Online Appendix for
"Worker-Side Discrimination: Beliefs and Preferences"

Md Moshi Ul Alam   Mehreen Mookerjee   Sanket Roy

July 22, 2025

A Appendix: Additional Information on the Survey and the Context

A.1 Institutional Context for Campus Recruitment

In most elite institutions in India, "campus recruitment" is a common practice. Representatives of recruiting firms—current employees including managers and at times vice-presidents—travel to university campuses and interview final-year students during pre-scheduled sessions. The recruitment process usually involves several stages: written aptitude tests, followed by one or more rounds of interviews. The final stages of the interview process, aside from the HR round, are often conducted by a panel that includes the hiring manager(s) under whom the selected candidate would work. This institutional feature typically makes jobseekers aware of the gender of their potential manager before accepting a job offer.

A.2 Details on survey administration and data collection

The online survey was designed and implemented on Qualtrics. We collected data in the second week of April 2020, using an online survey administered to students of a highly selective public university in India. Access to the internet was not a concern for our sample of students studying in a premier university in one of the largest metropolitan cities of India. However, we paid special attention in designing the survey to ensuring that it was mobile-friendly, in consideration of the fact that a small but significant proportion of the target population might not have access to a computer. Recruitment of students was done by the research assistants (RAs), who had previous experience in recruiting students for surveys and RCTs. We administered the online survey in three key steps.

Step 1: The RAs, based on their previous experience, sent a sign-up link to each department's student class representative, who distributed the link in the class lists. The sign-up sheet, in addition to containing the consent form for them to sign, asked for e-mail addresses to which the link of the survey would be sent, as well as basic demographic information, department affiliation, faculty of study (arts/science/engineering) and level (bachelor's or master's) and year of study. The sign-up sheet described the survey as *"...an online survey on hypothetical job choices"* with the purpose described as *"...to better understand the* *preferences for the different attributes of jobs.".* Students were also allowed to choose the date and the time at which they would like to take the survey. They had a choice among 4 dates from April 8th to April 11th. On their selected date, they had a choice among 6 time slots: 10 am, 12 noon, 5 pm, 7 pm, 9 pm and 11 pm. We observed that our pilots that had specified time slots along with dates had higher completion rates than those with just dates. The sign-up form ended with a summary. The sign-up form was designed to automatically send respondents' enrollment form to their email address. This enabled us to automatically have the signed copy of the consent form sent to the participant.

Step 2: Upon receiving the sign-ups, we scheduled emails to be sent out with unique links to the survey for each participant an hour before each he or she was scheduled to participate in the survey. Hence, the link could not be used on two different devices to fill in the survey. The survey was also designed to prevent ballot-boxing; i.e., once the survey was completed from one link, when clicked again, that link would show a confirmation that the survey had already been completed. The links were designed to expire within 24 hours. Thanks to the extensive pilots done before, we did not face any technical difficulties while implementing the survey. Debriefs with pilot participants were extremely helpful for rewording the questions to optimize communication and maximize participants' understanding.

Step 3: The mode and details of online payment were selected in the last section of the survey. The options included direct bank transfers, PayPal and UPI (unified payment interface). The payment was processed for the list of verified students within the prestated timeline for each payment mode.

B Appendix B: Gender gap at management levels

B.1 Workers

Utility of worker i at firm j is given by:

$$U_{ij} = w_j + \varphi\, q_j + \psi\, f_j + \beta^{-1}\epsilon_{ij},$$

where $w_j \equiv logW_j$ is the posted wage, $f_j \in [0,1]$ is the observable share of managers who are female, and q_j is (true) average mentoring quality in the firm. The parameters $\varphi > 0$ and ψ capture, respectively, the value the worker places on mentoring quality and an intrinsic taste for female managers. Worker horizontal preferences are i.i.d. across workers and firms with $\epsilon_{ij} \sim \text{Gumbel}(\text{location} = 0, \text{scale} = 1/\beta)$ i,.e., $\text{CDF}\,(\epsilon_{ij}) = \exp\left(-\exp\left(-\beta\epsilon_{ij}\right)\right)$. Let $\varphi\, q_j + \psi\, f_j \equiv logG_j \equiv g_j$. This makes the utility of worker i at firm j:

$$U_{ij} = V_{ij} + \beta^{-1}\epsilon_{ij} \qquad \text{where} \qquad V_{ij} \equiv w_j + g_j$$

representing simple monopsony model with vertical differentiation in mentoring quality and fraction of female managers, which can be thought of as amenities in the spirit of Lamadon, Mogstad and Setzler (2022) in absence of taxes and worker unobserved heterogeneity. The probability that worker i chooses firm j is given by::

$$P_{ij} = \frac{\exp(\beta V_{ij})}{\sum_{k\in\mathcal{J}} \exp(\beta V_{ik})},$$

where $\mathcal{J}$ is the set of all firms available to worker i. Summing across an economy of size N we obtain the aggregate labour supply curve for firm j:

$$L_j = NP_{ij} = N\frac{\exp(\beta w_j + \beta g_j)}{\sum_{k\in\mathcal{J}} \exp(\beta w_k + \beta g_k)}.$$

LMS show that this leads to a log-additive labor supply function:

$$\ell_j = \beta w_j + \beta g_j + \text{ constant} \qquad \text{(labor supply)}$$

Correspondingly, β captures the labor supply elasticity assuming firms are strategically small in spirit of Dixit and Stiglitz (1977). Although the labor supply elasticity is the same for all firms, the labor supply curve varies across firms because of vertical differentiation in amenities.

Workers observe f_j but in absence of information on mentorship quality they under-rate female mentors by fraction $b \in (0,1)$. Without loss of generality, normalize true quality so that male and female mentors are both quality 1. The biased average mentorship at firm j is then $\tilde{q}_j = (1 - f_j)\,\cdot 1 \;+\; f_j\,(1-b) \;=\; 1 - bf_j$.

B.2 Firms

Suppose firm production function $f_j(\cdot)$ has constant elasticity in labor with TFP A_j. Let the firm face a quadratic cost $c(f_j) = \frac{\gamma}{2}(f_j - \bar{f})^2$ of maintaining fraction f_j of female managers, where $\bar{f}$ is the average female manager share in the qualified labor pool and with $\gamma > 0$ firms face adjustment costs deviating from this average. Taking product market prices as given and normalized to 1, firm j's profit function becomes:

$$\pi_j(L_j, f_j) = A_j\,(L_j)^{1-\alpha} - W_j L_j - \frac{\gamma}{2}(f_j - \bar{f})^2$$

where W_j is the wage paid and L_j is the amount of labor hired by firm j.[1]

Equating marginal revenue of labor to the marginal cost of the labor and taking logs, we obtain firm demand for labor:

$$w_j = a_j - \alpha \ell_j + \text{ constant} \tag{labor demand}$$

B.3 Equilibrium

Solving for w_j from the labor supply and demand gives us the equilibrium wage and the labor supply functions:

$$\begin{aligned} w_j &= \frac{1}{1+\alpha\beta} a_j - \frac{\alpha\beta}{1+\alpha\beta} g_j + \text{ constant} \\ \ell_j &= \frac{\beta}{1+\alpha\beta} a_j + \frac{\beta}{1+\alpha\beta} g_j + \text{ constant} \end{aligned}$$

Incorporating the expression for g_j into the optimal wage and labor, we have:

$$w_j = \frac{1}{1+\alpha\beta} a_j - \frac{\alpha\beta}{1+\alpha\beta} (\varphi q_j + \psi f_j) + \text{ constant} \tag{1}$$

$$\ell_j = \frac{\beta}{1+\alpha\beta} a_j + \frac{\beta}{1+\alpha\beta} (\varphi q_j + \psi f_j) + \text{ constant} \tag{2}$$

Wage and labor elasticities with respect to f_j are:

$$\frac{\partial w_j}{\partial f_j} = -\frac{\alpha\beta}{1+\alpha\beta}\psi \tag{3}$$

$$\frac{\partial \ell_j}{\partial f_j} = \frac{\beta}{1+\alpha\beta}\psi \tag{4}$$

In absence of information on mentorship quality the wage equation in absence of information on mentorship becomes:

[1] Note these costs are symmetric in $f_j, \bar{f}$ and $m_j, \bar{m}$, where m_j is the share of male managers in the firm. So the costs that the firm faces is an adjustment cost for deviating from the averages. The quadratic structure is a convenience that prevents corner solutions.

$$w_j^I = \frac{1}{1+\alpha\beta} a_j - \frac{\alpha\beta}{1+\alpha\beta}(\varphi \tilde{q}_j + \psi f_j) + \text{ constant}$$

$$\begin{aligned}
\ell_j^I &= \frac{\beta}{1+\alpha\beta} a_j + \frac{\beta}{1+\alpha\beta}(\varphi \tilde{q}_j + \psi f_j) + \text{ constant} \\
&= \frac{\beta}{1+\alpha\beta} a_j + \frac{\beta}{1+\alpha\beta}(\psi + (\psi - b\varphi) f_j) + \text{ constant}
\end{aligned}$$

Wage and labor elasticities with respect to f_j are:

$$\begin{aligned}
\frac{\partial w_j^I}{\partial f_j} &= -\frac{\alpha\beta}{1+\alpha\beta}(\psi - b\varphi) \\
\frac{\partial \ell_j^I}{\partial f_j} &= \frac{\beta}{1+\alpha\beta}(\psi - b\varphi)
\end{aligned}$$

Firm's optimal choice of female manager share

Taking the FOC of firm profits with respect to f_j, we have

$$\begin{aligned}
\frac{d\pi_j}{df_j} &= \frac{\partial \pi_j}{\partial f_j} + \frac{\partial \pi_j}{\partial L_j}\frac{\partial L_j}{\partial f_j} = 0 \\
&\implies -\frac{\partial W_j}{\partial f_j} L_j^* - W_j^* \frac{\partial L_j}{\partial f_j} + MRP_L \frac{\partial L_j}{\partial f_j} - \gamma(f_j^* - \bar{f}) = 0 \\
&\implies -\frac{\partial W_j}{\partial f_j} L_j^* - \gamma(f_j^* - \bar{f}) = 0 \qquad (\text{since} \quad MRP_L = W_j^* \text{ in equilibrium})
\end{aligned}$$

Since $w_j \equiv \log(W_j)$, we have, $\frac{\partial W_j}{\partial f_j} = \frac{\partial \exp(w_j)}{\partial f_j} = \exp(w_j)\frac{\partial w_j}{\partial f_j} = W_j \frac{\partial w_j}{\partial f_j}$. So the firm's FOC with respect to f_j becomes:

$$-\frac{\partial w_j}{\partial f_j} W_j^* L_j^* - \gamma(f_j^* - \bar{f}) = 0$$

Under full information, $\frac{\partial w_j}{\partial f_j} = \frac{\alpha\beta}{1+\alpha\beta}\psi$, so we have:

$$f_j^* = \bar{f} + \frac{W_j^* L_j^*}{\gamma}\frac{\alpha\beta}{1+\alpha\beta}\psi \tag{5}$$

Under incomplete information, $\frac{\partial w_j^I}{\partial f_j} = -\frac{\alpha\beta}{1+\alpha\beta}(\psi - b\varphi)$, so we have:

$$f_j^{I*} = \bar{f} + \frac{W_j^{I*} L_j^{I*}}{\gamma}\frac{\alpha\beta}{1+\alpha\beta}(\psi - b\varphi) \tag{6}$$

Since $\psi - b\varphi \approx 0$ from our empirical results, we have $f_j^{I*} \approx \bar{f}$. However, with information provision, $f_j^* > \bar{f}$ since $\psi > 0$. Specifically, the change in female manager share from information provision at firm j is:

$$\Delta f_j \equiv f_j^* - f_j^{I*} = \frac{W_j^* L_j^*}{\gamma} \frac{\alpha\beta}{1+\alpha\beta} b\varphi > 0 \tag{7}$$

This implies that with information provision, firms increase the share of female managers. Furthermore, larger firms (higher L_j) with larger wage bills (higher $W_j L_j$) have more to gain from information provision, and hence increase their female manager share more.

B.3.1 Extension: Discriminatory firms

Consider firms with discriminatory preferences against female managers, captured by an additional cost $d \cdot f_j$ where $d > 0$ represents the discrimination parameter. The profit function for discriminatory firm j becomes:

$$\pi_j^D(L_j, f_j) = A_j (L_j)^{1-\alpha} - W_j L_j - \frac{\gamma}{2}(f_j - \bar{f})^2 - d \cdot f_j \tag{8}$$

The FOC with respect to f_j yields:

With full information:

$$\frac{\partial \pi_j^D}{\partial f_j} = W_j^{D*} L_j^{D*} \frac{\alpha\beta}{1+\alpha\beta} \psi - \gamma(f_j^{D*} - \bar{f}) - d = 0$$

$$\implies f_j^{D*} = \bar{f} + \frac{1}{\gamma}\left(W_j^{D*} L_j^{D*} \frac{\alpha\beta}{1+\alpha\beta}\psi - d \right)$$

Without information:

$$\frac{\partial \pi_j^D}{\partial f_j} = W_j^{D,I*} L_j^{D,I*} \frac{\alpha\beta}{1+\alpha\beta}(\psi - b\varphi) - \gamma(f_j^{D,I*} - \bar{f}) - d = 0$$

$$\implies f_j^{D,I*} = \bar{f} + \frac{1}{\gamma}\left(W_j^{D,I*} L_j^{D,I*} \frac{\alpha\beta}{1+\alpha\beta}(\psi - b\varphi) - d \right)$$

Since $\psi - b\varphi \approx 0$ from our empirical results:

$$f_j^{D,I*} \approx \bar{f} - \frac{d}{\gamma} < \bar{f} = f_j^{I*}$$

So in absence of information, in equilibrium the share of female managers in discriminatory firms ($f_j^{D,I*}$) is lower than the share of female managers in non-discriminatory firms (f_j^{I*}). The change in female manager share from information provision in discriminatory firms is:

$$\Delta f_j^D = f_j^{D*} - f_j^{D,I*} = \frac{W_j^{D*} L_j^{D*}}{\gamma} \frac{\alpha\beta}{1+\alpha\beta} b\varphi > 0 \tag{9}$$

The increase in female manager share from information provision is identical for discriminatory and non-discriminatory firms. The equality results from the linear discrimination cost and can be easily extended. However, the levels differ. Specifically, discriminatory firms have lower female manager shares both with and without information. Information provision partially offsets discrimination by creating wage savings that compensate for discrimination costs.

C Appendix: Non-parametric empirical Bayes (NPEB) shrinkage with heteroskedastic errors

C.1 Setup

For each individual $i = 1, \dots, n$ we observe an estimate $\hat{\theta}_i$ together with its standard error s_i that we obtain by estimating the model separately for each individual. We model:

$$\hat{\theta}_i \;=\; \theta_i + \varepsilon_i, \qquad \varepsilon_i \,\big|\, s_i \;\sim\; \mathcal{N}\big(0,\, s_i^2\big)$$

Equivalently,

$$\hat{\theta}_i \;\big|\; \theta_i,\; s_i \;\sim\; \mathcal{N}\big(\theta_i,\; s_i^2\big)$$

where $\hat{\theta}_i$ is the noisy point estimate, θ_i is the latent effect, and s_i is the known individual-specific standard error. θ_i are i.i.d. draws from an unknown population distribution $G()$ with density $g()$. The conditional density of the observed estimate is

$$f_i\big(\hat{\theta}_i \mid \theta_i, s_i\big) \;=\; \frac{1}{s_i}\,\phi\Big(\frac{\hat{\theta}_i - \theta_i}{s_i}\Big)$$

where $\phi(\cdot)$ denotes the standard normal pdf. Integrating out θ_i gives us the marginal density of $\hat{\theta}_i$

$$f_i(\hat{\theta}_i) \;=\; \int_{\mathbb{R}} \frac{1}{s_i}\,\phi\Big(\frac{\hat{\theta}_i - \theta}{s_i}\Big) g(\theta)\, d\theta \tag{10}$$

C.2 Log-spline deconvolution estimator for $G()$

We approximate the integral in (10) on a grid of M support points $\tau_1 < \cdots < \tau_M$ with width $\Delta = \tau_{m+1} - \tau_m$ as

$$f_i(\hat{\theta}_i) \;\approx\; \sum_{m=1}^{M} K_{im}\, g_\alpha(\tau_m), \qquad \text{where} \quad K_{im} := \frac{1}{s_i}\phi\Big(\tfrac{\hat{\theta}_i - \tau_m}{s_i}\Big)\Delta$$

where α is a vector of parameters that we use to model the density $g()$ on the grid, $g_m()$ is the density at node m, and K_{im} is the likelihood contribution of node m to the observed estimate $\hat{\theta}_i$. The likelihood contribution of each node is scaled by the width Δ to ensure that the sum approximates the integral.

We approximate $G()$ in smooth exponential family with log density parameterized by log-splines (Efron 2016). Let Q be the $M \times J$ design matrix of cubic B-spline basis functions $B_j(\tau)$ evaluated on the grid. Give α, the mass at point m is given by

$$g_\alpha(\tau_m) := \exp\Big((Q\alpha)_m - \log \sum_{m=1}^{M} exp((Q\alpha)_m)\Big),$$

where $(Q\alpha)_m = \sum_{j=1}^{J} \alpha_j B_j(\tau_m)$ and α is the parameter vector parametrizing the cubic splines $B_j(\tau)$, $j = 1, \dots, J$. Denoting $g_\alpha \equiv (g_\alpha(\tau_1), \dots, g_\alpha(\tau_M))'$ and $K_i \equiv (K_{i1}, \dots, K_{iM})'$, define the mixture vector $f_i(\alpha) := K_i' g_\alpha$, which is the marginal likelihood of $\hat{\theta}_i$ under the log-spline model with parameters α. The parameters α governing $g()$ are estimated by a penalised log-likelihood:

$$\hat{\alpha} = \arg\max_{\alpha} \sum_{i=1}^{N} \log\big(f_i(\alpha)\big) \; - \; \lambda\sqrt{\alpha'\alpha}, \tag{11}$$

where the ridge penalty $\lambda \geq 0$ is a tuning parameter. The objective function is numerically optimized using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm to obtain $\hat{\alpha}$. The corresponding log-spline deconovolution estimate of $G()$ is :

$$\hat{G}(\theta) = \sum_{m=1}^{M} \mathbb{I}\{\tau_m \leq \theta\} g_{\hat{\alpha}}(\tau_m)$$

C.3 Posterior mean and shrinkage factor

Given $\hat{\alpha}$, the posterior density for individual i at τ_m on the grid is obtained via Bayes' rule:

$$p_{im} = \frac{K_{im} g_{\hat{\alpha}}(\tau_m)}{f_i(\hat{\alpha})} = \frac{K_{im} g_{\hat{\alpha}}(\tau_m)}{\sum_{j=1}^{M} K_{ij} g_{\hat{\alpha}}(\tau_j)}$$

This leads to individual i's posterior means $\hat{\theta}_i^{NPEB}$:

$$\hat{\theta}_i^{NPEB} = E[\theta|\hat{\theta}_i] = \sum_{m=1}^{M} p_{im}\, \tau_m \tag{12}$$

and the individual i's shrinkage: $\hat{\theta}_i^{NPEB} - \hat{\theta}_i = \sum_{m=1}^{M} p_{im}(\tau_m - \hat{\theta}_i)$.