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 provides 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 $\pi(c_w, c_h)$ will control an analogue of the FDR, while a decision rule based on classical hypothesis testing will not.

As in Section 9, let $\delta : \{0, ..., L_w\} \times \{0, ..., L_h\} \rightarrow \{0, 1\}$ 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: $pFDR_J = \mathbb{E} \left[ N_J^{-1} \sum_{j=1}^{J} \delta(C_{jw}, C_{jb})(1 - D_j)|N_J \geq 1 \right]$. In words, $pFDR_J$ 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 $pFDR_J$.

**Lemma 4 (pFDR$^*$ Control).** If $\delta(C_{jw}, C_{jb}) = 1 \{\pi(C_{jw}, C_{jb}) > \bar{p}\}$ then $pFDR_J \leq 1 - \bar{p}$.

**Proof.** Storey (2003, Theorem 1) showed that $pFDR_J = \Pr(D_j = 0|\delta(C_{jw}, C_{jb}) = 1)$ 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

\[
pFDR_J = \Pr(D_j = 0|\pi(C_{jw}, C_{jb}) > \bar{p}) \\
\leq \Pr(D_j = 0|\pi(C_{jw}, C_{jb}) = \bar{p}) = 1 - \bar{p}.
\]

By contrast, consider an alternative decision rule $\delta^+(C_{jw}, C_{jb})$ 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 non-discrimination so that

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

We can write the resulting $pFDR_J$ of this rule

\[
\Pr(D_j = 0|\delta^+(C_{jw}, C_{jb}) = 1) = \frac{\Pr(\delta^+(C_{jw}, C_{jb}) = 1|D_j = 0) (1 - \pi)}{\Pr(\delta^+(C_{jw}, C_{jb}) = 1|D_j = 0) (1 - \pi) + \Pr(\delta^+(C_{jw}, C_{jb}) = 1|D_j = 1) \frac{\tilde{\alpha}(1 - \pi)}{\tilde{\alpha}(1 - \pi) + \bar{\pi}}}
\]

To see that $\delta^+(C_{jw}, C_{jb})$ fails to control $pFDR_J$, note that $\lim_{\pi \downarrow 0} \frac{\tilde{\alpha}(1 - \pi)}{\tilde{\alpha}(1 - \pi) + \bar{\pi}} = 1$. 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 $FDR_J = pFDR_J \times \Pr(N_J \geq 1)$. Because $\Pr(N_J \geq 1) \leq 1$, Lemma 4 implies that the posterior cutoff rule also controls $FDR_J$.

**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:

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

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

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

which implies

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

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

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

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

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

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

where the $\eta_{ks}$ 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

$$\varrho(k, s) = \min \left\{ \frac{k - s}{K} diag, \frac{1 - s}{K(1 + K - y)} off-diag \right\}.$$  

34
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:

$$\bar{f}(c_w, c_b) = \left( \begin{array}{c} L \vspace{0.5ex} \\ c_w \end{array} \right) \left( \begin{array}{c} L \vspace{0.5ex} \\ c_b \end{array} \right) \sum_{k=1}^{K} \sum_{s=1}^{K} \eta_{ks} \varrho(k,s)^c \left( 1 - \varrho(k,s) \right) c_w \varrho(s,k)^c \left( 1 - \varrho(s,k) \right) c_b, \quad (10)$$

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

$$\min_{\{\eta_{ks}\}} 1 - \sum_{(c'_w, c'_b) : c'_w + c'_b = t} \bar{f}(c'_w, c'_b) \sum_{k=1}^{K} \eta_{kk} \varrho(k,k)^t \left( 1 - \varrho(k,k) \right)^{L-t},$$

subject to (10) 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 et al. (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 callback probabilities $\{\Pr(C_{jw} = c_w, C_{jb} = c_b | L_j = l)\}_{c_w,c_b}$ into the vector $\tilde{f}_l$, our model relates these probabilities to moments of the callback distribution via the linear system $\tilde{f}_l = B_l \mu$, for $B_l$ a fixed matrix of binomial coefficients. Letting $\tilde{f}$ denote the vector formed by “stacking” the $\{\tilde{f}_l\}$ across designs in an experiment, we write $\tilde{f} = B \mu$. Let $\eta$ be a $K^2 \times 1$ vector comprised 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 comprised of entries with typical element $\varrho(k,s)^m \varrho(s,k)^n$. We
then have the moment restriction $\bar{f} = BM\eta$.

Let $\hat{f}$ denote the vector of empirical call back probabilities with typical element:

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

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

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

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

where $W$ is a fixed weighting matrix. Note that because $G(\cdot, \cdot)$ 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 (11) with diagonal weights proportional to the number of jobs used in the application design and then choosing $W = \hat{\Sigma}^{-1}$ where $\hat{\Sigma} = \text{diag} \left( \hat{f}^{(1)} \right)^2 - \hat{f}^{(1)}\hat{f}^{(1)'}$ is an estimate of the variance-covariance matrix of the callback frequencies implied by the first step shape-constrained callback probability estimates $\hat{f}^{(1)}$.

**Hong and Li (forthcoming) standard errors**

Standard errors on the moment estimates $\hat{\mu}$ are computed via the numerical bootstrap procedure of Hong and Li (forthcoming) 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 $\hat{\mu}^*$ of $\hat{\mu}$ solves the quadratic programming problem in (11) where $\bar{f}$ has been replaced by $\left( \hat{f} + J^{-1/3} f^* \right)$. The bootstrap probabilities $f^*$ have typical element:

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

where $\{ \omega_j^* \}_{j=1}^J$ are a set of iid draws from an exponential distribution with mean and variance one.

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 $J^{-1/3} [\phi (\hat{\mu}^*) - \phi (\hat{\mu})]$.

**Chernozhukov et al. (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 et al. (2015). Our test statistic (the “$J$-test”) can be written:

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

36
\[ s.t. \quad \eta \geq 0, \quad 1^t \eta = 1. \]

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:

\[ T_n^* = \inf_{\eta, h} (\mathbb{F}^* - B M h)^t W^* (\mathbb{F}^* - B M h) \]

\[ s.t. \quad (\tilde{f} - B M \eta)^t W (\tilde{f} - B M \eta) = T_n, \quad \eta \geq 0, \quad 1^t \eta = 1, \quad h \geq -\eta, \quad 1^t h = 0. \]

As in the full sample problem, we conduct a two-step GMM procedure in each bootstrap replication, setting \( W^* = \left[ \text{diag}(B M \eta^{(1)*}) - (B M \eta^{(1)*})(B M \eta^{(1)*})^t \right]^{-1} \) 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 (12), we re-formulate the problem in two ways. First, we define primary and auxiliary vectors of errors for each moment condition. Letting \( \xi_h = \mathbb{F}^* - B M h \) and \( \xi_\eta = \tilde{f} - B M \eta \), the problem can be re-posed as:

\[ T_n^* = \inf_{\xi_h, \xi_\eta} \xi_h^t W^* \xi_h, \]

\[ s.t. \quad \xi_\eta^t W \xi_\eta = T_n, \quad B M h + \xi_h = \mathbb{F}^*, \quad B M \eta + \xi_\eta = \tilde{f}, \quad 1^t h = 0, \quad 1^t \eta = 1, \quad h \geq -\eta, \quad \eta \geq 0. \]

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

\[ T_n^* = \inf_{\xi_h, \xi_\eta} \xi_h^t W^* \xi_h, \]

\[ s.t. \quad \xi_\eta^t W \xi_\eta = T_n, \quad B M h^+ + \xi_h + \xi_\eta = \mathbb{F}^*, \quad B M \eta + \xi_\eta = \tilde{f}, \quad 1^t h^+ = 1, \quad 1^t \eta = 1, \quad h^+ \geq 0, \quad \eta \geq 0. \]

Note that 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 \( G \left( p_w^H, p_w^L, p_b^H, p_b^L \right) \) with the discretized distribution

\[ G_K \left( p_w^H, p_w^L, p_b^H, p_b^L \right) = \sum_{k=1}^K \sum_{s=1}^K \sum_{k'=1}^K \sum_{s'=1}^K \eta_{k,s,k',s'} 1 \left\{ p_w^H \leq g(k, s), p_w^L \leq g(k', s'), p_b^H \leq g(s, k), p_b^L \leq g(s', k') \right\}, \]

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 \( \delta(C_j, L_j, q) = 1 \{ P(C_j, L_j, G_{logit}) > q \} \). With this notation, we can write the risk function

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

where \( \mathcal{A}_1 \) is the set of all \( 2^5 = 36 \) binary quality permutations possible in a design with 5 white and 5 black applications and \( w_l = \left( \frac{5}{L_j^w} \right) \left( \frac{5}{L_j^b} \right) (1/2)^{10} \) 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:

\[
\mathbb{E} \left[ \delta(C_j, l, q) \left\{ \kappa - \Lambda \left( \sum_{x \in \{H, L\}} \frac{\Lambda^{-1}(p_{uw}) - \Lambda^{-1}(p_{ub})}{2} \right) \right\} \mid L_j = l \right] = \sum_{c_{lw}^H=0}^{a_{lw}^H} \sum_{c_{lw}^L=0}^{a_{lw}^L} c_{lw}^H c_{lw}^L \sum_{s,k} \sum_{s,k'} \sum_{s'_k} \sum_{s'_k} \sum_{s'_k} \sum_{s'_k} \delta(c, l, q) \eta_{sksk'} \left( \frac{L_{ub}^{H}}{c_{lw}^H} \right) \left( \frac{L_{ub}^{L}}{c_{lw}^L} \right) \left( \frac{L_{ub}^{b}}{c_{lw}^b} \right) \left( \frac{L_{ub}^{b}}{c_{lw}^b} \right)
\times \varrho(k, s) c_{lw}^H (1 - \varrho(k, s)) c_{lw}^H \varrho(s, k) c_{lw}^L (1 - \varrho(s, k)) c_{lw}^L
\times \varrho(k', s') c_{lw}^L (1 - \varrho(k', s')) c_{lw}^L \varrho(s', k') c_{lw}^L (1 - \varrho(s', k')) c_{lw}^L
\times \left( \kappa - \Lambda \left( \frac{\Lambda^{-1}(\varrho(k, s)) - \Lambda^{-1}(\varrho(s, k))}{2} \right) + \frac{\Lambda^{-1}(\varrho(k', s')) - \Lambda^{-1}(\varrho(s', k'))}{2} \right).
\]

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

\[
R^m(q) = \max_{\{\eta_{sksk'}\}} \sum_{l \in \mathcal{A}_1} w_l \mathbb{E} \left[ \delta(C_j, l, q) \left\{ \kappa - \Lambda \left( \sum_{x \in \{H, L\}} \frac{\Lambda^{-1}(p_{uw}) - \Lambda^{-1}(p_{ub})}{2} \right) \right\} \mid L_j = l \right]
\]

subject to the constraint that the \( \eta_{sksk'} \) are non-negative and sum to one and that the following moment restrictions hold:

\[
\text{Pr}(C_j = c \mid L_j = l) = \left( \frac{L_{ub}^{H}}{c_{lw}^H} \right) \left( \frac{L_{ub}^{L}}{c_{lw}^L} \right) \sum_{k=1}^{K} \sum_{s=1}^{K} \sum_{s'=1}^{K} \eta_{sksk'}
\times \varrho(k, s) c_{lw}^H (1 - \varrho(k, s)) c_{lw}^H \varrho(s, k) c_{lw}^L (1 - \varrho(s, k)) c_{lw}^L
\times \varrho(k', s') c_{lw}^L (1 - \varrho(k', s')) c_{lw}^L \varrho(s', k') c_{lw}^L (1 - \varrho(s', k')) c_{lw}^L.
\]

We impose these restrictions for the following set of designs, all of which are present in the Nunley et al. (2015) experiment: \( \mathcal{A}_2 = \{(2,0,2,0), (2,0,0,2), (0,2,2,0), (0,2,0,2)\} \). To operationalize
these constraints, we replace the unknown cell probabilities $\Pr(C_j = c | L_j = l)$ for all $c$ and $l$ in $\mathcal{A}_2$ 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.
Figure A.I: 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 X 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 grey points show frequencies for designs with 3 white and 1 black, 1 white and three black, 4 white and zero black, and 0 white and 4 black applications, respectively. The dashed line is the 45-degree line. The chi-squared statistic and $p$-value come from a test that all model-predicted and empirical frequencies match, treating the model predictions as fixed.
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

<table>
<thead>
<tr>
<th>Moment</th>
<th>No constraints (1)</th>
<th>Shape constraints (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E[p_w]$</td>
<td>0.094 (0.006)</td>
<td>0.094 (0.007)</td>
</tr>
<tr>
<td>$E[p_b]$</td>
<td>0.063 (0.006)</td>
<td>0.063 (0.006)</td>
</tr>
<tr>
<td>$E[(p_w - E[p_w])^2]$</td>
<td>0.040 (0.005)</td>
<td>0.040 (0.005)</td>
</tr>
<tr>
<td>$E[(p_b - E[p_b])^2]$</td>
<td>0.023 (0.004)</td>
<td>0.023 (0.004)</td>
</tr>
<tr>
<td>$E[(p_w - E[p_w])(p_b - E[p_b])]$</td>
<td>0.028 (0.004)</td>
<td>0.028 (0.003)</td>
</tr>
<tr>
<td>$E[(p_w - E[p_w])^2(p_b - E[p_b])]$</td>
<td>0.015 (0.003)</td>
<td>0.014 (0.002)</td>
</tr>
<tr>
<td>$E[(p_w - E[p_w])(p_b - E[p_b])^2]$</td>
<td>0.023 (0.003)</td>
<td>0.012 (0.002)</td>
</tr>
<tr>
<td>$E[(p_w - E[p_w])^2(p_b - E[p_b])^2]$</td>
<td>0.010 (0.003)</td>
<td>0.010 (0.002)</td>
</tr>
</tbody>
</table>

$J$-statistic: 0.0
$P$-value: 1.00

Sample size: 1,112

Notes: 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 come from a bootstrap test of the hypothesis that the model restrictions are satisfied.
Table A.II: Moments of callback rate distribution, NPRS data

<table>
<thead>
<tr>
<th>Moment</th>
<th>Design-specific estimates</th>
<th>Combined estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(2,2) design (1)</td>
<td>(3,1) design (2)</td>
</tr>
<tr>
<td>( E[p_w] )</td>
<td>0.174 (0.010)</td>
<td>0.199 (0.025)</td>
</tr>
<tr>
<td>( E[p_b] )</td>
<td>0.148 (0.010)</td>
<td>0.149 (0.015)</td>
</tr>
<tr>
<td>( E[(p_w - E[p_w])^2] )</td>
<td>0.089 (0.007)</td>
<td>0.108 (0.009)</td>
</tr>
<tr>
<td>( E[(p_b - E[p_b])^2] )</td>
<td>0.085 (0.007)</td>
<td>-</td>
</tr>
<tr>
<td>( E[(p_w - E[p_w])(p_b - E[p_b])] )</td>
<td>0.083 (0.006)</td>
<td>0.084 (0.009)</td>
</tr>
<tr>
<td>( E[(p_w - E[p_w])^3] )</td>
<td>-</td>
<td>0.051 (0.008)</td>
</tr>
<tr>
<td>( E[(p_b - E[p_b])^3] )</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( E[(p_w - E[p_w])^2(p_b - E[p_b])] )</td>
<td>0.044 (0.004)</td>
<td>0.043 (0.007)</td>
</tr>
<tr>
<td>( E[(p_w - E[p_w])(p_b - E[p_b])^2] )</td>
<td>0.047 (0.005)</td>
<td>-</td>
</tr>
<tr>
<td>( E[(p_w - E[p_w])^3(p_b - E[p_b])] )</td>
<td>-</td>
<td>0.034 (0.005)</td>
</tr>
<tr>
<td>( E[(p_w - E[p_w])(p_b - E[p_b])^3] )</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( E[(p_w - E[p_w])^2(p_b - E[p_b])^2] )</td>
<td>0.036 (0.004)</td>
<td>-</td>
</tr>
</tbody>
</table>

\[ J \text{-statistic: } 23.1 \]
\[ P \text{-value: } 0.190 \]

Sample size | 1,146 | 544 | 550 | 2,240

Notes: 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>
<thead>
<tr>
<th>Moment</th>
<th>No constraints (1)</th>
<th>Shape constraints (2)</th>
<th>Moment</th>
<th>No constraints (3)</th>
<th>Shape constraints (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E[p_f]$</td>
<td>0.136 (0.010)</td>
<td>0.137 (0.010)</td>
<td>$E \left[(p_f - E[p_f])^4 \right]$</td>
<td>0.024 (0.004)</td>
<td>0.026 (0.003)</td>
</tr>
<tr>
<td>$E[p_m]$</td>
<td>0.103 (0.009)</td>
<td>0.109 (0.009)</td>
<td>$E \left[(p_m - E[p_m])^4 \right]$</td>
<td>0.019 (0.004)</td>
<td>0.023 (0.003)</td>
</tr>
<tr>
<td>$E \left[(p_f - E[p_f])^2 \right]$</td>
<td>0.066 (0.006)</td>
<td>0.066 (0.006)</td>
<td>$E \left[(p_f - E[p_f])^4 (p_m - E[p_m]) \right]$</td>
<td>0.012 (0.003)</td>
<td>0.012 (0.002)</td>
</tr>
<tr>
<td>$E \left[(p_m - E[p_m])^2 \right]$</td>
<td>0.047 (0.005)</td>
<td>0.052 (0.006)</td>
<td>$E \left[(p_f - E[p_f]) (p_m - E[p_m]) \right]$</td>
<td>0.013 (0.003)</td>
<td>0.013 (0.002)</td>
</tr>
<tr>
<td>$E \left[(p_f - E[p_f])^3 \right]$</td>
<td>0.032 (0.005)</td>
<td>0.064 (0.007)</td>
<td>$E \left[(p_f - E[p_f])^3 (p_m - E[p_m]) \right]$</td>
<td>0.012 (0.003)</td>
<td>0.013 (0.002)</td>
</tr>
<tr>
<td>$E \left[(p_m - E[p_m])^3 \right]$</td>
<td>0.025 (0.005)</td>
<td>0.048 (0.007)</td>
<td>$E \left[(p_f - E[p_f]) (p_m - E[p_m]) \right]$</td>
<td>0.010 (0.002)</td>
<td>0.010 (0.002)</td>
</tr>
<tr>
<td>$E \left[(p_f - E[p_f])^2 (p_m - E[p_m]) \right]$</td>
<td>0.021 (0.004)</td>
<td>0.018 (0.003)</td>
<td>$E \left[(p_f - E[p_f])^2 (p_m - E[p_m]) \right]$</td>
<td>0.010 (0.002)</td>
<td>0.010 (0.002)</td>
</tr>
<tr>
<td>$E \left[(p_f - E[p_f]) (p_m - E[p_m])^2 \right]$</td>
<td>0.022 (0.004)</td>
<td>0.020 (0.003)</td>
<td>$E \left[(p_f - E[p_f])^3 (p_m - E[p_m]) \right]$</td>
<td>0.010 (0.002)</td>
<td>0.009 (0.002)</td>
</tr>
<tr>
<td>$E \left[(p_f - E[p_f])^3 (p_m - E[p_m]) \right]$</td>
<td>0.015 (0.003)</td>
<td>0.015 (0.002)</td>
<td>$E \left[(p_f - E[p_f])^4 (p_m - E[p_m]) \right]$</td>
<td>0.008 (0.002)</td>
<td>0.008 (0.001)</td>
</tr>
<tr>
<td>$E \left[(p_f - E[p_f]) (p_m - E[p_m])^3 \right]$</td>
<td>0.016 (0.003)</td>
<td>0.017 (0.002)</td>
<td>$E \left[(p_f - E[p_f])^3 (p_m - E[p_m]) \right]$</td>
<td>0.008 (0.002)</td>
<td>0.008 (0.002)</td>
</tr>
<tr>
<td>$E \left[(p_f - E[p_f])^2 (p_m - E[p_m])^2 \right]$</td>
<td>0.016 (0.003)</td>
<td>0.016 (0.002)</td>
<td>$E \left[(p_f - E[p_f])^4 (p_m - E[p_m]) \right]$</td>
<td>0.007 (0.002)</td>
<td>0.007 (0.001)</td>
</tr>
</tbody>
</table>

**Notes:** 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 come from a bootstrap test of the hypothesis that the model restrictions are satisfied.
Table A.IV: Sensitivity of moments and bounds to discretization grid

<table>
<thead>
<tr>
<th>Grid spacing</th>
<th>J-statistic</th>
<th>Share discriminating one call</th>
<th>Share discriminating two calls</th>
<th>Share discriminating three calls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>$K_1 = 50$</td>
<td>Quadratic</td>
<td>23.13</td>
<td>0.266</td>
<td>0.483</td>
</tr>
<tr>
<td>$K_1 = 100$</td>
<td>Quadratic</td>
<td>23.10</td>
<td>0.366</td>
<td>0.692</td>
</tr>
<tr>
<td>$K_1 = 150$</td>
<td>Quadratic</td>
<td>23.09</td>
<td>0.401</td>
<td>0.765</td>
</tr>
<tr>
<td>$K_2 = 200$</td>
<td>Quadratic</td>
<td>23.10</td>
<td>0.366</td>
<td>0.692</td>
</tr>
<tr>
<td>$K_2 = 300$</td>
<td>Quadratic</td>
<td>23.09</td>
<td>0.401</td>
<td>0.765</td>
</tr>
<tr>
<td>$K_2 = 600$</td>
<td>Quadratic</td>
<td>23.09</td>
<td>0.401</td>
<td>0.765</td>
</tr>
<tr>
<td>$K_2 = 900$</td>
<td>Quadratic</td>
<td>23.09</td>
<td>0.401</td>
<td>0.765</td>
</tr>
<tr>
<td>$K_2 = 900$</td>
<td>Quadratic</td>
<td>23.09</td>
<td>0.401</td>
<td>0.765</td>
</tr>
<tr>
<td>$K_2 = 900$</td>
<td>Quadratic</td>
<td>23.09</td>
<td>0.401</td>
<td>0.765</td>
</tr>
<tr>
<td>$K_2 = 900$</td>
<td>Quadratic</td>
<td>23.09</td>
<td>0.401</td>
<td>0.765</td>
</tr>
</tbody>
</table>

Notes: 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. $K_1$ refers to the number of mass points used in the quadratic programming SCGMM step, while $K_2$ 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 $K_1$. Column (2) displays the lower bound on the fraction of discriminating jobs for each combination of $K_1$ and $K_2$. 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.