Algorithmic Hiring and Diversity: Reducing Human-Algorithm Similarity for Better Outcomes

Prasanna Parasurama
Emory University

Panos Ipeirotis
New York University

May 15, 2025

Abstract

Algorithmic tools are increasingly used in hiring to improve fairness and diversity, often by enforcing constraints such as gender-balanced candidate shortlists. However, we show theoretically and empirically that enforcing equal representation at the shortlist stage does not necessarily translate into more diverse final hires, even when there is no gender bias in the hiring stage. We identify a crucial factor influencing this outcome: the correlation between the algorithm's screening criteria and the human hiring manager's evaluation criteria—higher correlation leads to lower diversity in final hires. Using a large-scale empirical analysis of nearly 800,000 job applications across multiple technology firms, we find that enforcing equal shortlists yields limited improvements in hire diversity when the algorithmic screening closely mirrors the hiring manager's preferences. We propose a complementary algorithmic approach designed explicitly to diversify shortlists by selecting candidates likely to be overlooked by managers, yet still competitive according to their evaluation criteria. Empirical simulations show that this approach significantly enhances gender diversity in final hires without substantially compromising hire quality. These findings highlight the importance of algorithmic design choices in achieving organizational diversity goals and provide actionable guidance for practitioners implementing fairness-oriented hiring algorithms.

1 Introduction

In recent years, organizations have widely adopted various policies to increase workforce diversity (Shi et al. 2018). A popular policy is to diversify candidate shortlists or interview pools—often referred to as *soft* affirmative action policies. Unlike *hard* affirmative action policies such as hiring quotas, which are explicitly prohibited by US employment law, soft policies aim only to increase minority representation in the initial interview stage without imposing quotas on final hires (Civil Rights Act of 1974; Schuck 2002). Prominent examples of soft affirmative action policies include the NFL's Rooney Rule (NFL Operations 2003), which requires interviewing at least one ethnic minority candidate for head coaching positions, and similar policies adopted by major tech firms such as Facebook (2021), Pinterest (2015), and Patreon (2017).

With the increasing use of algorithms in hiring, these diversity policies are frequently implemented as algorithmic fairness constraints. For instance, *LinkedIn Recruiter* has deployed fairness-aware ranking algorithms aimed at improving gender diversity among candidates presented to recruiters (Geyik et al. 2019). However, presenting more diverse candidate sets does not necessarily translate into greater diversity in hiring outcomes. LinkedIn's own analyses highlight uncertainty regarding whether improved gender representation in candidate recommendations leads to measurable improvements in outcomes, such as candidate contacts or interview requests (Geyik et al. 2019). Ultimately, algorithmic recommendations are integrated with human decisions, and the final hiring outcomes depend on human managers or recruiters.

Previous laboratory studies have indicated that the effectiveness of fairness constraints can vary significantly across job types (Sühr et al. 2021; Peng et al. 2019). When these policies fail, conventional wisdom typically attributes their ineffectiveness to human biases. While human biases undoubtedly affect outcomes, other important factors influencing fairness constraints' effectiveness remain underexplored.

To systematically explore these factors, we propose and analyze a two-stage hiring model comprising algorithmic screening followed by human hiring decisions. The model considers a hiring scenario with a higher number of male applicants than female applicants, reflecting conditions typical in firms using diversity policies. In the first stage, a screening algorithm shortlists candidates and applies an *equal selection* constraint, such that an equal number of men and women are shortlisted.

A hiring manager then evaluates the shortlisted candidates and hires the best candidates based on her own assessments.

Analytically solving this model reveals a crucial insight: the effectiveness of the equal selection constraint diminishes as the correlation between the screening algorithm's evaluation criteria and the hiring manager's evaluation criteria increases. Moreover, the expected quality of hires also decreases as the correlation increases. In other words, the better the screening algorithm matches the manager's preferences, the lower the expected quality of hires *and* the less effective the equal selection constraint becomes. Based on this insight, we propose a complementary algorithm designed explicitly to select candidates likely to be overlooked by hiring managers yet still be competitive according to their evaluation criteria.

We empirically validate our theoretical predictions on hire diversity using extensive hiring data from eight technology firms, including nearly 800,000 applicants and over 3,600 job postings. Through counterfactual simulations, we demonstrate two key findings:

1. Consistent with our theoretical predictions, enforcing equal selection constraints in the shortlist does not consistently improve hiring diversity and may have negligible effects in some scenarios.

2. The constraint's effectiveness varies substantially across job types, driven primarily by differences in the correlation between algorithmic screening and managerial assessment criteria.

Furthermore, when we benchmark our complementary algorithm against other traditional fairness constraints, we find it substantially more effective in improving workforce diversity without significant trade-offs in candidate quality.

Our Contributions. We theoretically and empirically show that the equal selection constraint, a common algorithmic fairness constraint, fails to increase workforce diversity when algorithmic screening evaluations correlate strongly with human hiring evaluations. To address this, we introduce and validate a complementary screening algorithm designed specifically to reduce these correlations, significantly improving hire diversity outcomes with minimal loss in candidate quality across various hiring contexts. Our study contributes to the literature on algorithmic fairness in hiring pipelines in two key ways. First, we provide a theoretical characterization of when equal shortlist constraints effectively enhance diversity, emphasizing the critical role of correlation between screening and hiring evaluations. Second, we empirically validate this theoretical insight and introduce a complementary

algorithmic design that significantly improves diversity outcomes in practice.

The remainder of the paper proceeds as follows. Section 2 reviews related literature. Section 3 describes the theoretical model, discusses our findings, and introduces our complementary algorithmic design. Section 4 outlines the empirical approach and data. Section 5 presents our empirical findings, and Section 6 concludes with implications for practice and future research directions.

2 Related Work

This paper is related to the algorithmic fairness literature, which studies the design and evaluation of algorithms aimed to mitigate bias and improve fairness in algorithmic decision-making (Dwork, Hardt, et al. 2012; Zemel et al. 2013; Hardt et al. 2016; Zafar, Valera, Gomez Rodriguez, et al. 2017; Zafar, Valera, Rodriguez, et al. 2017; Geyik et al. 2019; Blum et al. 2022). In this literature, two broad notions of fairness exist: *individual fairness*, which requires that similar individuals are treated similarly by the algorithm; and *group fairness*, which requires that some statistic of interest is on average equal across groups along the lines of *protected attributes*.[1] Within group fairness, different definitions of fairness exist, such as demographic (or statistical) parity, equal selection, equal false-positive rates, equal false-negative rates, equal odds, equal accuracy rates, and equal positive predictive values across groups (see Table 7 for precise definitions and Mitchell et al. (2021) for a review). Except in trivial cases, it is impossible to simultaneously satisfy all fairness criteria (Chouldechova 2017; Kleinberg, Mullainathan, et al. 2016), so the choice of fairness criteria depends on the context and is often informed by laws, policies, and desired outcomes.

Fairness constraints are not only used to mitigate any potential bias in the algorithm but can also be used as a tool to inscribe diversity policies that proactively correct for pre-existing societal and systemic bias. For example, in the hiring context, prior studies have shown that women are deterred from applying to male-dominated jobs because they anticipate discrimination in the hiring process (Storvik and Schøne 2008; Brands and Fernandez-Mateo 2017; Bapna et al. 2021). To address such pre-existing disparities, firms have adopted hiring diversity policies that increase or equalize the representation of minorities in the shortlist (Shi et al. 2018).[2] As hiring becomes

[1] Protected attributes are attributes that are protected under the law against discrimination. U.S. federal law prohibits employment discrimination based on race, gender, religion, national origin, age, disability, sexual orientation, and pregnancy.

[2] For example, the diversity hiring policies implemented in high-tech firms such as Facebook, Pinterest, Patreon,

increasingly aided by algorithms, these diversity policies are implemented as algorithmic fairness constraints. Of particular interest is the *equal selection* fairness constraint (Khalili et al. 2021; Jiang et al. 2023), which requires positive outcomes to be equal across groups regardless of the proportions in the baseline population. For example, in algorithmic hiring, an equal selection constraint might require that the screening algorithm shortlists an equal number of men and women, regardless of the proportion of women in the applicant pool.[3]

Although these constraints guarantee fairness on algorithmic outputs, when these outputs are used as inputs in downstream decisions, the overall effects of these constraints in either mitigating bias or increasing diversity are not guaranteed. An emerging line of literature studies the efficacy of algorithmic fairness constraints in "pipelines"—i.e., settings where decisions are made sequentially. Bower et al. (2017) analyze the equal opportunity constraint in a pipeline setting and shows that individually fair algorithms, when assembled sequentially, do not necessarily guarantee fair final outcomes with respect to equal opportunity. Similarly, Dwork and Ilvento (2019) analyze the individual fairness constraint and conditional parity constraints in composition settings and show that individually fair algorithms, when composed together, do not necessarily guarantee fair final outcomes. Blum et al. (2022) propose a fair algorithm that satisfies the equality of opportunity constraint across the entire selection pipeline. Our main contribution to this algorithmic fairness and fair pipelines literature is that we study the *equal selection* constraint in a hiring pipeline setting, where decisions are made sequentially. We propose an algorithmic design to increase the effectiveness of the equal selection constraint and demonstrate its effectiveness using empirical hiring data.

Outside the algorithmic fairness literature, our work is also related to a number of theoretical papers that study bias and fairness in hiring settings. Kleinberg and Raghavan (2018) provide a theoretical hiring model in the presence of implicit bias and show that the Rooney Rule can increase the proportion of minority hires while also increasing the payoff of the decision-maker (see also Celis et al. (2021)). Fershtman and Pavan (2021) present a model to study the effect of "soft" affirmative

and *LinkedIn Recruiter's* ranking algorithm (Geyik et al. 2019) all seek to increase the representation of minorities in the shortlist.

[3]This is in contrast to *demographic parity*, another common fairness constraint in the algorithmic hiring setting, which requires the proportion of positive outcomes across groups to be equal to the proportions in a baseline population (Raghavan et al. 2020). For example, in algorithmic screening, demographic parity may require that the proportion of women on the shortlist be equal to the proportion of women in the applicant pool. Whereas demographic parity ensures that bias is not introduced in the hiring process, it does not correct for pre-existing disparities.

action policies that increase the proportion of minority candidates in the candidate pool. Lee and Waddell (2021) study a 2-stage hiring setting with agents with different levels of interest in diversity and show that this difference can lower the likelihood of highly qualified candidates being hired even when they enhance diversity. Our contribution to this theoretical hiring literature is that we explicitly model the correlation in assessment criteria between the screener and the hiring manager, which we show to be a key determinant of the effectiveness of a common diversity policy.

3 Theoretical Framework and Implications

3.1 Model Setup

Consider a hiring context with n_a applicants, each characterized by their group membership $g \in \{m, f\}$, where m represents the majority group (male) and f the minority group (female). The female proportion among applicants is $p_a < 0.5$. Each candidate also has an unobservable true quality Q, which is measurable only post-hire (e.g., via job performance).

Two-Stage Hiring Process. The hiring involves two sequential stages:

1. *Algorithmic Screening*: An algorithm assigns each candidate a screening score Q^S and shortlists candidates exceeding a threshold. To enhance diversity, the algorithm implements an *equal selection* constraint, shortlisting an equal number of male and female candidates by setting gender-specific thresholds (τ_m^S, τ_f^S). Let p_s be the proportion of women in the shortlist.

2. *Human Evaluation*: The shortlisted candidates are evaluated by a hiring manager who assigns a score (Q^H) and hires those exceeding a common threshold (τ^H), independent of gender.[4] Let p_h be the proportion of women in the hired pool.

Table 1 summarizes these selection rules.

Quality Scores Model. We model the scores (Q, Q^S, Q^H) as following a multivariate Gaussian

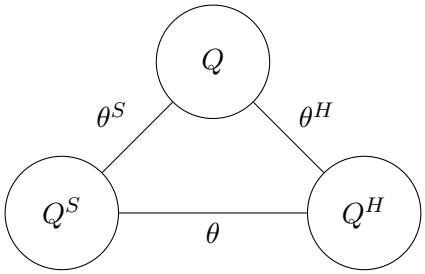
[4]Indeed implementing a constraint on the hiring manager to hire an equal number of men and women would trivially increase the gender diversity of hires; however, such a constraint on the hiring manager would be considered a hiring quota, which is prohibited under US Employment Law (Title VII, Civil Rights Act of 1974). This is the reason many diversity-focused hiring policies (e.g., Rooney Rule, Facebook's hiring policy (Huet 2017), LinkedIn's screening algorithm (Geyik et al. 2019)) target the initial screening decision rather than the final hiring decision.

Table 1: Stages of the hiring model

Stage	Constraint	Selection Rule
(1)	Equal Selection $\mathbb{P}(g = f \mid y^S = 1) = \mathbb{P}(g = m \mid y^S = 1)$	$y^S = \begin{cases} 1 & \text{if } Q^S > \tau_m^S, g = m \\ 1 & \text{if } Q^S > \tau_f^S, g = f \\ 0 & \text{otherwise} \end{cases}$
(2)	None	$y^H = \begin{cases} 1 & \text{if } Q^H > \tau^H \\ 0 & \text{otherwise} \end{cases}$

Notes: y^S and y^H are binary indicators of selection in the screening and hiring stages, respectively. Gender-specific thresholds τ_m^S and τ_f^S ensure equal selection, while τ^H is gender-neutral.

Figure 1: The correlation structure between Q, Q^S, Q^H



Notes: This figure illustrated the correlation structure between the candidates' true quality (Q), the algorithm's quality estimate (Q^S), and the hiring manager's quality estimate (Q^H). The θ values represent the correlations between these scores.

distribution, potentially with distinct distributions for male and female candidates:

$$(Q_m, Q_m^S, Q_m^H) \sim \mathcal{N}\left(\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & \theta^S & \theta^H \\ \theta^S & 1 & \theta \\ \theta^H & \theta & 1 \end{bmatrix} \right) \tag{3.1}$$

$$(Q_f, Q_f^S, Q_f^H) \sim \mathcal{N}\left(\begin{bmatrix} \alpha & \alpha + \beta^S & \alpha + \beta^H \end{bmatrix}, \begin{bmatrix} 1 & \theta^S - \delta^S & \theta^H - \delta^H \\ \theta^S - \delta^S & 1 & \theta - \delta \\ \theta^H - \delta^H & \theta - \delta & 1 \end{bmatrix} \right) \tag{3.2}$$

Without loss of generality, male candidates have a mean vector of zero, and female candidates may differ by parameters $(\alpha, \beta^S, \beta^H)$. The correlation structure is defined by a positive semi-definite covariance matrix with parameters $(\theta, \theta^S, \theta^H)$. The correlation structure may differ by gender, where the difference is parameterized by $(\delta, \delta^S, \delta^H)$. For now, we assume that $(\delta, \delta^S, \delta^H) = 0$. Table 2 summarizes all model parameters and assumptions, and Figure 1 provides a visual representation of

Table 2: Model parameters, definitions, and assumptions

Parameter	Definition	Assumption
θ^S	Correlation between Q and Q^S; measure of how good the screening algorithm is at predicting true quality	$\theta^S \in [0, 1)$
θ^H	Correlation between Q and Q^H; measure of how good the hiring manager is at predicting true quality	$\theta^H \in [0, 1)$
θ	Correlation between Q^S and Q^H; degree to which the screening algorithm and the hiring manager agree in their quality assessment	$\theta \in [0, 1)$
τ^S	Quality cutoff for the screener to pass the candidate to the next round ($Q_S \geq \tau_S$)	—
τ^H	Quality cutoff for the hiring manager to hire a candidate ($Q_H \geq \tau_H$)	—
$\delta := \theta_m - \theta_f$	Gender difference in correlation between the screener and the hiring manager	$\delta = 0$ (for now)
$\delta^S := \theta^S_m - \theta^S_f$	Predictive gender bias of screening scores	$\delta^S = 0$
$\delta^H := \theta^H_m - \theta^H_f$	Predictive gender bias of the hiring manager scores	$\delta^H = 0$
α	Mean quality difference between men and women; positive α implies women have higher mean quality than men	$\alpha = 0$. We extend the model in Appendix A.5
β^S	Systematic gender bias of screening scores	$\beta^S = 0$
β^H	Systematic gender bias of hiring manager scores	$\beta^H = 0$

the quality score model.

3.2 Theoretical results

We analyze how the gender diversity of hires, p_h, and the expected quality of hires, $E[Q_h]$, vary as functions of the firm's design parameters with respect to the screening algorithm—θ, δ, θ^S.[5]

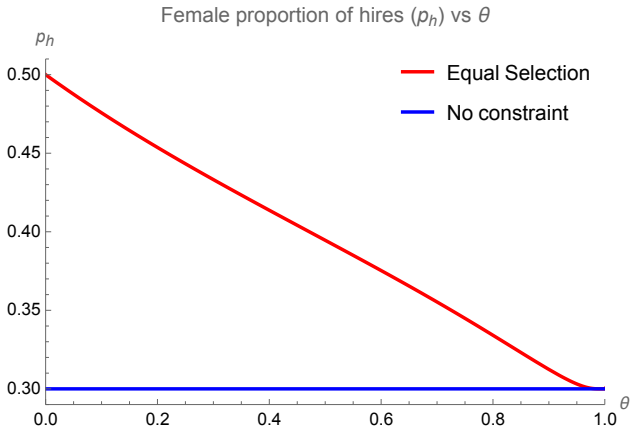
[5]Not all model parameters are design parameters that can be controlled by the firm. For a given candidate, Q is fixed, and estimation of Q^H is delegated to the hiring manager, which fixes θ^H. The firm has control over the screening algorithm, and thus how Q^S is estimated. Therefore, the design parameters that the firm can control are θ^S (i.e., how good the screening algorithm is in predicting true quality), and θ (how similar the screening algorithm is compared to the hiring manager in assessing quality).

3.2.1 Effects on hire diversity

Proposition 1. *The effectiveness of the equal selection constraint (p_h) decreases as the correlation (θ) between algorithmic scores and hiring manager scores increases.*

Corollary. *When screening and hiring manager scores are perfectly uncorrelated ($\theta = 0$), the equal selection constraint effectively balances the gender proportion of hires. In contrast, when the scores are perfectly correlated ($\theta = 1$), equal selection has no effect on the gender proportion of hires. Under partial correlation ($0 < \theta < 1$), higher values of θ lead to decreasing effectiveness of the constraint.*

Figure 2: Female proportion of hires (p_h) vs. correlation parameter (θ)



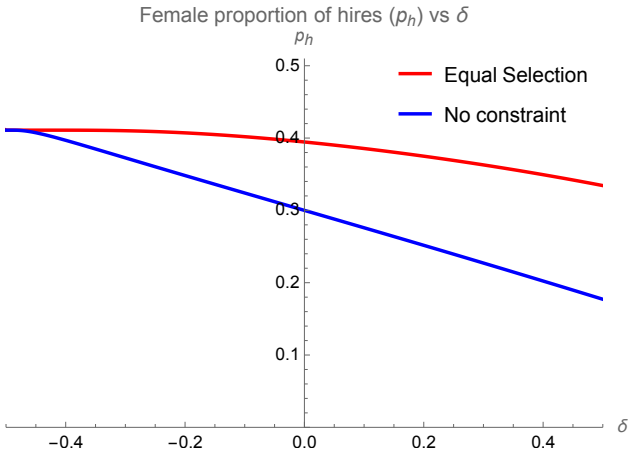
Notes: This figure plots the female proportion of hires, p_h, as a function of the correlation parameter, θ. The proportion of women in the applicant pool is fixed at $p_a = 0.3$.

A formal proof is provided in Appendix A.2. Here, we provide an intuitive explanation. Under equal selection, the female shortlist threshold (τ_f^S) is adjusted such that an equal number of women and men are shortlisted. Since there are more men than women in the applicant pool, the shortlist threshold for women will be lower compared to men ($\tau_f^S < \tau_m^S$). This means that the average Q^S score of women will be lower compared to men in the shortlist. When the screening and hiring manager scores are perfectly correlated ($\theta = 1$), $Q^S = Q^H$, this translates to lower average Q^H score for women. So, even though there are an equal number of male and female candidates in the shortlist, the shortlisted female candidates will be less likely to get hired compared to the male candidates. On the other extreme, when the two scores are perfectly uncorrelated ($\theta = 0$), the Q^S scores are independent of Q^H. Even though shortlisted female candidates have lower average Q^S score than male candidates, they have the same average Q^H score. Therefore, when $\theta = 0$, the

probability that a female is hired equals $\frac{1}{2}$. In partially correlated cases, the outcomes lie between these two extremes: the gender diversity outcomes will be worse when algorithmic and human evaluations align closely.

Proposition 2. *The female proportion of hires (p_h) decreases with the gender difference in the correlation parameter (δ).*

Figure 3: Female proportion of hires (p_h) vs. gender difference in correlation parameter (δ)



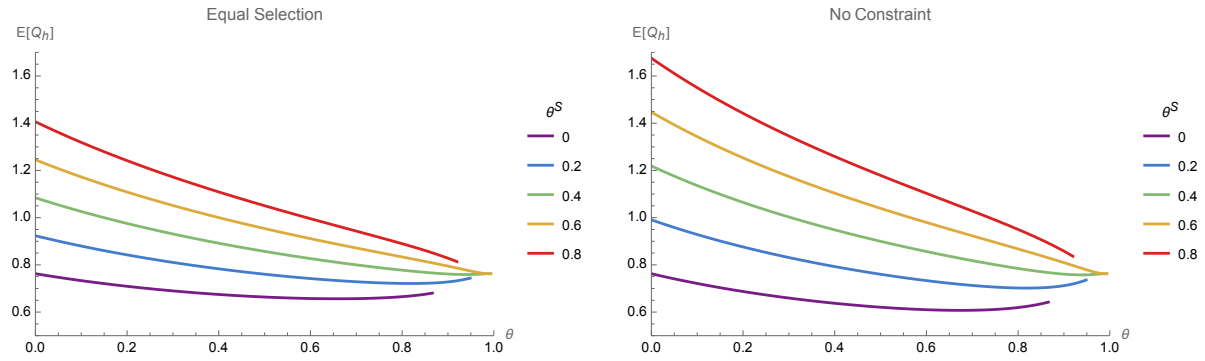
Notes: This figure plots the female proportion of hires, p_h, as a function of the gender difference in correlation parameter, δ. The proportion of women in the applicant pool is $p_a = 0.3$.

We provide the proof in Appendix A.3. The intuition is as follows: $\delta > 0$ means that the screening algorithm is less predictive of the hiring manager's evaluation for female candidates, which means lower Q^H scores for female candidates compared to men in the shortlist. This in turn means that the probability of a female being hired decreases. Therefore, any gender-specific discrepancies in evaluation consistency reduces female proportion of hires, both with and without the equal selection constraint.

3.2.2 Effects on hire quality

Proposition 3. *Conditional on the predictive accuracy of the screening algorithm (θ^S) and the hiring manager (θ^H), the average hire quality decreases as the correlation (θ) between algorithmic scores and hiring manager scores increases in the space $\theta \in [0, \min\{\frac{\theta^S}{\theta^H}, \frac{\theta^H}{\theta^S}\}]$, with hire quality reaching a global maximum at $\theta = 0$.*

Figure 4: Expected quality of hire ($E[Q_h]$) vs. θ, θ^S



Notes: This figure plots the expected quality of hires, $E[Q_h]$, as a function of the correlation parameter, θ for different θ^S values. The rest of the parameters are fixed at $\theta^H = 0.5, p_a = 0.3, \delta = 0$.

We provide the formal proof in Appendix A.4 and provide an intuitive explanation here. The screening score (Q^S) and hiring manager score (Q^H) act as two noisy signals providing information about the candidate's true quality (Q). The individual informativeness of these signals (i.e., θ^S, θ^H) is fixed. A key principle, discussed by Clemen and Winkler (1985), states that when combining signals, *less correlated* signals collectively provide more information about the underlying value than *highly correlated* signals (for a fixed level of individual signal informativeness). Thus, higher correlation leads to redundant information, reducing the overall quality gains from screening.

3.3 Implications for Algorithm Design

Our theoretical analysis offers clear guidance on designing screening algorithms to simultaneously maximize hire quality and workforce diversity. Specifically, our findings highlight three key points: (1) hire quality increases with the predictive accuracy of the screening algorithm (θ^S), (2) the proportion of female hires decreases with increased correlation (θ) between algorithmic scores and hiring manager evaluations under equal selection constraints (Proposition 1), and (3) conditional on predictive accuracy (θ^S), higher correlation (θ) reduces expected hire quality (Proposition 3). Therefore, the optimal strategy involves selecting screening algorithms with high θ^S (good at predicting true quality) but low θ (distinct from human evaluations), making the algorithms *complementary* to human assessments.

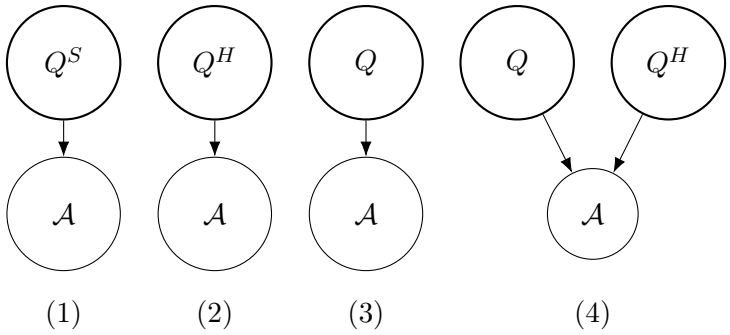


Figure 5: Target variable options for training a screening algorithm $\mathcal{A}$.

3.3.1 Algorithmic Selection with Independent Parameters

Consider a scenario where a firm chooses between two screening algorithm vendors. Both algorithms are equally accurate at predicting true quality (i.e., $\theta_1^S = \theta_2^S$) but differ in their correlation with hiring managers' evaluations (i.e., $\theta_1 \neq \theta_2$). Such differences can arise if algorithms rely on varying feature sets.

The firm should choose the algorithm with lower correlation (θ) to hiring manager assessments. Despite similar predictive performance, lower correlation algorithms offer less redundant information, thus enhancing both diversity and expected hire quality under equal selection constraints.

3.3.2 Balancing Predictive Accuracy and Managerial Complementarity

In practice, a firm often designs a screening algorithm with a fixed information source, such as resumes, creating inherent trade-offs between predictive accuracy (θ^S) and complementarity to human evaluations (θ). We outline several algorithm training strategies based on available target variables (Figure 5 illustrates these visually):

- **Option 1 (Historical Human Screener Scores):** Training on historical screening scores (Q^S) provides abundant data but has limited control over both θ^S and θ since past screening scores are proxies, not perfect predictors of true quality.
- **Option 2 (Hiring Manager Scores):** Training directly on hiring manager evaluations (Q^H) maximizes correlation (θ), diminishing the diversity benefits of equal selection.
- **Option 3 (True Quality Scores):** Training on actual job performance data (Q) maximizes predictive accuracy (θ^S) but offers no direct control over managerial correlation (θ).

- **Option 4 (Multi-objective Learning):** Training simultaneously on true quality (Q) and hiring manager evaluations (Q^H) using techniques such as adversarial learning. This balances high predictive accuracy and managerial complementarity, optimizing both diversity and hire quality simultaneously.

While these practical strategies describe how different target variables affect predictive accuracy and managerial correlation, it is also crucial to consider theoretical limitations in simultaneously optimizing these parameters when the overall information available is fixed.

We now specifically examine this scenario, where the total information regarding true candidate quality (Q), conditional on the algorithm's scores (Q^S) and the hiring manager's evaluations (Q^H), remains constant. Formally, the information given by Q^S and Q^H about Q is given by the conditional entropy $H(Q|Q^S, Q^H)$ which is:

$$H(Q|Q^S, Q^H) = \frac{1}{2} \cdot \log \left(\frac{2e\pi \cdot \text{Det} \left(\begin{bmatrix} 1 & \theta^S & \theta^H \\ \theta^S & 1 & \theta \\ \theta^H & \theta & 1 \end{bmatrix} \right)}{1 - \theta^2} \right) \tag{3.3}$$

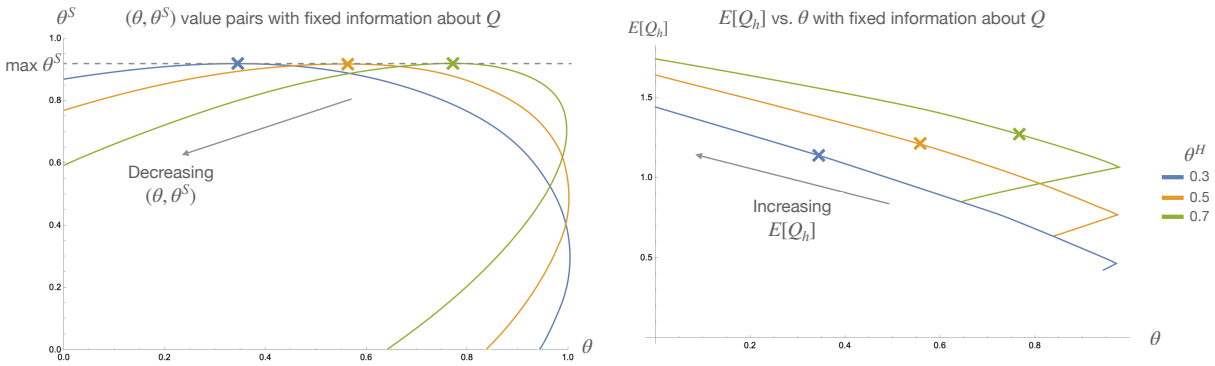
Under these fixed-information conditions, any attempt to maximize predictive accuracy (θ^S) will inherently constrain efforts to reduce correlation (θ) and vice versa. If we express θ^S as a function of θ, with fixed information H_0, we get:

$$\theta^S = \theta \cdot \theta^H \pm \sqrt{(1 - (\theta^H)^2) \cdot (1 - \theta^2) - \frac{(1 - \theta)\, e^{2H_0}}{2e\pi}} \tag{3.4}$$

Figure 6 demonstrates these trade-offs, showing pairs of (θ, θ^S) that yield constant information. Maximizing θ^S initially enhances predictive performance, but further reductions in correlation (θ) inevitably decrease θ^S. But, notably, even after reaching peak θ^S, further reducing θ can still increase expected hire quality due to the reduction in redundant information in a single stage.[6]

[6] Decreasing θ affects the expected quality of hires via two channels: (1) there is a direct effect of θ, where decreasing θ increases $E[Q_h]$, and (2) there is an indirect effect via θ^S, where decreasing θ also decreases θ^S, which in turn decreases $E[Q_h]$. Interestingly, the net effect of decreasing θ still increases $E[Q_h]$ even though θ^S is simultaneously decreasing—meaning that the direct increase in $E[Q_h]$ due to the decrease in θ offsets the indirect decrease in $E[Q_h]$ due to decreasing θ^S.

Figure 6: Simultaneously optimizing θ^S and θ with fixed information about Q



Notes: The left panel plots equal-information (θ, θ^S) pairs yielding the same conditional entropy $H(Q|Q^S, Q^H) = 0.5$. The 'x' marks the point where θ^S is maximized. The right panel plots expected hire quality $E[Q_h]$ as a function of θ using these equal-information pairs. θ^H is held fixed at $\theta^H \in 0.3, 0.5, 0.7$.

3.3.3 Directly Minimizing Gender Differences in Q^H Scores

Another approach is to directly minimize gender differences in Q^H scores in the shortlist. This can be implemented by training two predictors—one for Q and another for Q^H—and then shortlisting candidates with high predicted Q scores while minimizing gender differences in predicted Q^H scores.

This approach deliberately selects male and female candidates with similar hiring manager scores, maximizing the likelihood of gender-balanced hiring. While not minimizing θ to zero (and thus not maximizing hire quality to the fullest extent), this method effectively balances diversity and quality goals while being simpler to implement than adversarial approaches.

In summary, the firm's goal in algorithm design should be to create a screening system that complements, rather than replicates, managerial evaluations, thereby maximizing both quality and diversity outcomes under equal selection constraints.

4 Data and Empirical Methodology

Our theoretical analysis demonstrates that the effectiveness of the equal selection constraint depends significantly on parameters such as the correlation between screening algorithms and hiring manager evaluations (θ), and gender differences in these evaluations (δ). The optimal screening algorithm is one that is trained on both true quality and the hiring manager's assessment of quality. However, in practice, screening algorithms are typically trained on historical human screening decisions rather

than true job performance, primarily due to data availability.[7] This raises an empirical question: if firms train screening algorithms based on historical recruiter decisions, how effective will equal selection constraints be in improving diversity outcomes?

In this section, we describe our empirical approach to addressing this question. We estimate model parameters using actual hiring data from multiple firms, and then use these parameters to estimate the effectiveness of equal selection constraints across different job contexts. We also benchmark these outcomes against our proposed complementary screening algorithm and other commonly used fairness metrics using simulation.

4.1 Data Description

We use Applicant Tracking System (ATS) data from eight U.S.-based technology companies, provided by an HR analytics software vendor. This dataset includes detailed records for 799,000 external job applicants (60% male, 40% female) across 3,608 unique job postings. Each record captures candidate attributes (such as gender and experience), resumes, job posting details, and outcomes at each hiring stage (screening, first interview, subsequent interviews, and offers).

Table 3 summarizes the number of applicants and job postings by job category. Although specific hiring processes can vary slightly across firms, the typical hiring sequence for external applicants, as shown in Figure 7, involves four main stages: Screening, First Interview, Subsequent Interviews, and Offer. Initially, applicants undergo a screening stage. Those who pass the screening proceed to the first interview stage, followed by subsequent interviews, and finally receive an offer if selected.[8] On average, a typical job posting attracts 233 applicants, with about 36 advancing past the initial screening, approximately 7 candidates progressing beyond the first interview, and around 2 receiving offers.

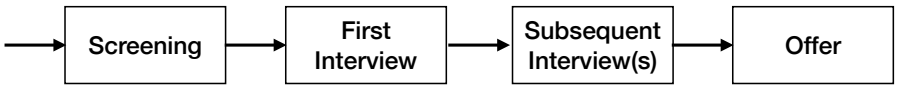
[7] For example, *LinkedIn Recruiter's* recommendation algorithm is trained on the human recruiter's decisions since it has no visibility into the true job performance of the candidates.

[8] For our empirical analysis, we specifically focus on two critical stages: the initial screening stage—where the equal selection constraint is applied—and the subsequent first interview stage. Although the actual hiring process is multi-staged, this simplified focus remains appropriate. To illustrate, consider a scenario with an applicant pool comprising 70% males and 30% females. With an equal selection constraint, the shortlisted candidates following the screening stage would consist of an equal gender split (50/50). However, the hiring manager in the first interview stage might partially reverse this constraint, resulting in a gender ratio such as 60/40. Provided that selections in subsequent stages are unbiased—a central assumption in our theoretical model—this revised gender ratio would persist throughout the remainder of the hiring process.

Table 3: Number of applicants and job postings by job category

Job Category	N Applicants	N Jobs
Engineering & Technical	214,943	1,178
Product & Design	130,669	534
Sales & Marketing	92,559	391
Legal & PR	75,955	332
Other	70,864	53
Finance & Accounting	69,536	308
Biz Dev & Operations	51,523	299
Human Resources	48,122	246
Customer Service & Acct Management	42,199	238
Overall	799,108	3,608

Figure 7: Hiring funnel



4.2 Empirical Strategy

Our empirical approach consists of three key steps:

Step 1: Estimating Screening and Hiring Manager Scores. The Applicant Tracking System (ATS) provides only binary outcomes (screening and interview decisions). To analyze the effectiveness of equal selection constraints, we first derive continuous quality scores for both the screening and hiring manager stages. We achieve this by training two machine learning (ML) models separately: one predicting screening decisions and the other predicting hiring manager decisions, using candidate resume texts and job descriptions as inputs. Further technical details about these models are discussed in Section 4.3.[9]

We employ BigBird, a transformer model specifically optimized for processing long text documents (Zaheer et al. 2020). To address selection bias arising from observing hiring manager decisions only for shortlisted candidates, we apply inverse propensity weighting to re-weight observations based on their probability of passing the screening stage. This provides unbiased estimates of hiring

[9]Note that we observe binary screening decisions and not the assessed scores by the screener and we observe binary hiring manager decisions only for shortlisted candidates. By building the ML models, we can infer continuous quality scores for all candidates and all stages.

manager evaluations for all candidates.

Predicted decision probabilities from the ML models are converted into quality scores using a Gaussian copula transformation, aligning with our theoretical assumption of multivariate Gaussian distributions for quality scores.[10] We show in Appendix C that the Gaussian copula has good goodness-of-fit measures on our empirical data compared to other copulas. The predicted probabilities are then transformed into quality scores via quantiles:

$$\hat{q}^S_{i,j} = Quantile(\hat{p}^S_{i,j}, \hat{\boldsymbol{p}}^S_j), \tag{4.1}$$

$$\hat{q}^H_{i,j} = Quantile(\hat{p}^H_{i,j}, \hat{\boldsymbol{p}}^H_j). \tag{4.2}$$

Step 2: Parameter Estimation (θ and δ). Using the recovered continuous quality scores, we estimate the critical parameters:

- **Correlation Parameter (θ)**: We estimate the correlation between screening scores and hiring manager evaluations for each job posting using the Spearman rank correlation coefficient. This measure is robust to transformations and widely used in practice.[11]

$$\hat{\theta} j = \text{Spearman}(\hat{\boldsymbol{q}} j^S, \hat{\boldsymbol{q}}^H_j). \tag{4.3}$$

- **Gender Difference Parameter (δ)**: We calculate the difference in correlation between male and female candidates' evaluations for each job posting.

These parameters are aggregated across job postings, weighted by the number of applicants per job, to derive representative average values. The rest of the parameters (p_a, τ^S, τ^H), are observed directly from the data.[12]

[10] Gaussian copulas are multivariate Gaussian distributions, whose marginals are uniformly distributed. They offer a flexible way to disentangle multivariate Gaussian distribution as a product of uniform marginal distributions and a Gaussian copula that "couples" them (See Joe (2014) and Nelsen (2007) for a reference on copulas). Formally, the joint empirical distribution of quality scores $(\hat{q}, \hat{q}^S, \hat{q}^H)$ has CDF $F_{\hat{q},\hat{q}^S,\hat{q}^H}(x, y, z; \Sigma) = C(F_{\hat{q}}(x), F_{\hat{q}^S}(y), F_{\hat{q}^H}(z))$. Here, C is the 3-dimensional Gaussian copula, $C(u, v, k) = \Phi(\Phi^{-1}(u), \Phi^{-1}(v), \Phi^{-1}(k))$, and Φ is the CDF of a multivariate Gaussian distribution. This transformation ensures that we stay close to the theoretical model, which assumes that the quality scores have a multivariate normal distribution.

[11] Using pearson correlation leads to highly similar results.

[12] We set the job-specific shortlist and hiring manager thresholds based on the actual size of the shortlist and finalist observed in the data. In doing so, we conceive the thresholds as exogenous variables. For example, the firm may have a limited budget to interview candidates and can only afford to interview a certain number of candidates. The shortlist and hiring manager threshold is therefore set based on the observed size of the shortlist and finalist respectively.

Step 3: Counterfactual Policy Simulation. Using the estimated parameters, we conduct counterfactual simulations to evaluate the effectiveness of equal selection constraints in enhancing workforce diversity. We compare the performance of these constraints against our proposed complementary screening algorithm and other widely adopted fairness criteria, allowing us to empirically demonstrate their relative effectiveness.

4.3 ML Model Details and Performance

To train the screening and hiring manager models, we consolidate each candidate's resume with relevant job information—such as company name, job title, business unit, employment type, location, skills, and keywords—into a unified input document. The skills and keywords are sourced from a comprehensive skills dictionary developed through an extensive analysis of LinkedIn profile data.

We partition the dataset into training (80%), validation (10%), and hold-out test sets (10%), stratifying by job postings to ensure representativeness and robust model evaluation. We follow Sun et al. (2019) for picking the optimal hyperparameters and select them based on validation performance (area under the ROC curve): `Epochs=3, Batch Size=14, Learning Rate=2e-5, Weight Decay=2e-5`.

The screening model training/evaluation set consists of 725,351 observations, with a hold-out test set of 73,757 observations. The hiring manager model training/evaluation set includes 106,419 observations, with a hold-out test set comprising 11,357 observations.

Model evaluation indicates strong predictive performance: the screening model achieves an AUC score of 0.83, while the hiring manager model achieves an AUC of 0.68. We see no significant differences in predictive performance between male and female candidates. Additional model performance details can be found in Appendix B.

4.4 Inverse Propensity Weighting

Since hiring manager decisions are only available for candidates who have passed the screening stage, we employ inverse propensity weighting to mitigate selection bias. Specifically, candidates less likely to be shortlisted (based on screening predictions) receive higher weights, while those more likely receive lower weights. This adjustment ensures unbiased estimation of hiring manager scores across the full applicant pool as long as there is noise in the selection process (see Cowgill (2020)).

Figure 8: Distribution of parameter estimates across job postings

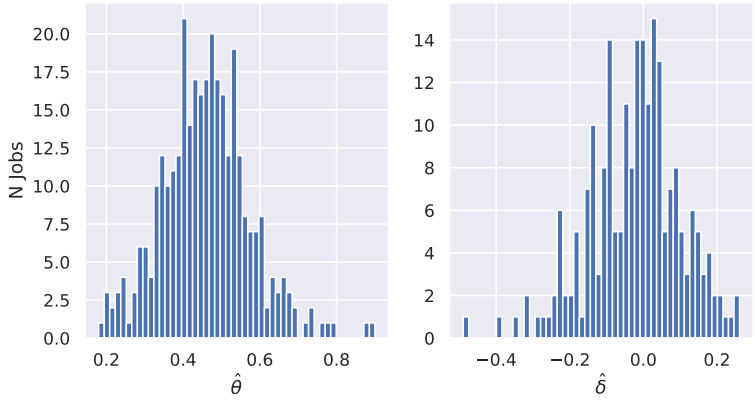


Table 4: Average parameter estimates

Job Category	$\hat{\theta}$	$\hat{\delta}$
Finance & Accounting	0.493	-0.019
Engineering & Technical	0.449	-0.009
Sales & Marketing	0.442	0.003
Product & Design	0.441	0.031
Customer Service & Acct Management	0.436	-0.02
Biz Dev & Operations	0.413	-0.014
HR	0.403	-0.074
Legal & PR	0.348	-0.016
Other	0.245	-0.332
Average	0.434	-0.007

Notes: This table reports the average parameter estimates for each job category. We estimate the parameters at the job posting level and aggregate it up to the job category level for all jobs in the hold-out test set.

5 Empirical Results

This section presents empirical findings based on our analysis of the hold-out test set.

5.1 Parameter estimates

We estimate model parameters (θ, δ) separately for each job posting, using the methodology in Section 4.2, and plot their distribution in Figure 8. Table 4 reports average parameter values aggregated by job category.

The average correlation parameter estimate, $\hat{\theta}$, is 0.43, but varies considerably across jobs.

Higher correlation is typically observed in technical roles requiring "hard skills" (e.g., Finance & Accounting, Engineering & Technical). In contrast, lower correlation occurs in roles emphasizing "soft skills" (e.g., HR, Legal & PR). A likely explanation is that hard skills are more readily assessable from resumes, thus aligning screener and hiring manager evaluations more closely.

Regarding gender differences, the overall average estimate for δ is -0.007, suggesting that, on average, the screening criteria align similarly for men and women.[13] But, there is substantial variability across job postings, with estimates ranging from -0.4 to 0.2.

5.2 Effectiveness of the Equal Selection Constraint

We estimate the impact of equal selection constraints using counterfactual simulations. For each job posting with underrepresented female applicants ($p_a < 0.5$), we simulate outcomes using the estimated job-specific parameters.

Table 6 reports the aggregate results by job category. The equal selection constraint raises the proportion of women from an average of 31% in the applicant pool to 50% in the shortlist (by design). However, this proportion decreases to 41% among finalists, showing a modest overall improvement.

The effectiveness varies notably across job categories. For example, Engineering & Technical roles achieve only a 36% representation of women in the finalist stage, indicating limitations in equal selection effectiveness in highly technical fields.

5.3 Test of propositions

We next exploit the variation in parameter estimates across jobs to empirically test Proposition 1 and Proposition 2 using the following regression specification:

$$p_{h,j} = \beta_0 + \beta_1 p_{a,j} + \beta_2 \theta_j + \beta_3 \theta_j^2 + \beta_4 \delta_j + \beta_5 \delta_j^2 + \epsilon_j \tag{5.1}$$

where $p_{h,j}$ is the estimated proportion of women in the finalist pool for job j, $p_{a,j}$ is the observed proportion of women in the applicant pool for job j, θ_j and δ_j are the estimated correlation and

[13] Note that the "Other" category has a high estimate of δ, but this is likely due to the small sample size (only 214 applicants in this category in the hold-out test set).

Table 5: Empirical test of propositions

Model:	No Constraint (1)	Equal Selection (2)
Variables		
p_a	1.359***	1.023***
	(0.1402)	(0.1400)
θ	-0.2409	-0.7005**
	(0.2757)	(0.2754)
θ^2	0.2035	0.6057**
	(0.2840)	(0.2837)
δ	-0.3333***	-0.2896***
	(0.0635)	(0.0634)
δ^2	0.0626	0.0102
	(0.0858)	(0.0857)
Constant	0.0089	0.2784***
	(0.0859)	(0.0858)
Fit statistics		
Observations	254	254
R^2	0.33790	0.24751

Standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: This table reports the OLS estimates of specification 5.1 with (Model (2)) and without the equal selection constraint (Model (2)). The outcome variable is the proportion of women in the finalist pool, p_h. The independent variables are the proportion of women in the applicant pool, p_a, the correlation between screening and hiring manager scores, θ, and the gender difference in correlation, δ.

gender difference parameters for job j respectively. We include the squared terms of θ_j and δ_j to capture non-linear relationships.

We estimate the model using the hold-out test set with and without the equal selection constraint and report the results in Table 5.

Consistent with the theoretical predictions, we find a negative relationship between θ and p_h under the equal selection constraint and a negative relationship between δ and p_h with and without the equal selection constraint.

5.4 Benchmarking Screening Algorithms and Fairness Constraints

We benchmark the equal selection constraint against other fairness criteria and our proposed complementary screening approach. Specifically, we evaluate seven screening algorithms:

Table 6: Estimated effectiveness of the equal selection constraint

Job Category	Equal Selection	Applied p_a	Screened p_s	Hired p_h
Biz Dev & Operations	False	0.38	0.38	0.38
	True	0.38	0.50	0.45
Customer Service & Acct Management	False	0.38	0.38	0.38
	True	0.38	0.50	0.47
Engineering & Technical	False	0.23	0.23	0.23
	True	0.23	0.50	0.36
Finance & Accounting	False	0.35	0.35	0.35
	True	0.35	0.50	0.45
HR	False	0.41	0.41	0.41
	True	0.41	0.50	0.46
Legal & PR	False	0.35	0.35	0.35
	True	0.35	0.50	0.44
Product & Design	False	0.33	0.33	0.33
	True	0.33	0.50	0.43
Sales & Marketing	False	0.36	0.36	0.36
	True	0.36	0.50	0.44
Overall	False	0.31	0.31	0.31
	True	0.31	0.50	0.41

Notes: This table reports the proportion of women in the applicant pool p_a, shortlist p_s and hired pool p_h — with and without the equal selection constraint. p_a is observed in the data. p_s and p_h are estimated by first estimating the job-specific model parameters $(\hat{\theta}_j, \hat{\delta}_j)$, and imputing the model parameters into the theoretical model. We estimate at the job posting level and aggregate it up to the job category level for all jobs with $p_a < 0.5$ in the hold-out test set.

1. No Constraint: Baseline without fairness constraints.
2. Equal Selection: Equal gender representation in the shortlist.
3. Demographic Parity: Shortlist gender proportions match applicant pool proportions.
4. Error Rate Parity: Equal error rates for men and women.[14]
5. Equalized Odds: Equal true and false positive rates across genders.
6. Equal Selection min Q^S Diff: Equal selection while minimizing gender differences in screening scores.
7. Complementary Equal Selection: Equal selection while minimizing gender differences in hiring manager scores.

We assess the impact of each screening algorithm on both diversity and expected hire quality through agent-based hiring simulations. In these simulations, ML models for screening and hiring

[14]We use fairlearn's implementation https://fairlearn.org/v0.10/user_guide/mitigation/reductions.html

Table 7: Screening algorithms

Screening Algorithm	**Constraint**
No Constraint	None
Equal Selection	$\mathbb{P}(g = f \mid \hat{y}^S = 1) = \mathbb{P}(g = m \mid \hat{y}^S = 1)$
Demographic Parity	$\mathbb{P}(\hat{y}^S \mid g = f) = \mathbb{P}(\hat{y}^S \mid g = m)$
Error Rate Parity	$\mathbb{P}(\hat{y}^S \neq y^S \mid g = f) = \mathbb{P}(\hat{y}^S \neq y^S \mid g = m)$
Equalized Odds	$\mathbb{P}(\hat{y}^S = 1 \mid y^S, g = f) = \mathbb{P}(\hat{y}^S = 1 \mid y^S, g = m),$ $y^S \in \{0, 1\}$
Equal Selection min Q^S Diff.	$\mathbb{P}(g = f \mid \hat{y}^S = 1) = \mathbb{P}(g = m \mid \hat{y}^S = 1)$ $\min \mathbb{E}[\hat{q}_s^S \mid g = f] - \mathbb{E}[\hat{q}_s^S \mid g = m]$
Complementary Equal Selection	$\mathbb{P}(g = f \mid \hat{y}^S = 1) = \mathbb{P}(g = m \mid \hat{y}^S = 1)$ $\min \mathbb{E}[\hat{q}_s^H \mid g = f] - \mathbb{E}[\hat{q}_s^H \mid g = m]$

Notes: This table summarizes the screening algorithms used for benchmarking. $\hat{y}^S$ is the predicted screening outcome, and y^S is the true screening outcome observed in the data. $\hat{q}_s^S$ is the predicted screening score of the shortlisted candidates, and $\hat{q}_s^H$ is the predicted hiring manager score of the shortlisted candidates. The primary objective of all the algorithms is to shortlist candidates with the highest screening score.

manager evaluations serve as agents. The screening agent shortlists the candidates with the highest screening scores, $\hat{q}^S$, while satisfying the constraint outlined in Table 7, and the hiring manager agent selects the candidates with the highest hiring manager scores, $\hat{q}^H$. Unlike the theoretical model, these simulations do not assume identical quality distributions for men and women, unbiased hiring managers, or normal distributions, making them robust to potential real-world deviations.

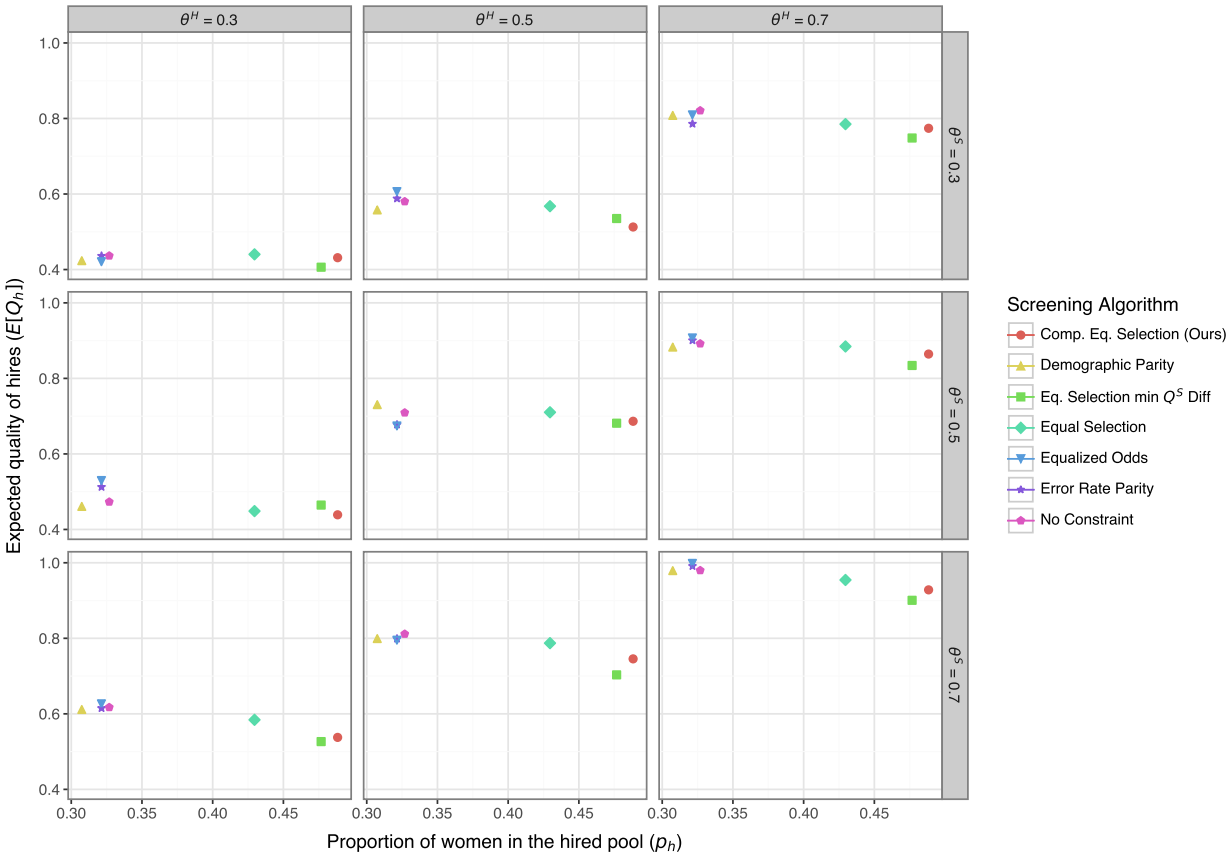
Simulating true quality Q. To measure the expected hire quality (unobservable in actual data), we generate semi-synthetic quality scores for candidates based on different assumed values of θ^S and θ^H, while holding the empirically estimated correlation θ fixed.[15] We present the results from these simulations in Figure 9.

Our results demonstrate significant variability in the diversity outcomes among the algorithms:

- The Complementary Equal Selection algorithm consistently achieves the highest diversity of hires, surpassing all other constraints, including the Equal Selection min Q^S Diff. This performance advantage arises because it directly aligns shortlisted candidates to be

[15] For each job we observe the vectors $\boldsymbol{q}_j^S$ and $\boldsymbol{q}_j^H$, which fixes θ. To generate $\boldsymbol{q}$, we first generate a random vector. We then orthogonalize it with respect to $\boldsymbol{q}_j^S$ and $\boldsymbol{q}_j^H$. We then transform the vector for a given value for θ^S and θ^H. This produces a random vector $\boldsymbol{q}_j$ that has the defined correlation structure Σ.

Figure 9: Quality and diversity of hires using different screening algorithms



Notes: This figure plots the expected quality of hires (y-axis) and the proportion of women in the hired pool (x-axis) for different screening algorithms using agent-based hiring simulation experiments. We use semi-synthetic data for true quality Q. Each grid in the facet corresponds to a different θ^S and θ^H value. Error bars in both the x and y axes represent bootstrapped 95% Error bars are not visible because they are narrow.

complementary to the hiring manager's evaluation criteria, ensuring greater diversity in the actual hires.

- Algorithms based on Demographic Parity, Error Rate Parity, and Equalized Odds do not substantially enhance diversity compared to the baseline. Demographic Parity actually reduces diversity since, empirically, women had higher shortlisting rates in our training data (see Appendix C.2), a disparity eliminated by enforcing equal representation with the applicant pool. Likewise, Error Rate Parity and Equalized Odds have limited effect because our ML models exhibit minimal to no gender differences in error rates or ROC curves (as confirmed in Appendix C.2 and Appendix B).

- While the Complementary Equal Selection approach significantly improves diversity, it does so with minimal reduction in the expected quality of hires, particularly under lower values of θ^S and θ^H.

6 Discussion and Conclusion

This paper examines the effectiveness of diversity policies implemented as algorithmic fairness constraints within Human+AI hiring systems. We develop a theoretical model of the hiring process, showing that the success of a common diversity policy—equal selection in the shortlist—is contingent upon key parameters such as the correlation between the screening algorithm's and the hiring manager's assessment criteria. Using real hiring data from technology firms, we empirically estimate these parameters and evaluate the impact of equal selection through counterfactual policy simulations.

Our findings indicate that enforcing equal selection in shortlists modestly improves gender diversity among hires but does not achieve parity. Moreover, the effectiveness of this constraint varies significantly across job categories. To address these limitations, we propose a complementary screening algorithm, designed explicitly to differ from the hiring manager's assessments, and demonstrate its superior performance in enhancing workforce diversity compared to traditional fairness constraints.

We highlight several critical managerial and algorithmic design implications arising from our results:

- First, achieving gender parity at the shortlist stage does not inherently guarantee gender parity in final hires, even if hiring managers are gender-unbiased. Without this understanding, stakeholders may mistakenly interpret the post-shortlist disparities as biases introduced by hiring managers, undermining trust in the fairness policy.

- Second, the effectiveness of the equal selection constraint is highly job-specific, driven by the correlation between screener and hiring manager evaluations. Notably, technical roles requiring measurable "hard skills" (e.g., software engineering) tend to exhibit higher correlations, diminishing the effectiveness of equal selection precisely in fields where women are most

underrepresented.

- Third, equal predictive accuracy of screening algorithms across genders is insufficient in multi-stage hiring processes. It is equally important for screening algorithms to maintain gender neutrality concerning their alignment with hiring managers' criteria—specifically, algorithms should exhibit no gender differences in their correlation with managerial assessments ($\delta = 0$).

- Lastly, we show theoretically that higher correlations between screeners' and hiring managers' assessments not only reduce the effectiveness of equal selection constraints but also negatively affect the expected quality of hires. This suggests a critical design insight: screening algorithms should be constructed to complement, rather than replicate, managerial evaluations.

Limitations and Future Directions

Our recommendation—that screening algorithms should complement hiring managers' assessments-may encounter practical organizational challenges. Hiring managers often perceive AI tools as substitutes to replicate their decision-making processes. Previous literature has identified similar tensions in algorithmic hiring contexts (van den Broek et al. 2021). Future research could explore strategies for effectively integrating AI tools explicitly designed to complement, rather than duplicate, human judgment.

Another limitation of our model is the assumption that hiring managers do not adapt their decision-making in response to diversity constraints. Future studies should investigate whether introducing fairness constraints could inadvertently induce biases among previously unbiased hiring managers. Existing psychology and management literature highlights that affirmative action programs (AAPs) can lead to stigma and stereotyping of minority candidates, even those not directly benefiting from AAPs (Heilman et al. 1997; Leslie et al. 2013). Given that algorithmic fairness constraints might be perceived similarly to AAPs, further research is needed to understand such policies' potential psychological and organizational impacts.

References

Bapna, Sofia, Alan Benson, and Russell Funk (Oct. 2021). *Rejection Communication and Women's Job-Search Persistence*. SSRN Scholarly Paper. Rochester, NY.

Blum, Avrim, Kevin Stangl, and Ali Vakilian (June 20, 2022). "Multi Stage Screening: Enforcing Fairness and Maximizing Efficiency in a Pre-Existing Pipeline". In: *2022 ACM Conference on Fairness, Accountability, and Transparency*. FAccT '22. New York, NY, USA: Association for Computing Machinery, pp. 1178–1193.

Bower, Amanda et al. (July 2017). "Fair Pipelines". In: *Workshop on Fairness, Accountability, and Transparency in Machine Learning*. arXiv. arXiv: 1707.00391 [cs, stat].

Brands, Raina A. and Isabel Fernandez-Mateo (Sept. 1, 2017). "Leaning Out: How Negative Recruitment Experiences Shape Women's Decisions to Compete for Executive Roles". In: *Administrative Science Quarterly* 62.3, pp. 405–442.

Celis, L. Elisa et al. (Mar. 3, 2021). "The Effect of the Rooney Rule on Implicit Bias in the Long Term". In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. FAccT '21. New York, NY, USA: Association for Computing Machinery, pp. 678–689.

Chouldechova, Alexandra (June 1, 2017). "Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments". In: *Big Data* 5.2, pp. 153–163.

Clemen, Robert T. and Robert L. Winkler (Apr. 1985). "Limits for the Precision and Value of Information from Dependent Sources". In: *Operations Research* 33.2, pp. 427–442.

Cowgill, Bo (2020). "Bias and Productivity in Humans and Algorithms: Theory and Evidence from Re´sume´ Screening".

Dwork, Cynthia, Moritz Hardt, et al. (Jan. 8, 2012). "Fairness through Awareness". In: *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*. ITCS '12. New York, NY, USA: Association for Computing Machinery, pp. 214–226.

Dwork, Cynthia and Christina Ilvento (2019). "Fairness Under Composition". In: *LIPIcs, Volume 124, ITCS 2019* 124, 33:1–33:20. arXiv: 1806.06122 [cs, stat].

Facebook (2021). *Facebook Diversity Efforts*. https://about.fb.com/news/2021/07/facebook-diversity-report-2021/. Accessed: 2025-04-01.

Fershtman, Daniel and Alessandro Pavan (Mar. 1, 2021). ""Soft" Affirmative Action and Minority Recruitment". In: *American Economic Review: Insights* 3.1, pp. 1–18.

Geyik, Sahin Cem, Stuart Ambler, and Krishnaram Kenthapadi (Apr. 30, 2019). "Fairness-Aware Ranking in Search & Recommendation Systems with Application to LinkedIn Talent Search". arXiv: 1905.01989 [cs].

Hardt, Moritz, Eric Price, and Nathan Srebro (Oct. 7, 2016). "Equality of Opportunity in Supervised Learning". arXiv: 1610.02413 [cs].

Heilman, Madeline E., Caryn J. Block, and Peter Stathatos (June 1, 1997). "The Affirmative Action Stigma Of Incompetence: Effects Of Performance Information Ambiguity". In: *Academy of Management Journal* 40.3, pp. 603–625.

Huet, Ellen (Jan. 10, 2017). "Facebook's Hiring Process Hinders Its Effort to Create a Diverse Workforce". In.

Jiang, Xiangyu, Yucong Dai, and Yongkai Wu (June 2023). "Fair Selection through Kernel Density Estimation". In: *2023 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8.

Joe, Harry (June 26, 2014). *Dependence Modeling with Copulas*. CRC Press. 483 pp. Google Books: 09ThAwAAQBAJ.

Khalili, Mohammad Mahdi, Xueru Zhang, and Mahed Abroshan (2021). "Fair Sequential Selection Using Supervised Learning Models". In: *Advances in Neural Information Processing Systems*. Vol. 34. Curran Associates, Inc., pp. 28144–28155.

Kleinberg, Jon, Sendhil Mullainathan, and Manish Raghavan (Nov. 17, 2016). "Inherent Trade-Offs in the Fair Determination of Risk Scores". arXiv: 1609.05807 [cs, stat].

Kleinberg, Jon and Manish Raghavan (Jan. 4, 2018). "Selection Problems in the Presence of Implicit Bias". arXiv: 1801.03533 [cs, stat].

Lee, Logan M. and Glen R. Waddell (Apr. 1, 2021). "Diversity and the Timing of Preference in Hiring Decisions". In: *Journal of Economic Behavior & Organization* 184, pp. 432–459.

Leslie, Lisa M., David M. Mayer, and David A. Kravitz (July 23, 2013). "The Stigma of Affirmative Action: A Stereotyping-Based Theory and Meta-Analytic Test of the Consequences for Performance". In: *Academy of Management Journal* 57.4, pp. 964–989.

Mitchell, Shira et al. (2021). "Algorithmic Fairness: Choices, Assumptions, and Definitions". In: *Annual Review of Statistics and Its Application* 8.1, pp. 141–163.

Nelsen, Roger B. (June 10, 2007). *An Introduction to Copulas*. Springer Science & Business Media. 277 pp. Google Books: yexFAAAAQBAJ.

NFL Operations (2003). *The Rooney Rule*. https://operations.nfl.com/inside-football-ops/diversity-inclusion/the-rooney-rule/. Accessed: 2025-04-01.

Patreon (2017). *Patreon Culture Deck*. https://www.slideshare.net/TarynArnold/patreon-culture-deck-april-2017. Accessed: 2025-04-01.

Peng, Andi et al. (Oct. 28, 2019). "What You See Is What You Get? The Impact of Representation Criteria on Human Bias in Hiring". In: *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* 7.1 (1), pp. 125–134.

Pinterest (2015). *Our plan for a more diverse Pinterest*. https://newsroom-archive.pinterest.com/our-plan-for-a-more-diverse-pinterest-2015. Accessed: 2025-04-01.

Raghavan, Manish et al. (Jan. 27, 2020). "Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices". In: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. FAT* '20. New York, NY, USA: Association for Computing Machinery, pp. 469–481.

Schuck, Peter H. (2002). "Affirmative Action: Past, Present, and Future". In: *Yale Law & Policy Review* 20.1, pp. 1–96.

Shi, Wei et al. (2018). "The Adoption of Chief Diversity Officers among S&P 500 Firms: Institutional, Resource Dependence, and Upper Echelons Accounts". In: *Human Resource Management* 57.1, pp. 83–96.

Storvik, Aagoth Elise and Pål Schøne (2008). "In Search of the Glass Ceiling: Gender and Recruitment to Management in Norway's State Bureaucracy1". In: *The British Journal of Sociology* 59.4, pp. 729–755.

Sühr, Tom, Sophie Hilgard, and Himabindu Lakkaraju (July 21, 2021). "Does Fair Ranking Improve Minority Outcomes? Understanding the Interplay of Human and Algorithmic Biases in Online Hiring". In: *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. AIES '21. New York, NY, USA: Association for Computing Machinery, pp. 989–999.

Sun, Chi et al. (2019). "How to Fine-Tune BERT for Text Classification?" In: *Chinese Computational Linguistics*. Ed. by Maosong Sun et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, pp. 194–206.

Tallis, Georges M (1961). "The moment generating function of the truncated multi-normal distribution". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 23.1, pp. 223–229.

Tong, Yung Liang (2012). *The multivariate normal distribution*. Springer Science & Business Media.

Van den Broek, Elmira, Anastasia Sergeeva, and Marleen Huysman (Sept. 2021). "When the Machine Meets the Expert: An Ethnography of Developing AI for Hiring". In: *MIS Quarterly* 45.3, pp. 1557–1580.

Zafar, Muhammad Bilal, Isabel Valera, Manuel Gomez Rodriguez, et al. (Apr. 3, 2017). "Fairness Beyond Disparate Treatment & Disparate Impact: Learning Classification without Disparate

Mistreatment". In: *Proceedings of the 26th International Conference on World Wide Web*. WWW '17. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, pp. 1171–1180.

Zafar, Muhammad Bilal, Isabel Valera, Manuel Gomez Rodriguez, et al. (Nov. 28, 2017). "From Parity to Preference-based Notions of Fairness in Classification". arXiv: 1707.00010 [cs, stat].

Zaheer, Manzil et al. (2020). "Big Bird: Transformers for Longer Sequences". In: Neural Information Processing Systems (NeurIPS).

Zemel, Rich et al. (May 26, 2013). "Learning Fair Representations". In: *International Conference on Machine Learning*. International Conference on Machine Learning. PMLR, pp. 325–333.

Appendix

A Proofs

A.1 Setup, Definitions, and Key Identities

We use the following definitions and identities throughout the proofs.

Setup Applicants are evaluated using the screening score (Q^S) in the first stage, and the hiring score (Q^H) in the second. Unless otherwise specified, we assume that $(\delta, \delta^S, \delta^H, \alpha, \beta^S, \beta^H) = 0$; the distribution of the scores are thus the same for men and women before any selection.

$$(Q, Q^S, Q^H) \sim \mathcal{N}\left(\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & \theta^S & \theta^H \\ \theta^S & 1 & \theta \\ \theta^H & \theta & 1 \end{bmatrix}\right) \tag{A.1}$$

Definitions

1. $\phi(x)$ is the standard normal Probability Density Function (PDF). Q, Q^S, Q^H have marginal PDFs $\phi(q), \phi(q^S), \phi(q^H)$.

2. $\Phi(x)$ is the standard normal Cumulative Distribution Function (CDF). Q, Q^S, Q^H have marginal CDFs $\Phi(q), \Phi(q^S), \Phi(q^H)$.

3. $\phi_2(x, y; \rho)$ is the standard bivariate normal PDF evaluated at (x, y) with correlation ρ. Q^S, Q^H have joint PDF $\phi_2(q^S, q^H; \theta)$.

4. $\Phi_2(x, y; \rho)$ is the standard bivariate normal CDF, representing $P(X \leq x, Y \leq y)$ where (X, Y) have standard normal marginal distributions with correlation ρ. Q^S, Q^H have joint CDF $\Phi_2(q^S, q^H; \theta)$.

5. $\overline{\Phi}_2(x, y; \rho)$ is the complement of the standard bivariate normal CDF (tail distribution), representing $P(X > x, Y > y)$ where (X, Y) have standard normal marginal distributions with correlation ρ. The probability that a candidate is hired is therefore given by $\overline{\Phi}_2(\tau^S, \tau^H; \theta)$.

Key Identities

1. Plackett's Identity for the derivative of the bivariate normal CDF with respect to correlation (see Tong (2012)):
$$\frac{\partial}{\partial \rho}\Phi_2(a, b; \rho) = \phi_2(a, b; \rho)$$

2. Property of the bivariate normal PDF:
$$\phi_2(-x, -y; \rho) = \phi_2(x, y; \rho)$$
This is because the exponent $-\frac{(-x)^2 - 2\rho(-x)(-y) + (-y)^2}{2(1-\rho^2)} = -\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}$ is unchanged.

3. Relation between univariate and bivariate normal PDFs:
$$\phi(x)\phi\left(\frac{y - \rho x}{\sqrt{1-\rho^2}}\right) = \sqrt{1-\rho^2}\phi_2(x, y; \rho)$$
And similarly, by symmetry of $\phi_2(x, y; \rho) = \phi_2(y, x; \rho)$:
$$\phi(y)\phi\left(\frac{x - \rho y}{\sqrt{1-\rho^2}}\right) = \sqrt{1-\rho^2}\phi_2(x, y; \rho)$$

4. Symmetry of the univariate normal PDF: $\phi(-x) = \phi(x)$.

5. Symmetry relating CDF and tail distribution: $\overline{\Phi}_2(x, y; \rho) = \Phi_2(-x, -y; \rho)$.

A.2 Proof of Proposition 1

Proposition. *The effectiveness of the equal selection constraint (p_h) decreases as the correlation (θ) between algorithmic scores and hiring manager scores increases.*

Proof. The goal is to show that the derivative of female proportion of hires, p_h, with respect to θ is negative for $\theta \in [0, 1)$ — i.e., $\frac{\partial p_h}{\partial \theta} < 0$ for $\theta \in [0, 1)$.

1. **Define Female Proportion of Hires, $p_h(\theta)$.**

   Let $N(\theta) = \Pr(\text{female is hired})$ and $M(\theta) = \Pr(\text{male is hired})$. Then:

$$p_h(\theta) = \frac{N(\theta)}{N(\theta) + M(\theta)}$$

$$N(\theta) = p_a \int_{\tau_f^S}^{\infty} \Pr(Q^H > \tau^H | Q^S = q^S, \theta) f_{Q^S}(q^S) dq^S$$
$$M(\theta) = (1 - p_a) \int_{\tau_m^S}^{\infty} \Pr(Q^H > \tau^H | Q^S = q^S, \theta) f_{Q^S}(q^S) dq^S$$

   where p_a is the proportion of women in the applicant pool ($p_a < 0.5$); τ_f^S and τ_m^S are the screening thresholds for women and men, respectively; and τ^H is the hiring threshold. $f_{Q^S} = \phi(q^S)$ is the standard normal PDF.

2. **Shortlisting Thresholds under Equal Selection.** Under the *equal selection* constraint, the shortlisting thresholds τ_f^S for women and τ_m^S for men are adjusted such that the number of shortlisted women equals the number of shortlisted men:

$$p_a \Pr(Q^S > \tau_f^S | \text{female}) = (1 - p_a) \Pr(Q^S > \tau_m^S | \text{male})$$

   Let F_{Q^S} be the CDF of Q^S. Then:

$$p_a(1 - F_{Q^S}(\tau_f^S)) = (1 - p_a)(1 - F_{Q^S}(\tau_m^S))$$

   Given $p_a < 0.5$, then $1 - F_{Q^S}(\tau_f^S) > 1 - F_{Q^S}(\tau_m^S)$, which means $\tau_f^S < \tau_m^S$. Women face a lower

bar for shortlisting. In more explicit terms:

$$\tau_f^S = F_{Q^S}^{-1}\left(1 - \frac{1-p_a}{p_a}(1 - F_{Q^S}(\tau_m^S))\right) \tag{A.2}$$

3. **Conditional Hiring Probability.** The conditional distribution of hiring scores is $Q^H|Q^S = q^S \sim \mathcal{N}(\theta q^S, 1-\theta^2)$. The probability that a candidate is hired given their screening score is:

$$\Pr(Q^H > \tau^H | Q^S = q^S, \theta) = 1 - \Phi\left(\frac{\tau^H - \theta q^S}{\sqrt{1-\theta^2}}\right) = \Phi\left(\frac{\theta q^S - \tau^H}{\sqrt{1-\theta^2}}\right)$$

4. **Derivative of Conditional Hiring Probability.** Let $P(H|q^S, \theta) = \Pr(Q^H > \tau^H | Q^S = q^S, \theta)$.

$$\frac{\partial P(H|q^S,\theta)}{\partial \theta} = \phi\left(\frac{\theta q^S - \tau^H}{\sqrt{1-\theta^2}}\right) \cdot \frac{d}{d\theta}\left(\frac{\theta q^S - \tau^H}{\sqrt{1-\theta^2}}\right)$$

Calculating the derivative of the argument:

$$\frac{d}{d\theta}\left(\frac{\theta q^S - \tau^H}{\sqrt{1-\theta^2}}\right) = \frac{q^S\sqrt{1-\theta^2} - (\theta q^S - \tau^H)\frac{-\theta}{\sqrt{1-\theta^2}}}{1-\theta^2} = \frac{q^S(1-\theta^2) + \theta(\theta q^S - \tau^H)}{(1-\theta^2)^{3/2}} = \frac{q^S - \theta\tau^H}{(1-\theta^2)^{3/2}}$$

So,

$$\frac{\partial P(H|q^S,\theta)}{\partial \theta} = \phi\left(\frac{\theta q^S - \tau^H}{\sqrt{1-\theta^2}}\right)\frac{q^S - \theta\tau^H}{(1-\theta^2)^{3/2}}$$

Since $\phi(\cdot) > 0$ and $(1-\theta^2)^{3/2} > 0$ for $\theta \in [0, 1)$, the sign depends on $(q^S - \theta\tau^H)$. The derivative is larger for larger values of q^S.

5. **Derivative of $p_h(\theta)$.** Using the quotient rule, $\frac{dp_h}{d\theta}$ has the same sign as $N'(\theta)(N(\theta) + M(\theta)) - N(\theta)(N'(\theta) + M'(\theta)) = N'(\theta)M(\theta) - N(\theta)M'(\theta)$. We want to show this is negative, which is equivalent to showing:

$$\frac{N'(\theta)}{N(\theta)} < \frac{M'(\theta)}{M(\theta)}$$

This means the relative rate of change of the hiring probability with respect to θ is smaller for women than for men.

6. **Connecting to Thresholds.** The goal is to understand how the difference in thresholds ($\tau_f^S < \tau_m^S$) interacts with the derivative of the conditional hiring probability to make $p_h(\theta)$

decrease as θ increases.

$$N'(\theta) = p_a \int_{\tau_f^S}^{\infty} \frac{\partial P(H|q^S, \theta)}{\partial \theta} \phi(q^S) dq^S$$
$$M'(\theta) = (1 - p_a) \int_{\tau_m^S}^{\infty} \frac{\partial P(H|q^S, \theta)}{\partial \theta} \phi(q^S) dq^S$$

Define the "boost function" as $g(q^S, \theta) = \frac{\partial P(H|q^S,\theta)}{\partial \theta}$. This function represents how much the conditional hiring probability increases for a small increase in correlation θ, given a screening score q^S. As shown before (Step 5 of the proof), $g(q^S, \theta)$ increases with q^S (i.e., $\frac{\partial g}{\partial q^S} > 0$).

Now, compare the integrals for $N'(\theta)$ and $M'(\theta)$:

- $N'(\theta)$ involves integrating the boost function $g(q^S, \theta)$ over the range $q^S \in [\tau_f^S, \infty)$.
- $M'(\theta)$ involves integrating the same boost function $g(q^S, \theta)$ over the range $q^S \in [\tau_m^S, \infty)$.
- Since $p_a < 0.5$, we have $\tau_f^S < \tau_m^S$. The integration range for men is shifted to include only higher screening scores compared to the range for women.
- Because the boost function $g(q^S, \theta)$ is larger for higher q^S, the *average value* of the boost function over the men's integration range $[\tau_m^S, \infty)$ will be greater than its average value over the women's integration range $[\tau_f^S, \infty)$. Let $\bar{g}_f(\theta)$ and $\bar{g}_m(\theta)$ represent these average boosts for shortlisted women and men, respectively. Then $\bar{g}_f(\theta) < \bar{g}_m(\theta)$.
- The overall derivatives $N'(\theta)$ and $M'(\theta)$ are related to these average boosts multiplied by the respective probabilities of being shortlisted. Specifically, the average boost experienced by shortlisted men ($\bar{g}_m$) is higher than that for women ($\bar{g}_f$).

As θ increases, the hiring probability for men increases relatively more strongly. That is, the higher average boost $\bar{g}_m$ leads to $\frac{M'(\theta)}{M(\theta)}$ being larger than $\frac{N'(\theta)}{N(\theta)}$.

Since the men's hiring probability increases relatively faster (or decreases slower) than the women's hiring probability as θ increases, the proportion of women in the hired pool, $p_h(\theta) = \frac{N(\theta)}{N(\theta)+M(\theta)}$, must decrease.

7. **Conclusion.** Because the relative increase in hiring probability with θ is greater for men than for women (when $p_a < 0.5$ and $\tau_f^S < \tau_m^S$), the proportion of women in the hired pool, $p_h(\theta)$, decreases as θ increases from 0 towards 1.

□

A.3 Proof for Proposition 2

Proposition. *The female proportion of hires (p_h) decreases with the gender difference in the correlation parameter (δ).*

Proof. The proof directly comes Plackett's identity, which is used to show a known result that for a bivariate normal distribution with correlation ρ, $\Phi_2(x, y; \rho)$ is increasing in ρ (see Tong (2012)).

The probability that a candidate is hired is:

$$P(Q^H > \tau^H, Q^S > \tau^S) = \overline{\Phi}_2(\tau^H, \tau^S; \rho) \ = \Phi_2(-\tau^H, -\tau^S; \rho)$$

where $\overline{\Phi}_2(\tau^H, \tau^S; \rho)$ is the complementary joint standard normal CDF of (Q^S, Q^H) with correlation ρ. By symmetry, we have $\overline{\Phi}_2(\tau^H, \tau^S; \rho) \ = \Phi_2(-\tau^H, -\tau^S; \rho)$

Here, $\rho_{male} = \theta$ and $\rho_{female} = \theta - \delta$. Since $\theta > \theta - \delta$, the probability that a female is hired decreases with δ. Consequently, the proportion of women in the hired pool decreases with δ. □

A.4 Proof for Proposition 3

Proposition. *Conditional on the predictive accuracy of the screening algorithm (θ^S) and the hiring manager (θ^H), the average hire quality decreases as the correlation (θ) between algorithmic scores and hiring manager scores increases in the space $\theta \in [0, \min\{\frac{\theta^S}{\theta^H}, \frac{\theta^H}{\theta^S}\}]$, with hire quality reaching a global maximum at $\theta = 0$.*

Proof. The proof is of two parts. First, the goal is to compute the derivative of the conditional expectation of hires $E[Q_h] \equiv E[Q|Q^S > \tau^S, Q^H > \tau^H]$ with respect to θ, and show that it decreases in the specified range. Second, we show that $E[Q_h]$ is globally maximized at $\theta = 0$

Part 1: Derivative of $E[Q_h]$. The goal of the first part is to show that the derivative of $E[Q_h]$ wrt θ is negative whenever $0 \leq \theta \leq \min\{\frac{\theta^S}{\theta^H}, \frac{\theta^H}{\theta^S}\}$.

1. **Conditional Expectation of Truncated Multi-Normal Distribution.** Tallis (1961) shows that the expected value of a truncated multi-normal distribution is given by:

$$E[Q \mid Q^H > \tau^H, Q^S > \tau^S] = \frac{\theta^H \, \phi(\tau^H) \, \overline{\Phi}\Big(\frac{\tau^S - \theta \, \tau^H}{\sqrt{1-\theta^2}}\Big) \; + \; \theta^S \, \phi(\tau^S) \, \overline{\Phi}\Big(\frac{\tau^H - \theta \, \tau^S}{\sqrt{1-\theta^2}}\Big)}{\overline{\Phi}_2(\tau^H, \tau^S; \theta)} \tag{A.3}$$

By symmetry, we have:

$$E[Q_h] = \frac{\theta^H \phi(\tau^H) \Phi\left(\frac{\theta\tau^H - \tau^S}{\sqrt{1-\theta^2}}\right) + \theta^S \phi(\tau^S) \Phi\left(\frac{\theta\tau^S - \tau^H}{\sqrt{1-\theta^2}}\right)}{\Phi_2(-\tau^H, -\tau^S; \theta)}$$

Let $N(\theta)$ be the numerator and $D(\theta)$ be the denominator:

$$N(\theta) = \theta^S \phi(\tau^S) \Phi\left(\frac{\theta\tau^S - \tau^H}{\sqrt{1-\theta^2}}\right) + \theta^H \phi(\tau^H) \Phi\left(\frac{\theta\tau^H - \tau^S}{\sqrt{1-\theta^2}}\right) \tag{A.4}$$

$$D(\theta) = \Phi_2(-\tau^S, -\tau^H; \theta) \tag{A.5}$$

We will use the quotient rule for differentiation:

$$\frac{\partial E[Q_h]}{\partial \theta} = \frac{\frac{\partial N}{\partial \theta} D(\theta) - N(\theta) \frac{\partial D}{\partial \theta}}{[D(\theta)]^2}$$

2. **Derivative of the Denominator $D(\theta)$.** Using Plackett's Identity:

$$\frac{\partial D}{\partial \theta} = \frac{\partial}{\partial \theta} \Phi_2(-\tau^S, -\tau^H; \theta) = \phi_2(-\tau^S, -\tau^H; \theta)$$

Using Identity 2:

$$\frac{\partial D}{\partial \theta} = \phi_2(\tau^S, \tau^H; \theta)$$

Let's denote this as $D'(\theta) = \phi_2(\tau^S, \tau^H; \theta)$.

3. **Derivative of the Numerator $N(\theta)$.** Let the arguments of $\Phi(\cdot)$ in $N(\theta)$ be:

$$k_S(\theta) = \frac{\theta\tau^S - \tau^H}{\sqrt{1-\theta^2}} \quad \text{and} \quad k_H(\theta) = \frac{\theta\tau^H - \tau^S}{\sqrt{1-\theta^2}}$$

We need the derivatives $\frac{dk_S}{d\theta}$ and $\frac{dk_H}{d\theta}$. For $k_S(\theta) = (\theta\tau^S - \tau^H)(1-\theta^2)^{-1/2}$:

$$\begin{aligned}\frac{dk_S}{d\theta} &= (\tau^S)(1-\theta^2)^{-1/2} + (\theta\tau^S - \tau^H)\left(-\frac{1}{2}\right)(1-\theta^2)^{-3/2}(-2\theta) \\ &= \frac{\tau^S(1-\theta^2) + \theta(\theta\tau^S - \tau^H)}{(1-\theta^2)^{3/2}} \\ &= \frac{\tau^S - \theta^2\tau^S + \theta^2\tau^S - \theta\tau^H}{(1-\theta^2)^{3/2}} = \frac{\tau^S - \theta\tau^H}{(1-\theta^2)^{3/2}}\end{aligned}$$

Similarly, for $k_H(\theta) = (\theta\tau^H - \tau^S)(1-\theta^2)^{-1/2}$ (by swapping $\tau^S \leftrightarrow \tau^H$ in the expression for $\frac{dk_S}{d\theta}$):

$$\frac{dk_H}{d\theta} = \frac{\tau^H - \theta\tau^S}{(1-\theta^2)^{3/2}}$$

Now, we differentiate $N(\theta)$ using the chain rule $\frac{d}{d\theta}\Phi(w(\theta)) = \phi(w(\theta))\frac{dw}{d\theta}$:

$$\frac{\partial N}{\partial \theta} = \theta^S\phi(\tau^S)\left[\phi(k_S(\theta))\frac{dk_S}{d\theta}\right] + \theta^H\phi(\tau^H)\left[\phi(k_H(\theta))\frac{dk_H}{d\theta}\right]$$

Substitute $\frac{dk_S}{d\theta}$ and $\frac{dk_H}{d\theta}$:

$$\frac{\partial N}{\partial \theta} = \theta^S\phi(\tau^S)\phi\left(\frac{\theta\tau^S - \tau^H}{\sqrt{1-\theta^2}}\right)\frac{\tau^S - \theta\tau^H}{(1-\theta^2)^{3/2}} + \theta^H\phi(\tau^H)\phi\left(\frac{\theta\tau^H - \tau^S}{\sqrt{1-\theta^2}}\right)\frac{\tau^H - \theta\tau^S}{(1-\theta^2)^{3/2}}$$

We use Identity 3. Let $u(\theta) = \frac{\tau^H - \theta\tau^S}{\sqrt{1-\theta^2}}$, so $k_S(\theta) = -u(\theta)$. And let $v(\theta) = \frac{\tau^S - \theta\tau^H}{\sqrt{1-\theta^2}}$, so $k_H(\theta) = -v(\theta)$. By Identity 4, $\phi(k_S(\theta)) = \phi(-u(\theta)) = \phi(u(\theta))$ and $\phi(k_H(\theta)) = \phi(-v(\theta)) = \phi(v(\theta))$. So,

$$\begin{aligned}\phi(\tau^S)\phi(k_S(\theta)) &= \phi(\tau^S)\phi\left(\frac{\tau^H - \theta\tau^S}{\sqrt{1-\theta^2}}\right) = \sqrt{1-\theta^2}\phi_2(\tau^S, \tau^H; \theta) = \sqrt{1-\theta^2}D'(\theta) \\ \phi(\tau^H)\phi(k_H(\theta)) &= \phi(\tau^H)\phi\left(\frac{\tau^S - \theta\tau^H}{\sqrt{1-\theta^2}}\right) = \sqrt{1-\theta^2}\phi_2(\tau^H, \tau^S; \theta) = \sqrt{1-\theta^2}D'(\theta)\end{aligned}$$

Substituting these into $\frac{\partial N}{\partial \theta}$:

$$\begin{aligned}\frac{\partial N}{\partial \theta} &= \theta^S \left(\sqrt{1-\theta^2} D'(\theta)\right) \frac{\tau^S - \theta\tau^H}{(1-\theta^2)^{3/2}} + \theta^H \left(\sqrt{1-\theta^2} D'(\theta)\right) \frac{\tau^H - \theta\tau^S}{(1-\theta^2)^{3/2}} \\ &= \frac{D'(\theta)}{1-\theta^2} \left[\theta^S(\tau^S - \theta\tau^H) + \theta^H(\tau^H - \theta\tau^S)\right] \\ &= \frac{D'(\theta)}{1-\theta^2} \left[\theta^S\tau^S - \theta^S\theta\tau^H + \theta^H\tau^H - \theta^H\theta\tau^S\right] \\ &= \frac{D'(\theta)}{1-\theta^2} \left[(\theta^S\tau^S + \theta^H\tau^H) - \theta(\theta^S\tau^H + \theta^H\tau^S)\right]\end{aligned}$$

Let this be $N'(\theta)$.

4. **Assembling the Derivative.** Using the quotient rule formulation $\frac{\partial E}{\partial \theta} = \frac{N'(\theta)}{D(\theta)} - E\frac{D'(\theta)}{D(\theta)}$:

$$\begin{aligned}\frac{\partial E}{\partial \theta} &= \frac{1}{D(\theta)} \left(\frac{D'(\theta)}{1-\theta^2} \left[(\theta^S\tau^S + \theta^H\tau^H) - \theta(\theta^S\tau^H + \theta^H\tau^S)\right]\right) - E\frac{D'(\theta)}{D(\theta)} \\ &= \frac{D'(\theta)}{D(\theta)} \left\{\frac{(\theta^S\tau^S + \theta^H\tau^H) - \theta(\theta^S\tau^H + \theta^H\tau^S)}{1-\theta^2} - E\right\}\end{aligned}$$

Substituting back the full forms for $D(\theta) = \Phi_2(-\tau^S, -\tau^H; \theta)$, $D'(\theta) = \phi_2(\tau^S, \tau^H; \theta)$:

$$\frac{\partial E[Q_h]}{\partial \theta} = \frac{\phi_2(\tau^S, \tau^H; \theta)}{\Phi_2(-\tau^S, -\tau^H; \theta)} \left[\frac{(\theta^S\tau^S + \theta^H\tau^H) - \theta(\theta^S\tau^H + \theta^H\tau^S)}{1-\theta^2} - E[Q_h]\right]$$

Note that the first term inside the bracket is the conditional expectation of $E[Q|Q^S = \tau^S, Q^H = \tau^H]$. Rewriting the formula, we get:

$$\boxed{\frac{\partial E[Q_h]}{\partial \theta} = \frac{\phi_2(\tau^S, \tau^H; \theta)}{\Phi_2(-\tau^S, -\tau^H; \theta)} \left[E[Q|Q^S = \tau^S, Q^H = \tau^H] - E[Q|Q^S > \tau^S, Q^H > \tau^H]\right]}$$

5. **Sign of the Derivative.** Since $\phi_2 > 0$ and $\Phi_2 > 0$, the sign of the derivative is the same as the sign of $E[Q|Q^S = \tau^S, Q^H = \tau^H] - E[Q|Q^S > \tau^S, Q^H > \tau^H]$. The derivative is therefore negative whenever $E[Q|Q^S = \tau^S, Q^H = \tau^H] > E[Q|Q^S > \tau^S, Q^H > \tau^H]$ — i.e., when the corner solution is lower than the average solution in the quadrant above the thresholds.

   The corner solution is guaranteed to be lower than the average whenever the conditional

expected quality increases with thresholds:

$$\frac{\partial}{\partial \tau^S} E[Q|Q^S = \tau^S, Q^H = \tau^H] > 0$$
$$\frac{\partial}{\partial \tau^H} E[Q|Q^S = \tau^S, Q^H = \tau^H] > 0$$

Substituting the full form of $E[Q|Q^S = \tau^S, Q^H = \tau^H]$:

$$\frac{\partial}{\partial \tau^S} E[Q|Q^S = \tau^S, Q^H = \tau^H] = \frac{\partial}{\partial \tau^S} \frac{(\theta^S \tau^S + \theta^H \tau^H) - \theta(\theta^S \tau^H + \theta^H \tau^S)}{1 - \theta^2} = \frac{\theta^S - \theta\theta^H}{1 - \theta^2} > 0$$
$$\frac{\partial}{\partial \tau^H} E[Q|Q^S = \tau^S, Q^H = \tau^H] = \frac{\partial}{\partial \tau^H} \frac{(\theta^S \tau^S + \theta^H \tau^H) - \theta(\theta^S \tau^H + \theta^H \tau^S)}{1 - \theta^2} = \frac{\theta^H - \theta\theta^S}{1 - \theta^2} > 0$$

This gives a bound for when the derivative is guaranteed to be negative.

$$0 \leq \theta \leq \min\{\frac{\theta^S}{\theta^H}, \frac{\theta^H}{\theta^S}\} \tag{A.6}$$

Without loss of generality, we assume that $\theta^S \leq \theta^H$ (i.e., the screener has lower quality than the hiring manager in predicting overall quality), so the condition gets simplified to:

$$0 \leq \theta \leq \frac{\theta^S}{\theta^H} \tag{A.7}$$

Intuitively, this is the setting where the screener's correlation to predict the hiring manager's preferences is lower than the screener's relative ability to predict the candidate's actual quality compared to the hiring manager's ability to do so. In other words, we want the screener to be better at predicting the true quality than it is at predicting the hiring manager's preferences. This completes the first part of the proof.

Part 2: Global Maximum. The goal of part 2 is to show that $E[Q|Q^S > \tau^S, Q^H > \tau^H]_{\theta=0} > E[Q|Q^S > \tau^S, Q^H > \tau^H]_{1>\theta>0}$.

1. **Expected Quality of Hires.** Using Tallis (1961) again, we have:

$$E[Q \mid Q^H > \tau^H, Q^S > \tau^S] = \frac{\theta^H \, \phi(\tau^H) \, \overline{\Phi}\Big(\frac{\tau^S - \theta \, \tau^H}{\sqrt{1-\theta^2}}\Big) \; + \; \theta^S \, \phi(\tau^S) \, \overline{\Phi}\Big(\frac{\tau^H - \theta \, \tau^S}{\sqrt{1-\theta^2}}\Big)}{\overline{\Phi}_2(\tau^H, \tau^S; \theta)}$$

2. **Case when $\theta = 0$.** When $\theta = 0$, Q^H and Q^S are independent. Then,

$$\overline{\Phi}_2(\tau^H, \tau^S; 0) = \overline{\Phi}(\tau^H)\,\overline{\Phi}(\tau^S)$$

This simplifies the above expression to:

$$E[q \mid h > \tau^H, s > \tau^S]_{\theta=0} = \frac{\theta^H\,\phi(\tau^H)\,\overline{\Phi}(\tau^S) + \theta^S\,\phi(\tau^S)\,\overline{\Phi}(\tau^H)}{\overline{\Phi}(\tau^H)\,\overline{\Phi}(\tau^S)} = \theta^H\,\frac{\phi(\tau^H)}{\overline{\Phi}(\tau^H)} + \theta^S\,\frac{\phi(\tau^S)}{\overline{\Phi}(\tau^S)} \quad \text{(A.8)}$$

3. **Difference in expected quality $\Delta(\theta)$.** Define the difference in expected quality as:

$$\Delta(\theta) := E[Q \mid Q^H > \tau^H, Q^S > \tau^S]_{\theta=0} \;-\; E[Q \mid Q^H > \tau^H, Q^S > \tau^S]_{0<\theta<1}$$

This can be rearranged as:

$$\Delta(\theta) = \theta^H\,\phi(\tau^H)\,D_H(\theta) \;+\; \theta^S\,\phi(\tau^S)\,D_S(\theta),$$

where

$$D_H(\theta) = \frac{1}{\overline{\Phi}(\tau^H)} \;-\; \frac{\overline{\Phi}\big(\frac{\tau^S - \theta\tau^H}{\sqrt{1-\theta^2}}\big)}{\overline{\Phi}_2(\tau^H, \tau^S; \theta)}, \quad \text{(A.9)}$$

and $D_S(\theta)$ is the analogous term swapping $Q^H \leftrightarrow Q^S$.

$$D_S(\theta) = \frac{1}{\overline{\Phi}(\tau^S)} \;-\; \frac{\overline{\Phi}\big(\frac{\tau^H - \theta\tau^S}{\sqrt{1-\theta^2}}\big)}{\overline{\Phi}_2(\tau^H, \tau^S; \theta)},$$

Because θ^H, θ^S, $\phi(\cdot)$ are all positive, the sign of $\Delta(\theta)$ depends on the sign of $D_H(\theta)$ and $D_S(\theta)$. Thus, to prove that $\Delta(\theta) > 0$ for all $\theta \in (0, 1)$, it is sufficient to show that $D_H(\theta) > 0$ and $D_S(\theta) > 0$ for all $\theta \in (0, 1)$.

4. **Showing $D_H(\theta) > 0$ and $D_S(\theta) > 0$.** First consider $D_H(\theta)$. Multiplying both sides of A.9 by $\overline{\Phi}_2(\tau^H, \tau^S; \theta)$, we get:

$$\overline{\Phi}_2(\tau^H, \tau^S; \theta)\,D_H(\theta) = \frac{\overline{\Phi}_2(\tau^H, \tau^S; \theta)}{\overline{\Phi}(\tau^H)} - \overline{\Phi}\Big(\frac{\tau^S - \theta\tau^H}{\sqrt{1-\theta^2}}\Big).$$

Note that:

$$\frac{\overline{\Phi}_2(\tau^H, \tau^S; \theta)}{\overline{\Phi}(\tau^H)} = P(Q^S > \tau^S \mid Q^H > \tau^H),$$

For the second term, since in a bivariate normal distribution with correlation θ, the conditional probability of $Q^S \mid Q^H = \tau^H$ is normal with mean $\theta\tau^H$ and variance $1 - \theta^2$, we get:

$$\overline{\Phi}\Big(\frac{\tau^S - \theta\tau^H}{\sqrt{1-\theta^2}}\Big) = P(Q^S > \tau^S \mid Q^H = \tau^H).$$

Therefore,

$$\overline{\Phi}_2(\tau^H, \tau^S; \theta)\, D_H(\theta) = P(Q^S > \tau^S \mid Q^H > \tau^H) - P(Q^S > \tau^S \mid Q^H = \tau^H).$$

Because $\overline{\Phi}_2(\cdot)$ is positive, the sign of $D_H(\theta)$ is the same as the sign of the above difference. To show that $D_H(\theta) > 0$, we need to show that:

$$P(Q^S > \tau^S \mid Q^H > \tau^H) > P(Q^S > \tau^S \mid Q^H = \tau^H). \tag{A.10}$$

Next, define:

$$g(q^H) = P(Q^S > \tau^S \mid Q^H = q^H) = \overline{\Phi}\Big(\frac{\tau^S - \theta q^H}{\sqrt{1-\theta^2}}\Big),$$

Note that the right-hand side term in A.10 is the point probability that $Q^S > \tau^S$ given that $Q^H = \tau^H$. The left-hand side term is the *average* probability that $Q^S > \tau^S$ given that $Q^H > \tau^H$ over the range of $\tau^H < Q^H < \infty$. We can rewrite A.10 as:

$$\mathbb{E}[g(q^H) \mid q^H > \tau^H] > g(\tau^H).$$

Taking the derivative of $g(q^H)$ with respect to q^H, we get:

$$\frac{d}{dq^H} g(q^H) = \phi\left(\frac{\theta q^H - \tau^S}{\sqrt{1-\theta^2}}\right) \frac{\theta}{\sqrt{1-\theta^2}} \tag{A.11}$$

The derivative is always positive since $\phi(\cdot) > 0$ and $\theta > 0$. Therefore, $g(q^H)$ is strictly increasing in q^H.

Since $g(q^H)$ is strictly increasing, the expected value of $g(q^H)$ over the range of $\tau^H < q^H < \infty$ is greater than $g(\tau^H)$. Therefore, $D_H(\theta) > 0$. A symmetric argument swapping $Q^H \leftrightarrow Q^S$ shows that $D_S(\theta) > 0$.

It follows that $\Delta(\theta) > 0$ for all $\theta \in (0, 1)$. Hence,

$$E[Q \mid Q^H > \tau^H, Q^S > \tau^S]_0 > E[Q \mid Q^H > \tau^H, Q^S > \tau^S]_\theta \quad \forall\, \theta \in (0,1), \tag{A.12}$$

Equal Selection Constraint. The above proofs do not explicitly consider the equal selection constraint, where the screening threshold τ^S differs for men (τ^S_m) and women (τ^S_f). The overall expected quality is simply a weighted average based on the proportion $p_h(\theta)$ of each group in the hired pool:

$$E[Q_h] = p_h(\theta) \cdot E[Q_{h,\text{female}}] + (1 - p_h(\theta))E[Q_{h,\text{male}}]$$

Since A.7 and A.12 apply to each subgroup (using their specific thresholds), it extends to the overall weighted average, completing the proof. □

A.5 Difference in quality between men and women

So far we have considered the case where male and female applicants are equally qualified. We now consider the case where female applicants can be more/less qualified than men on average. We parametrize this difference using the location parameter α, as follows:

$$\begin{aligned}
(Q_m, Q^S_m, Q^H_m) &\sim \mathcal{N}\left(\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & \theta^S & \theta^H \\ \theta^S & 1 & \theta \\ \theta^H & \theta & 1 \end{bmatrix} \right) \\
(Q_f, Q^S_f, Q^H_f) &\sim \mathcal{N}\left(\begin{bmatrix} \alpha & \alpha & \alpha \end{bmatrix}, \begin{bmatrix} 1 & \theta^S & \theta^H \\ \theta^S & 1 & \theta \\ \theta^H & \theta & 1 \end{bmatrix} \right)
\end{aligned} \tag{A.13}$$

A positive α implies that women are more qualified on average than men, and a negative α implies the opposite.

The probability that a candidate is hired is given by:

$$Pr(\text{male is hired}) = \overline{\Phi}_2(\tau_m^S, \tau^H; 0, \theta) \tag{A.14}$$

$$Pr(\text{female is hired}) = \overline{\Phi}_2(\tau_f^S, \tau^H; \alpha, \theta) \tag{A.15}$$

where $\overline{\Phi}_2(\cdot; \alpha, \theta)$ is the bivariate tail CDF of Q^S and Q^H with mean α and correlation θ.

The proportion of women in the hired pool is given by:

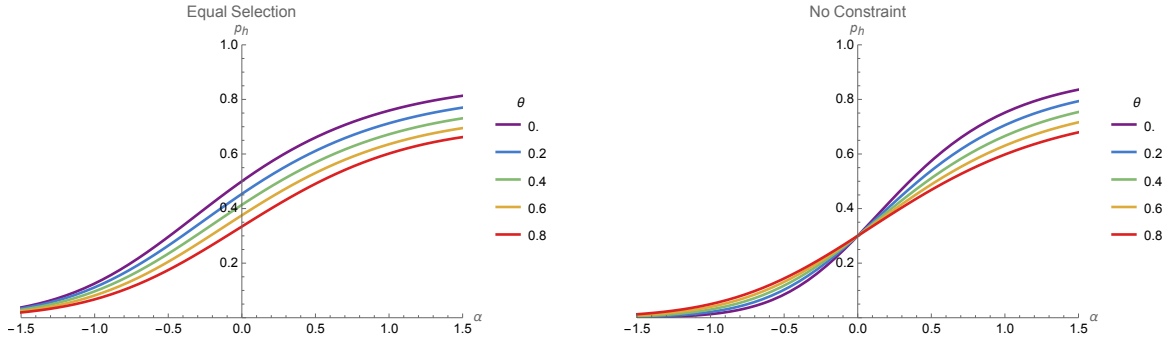
$$p_h = \frac{p_a \cdot Pr(\text{female is hired})}{p_a \cdot Pr(\text{female is hired}) + (1 - p_a) \cdot Pr(\text{male is hired})} \tag{A.16}$$

We solve for this numerically, and plot the proportion of women in the hired pool p_h as a function of α and θ in Figure 10.

When women are more qualified than men ($\alpha > 0$), the equal selection constraint becomes redundant. Higher mean quality compensates for the lower proportion of women in the applicant pool. This implies that the proportion of women in the hired pool p_h increases with α. Therefore, the equal selection constraint becomes redundant.

When women are less qualified than men ($\alpha < 0$), equal selection becomes even less effective.

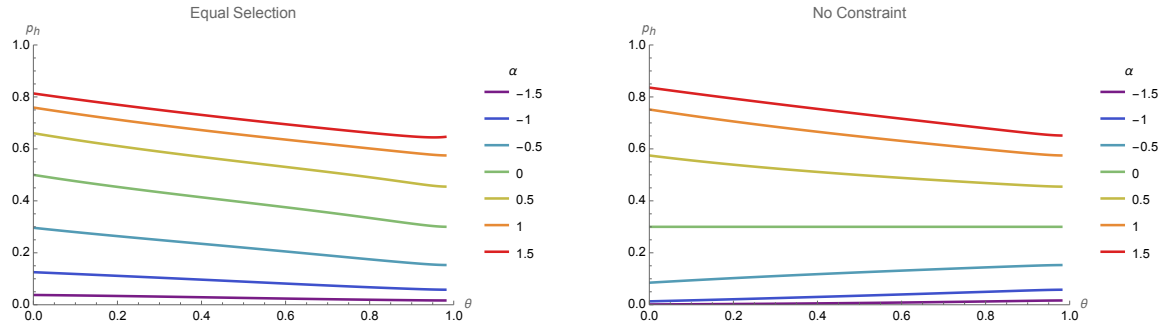
Figure 10: Female proportion of hires (p_h) vs. quality difference parameter (α)



Notes: This figure plots the female proportion of hires, p_h, as a function of the quality difference parameter, α. The proportion of women in the applicant pool is $p_a = 0.3$. This result does not depend on the θ^S and θ^H parameters.

Interestingly, without the equal selection constraint, the female proportion of hires decreases with θ when women have higher mean quality ($\alpha > 0$), and vice versa when women have lower mean quality.

Figure 11: Female proportion of hires (p_h) vs. correlation parameter (θ) for different α values



Notes: This figure plots the female proportion of hires, p_h, as a function of the correlation parameter, θ for different values of α. The proportion of women in the applicant pool is $p_a = 0.3$. This result does not depend on the θ^S and θ^H parameters.

B Additional details on the ML models

B.1 Predictive performance

We measure the predictive performance of the ML models using the Area Under ROC curve (AUC) criteria[16] on the hold-out test set and report the results in Table 8.

For the screening model, the overall AUC score is 0.83, and there is no difference in AUC scores between the male and female candidates. We also find that there is some heterogeneity in performance across job types, as reported in Table 9.

For the hiring manager model, the predictive performance is lower compared to the screening model since the hiring manager has more information from the interview, which we do not observe. Nonetheless, the predictive performance based on just resume characteristics is still reasonably high at 0.68, and there is no difference between genders. Note that the hiring manager model is evaluated on a subset of applicants in the hold-out test set who were, in fact, shortlisted.

[16]AUC is a widely-used measure for predictive performance for classification models since it is agnostic to both imbalanced classes and classification thresholds. The score ranges from 0.5 to 1, where 0.5 corresponds to a random classifier, and 1 corresponds to a perfect classifier.

Table 8: Predictive model performance by gender on hold-out test set

	Screening		Hiring Manager	
Group	AUC	Support	AUC	Support
Female	0.83	31,364	0.68	4,679
Male	0.83	42,393	0.68	6,678
Overall	0.83	73,757	0.68	11,357

Notes: This table reports the predictive performance of the screening and hiring manager classification models on the hold-out test set broken down by male and female candidates.

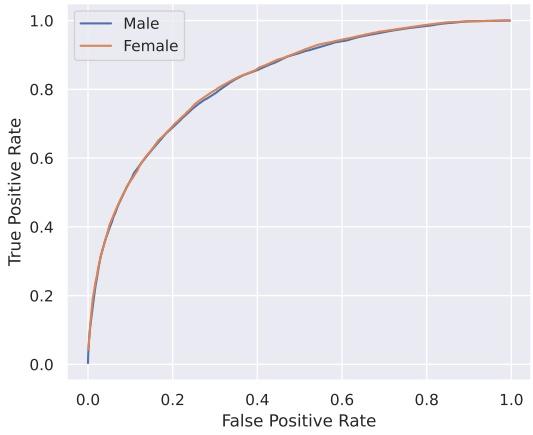
Table 9: Predictive model performance by job category on hold-out test set

	Screening		Hiring Manager	
Job Category	AUC	Support	AUC	Support
Legal & PR	0.86	7,337	0.65	926
Product & Design	0.85	12,519	0.66	1,863
Sales & Marketing	0.85	15,169	0.67	2,120
Other	0.83	214	0.59	25
Engineering & Technical	0.82	16,506	0.66	3,315
Finance & Accounting	0.82	7,026	0.67	886
Biz Dev & Operations	0.81	4,836	0.65	628
HR	0.79	3,355	0.69	452
Customer Service & Acct Management	0.78	6,734	0.7	1,112
Overall	0.83	73,757	0.68	11,357

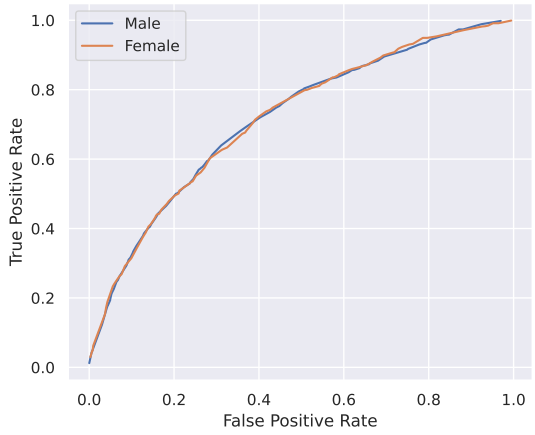
Notes: This table reports the predictive performance of the screening and hiring manager classification models on the hold-out test set. We estimate the metrics at the job posting level and aggregate up to the job category level.

Figure 12: ROC Curves for Screening and Hiring Manager Models

(a) Screening Model ROC Curve



(b) Hiring Manager Model ROC Curve



C Additional empirical analyses

C.1 Measures of observable quality differences between men and women

In this section we empirically assess differences in observable quality measures between men and women. To do so, we first identify four measures of observable quality: job-resume skill similarity, years of experience, attended a top 100 school, and educational attainment. We operationalize these measures as follows:

- **Job-Resume skill similarity:** We measure the average cosine similarity between skills listed in the job description and the applicant's resume. To get the cosine similarity, we first tokenize the job description and resume text. We then filter the tokens to extract only skills-related tokens (e.g., `python`, `data_analysis`, `project_management`) using a dictionary of skills[17]. We then get the vector representation of each skill token using a custom word2vec model trained on resumes, and take the average cosine similarity between the job description and resume skill vectors.

- **Years of experience:** We get the applicant's years of experience from the ATS.

[17]This dictionary was created using the skills section of LinkedIn profiles in a separate analysis.

- **Attended a top 100 school:** We create a binary variable indicating if the applicant attended a top 100 school based on the undergraduate institution listed in the resume. We use U.S. News and World Report's ranking of top 100 schools as the reference.

- **Educational attainment:** We create binary variables indicating if the applicant has a bachelor's, master's, or doctorate degree based on the highest degree listed in the resume.

For each of these outcomes, we estimate a linear regression model with job posting fixed effects

$$y_{ij} = \beta_{Female} \cdot Female_i + \alpha_j + \epsilon_{ij}$$

where y_{ij} is the observable quality measure for applicant i applying to job j, $Female_i$ is a binary variable indicating if the applicant is female, α_j is the job posting fixed effect, and ϵ_{ij} is the error term.

We report the coefficients and percentage differences below. Compared to male applicants, female applicants have roughly the same job-resume skill similarity, fewer years of experience, are more likely to have attended a top 100 school, more likely to have a bachelor's or master's degree, and less likely to have a doctorate degree.

Table 10: Regression coefficients of observable quality measures

Variable	β_{Female}	% Difference	p-value
Job-Resume skill similarity	0.003	0.51%	< 0.001
Yrs exp	-0.53	-6.1%	< 0.001
Attended top 100 school	0.013	4.66%	< 0.001
Has bachelor's degree	0.015	1.79%	< 0.001
Has master's degree	0.03	7.16%	< 0.001
Has doctorate	-0.004	-6.98%	< 0.001

Notes: This table shows the estimated regression coefficients and percentage differences of various observable quality measures. Job-Resume Similarity is the cosine similarity between the job description and the resume. Yrs Exp is the number of years of experience. Attended Top 100 School is a binary variable indicating if the candidate attended a top 100 school. Has Bachelor's Degree, Has Master's Degree, and Has Doctorate are binary variables indicating if the candidate has a bachelor's, master's, or doctorate degree, respectively.

C.2 Regression estimates on the likelihood of being shortlisted

Table 11: Likelihood of being shortlisted, OLS estimates

Dependent Variable:	Shortlisted (1=YES)
Variables	
Male	-0.0128***
	(0.0012)
Yrs Exp	0.0012***
	(0.0002)
Job Resume Similarity	0.3918***
	(0.0107)
Fixed-effects	
Job Posting	Yes
School Rank	Yes
Degree	Yes
Fit statistics	
Observations	595,246

Clustered (Job Posting) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: This table shows the OLS estimates of the likelihood of being shortlisted. Each observation corresponds to an application. Male applicants are more likely to be shortlisted than female applicants after controlling for job-resume skill similarity, years of experience, education, and job posting.

C.3 Goodness of fit using different copulas

Below we provide Kolmogorov-Smirnov (KS) statistics of empirical quality scores q^S and q^H fit against commonly used copulas. Lower KS statistics indicate a better fit. The Gaussian copula has the 2nd best fit after the Frank copula.

Table 12: Copula goodness-of-fit measures

Copula	KS-Statistic	p-value
Gaussian	2.62	< 0.0001
Gumbel	6.25	< 0.0001
Frank	2.29	< 0.0001
Clayton	6.07	< 0.0001
Joe	11.84	< 0.0001
AMH	5.37	< 0.0001

Notes: This table shows the goodness-of-fit measures of empirical quality scores q^S and q^H fit against various copulas.