# SUPPLEMENT TO "REASONABLE DOUBT: EXPERIMENTAL DETECTION OF JOB-LEVEL EMPLOYMENT DISCRIMINATION"
## (*Econometrica*, Vol. 89, No. 2, March 2021, 765–792)

PATRICK KLINE
Department of Economics, UC Berkeley and NBER

CHRISTOPHER WALTERS
Department of Economics, UC Berkeley and NBER

### APPENDIX A: CONNECTIONS TO LARGE SCALE TESTING

THE PROBLEM OF DETECTING INDIVIDUAL DISCRIMINATORS based on correspondence evidence is closely related to the literature on large scale testing, which is concerned with deciding which hypotheses to reject based upon the results of a very large number of tests (Efron (2012) provided a review). A seminal contribution to this literature comes from Benjamini and Hochberg (1995), who proposed controlling the False Discovery Rate (FDR): the expected share of rejected null hypotheses that are true. We next show that a decision rule based on the posterior probability $\pi(c_w, c_b)$ will control an analogue of the FDR, while a decision rule based on classical hypothesis testing will not.

As in Section 9, let $\delta : \{0, \ldots, L_w\} \times \{0, \ldots, L_b\} \to \{0, 1\}$ represent an auditing rule that maps the evidence vector $(C_{jw}, C_{jb})$ to a binary investigation decision. Letting $N_J \equiv \sum_{j=1}^{J} \delta(C_{jw}, C_{jb})$ denote the total number of investigations in a sample of $J$ jobs, we can define the *Positive False Discovery Rate* (Storey (2003)) as: $p\,\text{FDR}_J = \mathbb{E}[N_J^{-1} \sum_{j=1}^{J} \delta(C_{jw}, C_{jb})(1 - D_j) | N_J \geq 1]$. In words, $p\,\text{FDR}_J$ gives the proportion of investigated jobs that are not discriminating, conditional on at least one investigation taking place. The following lemma establishes that a posterior cutoff decision rule controls $p\,\text{FDR}_J$.

LEMMA A1: *If* $\delta(C_{jw}, C_{jb}) = 1\{\pi(C_{jw}, C_{jb}) > \bar{p}\}$, *then* $p\,\text{FDR}_J \leq 1 - \bar{p}$.

PROOF: Storey (2003, Theorem 1) showed that $p\,\text{FDR}_J = \Pr(D_j = 0 | \delta(C_{jw}, C_{jb}) = 1)$ for any deterministic decision rule $\delta(\cdot)$ obeying $\Pr(\delta(C_{jw}, C_{jb}) = 1) > 0$. Then the posterior cutoff rule $\delta(C_{jw}, C_{jb}) = 1\{\pi(C_{jw}, C_{jb}) > \bar{p}\}$ yields

$$p\,\text{FDR}_J = \Pr\big(D_j = 0 | \pi(C_{jw}, C_{jb}) > \bar{p}\big)$$
$$\leq \Pr\big(D_j = 0 | \pi(C_{jw}, C_{jb}) = \bar{p}\big) = 1 - \bar{p}. \qquad Q.E.D.$$

By contrast, consider an alternative decision rule $\delta^{\dagger}(C_{jw}, C_{jb})$ based on a classical hypothesis test that controls size at a fixed level $\tilde{\alpha} < 1$. To simplify exposition, suppose that the test is pivotal under the null of nondiscrimination so that

$$\Pr\big(\delta^{\dagger}(C_{jw}, C_{jb}) = 1 | p_{jw} = p, p_{jb} = p\big) = \tilde{\alpha}, \quad \forall p \in [0, 1].$$

Patrick Kline: pkline@berkeley.edu
Christopher Walters: crwalters@econ.berkeley.edu

We can write the resulting $p\,\text{FDR}_J$ of this rule:

$$\Pr(D_j = 0 | \delta^\dagger(C_{jw}, C_{jb}) = 1)$$

$$= \frac{\Pr(\delta^\dagger(C_{jw}, C_{jb}) = 1 | D_j = 0)(1 - \bar{\pi})}{\Pr(\delta^\dagger(C_{jw}, C_{jb}) = 1 | D_j = 0)(1 - \bar{\pi}) + \Pr(\delta^\dagger(C_{jw}, C_{jb}) = 1 | D_j = 1)\bar{\pi}}$$

$$\geq \frac{\tilde{\alpha}(1 - \bar{\pi})}{\tilde{\alpha}(1 - \bar{\pi}) + \bar{\pi}}.$$

To see that $\delta^\dagger(C_{jw}, C_{jb})$ fails to control $p\,\text{FDR}_J$, note that $\lim_{\bar{\pi} \downarrow 0} \frac{\tilde{\alpha}(1-\bar{\pi})}{\tilde{\alpha}(1-\bar{\pi})+\bar{\pi}} = 1$. That is, when almost no jobs are discriminating, classical hypothesis testing will result in the vast majority of investigations being false accusations.

The *False Discovery Rate* of Benjamini and Hochberg (1995) can be written $\text{FDR}_J = p\,\text{FDR}_J \times \Pr(N_J \geq 1)$. Because $\Pr(N_J \geq 1) \leq 1$, Lemma A1 implies that the posterior cutoff rule also controls $\text{FDR}_J$.

## APPENDIX B: PROOF OF LEMMA 2

By the law of total probability, the share of jobs calling $c_w$ white and $t - c_w$ black applications among those calling $t$ total can be written

$$\bar{f}_t(c_w) = (1 - \bar{\pi}_t)\bar{f}_t^0(c_w) + \bar{\pi}_t\bar{f}_t^1(c_w),$$

where $\bar{f}_t^d(c_w) = \Pr(C_{jw} = c_w | C_{jw} + C_{jb} = t, D_j = d)$ for $d \in \{0, 1\}$. Since $\bar{f}_t^1(c_w) \in [0, 1]$, we have

$$\bar{f}_t(c_w) \geq (1 - \bar{\pi}_t)\bar{f}_t^0(c_w), \qquad \bar{f}_t(c_w) \leq (1 - \bar{\pi}_t)\bar{f}_t^0(c_w) + \bar{\pi}_t,$$

which implies

$$\bar{\pi}_t \geq \max\left\{ \frac{\bar{f}_t^0(c_w) - \bar{f}_t(c_w)}{\bar{f}_t^0(c_w)}, \frac{\bar{f}_t(c_w) - \bar{f}_t^0(c_w)}{1 - \bar{f}_t^0(c_w)} \right\}.$$

Taking the maximum of these lower bounds over $c_w \in \{0, \ldots, t\}$ yields the bound on $\bar{\pi}_t$ in part (i) of Lemma 2.

By Bayes's rule, the share of discriminators among jobs calling $c_w$ white and $t - c_w$ black applications is given by

$$\pi(c_w, t - c_w) = 1 - \frac{\bar{f}_t^0(c_w)(1 - \bar{\pi}_t)}{\bar{f}_t(c_w)}.$$

Plugging the bound on $\bar{\pi}_t$ from part (i) of the lemma into this expression gives the bound on $\pi(c_w, t - c_w)$ in part (ii).

## APPENDIX C: DISCRETIZATION OF $G$ AND LINEAR PROGRAMMING BOUNDS

To compute the solution to the problem in (5), we approximate the CDF $G(p_w, p_b)$ with the discrete distribution

$$G_K(p_w, p_b) = \sum_{k=1}^{K} \sum_{s=1}^{K} \eta_{ks} 1\{p_w \leq \varrho(k, s), p_b \leq \varrho(s, k)\},$$

where the $\{\eta_{ks}\}_{k=1,s=1}^{K,K}$ are probability masses and $\{\varrho(k,s), \varrho(s,k)\}_{k=1,s=1}^{K,K}$ comprise a set of mass point coordinates generated by the function

$$\varrho(k,s) = \underbrace{\frac{\min\{k,s\}-1}{K}}_{\text{diagonal}} + \underbrace{\frac{\max\{0, k-s\}^2}{K(1+K-y)}}_{\text{off-diagonal}}.$$

This discretization scheme can be visualized as a two-dimensional grid containing $K^2$ elements. The diagonal entries on the grid represent jobs where no discrimination is present. The first term above ensures the mass points are equally spaced along the diagonal from $(0,0)$ to $(\frac{K-1}{K}, \frac{K-1}{K})$. The second term spaces off diagonal points quadratically according to their distance from the diagonal in order to accommodate jobs with very low levels of discrimination while economizing on the number of grid points. We use a spacing scheme that places more points near the diagonal because we are particularly interested in the mass exactly on the diagonal. Note that $\lim_{K\to\infty} \varrho(K,s) = 1$, ensuring the grid asymptotically spans the unit square.

With this notation, the constraints in (6) can be written

$$\bar{f}(c_w, c_b) = \binom{L_w}{c_w}\binom{L_b}{c_b} \sum_{k=1}^{K}\sum_{s=1}^{K} \eta_{ks}\varrho(k,s)^{c_w}\big(1-\varrho(k,s)\big)^{L_w-c_w}$$
$$\times\, \varrho(s,k)^{c_b}\big(1-\varrho(s,k)\big)^{L_b-c_b}, \tag{1}$$

for $c_w = (1, \ldots, L_w)$ and $c_b = (0, \ldots, L_b)$. Hence, our composite discretized optimization problem is to

$$\min_{\{\eta_{ks}\}} 1 - \frac{\binom{L}{t}}{\sum\limits_{c_w'=0}^{t} \bar{f}(c_w', t-c_w')} \sum_{k=1}^{K} \eta_{kk}\varrho(k,k)^t\big(1-\varrho(k,k)\big)^{L-t},$$

subject to (1) and

$$\sum_{k=1}^{K}\sum_{s=1}^{K} \eta_{ks} = 1, \quad \eta_{ks} \geq 0,$$

for $k = 1, \ldots, K$ and $m = 1, \ldots, K$. We solve this problem numerically using the Gurobi software package. Because setting $K$ too low will tend to yield artificially tight bounds, we set $K = 900$ in all bound computation steps, which yields $(900)^2 = 810,000$ distinct mass points.

Appendix Table A.IV reports linear programming bounds for various choices of $K$. As expected, the bounds stabilize with a sufficiently large $K$, and the quadratic spacing described above produces more accurate results than an equally-spaced grid: we obtain similar estimates for a quadratic grid with $300^2$ grid points and a rectangular grid with $900^2$ points.

## APPENDIX D: SHAPE-CONSTRAINED GMM

To accommodate the Nunley, Pugh, Romero, and Seals (2015) study which employs multiple application designs, we introduce the variable $L_j = (L_{jw}, L_{jb})$ which gives the

number of white and black applications sent to job $j$. Collecting the design-specific call-back probabilities $\{\Pr(C_{jw} = c_w, C_{jb} = c_b | L_j = l)\}_{c_w, c_b}$ into the vector $\bar{f}_l$, our model relates these probabilities to moments of the callback distribution via the linear system $\bar{f}_l = B_l \mu$, for $B_l$ a fixed matrix of binomial coefficients. Letting $\bar{f}$ denote the vector formed by "stacking" the $\{\bar{f}_l\}$ across designs in an experiment, we write $\bar{f} = B\mu$. Let $\eta$ be a $K^2 \times 1$ vector composed of the probability masses $\{\eta_{ks}\}_{k=1, s=1}^{K, K}$ (see Appendix C). For GMM estimation, we set $K = 150$ (larger values yield very similar results). From (3), we can write $\mu = M\eta$ where $M$ is a $\dim(\mu) \times K^2$ matrix composed of entries with typical element $\varrho(k, s)^m \varrho(s, k)^n$. We then have the moment restriction $\bar{f} = BM\eta$.

Let $\tilde{f}$ denote the vector of empirical callback probabilities with typical element

$$\frac{J^{-1} \sum_{j=1}^{J} 1\{C_{jw} = c_w, C_{jb} = c_b, L_j = l\}}{J^{-1} \sum_{j=1}^{J} 1\{L_j = l\}}.$$

Our shape-constrained GMM estimator of $\eta$ can be written as the solution to the following quadratic programming problem:

$$\hat{\eta} = \arg\inf_{\eta} (\tilde{f} - BM\eta)' W (\tilde{f} - BM\eta)$$

$$\text{s.t. } \eta \geq 0, \mathbf{1}'\eta = 1, \tag{2}$$

where $W$ is a fixed weighting matrix. Note that because $G$ is not identified, there are many possible solutions $\hat{\eta}$ to this problem, but these solutions will all yield the same values of $BM\hat{\eta}$. Our shape-constrained estimate of the moments is $\hat{\mu} = M\hat{\eta}$, while our estimator of the callback probabilities is $\hat{\bar{f}} = BM\hat{\eta}$. We follow a two-step procedure, solving (2) with diagonal weights proportional to the number of jobs used in the application design and then choosing $W = \hat{\Sigma}^{-1}$, where $\hat{\Sigma} = \text{diag}(\hat{\bar{f}}^{(1)}) - \hat{\bar{f}}^{(1)} \hat{\bar{f}}^{(1)\prime}$ is an estimate of the variance-covariance matrix of the callback frequencies implied by the first-step shape-constrained callback probability estimates $\hat{\bar{f}}^{(1)}$.

### Hong and Li (2020) Standard Errors

Standard errors on the moment estimates $\hat{\mu}$ are computed via the numerical bootstrap procedure of Hong and Li (2020) using a step size of $J^{-1/3}$ (we found qualitatively similar results with a step size of $J^{-1/4}$). Our implementation of the numerical bootstrap proceeds as follows: the bootstrap analogue $\mu^*$ of $\hat{\mu}$ solves the quadratic programming problem in (2) where $\tilde{f}$ has been replaced by $(\tilde{f} + J^{-1/3} f^*)$. The bootstrap probabilities $f^*$ have typical element

$$J^{1/2} \left( \frac{\sum_{j=1}^{J} \omega_j^* 1\{C_{jw} = c_w, C_{jb} = c_b, L_j = l\}}{\sum_{j=1}^{J} \omega_j^* 1\{L_j = l\}} - \frac{\sum_{j=1}^{J} 1\{C_{jw} = c_w, C_{jb} = c_b, L_j = l\}}{\sum_{j=1}^{J} 1\{L_j = l\}} \right),$$

where $\{\omega_j^*\}_{j=1}^J$ are a set of i.i.d. draws from an exponential distribution with mean and variance 1. For any function $\phi(\hat{\mu})$ of the moment estimates $\hat{\mu}$ reported, we use as our standard error estimate the standard deviation across bootstrap replications of $J^{-1/3}[\phi(\mu^*) - \phi(\hat{\mu})]$.

### Chernozhukov, Newey, and Santos (2015) Goodness-of-Fit Test

To formally test whether there exists an $\eta$ in the $K^2$-dimensional probability simplex such that $f = BM\eta$ holds, we rely on the procedure of Chernozhukov, Newey, and Santos (2015). Our test statistic (the "$J$-test") can be written

$$T_n = \inf_{\eta}(\tilde{f} - BM\eta)'\hat{\Sigma}^{-1}(\tilde{f} - BM\eta)$$

$$\text{s.t. } \eta \geq 0, \mathbf{1}'\eta = 1.$$

Letting $\mathbb{F}^* = f^* - \tilde{f}$ denote the (centered) bootstrap analogue of the callback frequencies $\tilde{f}$ and $W^*$ a corresponding bootstrap weighting matrix, our bootstrap test statistic takes the form

$$T_n^* = \inf_{\eta,h}\left(\mathbb{F}^* - BMh\right)'W^*\left(\mathbb{F}^* - BMh\right)$$

$$\text{s.t. } (\tilde{f} - BM\eta)'W(\tilde{f} - BM\eta) = T_n, \eta \geq 0, \mathbf{1}'\eta = 1, h \geq -\eta, \mathbf{1}'h = 0. \qquad (3)$$

As in the full sample problem, we conduct a two-step GMM procedure in each bootstrap replication, setting $W^* = [\text{diag}(BM\eta^{(1)*}) - (BM\eta^{(1)*})(BM\eta^{(1)*})']^{-1}$, where $\eta^{(1)*}$ is a first-step diagonally weighted estimate of the probabilities in the bootstrap sample. The goodness-of-fit $p$-value reported is the share of bootstrap samples for which $T_n^* > T_n$.

To simplify computation of (3), we reformulate the problem in two ways. First, we define primary and auxiliary vectors of errors for each moment condition. Letting $\xi_h = \mathbb{F}^* - BMh$ and $\xi_\eta = \tilde{f} - BM\eta$, the problem can be re-posed as

$$T_n^* = \inf_{\xi_h,\xi_\eta} \xi_h'W^*\xi_h,$$

$$\text{s.t. } \xi_\eta'W\xi_\eta = T_n, BMh + \xi_h = \mathbb{F}^*, BM\eta + \xi_\eta = \tilde{f}, \mathbf{1}'h = 0, \mathbf{1}'\eta = 1, h \geq -\eta, \eta \geq 0.$$

Now letting $h^+ = h + \eta$, we can further rewrite the problem as

$$T_n^* = \inf_{\xi_h,\xi_\eta} \xi_h'W^*\xi_h,$$

$$\text{s.t. } \xi_\eta'W\xi_\eta = T_n, BMh^+ + \xi_h + \xi_\eta = \mathbb{F}^*, BM\eta + \xi_\eta = \tilde{f},$$

$$\mathbf{1}'h^+ = 1, \mathbf{1}'\eta = 1, h^+ \geq 0, \eta \geq 0.$$

This final representation replaces a $K^2 \times K^2 + 1$ (inequality) constraint matrix encoding $\xi_h \geq -\xi_\eta$ and $\xi_\eta \geq 0$ with a $2K^2 \times 1$ vector encoding $h^+ \geq 0$ and $\eta \geq 0$. Because this problem still involves a quadratic constraint in $\xi_\eta$, we make use of Gurobi's Second Order Cone Programming (SOCP) solver to obtain a solution.

## APPENDIX E: COMPUTING MAXIMUM RISK

We approximate $G(p_w^H, p_w^L, p_b^H, p_b^L)$ with the discretized distribution

$$G_K(p_w^H, p_w^L, p_b^H, p_b^L) = \sum_{k=1}^{K} \sum_{s=1}^{K} \sum_{k'=1}^{K} \sum_{s'=1}^{K} \eta_{ksk's'} 1\{p_w^H \leq \varrho(k,s), p_w^L \leq \varrho(k',s'), p_b^H \leq \varrho(s,k),$$
$$p_b^L \leq \varrho(s',k')\},$$

which has $K^4$ mass points. In practice, we choose $K = 30$, which yields the same number of points as the approximation described in Appendix C.

Generalizing the notation of Appendix D, let the vector $L_j = (L_{jw}^H, L_{jw}^L, L_{jb}^H, L_{jb}^L)$ record the number of high quality and low quality applications of each race sent to job $j$ and let $C_j = (C_{jw}^H, C_{jw}^L, C_{jb}^H, C_{jb}^L)$ record the corresponding numbers of callbacks. The space of auditing rules we consider is of the form $\delta(C_j, L_j, q) = 1\{\mathcal{P}(C_j, L_j, G_{logit}) > q\}$. With this notation, we can write the risk function

$$\mathcal{R}(q) = \sum_{l \in \mathscr{A}_1} w_l \mathbb{E}\left[\delta(C_j, l, q)\left\{\kappa - \Lambda\left(\sum_{x \in \{H,L\}} \frac{\Lambda^{-1}(p_{wj}^x) - \Lambda^{-1}(p_{bj}^x)}{2}\right)\right\}\bigg| L_j = l\right],$$

where $\mathscr{A}_1$ is the set of all $2^5 = 36$ binary quality permutations possible in a design with 5 white and 5 black applications and $w_l = \binom{5}{l_w^H}\binom{5}{l_b^H}(1/2)^{10}$ is the set of weights that arise when quality is assigned at random within race.

To further evaluate the above risk expression, we can write

$$\mathbb{E}\left[\delta(C_j, l, q)\left\{\kappa - \Lambda\left(\sum_{x \in \{H,L\}} \frac{\Lambda^{-1}(p_{wj}^x) - \Lambda^{-1}(p_{bj}^x)}{2}\right)\right\}\bigg| L_j = l\right]$$

$$= \sum_{c_w^H=0}^{a_{jw}^H} \sum_{c_b^H=0}^{a_{jb}^H} \sum_{c_w^L=0}^{a_{jw}^L} \sum_{c_b^L=0}^{a_{jb}^L} \sum_{k=1}^{K} \sum_{s=1}^{K} \sum_{k'\geq k}^{K} \sum_{s'\geq s}^{K} \delta(c, l, q) \eta_{ksk's'} \binom{l_w^H}{c_w^H}\binom{l_b^H}{c_b^H}\binom{l_w^L}{c_w^L}\binom{l_b^L}{c_b^L}$$

$$\times \varrho(k,s)^{c_w^H}(1-\varrho(k,s))^{l_w^H - c_w^H} \varrho(s,k)^{c_b^H}(1-\varrho(s,k))^{l_b^H - c_b^H}$$

$$\times \varrho(k',s')^{c_w^L}(1-\varrho(k',l=s'))^{l_w^L - c_w^L} \varrho(s',k')^{c_b^L}(1-\varrho(s',k'))^{l_b^L - c_b^L}$$

$$\times \left\{\kappa - \Lambda\left(\frac{\Lambda^{-1}(\varrho(k,s)) - \Lambda^{-1}(\varrho(s,k))}{2} + \frac{\Lambda^{-1}(\varrho(k',s')) - \Lambda^{-1}(\varrho(s',k'))}{2}\right)\right\}.$$

Using this expression, maximal risk can therefore be written as the solution to the following linear programming problem:

$$\mathcal{R}^m(q) = \max_{\{\eta_{ksk's'}\}} \sum_{l \in \mathscr{A}_1} w_l \mathbb{E}\left[\delta(C_j, l, q)\left\{\kappa - \Lambda\left(\sum_{x \in \{H,L\}} \frac{\Lambda^{-1}(p_{wj}^x) - \Lambda^{-1}(p_{bj}^x)}{2}\right)\right\}\bigg| L_j = l\right]$$

subject to the constraint that the $\eta_{klk'l'}$ are nonnegative and sum to 1 and that the following moment restrictions hold:

$$\Pr(C_j = c | L_j = l) = \binom{l_w^H}{c_w^H}\binom{l_b^H}{c_b^H}\binom{l_w^L}{c_w^L}\binom{l_b^L}{c_b^L} \sum_{k=1}^{K}\sum_{s=1}^{K}\sum_{k'=1}^{K}\sum_{s'=1}^{K}\eta_{ksk's'}$$

$$\times \varrho(k, s)^{c_w^H}\left(1 - \varrho(k, s)\right)^{l_w^H - c_w^H}\varrho(s, k)^{c_b^H}\left(1 - \varrho(s, k)\right)^{l_b^H - c_b^H}$$

$$\times \varrho(k', s')^{c_w^L}\left(1 - \varrho(k', s')\right)^{l_w^L - c_w^L}\varrho(s', k')^{c_b^L}\left(1 - \varrho(s', k')\right)^{l_b^L - c_b^L}.$$

We impose these restrictions for the following set of designs, all of which are present in the Nunley et al. (2015) experiment: $\mathscr{A}_2 = \{(2, 0, 2, 0), (2, 0, 0, 2), (0, 2, 2, 0), (0, 2, 0, 2)\}$. To operationalize these constraints, we replace the unknown cell probabilities $\Pr(C_j = c | L_j = l)$ for all $c$ and $l$ in $\mathscr{A}_2$ with their predictions under the logit model reported in column (2) of Table V. Using the logit predictions serves as a form of smoothing that allows us to avoid problems that arise with small cells when considering quality variation due to covariates.

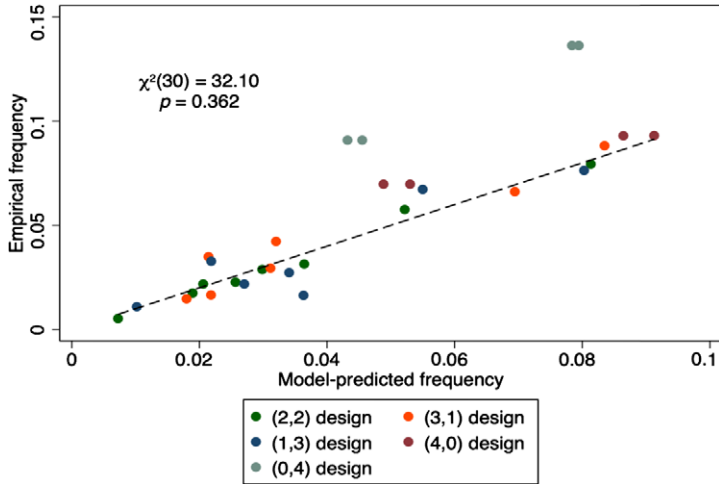## APPENDIX: FIGURES AND TABLES



FIGURE A.1.—Mixed logit model fit. *Notes*: This figure compares mixed logit predicted frequencies for callback events in the Nunley et al. (2015) data with corresponding empirical frequencies. The horizontal axis plots model-predicted probabilities for each possible combination of white and black callback counts (excluding zero total callbacks), separately by experimental design. Model predictions are calculated by simulating the logit model in column (2) of Table V 10,000 times for each job in the Nunley et al. data set. The vertical axis plots the observed frequency of each event. Green dots show frequencies for a design with two white and two black applications, while orange, blue, red, and gray points show frequencies for designs with 3 white and 1 black, 1 white and 3 black, 4 white and zero black, and 0 white and 4 black applications, respectively. The dashed line is the 45-degree line. The test statistic and *p*-value come from a Wald test that all model-predicted and empirical frequencies match, treating the model predictions as fixed.
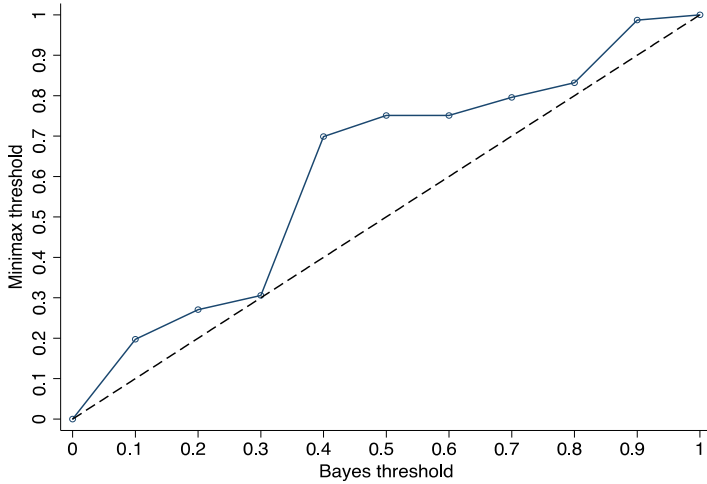
FIGURE A.2.—Bayes and minimax investigation thresholds. *Notes*: This figure compares Bayes and minimax decisions for various values of the investigation cost parameter $\kappa$. The horizontal axis displays the posterior investigation threshold for a Bayes regulator for each value of $\kappa$, and the vertical axis shows the corresponding threshold for a minimax regulator. The dashed line is the 45-degree line.

TABLE A.I

MOMENTS OF CALLBACK RATE DISTRIBUTION, BM DATA[a]

| Moment | No Constraints (1) | Shape Constraints (2) |
|---|---|---|
| $E[p_w]$ | 0.094 | 0.094 |
|  | (0.006) | (0.007) |
| $E[p_b]$ | 0.063 | 0.063 |
|  | (0.006) | (0.006) |
| $E[(p_w - E[p_w])^2]$ | 0.040 | 0.040 |
|  | (0.005) | (0.005) |
| $E[(p_b - E[p_b])^2]$ | 0.023 | 0.023 |
|  | (0.004) | (0.004) |
| $E[(p_w - E[p_w])(p_b - E[p_b])]$ | 0.028 | 0.028 |
|  | (0.004) | (0.003) |
| $E[(p_w - E[p_w])^2(p_b - E[p_b])]$ | 0.015 | 0.014 |
|  | (0.003) | (0.002) |
| $E[(p_w - E[p_w])(p_b - E[p_b])^2]$ | 0.023 | 0.012 |
|  | (0.003) | (0.002) |
| $E[(p_w - E[p_w])^2(p_b - E[p_b])^2]$ | 0.010 | 0.010 |
|  | (0.003) | (0.002) |
| $J$-statistic: |  | 0.0 |
| $p$-value: |  | 1.00 |
| Sample size | 1112 | |

[a]This table reports generalized method of moments (GMM) estimates of moments of the joint distribution of job-specific white and black callback rates in the Bertrand and Mullainathan (2004) data. Estimates in column (2) come from a shape-constrained GMM procedure imposing that the moments are consistent with a well-defined probability distribution. The $J$-statistic is the minimized shape-constrained GMM criterion function. The $p$-value comes from a bootstrap test of the hypothesis that the model restrictions are satisfied.

TABLE A.II

MOMENTS OF CALLBACK RATE DISTRIBUTION, NPRS DATA[a]

| Moment | Design-Specific Estimates | | | | Combined Estimates |
|---|---|---|---|---|---|
| | (2, 2) Design | (3, 1) Design | (1, 3) Design | P-Value | |
| | (1) | (2) | (3) | (4) | (5) |
| $E[p_w]$ | 0.174 | 0.199 | 0.142 | 0.027 | 0.176 |
| | (0.010) | (0.025) | (0.015) | | (0.007) |
| $E[p_b]$ | 0.148 | 0.149 | 0.157 | 0.854 | 0.153 |
| | (0.010) | (0.015) | (0.013) | | (0.007) |
| $E[(p_w - E[p_w])^2]$ | 0.089 | 0.108 | – | 0.097 | 0.095 |
| | (0.007) | (0.009) | | | (0.005) |
| $E[(p_b - E[p_b])^2]$ | 0.085 | – | 0.083 | 0.857 | 0.084 |
| | (0.007) | | (0.008) | | (0.005) |
| $E[(p_w - E[p_w])(p_b - E[p_b])]$ | 0.083 | 0.084 | 0.080 | 0.926 | 0.084 |
| | (0.006) | (0.009) | (0.009) | | (0.004) |
| $E[(p_w - E[p_w])^3]$ | – | 0.051 | – | | 0.106 |
| | | (0.008) | | | (0.007) |
| $E[(p_b - E[p_b])^3]$ | – | – | 0.044 | | 0.091 |
| | | | (0.007) | | (0.006) |
| $E[(p_w - E[p_w])^2(p_b - E[p_b])]$ | 0.044 | 0.043 | – | 0.875 | 0.040 |
| | (0.004) | (0.007) | | | (0.002) |
| $E[(p_w - E[p_w])(p_b - E[p_b])^2]$ | 0.047 | – | 0.045 | 0.819 | 0.042 |
| | (0.005) | | (0.007) | | (0.002) |
| $E[(p_w - E[p_w])^3(p_b - E[p_b])]$ | – | 0.034 | – | – | 0.035 |
| | | (0.005) | | | (0.002) |
| $E[(p_w - E[p_w])(p_b - E[p_b])^3]$ | – | – | 0.037 | – | 0.037 |
| | | | (0.006) | | (0.002) |
| $E[(p_w - E[p_w])^2(p_b - E[p_b])^2]$ | 0.036 | – | – | – | 0.037 |
| | (0.004) | | | | (0.002) |
| | | | | J-statistic: | 23.0 |
| | | | | p-value: | 0.226 |
| Sample size | 1146 | 544 | 550 | | 2240 |

[a]This table reports generalized method of moments (GMM) estimates of moments of the joint distribution of job-specific white and black callback rates in the Nunley et al. (2015) data. Columns (1), (2), and (3) show estimates based on jobs that received 2 white and 2 black, 3 white and 1 black, and 1 white and 3 black applications, respectively. Column (4) shows p-values from tests that the moments are the same in each design. Estimates in column (5) come from a shape-constrained GMM procedure imposing that the moments are consistent with a well-defined probability distribution. The J-statistic is the minimized shape-constrained GMM criterion function. The p-value come from a bootstrap test of the hypothesis that the model restrictions are satisfied.

TABLE A.III

MOMENTS OF CALLBACK RATE DISTRIBUTION, AGCV DATA[a]

| Moment | No Constraints (1) | Shape Constraints (2) | Moment | No Constraints (3) | Shape Constraints (4) |
|---|---|---|---|---|---|
| $E[p_f]$ | 0.136 (0.010) | 0.137 (0.010) | $E[(p_f - E[p_f])^4]$ | 0.024 (0.004) | 0.026 (0.003) |
| $E[p_m]$ | 0.103 (0.009) | 0.109 (0.009) | $E[(p_m - E[p_m])^4]$ | 0.019 (0.004) | 0.023 (0.003) |
| $E[(p_f - E[p_f])^2]$ | 0.066 (0.006) | 0.066 (0.006) | $E[(p_f - E[p_f])^4(p_m - E[p_m])]$ | 0.012 (0.003) | 0.012 (0.002) |
| $E[(p_m - E[p_m])^2]$ | 0.047 (0.005) | 0.052 (0.006) | $E[(p_f - E[p_f])(p_m - E[p_m])^4]$ | 0.013 (0.003) | 0.013 (0.002) |
| $E[(p_f - E[p_f])(p_m - E[p_m])]$ | 0.043 (0.005) | 0.043 (0.004) | $E[(p_f - E[p_f])^3(p_m - E[p_m])^2]$ | 0.012 (0.003) | 0.011 (0.002) |
| $E[(p_f - E[p_f])^3]$ | 0.032 (0.005) | 0.064 (0.007) | $E[(p_f - E[p_f])^2(p_m - E[p_m])^3]$ | 0.012 (0.003) | 0.013 (0.002) |
| $E[(p_m - E[p_m])^3]$ | 0.025 (0.005) | 0.048 (0.007) | $E[(p_f - E[p_f])^4(p_m - E[p_m])^2]$ | 0.010 (0.002) | 0.010 (0.002) |
| $E[(p_f - E[p_f])^2(p_m - E[p_m])]$ | 0.021 (0.004) | 0.018 (0.003) | $E[(p_f - E[p_f])^2(p_m - E[p_m])^4]$ | 0.010 (0.002) | 0.010 (0.002) |
| $E[(p_f - E[p_f])(p_m - E[p_m])^2]$ | 0.022 (0.004) | 0.020 (0.003) | $E[(p_f - E[p_f])^3(p_m - E[p_m])^3]$ | 0.010 (0.002) | 0.009 (0.002) |
| $E[(p_f - E[p_f])^3(p_m - E[p_m])]$ | 0.015 (0.003) | 0.015 (0.002) | $E[(p_f - E[p_f])^4(p_m - E[p_m])^3]$ | 0.008 (0.002) | 0.008 (0.001) |
| $E[(p_f - E[p_f])(p_m - E[p_m])^3]$ | 0.016 (0.003) | 0.017 (0.002) | $E[(p_f - E[p_f])^3(p_m - E[p_m])^4]$ | 0.008 (0.002) | 0.008 (0.002) |
| $E[(p_f - E[p_f])^2(p_m - E[p_m])^2]$ | 0.016 (0.003) | 0.016 (0.002) | $E[(p_f - E[p_f])^4(p_m - E[p_m])^4]$ | 0.007 (0.002) | 0.001 (0.001) |
| | | | $J$-statistic: | 2.7 | |
| | | | $p$-value: | 0.891 | |
| | | | Sample size: | 799 | |

[a]This table reports generalized method of moments (GMM) estimates of moments of the joint distribution of job-specific white and black callback rates in the Arceo-Gomez and Campos-Vasques (2014) data. Estimates in columns (2) and (4) come from a shape-constrained GMM procedure imposing that the moments are consistent with a well-defined probability distribution. The $J$-statistic is the minimized shape-constrained GMM criterion function. The $p$-value comes from a bootstrap test of the hypothesis that the model restrictions are satisfied.

TABLE A.IV

SENSITIVITY OF MOMENTS AND BOUNDS TO DISCRETIZATION GRID[a]

| | | Grid Spacing | J-Statistic (1) | Share Discriminating (2) | Share Disc., One Call (3) | Share Disc., Two Calls (4) | Share Disc., Three Calls (5) |
|---|---|---|---|---|---|---|---|
| | | | | *A. Nunley et al. data* | | | |
| $K_1 = 50$ | $K_2 = 200$ | Quadratic | 23.00 | 0.267 | 0.485 | 0.553 | 0.680 |
| $K_1 = 100$ | $K_2 = 200$ | Quadratic | 22.97 | 0.369 | 0.701 | 0.716 | 0.845 |
| | $K_2 = 400$ | Quadratic | | 0.332 | 0.619 | 0.652 | 0.788 |
| | $K_2 = 600$ | Quadratic | | 0.322 | 0.597 | 0.636 | 0.769 |
| $K_1 = 150$ | $K_2 = 300$ | Quadratic | 22.96 | 0.404 | 0.771 | 0.778 | 0.884 |
| | $K_2 = 600$ | Quadratic | | 0.371 | 0.699 | 0.717 | 0.840 |
| | **$K_2 = 900$** | **Quadratic** | | **0.361** | **0.679** | **0.700** | **0.826** |
| | | Rectangular | | 0.412 | 0.787 | 0.793 | 0.890 |
| | | | | *B. Arceo-Gomez & Campos-Vasquez data* | | | |
| $K_1 = 50$ | $K_2 = 200$ | Quadratic | 3.5 | 0.220 | 0.762 | 0.709 | 0.570 |
| $K_1 = 100$ | $K_2 = 200$ | Quadratic | 2.8 | 0.218 | 0.738 | 0.717 | 0.599 |
| | $K_2 = 400$ | Quadratic | | 0.209 | 0.732 | 0.710 | 0.583 |
| | $K_2 = 600$ | Quadratic | | 0.209 | 0.730 | 0.708 | 0.579 |
| $K_1 = 150$ | $K_2 = 300$ | Quadratic | 2.7 | 0.220 | 0.727 | 0.718 | 0.606 |
| | $K_2 = 600$ | Quadratic | | 0.208 | 0.722 | 0.709 | 0.587 |
| | **$K_2 = 900$** | **Quadratic** | | **0.207** | **0.721** | **0.708** | **0.584** |
| | | Rectangular | | 0.215 | 0.713 | 0.707 | 0.590 |

[a]This table explores the sensitivity of our shape-constrained generalized method of moments (SCGMM) and linear programming bounds results to the number of grid points used to approximate the joint distribution of callback probabilities. K1 refers to the number of mass points used in the quadratic programming SCGMM step, while K2 refers to the number of mass points used in the linear programming bounds step. Quadratic grid spacing refers to the scheme described in Appendix A, and rectangular spacing refers to a grid with equally spaced points. Column (1) shows the minimized SCGMM criterion function for each value of K1. Column (2) displays the lower bound on the fraction of discriminating jobs for each combination of K1 and K2. Columns (3)–(5) show corresponding bounds conditional on the total number of callbacks. Panel A displays results for an application design with two white and two black applicants in the Nunley et al. (2015) data, and panel B displays results for the Arceo-Gomez and Campos-Vasquez (2014) data. Bold lines indicate the preferred specification used in the main text.

## REFERENCES

ARCEO-GOMEZ, E. O., AND R. M. CAMPOS-VASQUEZ (2014): "Race and Marriage in the Labor Market: A Discrimination Correspondence Study in a Developing Country," *American Economic Review: Papers & Proceedings*, 104, 376–380. [10,11]

BENJAMINI, Y., AND Y. HOCHBERG (1995): "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society*, 57, 289–300. [1,2]

BERTRAND, M., AND S. MULLAINATHAN (2004): "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *American Economic Review*, 94, 991–1013. [8]

CHERNOZHUKOV, V., W. K. NEWEY, AND A. SANTOS (2015): "Constrained Conditional Moment Restriction Models,". arXiv:1509.06311. ArXiv preprint. [5]

EFRON, B. (2012): *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*, Vol. 1. Cambridge University Press. [1]

HONG, H., AND J. LI (2020): "The Numerical Bootstrap," *Annals of Statistics*, 48, 397–412. [4]

NUNLEY, J. M., A. PUGH, N. ROMERO, AND R. A. SEALS (2015): "Racial Discrimination in the Labor Market for Recent College Graduates: Evidence From a Field Experiment," *B.E. Journal of Economic Analysis and Policy*, 15, 1093–1125. [3,7,9,11]

STOREY, J. D. (2003): "The Positive False Discovery Rate: A Bayesian Interpretation and the q-Value," *The Annals of Statistics*, 31, 2013–2035. [1]