Online Appendix

Appendix A: Data Construction

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

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

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

---

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

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

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

---

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

Appendix B: A Simple Model of Monopsonistic Wage Setting

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

Worker Preferences

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

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

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

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

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

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

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

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

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

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

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

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

where d_1 and d_2 are market specific constants.

Firm Optimization

Firms have production functions of the form:

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

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

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

The associated first order conditions can be written:

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

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

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

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

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

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

4

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

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

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

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

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

A Simple Benchmark: Linear Production and Fixed Output Price

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

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

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

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

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

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

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

Now the first order conditions can be re-written:

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

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

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

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

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

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

Implications for AKM-Style Wage Models

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

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

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

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

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

References

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

Figure C1: Brazil's Regions and Racial Composition of its Micro-Regions
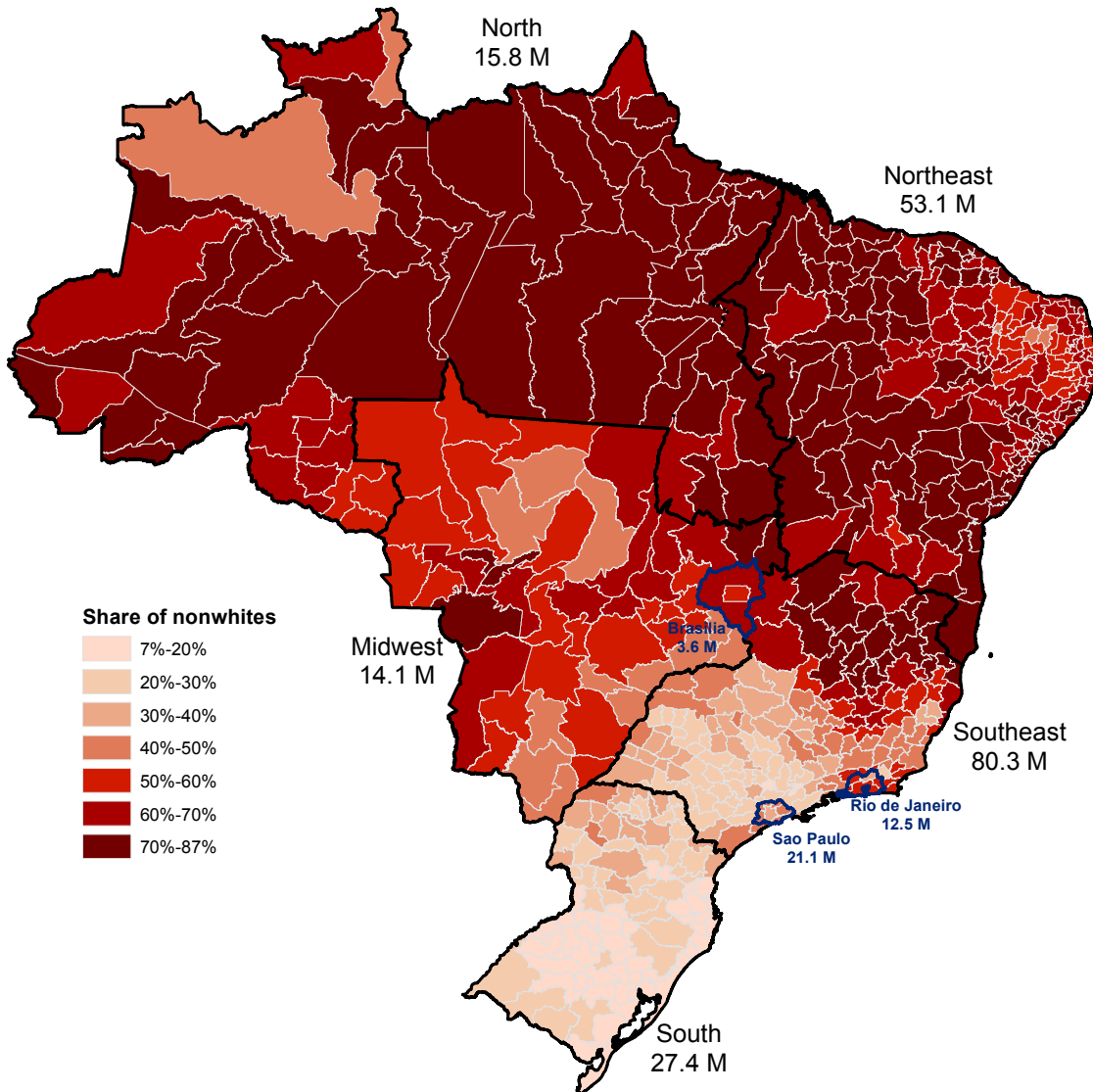
Notes: The figure displays a map of Brazil; the black lines correspond to the borders of Brazil's five regions (North, Northeast, Southeast, South, and Midwest; a region's population according to the 2010 census is reported under its name); the white lines correspond to the borders of Brazil's 557 micro-regions; and blue lines identify Brazil's two largest cities (São Paulo and Rio de Janeiro) and capital (Brasilia). The coloring provides information on the share of nonwhites in a micro-region's population according to the 2010 census.

Figure C2: Log Hourly Wage Distributions Among Private-Sector Employees (Females)

(a) All education levels (Southeast)

(b) All education levels (Northeast)

(c) No high school (Southeast)

(d) No high school (Northeast)

(e) Completed high school (Southeast)

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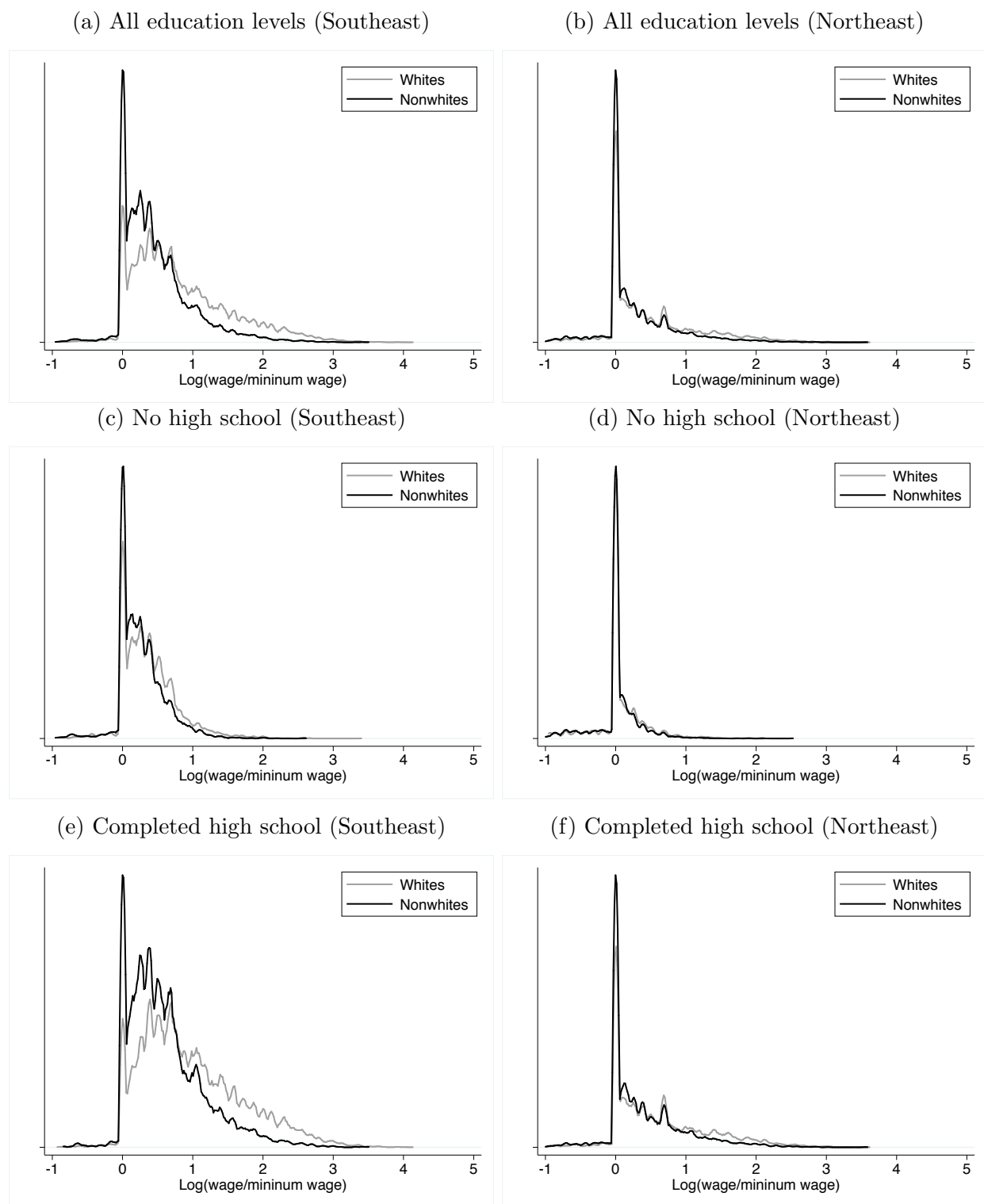
Notes: The figure displays kernel densities (Epanechnikov kernel with a 0.025 half-width) of the log wage-to-minimum-wage ratio for whites and nonwhites, based on PNAD 2002-2014 and constructed using survey weights (PNAD was not conducted in 2010). The samples include female non-farm private-sector employees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

Figure C3: Age-Wage Profiles by Cohort

(a) White males

(b) Non-white males

(c) White females

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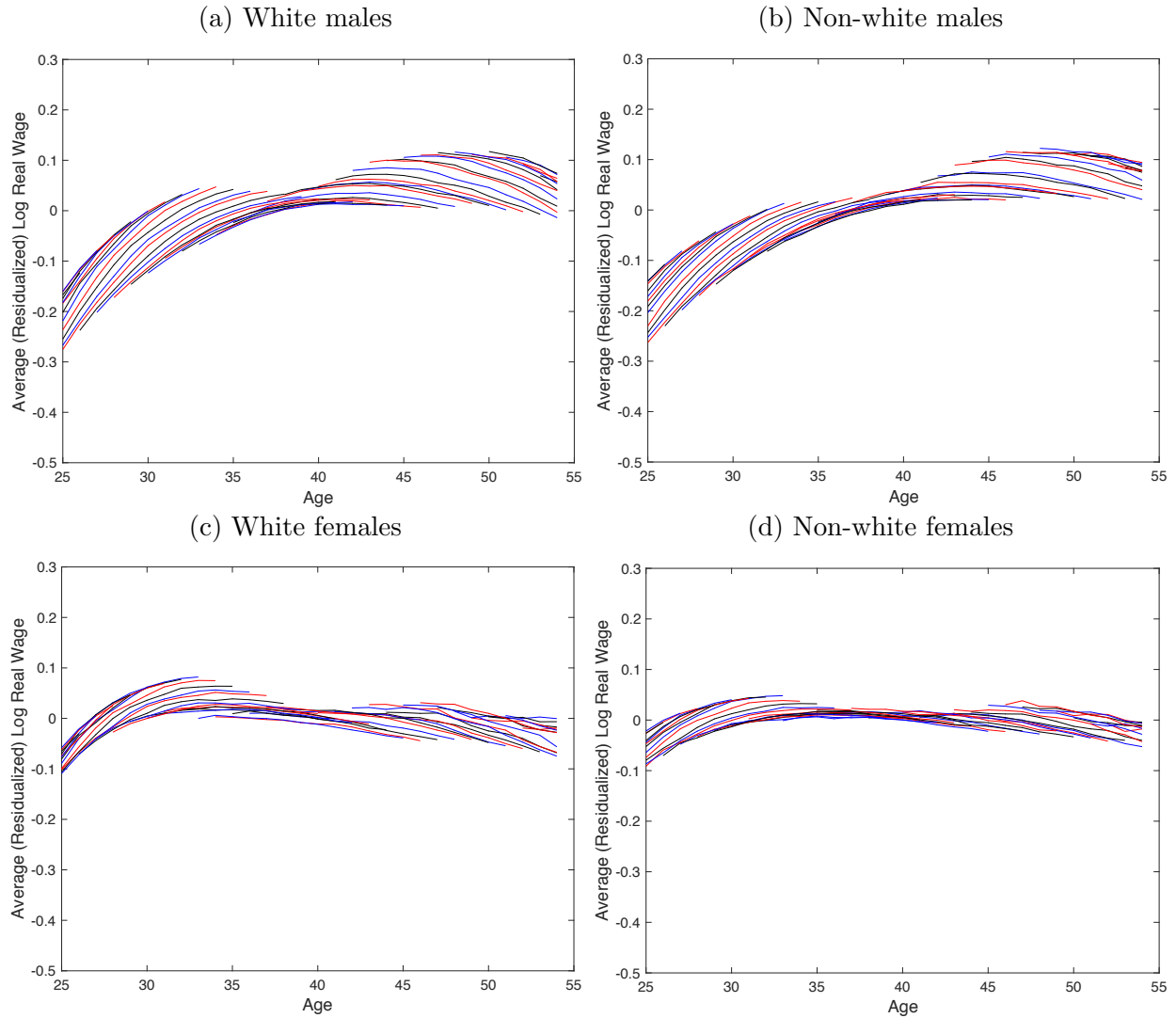
Notes: The figure displays the age-wage profile by cohort, separately for each race-gender group. Specifically, each panel shows the residuals of a regression of mean log wages by age and cohort on year fixed effects; the lines correspond to age-wage profiles for different cohorts. The mean log wages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

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

(a) White males

Slope: .994 (.000)

Correlation: .994

(b) Non-white males

Slope: .987 (.000)

Correlation: .990

(c) White females

Slope: .994 (.000)

Correlation: .995

(d) Non-white females

Slope: .989 (.000)

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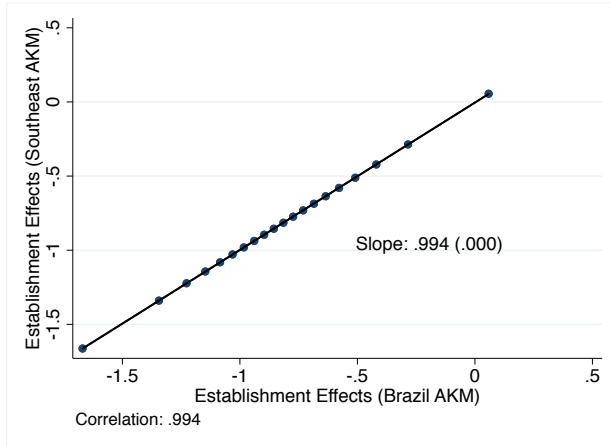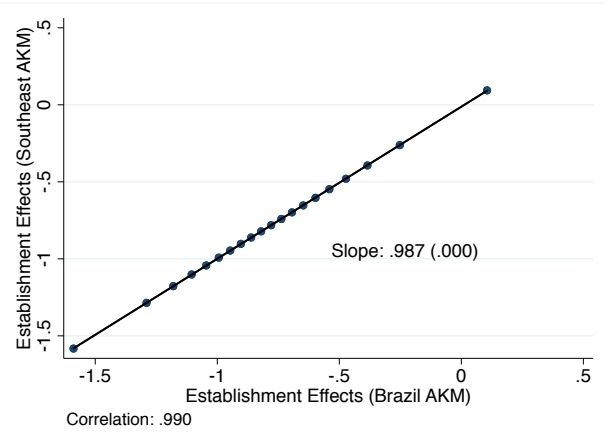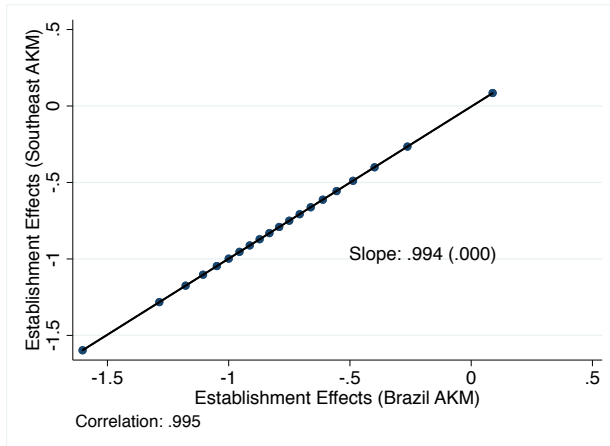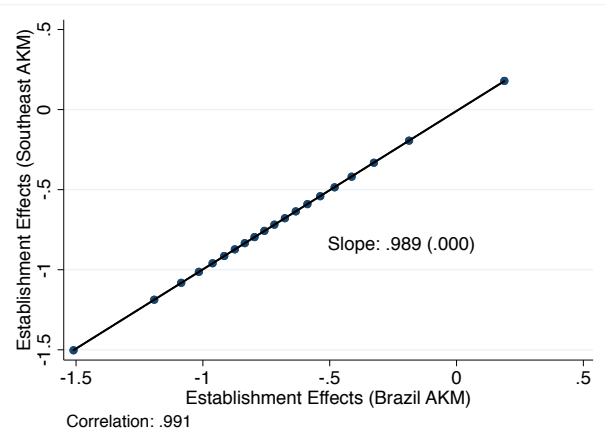
Notes: The figure displays binned scatterplots of establishment effects from AKM models estimated on the dual connected set of each race-gender group in the Southeast region versus in all of Brazil. Observations are unique establishments that appear in both the Southeast and Brazil samples. Scatterplots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

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

(a) White males

(b) Non-white males

(c) White females

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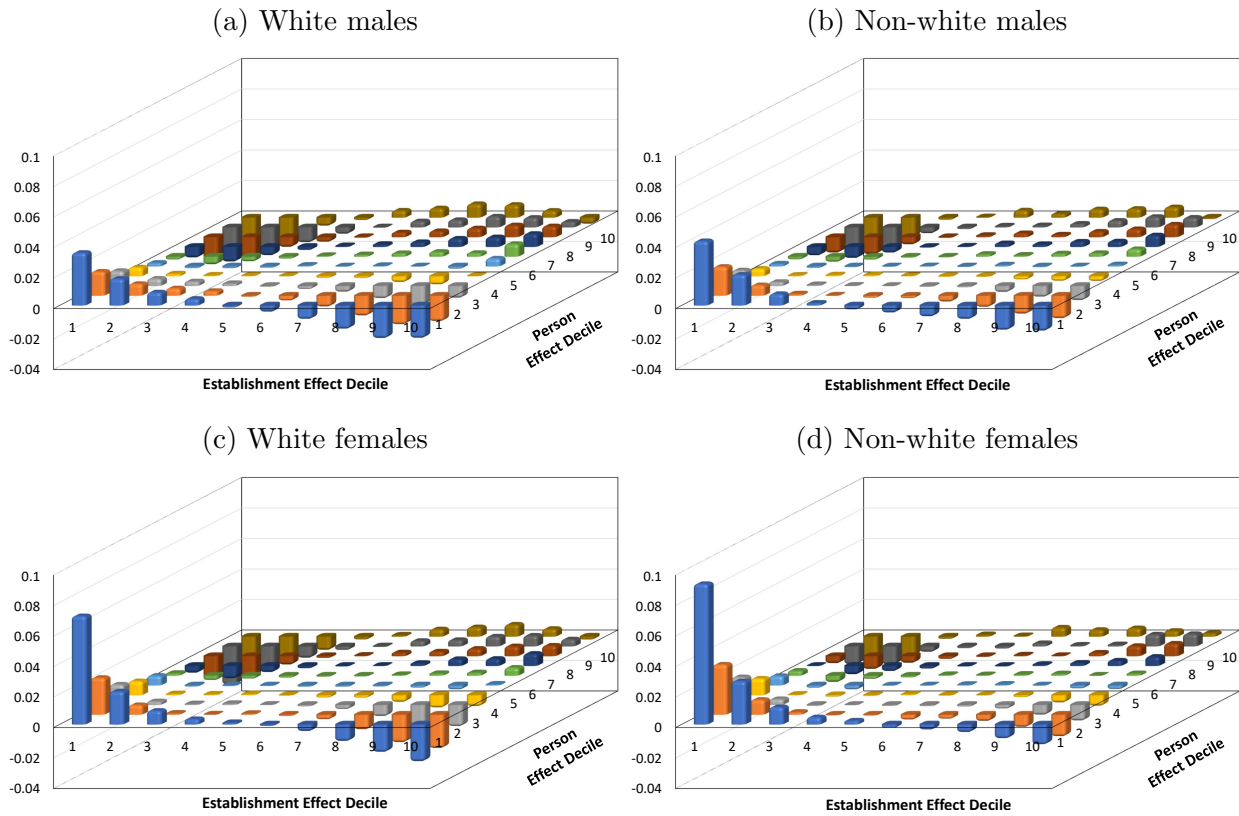
Notes: The figure displays mean residuals from AKM models estimated on the largest connected set of each race-gender group in the Southeast region, for 100 cells defined by deciles of estimated establishment effects interacted with deciles of estimated person effects. The mean residuals in each cell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

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

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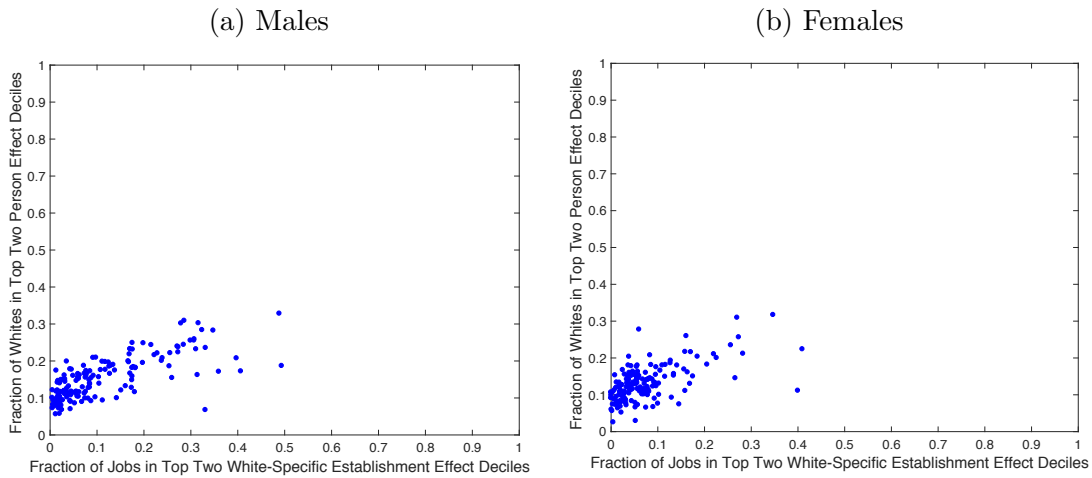
Notes: The figure displays scatterplots of the relationship between the share of jobs in high-premium establishments and the share of high-skilled white workers across micro-regions, based on person-year observations in the dual-connected set of each gender in the Southeast region. Each dot corresponds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

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

(a) Males (all education)

(b) Females (all education)

(c) Males (no high school)

(d) Females (no high school)

(e) Males (high school)

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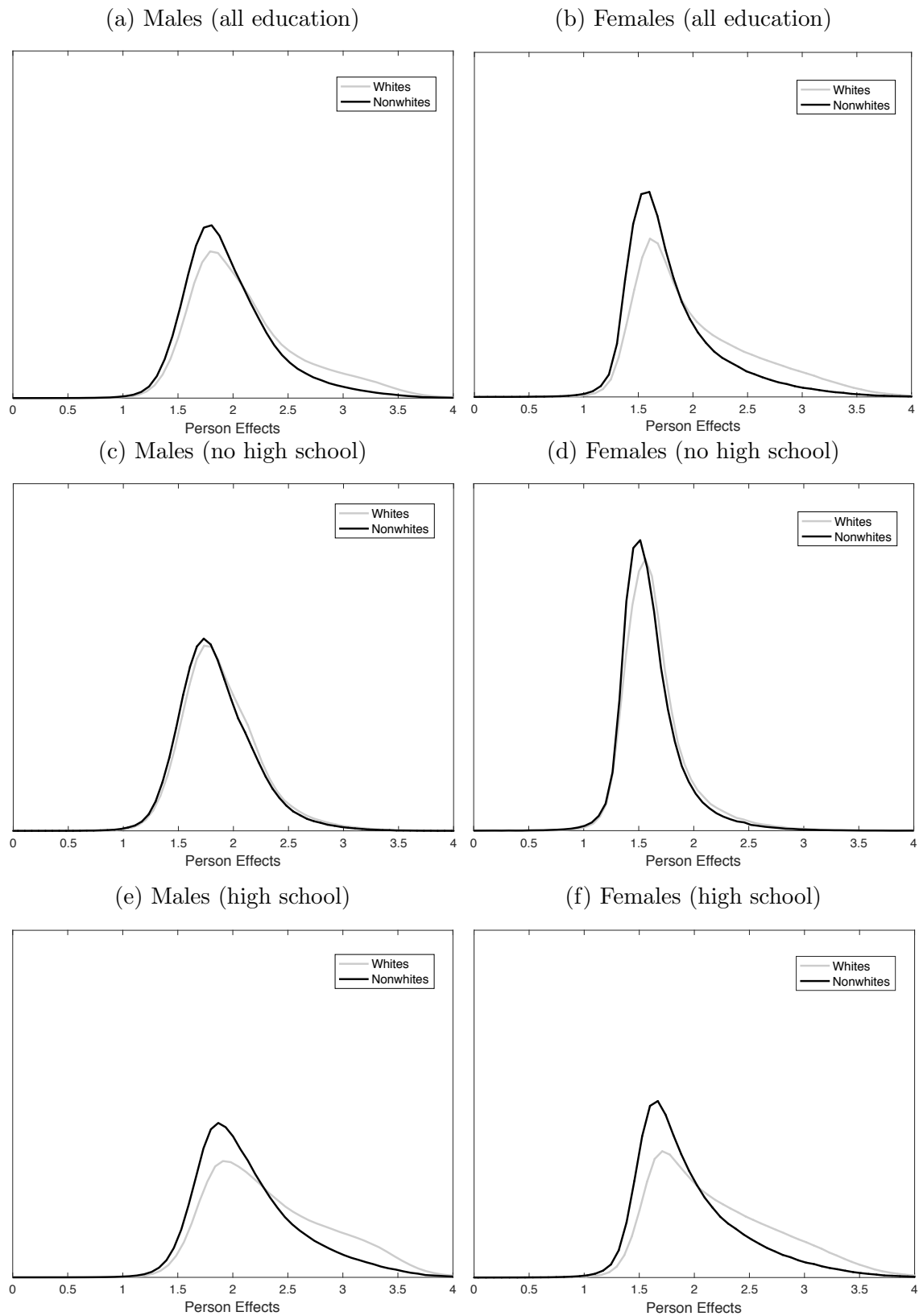
Notes: The figure displays kernel densities (normal kernel with optimal bandwidth) of the estimated person effects, using person-year observations in the dual-connected set of each gender (and education) group in the Southeast region. Nonwhites are reweighted so as to have the same distribution across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

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

(a) Males (all education)

(b) Females (all education)

(c) Males (completed high school)

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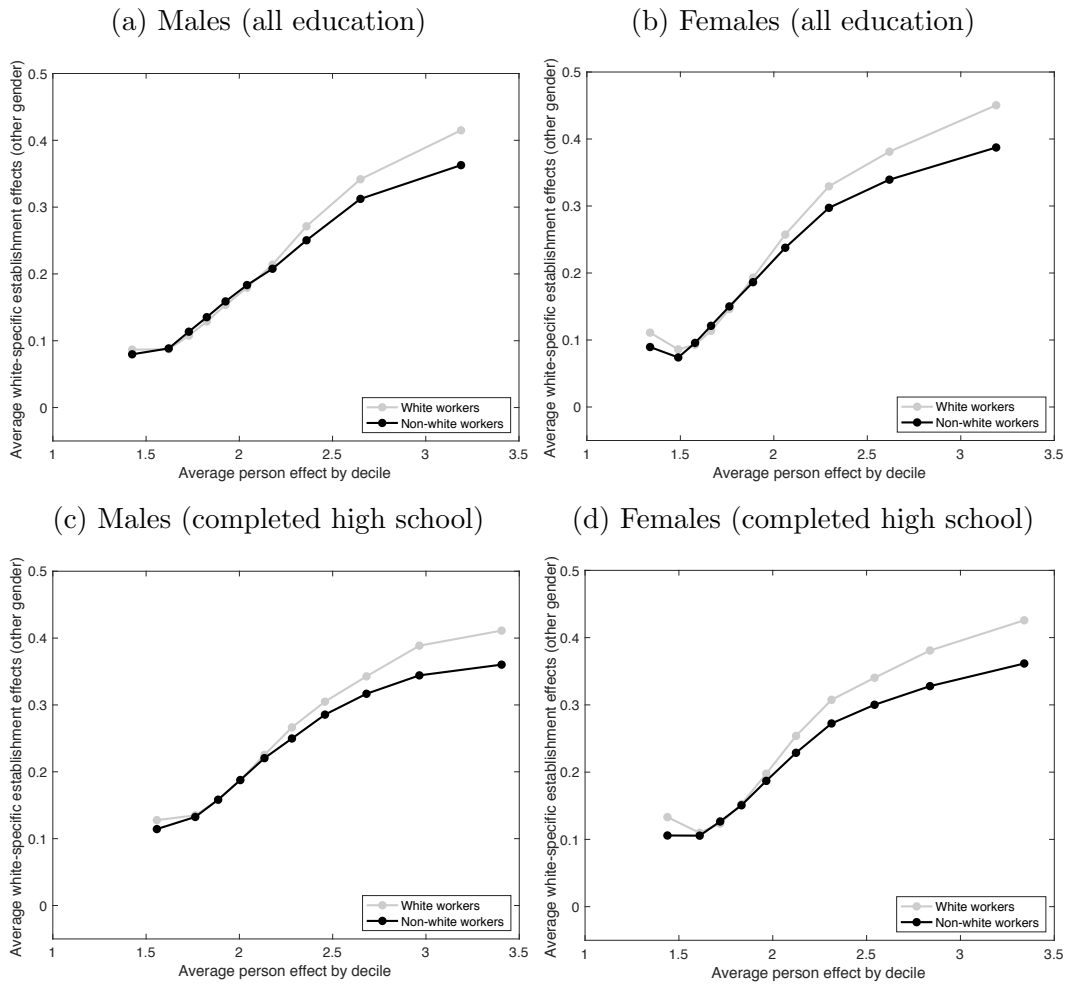
Notes: The figure is constructed similarly as Figure 8, but uses the establishment effects of the other gender to show that the main patterns in Figure 8 are robust to the sampling error issue discussed in the text. The figure displays the means of the estimated white-specific establishment effects for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

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

(a) Males (all education)

(b) Females (all education)

(c) Males (completed high school)

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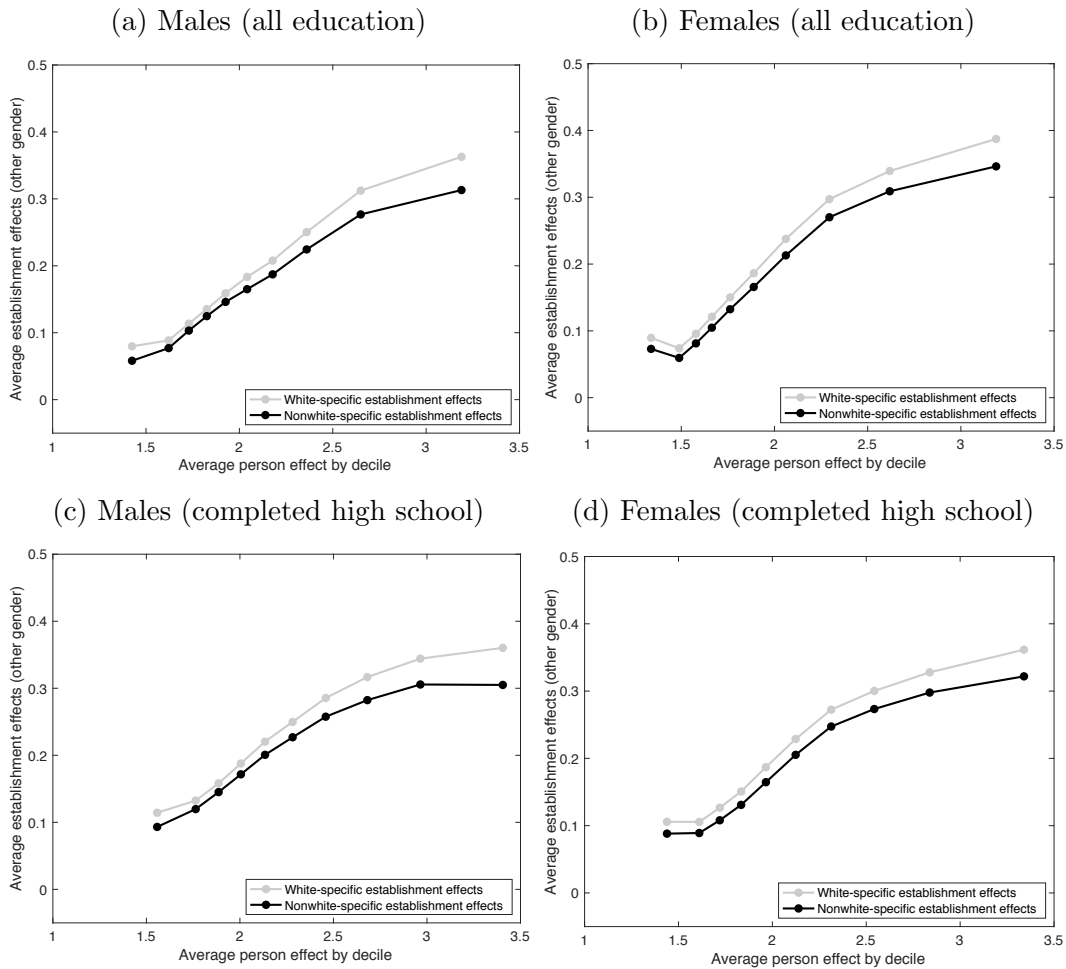
Notes: The figure is constructed similarly as Figure 9, but uses the establishment effects of the other gender to show that the main patterns in Figure 9 are robust to the sampling error issue discussed in the text. The figure displays the means of the estimated white-specific and nonwhite-specific establishment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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Table D7: Rank Correlation of Sector Premiums

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

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

Table D8: Relating Decomposition Results to Standard Mincerian Model

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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