unexplained sorting and differential wage setting are particularly large for college-educated workers, suggesting that the allocative costs of these policies may be important. Moreover, as we show later in the paper, the relative importance of these two components is even larger (for all education groups) controlling for observed occupational differences between whites and nonwhites.

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.[8] 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 are large racial gaps in education levels, and different race groups are differentially concentrated in higher- and lower-wage areas. 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, particularly at the top of the skill distribution.

[8]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 information on race.

I. Background and Data

A. Institutional Setting

In contrast to the rigid black-white divide enforced in post-reconstruction U.S., a more fluid notion of race emerged in Brazil after the end of slavery, marked 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 (see, e.g., Skidmore, 1974; Andrews, 1992; Marx, 1998; Telles, 2004; Travassos and Williams, 2004). Although racial differences remained highly salient in Brazil throughout the following century (Skidmore, 1992), legal intervention in the labor market emerged relatively late.[9] Indeed, it was only with the adoption of the 1988 Constitution and the passage of laws in 1989 and 1995 that racial discrimination in employment and pay setting became illegal in Brazil. A review of recent case law suggests that even into the early 2000's most claims of discrimination were dismissed (Equal Rights Trust, 2009).

The recent 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 enacted in 2010. These changes suggest that employer policies regarding employment and wage setting may have evolved during the 13 years (2002-2014) included in our sample. We explore this possibility as part of our robustness analysis later in the paper.

B. Initial Descriptive Analysis

Our main analysis uses administrative records for formal-sector workers from the Southeast region of Brazil. To provide a context for these data and address concerns about selection into formality, we begin by examining data from the *Pesquisa Nacional por Amostra de Domicílios* (PNAD), a yearly national survey that collects information on labor market outcomes for both formal and informal workers. For consistency with our main analysis, we pool the 2002-2014 PNAD surveys (IBGE, 2016*b*), and limit attention to men and women ages 25-54 with at least one year of potential labor market experience (based on age and years of schooling).[10] We present data both for Brazil as a whole and for the Southeast region, which includes the states of Espírito Santo, Minas Gerais, Rio de Janeiro, and São Paulo, and is the largest and richest region of the country. Figure C.1 presents a map showing Brazil's regions and micro-regions (our definition of local labor markets), as well as their racial composition based on the Brazilian Population Census of 2010 (IBGE, 2016*a*).

[9]The so-called Afonso Arinos Law, passed in 1951, allowed individual plaintiffs to sue for racial discrimination, but is widely believed to have had at most a symbolic effect (Campos, 2015).

[10]Ferreira, Firpo and Messina (2014) present an analysis of PNAD data from 1995 to 2012 and document trends in returns to education, racial wage gaps, and overall wage inequality.

As shown in the first row of Panels A and B in Table 1, about 50% of working-age Brazilian men and women are white ("branco"), 42% are mixed race ("pardo"), and 8% are black ("preto"). In the Southeast region the share of whites is higher (57%) while the share of mixed race individuals is lower (33%). Nationally and in the Southeast region these three main groups account for about 99% of the population, with Asians and indigenous groups making up the remainder.

Given our focus on firm-specific hiring and wage setting policies, we limit our attention to private-sector employees—i.e., those who are neither self-employed nor working in the public sector. The second row of each panel in Table 1 shows that 41% of men and 21% of women are in this class in Brazil. The private-sector employment rate of men differs by only a few percentage points across the three main race groups, but is more variable for women (25% for whites, 20% for blacks, and 17% for mixed-race women). In the Southeast, private-sector employment rates of both genders are higher and also less variable across race groups. This similarity implies that there is little room for differential selection bias caused by differences in the fractions of each race group observed in private-sector employment, particularly in the Southeast.

On average, 45% of Brazilian men employed in the private sector during our sample period had completed high school, with a higher rate for whites (53%) than either black or mixed-race men (37-38%). Employed women are better educated than employed men, with a 65% high school completion rate. Again, however, there are large racial gaps, with a completion rate of 69% for whites versus 58-59% for nonwhites. Average education levels are higher in the Southeast region but the gaps between race groups are similar to those at the national level.

Although 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 the gaps in other Latin American countries. In Chile, the gap in high school completion 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 Spanish speakers and indigenous-language speakers is about 23 ppts. (Esteve and López-Ruiz, 2010).

Looking next at the mean log hourly wage statistics presented in Table 1, three important facts stand out. First, white workers of both genders earn about 30-35% more than nonwhites. Second, wage levels are more than 10 log points higher in the Southeast than in the country as a whole, but the racial wage gaps remain similar. Third, as is true for education, mean wages of mixed-race and black workers are within a few percentage points of each other for both men and women.

In Table D.1 we present a more detailed analysis of wage differences between the three main race groups, focusing on two questions: (1) can we combine mixed-race and black individuals into a single nonwhite group? (2) how large are the racial wage gaps conditioning on education and potential labor market experience? Consistent with many previous studies (e.g., Oliveira, Porcaro and Araújo, 1981; Silva, 1978, 1980, 1985), we find that mixed-race and black workers receive very similar average wages whether we condition on education and experience or not. In the remainder of our analysis we therefore pool these two groups into a single nonwhite group. As expected given the differences in education between whites and nonwhites, we also find that racial wage gaps are reduced from 27%-30% when we only control for year and state fixed effects to 11%-13% when we add controls for education and experience. Interestingly, the magnitudes of these unexplained wage gaps are quite similar in the Southeast.

C. *Informality*

Like other Latin American countries, Brazil has a relatively large informal sector. As shown in the last row of each panel in Table 1, only about 80% of private-sector employees report having a valid "carteira de trabalho," which indicates that they are formally employed in Brazil. Importantly, however, the rates of formality are quite similar across race groups, particularly in the Southeast region. Again, this similarity suggests that there is relatively little scope for differential selection into formality to drive differences in racial wage gaps within the formal sector.

To assess this issue more formally, we report two additional analyses in Table D.2. First, we estimated simple linear probability models for the incidence of formality using the full set of controls used in Table D.1. These models show precisely estimated *zero effects* of nonwhite race on the likelihood of formality. Second, we compared the size of the unexplained white-nonwhite wage gaps (using the same controls) based on samples that include all private-sector employees and only those in the formal sector. Consistent with the zero effect of race on formality, the estimated wage gaps are virtually identical in the two samples (at 11%). Given these findings, we believe that conditioning on formality has little impact on the magnitude of racial wage differences among private-sector employees in Brazil.

D. *Impact of the Minimum Wage*

Brazil's minimum wage was set at a relatively high level during our sample period. The ratio of the minimum wage to the median wage for all private-sector employees—a standard index of the potential impact of the minimum wage—rose from 58% in 2002 to 70% in 2006, and remained relatively constant over the following 8 years.[11] The impact of the high minimum wage is shown visually in Figure 2. It displays density plots of log wages normalized to the minimum wage—i.e., $\log(w/min)$—for white and nonwhite workers of each gender in all Brazil (upper panels) and in the Southeast (lower panels) based on our pooled 2002-2014 PNAD samples. The graphs in the upper panels show large spikes

[11] 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 (Dube, 2014).

at the relative wage of 1, particularly for women, suggesting that the minimum wage has pushed up the lower tail of wages. For both genders, the spike is also larger for nonwhites than for whites, suggesting that the minimum wage helps to compress the white-nonwhite wage gap.

Figure 3 provides further insight into this compression effect. It plots the unexplained white-nonwhite wage gap by gender (using the same controls as above) for each of the 27 states of Brazil, against the ratio of the minimum wage to the median wage of white workers in the state. The wide variation in wage levels across states is illustrated by the range of the relative minimum wage (on the x-axis), which varies from 0.35 to 0.80 for men and from 0.45 to 0.90 for women. The four states of the Southeast region, which have relatively higher wage levels, are located near the bottom of the range in each panel (solid markers). The scatter of points for both genders suggests a fairly strong compression effect: the estimated slopes suggest that white-nonwhite wage gaps would be about 3 ppts. higher if the ratio of the minimum wage to the median wage were at its level in the highest-wage states (i.e., 0.35 for men and 0.45 for women) rather than at its level in an average state.

The more modest impact of the minimum wage in the Southeast is confirmed by the density plots in the lower panels of Figure 2. Relative to the national density plots in the upper panels, the spike at the minimum wage is smaller and the "distortion" of the lower tail of the distribution relative to the upper tail is less severe.

Given these potential effects of the minimum wage on the racial wage gaps, we follow two strategies. First, we focus on the Southeast for our main analysis, but report results for Brazil as a whole in the robustness section at the end of the paper. Second, we present results for all workers and by education group separately. Density plots reported in Gerard et al. (2018) show that the impacts of the minimum wage appear to be relatively small for workers with at least a high school education in the Southeast. Thus, our findings for this group arguably give a clearer picture of what could be expected in the absence of a binding minimum wage. As shown in Table D.2, the unexplained racial wage gaps are higher in this group (14%-17%).

E. Overview of RAIS

To estimate the impacts of firm policies on racial wage gaps, we use the *Relação Anual de Informações Sociais* (RAIS), a longitudinal dataset that provides nearly universal coverage of formal jobs in Brazil (MTE, 2015). Firms report annual information to RAIS on all employees who were on the payroll in the previous year, including their hiring and separation dates, monthly earnings in December, contracted hours, education, occupation, and race.[12] Race is classified into the same categories used in PNAD, but is only available after 2002. Hence, we use

[12]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

the RAIS files from 2002 to 2014 (the last year to which we have access for this study).

To construct an hourly wage, we use information on contracted monthly hours and monthly earnings in December of each year, restricting attention to individuals who worked for their employer for the full month.[13] This wage measure is similar to the one in PNAD, which also measures earnings and hours for a cross-section of jobs in the survey month. Finally, we exclude farm workers, those outside the 25-54 age range, workers on temporary contracts, the small fraction of workers who are not paid on a monthly basis (the standard pay period in Brazil), and those with very low or very high wages (see details in *Appendix A*).

An issue with RAIS is that race can be recorded differently by different employers. A worker in the Southeast whose modal race is white, for example, has their race recorded as mixed-race or black about 8% of the time, while one whose modal race is either mixed or black has their race recorded as white about 12% of the time.[14] 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. We also evaluate the robustness of our results to the coding of race in Section VI.

Estimates of racial wage gaps in the RAIS sample are similar to those obtained in the PNAD sample above. As shown in Table D.2, the white-nonwhite wage gap for female workers in the Southeast (controlling again for year and state fixed effects, education, and potential labor market experience) is 9% in the RAIS sample compared to 11% in the PNAD sample.[15] For male workers, the parallel gaps are 7% and 11%, respectively.[16]

II. Econometric Framework

A. 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

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") and social security benefits. Compliance with the mandatory reporting requirements is high because of large penalties when the data are late or incomplete.

[13]For the small fraction of workers with more than one job in December, we first select the job with the highest contracted hours, then the job with the highest wage; we break the few remaining ties by selecting one job at random. RAIS does not report actual hours worked, but we found no racial gap in hours worked reported by formal employees in PNAD (see Table D.3).

[14]As shown by Cornwall, Rivera and Schmutte (2017), the inconsistent reporting of race is not purely random: workers moving to better-paying jobs are more likely to become white in the data (and vice-versa).

[15]For consistency with the analysis below, these estimates are obtained using the sample of workers employed at establishments in 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.

[16]The smaller magnitude of the white-nonwhite wage gaps in the RAIS sample might be a result of measurement error in the reporting of race by employers. In PNAD, race is self-reported instead.

December of year t (y_{git}) is generated by a model of the form:

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

where α_{gi} is a person fixed effect representing the fully portable component of earnings capacity of individual i, X_{git} is a set of time varying controls (including a polynomial of age and year fixed effects), ψ^g_j is a *wage premium* paid at establishment j to workers in group g, $J(g,i,t)$ is an index function indicating the workplace for worker i in group g in year t, and ε_{git} is an error component capturing all other factors.[17] Note that the pay premium is allowed to vary by race and gender, but is constant within a group. We present some evidence supporting this assumption in Section IV below.

Assuming that the conditional expectation of the error term ε_{git} in equation (1) is independent of the job history of the worker—a so-called "exogenous mobility" assumption—OLS estimation will yield unbiased estimates of the establishment wage premiums in equation (1). Although this assumption has been widely criticized, a series of specification checks developed by Card, Heining and Kline (2013) and implemented in subsequent studies (e.g., Card, Cardoso and Kline, 2016; Macis and Schivardi, 2016; Song et al., 2019) suggest that the earnings changes experienced by job changers in Germany, Italy, Portugal and the U.S. are broadly consistent with exogenous mobility. In the next section, we confirm that this is also the case for Brazil using RAIS data.

Estimates of AKM-style models in many different settings find that the establishment effects contribute significantly to the overall variance of wages, and that employers that pay higher wage premiums tend to hire workers with larger person effects. Such positive assortative matching has implications for wage inequality because the variance of log wages for workers in group g can be decomposed as:

$$\begin{aligned} Var[\ln y_{git}] \quad = \quad & Var[\alpha_{gi}] + Var[\psi^g_{J(g,i,t)}] \\ & + Var[X'_{git}\beta_g] + Var[\varepsilon_{git}] \\ & + 2Cov[\alpha_{gi}, \psi^g_{J(g,i,t)}] + 2Cov[\alpha_{gi}, X'_{git}\beta_g] \\ & + 2Cov[\psi^g_{J(g,i,t)}, X'_{git}\beta_g]. \end{aligned} \tag{2}$$

A positive covariance between the worker and establishment effects magnifies the impacts of both 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.

[17] The person fixed effects and the time-varying $X's$ are not separately identified without a normalizing assumption. Following Card et al. (2018), we assume that in the baseline year $X'_{git}\beta_g = 0$ for 40-year old males and 35-year old females—i.e., the person effects are measured as of age 40 for men and 35 for women, which correspond to the approximate peaks of their experience profiles (see Figure C.2).

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

How do establishment-specific wage premiums contribute to the racial pay gap? Assume that 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. Taking expectations of equation (1), the mean white-nonwhite pay gap can be expressed as:

$$E[\ln y_{Wit}] - E[\ln y_{Nit}] = \alpha_W - \alpha_N \\ + \overline{X}'_W \beta_W - \overline{X}'_N \beta_N + \sum_j \psi_j^W \pi_{Wj} - \sum_j \psi_j^N \pi_{Nj}, \tag{3}$$

where $\alpha_g = E[\alpha_{gi}]$ and $\overline{X}_g = E[X_{git}]$. Assuming for expository purposes that $\overline{X}'_W \beta_W \approx \overline{X}'_N \beta_N$,[18] and adding and subtracting $\sum_j \psi_j^W \pi_{Nj}$, we obtain a simple expression for the net impact of the wage premiums and employment shares at different workplaces:

$$E[\ln y_{Wit}] - E[\ln y_{Nit}] = \alpha_W - \alpha_N \\ + \sum_j \psi_j^W (\pi_{Wj} - \pi_{Nj}) + \sum_j (\psi_j^W - \psi_j^N) \pi_{Nj}. \tag{4}$$

The first term in equation (4) is the difference in the mean person effects for the two groups, i.e., the difference in average wages that would prevail if there were no establishment-specific pay premiums. It can be interpreted as the gap in average productivity (or "skill") between whites and nonwhites (measured at a standardized age). The second term is a weighted average of the difference in employment shares of whites and nonwhites at different workplaces, using as weights the wage premiums for whites at each establishment. This is an estimate of the effect of *differential sorting* of the two groups across workplaces, evaluated under the assumption that nonwhites receive the same premiums as whites—a counterfactual that we believe is natural. The third term is a weighted average of the difference in wage premiums of whites and nonwhites across workplaces, weighted by the employment shares of nonwhites. This is an estimate of the effect of *differential pay-setting* between the two groups, evaluated using the actual distribution of nonwhites across establishments.[19]

[18] 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.

[19] As noted by Oaxaca (1973), there is an alternative decomposition that weights the difference in employment shares by the wage premium of nonwhites and weights the difference in pay premiums by the employment share of whites—counterfactuals that we believe are less natural.

C. Sorting Effect and Assortative Matching

The between-workplace *sorting effect* $\sum_j \psi_j^W(\pi_{Wj} - \pi_{Nj})$ in equation (4) will be positive if white workers are more likely than nonwhites to be employed at high-premium workplaces. There are three reasons to suspect that this will be true.

The first reason is that the white population in Brazil tends to live in areas with higher average wage premiums, potentially reflecting compensating differentials for cost of living differences, as well as other factors. To address this issue, we use a simple reweighting procedure that adjusts the distribution of nonwhites across micro-regions to match the distribution of whites.[20] We then use (4) to decompose the geographically reweighted average difference in wages between whites and nonwhites.

A second explanation is suggested by positive assortative matching between workers and establishments. If whites tend to have higher overall human capital than nonwhites (e.g., given the average education gaps), and higher-paying establishments tend to hire relatively more skilled workers, we would expect to see more whites at these establishments, even in the absence of other (race-based) factors.

To account for skill-biased hiring patterns, we classify individuals 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}, 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 then form the counterfactual *skill-based sorting effect*:

$$\sum_j \psi_j^W(\pi^*_{Wj} - \pi^*_{Nj}). \tag{5}$$

This gives the net effect of (race-neutral) skill-based employment probabilities on the racial wage gap, holding constant the skill distribution at each workplace, the wage premiums paid to white workers, and the racial composition of each skill group in the different local labor markets. We also implement a variant of this procedure in which we narrow the definition of skill groups to include occupation.

A third explanation for an under-representation of nonwhites at higher-paying establishments is discriminatory hiring and/or retention policies. We cannot directly test this explanation. We can, however, calculate the difference between the actual sorting effect and the counterfactual skill-based sorting effect:

[20] A micro-region ("microrregião") is a legally defined geographic entity roughly equivalent to a county. The 557 micro-regions (160 of them in the Southeast) are shown in Figure C.1.

$$\sum_j \psi_j^W(\pi_{Wj} - \pi_{Nj}) - \sum_j \psi_j^W(\pi_{Wj}^* - \pi_{Nj}^*) = \sum_j \psi_j^W[(\pi_{Wj} - \pi_{Wj}^*) - (\pi_{Nj} - \pi_{Nj}^*)]. \tag{6}$$

If higher-premium establishments tend to employ fewer nonwhites than would be expected given the skill composition of their workforce and the nonwhite share in each skill group in their local labor market, this *residual sorting effect* will be positive.

D. *Relative Wage-Setting Effect*

The *relative wage-setting effect* $\sum_j(\psi_j^W - \psi_j^N)\pi_{Nj}$ in equation (4) will be positive if nonwhites tend to receive lower pay premiums than whites at their place of employment. To gain some intuition for the likely pattern of pay premiums, it is useful to consider a monopsonist wage setting model in which each employer sets a race-specific pay premium (see Card et al., 2018, for a review). Under some simplifying assumptions about workers' preferences over jobs and the substitutability between groups of workers, we show in *Appendix B* that an optimal wage-setting policy will be characterized by a set of group-specific pay premiums:

$$\psi_j^g = \delta_g R_j, \tag{7}$$

where R_j is a measure of productivity at establishment j and $\delta_g > 0$ is a group-specific taste parameter that determines the elasticity of supply of group g to different firms. The wage-setting effect will be positive if $\delta_N < \delta_W$, i.e., if nonwhites gain less than whites from moving from a lower-productivity to a higher-productivity workplace. In particular, letting $\gamma = \delta_N/\delta_W$ represent the relative size of the wage premium for nonwhites, the wage-setting effect is

$$\sum_j(\psi_j^W - \psi_j^N)\pi_{Nj} = \frac{1-\gamma}{\gamma} \sum_j \psi_j^N \pi_{Nj}. \tag{8}$$

E. *Normalizing the Pay Premiums*

The worker and establishment effects in equation (4) are not identified without a normalization assumption. In essence, one has to identify a set of establishments that pay a zero premium in order to then decompose the wages of each worker into person and establishment effects. The numerical value of the estimated sorting effect in equation (4) is invariant to the particular normalization adopted for the establishment premiums paid to whites.[21] The value of the wage-setting effect, however, depends on the *relative normalization* of the premiums for whites and nonwhites, since this effect is a weighted average of the difference in premiums received by the two groups. Moreover, any difference between whites and nonwhites in the average premiums paid by the establishments used for normalization will be reflected in the difference in mean person effects. As a result, the normalization also affects the value of the skill-based sorting effect (we use the person effects to define the skill groups).

Our normalization assumes that the pay premiums for both whites and nonwhites are zero in the restaurant industry, i.e., that any wage gap between whites and nonwhites in this industry is due to differences in productivity. This will be true if white and nonwhite restaurant workers are perfect substitutes and restaurants have no wage-setting power. As a robustness check, we consider the alternative assumption that the racial pay gap in the restaurant industry is entirely due to differential pay-setting (i.e., that white and nonwhite restaurant workers are equally productive). Relative to the baseline, this will increase the magnitude of the pay-setting effect, and decrease the difference in mean person effects, by an amount equal to the racial pay gap in the restaurant industry. It will thus increase the overall contribution of firms to the racial wage gap by the same amount, and reduce the role of skill differentials between whites and nonwhites for the sorting effect. We believe that these two alternatives—one assuming that 100% of the racial wage gap in restaurants is due to productivity differences, the other assuming that 100% is due to discriminatory wage setting—represent plausible bounds on the size of these effects.

Fortunately, in both the PNAD and RAIS data, the observed racial wage gaps in the restaurant industry are small (see Table D.4). Specifically, models that control for year and state fixed effects, education, and potential labor market experience show racial wage gaps in our RAIS sample of 2.4 and 2.7 log points for male and female workers, respectively. This means that the range between estimates based on our two normalization assumptions for the contribution of firms to the racial wage gap is 2.4-2.7 log points.

III. RAIS Samples and Specification Tests

A. RAIS Sample Overview

Our main RAIS sample for the Southeast region includes about 8.5 million white men, 4 million nonwhite men, 6.5 million white women, and 2.5 million nonwhite women observed over the period 2002-2014. The characteristics of these groups are presented in the first four columns of Table D.5. On average each male worker contributes about 4.7 observations over the sample period, while each

[21]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.

female contributes about 4.3 observations, yielding a total sample of around 100 million person-year observations across the four groups. Nearly everyone in the sample (98% of women and 100% of men) works full time, with an average of around 185 hours per month ($\approx$ 43 hours per week) for both men and women.

A typical establishment in the Brazilian formal sector is relatively small, but mean establishment size from a worker's perspective is relatively large, about 500 employees for men and 600 for women, and larger for nonwhites than for whites. Establishments are also highly segregated by race and gender. The mean fraction of white coworkers at a white worker's establishment is about 75% for both men and women. The mean fraction for nonwhites, by comparison, is only 55%. Likewise, for both race groups, a typical male works at an establishment where only 25% of workers are female, whereas the coworkers of a typical female are 60% female.

As 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. Characteristics of the largest connected sets for our four race-gender groups are presented in the middle columns of Table D.5. The largest connected set includes 96% of person-year observations for white men, 94% for non-white men, 94% for white women, and 90% for non-white women. Mean wages for all groups are 1-2% higher for observations in these samples, and mean establishment size increases by 4%-11%, but their other characteristics remain very similar.

The decomposition in equation (4) 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. In our analysis below, we therefore focus on workers who are 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). Among males, the dual connected set (summarized in the last columns in Table D.5) includes 91% of person-year observations for nonwhites, but only 80% for whites, reflecting the higher share of whites working at all-white establishments. Among females, the corresponding rates are 86% for nonwhites and 71% for whites.

Narrowing the samples to workers at dual-connected establishments has little impact on the observed means of workers' age or education, but leads to an increase in average wages and mean establishment size for both race groups. This reflects the fact that single-race establishments tend to be smaller and have relatively low pay. More whites are employed at such workplaces, so the rise in wages is larger for whites (about 5 log points for both genders) than for nonwhites (about 1 log points for both genders). Thus, the (unadjusted) white-nonwhite gaps are about 4 log points larger in the dual connected set than in the sample as a whole.

B. *Specification Tests for Exogenous Mobility*

A longstanding concern with the AKM model of wages 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 C.3 shows the results of this analysis using our four race-gender groups in RAIS.[22] The figures exhibit clear step-like patterns for all groups: when workers move to higher-wage establishments, their wages tend to rise; when they move to lower-wage establishments they fall. There is little evidence of differential trends before or after a move for workers who move up or down the job ladder, though there are clearly permanent differences in wages prior to a move that are correlated with the direction of the move.[23] Such differential mobility on the basis of the *permanent* component of wages is consistent with exogenous mobility, since the residual in equation (1) is conditioned on a worker fixed effect.

A sharp prediction of the AKM model under exogenous mobility is that the mean wage changes for movers up the job ladder (e.g., from a set of lower-premium firms to a set of higher-premium firms) will be equal in magnitude but opposite in sign to the mean changes for movers in the opposite direction. Figure C.4 presents some visual evidence in support of this prediction. Here we classify workers who change establishments into 20 quantiles of average coworker wages at their origin workplace, and 20 quantiles of average coworker wages at their destination. We then plot the mean wage changes experienced by movers in each of the 400 origin-destination pairs against the mean change in average coworker wages for the pair. Under symmetry, the points in the graph should lie on a line passing through the origin. For all four groups, this is approximately true. We conclude that a simple AKM model provides a parsimonious and interpretable working model of the wage premiums offered at different establishments for different race-gender groups.

[22]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).

[23]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.

IV. Estimation Results

A. Estimation Results and Implied Variance Decomposition

Table 2 summarizes the results from estimating AKM-style models by race-gender group for workers in the Southeast region. For estimation purposes we use all observations in the largest connected set for each group. In the decompositions in the next section, we then limit attention to workers in the dual-connected sets for each gender.

Panel A presents some simple descriptive statistics summarizing the variation in the estimated worker and firm effects, the fit of the AKM model, and the implied variance decomposition based on equation (2). In general, the two-way fixed effects models fit well, with adjusted R-squared statistics of around 0.90. The implied variance shares show that person effects account for 51%-62% of the variance of wages across the four groups, while the establishment effects account for 20%-23%. Worker and firm effects are also positively correlated *within* each race-gender group, which accounts for another 8%-11% of the overall variance of wages for nonwhites and 18% for whites. Together, the differences in wage premiums paid by different establishments and the strong pattern of positive assortative matching between workers and establishments explain about 30%-40% of the variation in wages for all race-gender groups. 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.

A problem in interpreting plug-in estimates of second moments of regression coefficients is that the estimated coefficients contain sampling errors (e.g., Krueger and Summers, 1988; Andrews et al., 2008). These sampling errors will lead to an *over-statement* of the variances of the person and firm effects in two-way fixed effects models, and an *under-statement* of their covariance (since the sampling errors in the person and firm effects are negatively correlated). Kline, Saggio and Sølvsten (2020)—hereafter, KSS—present an elegant solution to this problem based on a leave-out procedure that corrects for the biases attributable to sampling error. However, this method can only be implemented on the subset of observations in the connected set that remain connected when the data for any one worker is dropped from the estimation sample. Panel B presents the same statistics reported in Panel A, focusing on those capturing the role of firms, but calculated over the subset of observations in the corresponding leave-one-out connected sets. As shown in the bottom row of the table, the leave-one-out connected set includes 56% of all person-year observations for white males, 53% for nonwhite males, 50% for white females, and 42% for nonwhite females.

Despite the loss of observations in the leave-one-out connected sets, the unadjusted estimates of the variance shares of establishment effects are not too different from those in Panel A. They fall by 2-6 ppts., with the largest drop for nonwhite females, the group that has the largest fraction of "weakly connected" establishments. However, the shares attributed to the covariance of person and establishment effects increase almost symmetrically, leaving the overall contribution of firms unchanged.

Finally, in Panel C we present corrected versions of these estimates based on the KSS procedure.[24] Comparisons of these estimates with the naive plug-in estimates in Panel B suggest that, in our setting, failure to account for sampling errors has only a modest impact on the estimated variance shares of the establishment effects, which fall by 2-3 ppts. The downward bias in the estimated share of the variance of wages attributable to the covariance of person and establishment effects is also relatively modest, at about 2 ppts for all race-gender groups.

Thus, in our setting, adjusting for the sampling errors in the estimated worker and firm effects does not change the qualitative conclusions about the importance of firms in wage inequality. The wage premiums paid by different establishments explain about 30%-40% of the variation in wages, in part because of a relatively strong assortative matching between workers and establishments within each race-gender group. Moreover, the degree of assortativeness is stronger among white workers of both genders. In this respect, the combination of upward bias in the variances and downward bias in the covariances implies a relatively large increase in the estimated *correlations* between the worker and firm effects, to 0.47-0.48 for whites and 0.37-0.40 for nonwhites.

The correlation of the sampling errors in the estimated worker and firm effects does not affect our decomposition of the racial wage gap in equation (4), but it does lead to a potential bias in our estimates of the skill-based sorting component defined in equation (5). Consider a firm for which the sampling error in the estimated firm effect for white workers is positive, i.e., $\widehat{\psi}_j^W > \psi_j^W$. On average, the estimated person effects for the white workers at the firm will then be under-estimated, leading us to under-estimate the fractions of those workers at the firm in higher skill groups (and over-estimate the fractions in lower skill groups). Since on average whites in the outside labor market are more skilled, we will then under-estimate the share of whites that would be employed at the firm in the absence of discriminatory hiring and over-estimate the share of nonwhites, leading us to under-estimate the expected excess fraction of white workers at the firm, i.e., $(\widehat{\pi}^*_{Wj} - \widehat{\pi}^*_{Nj}) < (\pi^*_{Wj} - \pi^*_{Nj})$. On the other hand, when the sampling error in the estimated firm effect is negative, we over-estimate the fractions of workers in higher skill groups, and over-estimate the share of whites that would be employed at the firm in the absence of discriminatory hiring and over-estimate the share of nonwhites. Thus $\widehat{\psi}_j^W$ will be negatively correlated with $\widehat{\pi}^*_{Wj} - \widehat{\pi}^*_{Nj}$, leading us to under-estimate the skill-based sorting component.

To assess the magnitude of this problem we take two approaches. First, we compare our main decomposition results based on the largest connected sets to

[24]For computational convenience, we follow KSS and implement their procedure by first adjusting wages by subtracting off the value of the estimated covariate index (obtained using the entire connected set of observations). Since the overall variance contribution of the X's is small, the standard deviation of adjusted wages is very similar to the standard deviation of actual wages.

decomposition results based on the leave-one-out connected sets, which are by construction "better connected." Second, we use a two-sample approach, estimating separate AKM models for the earlier and later halves of our sample. We then implement our decompositions in the first half of the sample, but we use the estimated person effects from the second half of the sample to classify workers into skills groups when computing the skill-based sorting component (we restrict attention to workers present in both subsamples). Since the sampling errors of the person and firm effects from different half samples are independent, this yields an unbiased estimate of the skill-based sorting effect, albeit for a selective set of workers and firms.

B. *Additional Specification Checks*

Before we move to using the parameter estimates from the two-way fixed effects models to decompose the racial wage gaps, we present a last specification check. The additive specification of equation (1) implies that each establishment pays the same wage premium to all workers, regardless of their skill. We test this in two ways. First, following Card, Heining and Kline (2013), we examine the mean residuals from the estimated models in Figure C.5. We show the mean in each of 100 cells, defined by deciles of the estimated person effects and deciles of the estimated establishment effects. Importantly, there is no evidence from these graphs that the AKM model systematically under-estimates the earnings of high-skilled workers at high-premium establishments. For low-skilled workers (in the lowest decile of estimated person effects), however, the mean residuals tend to be positive at low-wage premium establishments, particularly for women. We interpret this as evidence of the effect of the minimum wage, rather than as a specific problem of the AKM model.

Second, we estimated separate AKM models (by race and gender) for workers with less than a high school education and workers with high school education or more. We then considered simple regression models of the form

$$\psi_j^{g,Ed} = \delta_0 + \delta_1 \psi_j^g + \zeta_j, \tag{9}$$

where $\psi_j^{g,Ed}$ is the pay premium at workplace j for workers in race-gender group g with education Ed, and ψ_j^g is the pay premium at workplace j for workers group g from our main specification (which pools the two education groups). The AKM specification implies that $\delta_1 = 1$. However, if firms pay higher premiums to higher skilled workers (i.e., a potential alternative explanation for the relative wage-setting effect), we would expect $\delta_1 < 1$ when the dependent variable is the premium for low-education workers, and $\delta_1 > 1$ when the model is fit for high-education workers.

Simple OLS estimation of models like (9) using the estimated pay premiums $\widehat{\psi}_j^g$ and $\widehat{\psi}_j^{g,Ed}$ is likely to be biased by measurement error in the independent variable that is correlated with measurement error in the dependent variable. We therefore present instrumental variables (IV) estimates, using the estimated wage premium for workers of the same race group but opposite gender as an instrument for the overall group-specific wage premium at each establishment. The results, reported in Table D.6, yield IV estimates of the coefficient δ_1 that are very close to 1 for higher- and lower-education workers of each race-gender groups.[25] Among men, estimates are even slightly higher for lower-education workers (of both race groups). Among women, estimates are slightly lower for lower-education workers, but this is likely due to the higher impact of the minimum wage at low-paying firms for women (as discussed above). As we show in the lower part of the table, these estimates are essentially equal to 1 for both education groups (and both race groups) when we exclude the bottom decile of the establishment effect distribution. We conclude that the AKM specification, while not literally true, provides a relatively good approximation to the observed wage premiums offered to higher- and lower-education groups.

V. Decomposition results

We now decompose the white-nonwhite wage gap into person and establishment effects, and evaluate the impacts of firms' employment and wage-setting policies. We begin by presenting results pooling workers of all education levels together, but we later explore heterogeneity by education group. We also present results in which we narrow our definition of labor markets to include occupation categories.

A. Decomposing the Racial Wage Gap into Person and Establishment Effects

As discussed in Section II, an initial step is to normalize the establishment effects. This allows us to decompose the wages of any individual—or group—into a component due to their person effect and a component attributable to the premiums paid by their employer (and time-varying characteristics). As a baseline, for each race-gender group, we assume that establishments in the restaurant industry pay zero wage premiums on average. Figure 4 displays the distribution of implied average pay premiums by 3-digit industry for white workers. The estimated industry premiums for white males range from near zero—thus near the restaurant industry—for, e.g., delivery services (-0.07), auto repair services (-0.03), and footwear manufacturing (-0.01) to around 0.80 for, e.g., auto manufacturing (0.79) and petroleum extraction (0.83). The ranking is similar for females; the rank correlation with male estimates is 0.87. Interestingly, the high- and low-premium industries correspond fairly closely to the high- and low-wage industries identified by Krueger and Summers (1988).

Using this normalization, Table 3 begins by presenting results from implementing the decomposition in equation (3) using individuals in the dual-connected set of male (Panel A) and female workers (Panel B). Column (1) reports the

[25] For comparison purposes, all the results in Table D.6 restrict attention to establishments included in the largest connected sets of both education groups.

mean white-nonwhite pay gaps. Columns (2), (3) and (4) then decompose these gaps into differences in mean person effects, differences in means of the covariate indexes, and differences in mean establishment effects.

The first row in each panel, which considers all workers in the dual-connected set, shows that the racial wage gap reaches 16.5 ppts. for men and 23.8 ppts. for women. A majority of this gap is attributed to differences in person effects (79% for men and 85% for women). This is illustrated graphically in Figure C.6, which shows that the distribution of person effects among whites is clearly shifted to the right relative to nonwhites (for both genders). For male workers, the difference in the covariate index (which incorporates year effects and adjustments for age) account for a relatively small share of the pay gap. For female workers, however, it reduces the white-nonwhite pay gap by about 2 ppts. This arises because the experience profiles of younger women are steeper for whites than nonwhites. Adjusting women to an age-35 basis, as we do, thus raises the average person effects for white women more than for nonwhite women, with a compensating negative gap in the covariate index. With this in mind, we conclude that differences in the mean of workers' "transferable" skills—combining person effects and time-varying covariates—account for 79% and 75% of the overall white-nonwhite wage gap among men and women, respectively, while differences in the establishment effects account for 21% and 25%.

As noted in Section II, these estimates use a reweighting procedure that adjusts the distribution of nonwhites across micro-regions to match the distribution of whites. Table D.7 shows that this procedure reduces the racial wage gaps by 3-4 ppts. That reduction comes mainly from the gap in mean establishment effects, 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.

The decompositions in Table 3, columns (2)-(4), also rely on the assumption that workers of both race groups are paid their true productivity in the restaurant industry. If one assumes instead that the 2.4-2.7 ppts. wage gap in that industry is due to differential pay premiums, one would *lower* all the estimated person effects of whites and *raise* all the estimated establishment effects for whites by that amount. This would lead to a 2.4-2.7 ppts. reduction in the gap in estimated person effects in column (2) and a 2.4-2.7 ppts. increase in the gap in estimated establishment effects in column (4). Under this alternative assumption, the share of the white-nonwhite pay gap explained by employment and wage-setting policies would rise to 36% for both men and women.

B. Decomposing the Effect of Employment and Wage-Setting Policies

Next, columns (5)-(6) of Table 3 present the estimated wage-setting and sorting effects from equation (4). The first row in each panel, again considering all workers in the dual-connected set, shows that most of the overall contribution of establishment effects is attributable to the sorting effect. Indeed, the sorting effect accounts for 2.9 ppts. among men and 4.8 ppts. among women, or 18% and 20% of their white-nonwhite wage gap, respectively. By comparison, the estimated wage-setting effects are modest in size, on the order of 4%-5% of the overall white-nonwhite wage gap.

Some insight into the size of the wage-setting effects is provided by equation (8) and the pattern in Figure C.7, which shows binned scatterplots of the relationship between the estimated pay premiums for whites and nonwhites. 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 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). This approach leads to estimates of $\gamma = 0.964$ for males and $\gamma = 0.930$ for females. These estimates imply that the expected size of the wage-setting effect for men is about 3.7% of the average wage premium among nonwhite men, while for women it is about 7.5% of the average wage premium among nonwhite women. Given the magnitudes of the average premiums for nonwhites – 0.165 for males and 0.078 for females – equation (8) predicts pay-setting effects that are close to the estimates in column (5), particularly for men.

As noted in Section II, the value of the wage-setting effects—but not of the sorting effects—depends on the normalization of the establishment effects. Under our alternative assumption that white-nonwhite pay differences in the restaurant industry are due to differential pay premiums (rather than differential productivity), we would increase the relative wage-setting effect by 2.4-2.7 ppts. As a result, it would reach 18% of the overall racial wage gap for men and 16% for women.

C. *Decomposing the Sorting Effect into Skill-Based and Residual Sorting*

The last two columns of Table 3 further decompose the sorting effect into a skill-based component—due to assortative (race-neutral) matching—and a residual component, using equations (5) and (6), respectively. As discussed in Section II, 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 (by 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). Next, we calculate the fraction of employees at each establishment in each bin in each year, and the nonwhite share of each bin in its local labor market (micro-region) in that year. We combine these to calculate the expected fractions of whites and nonwhites at the establishment, which we use to calculate the counterfactual employment shares π^*_{Wj} and π^*_{Nj}, and the skill-based sorting effect given by equation (5).

Figure 5 presents our results graphically. For both genders, the black line displays the actual share of nonwhites by decile of the (white-specific) establishment-effect distribution. The red and green lines display counterfactual shares under two scenarios. First, we assume that each establishment maintains the age structure of its workforce but selects workers at random within age categories (i.e., without regard for race or skill) from its local labor market. The red line shows this "naive" counterfactual. Second, we assume that each establishment maintains its joint distribution of age and skill but selects workers at random within age-skill categories (i.e., without regard for race), yielding the "full" counterfactual shown by the green line.

The actual shares of nonwhite workers are stable across the lower deciles of the establishment-effect distribution, but then decrease sharply from 0.34 to 0.25 for men and from 0.33 to 0.18 for women. The shares of nonwhites predicted by our naive benchmark remain nearly constant, as we move across the firm effect deciles. Age differences are thus unimportant for the differential sorting of whites and nonwhites to higher- and lower-premium establishments. In contrast, 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 workplaces to hire higher-skilled workers. The green lines fall between the red and the black lines, suggesting that skill-based employment policies explain some, but not all, of the under-representation of nonwhites at higher-premium workplaces.

These results are summarized quantitatively in the first row of each panel in columns (7)-(8) of Table 3. For both genders, skill-based sorting accounts for about 65% of the overall sorting effect, while residual sorting—which includes any effect of discriminatory employment policies by higher-premium establishments—accounts for about 35% of the sorting effect or 6%-7% of the overall white-nonwhite wage gap.

As noted in Section II, under our alternative assumption regarding the source of the pay gap in the restaurant industry, we would reduce the estimated gap in mean person effects, and thus reduce the importance of skill differentials for the sorting of whites and nonwhites across workplaces. Specifically, the importance of skill-based sorting would decrease to 56%-57% of the overall sorting effect, and the residual sorting effect would account for 8%-9% of the overall white-nonwhite wage gap.

Finally, we note that a counterfactual (race-neutral) sorting based either on the component of human capital that we observe in the data or on the estimated person effects leads to similar conclusions. Indeed, the decomposition of the sorting effect in the first row of each panel in Table 3 is unchanged if we recalculate the counterfactual employment shares using skill groups defined by five education categories (no education, elementary school, middle school, high school, college) rather than by quartiles of the distribution of person effects. In that case, the skill-based sorting and residual sorting components amount to 2 ppts. and 0.9 ppts. among men (compared to 1.9 ppts. and 1 ppts. in Table 3), and to 3 ppts.

and 1.9 ppts. among women (compared to 3.1 ppts. and 1.8 ppts. in Table 3).

D. Decomposition results by education category

The other rows in each panel of Table 3 replicate the same decompositions for three education categories separately: workers with less than a high-school education (52% of males, 33% of females), high-school graduates who did not complete college (40% of males, 50% of females), and college graduates (8% of males, 17% of females).[26] Column (1) shows that the white-nonwhite pay gap increases steeply across these three education categories for both men and women, ranging from about 5 ppts. for workers with no high school to 19-22 ppts. for those with completed college.[27]

The relationships between education and average estimated person and establishment effects are illustrated graphically in Figure 6. The gray and black solid lines in the upper panels (by gender) show how the means of the estimated person effects for whites and nonwhites, respectively, vary across five education levels (no education, elementary school, middle school, high school, college). Mean person effects for both whites and nonwhites rise only slightly across the three lowest levels of education, which together comprise the "less than high school" group reported in Table 3. In contrast, there is a substantial gradient across the three highest education levels. This is also true for the mean establishment effects for both whites and nonwhites, which are displayed by the gray and black solid lines in the lower panels of Figure 6. The increases in mean establishment effects across education categories imply that a sizable share of the "return to college" in Brazil is attributable to increased access to jobs at establishments that pay higher premiums. Specifically, the college-high-school gap in mean establishment effects is 18 log points for white men, 15.5 log points for nonwhite men, 17.5 log points for white women, and 13 log points for nonwhite women.

The gaps between the gray and black solid lines in the four panels of Figure 6 illustrate the decomposition of the white-nonwhite wage gaps into person and establishment effects. Across the three lowest levels of education, there is a modest gap in mean person effects for both genders, and a small gap in mean establishment effects but only for women. Consequently, Table 3 shows that most of the white-nonwhite pay gap for workers without a high school education is attributed to workers' "transferable" skills among men, while differences in mean establishment effects account for 25% of this gap among women. For the higher education

[26]There are three types of college degrees in Brazil: bachelor's degrees (about 70% of recent graduates); licenciatura degrees that train teachers (22%); and technology degrees (12%), which are typically awarded for shorter vocationally-oriented programs.

[27]For comparison, the hourly wage gap between white and black males in the U.S. – based on annual earnings and hours collected in the 2016-2018 March Current Population Surveys (U.S. Department of Commerce, 2021) – is 0.06 for workers with less than a high school education, 0.19 for workers with high school and no college, and 0.22 for those with a 4-year degree. The corresponding gaps for females in the U.S. are smaller in magnitude and more similar across education groups: -0.04, 0.12, and 0.11, respectively.

levels, Figure 6 shows that the increase in mean person effects and mean establishment effects is steeper for whites than nonwhites for both genders. This is especially true in the lower panels of Figure 6. As a result, the gaps attributed to differences in mean establishment effects increase together with the overall racial wage gap for high-school and college graduates in Table 3. Differences in mean establishment effects account for 2.6 ppts. and 3.2 ppts. among men and women with a high school education, or 22% and 28% of their white-nonwhite wage gap. These figures are 5.4 ppts. and 8.2 ppts. among college graduates, or 28% and 37% of their racial pay gap.

The role of employment and pay-setting policies is also illustrated in Figure 6. The dashed black line in the lower panels displays the mean white-specific establishment effect at the places of employment of nonwhite workers. The gap between the dashed black line and the solid black line thus captures the wage-setting effect by education group, while the gap between the solid gray line and the dashed black line captures the sorting effect. The patterns in Figure 6 show that both effects increase in the higher-education categories. The entries in Table 3 indicate that the contribution of the differential sorting of whites and nonwhites across workplaces remains relatively larger for higher-education categories. For instance, the sorting effect accounts for 16% and 23% of the white-nonwhite wage gap among men and women with a college degree in column (6), compared to 12% and 15% for the wage-setting effect in column (5). However, as noted earlier, the shares attributed to the wage-setting effect would rise under our alternative normalization assumption, and they would actually exceed the shares attributed to differential sorting for each education category—e.g., they would reach 25% and 27% for college graduates.

Table 3 decomposes the sorting effect into skill-based and residual sorting by education category as well. One point to note here is that we calculate the counterfactual employment shares π^*_{Wj} and π^*_{Nj} for the three education categories separately. In particular, we still divide workers into 16 bins and we still use the same four age categories, but we now use quartiles of the distribution of person effects for workers of the education level corresponding to each row of Table 3. The skill groups are thus much narrower than when we consider all workers in the dual-connected set.

The entries in column (7) of Table 3 show that skill differentials are unimportant for the differential sorting of whites and nonwhites without high school to higher-premium establishments, but their sorting effect is quite small to begin with. The impact of skill-based sorting increases among high-school and college graduates of both genders, reflecting the combination of positive assortative matching and the larger gap in mean person effects among higher-educated workers. However, it is the residual (non-skill-based) component of the sorting effect that becomes most important—both as a share of the sorting effect and as a share of the white-nonwhite wage gap—for college-educated workers. This suggests that discriminatory hiring practices have the largest impacts on the most highly-educated workers in Brazil.

It is worth emphasizing that the estimated person effects in an AKM-style model incorporate any unobserved components of human capital—such as differences in the quality of schooling or the choice of college major—provided that those components are rewarded (approximately) equally by different employers. Thus, differences in college quality or in college major between whites and nonwhites will likely be reflected in our measure of *skill-based* sorting, but will not bias our measures of residual sorting.

Finally, as a further illustration of the higher wage losses associated with unexplained sorting and relative wage-setting for higher-skilled nonwhites, Figure 7 plots the mean value of the establishment premiums for whites and nonwhites with different ranges of estimated person effects (by gender). To address the problem that the sampling errors in the estimated person and establishment effects are negatively correlated, we use the two-sample approach discussed in Section IV. We estimate separate AKM models for the first half (2002-2008) and the second half (2008-2014) of our sample. We then focus on people who appear in the dual-connected sets in both subsamples, divide them into deciles based on their person effects in the later period, and plot the mean establishment effects from the earlier period for each decile.

Figure 7 shows several interesting points. First, consistent with the (bias-corrected) correlations between worker and establishment effects reported in Table 2, workers with higher transferable skills are more likely to work at establishments that pay higher premiums. The mean establishment effect increases across the person effect deciles for whites (solid gray lines) and nonwhites (solid black lines) of both genders. Second, while whites tend to have higher transferable skills than nonwhites,[28] they also earn higher pay premiums than nonwhites for a given skill level. For the lower deciles of the person effects, the gap in mean establishment effects between whites and nonwhites is positive but relatively small for men and women. In contrast, for individuals with estimated person effects in the top three deciles, the gaps are relatively large, particularly for women. Third, the dashed black lines in Figure 7, which display the mean white-specific establishment effects at the places of employment of nonwhite workers, show that the larger gaps among higher-skilled workers arise because of both residual sorting and differential wage setting. The gaps between the solid gray lines and the dashed black lines—capturing the residual sorting effect—and the gaps between the dashed black lines and the solid black lines—capturing the relative wage-setting effect—all increase across the person effect deciles.

As mentioned in Section IV, we also implemented all our decompositions for the sample of Figure 7, but defining the skill groups for the computation of the skill-based sorting component based on the estimated person effects from the later sample. These results are shown in Table D.8. Pooling all education groups,

[28]This is also illustrated in Figure C.8, which shows that the number of white workers increases across the person effect deciles, while the number of nonwhite workers decreases.

skill-based sorting accounts again for about two thirds of the overall sorting effect with this unbiased estimate of the skill-based sorting effect. Moreover, the residual sorting component remains increasing in education levels. Our key findings with respect to the decomposition of the sorting effect in Table 3 are thus unlikely driven by any bias arising from the correlation of the sampling errors in the worker and firm effects.[29]

E. Including occupation in our definition of labor markets

A potential concern with our measurement of skill-based sorting is that a classification of skill based only on age and average wages ignores differences between occupations. For example, a sales worker and a skilled trade worker may earn similar wages, but one cannot easily fill the other's job. This may lead us to mismeasure the nonwhite shares in the "relevant" local labor markets for a given firm, leading to bias in our estimates of the counterfactual shares π^*_{Wj} and π^*_{Nj} that would be expected if a firm hired without regard for race but maintained its skill distribution.

To address this concern, we redefined labor markets based on micro-region and occupation, using six major occupation groups (managers, professional workers, technicians, administrative workers, service/sales workers, and blue collar workers). We then modified the procedures underlying Table 3 in two ways. First, we reweighted nonwhites to have the same joint distribution as whites across micro-regions *and occupations*. Second, we re-classified the workforce of each establishment in each year into four age cells, four quartiles of estimated person effects, *and six occupations*, and calculated the expected share of nonwhites at the establishment if it hired at random from its surrounding micro-region within these $4 \times 4 \times 6 = 96$ cells.

The results are presented in Table 4, which follows the same format as Table 3. A first observation is that reweighting nonwhites based on the joint distribution of whites across location and occupation narrows the racial wage gap substantially: e.g., by 6.1 ppts. for all men and 11.3 ppts. for all women. Most of this narrowing comes from a narrowing of the gap in person effects. The racial gap in establishment effects falls by only 0.9 ppts. for all men and 1.8 ppts. for all women when we control for occupation. Within education categories, the effects are even smaller. For example, the gap in establishment effects for college-educated men falls from 5.4 ppts. to 5.0 ppts., while the gap for college-educated women falls

[29] To add to this discussion, in Table D.8, we also compare estimates of the skill-based sorting component for which we define skill groups based on the estimated person effects in either the earlier sample or the later sample. The skill-based sorting component increases modestly when we use the person effects from the later sample. However, we note that the bias is likely larger in this case than for our baseline results, because the sample used in Table D.8 is "less connected." This is illustrated in Table D.9, which present similar statistics as in Table 2 but for the largest connected set of each race-gender group in the earlier sample (restricting attention to workers who also appear in the later sample). Indeed, the share of the overall variance of wages attributed to the variance of the establishment effects and to the covariance of person and establishment effects experience larger changes between Panel A and Panel B, and between Panel B and Panel C, than in Table 2.

from 8.2 ppts. to 8.0 ppts.

A second observation is that adjusting for occupation has little effect on the magnitude of the wage-setting effects. As a share of the white-nonwhite pay gap, the contribution of differential wage setting is therefore larger—on the order of 7%-10% for workers of all education levels together, and 22%-23% for college graduates.

A third observation is that controlling for occupation leads to some reduction in the estimated sorting effects: e.g., by 1 ppts. for all men and 1.9 ppts. for all women. Interestingly, virtually all of this reduction is attributable to a reduction in skill-based sorting. The residual sorting effects in column (8) of Table 4 are very similar in magnitude to those in Table 3. This indicates that sorting based on workers' occupations is mostly captured by our estimated skill-based sorting in Table 3.

We conclude that differential pay setting and unexplained sorting contribute importantly to racial pay gaps in Brazil, with particularly large impacts on nonwhites with the highest levels of education. 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.

VI. 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 use alternative choices.

We summarize our findings graphically in Figure 8 (by gender). 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 the first row of each panel in Table 3.

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. As shown in the second column of each panel, the overall wage gap is about 2 ppts. smaller for males and 3 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 AKM models, and repeated our decompositions, using only observations from the Northeast region where average education and wage levels are lower. In this case, the wage gaps are smaller, and the sorting and wage-setting effects are scaled down accordingly. We find that skill-based sorting is relatively less important in the Northeast than in our baseline 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 I, some individuals in RAIS are classified as white in some years and as nonwhite in others. For our baseline results, we resolved this issue by assigning individuals their modal race group. As a simple alternative, we drop any worker classified in more than one (binary) race group, which removes between-establishment movers disproportionately, since most changes in race occur with a change in employer.[30] The average racial wage gaps in this "consistent race" sample are about 9 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, but the decomposition shares are not too different. The contribution of establishment pay-premiums is larger with this sample, most notably because the share of the wage gaps attributed to the pay-setting effect rises to 8% for men and 7% for women.

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

Next, we show the impact of adopting our alternative assumption regarding the source of the racial wage gap in the restaurant industry (i.e., assuming that it is due to differential wage premiums earned by whites), which we discussed in Section V. We also show that using octiles rather than quartiles of the person effect distributions for the decomposition of the sorting effects into skill-based and residual sorting lead to only negligeable changes relative to our baseline results. Additionally, as discussed in Section IV, we show decomposition results using AKM estimates based on the leave-one-out connected sets described in Panel B of Table 2. The racial wage gaps are slightly larger in the leave-one-out dual connected sets, but the relative contribution of skill-based and residual sorting is almost identical. As the leave-one-out connected sets are "better connected," this indicates again that the correlation of the sampling errors in the worker and firm effects are unlikely to affect our conclusions regarding the role of skill-based

[30]The percentage of person-year observations lost when imposing this restriction is 54% for white males, 49% for nonwhite males, 57% for white females, and 53% for nonwhite females.

sorting based on the results in Table 3.[31]

The following columns display our result adjusting for observed differences in occupation categories (shown in Table 4) and the result of a similar exercise in which we adjust instead for observed differences in industry of employment. In that case, we reweighted nonwhites to have the same distribution as whites across micro-regions *and industries*.[32] We then re-classified the workforce of each establishment in each year into four age cells, four quartiles of estimated person effects, *and 13 industries*, and calculated the expected share of nonwhites at the establishment if it hired at random from its surrounding micro-region within these $4 \times 4 \times 13 = 208$ cells. Adjusting for industry differences reduces the white-nonwhite wage gaps to a smaller extent than adjusting for occupation differences: the reduction in the contribution of establishment effects is similar, but there is less of a closing of the gap in person effects. As a result, the share of white-nonwhite wage gaps attributed to differences in mean establishment effects decreases, to 18% among men and 21% among women. As when adjusting for occupation differences, the reduction in the contribution of establishment effects is driven by a reduction in the sorting effect. Part of the sorting effect in our baseline results thus implied a differential sorting across industries. Yet, adjusting for industry differences, the relative contribution of skill-based and residual sorting for the overall sorting effect is unchanged compared to the baseline results.[33]

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 in Section I, legal or social sanctions against discrimination in the Brazilian labor market may have evolved over time.

To evaluate this possibility, we re-estimated our AKM models and repeated our decompositions separately for two 7-year sub-periods, from 2002 to 2008 and from 2008 to 2014. 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, implementing our decompositions separately for high "face-time" industries (commerce, hospitality, financial services and insurance, real estate) and low "face-time" industries (extractive and transformation industries, utilities, construction), we see that the residual sorting and relative pay-setting effects are larger in the first group of industries.

[31]For this robustness check, we present the full decomposition results in Table D.10, which follows a format similar to Table 3.

[32]Farming and fishing, Extractive industries, Manufacturing, Electricity/gas/utilities, Construction, Trade, Accomodation and food, Transportation and communication, Banking and finance, Real estate, Public administration, Education, Other services and organizations.

[33]We also present decomposition results by industry in Table D.11.

VII. 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 assortative 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 employment 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 compared to white workers of similar skill levels, and tend to receive lower pay premiums in the establishments they are sorted into. 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 Table D.11 that our estimates of the "unexplained" under-representation of nonwhites is largest in the "banking and finance" industry.

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 consideration in their recruitment process. For instance, in contrast to the U.S., few Brazilian federal agencies require 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 in place in those corporations. It remains unclear, however, to which extent this 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.

Altonji, Joseph, and Rebecca Blank. 1999. "Race and Gender in the Labor Market." In *Handbook of Labor Economics, Vol. IIIc.* , ed. 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, 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.

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.

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

Bailey, Martha J., John DiNardo, and Bryan A. Stuart. Forthcoming. "The Economic Impact of a High National Minimum Wage: Evidence from the 1966 Fair Labor Standards Act." *Journal of Labor Economics.*

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.

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.

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

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 O. 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*, 5–21.

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, 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, 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.

Charles, Kerwin K., and Jonathan Guryan. 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. Forthcoming. "Minimum Wages and Racial Inequality." *Quarterly Journal of Economics.*

Dube, Arindrajit. 2014. "Designing Thoughtful Minimum Wage Policy at the State and Local Levels." *The Hamilton Project Policy Proposal.*

Dube, Arindrajit. 2019. "Impacts of Minimum Wages: Review of the International Evidence." London, UK: HM Treasury and Department for Business, Energy & Industrial Strategy Available at https://www.gov.uk/government/publications/impacts-of-minimum-wages-review-of-the-international-evidence.

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.

Francis, Andrew, and Maria Tannuri-Painto. 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.* , ed. Orley Ashenfelter and David Card, 885–971. Amsterdam: Elsevier.

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 No. 681.*

Gerard, Francois, Lorenzo Lagos, Edson Severnini, and David Card. 2018. "Assortative Matching or Exclusionary Hiring? The Impact of Firm Policies on Racial Wage Differences in Brazil." *NBER Working Paper No. 25176.*

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

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.

Giuliano, Laura, Jonathan Leonard, and David I. Levine. 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.

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.* , ed. 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 No. 807.*

Hirata, Guilherme, and Rodrigo R. Soares. 2020. "Competition and the Racial Wage Gap: Evidence from Brazil." *Journal of Development Economics*, 146: 102519.

IBGE, Instituto Brasileiro de Geografia e Estatística. 2016*a*. "Censo Demográfico (Censo) 2010." Access obtained on 2016-03-04.

IBGE, Instituto Brasileiro de Geografia e Estatística. 2016*b*. "Pesquisa Nacional por Amostra de Domicílios (PNAD) 2002-2014." Access obtained on 2016-06-24.

Kline, Patrick, Raffaele Saggio, and Mikkel Sølvsten. 2020. "Leave-out Estimation of Variance Components." *Econometrica*, Forthcoming.

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.*

Kristensen, Nicolai, and Wendy Cunningham. 2006. "Do Minimum Wages in Latin America and the Caribbean Matter? Evidence from 19 Countries." *World Bank Policy Research Working Paper 3870.*

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.*

Longhi, Simonetta. 2017. "Racial Wage Differentials in Developed Countries." *IZA World of Labor*, 365.

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

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.

MTE, Ministério do Trabalho. 2015. "Relatório Anual de Informações Sociais (RAIS) 2005-2014." Access granted on 2015-09-21.

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).

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.

Pendakur, Krishna, and Ravi Pendakur. 2002. "Colour My World: Have Earnings Gaps for Canadian-Born Ethnic Minorities Changed Over Time?" *Canadian Public Policy*, 28(4): 489–512.

Penner, Andrew M. 2008. "Race and Gender Differences in Wages: The Role of Occupational Sorting at the Point of Hire." *The Sociological Quarterly*, 49(3): 597–614.

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 No. 1114.*

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

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

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

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

Skidmore, Thomas E. 1992. "Fact and Myth: Discovering a Racial Problem in Brazil." *Kellogg Institute Working Paper No. 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 No. 769.*

Song, Jae, David J. Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter. 2019. "Firming Up Inequality." *Quarterly Journal of Economics*, 134(1): 1–50.

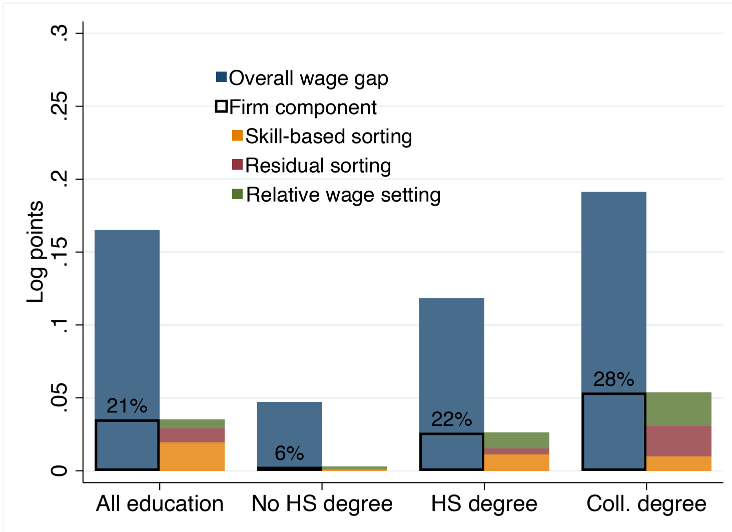
Telles, Edward E. 2004. *Race in Another America: The Significance of Skin Color in Brazil.* Princeton NJ: Princeton University Press.

Travassos, Claudia, and David R. Williams. 2004. "The Concept and Measurement of Race and Their Relationship to Public Health: A Review Focused on Brazil and the United States." *Cadernos de Saúde Pública*, 20(3): 660–678.

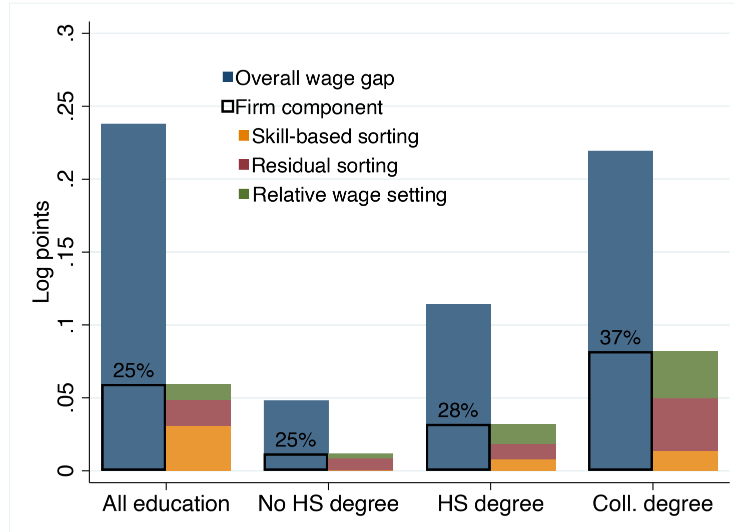
U.S. Department of Commerce, Bureau of the Census. 2021. "Annual Social and Economic Supplement (ASEC) 2016-2018." Last accessed on 2021-01-27.

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*, 1–19.

FIGURE 1. SUMMARY OF OUR MAIN DECOMPOSITION RESULTS



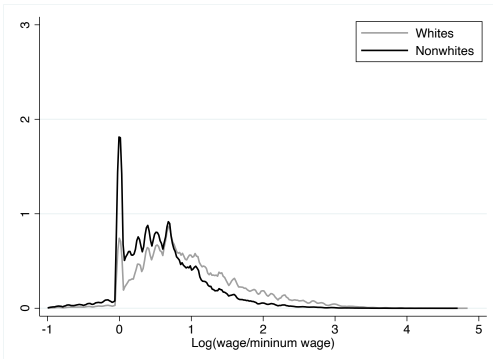
(a) Males



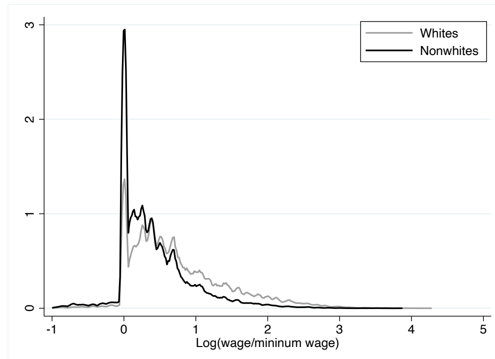
(b) Females

Notes: The figure summarizes our main results (presented in Table 3) for the decomposition of the racial wage gap by gender and education group—all education levels together, less than a high-school (HS) education, high-school graduates with no college degree, and college graduates. In each case, the height of the blue column on the left represents the mean white-nonwhite wage gap, with the black rectangle representing the portion of the gap attributed to firms (also reported as a percentage). The stacked column on the right displays the three components of the contribution of firms to racial wage gaps: skill-based sorting, residual sorting, and relative wage-setting. All samples are restricted to the dual-connected set of each gender group. Nonwhites are reweighted so as to have the same distribution across micro-regions as whites (of the same gender).

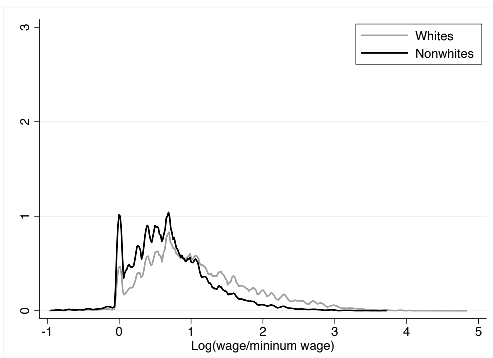
FIGURE 2. LOG HOURLY WAGE DISTRIBUTIONS AMONG PRIVATE-SECTOR EMPLOYEES



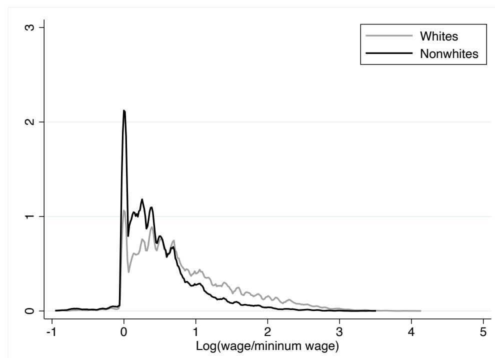
(a) Males (Brazil)



(b) Females (Brazil)



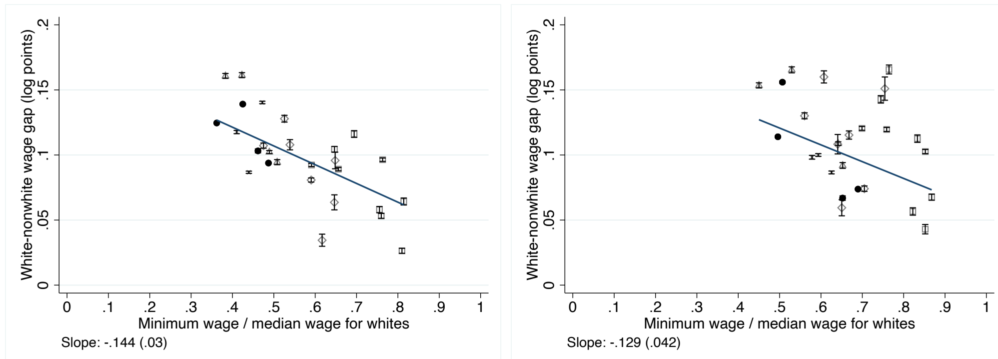
(c) Males (Southeast region)



(d) Females (Southeast region)

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 full-time nonfarm private-sector employees (either formal or informal), ages 25 to 54 with potential labor market experience of at least 1 year and non-zero tenure, as well as non-missing data on race, gender, education, wage, and hours worked. The left and right panels restrict the samples to male and female workers, respectively. The top panels pool all regions of Brazil together, while the bottom panels are restricted to the Southeast region.

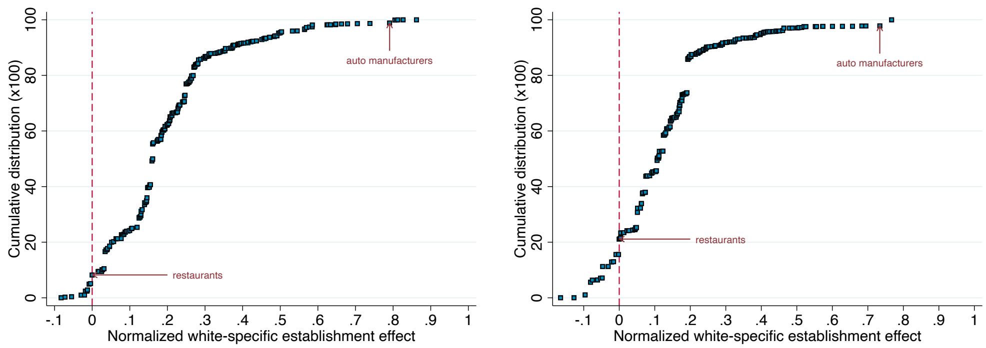
Figure 3. Racial Wage Gaps and Minimum-to-Median Wage Ratios by State



(a) Males (b) Females

Notes: The figure plots estimates of the "unexplained" white-nonwhite wage gap (with their 95% confidence intervals) for each of the 27 Brazilian states, against the ratio of the minimum wage to the median wage of white workers in the state (by gender). The figure uses the same PNAD 2002-2014 samples used in Figures 2a and 2b. The state-level regressions control for year fixed effects, five education dummies (incomplete elementary school, and complete elementary school, middle school, high school, or college), and a quadratic in potential experience, and use survey weights. The different markers identify the region of Brazil that a state belongs to: Northeast (squares), North (diamonds), Midwest (triangles), Southeast (black circles), and South (exes). The line in each panel displays the result of a WLS regression of the state-level estimates on the minimum-to-median wage ratios, weighting the estimates by the inverse of their standard error squared. The estimated slope is reported with its standard error at the bottom of the figure.

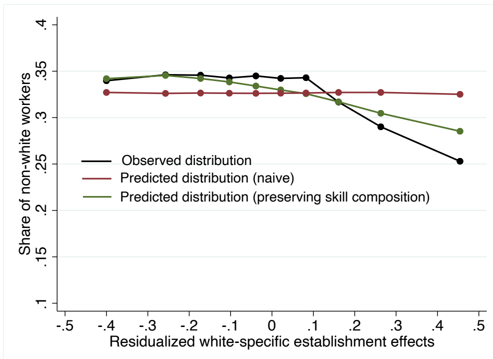
Figure 4. Cumulative Distributions of Normalized Establishment Effects by Industry



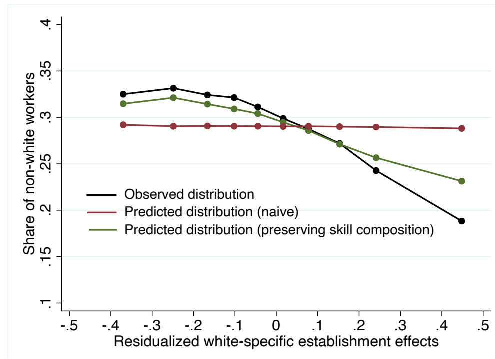
(a) Males (b) Females

Notes: The figure displays the cumulative distribution of average white-specific establishment effects for 3-digit industries (CNAE), after normalizing the establishment effect with respect to the average in the restaurant industry (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 D.5. Nonwhites are reweighted so as to have the same distribution across micro-regions as whites (of the same gender). The rank correlation between these industry-level averages for males and females is 0.87. Industries representing less that 0.05% of the sample are omitted.

Figure 5. Share of Nonwhite Workers by Establishment Effect Deciles



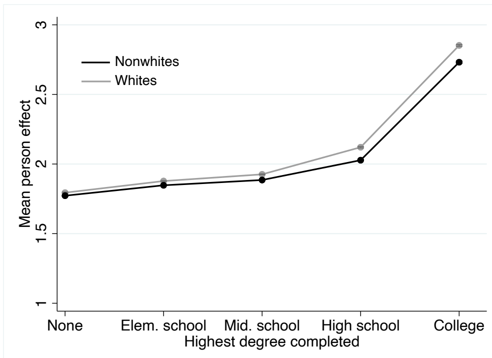
(a) Males



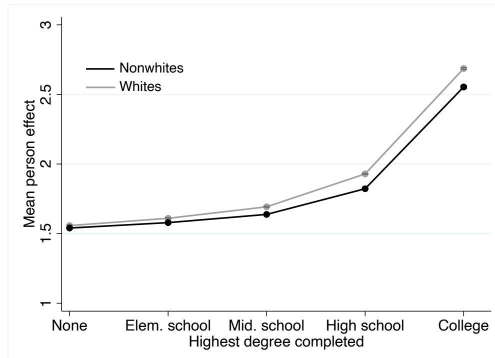
(b) Females

Notes: The figure displays the observed distribution of nonwhite person-year observations by deciles of the white-specific establishment effects (residualized of micro-region fixed effects), as well as two predicted distributions that reshuffle person-year observations—regardless of race—across establishments within a micro-region in each year. The naïve distribution maintains the age distribution of each establishment. The skill-preserving distribution maintains the joint age-skill distribution of each establishment. Samples are restricted to the dual-connected set of each gender group in the Southeast region. Nonwhites are reweighted so as to have the same distribution across micro-regions as whites (of the same gender).

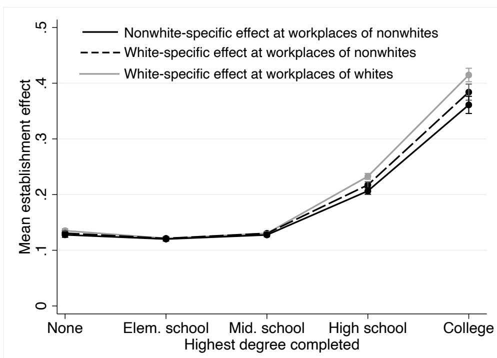
Figure 6. Mean Person Effects and Establishment Effects by Education Level



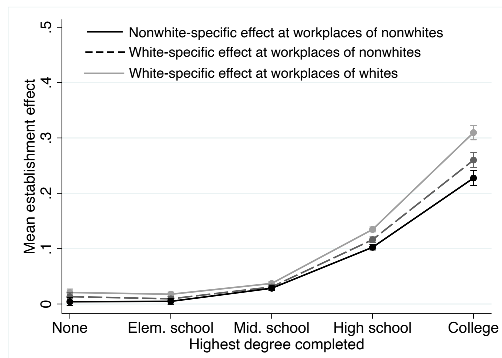
(a) Males (person effects)



(b) Females (person effects)



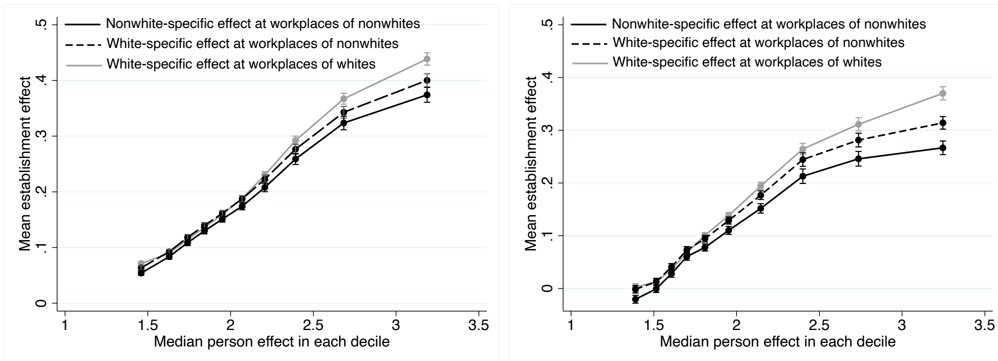
(c) Males (establishment effects)



(d) Females (establishment effects)

Notes: The top panels display the means of the person effects for whites and nonwhites by education level (and gender). The bottom panels display the means of the white-specific establishment effects at workplaces of whites (solid gray lines), of the white-specific establishment effects at workplaces of non-whites (dashed black lines), and of the nonwhite-specific establishment effect at workplaces of nonwhites (solid black lines). The point estimates (with their 95% confidence intervals) are obtained by regressing the corresponding variable on dummies for the five education groups (without a constant). Samples are restricted to the dual-connected set of each gender group in the Southeast region. Nonwhites are reweighted so as to have the same distribution across micro-regions as whites (of the same gender). The gaps between the gray and black solid lines illustrate the decomposition of the white-nonwhite wage gaps into person and establishment effects by education group. In the bottom panels, the gap between the dashed black line and the solid black line captures the wage-setting effect by education group, while the gap between the solid gray line and the dashed black line captures the sorting effect by education group.

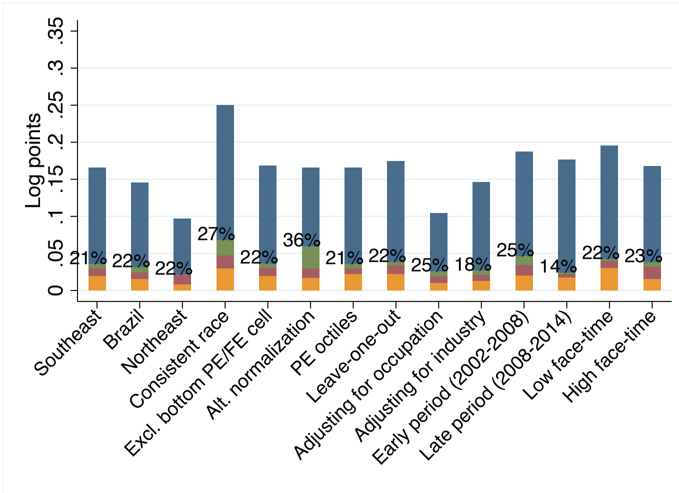
Figure 7. Mean Establishment Effects by Person Effect Deciles



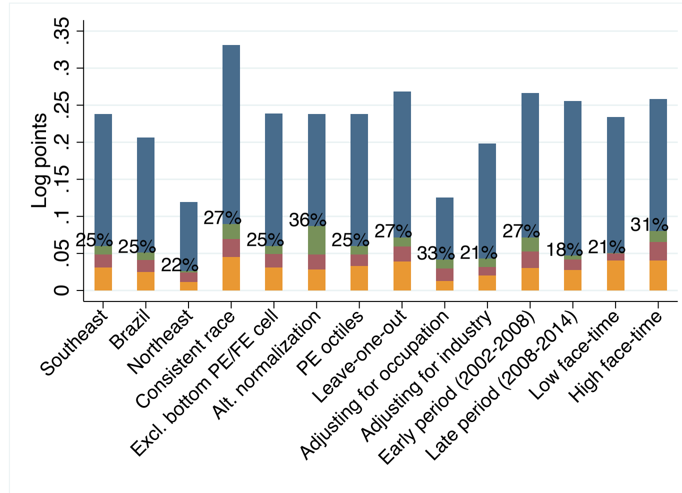
(a) Males (b) Females

Notes: The figure displays the means of the white-specific establishment effects at workplaces of whites (solid gray lines), of the white-specific establishment effects at workplaces of nonwhites (dashed black lines), and of the nonwhite-specific establishment effect at workplaces of nonwhites (solid black lines), estimated in the 2002-2008 subsample, by decile of the estimated person effect distribution in the 2008-2014 subsample (and by gender). The point estimates (with their 95% confidence intervals) are obtained by regressing the corresponding variable on dummies for the 10 deciles (without a constant). Samples are restricted to workers who appear in the dual-connected set of each gender group in the Southeast region in both the early period and late period subsamples. This two-sample approach addresses the problem that the sampling errors in the estimated person and establishment effects are negatively correlated. Nonwhites are reweighted so as to have the same distribution across micro-regions as whites (of the same gender). In each panel, the gap between the dashed black line and the solid black line captures the wage-setting effect by decile, while the gap between the solid gray line and the dashed black line captures the sorting effect by decile, i.e., a residual sorting effect.

Figure 8. Robustness and Heterogeneity of Racial Wage Gap Decompositions



(a) Males



(b) Females

Notes: The figure displays similar decomposition results as in the first row of each panel in Table 3 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 (in orange), to residual sorting (in maroon), to relative wage-setting (in green), and to differences in person effects and covariates (in blue), respectively. The overall contribution of firms—the sum of the first three components—as a proportion of the racial wage gap is reported in each column. For reference, the first column in each panel reproduces our baseline results from Table 3. All samples are restricted to the dual-connected set of each gender 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	8.45	9.24	7.59	7.76	8.92	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$)	1.61	1.78	1.42	1.46	1.75	1.89	1.57	1.55
Share in formal sector employment	0.79	0.82	0.75	0.78	0.83	0.84	0.81	0.82
B. Females								
Share of sample in column race group	1.00	0.51	0.41	0.08	1.00	0.59	0.32	0.08
Share of column race group in private employment	0.21	0.25	0.17	0.20	0.25	0.27	0.23	0.24
Characteristics of private-sector employees in column race group:								
Mean years of education	10.16	10.66	9.41	9.39	10.38	10.94	9.39	9.28
Fraction with high school or more	0.65	0.69	0.58	0.59	0.67	0.72	0.57	0.57
Mean log hourly wage (R$)	1.48	1.62	1.28	1.31	1.59	1.70	1.38	1.37
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), ages 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 to a 2010 base using the Brazilian CPI. Columns (1) and (5) pool 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—Summary of Estimated Two-Way Fixed Effects Models

	White male	Non-white male	White female	Non-white female
	(1)	(2)	(3)	(4)
Largest connected set				
Standard deviation of log wages	0.670	0.582	0.685	0.554
Mean log wages	1.939	1.768	1.782	1.557
A. AKM decomposition				
Std. dev. of person effects (across person-yr obs.)	0.484	0.415	0.527	0.437
Std. dev. of estab. effects (across person-yr obs.)	0.304	0.279	0.304	0.266
Std. dev. of covariates (across person-yr obs.)	0.175	0.181	0.181	0.185
Correlation of person/estab. effects	0.275	0.167	0.264	0.102
Adjusted R-squared of model	0.901	0.876	0.918	0.897
Percentage of variance of log wages due to:				
person effect	52.1%	50.9%	59.1%	62.1%
establishment effect	20.6%	23.1%	19.7%	23.1%
covariance of person and estab. effects	18.0%	11.4%	18.0%	7.7%
estab. effects+covariance person and estab. effects	38.6%	34.5%	37.7%	30.9%
Number of establishments	1,284,740	717,098	1,162,373	508,088
Number of movers	4,052,299	1,771,840	2,845,495	930,306
Number of person-year observations	39,661,514	16,605,082	27,814,349	8,900,093
Leave-one-out connected set				
Standard deviation of log wages	0.670	0.576	0.706	0.569
Mean log wages	1.961	1.775	1.847	1.582
B. AKM decomposition				
Std. dev. of estab. effects (across person-yr obs.)	0.287	0.253	0.293	0.236
Correlation of person/estab. effects	0.354	0.261	0.375	0.260
Percentage of variance of log wages due to:				
establishment effect	18.3%	19.3%	17.2%	17.2%
covariance of person and estab. effects	21.3%	15.7%	23.0%	16.2%
estab. effects+covariance person and estab. effects	39.6%	35.0%	40.2%	33.4%
C. KSS decomposition				
Std. dev. of estab. effects (across person-yr obs.)	0.271	0.233	0.273	0.213
Correlation of person/estab. effects	0.468	0.397	0.480	0.369
Percentage of variance of log wages due to:				
establishment effect	16.3%	16.4%	15.0%	14.0%
covariance of person and estab. effects	22.8%	17.6%	24.6%	18.4%
estab. effects+covariance person and estab. effects	39.1%	34.0%	39.6%	32.5%
Number of establishments	749,877	325,034	600,499	179,697
Number of movers	3,551,977	1,423,252	2,328,539	645,146
Number of person-year observations	22,305,141	8,761,529	13,972,235	3,708,699

Notes: The table summarizes the results from estimating two-way fixed effects models for log hourly wages using person-year observations for each race-gender group in the Southeast region. The models include dummies for individual workers and individual establishments, year dummies interacted with five education dummies, and quadratic and cubic terms in age interacted with the education dummies. Panels A and B fit a standard AKM on the largest and leave-one-out connected set, respectively. Panel C presents bias-corrected versions of the estimates in Panel B based on the KSS procedure, i.e., correcting for the negative correlation between sampling errors in the person and establishment effects.

TABLE 3—DECOMPOSITION OF THE RACIAL WAGE GAP

					Decomposition of gap in establishment effects		Decomposition of sorting effect	
	Overall racial wage gap	Gap in person effects	Gap in covariate index	Gap in establishment effects	Relative wage-setting	Sorting	Skill-based sorting	Residual sorting
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Males								
All education	0.165	0.131	0.000	0.035	0.006	0.029	0.019	0.010
		79%	*0%*	*21%*	*4%*	*18%*	*12%*	*6%*
No high school	0.047	0.039	0.005	0.003	0.002	0.001	0.001	0.000
(Proportion: 52%)		*83%*	*11%*	*6%*	*5%*	*1%*	*2%*	*-1%*
High school degree	0.118	0.093	-0.001	0.026	0.011	0.015	0.011	0.004
(Proportion: 40%)		*79%*	*-1%*	*22%*	*9%*	*13%*	*9%*	*4%*
College degree	0.191	0.121	0.016	0.054	0.023	0.031	0.010	0.021
(Proportion: 8%)		*63%*	*8%*	*28%*	*12%*	*16%*	*5%*	*11%*
B. Females								
All education	0.238	0.203	-0.024	0.060	0.011	0.048	0.031	0.018
		85%	*-10%*	*25%*	*5%*	*20%*	*13%*	*7%*
No high school	0.048	0.049	-0.012	0.012	0.004	0.008	0.000	0.008
(Proportion: 33%)		*101%*	*-25%*	*25%*	*8%*	*17%*	*0%*	*17%*
High school degree	0.115	0.107	-0.024	0.032	0.014	0.019	0.008	0.011
(Proportion: 50%)		*93%*	*-21%*	*28%*	*12%*	*16%*	*7%*	*10%*
College degree	0.219	0.133	0.004	0.082	0.032	0.050	0.014	0.036
(Proportion: 17%)		*60%*	*2%*	*37%*	*15%*	*23%*	*6%*	*16%*

Notes: The table displays the results from implementing the decomposition of the average white-nonwhite gap for men (Panel A) and women (Panel B) based on equations (3)-(6). 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) reports the mean white-nonwhite pay gaps. Columns (2), (3) and (4) first decompose these gaps into differences in mean person effects, differences in means of the covariate indexes, and differences in mean establishment effects. Column (5) and (6) then decompose the gap in column (4) into a relative wage-setting effect and a sorting effect. Columns (7) and (8) finally decompose the sorting effect into a skill-based sorting effect and a residual sorting effect. Entries in italics represent the share of the overall racial wage gap (in column 1) that is explained by the source in the column heading.

Table 4—Decomposition of the Racial Wage Gap Controlling for Occupation

					Decomposition of gap in establishment effects		Decomposition of sorting effect	
	Overall racial wage gap	Gap in person effects	Gap in covariate index	Gap in establishment effects	Relative wage-setting	Sorting	Skill-based sorting	Residual sorting
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Males								
All education	0.104	0.080	0.000	0.026	0.007	0.019	0.010	0.009
		77%	*0%*	*25%*	*7%*	*18%*	*10%*	*8%*
No high school	0.039	0.032	0.005	0.004	0.002	0.002	0.001	0.000
		81%	*12%*	*9%*	*5%*	*4%*	*4%*	*1%*
High school degree	0.071	0.055	-0.003	0.021	0.011	0.010	0.004	0.006
		78%	*-4%*	*30%*	*16%*	*14%*	*6%*	*8%*
College degree	0.106	0.050	0.008	0.050	0.023	0.027	0.004	0.023
		47%	*7%*	*48%*	*22%*	*26%*	*4%*	*22%*
B. Females								
All education	0.125	0.108	-0.022	0.042	0.012	0.029	0.012	0.017
		87%	*-18%*	*33%*	*10%*	*23%*	*10%*	*14%*
No high school	0.020	0.026	-0.012	0.008	0.003	0.005	0.000	0.005
		129%	*-61%*	*41%*	*13%*	*28%*	*0%*	*27%*
High school degree	0.060	0.063	-0.026	0.025	0.014	0.011	-0.001	0.012
		106%	*-43%*	*42%*	*23%*	*19%*	*-1%*	*20%*
College degree	0.135	0.063	-0.005	0.080	0.031	0.049	0.013	0.036
		47%	*-3%*	*59%*	*23%*	*36%*	*10%*	*26%*

Notes: The table displays the results from implementing a similar decomposition of the average white-nonwhite gap as in Table 3, but adjusting for observed differences in occupation, using six major occupation groups. Specifically, we modified the procedures underlying Table 3 in two ways. First, we reweighted nonwhites to have the same joint distribution as whites across micro-regions *and occupations*. Second, for the decompositions in column (7) and (8), we re-classified the workforce of each establishment in each year into four age cells, four quartiles of estimated person effects, *and six occupations*, and calculated the expected share of nonwhites at the establishment if it hired at random from its surrounding micro-region within these $4 \times 4 \times 6 = 96$ cells.