ONLINE APPENDIX

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

François Gerard (Queen Mary University of London and CEPR)
Lorenzo Lagos (Princeton University and IZA)
Edson Severnini (Carnegie Mellon University and IZA)
David Card (UC Berkeley and NBER)

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) nonfarm 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

[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.

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 above the 99th percentile of the state- and year-specific wage distribution. Fourth, the log hourly wage changes by more than 100 log points between consecutive observations. 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 D.5 is based 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 D.5.

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:

$$u_{igj} = \delta_g^0 \ln(w_{gj} - b_g) + a_{gj}^0 + v_{igj}, \tag{B1}$$

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

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

$$p_{gj} \equiv P(u_{igj} = \underset{k \in \{1,..,J\}}{\arg\max} \{u_{igk}\}) = \frac{\exp(\delta_g(\ln(w_{gj} - b_g) + a_{gj})}{\sum_{k=1}^{J} \exp(\delta_g \ln(w_{gk} - b_g) + a_{gk})}, \tag{B2}$$

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 B2 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:

$$\ln N_{1j}(w_{1j}) = d_1 + \delta_1 \ln(w_{1j} - b_1) + a_{1j} \tag{B3}$$

$$\ln N_{2j}(w_{2j}) = d_2 + \delta_2 \ln(w_{2j} - b_2) + a_{2j}, \tag{B4}$$

where d_1 and d_2 are market specific constants.

Firm Optimization

Firms have production functions of the form:

$$Y_j = T_j f(N_{1j}, N_{2j}), \tag{B5}$$

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 (B3) and (B4). 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:

$$w_{1j} = \frac{e_{1j}}{1 + e_{1j}} T_j f_1 \mu_j \tag{B6}$$

$$w_{2j} = \frac{e_{2j}}{1 + e_{2j}} T_j f_2 \mu_j \tag{B7}$$

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 (B6) and (B7) 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 (B3) and (B4), the elasticities of supply are:

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

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

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:

$$w_{1j} = \frac{1}{1+\delta_1} b_1 + \frac{\delta_1}{1+\delta_1} T_j f_1 \mu_j \tag{B8}$$

$$w_{2j} = \frac{1}{1+\delta_2} b_2 + \frac{\delta_2}{1+\delta_2} T_j f_2 \mu_j. \tag{B9}$$

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 (B8) and (B9) evaluate to:

$$\begin{aligned} w_{1j} &= \frac{1}{1+\delta_1} b_1 + \frac{\delta_1}{1+\delta_1} T_j P_j^0 \theta_1 \\ w_{2j} &= \frac{1}{1+\delta_2} b_2 + \frac{\delta_2}{1+\delta_2} T_j P_j^0 \theta_2. \end{aligned}$$

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:

$$\ln w_{1j} = \ln \frac{\theta_1 b}{1+\delta_1} + \ln(1 + \delta_1 R_j) \tag{B10}$$

$$\ln w_{2j} = \ln \frac{\theta_2 b}{1+\delta_2} + \ln(1 + \delta_2 R_j) \tag{B11}$$

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:

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

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 B10 and B11, suppose that $\delta_g R_j$ is relatively small. In this case:

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

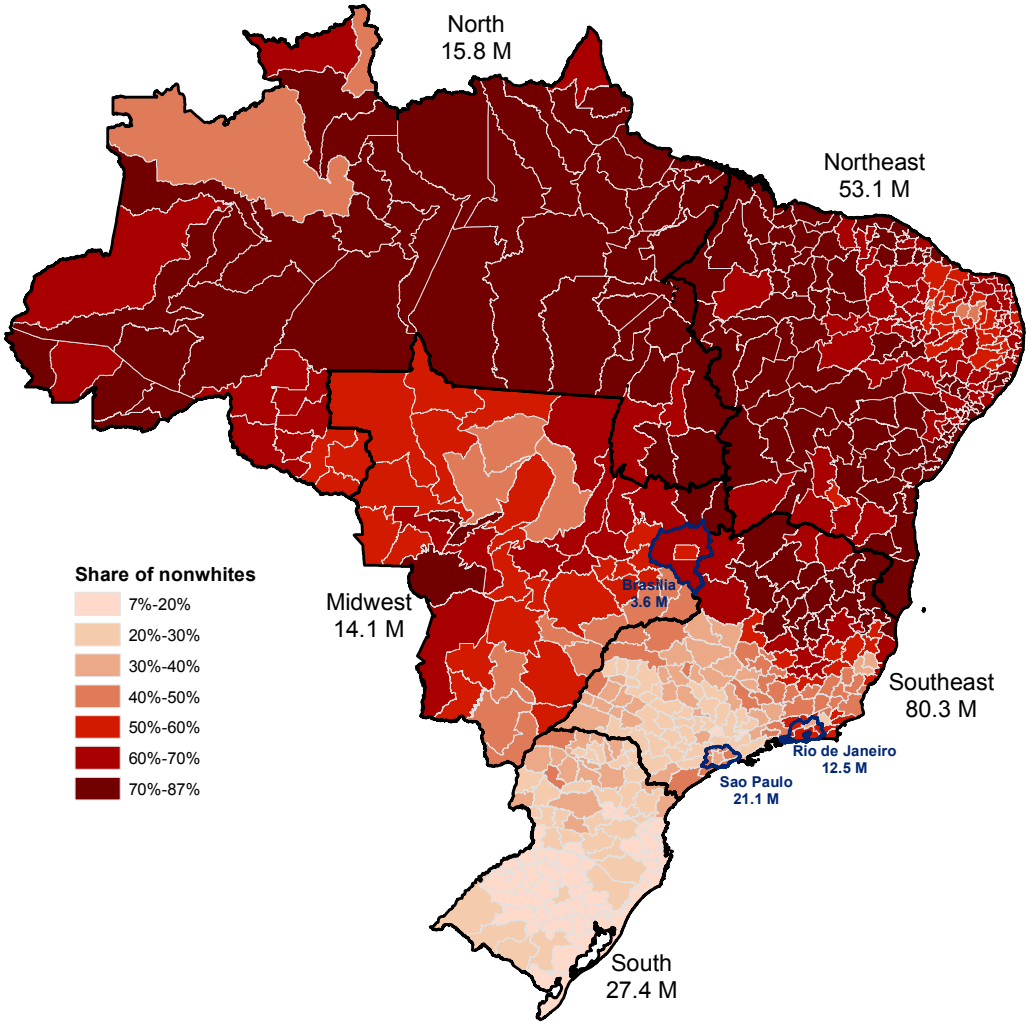
These equations imply that the wages of workers at different firms can be written in the form:

$$\ln w_{gj} = \alpha_g + \psi_j^g, \tag{B12}$$

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.

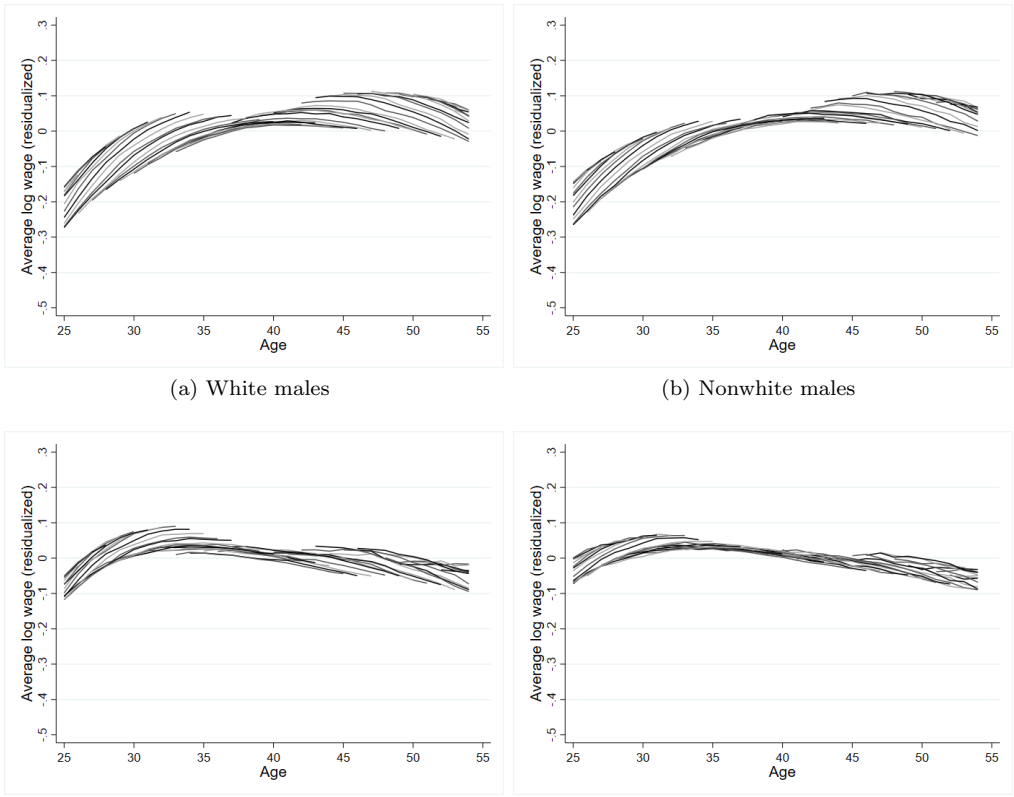
Appendix C: Additional Figures

FIGURE C.1. 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; each 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 the blue lines identify Brazil's two largest cities (São Paulo and Rio de Janeiro) and the capital (Brasilia). The coloring provides information on the share of nonwhites in a micro-region's population according to the 2010 census.

Figure C.2. Age-wage Profiles by Cohort

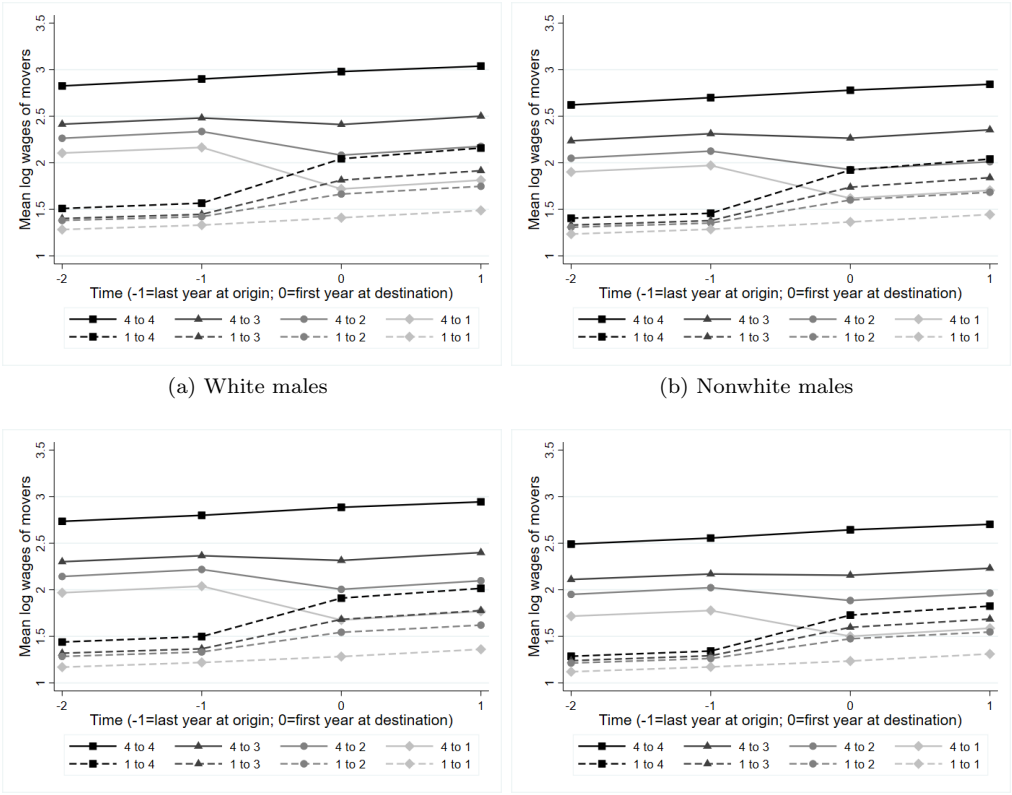


(a) White males (b) Nonwhite males

(c) White females (d) Nonwhite 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 D.5. 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 C.3. Event Studies Around Job Moves

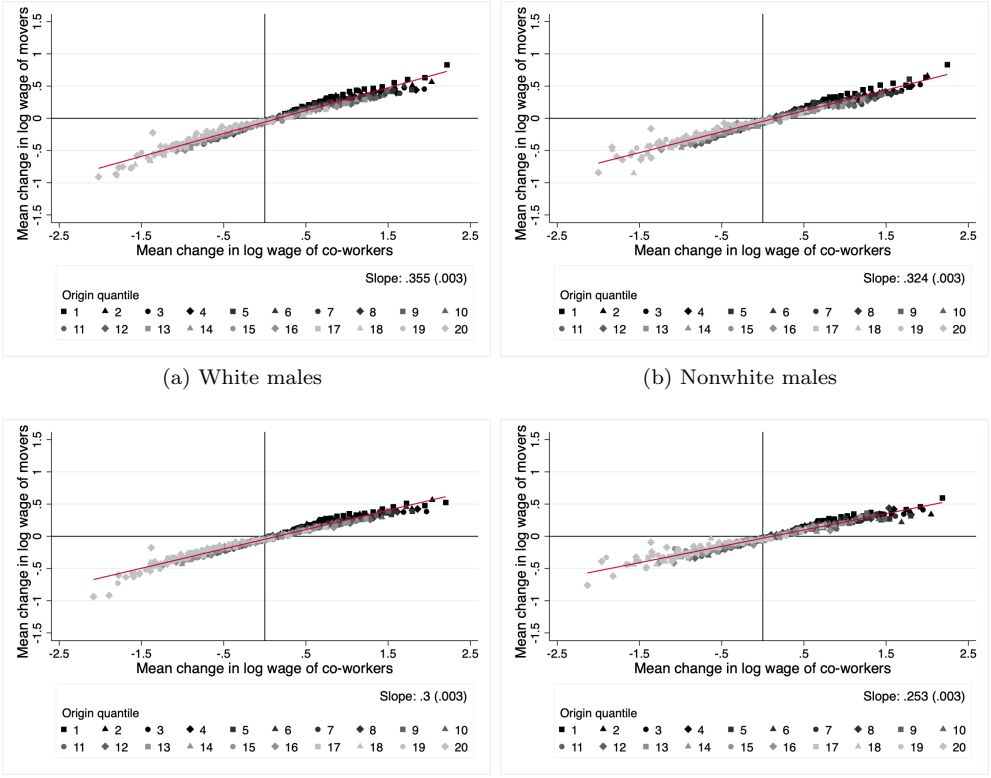


(a) White males (b) Nonwhite males

(c) White females (d) Nonwhite females

Notes: The figure displays the evolution of wages for workers in the Southeast region who moved from origin establishments in the top and bottom quartile groups to destination establishments in any of the other quartile groups. We use the samples described in columns (1)-(4) in Table D.5; the movers are defined as workers at establishments employing at least one worker of each race-gender group, who separated from the origin establishment in 2003-2012, were reemployed in the destination establishment the next or the following year, and were employed at the origin and destination establishments for 2+ consecutive years. Origin/destination groups are based on quartiles of co-worker wages during the calendar year of separation/hiring.

Figure C.4. Wage Changes of Movers vs. Changes of Co-worker Wages

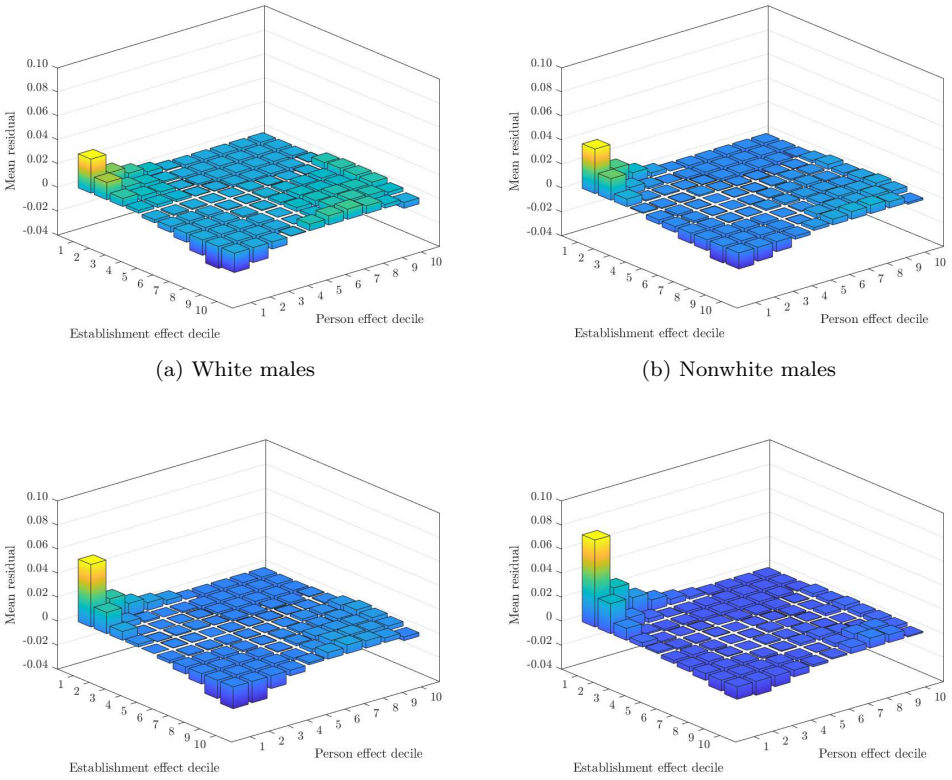


(a) White males (b) Nonwhite males

(c) White females (d) Nonwhite females

Notes: The figure plots the mean change in movers' wages between the years prior to separation and after hiring against the mean change in co-worker wages between origin and destination establishments. Origin and destination establishments are grouped by 20 quantiles of co-worker wages; each of the 20x20 dots corresponds to movers from/to given origin/destination quantiles. Movers are defined as in Figure C.3; their wage changes are adjusted for trends based on coefficients from a regression estimated on the sample of stayers, i.e., workers who remain at the same origin establishments over the years around a move. The model includes the same education dummies as in Table D.1 and a quadratic in age fully interacted with these dummies.

Figure C.5. Mean AKM Residuals by Person Effect and Establishment Effect Deciles



(a) White males

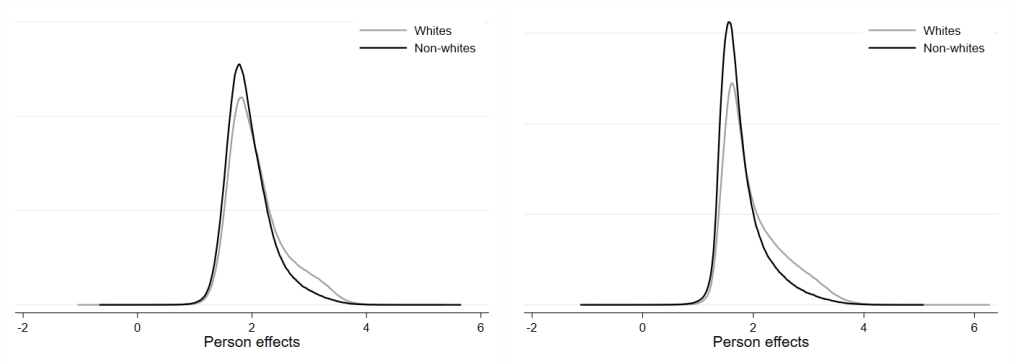
(b) Nonwhite males

(c) White females

(d) Nonwhite 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 VI.

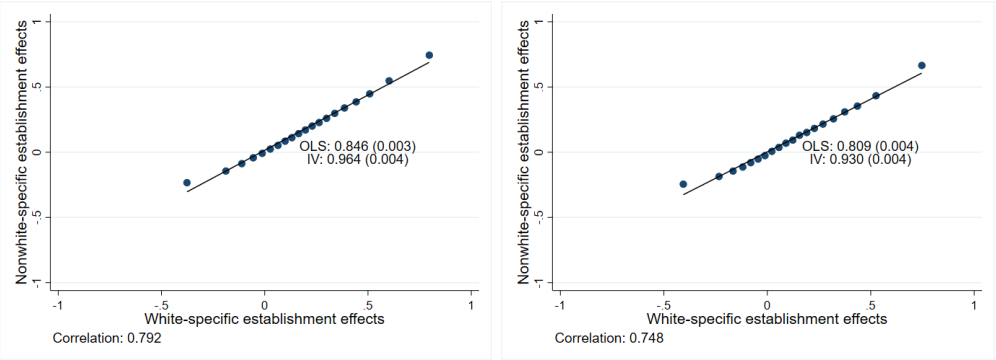
Figure C.6. Distribution of Estimated Person Effects for White and Nonwhite Workers



(a) Males (b) Females

Notes: The figure displays kernel densities (Epanechnikov kernel with optimal bandwidth) of the estimated person effects, using person-year observations in 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).

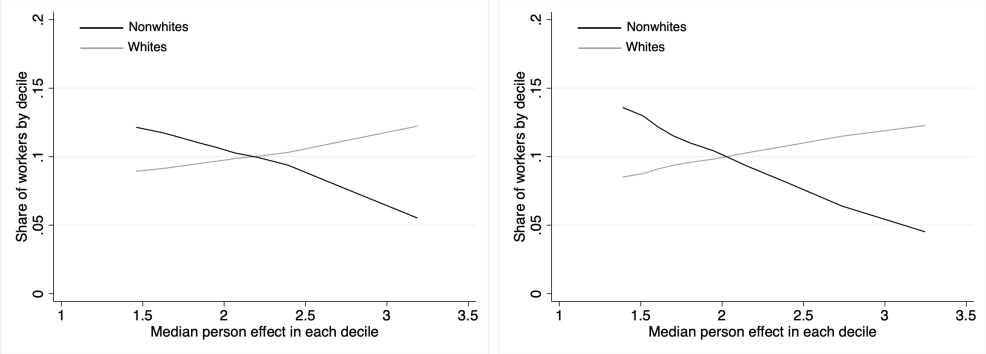
Figure C.7. Correlation of Establishment Effects for Whites and Nonwhites



(a) Males (b) Females

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

FIGURE C.8. DISTRIBUTION OF PERSON-YEAR OBSERVATIONS IN THE EARLIER SAMPLE (2002-2008) BY PERSON EFFECT DECILES BASED ON ESTIMATES FROM THE LATER SAMPLE (2008-2014)



(a) Males (b) Females

Notes: The figure displays the distribution of person-year observations from the 2002-2008 subsample by the median person effect in each decile (of the estimated person effects in the 2008-2014 subsample). Samples are restricted to workers who appear in the dual-connected set of each gender group in the Southeast region in both the earlier period and later period subsamples. Nonwhites are reweighted so as to have the same distribution across micro-regions as whites (of the same gender). These distributions correspond to the distributions underlying the analyses in Figure 7.

Appendix D: Additional Tables

TABLE D.1—RACIAL DIFFERENCES IN LOG WAGES IN PNAD (PRIVATE-SECTOR EMPLOYEES)

	Brazil		Southeast region	
	(1)	(2)	(3)	(4)
A. Males				
Dummy if mixed race	-0.27	-0.11	-0.30	-0.11
(std err.)	(0.00)	(0.00)	(0.00)	(0.00)
Dummy if black	-0.29	-0.13	-0.33	-0.14
(std err.)	(0.00)	(0.00)	(0.00)	(0.00)
Year and state fixed effects	yes	yes	yes	yes
Education and experience	no	yes	yes	yes
B. Females				
Dummy if mixed race	-0.28	-0.11	-0.31	-0.11
(std err.)	(0.00)	(0.00)	(0.00)	(0.00)
Dummy if black	-0.30	-0.11	-0.33	-0.11
(std err.)	(0.00)	(0.00)	(0.00)	(0.00)
Year and state fixed effects	yes	yes	yes	yes
Education and experience	no	yes	no	yes

Notes: The table displays the results of regressing log hourly wages on a series of race group dummies, using data from PNAD 2002-2014 (PNAD was not conducted in 2010). The samples include male (panel A) and female (panel B) nonfarm 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. The samples in columns (1)-(2) use data for the whole country; columns (3)-(4) restrict the samples to the Southeast region only.

Table D.2—Racial Differences in Formality and Log Wages in PNAD and RAIS (Private-Sector Employees)

	Brazil				Southeast region			
	Formality	Log hourly wage			Formality	Log hourly wage		
	PNAD	PNAD	PNAD-formal	RAIS	PNAD	PNAD	PNAD-formal	RAIS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Males - All education								
Dummy if nonwhite	0.00	-0.11	-0.11	-0.06	0.00	-0.12	-0.11	-0.07
(std err.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Mean for whites	0.82	1.78	1.85	1.92	0.84	1.89	1.94	1.98
B. Males - Completed high school								
Dummy if nonwhite	0.01	-0.16	-0.16	-0.10	0.01	-0.17	-0.16	-0.10
(std err.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Mean for whites	0.86	2.10	2.13	2.14	0.87	2.19	2.21	2.22
C. Females - All education								
Dummy if nonwhite	-0.01	-0.11	-0.10	-0.08	0.00	-0.11	-0.11	-0.09
(std err.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Mean for whites	0.82	1.62	1.67	1.77	0.83	1.70	1.75	1.84
D. Females - Completed high school								
Dummy if nonwhite	-0.01	-0.14	-0.13	-0.10	0.00	-0.14	-0.13	-0.11
(std err.)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Mean for whites	0.85	1.82	1.85	1.95	0.86	1.89	1.92	2.03

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

Table D.3—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	0.02	0.00	0.02	0.00	0.03	0.01	0.03	0.01
(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	0.03	0.01	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 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) nonfarm private-sector formal employees, age 25 to 54, with potential labor market experience of at least 1 year, and non-missing data on race, gender, education, wage, and hours worked. All specifications include year and state fixed effects and use survey weights. Education and experience controls include five education dummies (incomplete elementary school, and complete elementary school, middle school, high school, or college) and a quadratic in potential experience. The omitted race group is white. Other race dummies not reported are indigenous and asian. The samples in columns (1)-(4) use data for the whole country; columns (5)-(8) restrict the samples to the Southeast region only. In each case, the first two columns pool workers of all education levels; the other columns restrict the sample to workers with completed high school.

TABLE D.4—WAGE DIFFERENTIALS IN THE NORMALIZING INDUSTRY AND OTHER INDUSTRIES

	PNAD		RAIS	
	Restaurant industry	All other industries	Restaurant industry	All other industries
	(1)	(2)	(3)	(4)
Dummy if nonwhite	-0.034	-0.120	-0.024	-0.067
Dummy if female	-0.261	-0.326	-0.181	-0.298
Nonwhite *x* female	-0.002	0.016	-0.003	-0.035

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 D.2 column (7) and using the pooled RAIS dual-connected sets for males and females from the Southeast region as in Table D.2 column (8)—and described in Table D.5 columns (9)-(12). All specifications include year and state effects, the same education dummies as in Table D.1, 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 D.5—Descriptive Statistics Comparing the Analysis Samples, Largest Connected Sets, and Dual-Connected Sets

	Analysis samples (all valid observations)				Largest connected sets (by race-gender group)				Dual-connected sets (by gender)			
	White male	Non-white male	White female	Non-white female	White male	Non-white male	White female	Non-white female	White male	Non-white male	White female	Non-white female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Age												
Mean age	36.5	36.3	35.6	35.6	36.4	36.2	35.5	35.6	36.5	36.2	35.5	35.6
Share ≤ 30 years old	0.30	0.30	0.34	0.33	0.30	0.30	0.34	0.33	0.30	0.30	0.33	0.33
Share ≥ 50 years old	0.08	0.08	0.06	0.06	0.08	0.08	0.06	0.06	0.08	0.08	0.06	0.06
Education												
Mean years of schooling	9.3	8.4	10.6	9.6	9.3	8.5	10.6	9.7	9.4	8.5	10.7	9.7
Share completed high school	0.51	0.41	0.69	0.59	0.51	0.41	0.69	0.60	0.52	0.41	0.70	0.60
Share completed college	0.09	0.04	0.18	0.08	0.10	0.04	0.18	0.08	0.10	0.04	0.20	0.08
Wages and Hours												
Mean log hourly wage (R$)	1.92	1.76	1.76	1.54	1.94	1.77	1.78	1.56	1.98	1.78	1.84	1.57
(standard dev.)	(0.67)	(0.58)	(0.68)	(0.55)	(0.67)	(0.58)	(0.69)	(0.55)	(0.68)	(0.58)	(0.70)	(0.56)
Mean monthly hours	188.4	188.6	184.6	185.5	188.3	188.5	184.3	185.1	188.1	188.5	183.2	184.9
Share full-time	1.00	1.00	0.98	0.99	1.00	1.00	0.98	0.99	1.00	1.00	0.98	0.99
Establishment Characteristics												
Mean establishment size	460	595	486	708	477	630	514	787	571	654	681	826
Share females at establishment	0.26	0.24	0.58	0.59	0.26	0.23	0.58	0.59	0.25	0.23	0.59	0.60
Share whites at establishment	0.76	0.54	0.77	0.55	0.75	0.53	0.77	0.54	0.73	0.54	0.74	0.55
Sample Sizes												
No. person-year obs.	41,141,880	17,591,624	29,435,312	9,906,003	39,661,512	16,605,082	27,814,348	8,900,093	33,005,160	15,996,350	20,727,960	8,478,319
No. persons	8,618,385	3,918,056	6,679,431	2,495,441	8,187,549	3,633,690	6,171,704	2,169,170	7,224,486	3,566,348	4,996,945	2,113,524
No. establishments	1,617,459	955,214	1,556,637	777,189	1,284,740	717,098	1,162,373	508,088	578,107	578,107	400,464	400,464

Notes: The table displays descriptive statistics by race-gender group in three samples based on RAIS 2002-2014. The analysis samples in columns (1)-(4) include nonfarm private-sector formal employees in the Southeast region, age 25 to 54, with potential labor market experience of at least 1 year, tenure of at least 1 month, and non-missing data on race, gender, education, wage, and hours worked, who are employed on December 31st of each year with an open-ended contract in which they are paid on a monthly basis. The 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 D.6—LOG ADDITIVITY TEST USING AKM ESTIMATES BY EDUCATION

	White		Non-white	
	Completed high school on All education	No high school on All education	Completed high school on All education	No high school on All education
	(1)	(2)	(3)	(4)
A. Males				
OLS (OLS sample)	1.0011	1.0225	0.9919	1.0169
	(0.0010)	(0.0039)	(0.0019)	(0.0058)
OLS (IV sample)	0.9934	1.0284	0.9834	1.0306
	(0.0013)	(0.0054)	(0.0027)	(0.0086)
IV (IV sample)	1.0001	1.0297	0.9875	1.0393
	(0.0016)	(0.0063)	(0.0034)	(0.0097)
B. Females				
OLS (OLS sample)	1.0241	0.9239	1.0277	0.9315
	(0.0013)	(0.0088)	(0.0024)	(0.0078)
OLS (IV sample)	1.0149	0.9424	1.0145	0.9533
	(0.0016)	(0.0104)	(0.0028)	(0.0097)
IV (IV sample)	1.0134	0.9613	1.0069	0.9766
	(0.0016)	(0.0112)	(0.0032)	(0.0114)
C. Females, dropping bottom 10% of estab. effects distribution				
OLS (OLS sample)	1.0065	0.9603	1.0065	0.9691
	(0.0018)	(0.0110)	(0.0031)	(0.0098)
OLS (IV sample)	1.0014	0.9698	0.9987	0.9785
	(0.0021)	(0.0124)	(0.0034)	(0.0112)
IV (IV sample)	1.0011	0.9887	0.9919	0.9996
	(0.0021)	(0.0132)	(0.0038)	(0.0132)

Notes: The table displays coefficients and standard errors (in parentheses) from regressing the estimated establishment fixed effects from an AKM decomposition using only workers with a completed high school degree—columns (1) and (3)—or using only workers without a completed high school degree—columns (2) and (4)—on the estimated establishment effects used in the main analyses, i.e., pooling all education levels and restricting to dual-connected sets. Samples are restricted to observations in the intersection of the largest connected sets of both education levels for each race-gender group in the Southeast region ("OLS sample"). The table also displays results form an IV specification using as instrument the estimated establishment effect for the other gender (but same race), further restricting attention to establishments that also appear in the largest connected set of the other gender ("IV sample"). All establishment effects are normalized relative to the restaurant industry. Observations without completed high school are reweighted so that they have the same distribution across micro-regions as workers with completed high school (of the same race-gender group). Columns (1) and (2) use white-specific establishment effects, while columns (3) and (4) use nonwhite-specific establishment effects. Panels A and B show results for males and females, respectively. Panel C presents the results for females after dropping the bottom 10% of the establishment effect distribution to minimize the impact of the minimum wage.

TABLE D.7—ROLE OF LOCATIONAL DIFFERENCES IN THE DECOMPOSITION OF THE RACIAL WAGE GAP

	Overall racial wage gap	Component attributable to: Person effects	Covariate index	Establishment effects
	(1)	(2)	(3)	(4)
A. Males				
Unadjusted for locational differences	0.204	0.145 *71%*	0.001 *1%*	0.058 *28%*
Adjusted for locational differences (nonwhites reweighted to white distribution)	0.165	0.131 *79%*	0.000 *0%*	0.035 *21%*
B. Females				
Unadjusted for locational differences	0.271	0.203 *75%*	-0.020 *-7%*	0.087 *32%*
Adjusted for locational differences (nonwhites reweighted to white distribution)	0.238	0.203 *85%*	-0.024 *-10%*	0.060 *25%*

Notes: The table displays the results from implementing the decomposition of the average white-nonwhite pay gap based on equation (3) using person-year observations in the dual-connected set of each gender in the Southeast region, without and with reweighting so that nonwhites have the same distribution across micro-regions as whites (of the same gender). See text for a detailed explanation of the decomposition. Entries in italics represent the share of the overall racial wage gap (in column 1) that is explained by the source in the column heading.

Table D.8—Decomposition of the Racial Wage Gap in the Sample of Figure 7

					Decomposition of gap in establishment effects		Decomposition of sorting effect		Decomposition of sorting effect using person effects in later sample to classify workers by skill group	
	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	Skill-based sorting	Residual sorting
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A. Males										
All education	0.193	0.147	-0.004	0.051	0.013	0.038	0.023	0.015	0.026	0.012
		76%	*-2%*	*26%*	*7%*	*20%*	*12%*	*8%*	*13%*	*6%*
No high school	0.069	0.044	0.007	0.018	0.011	0.007	0.002	0.005	0.004	0.003
		64%	*10%*	*26%*	*16%*	*10%*	*3%*	*7%*	*6%*	*5%*
High school degree	0.148	0.108	0.000	0.040	0.015	0.026	0.015	0.011	0.017	0.008
		73%	*0%*	*27%*	*10%*	*17%*	*10%*	*7%*	*12%*	*6%*
College degree	0.194	0.142	0.004	0.047	0.024	0.023	0.006	0.017	0.010	0.013
		73%	*2%*	*24%*	*13%*	*12%*	*3%*	*9%*	*5%*	*7%*
B. Females										
All education	0.279	0.215	-0.013	0.077	0.020	0.057	0.033	0.023	0.040	0.017
		77%	*-5%*	*28%*	*7%*	*20%*	*12%*	*8%*	*14%*	*6%*
No high school	0.064	0.040	0.000	0.024	0.016	0.008	-0.001	0.010	0.003	0.005
		63%	*0%*	*38%*	*25%*	*13%*	*-2%*	*15%*	*5%*	*8%*
High school degree	0.134	0.097	-0.008	0.046	0.021	0.024	0.008	0.017	0.012	0.012
		72%	*-6%*	*34%*	*16%*	*18%*	*6%*	*12%*	*9%*	*9%*
College degree	0.229	0.139	0.008	0.082	0.036	0.046	0.008	0.038	0.011	0.035
		60%	*4%*	*36%*	*16%*	*20%*	*3%*	*16%*	*5%*	*15%*

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, but for the first half of our sample (2002-2008). We further restrict attention to workers who also appear in the dual-connected set of each gender group in the later sample (2008-2014), i.e., workers for whom we can estimate person effects in both the earlier sample and the later sample. 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) decompose these gaps into differences in mean person effects, differences in means of the covariate indexes, and differences in mean establishment effects. Columns (5) and (6) decompose the gap in column (4) into a relative wage-setting effect and a sorting effect. Columns (7) and (8) decompose the sorting effect into a skill-based sorting effect and a residual sorting effect, defining skill groups based on estimated person effects in the earlier sample. Finally, columns (9) and (10) decompose again the sorting effect into a skill-based sorting effect and a residual sorting effect, but defining skill groups based on estimated person effects in the later sample. 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 D.9—Summary of Estimated Two-Way Fixed Effects Models Underlying the Analysis in Figure 7

	White male	Non-white male	White female	Non-white female
	(1)	(2)	(3)	(4)
Largest connected set				
Standard deviation of log wages	0.676	0.588	0.710	0.583
Mean log wages	1.868	1.662	1.750	1.473
A. AKM decomposition				
Std. dev. of person effects (across person-yr obs.)	0.516	0.445	0.571	0.484
Std. dev. of estab. effects (across person-yr obs.)	0.321	0.300	0.335	0.303
Std. dev. of covariates (across person-yr obs.)	0.114	0.109	0.098	0.096
Correlation of person/estab. effects	0.219	0.117	0.183	0.043
Adjusted R-squared of model	0.920	0.898	0.937	0.922
Percentage of variance of log wages due to:				
person effect	58.2%	57.3%	64.8%	68.9%
establishment effect	22.5%	26.0%	22.2%	26.9%
covariance of person and estab. effects	15.8%	9.1%	13.9%	3.7%
estab. effects+covariance person and estab. effects	38.3%	35.1%	36.1%	30.6%
Number of establishments	557,614	267,458	424,505	137,149
Number of movers	1,511,421	625,361	876,885	232,964
Number of person-year observations	13,700,443	5,247,252	8,345,180	2,159,068
Leave-one-out connected set				
Standard deviation of log wages	0.680	0.574	0.744	0.594
Mean log wages	1.889	1.652	1.841	1.469
B. AKM decomposition				
Std. dev. of estab. effects (across person-yr obs.)	0.296	0.263	0.313	0.253
Correlation of person/estab. effects	0.341	0.245	0.362	0.250
Percentage of variance of log wages due to:				
establishment effect	18.9%	21.0%	17.7%	18.1%
covariance of person and estab. effects	22.0%	16.2%	23.4%	16.9%
estab. effects+covariance person and estab. effects	40.9%	37.3%	41.1%	35.0%
C. KSS decomposition				
Std. dev. of estab. effects (across person-yr obs.)	0.261	0.241	0.285	0.218
Correlation of person/estab. effects	0.502	0.372	0.491	0.424
Percentage of variance of log wages due to:				
establishment effect	14.8%	17.7%	14.7%	13.5%
covariance of person and estab. effects	25.5%	19.7%	26.8%	22.4%
estab. effects+covariance person and estab. effects	40.3%	37.4%	41.4%	35.9%
Number of establishments	291,144	109,640	187,208	41,101
Number of movers	1,256,294	478,541	653,608	145,760
Number of person-year observations	5,709,478	2,131,029	2,964,021	652,689

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, but for the first half of our sample (2002-2008). We further restrict attention to workers who also appear in the largest connected set of each race-gender group in the later sample (2008-2014). 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 D.10—Decomposition of the Racial Wage Gap in the Leave-one-out Sample

				Means of establishment effects			Decomposition of gap in establishment effects		Decomposition of sorting effect	
	Overall racial wage gap	Gap in person effects	Gap in covariate index	Whites	Nonwhites	Gap in establishment effects	Relative wage-setting	Sorting	Skill-based sorting	Residual sorting
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A. Males										
All education	0.175	0.132	0.004	0.206	0.167	0.038	0.005	0.033	0.022	0.011
		76%	*3%*			*22%*	*3%*	*19%*	*13%*	*6%*
No high school	0.045	0.032	0.011	0.127	0.125	0.002	0.000	0.002	0.001	0.001
		71%	*25%*			*5%*	*0%*	*5%*	*2%*	*2%*
High school degree	0.117	0.085	0.005	0.237	0.209	0.027	0.010	0.018	0.012	0.006
		72%	*4%*			*23%*	*8%*	*15%*	*10%*	*5%*
College degree	0.196	0.108	0.025	0.417	0.357	0.060	0.025	0.035	0.010	0.025
		55%	*13%*			*31%*	*13%*	*18%*	*5%*	*13%*
B. Females										
All education	0.268	0.212	-0.015	0.155	0.083	0.071	0.013	0.059	0.039	0.020
		79%	*-6%*			*27%*	*5%*	*22%*	*15%*	*7%*
No high school	0.050	0.035	0.001	0.021	0.006	0.015	0.003	0.012	0.001	0.010
		70%	*3%*			*30%*	*6%*	*23%*	*2%*	*21%*
High school degree	0.117	0.094	-0.015	0.151	0.112	0.039	0.016	0.023	0.010	0.013
		80%	*-12%*			*33%*	*14%*	*19%*	*8%*	*11%*
College degree	0.221	0.123	0.012	0.325	0.242	0.083	0.034	0.050	0.009	0.041
		56%	*5%*			*38%*	*15%*	*23%*	*4%*	*18%*

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 for the leave-one-out connected set in Panel B of Table 2 (Table 3 in the paper uses instead the largest connected set in Panel A of Table 2). 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 (6) first decompose these gaps into differences in mean person effects, differences in means of the covariate indexes, and differences in mean establishment effects. For reference, we also show the means of the estimated establishment effects for whites and nonwhites in columns (4) and (5). Column (7) and (8) then decompose the gap in column (6) into a relative wage-setting effect and a sorting effect. Columns (9) and (10) 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 D.11—Across- and Within-Industry Decompositions

	Across-sector analysis				Within-sector analysis				
	White-specific establishment effect	White-specific person effect	Observed share non-white	Predicted minus observed share non-white	Overall racial wage gap	Relative wage-setting	Sorting	Skill-based sorting	Residual sorting
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A. Males									
Banking and finance (2%)	0.652	2.568	0.183	0.085	0.204	0.008	0.031	0.017	0.014
Manufacturing (23%)	0.288	2.190	0.304	0.011	0.187	0.000	0.029	0.023	0.006
Other services and organizations (6%)	0.216	2.128	0.325	-0.005	0.191	0.014	0.024	0.012	0.012
Transportation and communications (12%)	0.199	2.083	0.321	0.002	0.103	0.012	0.016	0.009	0.007
Real estate (19%)	0.189	2.028	0.352	-0.017	0.179	0.011	0.033	0.015	0.018
Construction (9%)	0.149	1.993	0.403	-0.067	0.088	0.009	0.000	0.005	-0.005
Retail/wholesale (22%)	0.099	2.003	0.321	0.013	0.111	0.003	0.013	0.005	0.008
Accomodation and food (4%)	0.013	1.847	0.304	0.046	0.034	-0.003	-0.008	-0.002	-0.006
B. Females									
Banking and finance (4%)	0.554	2.438	0.161	0.058	0.251	0.030	0.060	0.035	0.025
Transportation and communications (4%)	0.206	2.140	0.252	0.021	0.306	0.029	0.073	0.037	0.036
Manufacturing (17%)	0.160	1.958	0.274	0.015	0.198	-0.002	0.034	0.025	0.009
Other services and organizations (20%)	0.153	2.079	0.281	-0.008	0.176	0.013	0.015	0.012	0.003
Retail/wholesale (21%)	0.089	1.869	0.304	0.001	0.117	0.006	0.012	0.006	0.006
Education (5%)	0.088	2.220	0.215	0.046	0.330	0.021	0.036	0.022	0.015
Real estate (19%)	0.080	1.963	0.338	-0.031	0.285	0.022	0.062	0.038	0.024
Accomodation and food (7%)	0.015	1.700	0.338	-0.006	0.060	-0.006	0.006	-0.001	0.007

Notes: The table displays key decomposition statistics by broad industries. Columns (1) and (2) report the mean white-specific establishment effect and the mean person effect among whites (using white person-year observations), respectively, for a given industry. Columns (3) to (4) present results analogous to the analysis in Figure 5, but where shares of nonwhites are calculated separately for each industry rather than by establishment effect deciles. That is, column (3) is the average observed share of nonwhites in the industry—similar to the black line in Figure 5—while column (4) is the difference between the predicted and observed share of nonwhites in the industry—akin to the difference between the green and black lines in Figure 5. 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). Broad industries in each panel (A for males; B for females) are sorted in descending order of their mean establishment effect in column (1), ommitting those that account for less than 2% of the person-year observations in the sample. Columns (5)-(9) are analogous to columns (1), (7), (8), (9), and (10) in Table 3 but present decomposition results *within each industry*. That is, nonwhites are reweighted so as to have the same distribution across micro-regions as whites (of the same gender) employed in the same industry and the pool of "suitable" workers underlying the computation of the skill-based sorting effect is defined by workers of the same age group, same person-effect group, employed in the same local labor market, in the same year, and in the same industry.