Online Appendix for "Systemic Discrimination Among Large U.S. Employers"

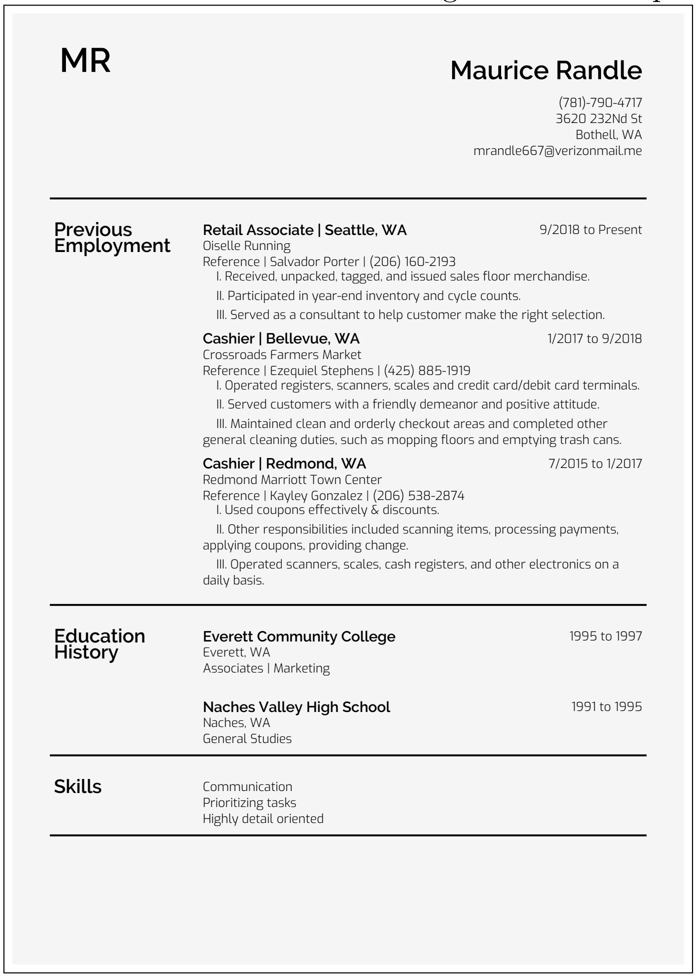
Patrick Kline

Evan K. Rose

Christopher R. Walters

Appendix A Additional Figures and Tables

Figure A1: Examples of applicant resumes



MR

Maurice Randle

(781)-790-4717
3620 232Nd St
Bothell, WA
mrandle667@verizonmail.me

Previous Employment

Retail Associate | Seattle, WA 9/2018 to Present
Oiselle Running
Reference | Salvador Porter | (206) 160-2193
I. Received, unpacked, tagged, and issued sales floor merchandise.
II. Participated in year-end inventory and cycle counts.
III. Served as a consultant to help customer make the right selection.

Cashier | Bellevue, WA 1/2017 to 9/2018
Crossroads Farmers Market
Reference | Ezequiel Stephens | (425) 885-1919
I. Operated registers, scanners, scales and credit card/debit card terminals.
II. Served customers with a friendly demeanor and positive attitude.
III. Maintained clean and orderly checkout areas and completed other general cleaning duties, such as mopping floors and emptying trash cans.

Cashier | Redmond, WA 7/2015 to 1/2017
Redmond Marriott Town Center
Reference | Kayley Gonzalez | (206) 538-2874
I. Used coupons effectively & discounts.
II. Other responsibilities included scanning items, processing payments, applying coupons, providing change.
III. Operated scanners, scales, cash registers, and other electronics on a daily basis.

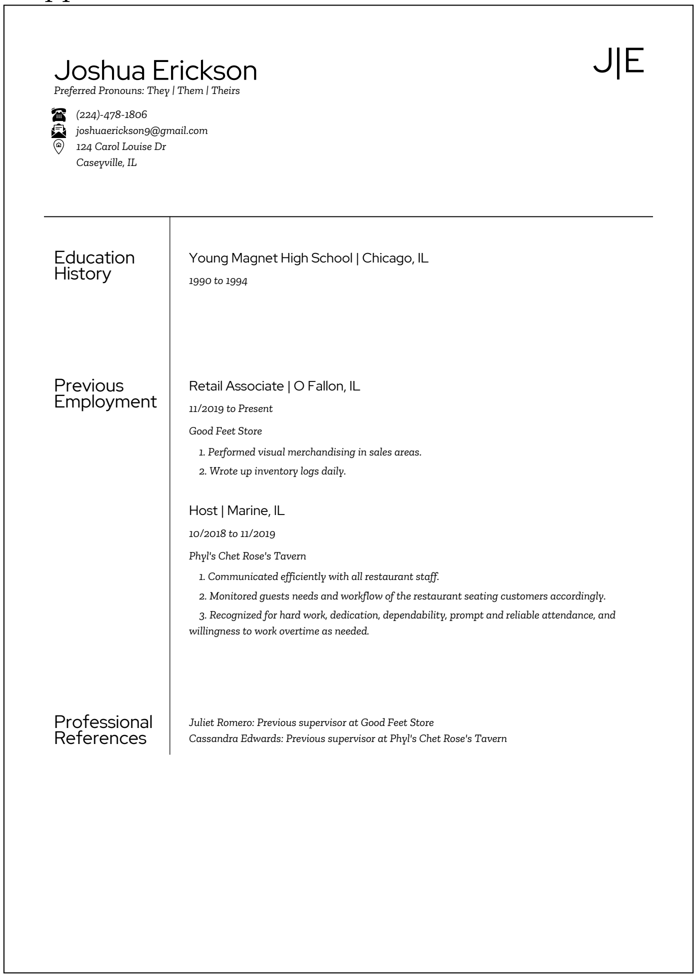
Education History

Everett Community College 1995 to 1997
Everett, WA
Associates | Marketing

Naches Valley High School 1991 to 1995
Naches, WA
General Studies

Skills

Communication
Prioritizing tasks
Highly detail oriented



Joshua Erickson

J|E

Preferred Pronouns: They | Them | Theirs

(224)-478-1806
joshuaerickson9@gmail.com
124 Carol Louise Dr
Caseyville, IL

Education History

Young Magnet High School | Chicago, IL
1990 to 1994

Previous Employment

Retail Associate | O Fallon, IL
11/2019 to Present
Good Feet Store
1. *Performed visual merchandising in sales areas.*
2. *Wrote up inventory logs daily.*

Host | Marine, IL
10/2018 to 11/2019
Phyl's Chet Rose's Tavern
1. *Communicated efficiently with all restaurant staff.*
2. *Monitored guests needs and workflow of the restaurant seating customers accordingly.*
3. *Recognized for hard work, dedication, dependability, prompt and reliable attendance, and willingness to work overtime as needed.*

Professional References

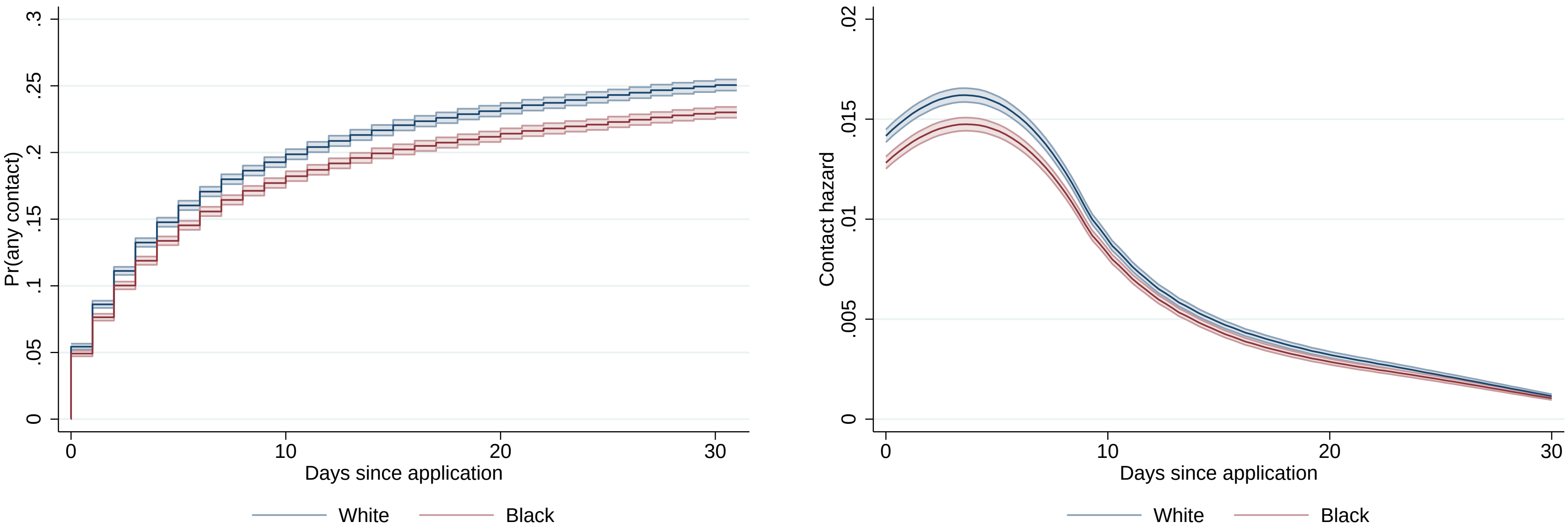
Juliet Romero: Previous supervisor at Good Feet Store
Cassandra Edwards: Previous supervisor at Phyl's Chet Rose's Tavern

Notes: This figure presents two examples of randomly generated resumes used in the experiment. Resumes are formatted using a combination of pre-set options specifying length, fonts, text sizes, section header names, and layouts, with controls to ensure resumes that overflow one page are not generated. The resume on the right features gender-neutral pronouns displayed below the name.

Figure A2: Kaplan-Meier estimates of contact probability and smoothed hazard

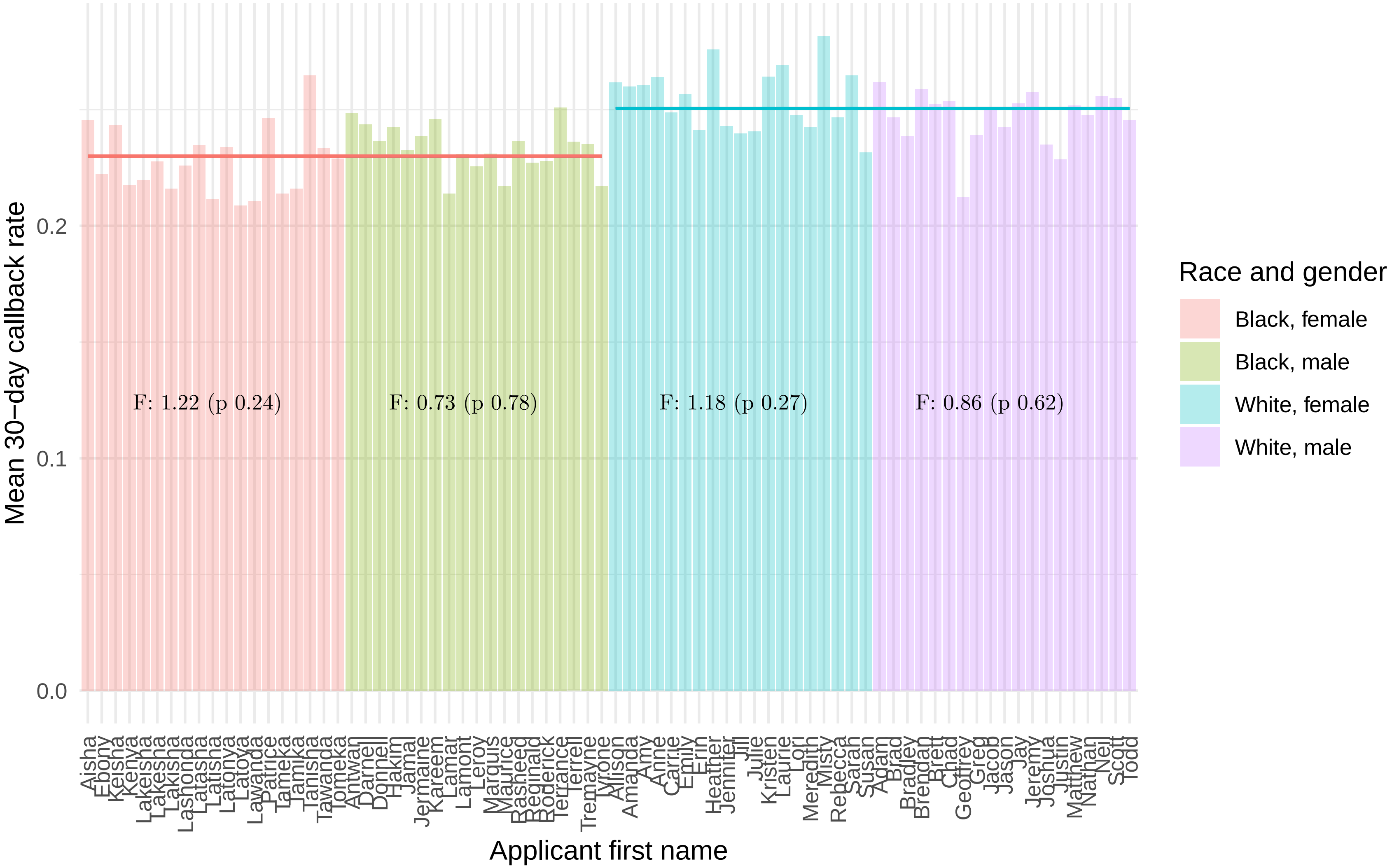
a) Contact probability

b) Smoothed contact hazard



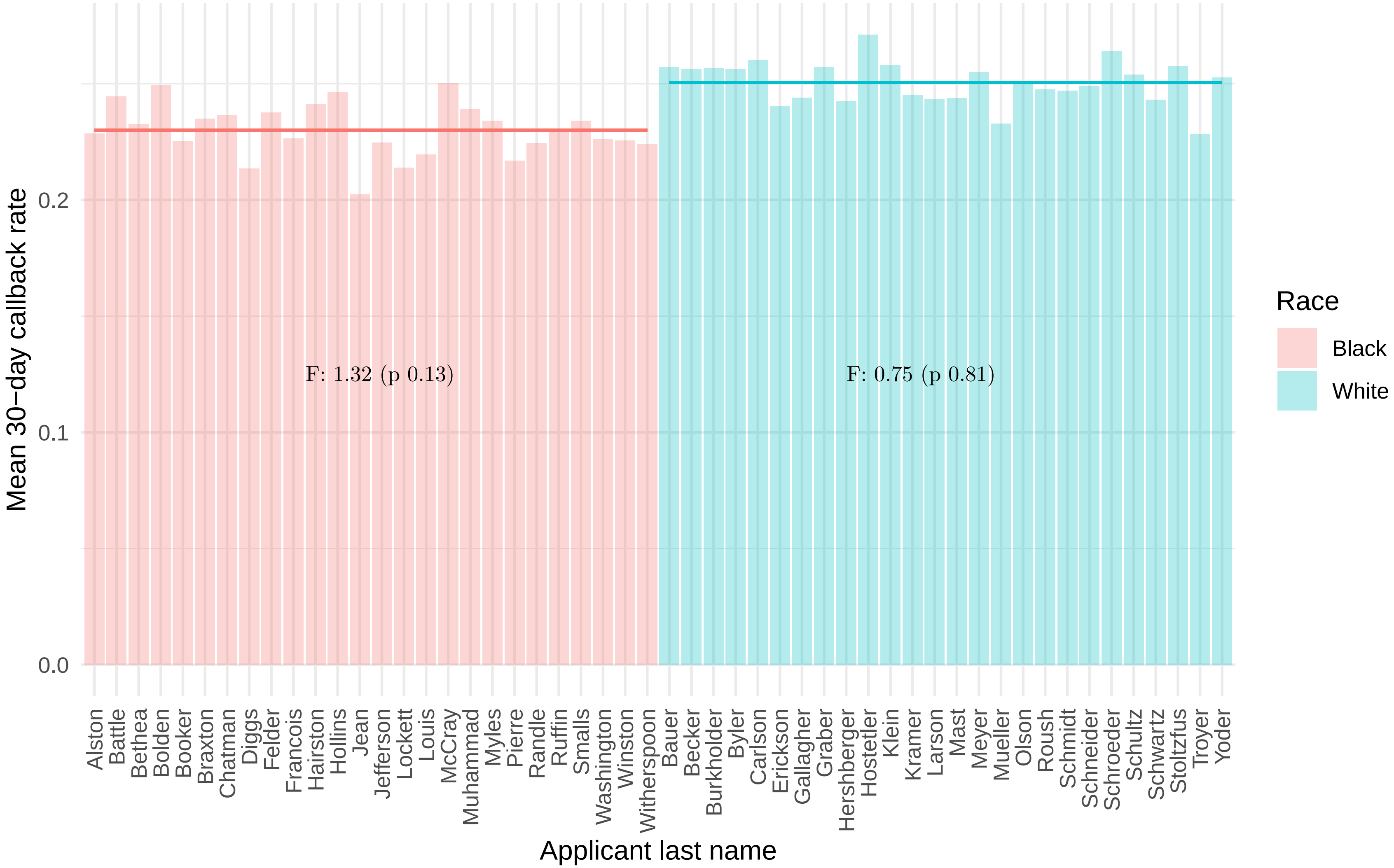
Notes: This figure plots contact probabilities and hazards as functions of days since application. Contact probabilities correspond to Kaplan-Meier failure function estimates. Hazards are Kaplan-Meier hazard estimate smoothed using the Epanechnikov kernel. Shaded areas represent pointwise 95% confidence bands.

Figure A3: Callbacks by applicant first name



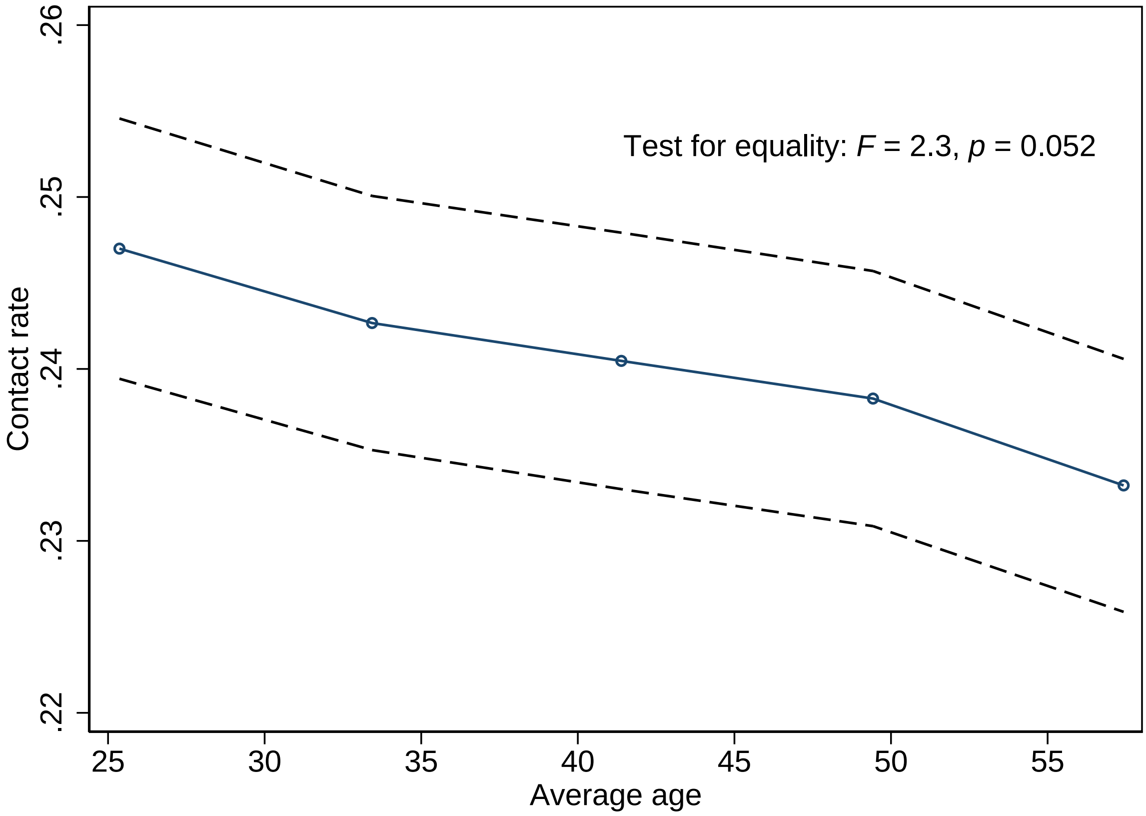
Notes: This figure shows mean contact rates by applicant first name, organized by race and gender group. The horizontal bars show race group mean contact rates. F-tests and p-values come from joint tests of the hypothesis that contact rates are equal across names separately by race and gender group.

Figure A4: Callbacks by applicant last name



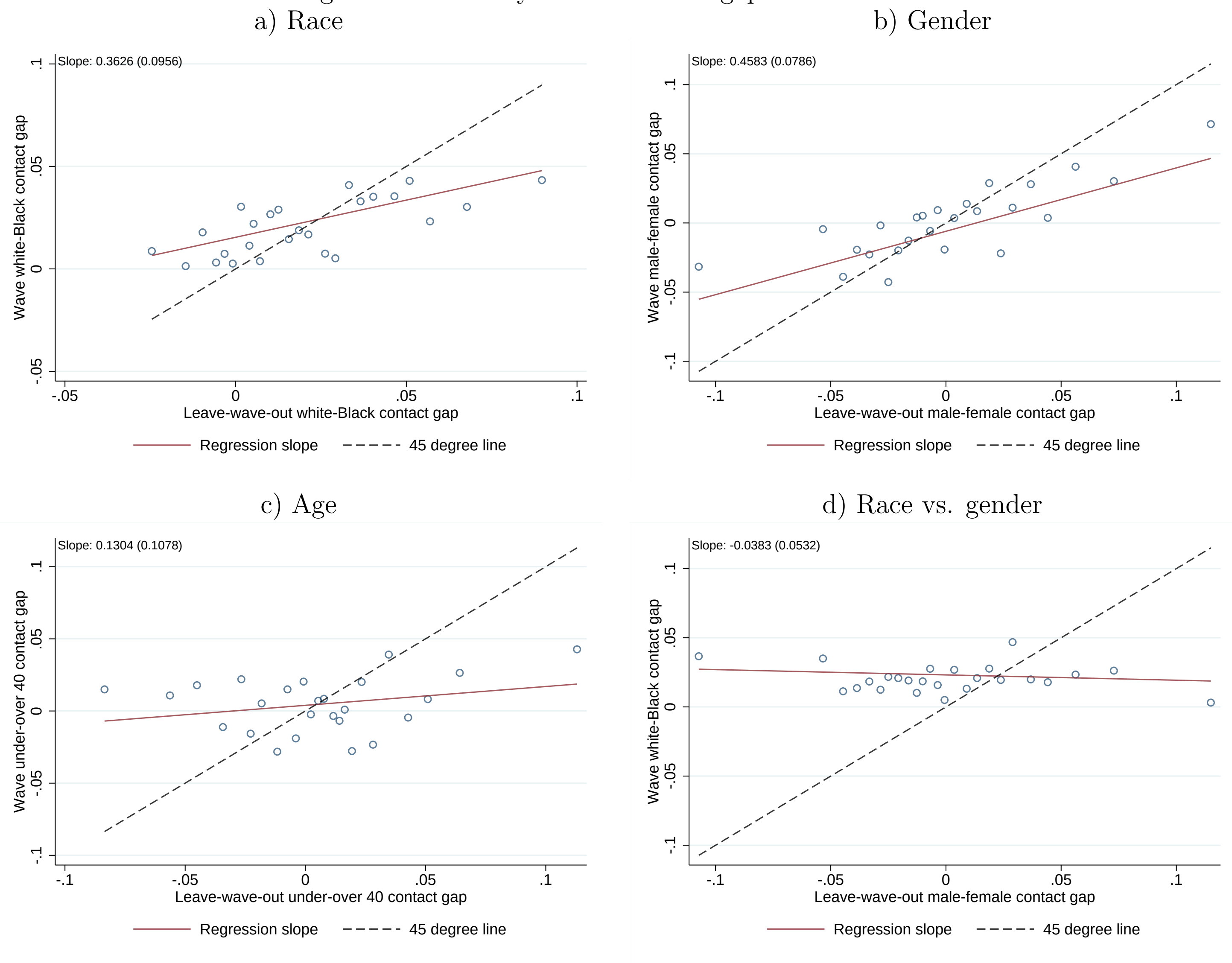
Notes: This figure shows mean contact rates by applicant last name, organized by race. The horizontal bars show race group mean contact rates. F-tests and p-values come from joint tests of the hypothesis that contact rates are equal across names separately by race.

Figure A5: Contact rates by age category



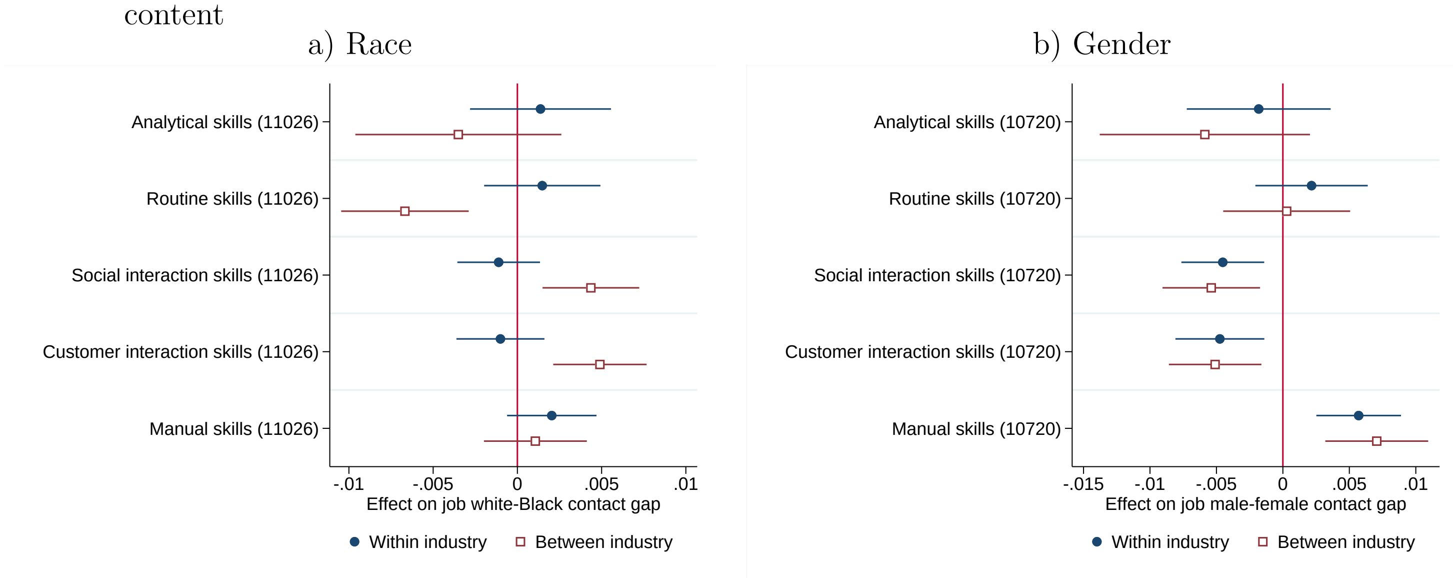
Notes: This figure plots average 30-day contact rates by quintile of applicant age at the time of application. Estimates come from regressions of a contact indicator on indicators for age quintile, controlling for wave indicators. The horizontal axis plots average age in each quintile. The vertical axis plots the mean contact rate, calculated as the sum of the quintile coefficient and mean wave effect. Dashed lines indicate 90% confidence intervals. F-statistic and p-value come from a Wald test that contact rates are equal across quintiles, clustering standard errors by job.

Figure A6: Stability of firm contact gaps across waves



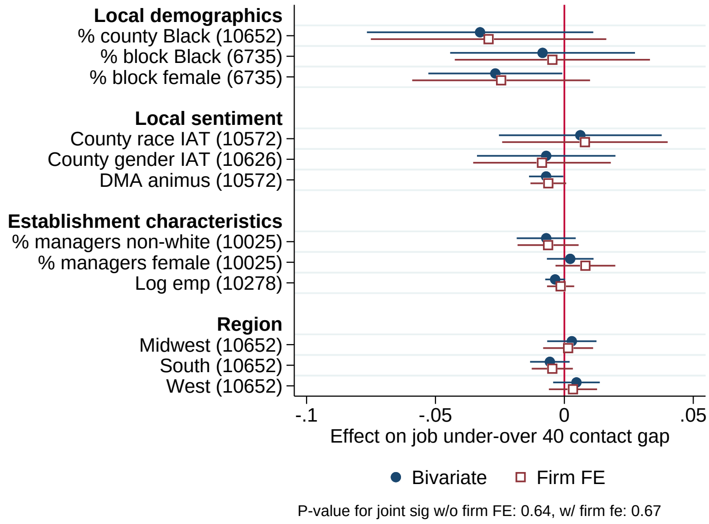
Notes: This figure presents binned scatter plots of firm-specific wave-average contact gaps vs. leave-wave-out firm-specific average contact gaps. Panel (a) reports results for the white/Black difference in contact rates. Panel (b) shows results for the male/female difference in contact rates. Panel (c) displays results for the difference between contact rates for applicants under and over age 40. Panel (d) plots the correlation between race and gender contact gaps. The points are means of the dependent and independent variables within vingtiles of the independent variable. The dotted line has a slope of 1 and passes through the origin. The red line corresponds to the regression slope reported on the figure, with firm-clustered standard errors reported in parentheses. All firms present in at least 2 waves are included.

Figure A7: Within and between industry relationship between contact gaps and task content



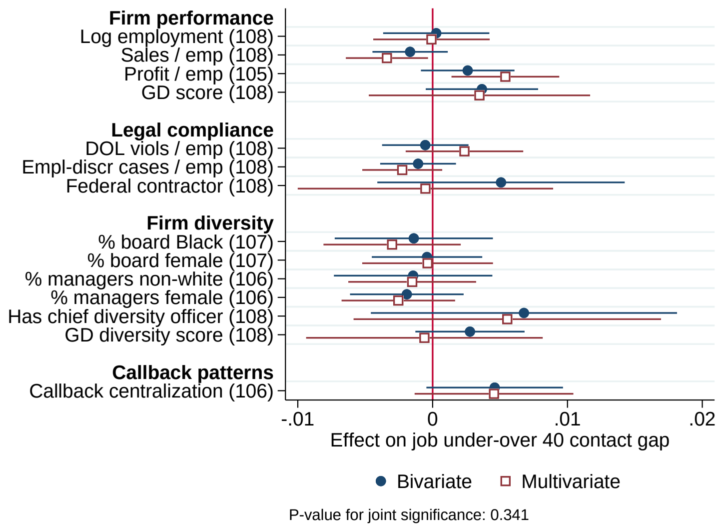
Notes: This figure plots the relationship between O*Net measures of job-level task content and contact gaps for race and gender within and between industry. The within relationship is estimated with a linear regression with job-level contact gaps as the outcome and two-digit industry fixed effects. The between relationship is estimated by instrumenting job task content with industry dummies. All jobs with defined contact gaps for each attribute are included. The number of jobs in each regression is in parentheses. Task measures are normalized to have standard deviation one in sample. Bars indicate 95% confidence intervals based on robust standard errors. Appendix C provides a complete description of task definitions and sources.

Figure A8: Relationships between age contact gaps and establishment characteristics



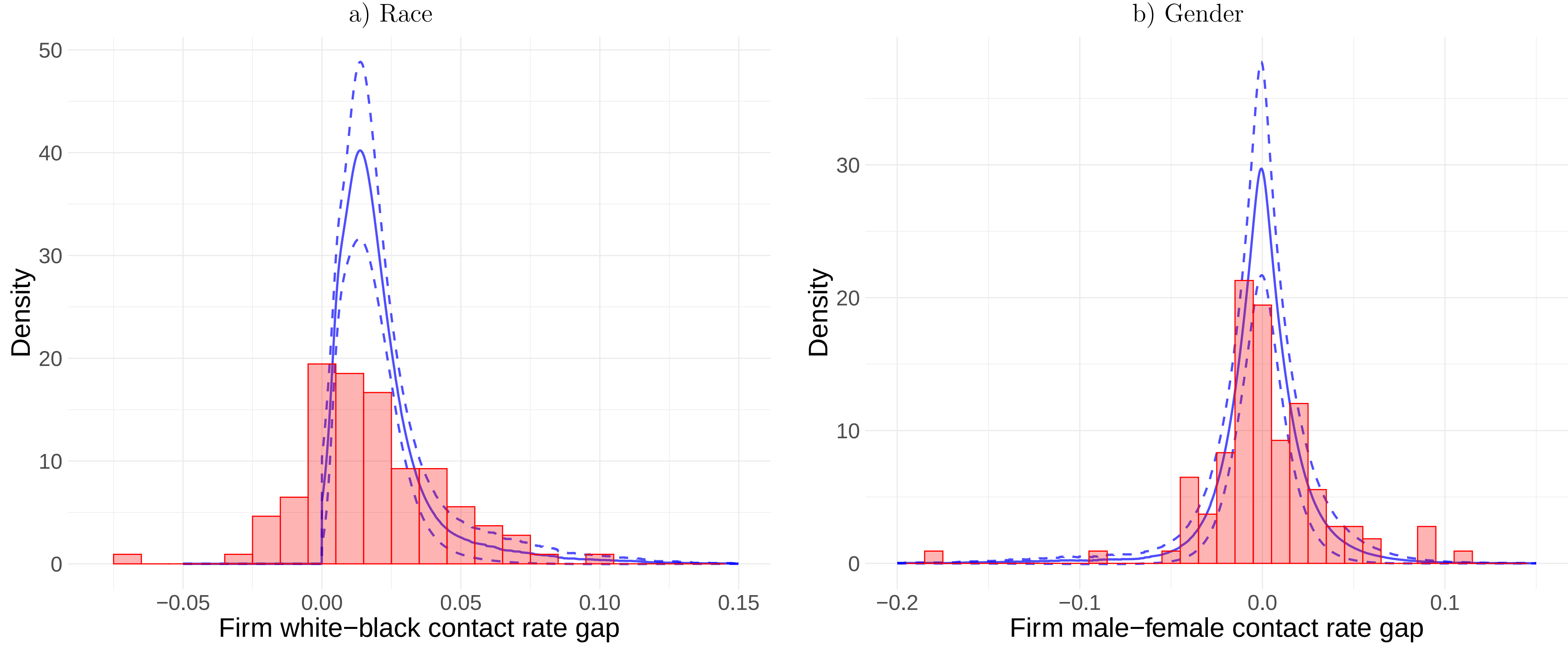
Notes: This figure plots the relationship between establishment-level covariates and contact gaps for applicant age under vs. over 40. Each relationship is estimated with a linear regression with job-level contact gaps as the outcome. All jobs with defined contact gaps for age and matched to the listed covariate are included. "Bivariate" points plot coefficients from a regression of contact gaps on the covariate alone. "Firm FE" points include firm fixed effects. Bars indicate 95% confidence intervals based on robust standard errors. Appendix C provides a complete description of covariate definitions and sources.

Figure A9: Relationships between age contact gaps and firm characteristics



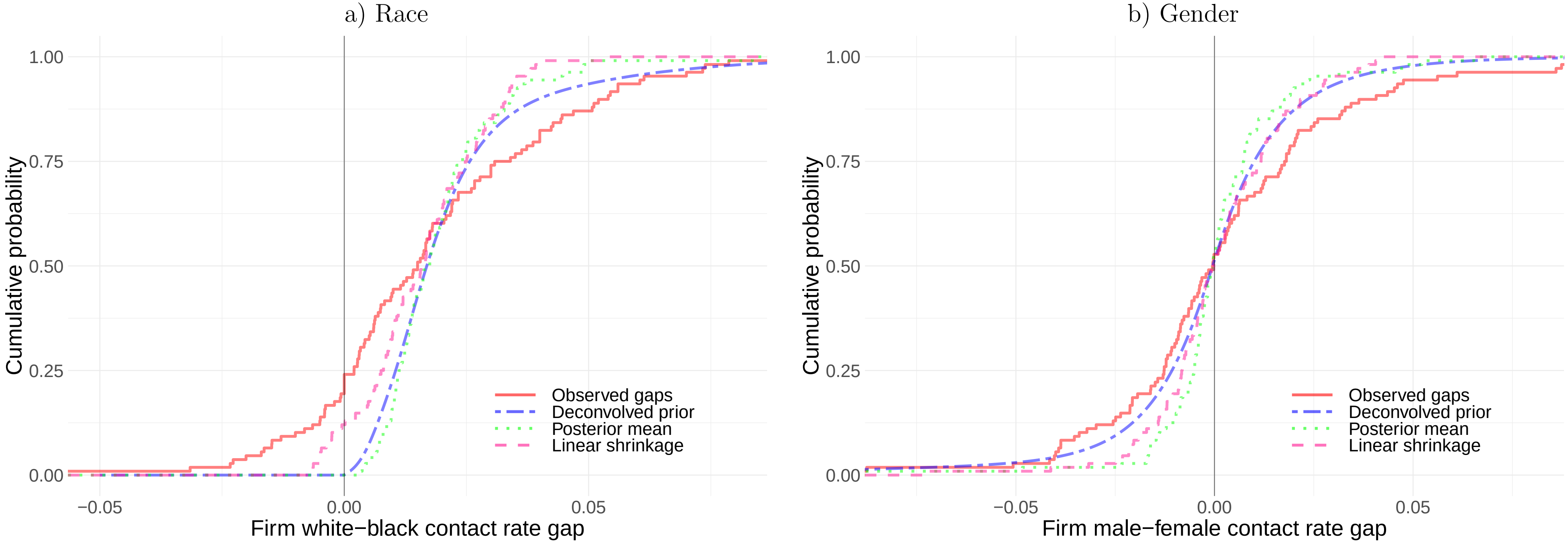
Notes: This figure plots relationships between firm-level covariates and contact gaps for application age under vs. over 40. Each relationship is estimated with a linear regression with job-level contact gaps as the outcome. All covariates are standardized to be mean zero, standard deviation 1 in sample."Bivariate" points plot coefficients from a regression of contact gaps on the covariate alone. "Multivariate" points plot effects when all covariates are entered simultaneously. Bars indicate 95% confidence intervals based on standard errors clustered at the firm level. Appendix C provides a complete description of covariate definitions and sources.

Figure A10: Confidence intervals on deconvolutions of firm-level discrimination



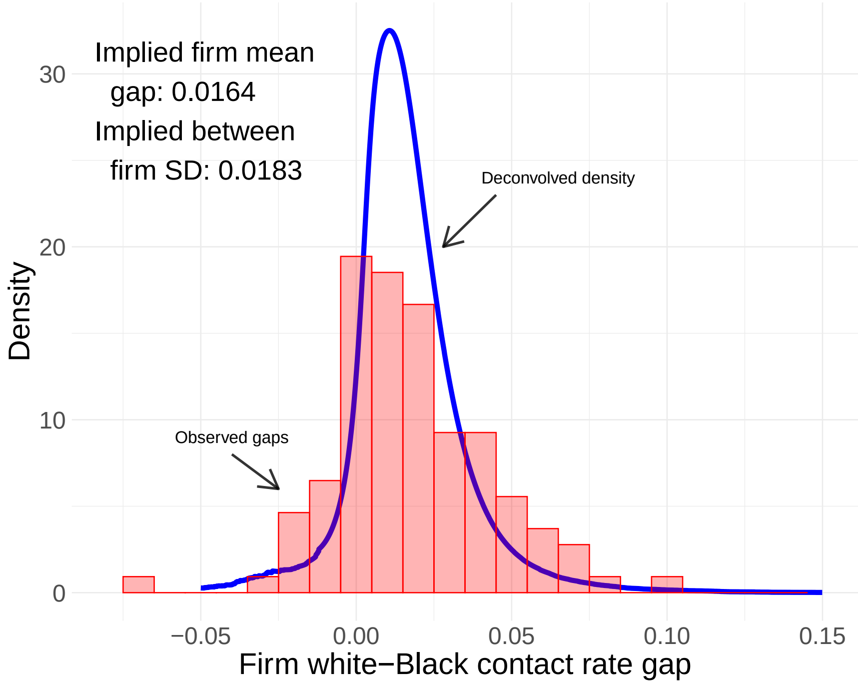
Notes: This figure presents non-parametric estimates of the distribution of firm-specific contact gaps and point-wise 95% confidence intervals. Panel (a) presents estimates for white-Black contact rate differences, and panel (b) presents estimates for male-female differences. Red histograms show the distribution of estimated firm contact gaps. Blue lines shows estimates of population contact gap distributions. The population distributions are estimated by applying the deconvolveR package (Narasimhan and Efron, 2020) to firm-specific z-score estimates, then numerically integrating over the empirical distribution of standard errors to recover the distribution of contact gaps. Since the estimated density of Δ is a linear combination of points in the density of $\hat{g}_\mu$, standard errors can be computed using the delta method and the variance-covariance of $\hat{g}_\mu$ produced by Narasimhan and Efron (2020). The penalization parameter in the deconvolution step is calibrated so that the resulting distribution matches the corresponding bias-corrected variance estimate from Table 4. In panel (a), the density of population z-scores is constrained to be weakly positive.

Figure A11: Distribution of observed, deconvolved, and posterior estimates



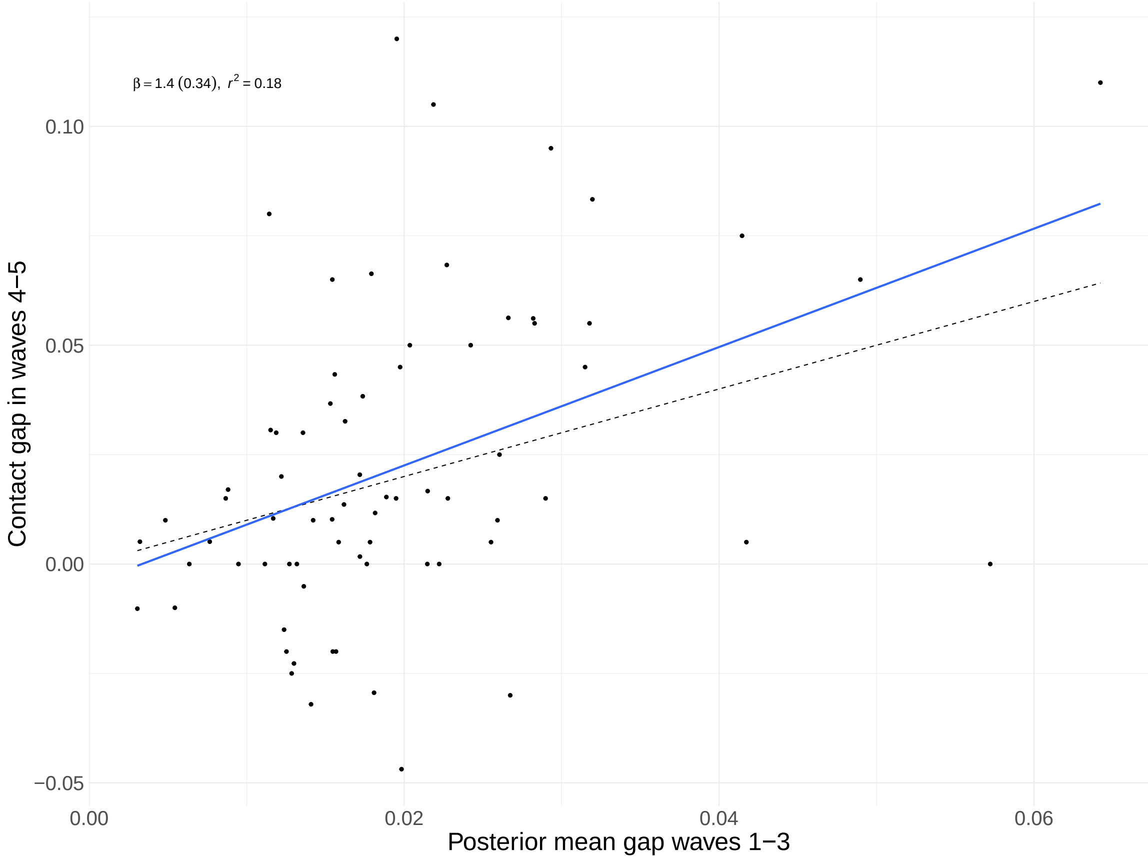
Notes: This figure presents estimates of the distribution of firm-specific contact gaps for race and gender. The red solid line presents the empirical CDF of estimated gaps. The blue dashed line plots the CDF of population contact gaps based on the deconvolution estimates in Figure 6. The green dotted line plots the empirical CDF of posterior means, constructed treating the deconvolved density as prior knowledge. The pink dashed line shows the empirical CDF of estimates shrunk linearly towards the grand mean with weights given by the signal-to-noise ratio $\hat{\theta}/(s_f^2 + \hat{\theta})$, where $\hat{\theta}$ is the bias-corrected estimate of the variance of contact gaps across firms.

Figure A12: Deconvolution of firm-level racial discrimination without support restriction



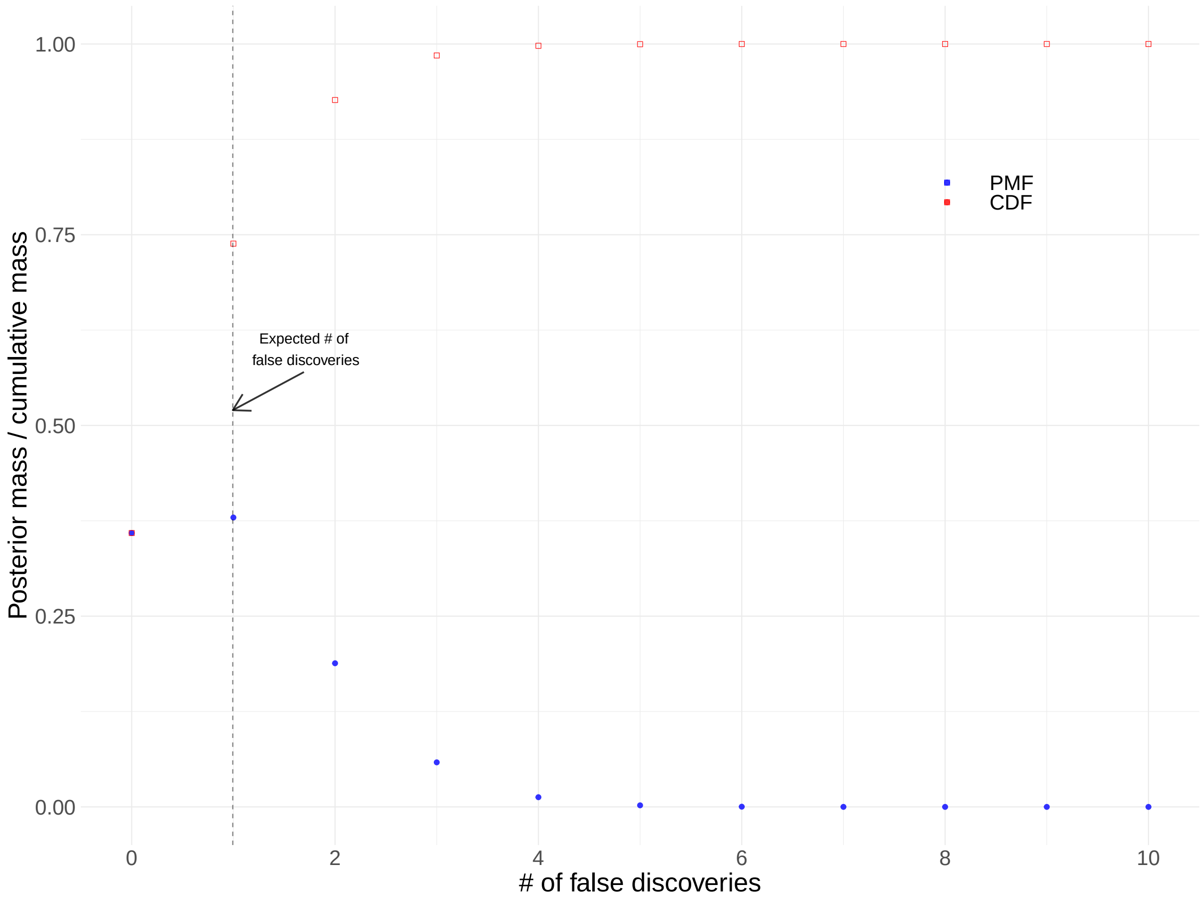
Notes: This figure presents non-parametric estimates of the distribution of firm-specific white-Black contact rate differences. The red histogram shows the distribution of estimated firm contact gaps. Blue line shows estimates of the population contact gap distributions. The population distributions are estimated by applying the deconvolveR package (Narasimhan and Efron, 2020) to firm-specific z-score estimates, then numerically integrating over the empirical distribution of standard errors to recover the distribution of contact gaps. The penalization parameter in the deconvolution step is calibrated so that the resulting distribution matches the corresponding bias-corrected variance estimate from Table 4.

Figure A13: Out-of-sample predictive power of racial contact gap posteriors



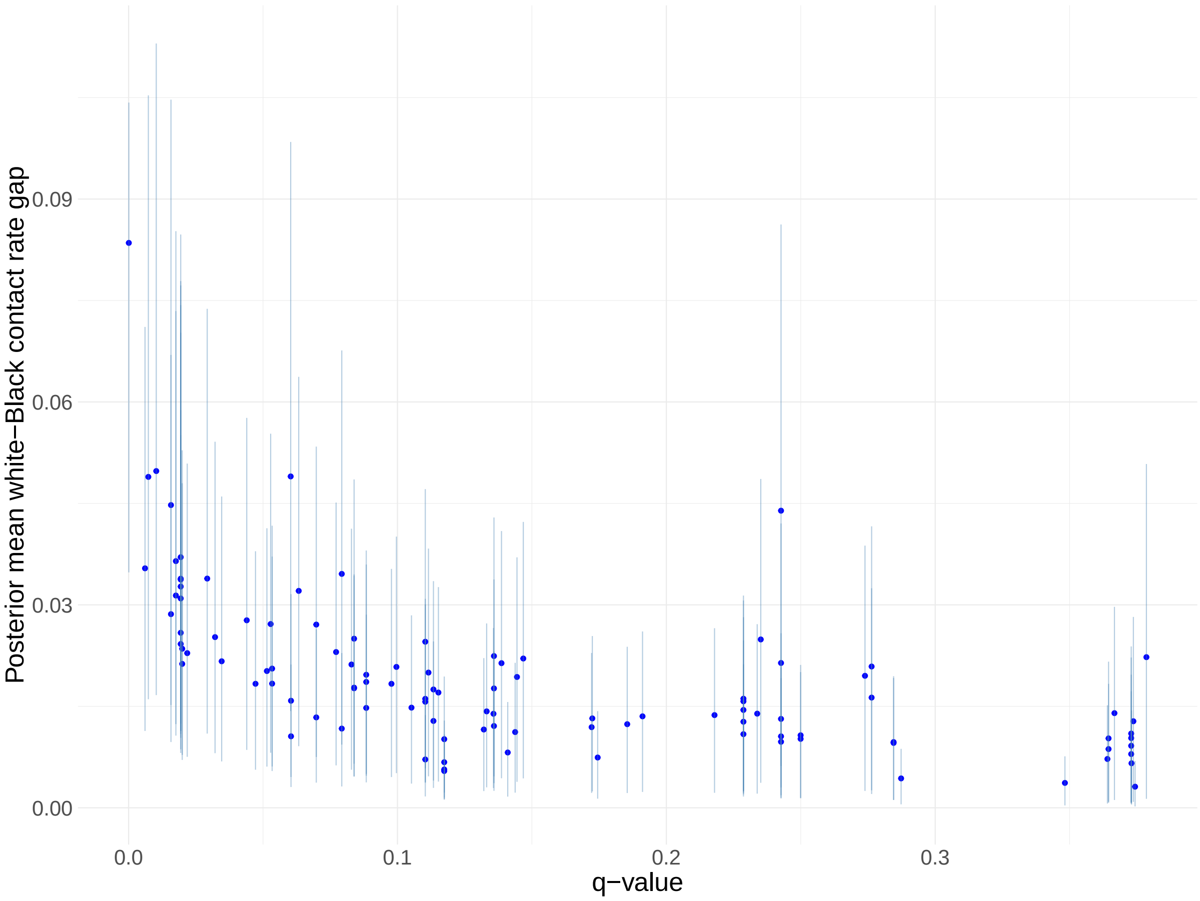
Notes: This figure plots posterior mean white-Black contact gaps computed using data from waves 1-3 against observed gaps in waves 4-5 for the sample of firms included in all five waves. Posterior means are computed using the population contact gap distributions estimated by applying the deconvolveR package (Narasimhan and Efron, 2020) to firm-specific z-score estimates from waves 1-3. The penalization parameter in the deconvolution step is calibrated so that the resulting distribution matches the corresponding bias-corrected variance estimate. The black dotted line is the 45 degree line. The blue line is the least squares fit. Adjusting for the noise in the wave 4 and 5 estimates yields a bias corrected R^2 of 0.5, or a correlation between predictions in later waves and the latent true contact gaps of $\sqrt{0.5} = 0.71$.

Figure A14: Posterior false discovery distribution among 23 firms with low q-values



Notes: This figure plots EB posterior estimates of the probability mass function and cumulative distribution of false discoveries among the 23 firms with q-values below 0.05 for race. Posterior was calculated using the Poisson binomial distribution implied by the 23 firms' LFDR estimates plotted in Figure 10. The dotted line denotes the expected number of false discoveries among these firms, which coincides with the mean of their LFDRs.

Figure A15: Posterior mean contact gaps vs. q-values



Notes: This figure plots posterior mean white/Black contact gaps $\bar{\Delta}_f$ for each firm against estimated q-values for racial discrimination. Vertical lines depict 95% empirical Bayes credible intervals (EBCIs).

Table A1: Balanced sample: Firm-level heterogeneity in discrimination

	(1) χ^2 test of heterogeneity	(2) p-value for no discrim against:	Contact gap SD (3) Bias-corrected	(4) Cross-wave	(5) Cross-state
Race	229.5	W: 1.00	0.0184	0.0171	0.0182
	[0.000]	B: 0.00	(0.0029)	(0.0032)	(0.0031)
Gender	124.2	M: 0.06	0.0207	0.0213	0.0200
	[0.000]	F: 0.03	(0.0044)	(0.0043)	(0.0045)
Over 40	90.2	Y: 0.15	0.0098	0.0096	0.0099
	[0.072]	O: 0.02	(0.0060)	(0.0067)	(0.0057)

Notes: This table presents estimated standard deviations of firm-level contact rate gaps and tests for heterogeneity in gaps using the balanced sample of firms present in all five waves. Column 1 displays χ^2 values and associated p-values from tests of the null hypothesis of no heterogeneity in discrimination. The test statistic is $\sum_f (\hat{\Delta}_f - \bar{\Delta})^2/s_f^2$, where $\hat{\Delta}_f$ is the contact cap estimate for firm f, s_f is the estimate's standard error, and $\bar{\Delta}$ is the equally-weighted average of contact gaps. Column 2 presents one-sided tests of no discrimination against white (W), black (B), male (M), female (F), aged under 40 (Y), and over 40 (O) applications using the methodology in Bai et al. (2021). Column 3 reports bias-corrected estimates of standard deviations of average contact gaps across firms based on covariances between job-level contact gaps. Columns 4 and 5 report cross-wave and cross-state estimates based on covariances between firm-by-wave and firm-by-state contact gaps. Details on these estimators appear in the Appendix. Standard errors for all variance estimators are produced by job-clustered weighted bootstrap.

Table A2: Variance components for other resume attributes

	(1) χ^2 test of heterogeneity	(2) p-value for no discrim against:	Contact gap SD (3) Bias-corrected	(4) Cross-wave	(5) Cross-state
LGBTQ Club Member	88.0	W: 1.00	-	-	-
	[0.885]	B: 0.98			
Gender Neutral Pronouns	126.5	Y: 0.92	0.0198	0.0177	0.0147
	[0.076]	O: 0.65	(0.0156)	(0.0176)	(0.0208)

Notes: This table presents estimated standard deviations of firm-level contact rate gaps by LGBTQ club member status and the presence of gender-neutral pronouns, along with tests for heterogeneity in these gaps. Column 1 displays χ^2 values and associated p-values from tests of the null hypothesis of no heterogeneity in discrimination. The test statistic is $\sum_f (\hat{\Delta}_f - \bar{\Delta})^2/s_f^2$, where $\hat{\Delta}_f$ is the contact cap estimate for firm f, s_f is the estimate's standard error, and $\bar{\Delta}$ is the equally-weighted average of contact gaps. Column 2 presents one-sided tests of no discrimination against applicants with the relevant attribute (Y) and those without the attribute (N) using the methodology in Bai et al. (2021). Column 3 reports bias-corrected estimates of standard deviations of average contact gaps across firms based on covariances between job-level contact gaps. Columns 4 and 5 report cross-wave and cross-state estimates based on covariances between firm-by-wave and firm-by-state contact gaps. Details on these estimators appear in the Appendix. Standard errors for all variance estimators are produced by job-clustered weighted bootstrap. Estimates include all 108 firms.

Table A3: Relationship between z-scores and standard errors

	Race		Gender	
	(1) Full sample	(2) Split sample	(3) Full sample	(4) Split sample
Z-score	33.98	18.06	11.50	4.52
	(24.07)	(11.35)	(14.12)	(6.74)
Squared residual	86.20	17.94	83.17	28.78
	(48.44)	(17.58)	(53.30)	(16.94)

Notes: This table assesses dependence between firm-specific z-score estimates and standard errors. Coefficients in the first row come from regressions of z-scores on standard errors, and coefficients in the second row come from regressions of the squared residuals from the first row on standard errors. Columns 1 and 2 display results for race, and columns 3 and 4 show results for gender. Columns 1 and 3 use z-scores and standard errors computed in the full sample of jobs. Columns 2 and 4 randomly split the jobs at each firm into two equally-sized samples and regress z-scores computed in one sample on standard errors computed in the other sample, stacking the two samples and clustering standard errors by firm identifier.

Table A4: Job-level discrimination prevalence bounds

	(1) Race	(2) Gender	(3) Over 40
Mean gap	0.020	0.002	0.004
	(0.002)	(0.002)	(0.002)
Total job-level variance	0.070	0.090	0.026
	(0.000)	(0.000)	(0.000)
Prevalence bound	0.073	0.000	0.014
	(0.012)	(0.001)	(0.021)

Notes: This table reports a bound on the job-level prevalence of discrimination assuming that a fixed fraction of jobs discriminate and the remaining jobs exhibit contact gaps of zero. The mean gap reported is the job-weighted average contact gap. The total job level variance is computed as the covariance of contact gaps among the first four and last four applications at every job. The prevalence bound is estimated as $(\hat{\Delta}^2 - s^2)/(\hat{\sigma}^2 + \hat{\Delta}^2 - s^2)$, where $\hat{\Delta}^2$ is the square of the estimated mean gap, s is the mean gap's estimated standard error, and $\hat{\sigma}^2$ is the estimated between-job variance. Bootstrap standard errors are reported in parentheses.

Appendix B Details of Experimental Design

Resume characteristics

Names: We draw racially distinctive first names from two sources. First, we use the same set of names in Bertrand and Mullainathan (2004), which are in turn drawn from Massachusetts birth records covering 1974 to 1979. Second, we supplement with names drawn from administrative records on speeding infractions and arrests provided by the North Carolina Administrative Office of the Courts and covering 2006 to 2018. We pick the most common names among drivers born between 1974 and 1979 with race- and gender-specific shares of at least 90%. The top names using this criterion substantially overlaps with Bertrand and Mullainathan (2004)'s list, with 6/9, 4/9, 4/9, and 3/9 names included in both sources for Black women, Black men, white women, and white men, respectively. We add 10 new names from the N.C. records for each race and gender group, leaving 19 total first names per group.

Table B1: First names assigned by race and gender

	Black male		White male		Black female		White female	
	Name	Source	Name	Source	Name	Source	Name	Source
1	Antwan	NC	Adam	NC	Aisha	Both	Allison	BM
2	Darnell	BM	Brad	Both	Ebony	Both	Amanda	NC
3	Donnell	NC	Bradley	NC	Keisha	BM	Amy	NC
4	Hakim	BM	Brendan	Both	Kenya	BM	Anne	BM
5	Jamal	Both	Brett	BM	Lakeisha	NC	Carrie	BM
6	Jermaine	Both	Chad	NC	Lakesha	NC	Emily	Both
7	Kareem	Both	Geoffrey	BM	Lakisha	Both	Erin	NC
8	Lamar	NC	Greg	BM	Lashonda	NC	Heather	NC
9	Lamont	NC	Jacob	NC	Latasha	NC	Jennifer	NC
10	Leroy	BM	Jason	NC	Latisha	NC	Jill	Both
11	Marquis	NC	Jay	BM	Latonya	Both	Julie	NC
12	Maurice	NC	Jeremy	NC	Latoya	Both	Kristen	Both
13	Rasheed	BM	Joshua	NC	Lawanda	NC	Laurie	BM
14	Reginald	NC	Justin	NC	Patrice	NC	Lori	NC
15	Roderick	NC	Matthew	Both	Tameka	NC	Meredith	BM
16	Terrance	NC	Nathan	NC	Tamika	Both	Misty	NC
17	Terrell	NC	Neil	BM	Tanisha	BM	Rebecca	NC
18	Tremayne	BM	Scott	NC	Tawanda	NC	Sarah	Both
19	Tyrone	Both	Todd	BM	Tomeka	NC	Susan	NC

Notes: This table lists the first names assigned by race and gender and their sources. "BM" indicates that the name appeared in original set of nine names used for each group in Bertrand and Mullainathan (2004). "NC" indicates the name was drawn from data on North Carolina speeding infractions and arrests. "Both" indicates the name appeared in both sources. Names from N.C. speeding tickets were selected from the most common names where at least 90% of individuals are reported to belong to the relevant race and gender group.

Last names are drawn from 2010 Decennial Census data. We use the names with highest race-specific shares that occur at least 10,000 times, picking 26 total for each race group. Each resume is assigned a first and last name from the appropriate race and gender group, sampling without replacement within firm. Each pair of applicants was assigned a white and Black first and last name, with the gender of the first name chosen randomly.

Table B2: Last names assigned by race

	Black			White		
	Name	Frequency	Race share	Name	Frequency	Race share
1	Alston	30,693	79.8	Bauer	65,004	95.1
2	Battle	26,432	77.3	Becker	87,859	94.89
3	Bethea	12,061	74.8	Burkholder	11,532	97.55
4	Bolden	21,819	72.3	Byler	13,230	98.19
5	Booker	36,840	65.2	Carlson	120,552	94.83
6	Braxton	12,268	72.4	Erickson	82,085	95.05
7	Chatman	15,473	79.2	Gallagher	69,834	94.62
8	Diggs	14,467	68.1	Graber	12,204	97.16
9	Felder	13,257	66.9	Hershberger	14,357	98.08
10	Francois	14,593	78	Hostetler	14,505	97.46
11	Hairston	16,090	80.9	Klein	81,471	95.41
12	Hollins	10,213	73.8	Kramer	63,936	95.35
13	Jean	21,140	70.3	Larson	122,587	94.79
14	Jefferson	55,179	74.2	Mast	15,932	96.99
15	Lockett	14,140	71.4	Meyer	150,895	94.84
16	Louis	23,738	65.5	Mueller	64,191	95.66
17	McCray	28,024	67.4	Olson	164,035	94.76
18	Muhammad	19,076	82.9	Roush	11,386	96.44
19	Myles	13,898	72.1	Schmidt	147,034	95.15
20	Pierre	33,913	86.7	Schneider	101,290	95.35
21	Randle	14,437	68.8	Schroeder	67,977	95.36
22	Ruffin	16,324	80.4	Schultz	104,888	94.81
23	Smalls	12,435	90.5	Schwartz	90,071	95.93
24	Washington	177,386	87.5	Stoltzfus	15,786	99
25	Winston	21,667	62.7	Troyer	16,981	97.96
26	Witherspoon	13,171	62.1	Yoder	56,410	97.77

Notes: This table reports the last names used in the experiment. Names are drawn from Decennial Census data. We pick names with the highest race-specific shares among those that occur more than 10,000 times. The table reports each name's frequency and the share of individuals with that surname who belong to each race group.

Dates of birth: Applicants were initially randomly assigned a date of birth between 1960 and 2000. Because these dates were fixed, as the experiment continued the average age of applicants increased. In wave 5 we began to assign dates of birth implying a uniform distribution of applicant ages between 20 and 60 at the time of application creation.

Social security numbers: Some applications required us to provide a social security number. We assigned all applicants a social security number from a publicly available database of numbers belonging to the deceased.

Emails: We manually created Gmail, Outlook, and Yahoo email accounts for roughly half of our applicants. To facilitate account creation and avoid account limits on these service, we also registered new domains designed to imitate common internet service providers' names: icloudlive.me, spectrumemail.org, fiosmail.net, and xfinity19.com. Each domain redirected to the relevant provider (e.g., icloudlive.me redirected to the icloud home page). Email addresses were creating using combinations of assigned first and last names and random integers. Each email was associated with a single first and last name combination. All emails were set up to automatically forward to a single inbox that was monitored for contacts.

Phone numbers: We provisioned phone numbers from Twilio. During each wave of the experiment, we rented roughly 200 numbers with SMS capabilities from area codes across the country. Each number was assigned to a single first and last name combination, ensuring that the same number was used only once at each company. We rented new numbers each wave so that each unique number was used at each firm at most once.

Phone calls to each number were automatically directed to a voicemail with a standard, non-personalized message. All calls were logged. Any voicemails were recorded and transcribed. We then used a combination of manual and automatic methods to tag voicemails as callbacks from particular employers using text searches on transcribed voicemails and by listening to voicemails. Text message callbacks were processed in the same way.

Addresses: We assigned each application a home address close to the job to which the application was submitted. Addresses were sourced from openaddresses.io and the U.S. Department of Transportation's National Address Database. We download the full set of addresses from both sources and manually eliminated unusual and non-residential addresses. Addresses were randomly assigned to applications without replacement for each job from the set of addresses in zip codes within 20 miles of the target job. If insufficient addresses were available with a 20 mile radius, a 40 mile radius was used instead.

Educational history: All applicants were assigned a high school in same state as the target job. We use the National Center for Educational Statistics to identify all non-specialized public schools with instruction in grades 9-12 and randomly select a school from zip codes with an absolute difference of less than 1,000 from the target job's zip code. If insufficient schools are available, we randomly assign a school from anywhere in the state. All applicants graduated from high school the same year they turned 18 years old.

We attempted to randomly assign half of our applicants an associate degree from a community college in the same state as the target job. We use the Department of

Education's College Scorecard data to identify all relevant degree-granting institutions, manually eliminating some specialty schools. Colleges were assigned in the same manner as high schools. Each applicant with a degree was also assigned a major from a list of common, non-specialized degrees, including Business Technology, Marketing, Information Technology, Communication Studies, and Sales Management. All applicants received their degree two years after finishing high school. Because appropriate colleges were not available in all geographies, slightly less than half of applicants were assigned a degree.

Club membership: Beginning in Wave 2, 20% of applicants were assigned a club to be listed on their resume as part of their educational experience. Half of applicants assigned a club listed clubs intended to signal LGBTQ affiliation: the Gay-Straight Alliance, the Lesbian, Gay, Bisexual, Transgender, and Queer Association, and the Queer-Straight Alliance. The remaining half were assigned either a generic club (History Club, Speech and Debate Club, Foreign Language Club, Outdoors Club, Model United Nations, Performing Arts Club, Student Government, or Music Club) or political club (Young Republican Club, Student Republican Association, Young Republican Club, Student Republican Association, Young Democrat Club, Student Democrat Association, Young Democrat Club, or Student Democrat Association). Applicants were randomly listed as the president, founder, secretary, vice-president or member of the assigned club.

Pronouns: Beginning in Wave 2, 10% of applicants were assigned preferred pronouns. Half of applicants with pronouns received gender-neutral pronouns (they/them/their), and half received pronouns reflecting the typical gender identity of their first name (he/him/his or she/her/hers). Pronouns were listed on the PDF resumes near name and contact information.

Employment history: Each applicant was assigned two to three previous employers. Employers were drawn from the universe of establishment names and addresses listed in the Reference USA dataset. As with addresses, we sample previous employers from zip codes within 20 miles of the target job's zip code, or 40 miles if insufficient employers are available within 20. We exclude any establishments from the same firm as the target job.

Each target job was assigned one of four employment categories: general, retail, clerical, and manual labor. Applicants to general category jobs were assigned previous employers from SIC codes 15, 24, 25, 34, 36, 42, 53, 54, 56, 58, 64, 65, 70, 73, and 80. Applicants to retail category jobs were assigned previous employers from codes 53, 54, 56, 58 and 70. Applicants to clerical jobs were assigned previous employers from codes 15, 24, 25, 34, 36, 64, 65, 73, and 80. Applicants to manual labor jobs were assigned previous employers from codes 34, 36, 25, 24, 15, and 42. Prior employers were assigned without replacement for all applications to the same target job.

Entry-level job titles were assigned for each previous employer appropriate to the industry and experience. Jobs at retail establishments were assigned job titles from Team Member, Retail Associate, Cashier, Stocker, and Customer Service Associate. Jobs at fast-food / quick-service restaurants were assigned titles from Crew Member, Cashier, Food prep / service, and Cook. Jobs at restaurants were assigned titles from Server, Dishwasher, Cashier, Host, and Cook. Jobs at manufacturers and wholesalers were assigned titles from Package Handler, Handler, Laborer, Delivery Driver / Courier, Dockworker, and Warehouse Associate. Office and clerical positions were assigned titles from Office Manager, Receptionist, and Assistant. Jobs at hotels were assigned titles of Housekeeper or Receptionist.

Each job was assigned a fictional supervisor with a first and last name drawn from the most common in the United States and a fictional phone number. Since some applications required us to list a reason for leaving each previous job, we populated a large list of sample reasons (e.g., insufficient hours, seeking promotion opportunity, etc.) and randomly assigned them to each previous job.

Tenure in previous jobs was selected uniformly from 9 to 24 months. No interruptions in employment history were assigned and all applicants reported being currently employed by their most recent prior employer.

We assigned a sample of two to three job duties scraped from online databases of resumes such as jobhero.com. We manually cleaned and formatted these duties to eliminate references to specific employer names or technologies. Duties were entered into "responsibilities / duties" sections of target job applications.

References: When required, applicants listed references using the fictional supervisors at their previous employers.

Personality and skills assessments: Some jobs required applicants to complete personality or skills assessments before they could be considered for an interview. We developed guides for each of these assessments that randomly specified acceptable answers within a range appropriate for the question. Our answers avoided providing an obviously negative signal about applicant quality (e.g., answering "Yes" to "Is it ever acceptable to steal from an employer?"). When questions had no obvious connection to applicant quality, we answered randomly but ensured that answers remained consistent across questions. We answered analytical-reasoning and skill-based questions to mimic the performance of our undergraduate volunteers.

Miscellaneous resume characteristics: Many applications required answering a large number of idiosyncratic questions, ranging from open-ended questions about why the applicant wants to work at the target employer to questions about willingness to comply with employer rules about dress, drug use, and conduct. We developed guides to answer each of these questions that either provided the most obviously "positive" answer or answered randomly from a bank of responses. Our applicants always answered "No" to any questions about possessing a prior criminal record.

Job sampling

We developed code that scraped all vacancies posted on each firm's proprietary hiring portal each day. We then manually identified the set of job titles that did not require a) a bachelor's or advanced degree, b) substantial prior experience, or c) a specialized license (at the time of application). When adding a new job for each firm, we selected randomly from among the most recently posted vacancies in counties from which we had not previously sampled a job for that firm. In rare cases no jobs were available in counties we had not previously sampled. In these cases we added new jobs in the same county but at different establishments to those sampled previously.

The *RandRes* platform automatically monitored scraped vacancies and added new jobs to the system. In each wave, we randomly sorted firms and worked through the sample by adding 5-10 jobs for each firm at a time to match maximum total application submission capacity.

Resume creation

RandRes features a PDF generator program that randomizes layout and design features to produce realistic resumes submitted as part of our application packages. The program parses an applicant's information generated by *RandRes*, include demographic details, employment history, and education history, and then randomly assigns a resume format including margins, font, text size, alignment, bullet shape, and other typical features. The process may redraw some features to ensure that resumes do not exceed one page in length or contain excessive white space.

The order and method in which information is presented is also random, meaning some applicants may list their education first while others list work experience first. Some resumes may include a separate section for references while others may include it as part of their employment history. Variations in language, such as whether or not to abbreviate U.S. state names, are also randomized.

The program tracks indicators of which special design attributes which have already been used in resumes for previous applicants at a particular job. This includes attributes such as off-white background coloring or a border around the contact information. Some resumes included monograms and watermarks as special attributes. A given resume may incorporate several of these design attributes together, but each special attribute is not used more than once at each job to ensure resumes are sufficiently differentiated. We find no evidence that special resume features increase contact rates.

We used the PDF resumes to signal characteristics not always collected in the online job application, such as year of high school graduation. When an applicant was assigned an LGBTQ or other student-club, the resume listed the club as part of educational experience. When an applicant was assigned preferred pronouns, they were listed in the resume below the applicant's name.

Application submission

The *RandRes* application platform automatically generated applications for all jobs active in the system. Applications were generated in pairs and new applications were generated whenever a job had fewer than two unsubmitted applications and no applications submitted within 24 hours. During Wave 1 of the experiment, applications were manually submitted by our team of undergraduate volunteers. *RandRes* instructed each volunteer which application to submit, provided the relevant details, and recorded submission status.

In subsequent waves, we developed software to automatically submit our applications to firms' job portals. By controlling a web browser, the software was able to visit the portal, fill out all application details, submit the application, and complete any assessments while operating at speeds designed to mimic human behavior. We used cloud computing providers to cycle through hundreds of IP addresses, user-agent strings, and other browser signatures to minimize our chances of detection.

We submitted up to 8 total applications to each job. Occasionally, vacancies would be closed or removed from hiring portals partway through our application process. Ninety-four percent of applicants were sent in complete groups of 8 and 88% of jobs received all 8 applications.

Appendix C Covariates

This Appendix provides details on sources and construction for the covariates used in Section 8.

Establishment-level covariates

- % county Black: Sourced from the U.S. Census's Longitudinal Employer-Household Dynamics Workplace Area Characteristics series. Measures the Black share of workers in 2015-2017 in the target job's county.

- % block Black / female: Same as above but defined at the census block level. Exact address data are not available for all jobs, making it impossible to match all jobs to census blocks. Only matched jobs are included.

- County IAT: Constructed using raw data from Harvard's Project Implicit. Defined as the average of all valid 2015 - 2020 IAT scores in each county, normalized to have a standard deviation of one within year. A higher value indicates more implicit bias against Black or female faces in the test. The female IAT used contrasts male vs. female faces with Science vs. Liberal Arts.

- DMA animus: Relative Google search rates for racially charged epithets as studied in Stephens-Davidowitz (2014). DMA refers to the target job's Designated Market Area. Higher values indicate more racially charged searchers. Normalized to have a standardized deviation of 1 within year and averaged over 2015-2019.

- State animus: Same as above but defined at state-level.

- White manager: Sourced from Reference USA establishment-level data. White manager indicates that Reference USA listed at least one "Manager", "Site Manager", or "Office Manager" as ethnically "Western European", "Eastern European", "Scandinavian", or "Mediterranean." Not all establishments were able to be linked to the Reference USA data, and not all establishments in Reference USA had manager ethnicity information. Only jobs with valid data are included. Constructed with the most recently available Reference USA data set.

- Male manager: Same as above but defined as at least one manager with gender listed as "Male."

- Log employment: Sourced from Reference USA establishment-level data. Normalized to have standard deviation of one in sample.

Firm-level covariates. All firm-level covariates are normalized to have a standard deviation of one in sample.

- Log employment: Total US employment scraped from most recent publicly available data online, including annual reports and firm websites.
- DOL viols/emp: Includes all wage and hour compliance violations since FY 2005 reported by the Department of Labor. Normalized by total employment.
- Empl-discr cases/emp: Data scraped from `https://www.goodjobsfirst.org/violation-tracker`. Defined as the total count of reported penalties since 2000 where the primary offense category is "Employment Discrimination" divided by employment. Firms with no penalties reported are coded as zeros.
- Sales / emp: Data from Dun and Bradstreet. Defined as total sales divided by DB-reported employment averaged over 2010-2018.
- Profit / emp: Data from Compustat. Defined as average gross profit divided by Compustat-reported employment averaged over 2010-2018. Three firms do not have Compustat data and are omitted.
- % board Black: Measures the average Black share of the corporate board over 2014-2019. Board member race sourced from blackenterprise.com and manual searches.
- Chief diversity officer: Binary indicator manually scraped from company websites. Includes C-Suite executives only.
- GD score: Overall company rating scraped from GlassDoor.com.
- GD diversity score: Diversity score ratings scraped from GlassDoor.com.
- Callback centralization: Defined as total number of unique phone numbers that contacted applicants the firm divided by the total number of jobs where applicants received at least one contact times minus 1. To avoid any mechanical correlation with outcomes, constructed as a leave-out mean omitting any contacts to own job.
- % managers white: Sourced from Reference USA. Measures share of managers at all establishments belonging to this firm with race reported as defined in establishment-level covariates. Two firms do not appear in the Reference USA data.
- % managers male: Same as above but defined as share of managers reported to be male.

Industry-level covariates.

- White adj wage, white - Black adj wage, male adj wage, male - female adj wage: Constructed using the CPS Monthly Outgoing Rotation Groups from 2009 to 2019,

Table C1: Summary statistics for firm-level covariates

	Mean	SD	Median
Firm performance			
Log employment	11.067	1.01	10.922
Sales / emp ($M)	0.331	0.36	0.238
Profit / emp ($M)	0.101	0.08	0.078
GD score	3.566	0.32	3.600
Legal compliance			
DOL viols / emp	0.136	0.37	0.002
Empl-discr cases / thousand emp	0.048	0.13	0.020
Federal contractor	0.667	0.47	1.000
Firm diversity			
% board Black	0.088	0.07	0.091
% board female	0.257	0.10	0.255
% managers non-white	0.257	0.09	0.250
% managers female	0.493	0.39	0.449
Has chief diversity officer	0.167	0.37	0.000
GD diversity score	3.816	0.33	3.800
Callback patterns			
Callback centralization	-1.117	0.38	-1.073
Observations	108		

Notes: This table reports summary statistics for firm-level covariates. See Appendix C for full details on the sources and construction of each variable.

extracted from IPUMS at `https://cps.ipums.org/cps/`. Sample includes individuals aged 20-60 who work full-time (35+ hours a week) in the private sector that do not have imputed earnings or hours worked. To obtain 2-digit SIC industry codes, we link IPUMS variable IND1990 with 1987 SIC industry codes using a crosswalk from Autor, Dorn, and Hanson (2019). Wage gaps are obtained from a regression of log hourly wages (equal to weekly earnings divided by usual hours worked per week) on indicators for each industry, for being black (female), their interaction, and a set of year indicators. Adjusted wage gaps correspond to the same coefficients from regressions with an indicator for female (or Black when constructing gender gaps), education (6 categories), and a quartic in age also included. All calculations use CPS household or earnings weights.

- % ind Black, % ind female: Constructed using the Equal Employment Opportunity Commission's 2018 public use file of EEO-1 data. Defined as the Black (female) share of workers in the NAICS 3-digit industry.

- % mgmt - % ind Black, % mgmt - % ind female: Constructed using same data as above. Defined as the Black (female) share of mid-level officers and managers less the total Black (female) share of workers in the NAICS 3-digit industry.

Table C2: Firm-level predictors of centralization

Firm performance	
Log employment	-0.144
	(0.116)
Sales / emp	-0.0682
	(0.0813)
Profit / emp	-0.00380
	(0.0797)
GD score	-0.211
	(0.173)
Legal compliance	
DOL viols / emp	-0.0836
	(0.121)
Empl-discr cases / emp	0.0576
	(0.0427)
Federal contractor	0.559**
	(0.265)
Firm diversity	
% board Black	0.178
	(0.116)
% board female	-0.0153
	(0.0982)
% managers non-white	0.00869
	(0.123)
% managers female	0.0553
	(0.0986)
Has chief diversity officer	0.148
	(0.211)
GD diversity score	0.223
	(0.157)
Observations	10500

Notes: This table reports the multivariate relationship between centralization and other firm-level predictors. All predictors except the binary indicators for federal contractor status and having a chief diversity officer are normalized to have standard deviation of 1. As with firm-level relationships reported in Figure 5, the regression is estimated on job-level data with firm-clustered standard errors. See Appendix C for full details on the sources and construction of each variable.

- White - Black col share, male - female col share: Constructed using the same CPS sample and data as adjusted wage gaps. College share gaps are equal to the Black-white difference in the share of workers with a college degree in each industry.

- Top 4 sales share: Defined as the share of total sales accounted for by the four largest firms at the NAICS 3-digit level. Sourced from 2017 Economic Census data.

Occupation-level covariates.

- O*NET occupation task measures: We follow Deming (2017) and use the Occupational Information Network (O*NET), available at `https://www.onetcenter.org/db_releases.html`, to measure characteristics of occupations in the U.S.[28] The O*NET database provides information on various components of an occupation, including the *skills*, *knowledge*, and *abilities* required to perform the work, the *activities* typically performed on the job, and the *context*, or characteristics and conditions, of the job. We use this information to create the following five composite variables:

  - Analytical: Our analytic measure combines the following three components: 1) *mathematical reasoning ability* (defined as "the ability to understand and organize a problem and then to select a mathematical method or formula to solve the problem"), 2) *mathematics knowledge* ("knowledge of numbers, their operations, and interrelationships including arithmetic, algebra, geometry, calculus, statistics, and their applications"), and 3) *mathematics skill* ("using mathematics to solve problems").
  - Social: Our social measure combines the following three skills: 1) *social perceptiveness* (defined as "being aware of others' reactions and understanding why they react the way they do"), 2) *coordination* ("adjusting actions in relation to others' actions"), 3) *persuasion* ("persuading others to approach things differently"), and 4) *negotiation* ("bringing others together and trying to reconcile differences").
  - Routine: Our routine measure combines two context variables, in particular 1) degree of automation (defined as "the level of automation of this job") and 2) importance of repeating same tasks ("how important is repeating the same physical activities or mental activities over and over, without stopping, to performing this job?").
  - Service: Our service measure measure combines the activity variable *assisting and caring for others* (defined as "providing assistance or personal care to

[28]Unlike Deming (2017), we use production release 25.3 of O*NET.

others") and the skill variable *service orientation* ("actively looking for ways to help people").

- Manual: Our manual measure combines two skill variables, specifically 1) *performing general physical activities* (defined as "performing physical activities that require considerable use of your arms and legs and moving your whole body, such as climbing, lifting, balancing, walking, stooping, and handling of materials") and 2) *handling and moving objects* ("using hands and arms in handling, installing, positioning, and moving materials, and manipulating things").
- Customer interaction: Our customer interaction measure averages two activities variables, one knowledge variable, and one context variable. The work activities variables include 1) *performing for or working directly with the public* (defined as "performing for people or dealing directly with the public") and 2) *establishing and maintaining interpersonal relationships* ("developing constructive and cooperative working relationships with others, and maintaining them over time"). We use the work knowledge variable *customer and personal service* ("knowledge of principles and processes for providing customer and personal services) and the work context variable *contact with others*, which answers the question "how much does this job require the worker to be in contact with others (face-to-face, by telephone, or otherwise) in order to perform it?"

Each composite variable is calculated as the average of its component variables. Since some of these component variables are measured on different scales, we first rescale all the component variables to fall between 0 and 10.

Appendix D Technical Appendix

Denote the realized contact gap at job $j \in \{1, ..., J_f\}$ of firm $f \in \{1, ..., F\}$ by $\hat{\Delta}_{fj}$. For most of our analysis $\hat{\Delta}_{fj}$ is measured as the difference between white and Black contact rates at job j, but the same construction is used to study other binary protected characteristics such as gender. Let $D_{fj} \in \Omega$ give the *design* (i.e., assigned characteristics) of the portfolio of resumes sent to job j. This design includes, for example, the mix of employment histories on each resume, the time of day each resume was sent, each applicant's year of high school graduation, and the formatting of the resumes. Define $\hat{\Delta}_{fj}(d)$ as the contact gap that would arise at job j if it had been assigned application design d. Realized contact gaps can be written $\hat{\Delta}_{fj} = \hat{\Delta}_{fj}(D_{fj})$. Population contact gaps are defined as

$$\Delta_{fj} \equiv \mathbb{E}\left[\hat{\Delta}_{fj}(D_{fj}) \mid \left\{\hat{\Delta}_{fj}(d)\right\}_{d\in\Omega}\right] = \sum_{d\in\Omega} \omega_{fjd} \hat{\Delta}_{fj}(d),$$

where $\omega_{fjd} \in (0, 1)$ is the probability that design d is assigned to job j of firm f. Note that the expression after the equals sign presumes that the assignment probabilities $\{\omega_{fjd}\}$ are independent of the potential contact gaps $\{\hat{\Delta}_{fj}(d)\}$, a property ensured by random assignment. Assignment probabilities may differ by f as, for example, applicant job histories were tailored to the firms being studied. The $\{\omega_{fjd}\}$ may also differ across jobs, as local educational institutions and references were listed on applicant resumes.

We now make two key assumptions:

Assumption 1 (Design uncertainty) *The errors* $\left\{\hat{\Delta}_{fj} - \Delta_{fj}\right\}_{f=1,j=1}^{F,J_f}$ *are mutually independent and have mean zero.*

Assumption 2 (Sampling uncertainty) *Each firm's population gaps* $\{\Delta_{fj}\}_{j=1}^{J_f}$ *are iid draws from a firm specific distribution* G_f *with mean* Δ_f.

Assumption 1 follows from random assignment of application characteristics. This condition also implicitly requires the behavioral assumption of no interference between jobs, an assumption made more plausible by the requirement that sampled jobs be located in different U.S. counties. Assumption 2 follows from *i.i.d.* sampling of jobs from the set of available vacancies posted on company job boards. The mean Δ_f, which is our measure of discrimination at firm f, gives the expected contact gap at an average job posting by firm f over the course of our study.

Together, these assumptions yield a hierarchical model with two sources of uncertainty. The first source ("design uncertainty") arises from randomness in the application design assigned to each job. The second ("sampling uncertainty") arises from randomness in the set of jobs sampled. We use the operator $\mathbb{E}[\cdot]$ to denote expectations with respect to both sorts of uncertainty; that is, to denote integration against G_f and the design probabilities $\{\omega_{fjd}\}_{d\in\Omega}$. Our assumptions thus far imply that

$$\mathbb{E}\left[\hat{\Delta}_{fj}|\Delta_{fj}\right] = \Delta_{fj}, \quad \mathbb{E}\left[\hat{\Delta}_{fj}\right] = \Delta_{f}.$$

Target parameter

The variance of the firm component of discrimination can be defined as

$$\begin{aligned}\theta &= \frac{1}{F}\sum_{f=1}^{F}\Delta_f^2 - \left(\frac{1}{F}\sum_{f=1}^{F}\Delta_f\right)^2 \\ &= \left(\frac{F-1}{F}\right)\left\{\frac{1}{F}\sum_{f=1}^{F}\Delta_f^2 - \frac{2}{F(F-1)}\sum_{f=2}^{F}\sum_{k=1}^{f-1}\Delta_f\Delta_k\right\}.\end{aligned}$$

Bias corrected estimator

The fundamental difficulty in estimating θ involves the first term in the curly brackets. Let $\hat{\Delta}_f = \frac{1}{J_f}\sum_{j=1}^{J_f}\hat{\Delta}_{fj}$ denote the mean contact gap at firm f. Both design and sampling uncertainty generate an upward bias in the "plug-in" estimator $\left(\hat{\Delta}_f\right)^2$ of Δ_f^2 because

$$\begin{aligned}\mathbb{E}\left[\left(\hat{\Delta}_f\right)^2\right] &= \mathbb{E}\left[\left(\hat{\Delta}_f - \Delta_f\right)^2\right] + \Delta_f^2 \\ &= \mathbb{E}\left[\left(\underbrace{\hat{\Delta}_f - \frac{1}{J_f}\sum_{j=1}^{J_f}\Delta_{fj}}_{\text{design error}} + \underbrace{\frac{1}{J_f}\sum_{j=1}^{J_f}\Delta_{fj} - \Delta_f}_{\text{sampling error}}\right)^2\right] + \Delta_f^2 \\ &> \Delta_f^2.\end{aligned}$$

The bias corrected estimator of θ is motivated by the approximation $\mathbb{E}\left[\left(\hat{\Delta}_f - \Delta_f\right)^2\right] \approx s_f^2$, where s_f is an estimated standard error. When this approximation holds exactly, we have $\mathbb{E}\left[\hat{\Delta}_f^2\right] = \Delta_f^2 + s_f^2$. The bias corrected estimator can be written

$$\begin{aligned}\hat{\theta} &= \left(\frac{F-1}{F}\right)\left\{\underbrace{\frac{1}{F-1}\sum_{f=1}^{F}\left(\hat{\Delta}_f - \frac{1}{F}\sum_{k=1}^{F}\hat{\Delta}_k\right)^2}_{\text{plug-in}} - \underbrace{\frac{1}{F}\sum_{f=1}^{F}s_f^2}_{\text{correction}}\right\} \\ &= \left(\frac{F-1}{F}\right)\left\{\frac{1}{F}\sum_{f=1}^{F}\left(\hat{\Delta}_f^2 - s_f^2\right) - \frac{2}{F(F-1)}\sum_{f=2}^{F}\sum_{k=1}^{f-1}\hat{\Delta}_f\hat{\Delta}_k\right\}.\end{aligned}$$

Variants of this estimator have been applied in several literatures (e.g., Krueger and Summers, 1988; Aaronson et al., 2007), though typically without the adjustment factor of $\frac{F-1}{F}$.

In our analysis, we employ the following standard error estimator

$$s_f = \sqrt{\frac{1}{J_f\left(J_f-1\right)} \sum_{j=1}^{J_f} \left(\hat{\Delta}_{fj} - \hat{\Delta}_f\right)^2}.$$

With this choice of s_f, $\hat{\theta}$ becomes an unbiased leave out variance component estimator of the sort proposed by Kline et al. (2020). In particular, it can be shown that

$$\hat{\Delta}_f^2 - s_f^2 = \frac{2}{J_f\left(J_f-1\right)} \sum_{j=2}^{J_f} \sum_{\ell=1}^{j-1} \hat{\Delta}_{fj}\hat{\Delta}_{f\ell} = \frac{1}{J_f} \sum_{j=1}^{J_f} \hat{\Delta}_{f(j)}\hat{\Delta}_{fj},$$

where $\hat{\Delta}_{f(j)} = \frac{1}{J_f-1}\sum_{\ell\neq j}\hat{\Delta}_{f\ell}$ is the leave-job out mean contact gap at firm f.

Independence of the errors across jobs guarantees that $\mathbb{E}[\hat{\Delta}_{fj}\hat{\Delta}_{f\ell}] = \mathbb{E}[\Delta_{fj}]\mathbb{E}[\Delta_{f\ell}] = \Delta_f^2$, with the second equality following from random sampling of jobs (Assumption 2). Likewise, independence of both design and sampling errors across firms ensures that $\mathbb{E}[\hat{\Delta}_f\hat{\Delta}_k] = \mathbb{E}[\hat{\Delta}_f]\mathbb{E}[\hat{\Delta}_k] = \Delta_f\Delta_k$. Consequently, $\mathbb{E}[\hat{\theta}] = \theta$. Lemma 3 of Kline et al. (2020) establishes consistency of $\hat{\theta}$ for θ as the total number of jobs $\sum_{f=1}^F J_f$ grows large. Asymptotic normality of $\hat{\theta}$ follows from Theorem 2 of Kline et al. (2020).

Cross-wave estimator

The cross wave estimator of θ is analogous to $\hat{\theta}$ but uses cross-products of wave level, as opposed to job-level, average gaps to estimate Δ_f^2. Suppose that for any two waves $(\tau_1,\tau_2) \in \{1,...,T_f\}^2$

$$\mathbb{E}\left[\hat{\hat{\Delta}}_{f\tau_1}\hat{\hat{\Delta}}_{f\tau_2}\right] = \Delta_f^2 \quad \text{if } \tau_1 \neq \tau_2,$$

where $\hat{\hat{\Delta}}_{f\tau}$ is the mean gap in wave τ. This moment condition would follow from Assumptions # 1 and # 2 if each firm's distribution of population job gaps were restricted to be time invariant. An unbiased estimator of Δ_f^2 is the (job-weighted) cross-wave analogue of this moment condition:

$$\widehat{\Delta_f^2} \equiv \frac{\sum_{\tau_1=2}^{T_f}\sum_{\tau_2=1}^{\tau_1-1} n_{f\tau_1}n_{f\tau_2}\hat{\Delta}_{f\tau_1}\hat{\Delta}_{f\tau_2}}{\sum_{\tau_1=2}^{T_f}\sum_{\tau_2=1}^{\tau_1-1} n_{f\tau_1}n_{f\tau_2}},$$

where $n_{f\tau}$ gives the number of jobs sampled from firm f in wave τ. Our corresponding unbiased cross-wave estimator of θ is

$$\left(\frac{F-1}{F}\right)\left\{\frac{1}{F}\sum_{f}\widehat{\Delta_f^2}-\frac{2}{F\,(F-1)}\sum_{f=2}^{F}\sum_{k=1}^{f-1}\hat{\Delta}_f\hat{\Delta}_k\right\}.$$

Cross-state estimator

The cross state estimator is identical to the cross-wave estimator except that cross-products between state averages of job contact gaps at each firm replace wave averages of job contact gaps at each firm. As with the cross-wave estimator, the cross-products of averages are job weighted.

Industry and portal intermediary variance components

Firm identifiers are "nested" within industry and job portal intermediary categories. Variance components for these alternate groupings of jobs can be defined as weighted analogues of the firm level component θ.

Working with industry as our focal example, let $\ddot{\Delta}_i$ denote the population contact gap in industry $i \in \{1,...,I\}$, which we define as the equally weighted average of the population contact gaps among firms in that industry. Letting F_i be the number of firms in industry i and $F=\sum_{i=1}^{I}F_i$ the total number of firms in the experiment, the industry component can be written:

$$\begin{aligned}\theta_I &= \frac{1}{F}\sum_{i=1}^{I}F_i\ddot{\Delta}_i^2-\left(\frac{1}{F}\sum_{i=1}^{I}F_i\ddot{\Delta}_i\right)^2\\ &= \left(\frac{F-1}{F}\right)\left\{\frac{1}{F(F-1)}\sum_{i=1}^{I}F_i(F-F_i)\ddot{\Delta}_i^2-\frac{2}{F\,(F-1)}\sum_{i=2}^{I}\sum_{k=1}^{i-1}F_iF_k\ddot{\Delta}_i\ddot{\Delta}_k\right\}\end{aligned}$$

The firm weighting used in this definition ensures that the ratio $\theta_I/\theta \in [0,1]$ possesses an R^2 interpretation. When $\theta_I=\theta$ industry explains all of the variation across firms.

Mirroring the firm-level analysis, an unbiased estimate of the squared mean $\ddot{\Delta}_i^2$ can be constructed as a weighted average of cross-products of job-level gaps in industry i. To preserve the interpretation of $\ddot{\Delta}_i$ as an equally weighted average of contact gaps across firms in an industry, we weight jobs inversely by "firm size" when computing these cross-products. Indexing jobs in industry i by $n \in \{1,...,N_i\}$, let $\hat{\Delta}_{in}$ give the estimated contact gap at that job. Using $f(i,n)$ to denote the parent company of job n our job weights can be written $w_{in}=1/J_{f(i,n)}$. Note that w_{in} gives the inverse of the total number of jobs at the parent firm containing job n. Hence, an unbiased estimator of $\ddot{\Delta}_i$

is $\left(\sum_{n=1}^{N_i} w_{in}\right)^{-1}\left(\sum_{n=1}^{N_i} w_{in}\hat{\Delta}_{in}\right)$. Our corresponding estimator for $\ddot{\Delta}_i^2$ can be written:

$$\widehat{\ddot{\Delta}_i^2} \equiv \frac{\sum_{n=2}^{N_i}\sum_{k=1}^{n-1} w_{in}w_{ik}\widehat{\Delta}_{in}\widehat{\Delta}_{ik}}{\sum_{n=2}^{N_i}\sum_{k=1}^{n-1} w_{in}w_{ik}}.$$

Plugging these unbiased estimators of $\ddot{\Delta}_i$ and $\ddot{\Delta}_i^2$ into the expression for θ_I yields the unbiased industry variance component estimator $\hat{\theta}_I$.

State and job title variance components

Defining state and job title variance components requires some additional notation, as these groupings of jobs do not nest firms. Working with state as our focal example, we index states by $s \in \{1, ..., S\}$ and jobs in states by $b \in \{1, ..., B_s\}$. Accordingly, we denote the population gap at job b of state s by Δ_{sb}. Letting $w_{f(s,b)} = 1/J_f$ denote the inverse size of the firm containing job b, and $W_s = \sum_{b=1}^{B_s} w_{f(s,b)}$, the sum of these weights, the overall population gap in state s is defined as

$$\dddot{\Delta}_s = \frac{1}{W_s}\sum_{b=1}^{B_s} w_{f(s,b)}\Delta_{sb}.$$

Letting $W = \sum_{s=1}^{S} W_s$ be the total number of firms in the experiment, our variance component of interest is:

$$\begin{aligned}
\theta_S &= \frac{1}{W}\sum_{s=1}^{S} W_s\dddot{\Delta}_s^2 - \left(\frac{1}{W}\sum_{s=1}^{S} W_s\dddot{\Delta}_s\right)^2 \\
&= \left(\frac{W-1}{W}\right)\left\{\frac{1}{W(W-1)}\sum_{s=1}^{S} W_s(W-W_s)\dddot{\Delta}_s^2 - \frac{2}{W\left(W-1\right)}\sum_{s=2}^{S}\sum_{k=1}^{s-1} W_sW_k\dddot{\Delta}_s\dddot{\Delta}_k\right\}.
\end{aligned}$$

To estimate θ_S we substitute $\widehat{\dddot{\Delta}}_s = \frac{1}{W_s}\sum_{b=1}^{B_s} w_{f(s,b)}\hat{\Delta}_{sb}$ for $\dddot{\Delta}_s$ in the second term in braces. The quantity $\dddot{\Delta}_s^2$ entering the first term in braces is replaced with the weighted average cross-product:

$$\frac{\sum_{b=2}^{B_s}\sum_{k=1}^{b-1} w_{f(s,b)}w_{f(s,k)}\widehat{\Delta}_{sb}\widehat{\Delta}_{sk}}{\sum_{b=2}^{B_s}\sum_{k=1}^{b-1} w_{f(s,b)}w_{f(s,k)}}.$$

Appendix E Alternative Deconvolution Estimates

This section explores the robustness of estimated population contact gap distributions to alternative models for the relationship between estimated gaps, $\hat{\Delta}_f$, and their standard errors, s_f. Our baseline analysis assumes that s_f is independent of the population z-score Δ_f/s_f. After applying the Efron (2016) deconvolution estimator to the sample z-scores $\hat{\Delta}_f/s_f$, we recover the distribution of Δ_f by numerically integrating against the empirical distribution of standard errors. Here we consider three alternatives: a variance stabilizating transformation approach, a "local deconvolution" approach that separately estimates population distributions among groups of firms with similar standard errors, and a non-parametric approach that estimates the joint distribution of contact gaps and standard errors.

Appendix E.1 Variance stabilizing transformation

If one assumes a parametric model for the dependence of the firm specific variances on the latent contact gaps of the form $s_f^2 = h(\Delta_f)$, then heteroscedasticity can be eliminated via the variance stabilizing transformation:

$$y(t) = \int_{\infty}^{t} h(x)^{-1/2} dx.$$

Note that $\frac{d}{dt}y(t) = h(t)^{-1/2}$. Hence, standard delta-method reasoning implies that $y(\hat{\Delta}_f)|\Delta_f \sim \mathcal{N}(y(\Delta_f), 1)$. Applying the deconvolution estimator of Efron (2016) to the transformed estimates $y(\hat{\Delta}_f)$, one can then generate an estimate of the population distribution of Δ_f using the change of variables $\hat{g}_\Delta(x) = \hat{g}_{y(\Delta)}(x)h'(x)$, where $\hat{g}_t(\cdot)$ is the estimated density of t.

To implement this approach, we allow for non-linear dependence of the (squared) standard errors on contact gaps by assuming that

$$h(\Delta) = \alpha + \beta_1\Delta + \beta_2\Delta^2 \quad \text{for } \Delta \in \mathcal{S},$$

where $\mathcal{S}$ is the support of the population contact gap under study. We use split-sample IV (Angrist and Krueger, 1995) to estimate the parameters (β_1, β_2). Splitting each firm's jobs into two groups $g \in \{0, 1\}$, we proxy each group's values of Δ_f and Δ_f^2 with $\hat{\Delta}_{fg}$ and $\hat{\Delta}_{fg}^2 - s_{fg}^2$, respectively, where s_{fg} is the standard error of $\hat{\Delta}_{fg}$. We then estimate β_1 and β_2 via a regression of s_{fg}^2 on $(\hat{\Delta}_{fg}, \hat{\Delta}_{fg}^2 - s_{fg}^2)$ using as instruments $(\hat{\Delta}_{f(g)}, \hat{\Delta}_{f(g)}^2 - s_{f(g)}^2)$, where $(g) = 1 - g$ refers to the omitted group. To minimize uncertainty attributable to the splitting process, we take the median across 1,000 iterations of this procedure.

Figure E1 presents the resulting deconvolved population distributions of contact gaps for race and gender. As in the primary estimates, the race gap distribution exhibits a peak close to zero and a fat right tail of heavy discriminators. The distribution of gender gaps continues to show concentrated mass near zero and severe discriminators in both tails. Figure E2 summarizes the concentration of discrimination based on the variance-stabilized approach by plotting the Lorenz curves implied by the deconvolved density $\hat{g}_{\Delta}$. Similar to our baseline estimates in Figure 7, these curves imply discrimination is concentrated in a relatively small share of firms for both race and gender. Finally, Figure E3 shows the estimated racial contact gap distribution based on the variance-stabilization approach without restricting the density of Δ_f to be weakly positive, which produces minimal changes to the results.

Appendix E.2 Local deconvolutions

A less parametric approach to dealing with heteroscedasticity in the contact gaps is to simply estimate the deconvolution within bins defined by ranges of s_f. This approach weakens the requirement that s_f be independent of the population z-score Δ_f/s_f in the full population to a requirement that independence only hold among firms with similar s_f. To implement this approach, we split firms into two groups $k \in \{H, L\}$ by whether their contact gap standard error falls above / below the sample median standard error. We then apply the deconvolution estimator of Efron (2016) to the sample z-scores in each group, $\hat{\Delta}_{fk}/s_{fk}$, and recover the group-specific population contact gap density, $g_{\Delta,k}$, by integrating against the empirical distribution of standard errors in that group. The marginal density of contact gaps is then given by the mixture:

$$\hat{g}_{\Delta}(x) = \frac{1}{2} g_{\Delta,H} + \frac{1}{2} g_{\Delta,L}$$

We use a common penalization parameter in the deconvolution step for both groups and calibrate it so that the resulting marginal distribution matches the corresponding bias-corrected variance estimate from Table 4.

Figure E4 shows the resulting group-specific densities for both race and gender. Figure E5 shows the corresponding marginal densities. As in the primary estimates, the race density shows concentrated mass close to zero and fat right tail. The gender density is strongly peaked at zero. Both densities continue to show that discrimination is strongly concentrated in a relatively small share of firms, as shown in Lorenz curves presented in Figure E6.

The close agreement of the top 20% share estimates found in Figures 7, E6, and E2 is reassuring and suggests our modeling of heteroscedasticity patterns is not an important factor driving our concentration results.

Appendix E.3 NPMLE

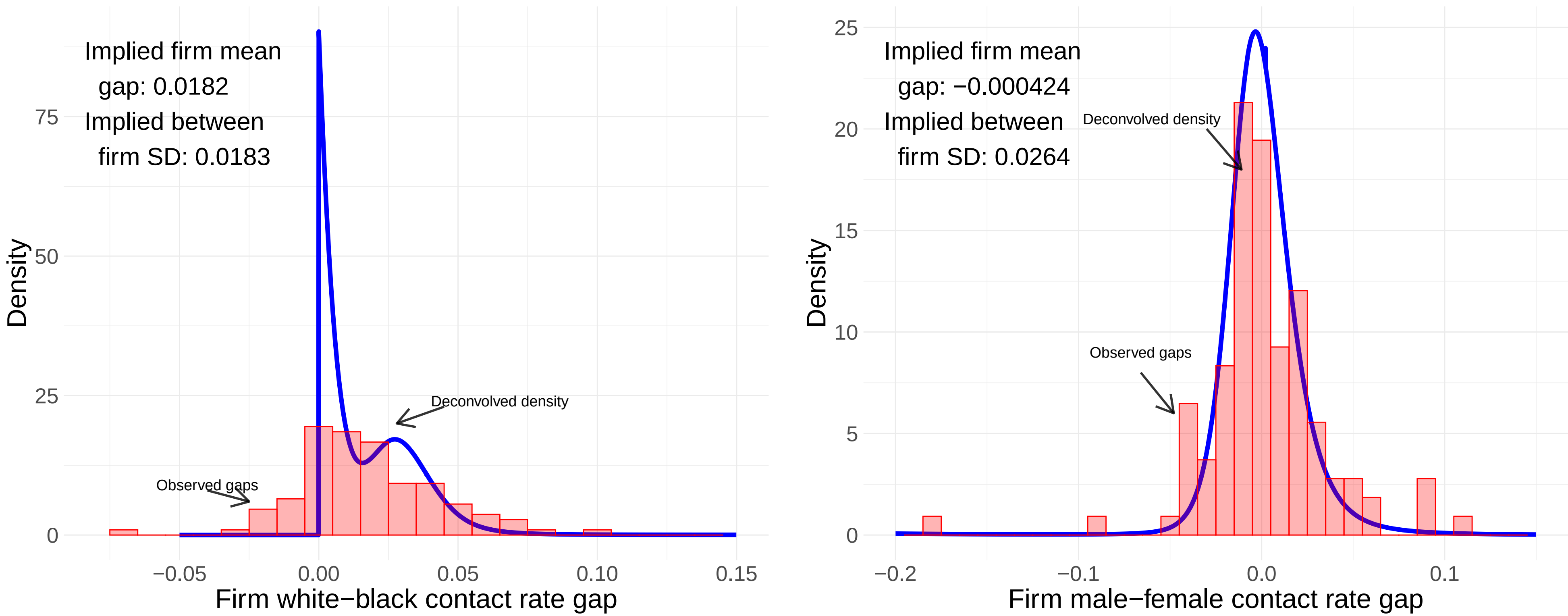
As a final approach to accounting for heteroscedasticity, we estimate a non-parametric mixing distribution over contact gaps and standard errors that allows for unrestricted dependence between these objects. To implement this approach, we use the approximation to the Kiefer-Wolfowitz non-parametric maximum likelihood estimator (NPMLE) developed by Koenker and Mizera (2014) and implemented in the REBayes package of Koenker and Gu (2017).

Figure E7 presents the resulting discrete marginal densities of contact gaps for race and gender. NPMLE estimates of the distribution of racial discrimination show similar patterns to our earlier spline approximations, with a concentrated mass of firms exhibiting limited discrimination and a fat tail of more heavy discriminators. Gender estimates show substantial mass near zero and smaller mass points in the extremes of both tails.

Figure E8 shows that these densities also imply substantial concentration of discrimination among a subset of employers. For comparability with prior results, we linearly interpolate between mass points, which yields kinks in the Lorenz curves. The interpolated top 20% shares are slightly higher than in our baseline specification utilizing spline approximations, suggesting again that our concentration finding is highly robust. The Gini coefficients are also close to those found in our baseline specification, with the race Gini slightly higher and the gender Gini slightly lower than the corresponding estimates in Figure 7.

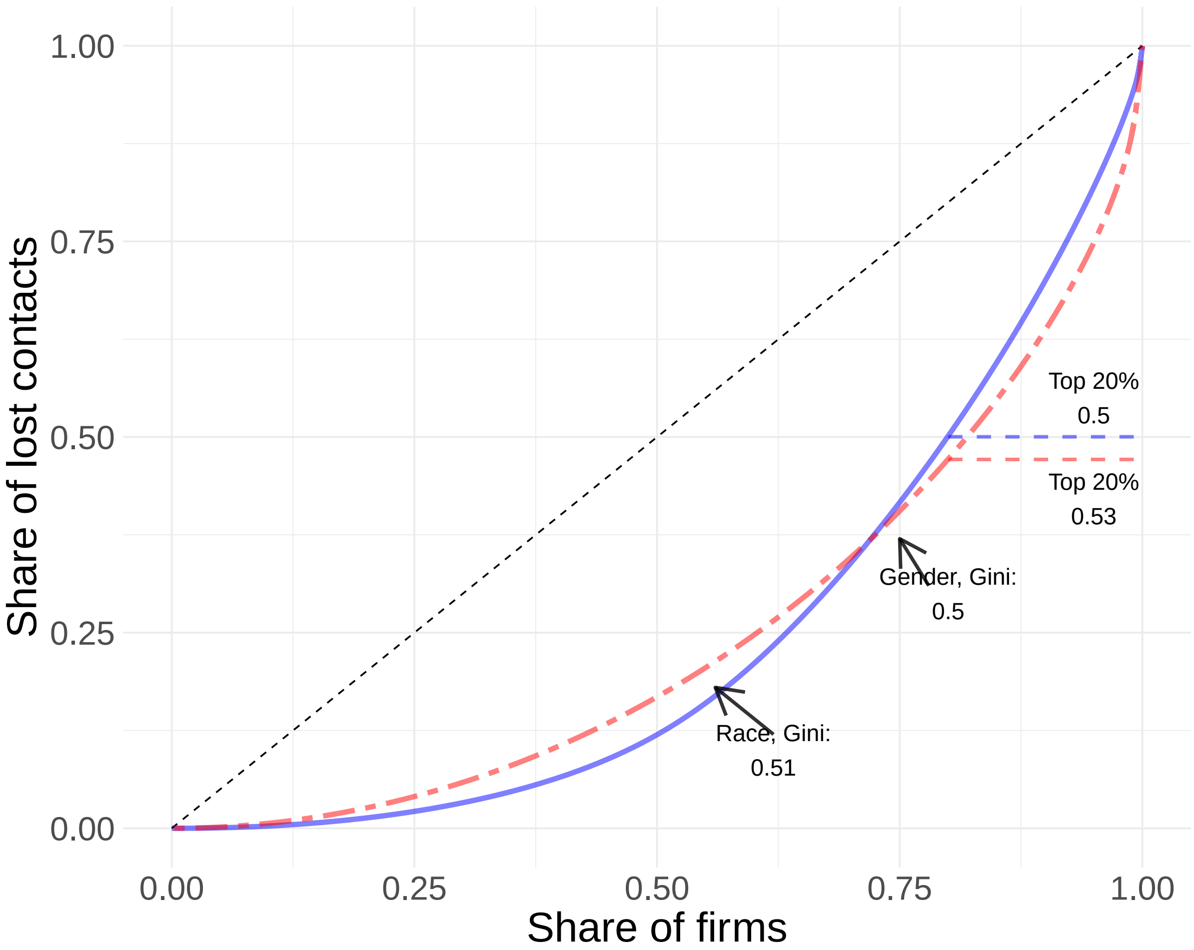
Figure E1: Variance-stabilized deconvolution of firm-level discrimination distributions

a) Race b) Gender



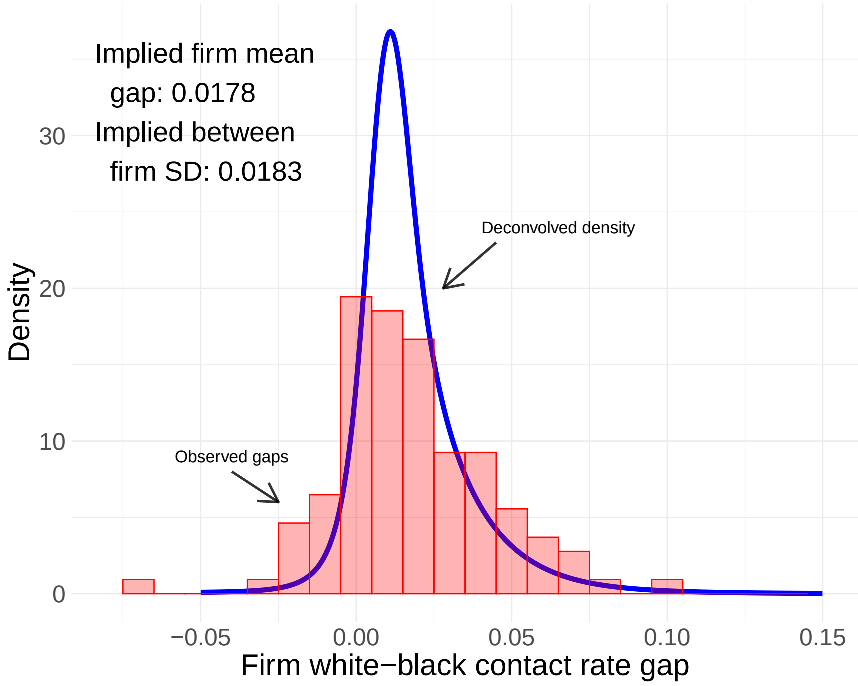
Notes: This figure presents non-parametric estimates of the distribution of firm-specific contact gaps. Panel (a) presents estimates for white-Black contact rate differences, and panel (b) presents estimates for male-female differences. Red histograms show the distribution of estimated firm contact gaps. Blue lines shows estimates of population contact gap distributions. The population distributions are estimated by applying the deconvolveR package (Narasimhan and Efron, 2020) to variance-stabilized estimates of firm-specific contact gaps. The variance-stabilizing transformation is constructed by assuming that $s_f^2 = \alpha + \beta_1 \Delta_f + \beta_2 \Delta_f^2$, with α, β_1, and β_2 estimated via split-sample IV (Angrist and Krueger, 1995). The estimated population distribution of transformed gaps is transformed into the distribution of Δ_f using the change of variables formula. The penalization parameter in the deconvolution step is calibrated so that the resulting distribution matches the corresponding bias-corrected variance estimate from Table 4. In panel (a), the density of population Δ_f is constrained to be weakly positive.

Figure E2: Variance-stabilized discrimination Lorenz curves



Notes: This figure displays Lorenz curves implied by the non-parametric deconvolution estimates of race and gender contact gap distributions in Figure E1. The solid blue curve is the Lorenz curve for the white/Black contact gap, and the dashed red curve is the Lorenz curve for the absolute value of the male/female contact gap. The Lorenz curve reports the share of lost contacts in the experiment attributable to firms below each contact gap percentile. The share of lost contacts equals the sum of contact gaps at firms below a particular contact gap percentile as a share of the sum of contact gaps across all firms. The dashed line is the 45 degree line. The labels for each curve also report Gini coefficients, equal to 1 minus twice the area under each curve.

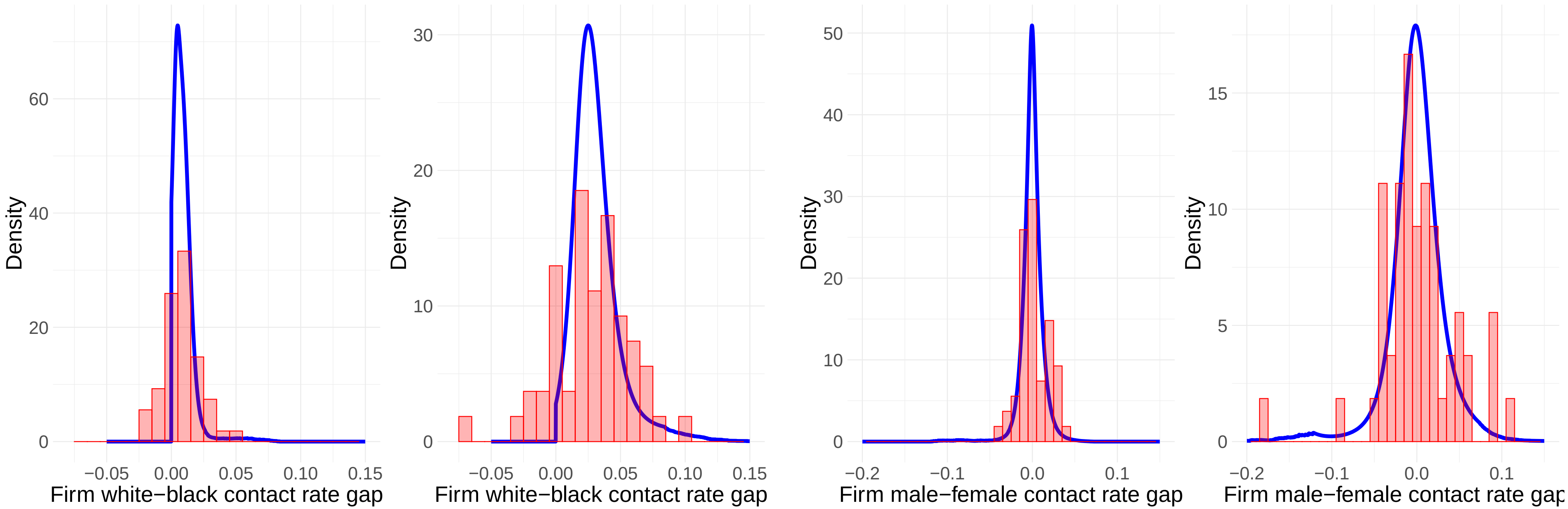
Figure E3: Variance-stabilized deconvolutions of racial discrimination without support restriction



Notes: This figure presents non-parametric estimates of the distribution of firm-specific white-Black contact rate differences. The red histogram shows the distribution of estimated firm contact gaps. Blue line shows an estimate of the population contact gap distribution constructed as in Panel (a) of Figure E1, but without the restriction that the density of Δ_f is weakly positive.

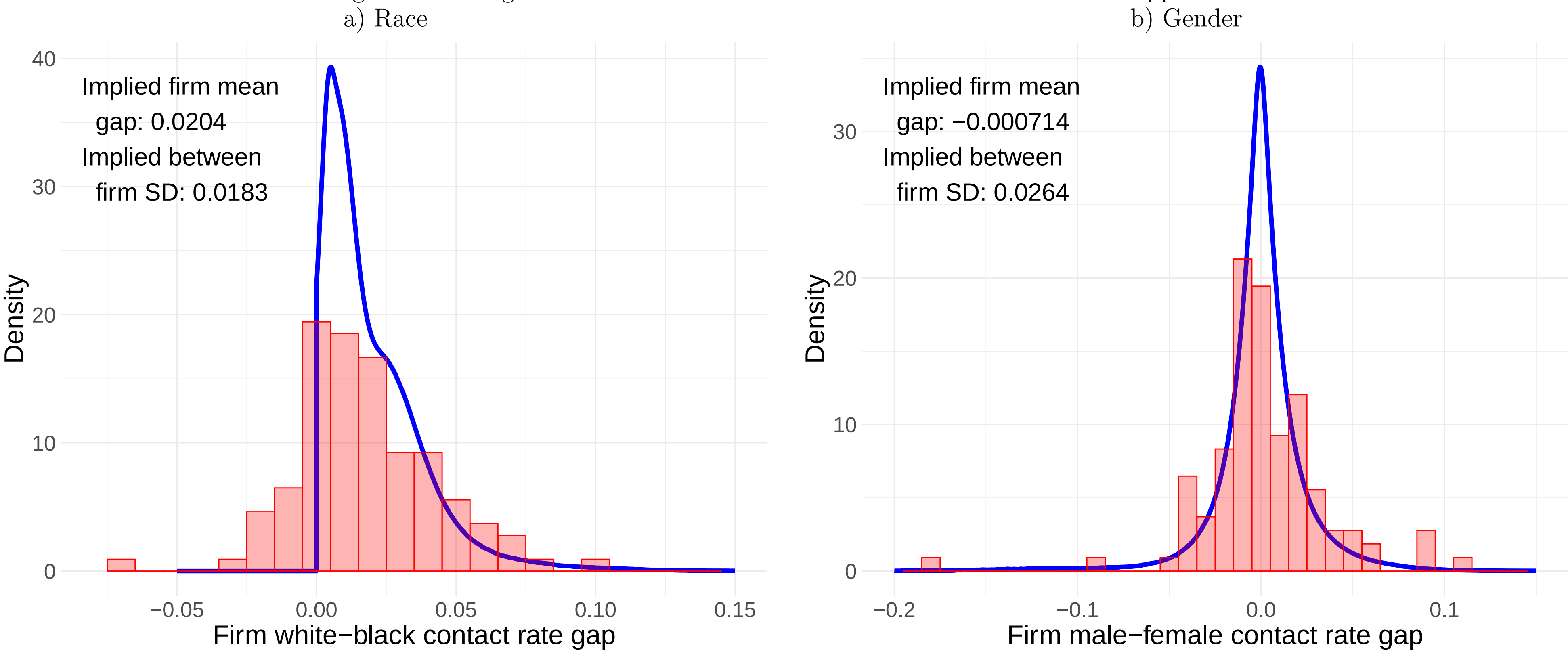
Figure E4: Local deconvolutions of firm-level discrimination distributions

a) Race b) Gender



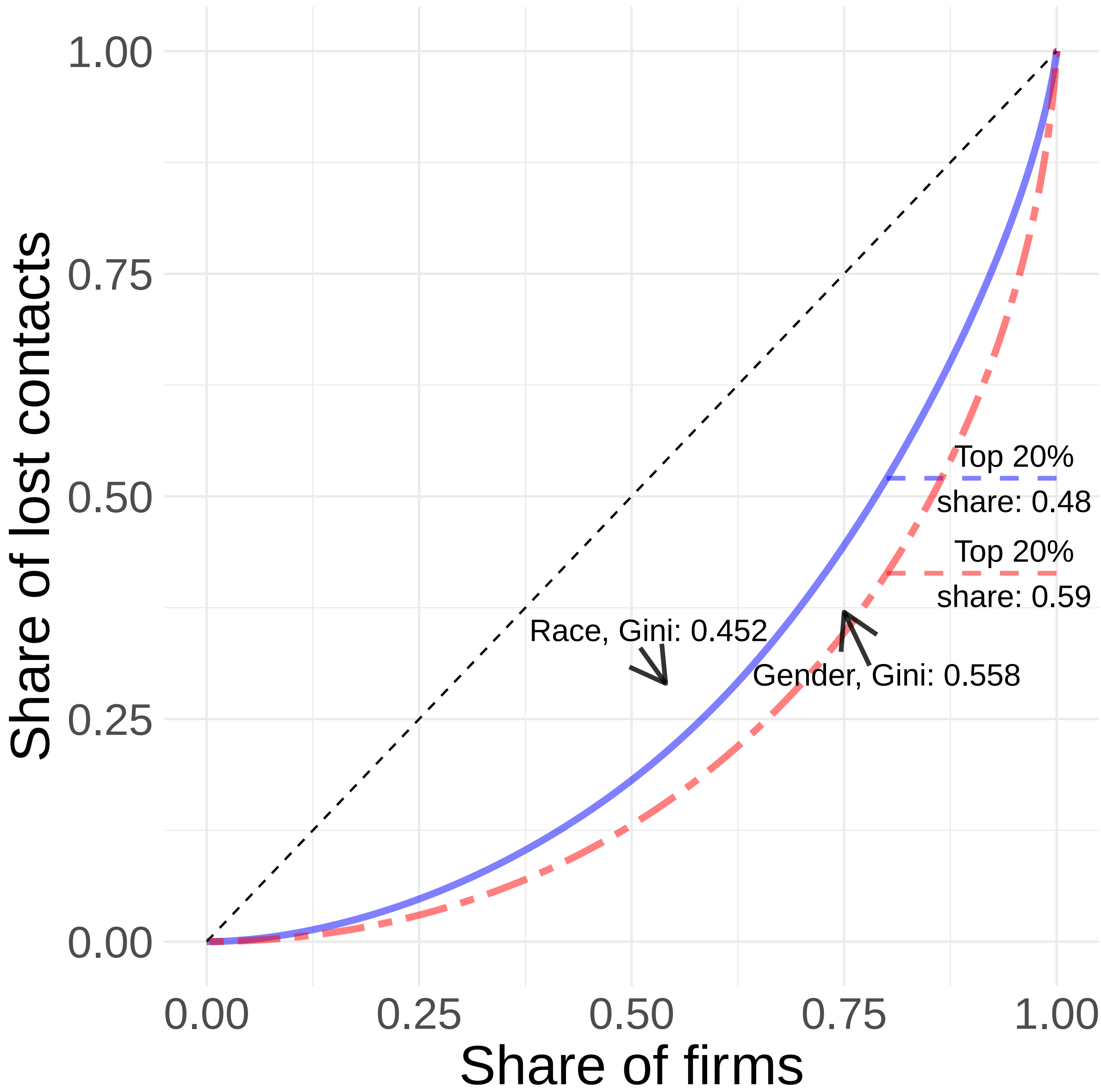
Notes: This figure presents non-parametric estimates of the distribution of firm-specific contact gaps estimated separately for firms with above / below median standard errors. Panel (a) presents estimates for white-Black contact rate differences, and panel (b) presents estimates for male-female differences. Red histograms show the distribution of estimated firm contact gaps in each group. Blue lines show estimates of population contact gap distributions for each group. The population distributions are estimated by applying the deconvolveR package (Narasimhan and Efron, 2020) to firm-specific z-score estimates within group, then numerically integrating over the group's empirical distribution of standard errors. A common penalization parameter is used in the deconvolution step for both groups and calibrated so that the resulting marginal distribution matches the corresponding bias-corrected variance estimate from Table 4. In panel (a), the density of population z-scores is constrained to be weakly positive in each group.

Figure E5: Marginal distributions of firm-level discrimination from local approach



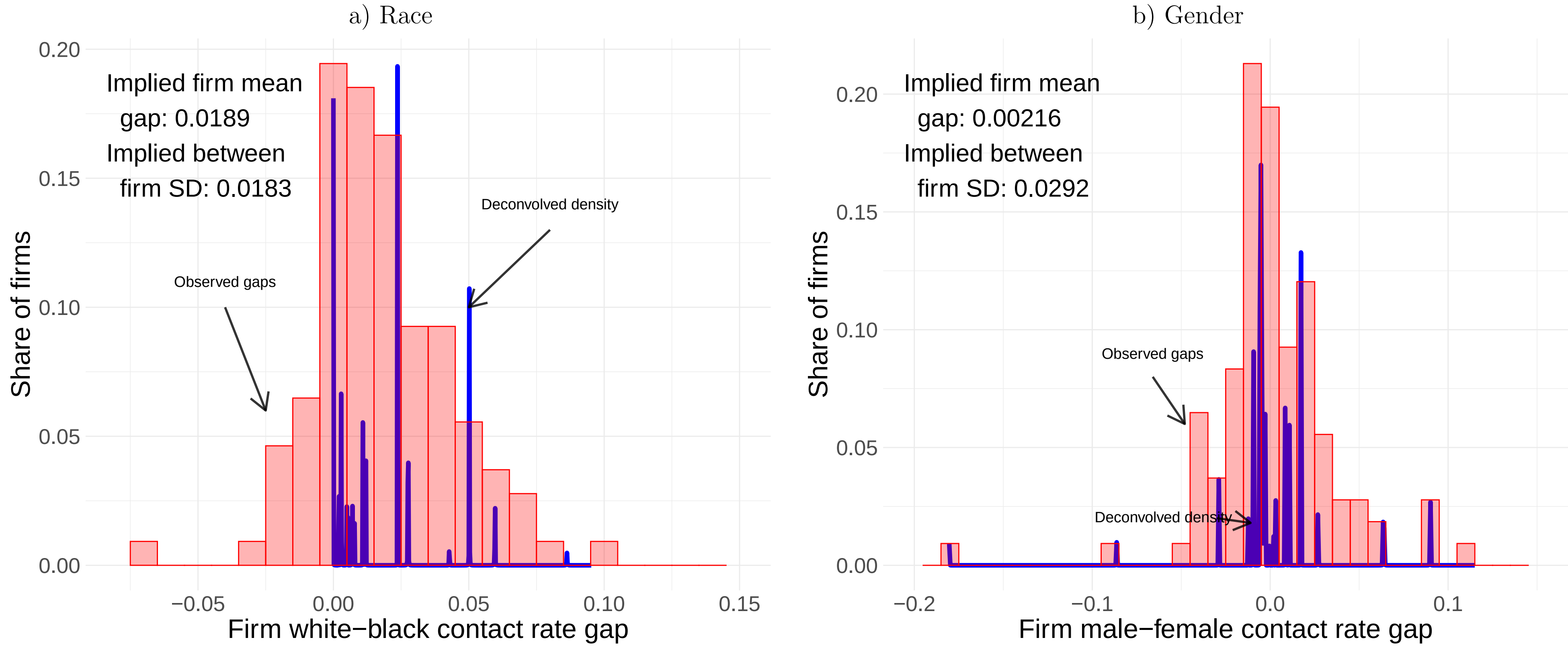
Notes: This figure presents non-parametric estimates of the marginal distribution of firm-specific contact gaps corresponding to the group-specific estimates in Figure E4. Panel (a) presents estimates for white-Black contact rate differences, and panel (b) presents estimates for male-female differences. Red histograms show the distribution of estimated firm contact gaps. Blue lines shows estimates of population contact gap distributions. The marginal density is compute as the average of the group-specific densities in Figure E4.

Figure E6: Local deconvolution Lorenz curves



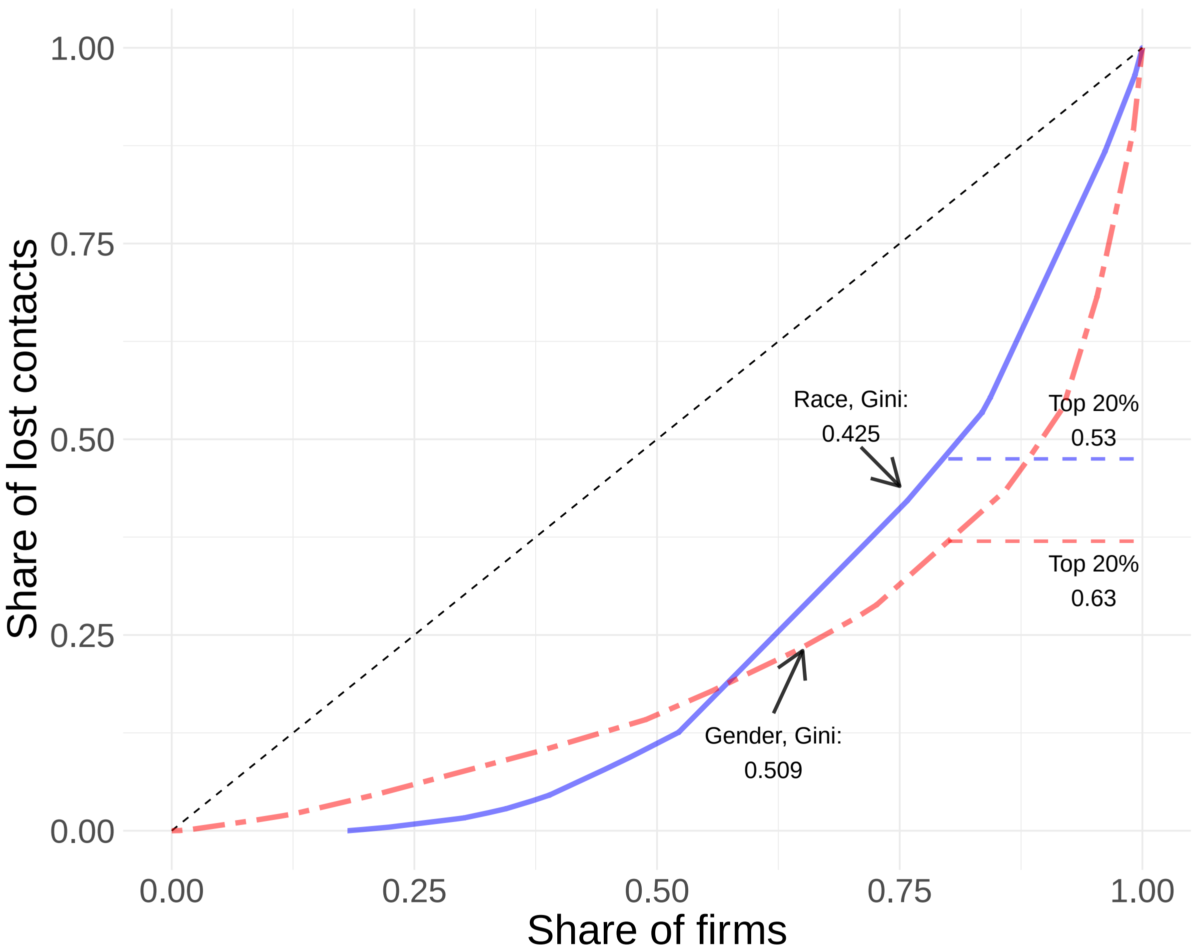
Notes: This figure displays Lorenz curves implied by the non-parametric deconvolution estimates of race and gender contact gap distributions in Figure E5. The solid blue curve is the Lorenz curve for the white/Black contact gap, and the dashed red curve is the Lorenz curve for the absolute value of the male/female contact gap. The Lorenz curve reports the share of lost contacts in the experiment attributable to firms below each contact gap percentile. The share of lost contacts equals the sum of contact gaps at firms below a particular contact gap percentile as a share of the sum of contact gaps across all firms. The dashed line is the 45 degree line. The labels for each curve also report Gini coefficients, equal to 1 minus twice the area under each curve.

Figure E7: NPMLE estimates of marginal distributions of firm-level discrimination



Notes: This figure presents non-parametric maximum likelihood estimates of the distribution of firm-specific contact gaps estimated using the approach in Koenker and Gu (2017). Panel (a) presents estimates for white-Black contact rate differences, where we impose the restriction that all contact gaps are weakly positive, and panel (b) presents estimates for male-female differences. Red histograms show the distribution of estimated firm contact gaps. Blue lines shows estimates of population contact gap distributions. Population distributions are estimated allowing a non-parametric bivariate distribution for the mixing distribution of contact gaps and standard errors. The figures plot the marginal distribution of contact gaps. Since the distribution is discrete, the blue lines plot the probability mass function in below, while the histogram reports the share of sample firms in each bin.

Figure E8: NPMLE Lorenz curves



Notes: This figure displays Lorenz curves implied by the NPMLE estimates of race and gender contact gap distributions reported in Figure E7. The solid blue curve is the Lorenz curve for the white/Black contact gap, and the dashed red curve is the Lorenz curve for the absolute value of the male/female contact gap. The Lorenz curve reports the share of lost contacts in the experiment attributable to firms below each contact gap percentile. The share of lost contacts equals the sum of contact gaps at firms below a particular contact gap percentile as a share of the sum of contact gaps across all firms. Linear interpolation has been used between mass points, which generates kinks in the curve. The dashed line is the 45 degree line. The labels for each curve also report Gini coefficients, equal to 1 minus twice the area under each curve.