Hiring as Exploration

Peter Bergman* Danielle Li † Lindsey Raymond ‡

January 3, 2020

Abstract

Modern hiring algorithms (AI) favor groups with proven track records, possibly to the detriment of those from under-represented groups or those with non-traditional qualifications. In this paper, we develop a new class of hiring algorithms that view hiring as a bandit—rather than a pure prediction—problem: in looking for the best workers over time, firms must balance exploitation (hiring from groups with proven track records) with exploration (hire non-traditional workers to learn about their quality). Using data on job applications for professional services positions within a large firm, we train an adaptive bandit algorithm for making interview decisions and evaluate its performance against an exploitation-only approach, as well as against human recruiting decisions. Our results suggest that an exploration-based approach can yield a Pareto improvement in diversity and quality (as measured by an applicant's likelihood of being hired conditional on being interviewed), relative to the firm's existing practices By contrast, we show that while an exploitation-only approach performs slightly better in terms of predicting hiring potential, it drastically reduces the percent of underrepresented minorities who are selected for an interview. Our results therefore highlight the importance exploration as a tool for constructing more efficient and equitable hiring algorithms.

*Columbia University
†MIT and NBER
‡MIT

The use of algorithms in hiring is estimated to have increased dramatically in the past decade.[1] Critics of this rise argue that automated approaches to screening may codify existing hiring patterns, possibly to the detriment of workers from under-represented groups or those with non-traditional qualifications. In particular, modern hiring algorithms are largely based on "supervised learning": given a training dataset, the algorithm forms predictions about the relation between worker performance and worker characteristics and applies its predictions to subsequent job applicants. Because this approach implicitly assumes that historical examples extend to future applicants, firms that rely on AI screening will tend to select from groups with proven track records, rather than taking risks to learn about non-traditional applicants.

Our paper is the first, that we know of, to develop and evaluate hiring algorithms that explicitly value exploration and learning. Our approach, which is based on "reinforcement learning," begins with the idea that hiring can be thought of as a bandit problem: in looking for the best workers over time, firms must balance exploitation (hiring from groups with proven track records) with exploration (hiring non-traditional workers to learn about their quality). In contrast with supervised learning, which focuses exclusively on exploitation, reinforcement learning places positive weight on exploration: each new hire provides the firm with an opportunity to learn about the productivity of future candidates with similar covariates.

We focus on the decision to grant interviews for high-skilled professional services positions: consulting, financial analysis, and data science. These sectors tend to offer well-paid jobs with opportunities for career advancement and, relatedly, also have attracted criticism for employing fewer women, minorities, and those with non-elite educations (Riviera, 2016). Using data from a large financial services firm, we train both supervised and reinforcement learning algorithms to identify job applicants who are likely to both receive and accept an offer, if interviewed. We show that, relative to what recruiters at this firm currently do, a standard supervised learning algorithm trained to maximize an applicant's likelihood of hire conditional on an interview would select candidates that are substantially less demographically diverse: most strikingly, the percent of interviewed candidates who are black or hispanic would fall from 10% under human selection to 1% under algorithmic selection. In contrast, the use of a reinforcement learning approach substantially increases the representation of black or hispanic applicants among those selected for an interview, to approximately 15%. This increase in diversity does not appear to decrease hiring rates among those interviewed: both the supervised and reinforcement learning algorithms *increase* average hiring rates among the selected cohort by 25% to 40% relative to human selection. Although the diversity

[1] While exact numbers are difficult to find, a recent survey of technology companies indicated that 60% plan on investing in AI-powered recruiting software in 2018, and over 75% of recruiters believe that artificial intelligence will transform hiring practices (Bogen and Rieke, 2018).

of selected applicants differs markedly between the supervised and reinforcement algorithms, we find minimal difference between their hiring rates (13.0% for supervised vs. 12.5% for reinforcement).

These findings show that there need not be an equity-efficiency tradeoff when it comes to expanding diversity; we are able to build an algorithm that increases hiring rates while also increasing diversity. However, our results also strongly caution against the idea that AI is generally equity and efficiency enhancing: in our setting, the standard exploitation-only approach achieves the same increases in hiring rates, but at the cost of virtually eliminating demographic diversity. This substantial difference in outcomes underscores the importance of building algorithms that value exploration. The remainder of this introduction discusses our approach and results in more detail.

Our data cover 100,008 job applications over the period 2016 to 2019. We observe the positions to which applicants apply, their demographics, work and education history, as well as an indicator for whether they receive an interview. Among those who are interviewed, we observe interview scores, whether an offer is made, and whether it is accepted. Our firm, like many others in this sector, is overwhelmed with initial options: it receives almost 200 applications for each opening it ends up filling, and 90% of applicants are rejected prior to being interviewed. Even following this, almost 90% of applicants who are interviewed either do not receive an offer or do not accept an offer. These "negative" interviews are wasteful: they take employees away from other productive work, and the firm receives little value from candidates it does not hire (Kuhn and Yu, 2019). The firm may therefore prefer to hire the same workers with fewer interviews, or to interview different candidates that are more likely to be hired.

We build two algorithms to predict an applicant's likelihood of being hired if interviewed. This is a potential outcome, which we denote H, that is only observed for applicants who are actually interviewed (denoted $I = 1$). As such, we train our algorithm on the subset of 2016-2017 job applicants who receive an interview. We first construct a supervised learning approach that can be thought of as "exploitation only": the goal of the algorithm is to identify candidates to interview based on what it has learned from its static training inputs. The algorithm places no weight on interviewing applicants in order to learn about their productivity over time.

Second, we implement a reinforcement learning approach to predict hiring potential. Our baseline model is trained on the same data as the supervised model, meaning that it begins with the same predictions about which types of applicants are likely to be successful. However, in addition to valuing a worker's immediate likelihood of being hired, the reinforcement algorithm also values learning about the productivity of other groups, and assigns an "exploration bonus" to selecting job applicants from groups with less representation in the algorithm's existing training data.[2] In each

[2]We assign exploration bonuses in two ways: first as a decreasing function of the number of training examples for an applicant with a given set of covariates and, second, using an "upper confidence bound" approach that assigns the

period of the 2018-2019 test sample, the reinforcement algorithm selects a subset of applicants that it would like to interview and, for the subset for whom we observe potential hiring outcomes (i.e. those who are actually interviewed), we record their actual interview performance and update our training data for the next period. We then re-train our algorithm on these new data and repeat this process.

Before continuing, we note that our approach yields the most reliable predictions when there is no selection on unobservables. This is a reasonable assumption in our setting because interview decisions are made entirely on the basis of resume data that are also available to us. Recruiters do not interact with candidates. If, however, there is selection on unobservables, then our model's predictions may be biased. We acknowledge the potential for such bias. We are not aware of any commercially-available hiring AI that corrects for this type of selection; in this sense our algorithms are representative of hiring AI as currently implemented.[3] Later in this introduction, we discuss how we use plausibly-random random variation in who is interviewed to account for selection concerns in evaluating human vs. algorithmic performance.

Our first set of results concerns the diversity of the candidates selected by these two algorithms relative to the human-based status quo. Assessing differences in diversity is straightforward: each algorithm assigns a score to each applicant, and we "select" the top N candidates under each ranking to be interviewed, where N is equal to the actual number of candidates who are interviewed. We then compare race, ethnicity, gender, and educational composition for each group. When it comes to race, we find that the industry standard, exploitation-only based approach would choose to interview far fewer under-represented minorities than the firm's current recruiting practices: the percent of interviewed applicants who are black falls from 5% to <1% and the percent who are Hispanic falls from 4% to <1%. This result suggests that, in the historical training sample, black and hispanic applicants who received an interview were, all else equal, somewhat less likely to be hired. A reinforcement-learning based approach takes our static model's initial predictions of hiring likelihood and adds exploration bonuses; doing this dramatically increases the representation of black and hispanic applicants, to 5% and 10% respectively, indicating that the the value the algorithm places on learning more about minority candidates outweighs any initial differences in predicted quality for these groups. Moreover, this gap in the minority interviewing rate persists even as exploration bonuses decline over time. Such a pattern is consistent with an algorithm learning

most optimistic guess for an applicants quality; this guess will be higher the noisier an estimate is. See Section 3.2 for more details.

[3]Raghavan, Barocas, Kleinberg, and Levy (2019) survey commercially available hiring "AI" and find that the minority that actually use ML tools simply attempt to predict performance using training data that is provided by either their client firms or their own in-house datasets.

over time that the distribution of hiring potential for minority candidates is comparable to that of majority candidates.

We see a related pattern when examining gender. At baseline, approximately 65% of interviewed applicants are male and 35% are female.[4]. This ratio essentially flips under the exploit-only algorithm: just under 60% of interviewed candidates would be female, indicating that, conditional on being interviewed, women were substantially more likely to be hired in the historical training data. Our reinforcement learning approaches lead to more equal representation, closer to 50-50; here, the exploration bonus actually applies to men.

Next, we assess the how these algorithms perform in terms of their ability to select candidates with a high likelihood of being hired. Here, our analysis is subject to concerns about sample selection because, while demographic variables are observed for all applicants, we can only observe hiring outcomes for applicants who are actually interviewed. In particular, we cannot directly observe the hiring outcomes of applicants chosen by either of our algorithms but not by a human. That is, while we want to compare $E[H|\tilde{I} = 1]$ for various interview policies $\tilde{I}$, we can only observe $E[H|\tilde{I} = 1, I = 1]$.

Our analysis addresses this concern in two ways. First, we consider the predictive power of our algorithms within the set of applicants who are actually interviewed: that is, among the set of applicants who are interviewed, do ones preferred by our AI models have a higher likelihood of being interviewed than the ones preferred by the human? To do this, we first need to distinguish human priorities within the set of interviewed candidates. We therefore construct a third algorithm, one designed to predict human interview decisions (rather than hiring likelihood); this means that we can rank every applicant in our test data by their likelihood of being chosen by a human recruiter to be interviewed; by their predicted likelihood of hire based on our supervised learning approach; and by their predicted likelihood of hire plus exploration bonus based on our reinforcement learning approach. Given this, we compare $E[H|\tilde{I} = 1, I = 1]$ where $\tilde{I}$ represents an alternative interview policy, such as interviewing the top 25% of applicants as ranked by each algorithm, and I denotes the firm's actual interview policy.

Using this approach, we show that average hiring rates are higher among applicants selected by either of our AI models compared to those selected by our human-mimicking algorithm: 9% of applicants selected by the human model are hired if interviewed compared to 13% among those selected by the supervised learning model designed to predict hiring likelihood and 12% among those selected by the reinforcement learning algorithm. In cases where the two algorithms explicitly disagree, applicants selected by either hiring AI but not by the human model are more likely to be hired than applicants selected by the human model but not the hiring AI. Similarly, we find a

[4]We exclude the 3.4% of candidates with missing or non-binary gender for this analysis.

positive correlation (among the set of interviewed applicants) between both hiring AI scores and actual hiring likelihood, but no correlation between an applicant's predicted likelihood of being interviewed and hiring outcomes.

Second, we allow for the possibility of selection on unobservables. In doing so, we are no longer able to derive an explicit estimate of hiring rates overall. However, we are able to show that moving *toward* AI-based recommendations can improve the firm's existing interview policy on the margin. To do this, we use the random assignment of applicants to initial screeners (who vary in their propensity to interview) as an instrument for being interviewed. Following the models of Abadie (2003), Angrist, Imbens, and Rubin (1996) we use this instrument to construct an explicit alternative interview policy $\tilde{I}$ that places more weight on algorithmic recommendations. Our new policy $\tilde{I}$ works as follows: it asks the firm to interview candidates with high algorithmic scores *as if* they were being evaluated by lenient initial screeners (those who are more likely to make interview offers) and, conversely, it asks that candidates with low algorithmic scores be interviewed as if they were being evaluated by strict initial screeners. Interview outcomes under I and $\tilde{I}$ differ only in their treatment of instrument compliers: candidates that would be interviewed (or rejected) regardless of their assigned screener are interviewed (or rejected) under both policies but compliers—those who would only be interviewed if they were assigned a lenient screener—change interview status. Specifically, $\tilde{I}$ interviews all compliers with high algorithmic scores and no compliers with low algorithmic scores. As such, we can evaluate differences in demographics and hiring likelihood between I and $\tilde{I}$ by comparing the characteristics of instrument compliers with high and low algorithmic scores.

We find that marginal applicants with high reinforcement-learning scores having a greater likelihood of being hired than marginal candidates with low scores. The same is true if we examine marginal applicants with high supervised-learning scores. Shifting the firm's interview policy closer to (either) algorithmic recommendation would therefore increase the quality of interviewed cohorts. In contrast, we find that while marginal applicants with high reinforcement-learning scores tend to be more diverse than marginal applicants with low scores, the same is not true for marginal applicants with high supervised-learning scores, who are less likely to be black or hispanic than marginal applicants with low supervised-learning scores. These results are consistent with our earlier results on the full interviewed sample.

In summary, we find that both supervised and reinforcement-learning based AI improves the accuracy of interview decisions relative to human decision-making, as measured by hiring rates. Yet despite having similar effects on quality, our algorithms diverge in terms of their impact on diversity: a standard exploitation only AI virtually stops interviewing black and hispanic applicants, while an adaptive approach doubles representation relative to the status quo. This stark contrast is likely driven by small differences in hiring outcomes among minority candidates in our training sample: if

minority candidates happen to be slightly less likely to be hired, then an exploitation-only based AI will tend to avoid such candidates. Our results highlight the importance of valuing exploration as a way to grant opportunities to under-represented candidates. Incorporating exploration into hiring algorithms can help identify qualified, diverse candidates that may be overlooked both by traditional human review as well as by common supervised-learning based approaches. While it is often thought that firms face a trade-off between equitable and efficient hiring practices, this tradeoff only exists at the frontier. In practice, firms may be inside this frontier, meaning that there is scope to improve both the quality and diversity of selected candidates.

Our paper is the first to develop and evaluate an adaptive algorithm that balances exploitation and exploration in the hiring process. In doing so, we build on a classic literature on bandit problems and exploration/exploitation tradeoffs in both economics and computer science (Thompson, 1933; Robbins, 1952; Gittins, 1979; Jovanovic, 1979; Lai and Robbins, 1985; March, 1991; Auer, 2003; Li et al., 2010; Bubeck and Cesa-Bianchi, 2012). Our specific implementation of reinforcement learning is an example of an Upper Confidence Bound (UCB) algorithm; algorithms based on this principle have been shown to be asymptotically efficient in terms of regret (Lai and Robbins, 1985; Audibert and Bubeck, 2009; Srinivas et al., 2012; Abbasi-Yadkori et al., 2019).[5] Related adaptive algorithms have been studied in the context of recommendation systems (Li et al., 2010), online ad placement (Russo et al., 2018), clinical trials (Berry, 2006), policy learning (Athey and Wager, 2019) and experimental designs (Kasy and Sautmann, 2019; Dimakopoulou, Athey, and Imbens, 2018).

In applying these tools, our paper connects a modern literature on contextual bandits (Li et al., 2010; Chu et al., 2011; Agrawal and Goyal, 2013; Dimakopoulou, Athey, and Imbens, 2018) to studies of automated screening and hiring practices. An early paper in this area, Autor and Scarborough (2008), shows that computerized job testing can improve the productivity of hires without impacting minority hiring rates. Hoffman, Kahn, Li (2019) demonstrate, further, that firms can improve the quality of hires by limiting human discretion and relying more on algorithmic recommendations. While these papers broadly demonstrate the potential of algorithms to improve decision-making, recent work has explored the role of algorithmic bias in settings ranging from healthcare (Mullainathan and Obermeyer, 2019; Obermeyer et al., 2019), to criminal justice (Kleinberg et al., 2018), to job search (Datta, Tschantz, and Datta, 2015).[6]

[5]For example, Abbasi-Yadkori et al. (2019) improves the theoretical analysis of UCB algorithms for linear stochastic bandits based on confidence ball sets to be $O(dlog(n)\sqrt{n} + \sqrt{dnlog(\frac{n}{\delta})}$ with probability of $1-\delta$ where d is the number of arms.

[6]In addition, there have been many publicized allegations of bias among recently deployed algorithms: "Amazon scraps secret AI recruiting tool that showed bias against women" https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-biasagainst-women-idUSKCN1MK08G; "Apple Card Investigated After Gender Discrimination Complaints." https://www.nytimes.com/2019/11/10/business/Apple-credit-card-investigation.html; "The racial bias built into photography" https://www.nytimes.com/2019/04/25/lens/sarah-lewis-racial-bias-photography.html

Existing studies of algorithmic bias focus primarily on supervised learning algorithms: due to their reliance on historical data, this approach (unintentionally) risks codifying existing biases (Barocas and Selbas, 2016; Lipton, Chouldechova and McAuley, 2019; Cowgill and Tucker, 2019)In fact, concern about fairness in algorithmic decision making led to a literature on the design Disparate Learning Processes (DLPs), algorithms designed to produce decision rules that equalize outcomes without engaging in disparate treatment of individuals. However, even in DLPs, treatment disparity results if other features correlate to group membership, identity is partly revealed by other features and, in general, DLPs provide a suboptimal trade-off between accuracy and impact parity (Lipton, Chouldechova and McAuley, 2019). [7] Based on our technical analysis, we argue that transparent treatment disparity is preferable to occluded methods for achieving impact parity. By showing that using a bandit approach can improve diversity outcomes relative to an exploitation-only approach, our paper provides empirical evidence that adding exploration can improve diversity with little impact on quality (relative to a supervised approach). This is consistent with the theoretical work of Cowgill (2019), which shows how noise in historical human hiring practices can mitigate machine bias by making an algorithm's training data more representative.

The remainder of the paper is organized as follows. Section 1 discusses our firm's hiring practices and its data. Section 2 presents a simple model of a firm's interview decision, and how human and algorithmic interview rules would operate in our setting. We next discuss how we explicitly construct and validate our algorithms in Section 3

1 Setting and Data

Our data come from a Fortune 500 financial services company in the United States. The data span four job families: Business Analysts, Data Analysts, Data Scientists, and Financial Analysts. Business Analysts' duties include pricing and market strategy, risk assessment, and product development and implementation. Data Analysts work in multiple areas of the company, building data infrastructure and optimizing business strategies. Data Scientists develop machine learning models for various business and product teams, write software, and analyze different types of data collected by the company. Finally, Financial Analysts' duties include forecasting, risk management, budgeting, and financial reporting. All of these positions require a Bachelor's degree with a preference for candidates graduating with a STEM major, a Master's degree, and often, experience programming in Python, R and SQL.

[7] For example, the Lipton et al., (2019) paper provide a toy example that given two attributes, work experience and hair length for male and female candidates, a algorithm constrained from using gender will differentiate on hair length (an irrelevant attribute).

Despite the employment growth in these types of mathematical and computer-oriented occupations, women, blacks, and hispanics are substantially under-represented relative to their overall shares of the workforce (**?**). This lack of diversity is acute enough that financial service companies such as JP Morgan Chase, Morgan Stanley, Bank of America and Citigroup offer scholarships and internship opportunities for historically under-represented minorities.[8] Similar disparities in hiring across demographic groups occur in the setting we study, as we show below.

1.1 Overview of the Hiring Process

Applications to these jobs primarily come from one of two sources: online postings and referrals. Openings are posted widely on websites such as Indeed, Glassdoor, the company's website, and job aggregators. Applicants frequently apply to more than one position within the company.

Irrespective of source, the hiring process is as follows. After a check to confirm an applicant meets the basic qualifications of the posted job (e.g. has a Bachelor's degree), a recruiter screens the application to assess jobs skills and fit using a pre-defined rubric. If an application passes this screen, they are invited for an interview. The interview consists of a case to evaluate the applicant's quantitative skills, ability to communicate, and conceptual understanding. If the candidate passes this stage of the interview process, they proceed to an on-site interview. The on-site interview consists of two cases and a product interview, each of which is administered by separate employees. Each interviewer rates the applicant's performance according to another rubric, and then they come together to form a consensus recommendation to offer the candidate the position or not. The applicant then decides whether to accept the offer or not to accept the offer.

1.2 Data

The data cover each stage of the hiring process from the year 2016 through the first quarter of 2019. For each job candidate, we know their race, gender, education history, work experience, referral status, application date, what job they applied to and its location. For each recruiter, we observe whose resume they reviewed and their decision to interview the candidate or not. For each interviewer, we observe whom they interviewed and their scores of that interview, and their recommendation. We observe the race and gender of the interviewer as well. Lastly, we observe whether a candidate received a job offer and whether they accepted.

[8] JP Morgan Chase created Launching Leaders; Morgan Stanley offers the Richard B. Fisher Scholarship Program; Citigroup offers the HSF/Citigroup Fellows Award; and Bank of America awards the Bank of America Merrill Lynch Diversity Scholarship of Distinction.

Because a candidate may apply to and be interviewed for multiple job openings, our unit of observation is aggregated to the applicant-by-job-family-by-seniority level. We aggregate our key outcome variables, interview status and hire status, to match this unit of observation.

Table 1 shows the characteristics of the sample. The data consist of 100,008 applications, mostly from asian (52%) and white (27%) applicants. Black and hispanic candidates comprise 12% of all applications. There is a gender disparity as well: roughly two thirds of the applications are from male job candidates. Almost all applicants attended a college in the United States, just over 60% have a Master's degree as well, and about one quarter of applicants have a quantitative background. [9] Lastly, 13% of candidates received a referral.

Table 2 shows the total number of candidates from the application stage through the hiring stage for applications and then broken down by job family. From 2016 to 2019, we observe 72,132 candidates submitting 100,008 applications. Interview rates are low: ranging from 2% for Data Scientists to 9% for Business Analysts. Hiring rates conditional on making it to the interview stage vary from 8% for Business Analysts up to 40% for Data Analysts.

Table 3 shows these numbers, but by the race and gender of the applicant instead of the job family. Interview rates are generally comparable across race and gender, except for black job candidates. Black job candidates are interviewed at roughly half the rate of other candidates. Conditional on receiving an interview, both black job candidates and hispanic job candidates receive an offer at a lower rates as well. The absolute number of minority hires matter as well; so few minority job candidates are interviewed and hired that any process that attempts to affect the diversity of employees must affect who is chosen to be interviewed first. Below, we also show that these historical hiring rates across demographic groups have implications for the estimation of hiring algorithms as well.

2 Conceptual Framework

We begin by providing a framework of the firm's interview and hiring decisions, as described to us by the firm. The firm receives a large number of applicants for a posted job opening and must decide which ones to interview. Each applicant has quality q, which is initially unobserved, but which is revealed to the firm if the candidate is interviewed. The firm wants to hire all applicants with $q > \tau$, where τ is a fixed quality threshold.

We assume that the firm receives a payoff of 1 for each candidate it interviews and hires and pays a fixed cost of c per interview. For simplicity, we assume that there are N fixed interview slots

[9] We define quantitative background by checking previous work history job titles jobs including scientist, programmer, software, engineer, developer, quant, statistics, technology and data.We also define a second variable, finance background, by checking job titles for finance, financial, bank and business.

and normalize $c = 0$. The firm's maximization problem is thus given by

$$\max_{I \in \mathbf{I}} \sum_i I_i \times \Pr(q_i > \tau) \tag{1}$$

where $\mathbf{I}$ is the set of interview policies that satisfy the slot constraint and $I_i \in \{0, 1\}$ is the interview status for applicant i. The firm's maximal payoff is N, which it obtains if it hires all the applicants it interviews.

Firms can make inferences about a candidate's quality based on his or her covariates X: that is, the firm forms a noisy estimate $f(X) = q + \varepsilon$. The optimal interview policy for the firm at any given time is therefore:

$$I^* = \mathbb{I}(f(X) > \underline{\tau}) \tag{2}$$

where $\underline{\tau}$ is set to meet the interview slot constraint. We can think of $f(X)$ as a "score" that model f gives candidates with covariates X. Given this policy, the firm's expected payoff can be expressed as

$$\sum_i \Pr(q_i > \tau | q_i + \varepsilon_i > \underline{\tau}) \tag{3}$$

This figure is weakly smaller than N and is decreasing in the variance of ε_i. That is, a firm's payoff depends on its ability to form an accurate estimate of q based on an applicant's observed covariates. We next describe three approaches for forming estimates of q.

Human recruiters

We assume that human recruiters form estimates of q based on the applicant's covariates X (which are also observed by the econometrician) and W (which are only observed by the recruiter) using a model that is based on data from the human recruiter's personal experience, a state variable denoted D^H. D^H could be, for instance, the total number of previous applicants a given recruiter has evaluated. We denote this human model by $f^H(X, W; D^H) = q + \varepsilon^H$. Going forward, we use the superscript H to refer to any inputs and outputs related to human HR decisions.

The recruiter's interview policy is therefore given by $I = \mathbb{I}(f^H(X, W; D^H) > \underline{\tau})$ and the expected payoff to a firm following interview policy I is thus given by

$$\sum_i \Pr(q_i > \tau | f^H(X, W; D^H) > \underline{\tau})$$

For notational simplicity going forward, we will refer to an applicant's "score" under the human hiring model as $s^H = f^H(X, W; D^H)$, suppressing notation related to covariates and the data state.

Supervised learning

In a standard supervised learning model, the firm attempts to estimate q using data on the characteristics and revealed quality of previously interviewed applicants. This approach differs from the human problem in two ways. First, we assume that algorithms cannot take advantage of a recruiter's private information, W, and must instead rely only on observables X. Second, we allow for the algorithm to make use of a different—and potentially much larger—training dataset of outcomes and covariates upon which to build its model. We call these data D^S, where, going forward, the superscript S denotes variables related this model.

Given a model of q, $f^S(X; D^S) = q + \varepsilon^S$, the optimal interview policy is given by $I^S = \mathbb{I}(f^S(X; D^S) > \underline{\tau})$, and the expected payoff to a firm following interview policy I^S is

$$\sum_i \Pr(q_i > \tau | f^S(X; D^S) > \underline{\tau})$$

The relative performance of the human and supervised learning interview policies depends on two factors: the value of the recruiter's private information W and the value of the supervised learning model's larger dataset: $|D^S| >> |D^H|$. In practice, both estimates of q could be subject to systematic biases; here, we include that in the error term ε. Again, we write $s^S = f^S(X; D^S)$ to reference an applicant's score under this supervised learning model.

Reinforcement learning

So far, we have only considered the firm's hiring in a given period. In practice, firms make interview decisions regularly over the course of many years in a sequential manner. Decisions made at $t = 0$ influence not just the firm's immediate productivity, but the firm's future productivity, via its impact on the firm's future training data. The firm is interested in maximizing a present discounted value of its interview decisions over time:

$$\sum_{t=0}^{\infty} \beta^t \max_{I_t^R \in \mathbf{I}_t} \sum_i I_{it}^R \times \Pr(q_i > \tau) \tag{4}$$

Here, the firm optimizes its interview policy in each period t. In particular, the firm's interview policy may change over time because its model of q may change: in any given period, it forms a model $f_t^R(X; D_t^R) = q + \varepsilon_t^R$ based on the state of its data D_t^R. A firm's data is a stock variable that evolves according to $D_{t+1}^R = D_t^R \cup I_t^R$. That is, the data available to predict applicant quality grows with each set of applicants who are interviewed.

The set up above is an example of a contextual bandit problem (Li et al., 2010; Agrawal and Goyal,. 2013b). A contextual bandit is a generalization of the classic multi-armed bandit, in which a decision-maker must decide which action to take each period (e.g. which arm to pull), but has the benefit of some contextual information (observable characteristics that may predict the rewards associated with each given arm). The decision-maker's goal is to minimize regret—the difference in payoffs between the sequence of actions she chose, and the best possible sequence in hindsight (Chu et al., 2011). As applied in our setting, applicants can be thought of as "arms" and their demographics or experience provide the potentially predictive "context." A screener must decide which applicants to interview each period, with the goal of minimizing the number of qualified candidates she misses over time.

The optimal solution to these types of contextual bandits requires some amount of exploration (Dimakopolou et al, 2018). Specifically, Bastani, Bayati, Khosravi (2019) show that exploration-free greedy algorithms (such as supervised learning) are generally sub-optimal. This means that a firm will sometimes prefer to interview candidates with lower expected q, especially if it is particularly uncertain about that candidate's quality.

Constructing an optimal algorithm in this type of setting, however, is often computationally intractable. As such, the computer science literature has instead focused on developing feasible heuristic algorithms that are asymptotically optimal (Bastani, Bayati, Khosravi 2019). Such classes of algorithms often apply an "exploration bonus" to certain applicants. In our case, we can write the firm's reinforcement-learning based interview policy as

$$I_t^R = \mathbb{I}(f_t^R(X; D_t^R) + B^R(X; D_t^R) > \underline{\tau}) \tag{5}$$

where $f_t^R(X; D_t^R)$ is an applicant's predicted quality given dataset D_t^R and $B^R(X; D_t^R)$ is a bonus value that depends on a worker's covariates and the firm's data. We write $s^R = f_t^R(X; D_t^R) + B^R(X; D_t^R)$ to refer to an applicant's reinforcement learning score. In Section 3.2 we discuss three asymptotically-optimal ways of defining these bonuses.

3 Algorithm Construction

In this paper, we compare outcomes associated with three different interview policies, which we denote I (or I^H), I^S, and I^R. Policy $I \in \{0, 1\}$ is the firm's status quo interview policy: the set of applicants who are actually interviewed in our data. We will additionally train an algorithm to mimic human decisions and denote the choices of this algorithm as $I^H \in \{0, 1\} = \mathbb{I}(s^H > \underline{\tau})$. Policy $I^S \in \{0, 1\} = \mathbb{I}(s^S > \underline{\tau})$ is based on using supervised learning to predict a worker's likelihood

of being hired conditional on being interviewed. Policy $I^R \in \{0,1\} = \mathbb{I}(s^R > \underline{\tau})$ is based on using reinforcement learning to learn about workers' hiring likelihoods over time. In our comparisons, we will vary the selection threshold $\underline{\tau}$, but make sure that each algorithm selects the same number of applicants to interview so that $E[I] = E[I^S] = E[I^R]$.

3.1 Training Data and Input Features

We have data on job applications from January 2016 to March 2019. We divide this sample up into a training dataset consisting of applications that arrive before 2018 and a test dataset of applications that arrive afterward. Our models are built on the training data and our analyses of diversity and hiring likelihood are based on out-of-sample model performance in the test data.

We have information on an applicants' educational background, work experience, referral status, basic demographics, as well as information on the type of position to which they applied. To transform this raw information into usable inputs for a machine learning model, we create a series of categorical and numerical variables that serve as "features" for each applicant. We standardize all non-indicator features to bring them into the same value range. Because we are interested in decision-making at the interview stage, we only use information available as of application date as predictive features. Our final model includes 106 input features.

3.2 Models

Human Decision-making

First, we train a model that predicts an applicant's actual interview outcome, as chosen by a human recruiter. Following Kaelbling (2019), we randomly subsample our training data to create a balanced sample such that half the sample was not interviewed and half was interviewed. We fit an ensemble model that combines predictions from a L1-regularized logistic regression (lasso) and a random forest. We first fit the lasso using three fold cross validation to choose the γ penalty

$$\min_{\beta \in \mathcal{R}^{\mathcal{D}}} \frac{1}{N} ||\mathcal{I} - X\beta||^2 + \gamma ||\beta||_1$$

We fit the second sub-component, a random forest model, on a second randomly selected balanced sample from our training set and use three fold cross validation to choose tree depth, number of trees, and the maximum number of features. We then average the predictions of the two individual models to generate a predicted probability of interview for each applicant.

We evaluate out-of-sample performance on randomly selected balanced samples from our testing period. Appendix Figure **??** plots the receiver operating characteristic (ROC) curve and its associated

AUC, or area under the curve. These are standard measure of predictive performances that quantify the trade-off between a model's true positive rate and its false positive rate. Formally, the value of the AUC is equal to $\Pr(\hat{I}_i > \hat{I}_j | I_i = 1, I_j = 0)$. [10] Our model ranks a randomly chosen interviewed applicant ahead of a randomly chosen applicant who is not interviewed 76% of the time. In the appendix, we report the AUC separately for each test sample year, as well as a standard confusion matrix which shows the percentage of correctly classified candidates and the number of type I and II errors (assuming that our model "selects" the top X candidates to be interviewed, where X is the same as the actual number of interviewees). We correctly classify candidates 70% of the time. Although a "good" AUC number is heavily context specific, a general rule of thumb is that tests in the AUC range of $0.75 - 0.85$ have intermediate to good discriminative properties depending on the specific context and shape of the curve (Fischer et al., 2013). In the interview context, given the smooth and symmetrical shape of the curve, this is a good result.

Algorithmic Decision-making: Supervised Learning

Next, instead of trying to understand what human recruiters do, we train a supervised learning model to predict an applicant's likelihood of being hired if interviewed. We use the same ensemble model (a lasso and random forest model) as we do when constructing our model of human interview decisions.[11] Because we are only able to observe an applicant's hiring outcome if that applicant is interviewed, we restrict our training data to the subset of applicants who are actually interviewed. Otherwise, the training procedure is exactly the same as described above. Appendix Figure **??** plots the ROC associated with this hiring model. Our model has an AUC of .67, meaning that it will rank an interviewed applicant who is hired higher than an interviewed but not hired applicant 67 percent of the time. In contrast to modelling the human decision making process, our goal is to provide a standard example of a supervised learning decision maker. [12]

Reinforcement Learning Decision Making

As stated in Equation (5), our reinforcement learning algorithm assigns priority to applicants based on the sum of their estimated quality and an exploration bonus that depends on their covariates and the state of the model's training data at time t: $I_t^R = \mathbb{I}(f_t^R(X; D_t^R) + B^R(X; D_t^R) > \underline{\tau})$. To implement such an algorithm, we need two components: a way of predicting a worker's probability of hire and way of calculating his or her exploration bonus.

[10] AUC is identical to the result of a Wilcoxon or Mann–Whitney U test (Hanley and McNeil, 1982).

[11] By choosing the same model structure for both interview and hire, we avoid differences in predictive accuracy driven by choice of machine learning algorithm.

[12] Ideally, we would compare our accuracy results to those of commercial providers, but none provide comparable metrics (Raghaven et al., 2019).

To generate predictions of probability of hire, we begin with the baseline supervised learning model described above and trained on the 2016-2017 sample. That is, by construction $f_0^R(X; D_0^R) = f^S(X; D^S)$ for period $t = 0$ of the test data. Whereas the supervised learning model is static, the reinforcement learning model is updated each period with data from applicants who were selected to be interviewed in the previous period. Because we are unable to observe hiring outcomes for $I^R = 1, I = 0$ applicants, we only add $I^R = 1, I = 1$ candidates into D_{t+1}^R. After this, we create a randomly selected balanced sample, and generate new estimates $f_{t+1}^R(X; D_{t+1}^R)$ on the new data, using the same ensemble approach as describe for our supervised learning algorithm.

The second part of the reinforcement algorithm involves calculating the exploration bonus $B^R(X; D_t^R)$. We do this in two standard ways.

1. **Experience Based**: Here, an applicant's exploration bonus depends on the number of applicants of the same type in D_t^R. Specifically, we choose

$$B^R(X; D_t^R) = \frac{\alpha_a}{\sqrt{N_{t,a}}}$$

where $N_{t,a}$ is the number of group a members in the model's prior data D_t^R, and $\alpha > 0$ is a parameter. This algorithm requires that we pre-specify what a group is. In our implementation, we define groups based on gender and race/ethnicity.[13] That is, white applicants receive exploration bonuses as a function of the number of prior white applicants who were interviewed. Bonuses are additive in the sense that a black women will receive a bonus for race and gender.

2. **Upper Confidence Bound**:Under this approach, the exploration bonus is calculated as an upper confidence bound:

$$B^R(X; D_t^R) = Z_{.95}\frac{\alpha\sigma_{t,a}}{\sqrt{N_{t,a}}}$$

where $N_{t,a}$ is the number of group a members in the model's prior data D_t^R; $se_{t,a}$ is the standard error of predictions for members of group a as of time t. At time t in our algorithm, all black applicants will receive the same $se_{t,black}$ value; and $Z_{.95}$ represents the Z value for the confidence level (we use the 95th). Again, $\alpha > 0$ is a parameter. For instance, in addition to deriving regret bounds comparable to those that are state of the art in the literature, Hao et al. 2019 show bootstrapped UCB performs well in simulations against other methods and is much easier to implement (Hao et al., 2019 and Kveton et al., 2018).

[13] See Li et al. (2011) for an example of using a contextual bandit problem with a set of predefined user groups.

4 Main Results

4.1 Impacts on Diversity of Interviewed Applicants

We begin by assessing the impact of each policy on the diversity of candidates selected for an interview in our test sample. This is given by comparing $E[X|I = 1]$, $E[X|I^S = 1]$, and $E[X|I^R = 1]$, for various demographic measures X. This analysis is straightforward in the sense that we observe demographic covariates such as race and gender for all applicants, so that we can easily examine differences in the composition of applicants selected by each of the interview policies described above.

Figure 1 shows the composition of selected applicants averaged over our test period sample. In Panel A, we display the distribution of applicants in our sample: of those who choose to report their race, 58% of applicants are asian, 29% are white, 9% are black, and 4% are hispanic. Among those interviewed (Panel B), the proportion of asians and hispanics stay the same at 57% and 4%, respectively, and the proportion who are white rises to 34%, largely at the expense of black applicants, whose representation falls in half to just under 5%.

Panels C and D display the counterfactual interview choices of our AI models. Panel C shows that a standard supervised ML algorithm would increase the proportion of white applicants who are interviewed by decreasing the representation of black applicants to 3% and virtually eliminating hispanic applicants. By contrast, a dynamic exploration based approach would increase the proportion of black and hispanic applicants among the interviewed set to 10% and 5%, respectively. These increases would largely come at the expense of Asian applicants, whose representation would fall to 49%.

Panel A of Figure 2 plots one key mechanism underlying this difference in diversity outcomes between our AI models: the exploration bonuses used in the reinforcement model. Here, the size of a bonus is tied to the number of previous examples in the data: there are very few black and hispanic candidates in the model's training data D_t^H, especially early on. As a result, black and hispanic applicants receive large initial exploration bonuses. As more black and hispanic applicants are interviewed, the algorithm's data are updated and exploration bonuses decline. Bonuses fall relatively quickly because of the $\frac{1}{n}$ term and that the applicant pool during the training periods contains nonzero numbers of diverse candidates. However, even as exploration bonuses decline significantly over the course of the test sample, we still see higher long term diversity among applicants selected by the reinforcement learning based algorithm: Panel B of Figure 2 shows higher levels of minority representation in the last 6 months of the sample.

Figure 3 plots the same set of results, for gender. Here, the majority of applicants are male (66%) and men and women are selected by human screeners at the same rate as they apply so

that the interviewed class maintains the same composition. Both our supervised and reinforcement ML algorithms, however, would interview women at a greater rate than men, increasing their representation among those selected to 46% (supervised) and 50% (reinforcement). Unlike race and ethnicity, our supervised and reinforcement algorithms are aligned. This suggests that, in the historical training data women who were interviewed were more likely to be hired than men; maximizing hiring likelihood in the static sense would lead to women being interviewed at greater rates. Figure 4 shows that exploration bonuses are decreasing over time for both men and women, although the overall level of exploration bonuses is smaller relative to the bonuses seen for blacks and hispanics, since their representation is higher in the training data. When we focus on examining outcomes in the last 6 months of the test sample, we still see a marked increase in representation among women.

4.2 Impacts on Quality of Interviewed Applicants

Next, we consider the relative quality of the interview classes selected by each algorithm, as well as by human recruiters. In particular, we focus on whether and to what extent the gains in diversity made by our exploration-based algorithm come at the cost of applicant quality. We use a worker's likelihood of being hired (if interviewed) as a proxy for his or her quality. In our model, there is a direct relation between an applicant's quality, q and his or her likelihood of being hired conditional on being interviewed. We denote this $H = \mathbb{I}(q > \tau)$. We would ideally like to compare the average hiring likelihood among applicants selected by each interview policy: the status quo human selection, $E[H|I = 1]$; the one based on the recommendations of the supervised learning algorithm, $E[H|I^S = 1]$; and the one based on reinforcement learning, $E[H|I^R = 1]$.

However, unlike a worker's demographics, which are always observed, a worker's hiring outcome if interviewed is a potential outcome that is only observed for applicants who are actually interviewed. In particular, we cannot directly observe hiring outcomes for applicants selected by either AI algorithm, but not by the human reviewer. Our analysis will address this concern in two ways. First, we simply restrict our analysis to the set of applicants who are interviewed. Among this set, we ask whether applicants preferred by one policy are more likely to be hired than those preferred by another. We note that although this sample is selected, we believe that there is relatively little scope for selection on unobservables in our setting because interview decisions are made on the basis of resume review, without any in-person contact. Because we have data on education, work history, and demographics, we are able to make predictions using largely the same information.

Second, we allow for selection on unobservables. Doing so means that we are no longer able to compare hiring outcomes between the human interview policy I and the algorithmic policies I^S and I^R. Instead, we use an instrumental variables approach to construct a modification of policy I

that more closely follows I^S and I^R in marginal cases. We then explicitly show that this alternative policy can yield a Pareto improvement in terms of both quality and diversity, relative to the firm's existing interview practices.

Predictive Accuracy Amongst Interviewed Sample

In this section, we compare the quality of applicants selected by our AI algorithms among the sample of applicants who are interviewed, for whom hiring outcomes are observed. Because all applicants in this sample are, by construction, selected by human recruiters, we cannot directly compare the accuracy of algorithmic. human choices within this sample. To get around this, we construct another algorithm, one designed to predict human interview decisions (rather than hiring likelihood), using the standard LASSO approach described in Section 3.2. This model allows us to order interviewed applicants in terms of their human score s^H, in addition to their algorithmic scores s^S and s^R.

Figure 5 plots a binned scatterplot depicting the correlation between algorithm scores and hiring outcomes among the set of interviewed applicants, where each dot represents the hiring likelihood for applicants in each selection score ventile. Panel A focuses on applicants as ordered by their human-predictive model score. We see a slightly negative correlation between human model scores and an applicant's actual likelihood of being hired: if anything, applicants whom we predict are more likely to be interviewed are in fact less likely to be hired. In contrast, Panels B-D show a consistently positive relation between algorithmic priority selection scores and an applicant's (out of sample) likelihood of being hired. Interviewed applicants with algorithmic scores in the bottom half the distribution have hiring rates of less than 10%, while those in the upper half of the distribution have hiring rates of closer to 20%. All of our algorithms are particularly good at identifying interviewed candidates with the greatest hiring likelihoods: over 30% of interviewed applicants with algorithmic scores in the top decile end up being hired.

Figure 6 translates these scores into actual selection and hiring outcomes for various selection policies. In Panel A, we consider an "interview" strategy that selects the top 10% of interviewed applicants ranked by each model, and compares hiring outcomes for this group. Consistent with the correlations in Figure 5, we see that fewer than 5% of top applicants under the human model are hired, while over 25% of top algorithmic scores are. Among those, our exploitation-only model performs slightly better, with an average hiring likelihood of 30%. In Panels B-D, we consider outcomes under less selective interview policies. In all cases, cohorts selected by our algorithms are more likely to be hired than those selected by the model trained to mimic human performance. As the selection threshold decreases, the gap between human and algorithmic performance narrows: this comes both from the fact that hiring likelihoods for human-selected applicants actually increases

(reflecting the negative correlation between human model scores and hiring outcomes) and from the fact that hiring likelihoods for algorithm-selected applicants increases (reflecting the positive correlation between model scores and hiring outcomes). Importantly, there do not appear to be significant differences in overall hiring likelihoods between the supervised and reinforcement models, even though, as discussed earlier, these models differ substantially in the diversity of the candidates they select.

Table 4 examines how much these algorithms agree on whom to interview, and which model is correct when they disagree. The first panel compares the supervised ML hire-prediction model to the human interview model. These algorithms tend to disagree, especially among more highly ranked candidates: there is only a 1.7% overlap amongst candidates in the top 10% of each model. The human model performs substantially worse in terms of predicting hiring likelihood when the models disagree: only 3.9% of candidates favored by the human model are eventually hired, compared to 30.0% of candidates favored by the algorithm. The second panel of Table 4 shows that the same pattern holds for humans versus the reinforcement learning model. The final panel, meanwhile, shows that both the supervised and reinforcement models agree on most candidates, with the supervised model having higher predictive accuracy when they disagree.

One may be concerned that the patterns we document on the sample of interviewed candidates cannot be extended to candidates who are not interviewed. This may be the case if there is selection on unobservables. Even though screeners do not interact with candidates, they may have access to more information than we include as model inputs. For example, an applicant may have a strong CV but write a disinterested cover letter. This disinterest is identified by the human screener who rejects the candidate, but not by algorithms trained only on CV inputs or with poor natural language processing. Had the applicant been interviewed, her disinterest would have likely resulted in rejection: dropping this applicant from our analysis would therefore lead us to overstate the accuracy of our supervised and reinforcement hiring scores, relative to the human screener.

We note that this type of selection bias implies that, conditional on an applicant's algorithmic score, applicants with a low predicted probability of being interviewed should be less like to be hired if interviewed. That is, comparing two applicants with the same algorithmic score, having a lower human score should be correlated with some unobservable that makes an applicant less likely to be hired. Table 5 shows that, if anything, the opposite is the case: there is a baseline negative correlation between human scores and hiring likelihood, and this correlation remains negative after controlling for algorithmic scores.

To further explore selection issues in our sample, we consider how our results would change if we vary the extent of selection into our sample of interviewed candidates. That is, suppose we observe interview outcomes for all applicants with a human score of $s^{H*} > \tau$ (where H^* denotes the true

human model, in contrast with our predicted human model score s^H). If τ is smaller, we would observe a less selected sample and if τ were larger, we would observe a more selected sample. We may be concerned that the gap in hiring likelihood between our algorithmic and human models—$E[H|I^S = 1, s^{H*} > \tau] - E[H|I^H = 1, s^{H*} > \tau]$ or $E[H|I^R = 1, s^{H*} > \tau] - E[H|I^H = 1, s^{H*} > \tau]$—may be decreasing in τ and may be negative when $\tau = 0$ and the sample is unselected. While we cannot observe this gap for less selective thresholds, we can redo our analysis for *more* selective thresholds, using s^H as a proxy for s^{H*}. If selection issues bias our estimates of accuracy differences between models, then we would expect this difference to change as we artificially contract our sample to make it more selected. If sample selection leads us to mistakenly conclude that algorithms perform better than humans, then we would expect the relative performance of algorithms to increase as we make the sample more selected.

Figure 7 plots the results of this exercise. The leftmost side of each panel shows hiring likelihoods when there is no sample contraction, that is, for the full set of interviewed applicants (this would be equivalent to the bar charts in Figure 4). Moving to the right, we order applicants in the sample by their human score s^H and gradually eliminate the bottom, in 5% increments up until we have eliminated half the sample. For each sample contraction, we select the top X% of remaining applicants, as ordered by their human and algorithmic scores, and compute average hiring rates for the selected group. Panels A-D plot these results when we select the top 10, 25, 50, and 75% of remaining applicants, respectively. In all cases, we see that the gap between algorithmic performance (the upper cluster of lines) and human performance (the lowest line in each figure) stays approximately constant and, if anything decreases as we contract the sample. The fact that sample selection does not seem to impact our findings within the set of interviewed applicants makes less likely that selection would lead our results to change sign in full applicant sample.

Improving on the status quo

The analysis in the previous sections suggest that firms could improve both the diversity and quality of the applicants they choose to interview by following the recommendations of an exploration-based algorithm. However, without information on the hiring outcomes of candidates who are not actually interviewed, it is impossible to make a direct comparison between $E[H|I = 1]$ and $E[H|I^S = 1]$ or $E[H|I^S = 1]$.

In this section, we take a different approach and show that there exists a Pareto improving alternative interview policy $\tilde{I}$ that is intermediate between the firm's actual practice I and a fully algorithmic policy I^S or I^R. That is, we show that the firm can improve diversity and quality comes by following algorithmic recommendations more *on the margin*.

To show this, we use random variation in assignment to initial resume screeners as an instrument for receiving an interview, following the methodology pioneered by Kling (2006). Applicants in our data are randomly assigned to screeners who review their resumes and make initial interview decisions. These screeners vary greatly in their propensity to pass applicants to the interview round: an applicant may receive an interview if she is assigned to a generous screener and that same applicant may not if she is assigned to a stringent one. For each applicant, we form the jackknife mean pass rate of their assigned screener and use this as an instrument, Z, for whether the applicant is interviewed. We use Z to identify applicants interviewed on the margin: while some applicants are always interviewed and others are never interviewed, compliers are marginal in the sense that they are only interviewed if they are lucky enough to draw a generous screener. With this in mind, we propose the following counterfactual interview policy:

$$\tilde{I} = \begin{cases} I^{Z=1} \text{ if } D^H(X) > \overline{\tau}, \\ I \text{ if } \underline{\tau} \leq D^H(X) \leq \overline{\tau}, \\ I^{Z=0} \text{ if } D^H(X) < \underline{\tau}, \end{cases}$$

where $D^H(X) = f_t^R(X; D_t^R) + B^R(X; D_t^R)$.[14]

The policy $\tilde{I}$ takes the firm's existing interview policy, I, and modifies it at the margin: $\tilde{I}$ favors applicants with high reinforcement model scores ($D^H(X) > \overline{\tau}$) by asking the firm to evaluate these applicants as if they were assigned to a generous screener ($Z = 1$). Similarly, $\tilde{I}$ penalizes applicants with low reinforcement scores ($D^H(X) < \underline{\tau}$) by treating them as if they face a stringent screener ($Z = 0$).

$\tilde{I}$ is constructed so that we can directly estimate its performance relative to I

$$\begin{aligned} E[H|\tilde{I}=1] - E[H|I=1] \quad = \quad & \Pr(I^{Z=1} > I^{Z=0}) E[H|I^{Z=1} > I^{Z=0}, D^H(X) > \underline{\tau}] \\ & - \Pr(I^{Z=1} > I^{Z=0}) E[H|I^{Z=1} > I^{Z=0}, D^H(X) < \overline{\tau}] \end{aligned} \tag{6}$$

Equation (6) follows because $\tilde{I}$ differs from the status quo I only in its treatment of instrument compliers: in this case, $\tilde{I}$ chooses to interview compliers with high reinforcement scores and chooses not to interview compliers with low reinforcement scores. The performance of $\tilde{I}$ (in selecting applicants with greater hiring potential) relative to I therefore depends entirely on whether high reinforcement score applicants who are interviewed on the margin turn out to have greater hiring potential than low score applicants who are also interviewed on the margin.

[14] For simplicity in exposition, we let Z be a binary instrument in this example (whether an applicant is assigned to an above or below median stringency screener) though in practice we will use a continuous variable.

The following proposition, analogous to a result in Benson, Li, and Shue (2019), shows that we can estimate the average hiring likelihoods of applicants interviewed on the margin:

Proposition 1. *Consider the regression*

$$H_i \times I_i = a_{0j} + a_{1j} I_i + \alpha_j X_i + \varepsilon_i \tag{7}$$

for applicants with reinforcement scores $D^H(X_i)$ falling in bin S_j. Suppose that we have a valid binary instrument Z[15] *such that $I_i^{Z_i=1} \geq I_i^{Z_i=0}$ and the exclusion restriction holds ($H_i \perp \{Z_i, I_i^{Z_i=1}, I_i^{Z_i=0}\} | X_i, D^H(X_i)$). Then it is the case that*

$$\hat{a}_{1j}^{IV} = E[H_i | D^H(X_i) \in S_j, I_i^{Z_i=1} > I_i^{Z_i=0}]. \tag{8}$$

See Appendix A for the full proof. This result is analogous to Angrist, Imbens, Rubin (1996), who show that IV estimates identify local average treatment effects for instrument compliers. Following Abadie (2003), Proposition 1 takes this same framework and focuses on estimating local selection rather than treatment effects: we estimate $E[Y^1 | D^{Z=1} > D^{Z=0}]$ rather than $E[Y^1 - Y^0 | D^{Z=1} > D^{Z=0}]$.

Figure A.2 plots the distribution of jackknife interview pass rates in our data, restricting to the 54 recruiters (2/3rd of the sample) who evaluate more than 50 applications (the mean in the sample overall is 156). Some of this variation reflects variation in hiring needs over time and across jobs. However, after controlling for job family, job level, and work location fixed effects, the 75th percentile screener has a 50% higher pass rate than the 25th percentile screener.

Table 6 tests the first stage and exclusion criterion for our instrument. In Column 1, we see that a screener's interview rate is strongly related to an applicant's likelihood of being interviewed.

Figure 8 plots characteristics of marginally interviewed applicants with high and low scores, under our reinforcement learning model of hiring. In Panel A, we see that marginal applicants with high reinforcement scores are more likely to be hired than marginal applicants with low scores. Similarly, Panels B-D show that marginal high score applicants are more likely to be black, hispanic, and female. As such, the interview policy defined by $\tilde{I}$ would increase quality and diversity on the margin, relative to the firm's current practices. Figure 9 repeats this exercise for applicants with high and low supervised learning scores. Consistent with our previous estimates, we find that

[15] In practice, our instrument will be continuous. In the continuous case, Equation (6) uses the notation $E[H|X, \lim_{z' \downarrow z} I^{z'} = 1, \lim_{z' \uparrow z} I^{z'} = 0]$. We focus on the binary case for intuition, because it can be interpreted as a close analogue to the Wald estimate. In general, we require that the probability of promotion is monotonic in the instrument, following Angrist, Imbens, and Rubin (1996) and Heckman and Vytlacil (2005).

marginal applicants with high scores are more likely to be hired, but they are less likely to be black or hispanic.

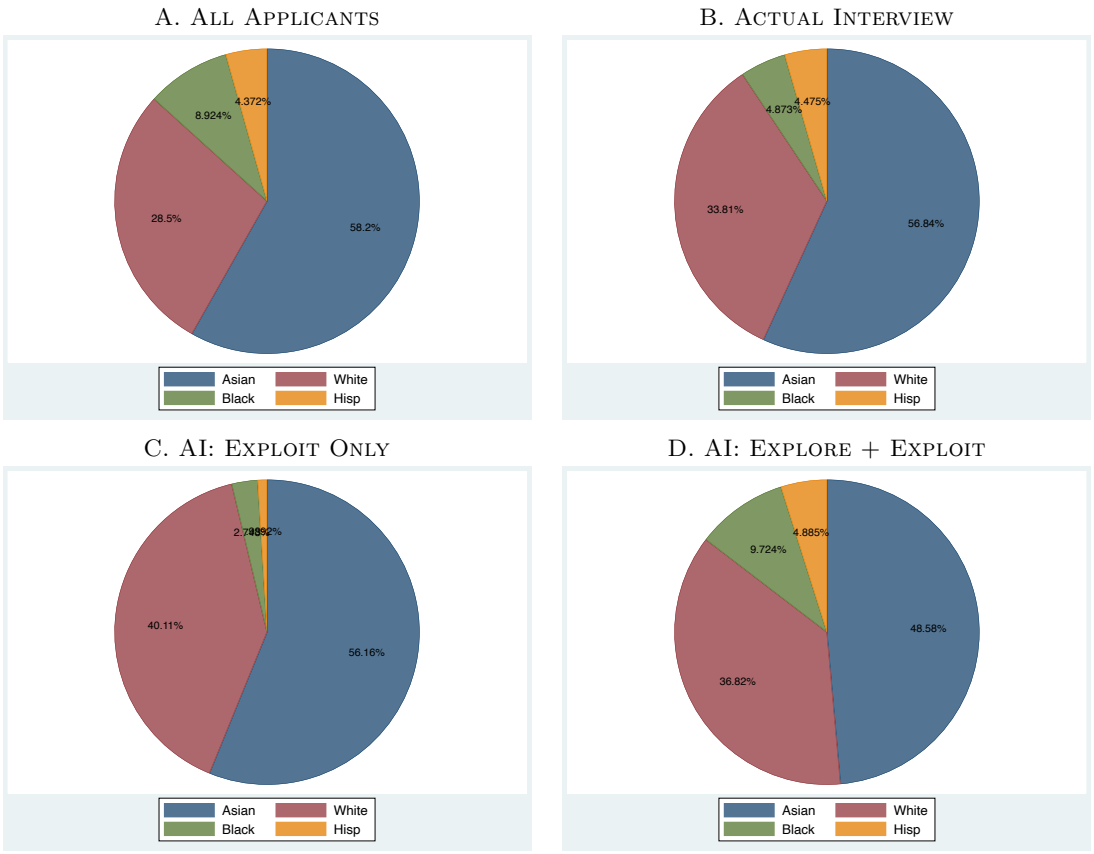
5 Conclusion

Our results highlight the possibility that the use of algorithmic screening tools can improve both the diversity and quality of the applicants firms choose to interview. This result goes against the idea that gains in equity must necessarily come at the expense of efficiency; rather, firms appear to be operating inside the Pareto frontier, making improvements along both dimensions possible.

While a growing body of papers have pointed out potential gains from following algorithmic recommendations, our paper goes further to compare outcomes between standard supervised learning approaches, and an adaptive reinforcement learning based approach. Our finding that these algorithms differ substantially in their diversity impacts (despite having similar positive impacts on quality) highlights the importance of algorithm design in achieving efficiency and equity enhancing outcomes.

Our results are consistent with a range of recent papers pointing out biases in supervised learning approaches. In our case, under-represented minorities in our training data were somewhat less likely to be hired conditional on being interviewed. Training an exploitation only algorithm on such data leads it to avoid these groups, even though quality differences are small. The fact that supervised learning decreases diversity relative to the firm's practices suggests that the firm is engaging in de-facto affirmative action, but doing so in a way that is less efficient—that is, interview weaker under-represented minorities ahead of stronger under-represented minorities. A reinforcement learning approach, by contrast, prioritizes hiring from groups with greater uncertainty (and therefore option value) but appears to be better at identifying stronger candidates within these groups.

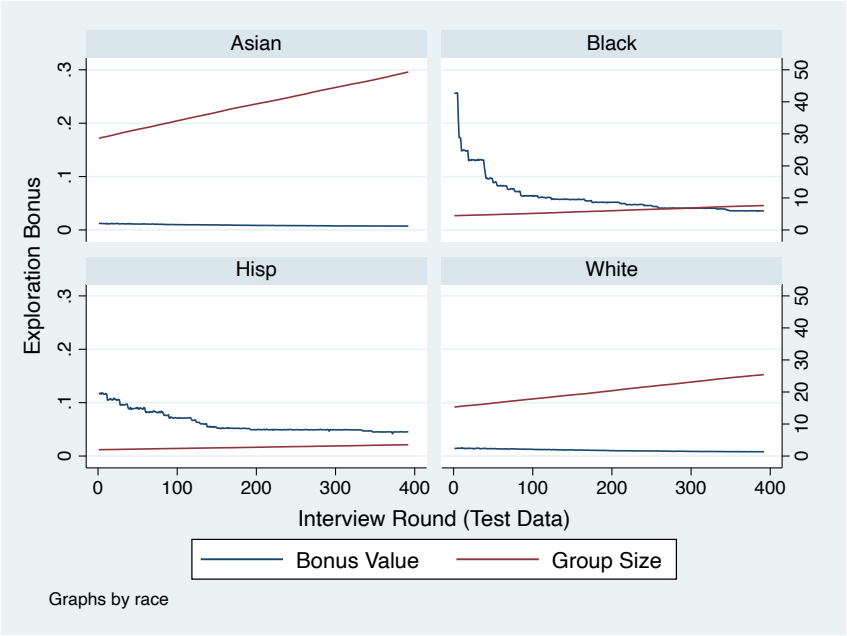
Figure 1: Racial Composition



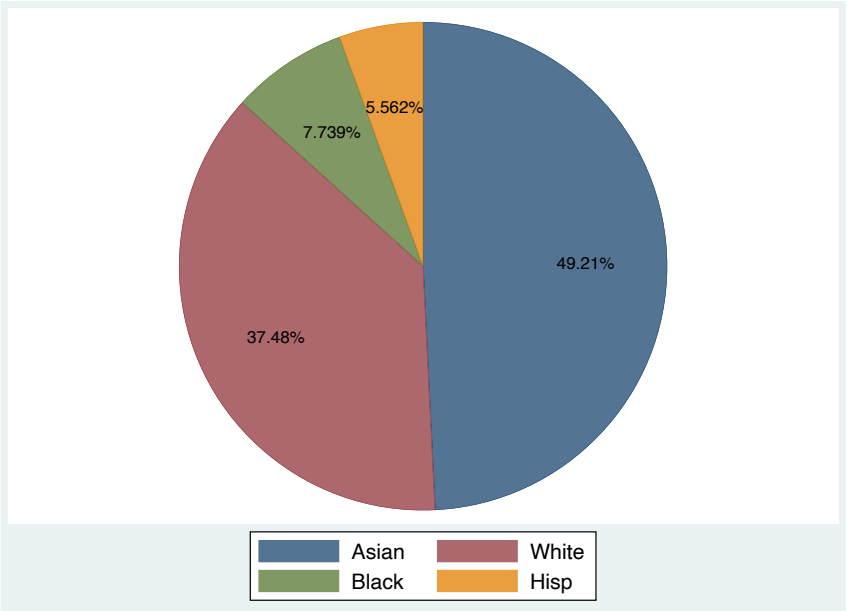
Notes:

Figure 2: Exploration over time, by Race

A. Exploration Bonuses

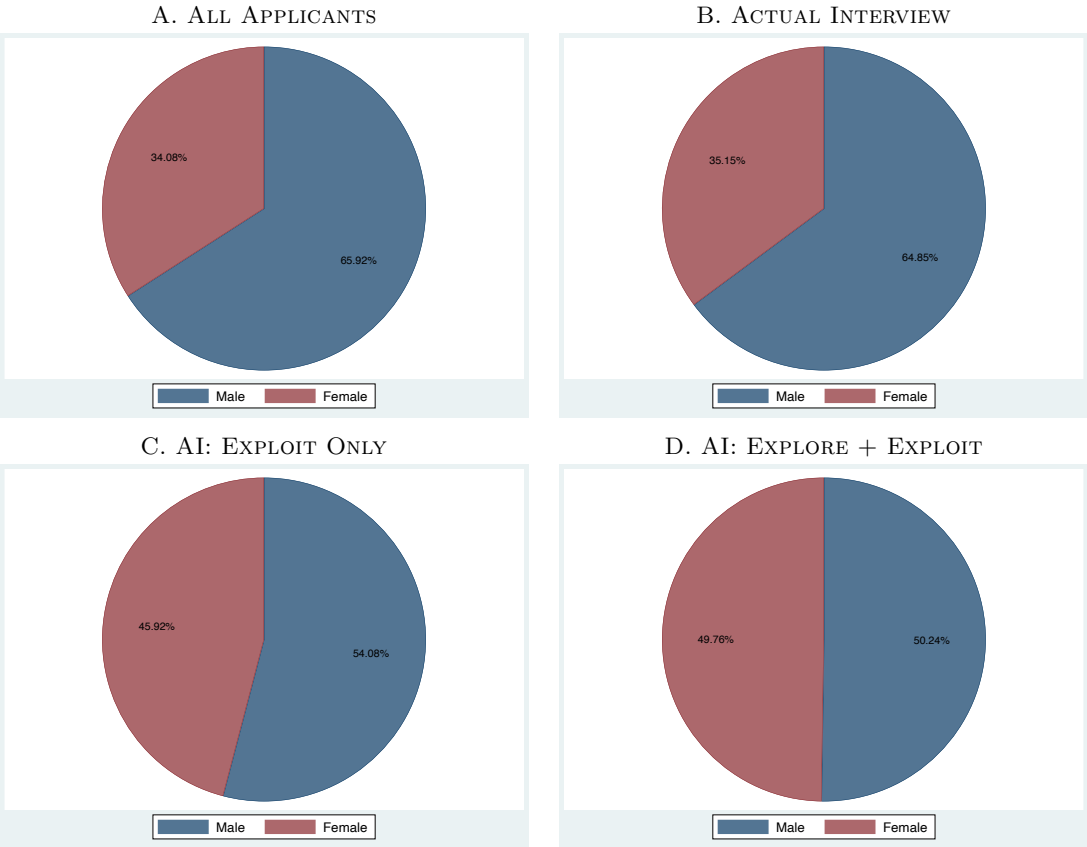


B. Sample End Composition



Notes:

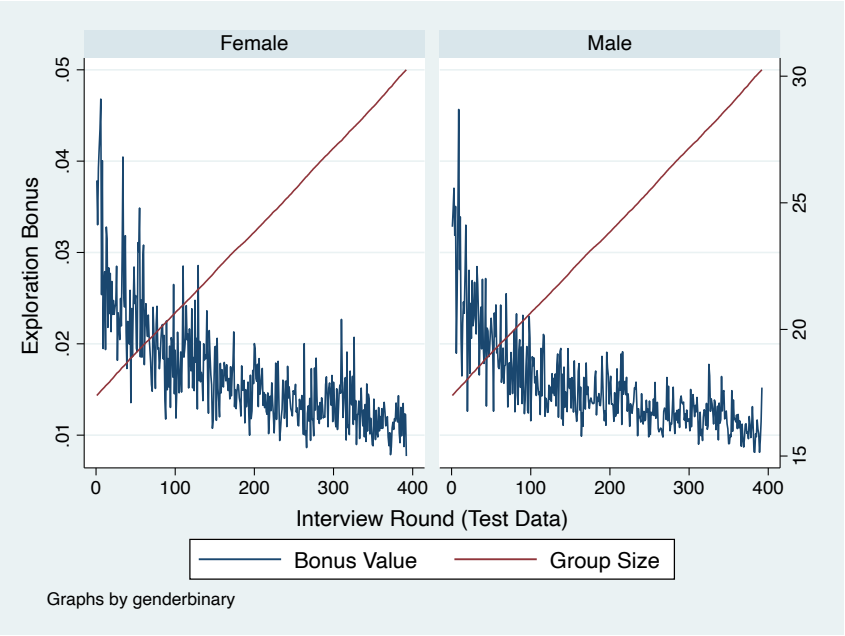
Figure 3: Gender Composition



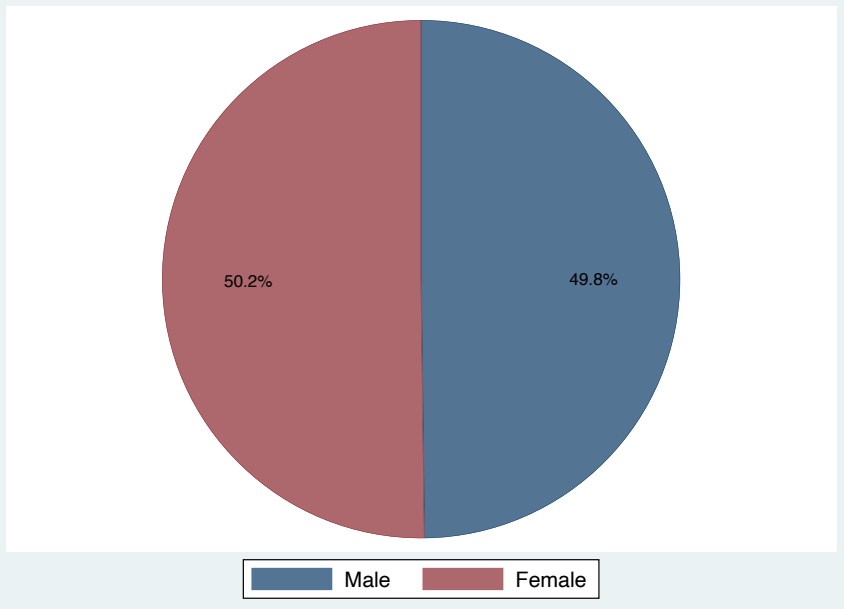
Notes:

FIGURE 4: EXPLORATION OVER TIME, BY GENDER

A. EXPLORATION BONUSES

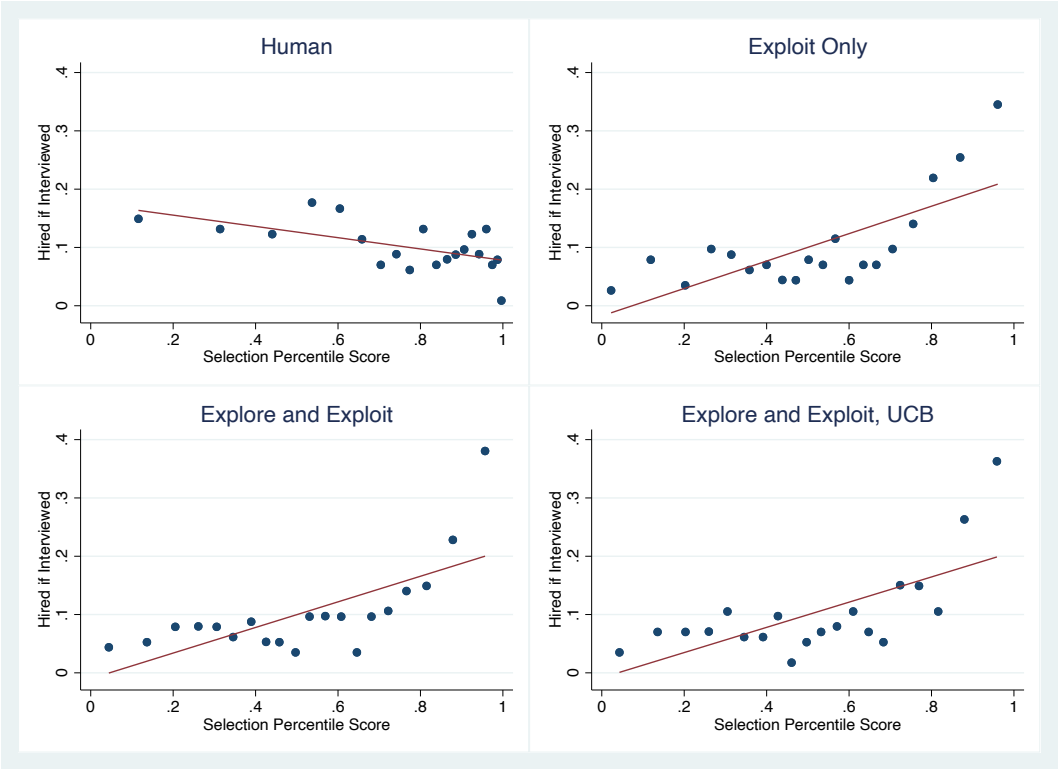


B. SAMPLE END COMPOSITION



NOTES:

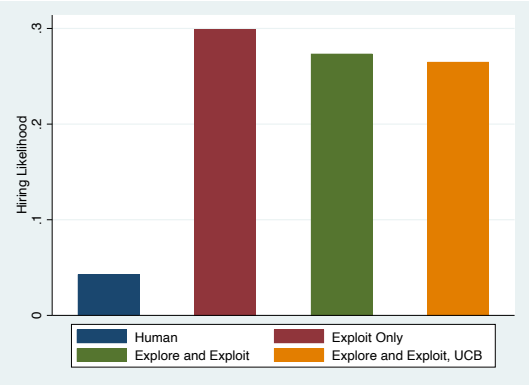
FIGURE 5: CORRELATION BETWEEN SCORES AND APPLICANT QUALITY



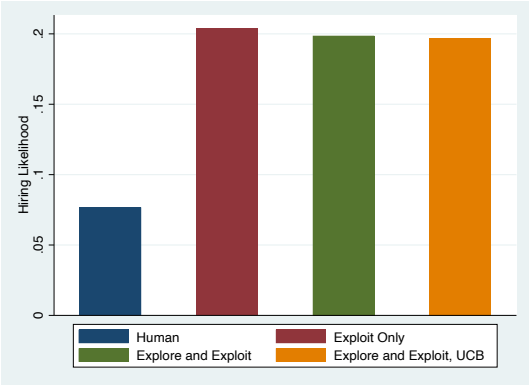
NOTES:

FIGURE 6: AVERAGE HIRING LIKELIHOOD, INTERVIEWED SAMPLE

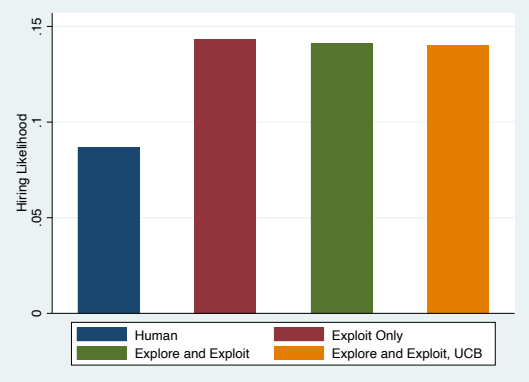
A. SELECT TOP 10%



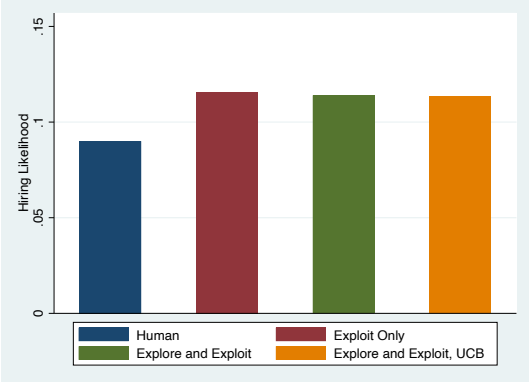
B. SELECT TOP 25%



C. SELECT TOP 50%



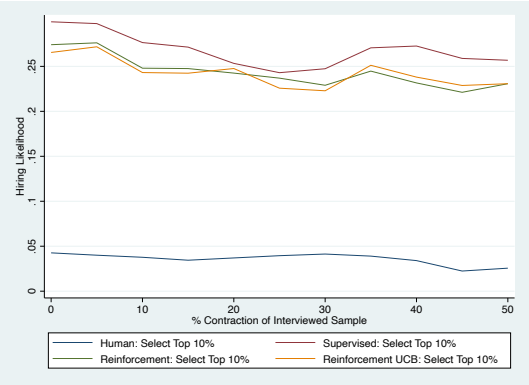
D. SELECT TOP 75%



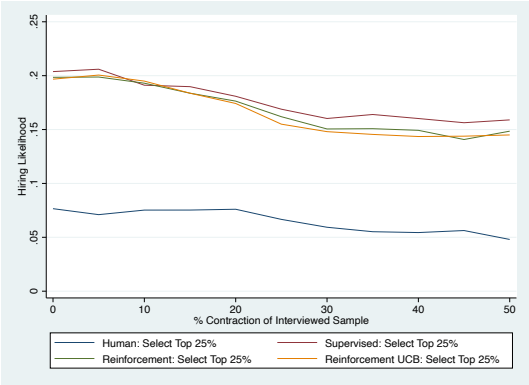
NOTES:

Figure 7: Model Accuracy, Contracting Interviewed Sample

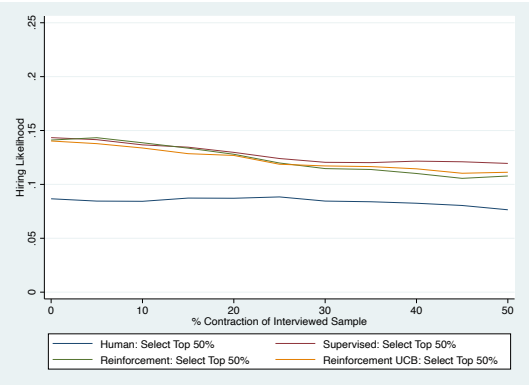
A. Select Top 10%



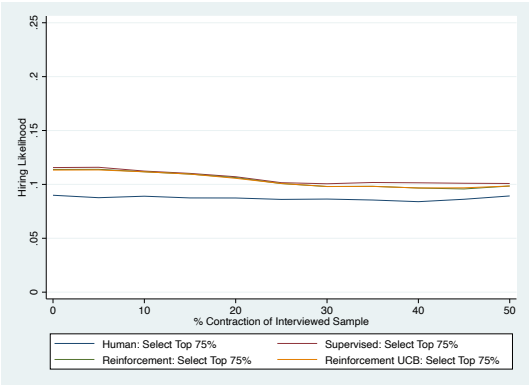
B. Select Top 25%



C. Select Top 50%

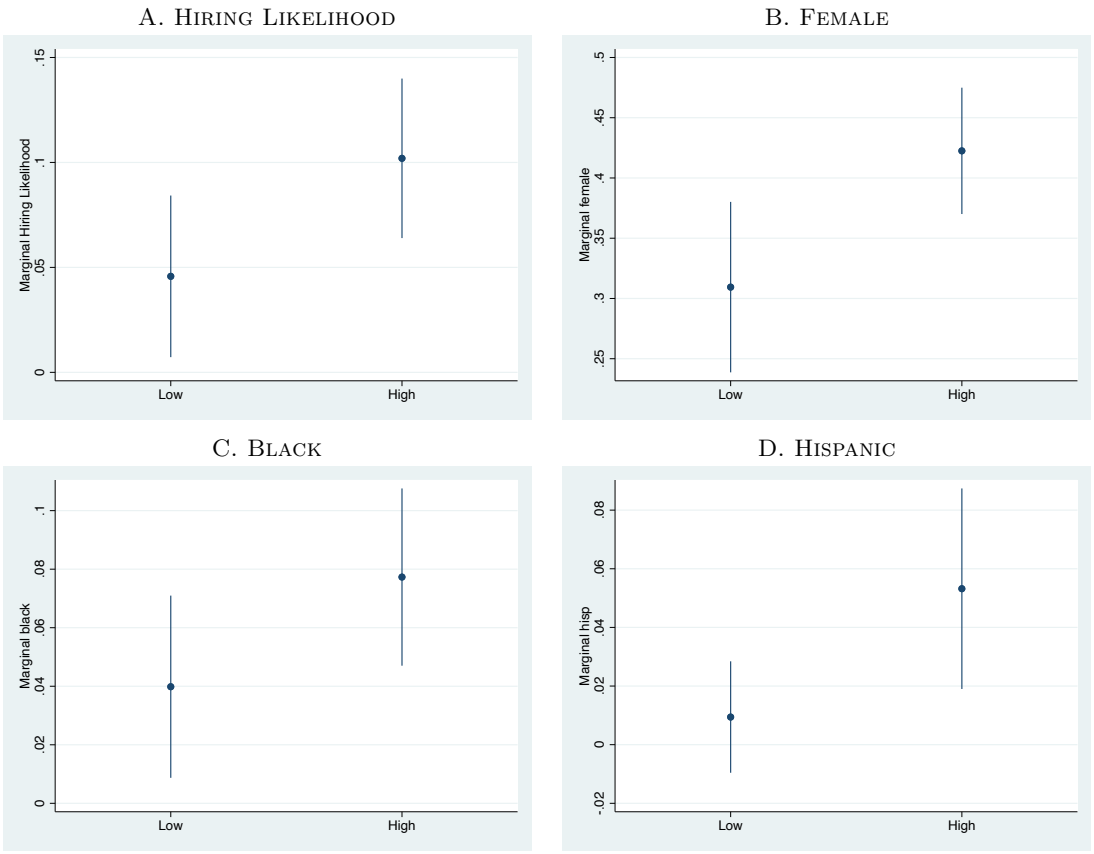


D. Select Top 75%



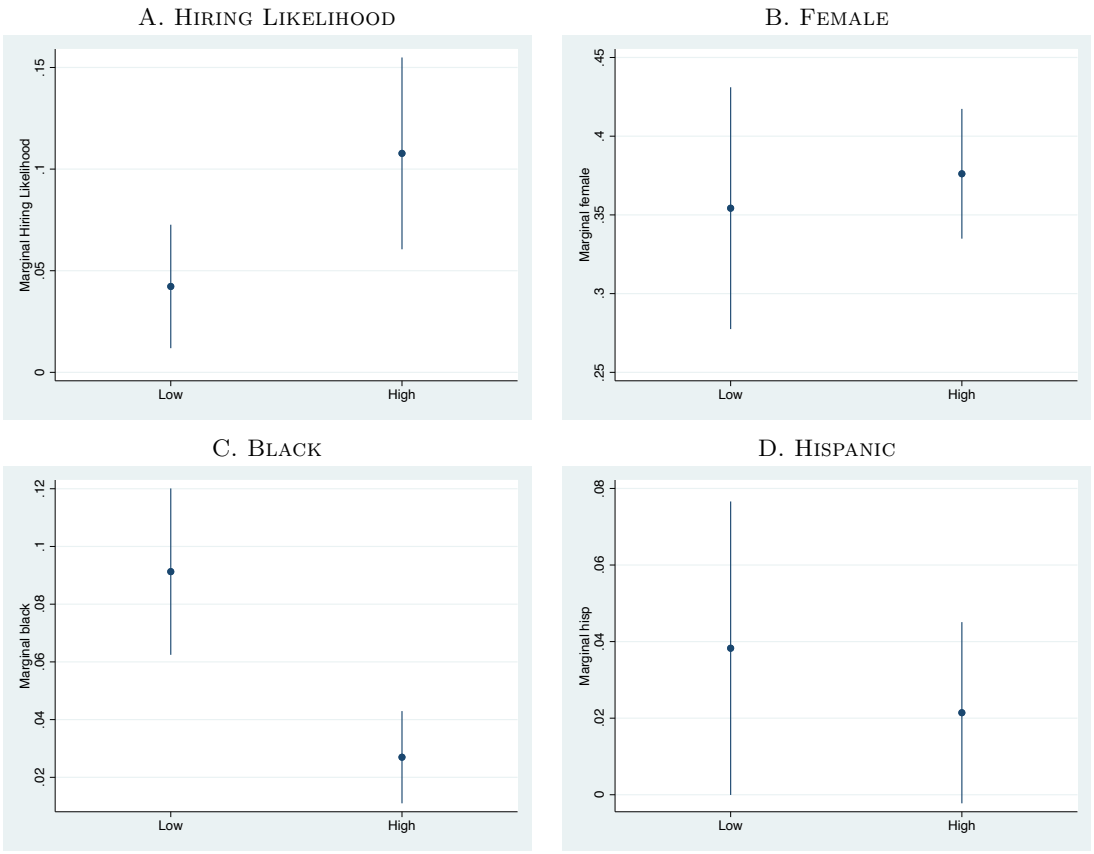
Notes:

Figure 8: Characteristics of marginal interviewees, by Reinforcement Score



Notes:

Figure 9: Characteristics of marginal interviewees, by Supervised Score



Notes:

Table 1: Applicant Summary Statistics

Variable	Mean	Observations
Black Applicants	0.08	100,008
Hispanic Applicants	0.04	100,008
Asian Applicants	0.52	100,008
White Applicants	0.27	100,008
Male Applicants	0.64	100,008
Female Applicants	0.32	100,008
Referred Applicants	0.13	100,008
B.A. Degree	0.24	100,008
Associate Degree	0.01	100,008
Master's Degree	0.62	100,008
Ph.D.	0.07	100,008
Attended a U.S. College	0.98	99,964
Attended Elite U.S. College	0.06	99,964
Worked at a Fortune 500 Co.	0.02	99,789
Has a Quantitative Background	0.24	99,789

Table 2: Job Funnel Summary Statistics

Variable	Observations	Share of Previous Rd.
All Job Families		
Applications	100,008	
Interviewed	5,351	0.05
Hired	560	0.10
Business Analysts		
Applications	39,133	
Interviewed	3,552	0.09
Hired	270	0.08
Data Analysts		
Applications	24,902	
Interviewed	839	0.03
Hired	192	0.23
Data Scientists		
Applications	24,202	
Interviewed	487	0.02
Hired	23	0.05
Financial Analysts		
Applications	11,771	
Interviewed	473	0.04
Hired	75	0.16

Table 3: Applicant Funnel

Variable	Observations	Share of Previous Rd.
Black Applicants		
Applications	7,944	
Interviewed	218	0.03
Hired	8	0.04
Hispanic Applicants		
Applications	3,714	
Interviewed	210	0.06
Hired	16	0.08
Asian Applicants		
Applications	52,323	
Interviewed	2,770	0.05
Hired	308	0.11
White Applicants		
Applications	26,503	
Interviewed	1,562	0.06
Hired	190	0.12
Female Applicants		
Applications	31,896	
Interviewed	1,633	0.05
Hired	191	0.12
Male Applicants		
Applications	64,402	
Interviewed	3,419	0.05
Hired	359	0.11

Table 4: Impact on New Development

Human vs. Supervised Learning

Selectivity (Top X%)	Overlap %	SL and Human	Human Only	SL Only	Human All	SL All
	(1)	(2)	(3)	(4)	(5)	(6)
10	1.71	25.00	3.91	30.00	4.27	29.91
25	14.98	13.95	6.56	21.52	7.67	20.38
50	47.81	12.27	5.37	16.22	8.67	14.34
75	78.62	10.01	4.92	16.94	9.00	11.57

Human vs. Reinforcement Learning

Selectivity	Overlap %	RL and Human	Human Only	RL Only	Human All	RL All
	(1)	(2)	(3)	(4)	(5)	(6)
10	2.99	14.29	3.96	27.75	4.27	27.35
25	17.42	15.00	6.12	20.89	7.67	19.86
50	47.63	11.21	6.35	16.75	8.67	14.11
75	78.27	9.78	6.18	17.20	9.00	11.39

Supervised vs. Reinforcement Learning

Selectivity	Overlap %	SL and RL	SL Only	RL Only	SL All	RL All
	(1)	(2)	(3)	(4)	(5)	(6)
10	68.38	33.75	21.62	13.35	29.91	27.35
25	74.39	23.19	12.24	10.20	20.38	19.86
50	78.93	16.50	6.22	5.04	14.34	14.11
75	91.06	11.93	7.84	5.88	11.57	11.39

Notes:

Table 5: Correlation between human scores and hiring likelihood

	(1)	(2)	(3)
Human Score	-0.0967	-0.106**	-0.0652***
	(0.0630)	(0.0219)	(0.00912)
R-squared	.006	.045	.092
Linear Algorithmic Score Controls	No	Yes	No
FE Ventiles of Algorithmic Scores	No	No	Yes
Observations	2275	2275	2275

Notes:

Table 6: Instrument Validity

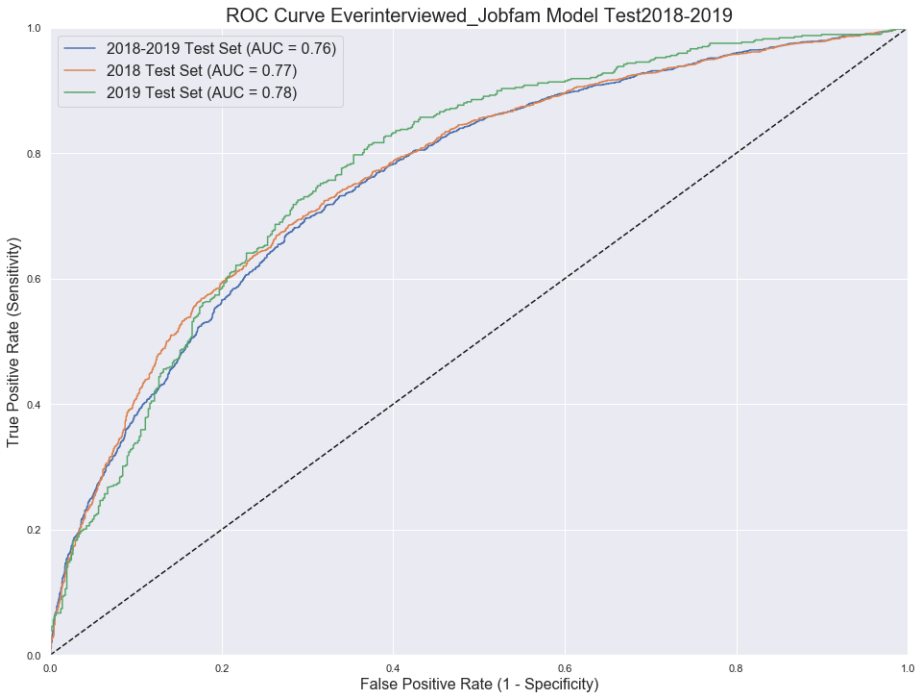
	Int.	Black	Hisp	Asian	White	Female	Ref.	MA	Elite	F500
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
JK interview rate	0.0695***	-0.000317	-0.000220	0.00671	-0.00579	-0.00439	0.00766	-0.0105	0.00321	0.000305
	(0.00543)	(0.00366)	(0.00168)	(0.00974)	(0.00711)	(0.00404)	(0.00662)	(0.00961)	(0.00222)	(0.000676)
Pre-promotion controls										
Pseudo R-squared										
Observations	37662	37662	37662	37662	37662	37662	37662	37662	37646	37519

Notes:

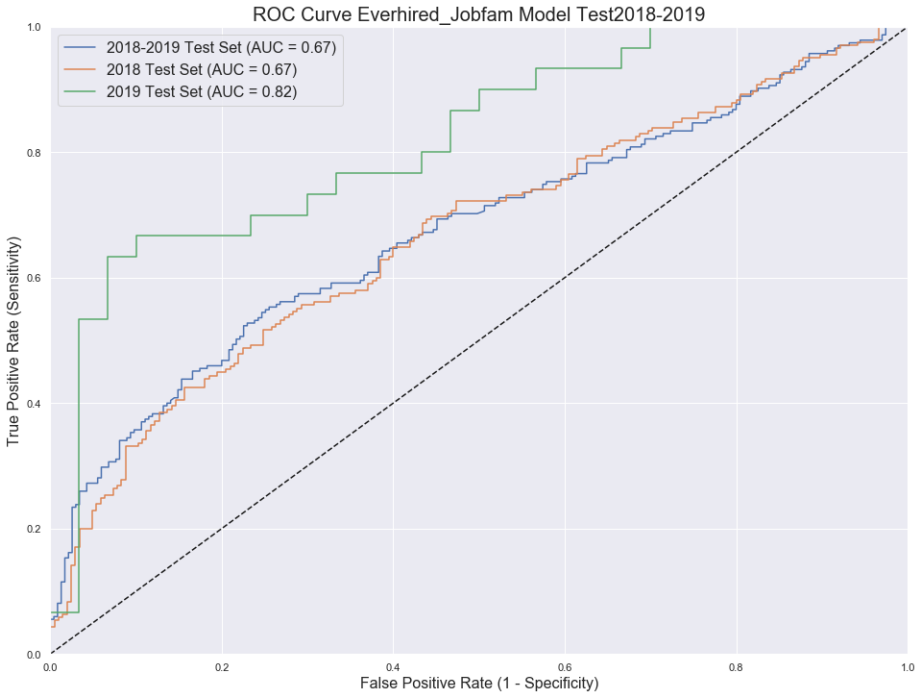
6 Appendix

FIGURE A.1: MODEL PERFORMANCE: ROC

A. INTERVIEW MODEL

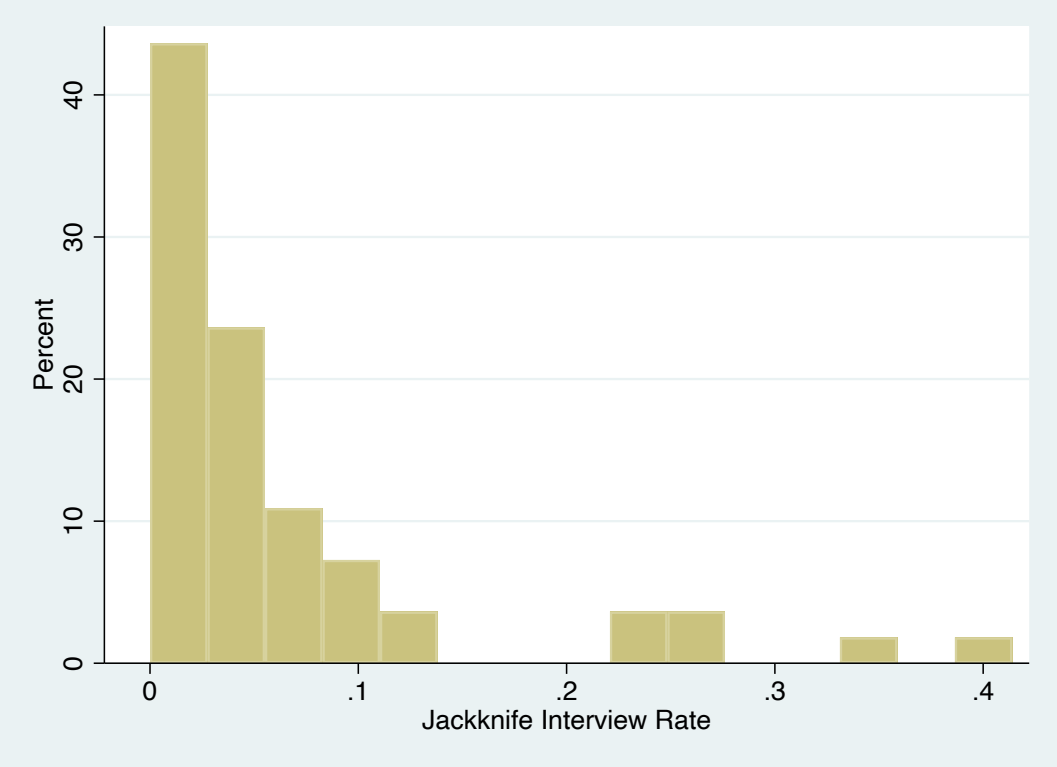


B. ACTUAL HIRING MODEL



NOTES:

FIGURE A.2: DISTRIBUTION OF INTERVIEW RATES



NOTES: Average of jack-knife interview rates for the 54 screeners in our data who evaluate more than 50 applicants.

A Proof of Proposition 1

Proof. For legibility, we write $I^1 = P^{Z=1}$ and $I^0 = P^{Z=0}$. We suppress subscripts i for individuals, as well as subscripts j, which denote bins of AI score. Finally, we suppress the existence of covariates X for simplicity. Given this set up, we have, by definition, that

$$\begin{aligned} a_1^{IV} &= \frac{Cov(HI, Z)}{Cov(I, Z)} \\ &= \frac{E[MPZ] - E[HI]E[Z]}{E[IZ] - E[I]E[Z]} \end{aligned}$$

$E[IZ]$ can be written as $E[IZ|Z=1]Pr(Z=1) + E[IZ|Z=0]Pr(Z=0)$, which, canceling terms and rewriting $Pr(Z=1) = E[Z]$, is equivalent to $E[I|Z=1]E[Z]$. This means that we can rewrite

the denominator of the expression above as:

$$\begin{aligned}
E[IZ] - E[I]E[Z] &= E[I|Z=1]E[Z] - E[I]E[Z] \\
&= E[I|Z=1]E[Z] - E[Z]\big(E[I|Z=1]Pr(Z=1) + E[I|Z=0][1-Pr(Z=1)]\big) \\
&= E[I|Z=1]E[Z] - E[Z]\big(E[I|Z=1]E[Z] + E[I|Z=0](1-E[Z])\big) \\
&= E[I|Z=1]E[Z] - E^2[Z]E[I|Z=1] - E[I|Z=0](1-E[Z])E[Z] \\
&= E[I|Z=1]E[Z]\big(1-E[Z]\big) - E[I|Z=0](1-E[Z])E[Z] \\
&= \big(1-E[Z]\big)E[Z]\big(E[I|Z=1] - E[I|Z=0]\big)
\end{aligned}$$

Similarly, the numerator can be written as

$$E[HIZ] - E[HI]E[Z] = \big(1-E[Z]\big)E[Z]\big(E[HI|Z=1] - E[HI|Z=0]\big)$$

This means that we have

$$a_1^{IV} = \frac{E[HI|Z=1] - E[HI|Z=0]}{E[I|Z=1] - E[I|Z=0]}$$

So far, we have not assumed that Z is a valid instrument. Now, we show that we can interpret a_1^{IV} as an estimate of the hiring potential of marginally interviewed applicants if Z is a valid instrument. That is, we want to show:

$$a_1^{IV} = \frac{E[HI|Z=1] - E[HI|Z=0]}{E[I|Z=1] - E[I|Z=0]} = E[M|I^1 > I^0]$$

To see this, note that we can express a person's interview status in terms of potential outcomes:

$$I = I^1 Z + I^0(1-Z)$$

where I^1 is a person's interview status in a world where he has a high value of the instrument and I^0 is a person's promotion status if he has a low value of the instrument. If Z is a valid instrument, it is orthogonal to a person's potential outcomes: Z is orthogonal to hiring outcomes if interviewed (H), as well as his interview status under differ values of the instrument (I^1 and I^0). The latter means that a person's "sensitivity" to the instrument does not impact the value of the instrument

that he receives. Given this, we can express

$$\begin{aligned} E[I|Z=1] - E[I|Z=0] &= E[I^1 Z + I^0(1-Z)|Z=1] - E[I^1 Z + I^0(1-Z)|Z=0] \\ &= E[I^1|Z=1] - E[I^0|Z=0] \end{aligned}$$

Using similar reasoning, it follows that:

$$E[HI|Z=1] - E[HI|Z=0] = E[HI^1|Z=1] - E[HI^0|Z=0]$$

Finally, we have:

$$\begin{aligned} a_1^{IV} &= \frac{E[HI|Z=1] - E[HI|Z=0]}{E[I|Z=1] - E[I|Z=0]} \\ &= \frac{E[MI^1|Z=1] - E[MI^0|Z=0]}{E[I^1|Z=1] - E[I^0|Z=0]} \\ &= \frac{E[MI^1] - E[MI^0]}{E[I^1] - E[I^0]} \quad \text{(assuming conditional independence of } Z\text{)} \\ &= \frac{E[M(I^1 - I^0)]}{E[I^1 - I^0]} \\ &= \frac{E[M|I^1 > I^0]Pr(I^1 > I^0) - E[M|I^1 < I^0]Pr(I^1 < I^0) + E[0|I^1 = I^0]Pr(I^1 = I^0)}{E[1]Pr(I^1 > I^0) - E[1]Pr(I^1 < I^0) + E[0]Pr(I^1 = I^0)} \\ &= \frac{E[M|I^1 > I^0]Pr(I^1 > I^0)}{Pr(I^1 > I^0)} \quad \text{(assuming monotonicity of } Z\text{)} \\ &= E[M|I^1 > I^0] \end{aligned}$$

■