The Diminishing Returns to Human Recruiting in Online Labor Markets

Emma Wiles* Boston University

John Horton† MIT & NBER

John Fallon ‡ Boston University

May 12, 2026

Most recent draft here.

Abstract

Employers often rely on outside recruiters to find workers, but it is unclear whether human intermediaries add value when employers already have access to algorithmic screening tools. We study a randomized experiment in a large online labor market that assigned human recruiting assistance to job postings. Treated employers received 16% more applications and conducted 35% more interviews but were no more likely to hire than control employers with access only to algorithmic tools. Treated employers spent less on hires and their workers completed fewer hours, suggesting match quality declined. We develop a model of delegated recruiting in which recruiters and employers rely on a common noisy signal, which can generate these patterns when their assessments are highly correlated. Consistent with this mechanism, recruited applicants are positively selected on observable characteristics yet do not improve hiring outcomes. These findings suggest that as algorithmic screening improves, the scope for intermediaries to add value shrinks because it becomes harder to access independent information.[1]

*emma.b.wiles@gmail.com

†johnjosephhorton@gmail.com

‡jfallon899@gmail.com

[1] We would like to thank the participants of the Columbia's Management, Analytics, & Data Conference, MIT's Conference on Digital Experiments, the Workshop on Information Systems & Economics, and Edward Wiles for helpful feedback. The content is solely the responsibility of the authors and does not necessarily represent the official views of Boston University, MIT, or the NBER.

1 Introduction

Hiring is costly, and firms routinely use outside help to find workers. Roughly two-thirds of US hiring managers report having used external staffing firms (Staffing Industry Analysts, 2022), and Human Resource (HR) and recruitment services generate nearly \$900 billion per year (IBISWorld, 2024). Theoretically, intermediaries diversify sampling risk, as recruiters and re-use information from prior searches (Bull et al., 1987; Howitt and McAfee, 1987). More intense searches fill vacancies more often and intermediaties offer an immediate escalation in intensity (Davis et al., 2013, 2012; Gavazza et al., 2018). Firms invest substantially in interviewing and screening (Barron et al., 1985), in hopes of identifying good candidates from available signals (Pellizzari, 2005; Jovanovic, 1979). Much of the hiring process now happens online either through job boards or applicant tracking systems (Carnevale et al., 2014). Therefore, algorithms increasingly perform parts of the search and screening process that were once central the role of human intermediaries like ranking applicants and recommending matches. As algorithmic matching improves, under what conditions do human recruiters still add value?

We study this question in an online labor market. In online labor markets, Employers must evaluate workers they have never met, relying heavily on platform-verified histories and ratings (Pallais, 2014; Stanton and Thomas, 2015).[2] The platform randomly assigned human hiring consultants to 83,017 job posts, with 25% serving as a control. Control employers had access to the platform's existing tools: worker search, algorithmic recommendations, detailed profiles with ratings and work history, and standard matching features. The key question is whether and under what conditions hiring consultants add value to employers beyond these algorithmic tools.

The intervention generated a strong first stage on both search and screening margins. Treated job posts received 16% more applications and 35% more interviews, driven by recruiting invitations and shortlisting. Treated posts received 3.2 more recruiting invitations on average; employers' own invitations remained unchanged. Total applications rose by 2.6, with recruited applications up 1.4 and organic applications up 1.2. Consultants also delivered selection help, with treated posts receiving more shortlisted applicants from both consultants and employers. Treated employers increased their own shortlisting effort in response, suggesting consultants complemented rather than crowded out employer engagement.

Despite this activity, employers assigned hiring consultants were no more likely to make a hire. About 29% of job posts led to a hire in both groups. We find an insignificant treatment effect of 0.6%. The null effect held for experienced and new employers alike. The recruiters may have simply found a better candidate to hire, but we find that human assistance actually harmed match quality. Treated employers spent about -9% less on hires in the first 30 days, and their hires worked significantly fewer total hours. The effect is concentrated among hourly jobs where wagebill proxies for the duration of the employment relationship. Recruited applicants are positively selected on some observable platform characteristics (engagement, prior

[2] We use the terms "worker," "job-seeker," "job post," and "application" for consistency with the economics literature and not as a commentary on the legal nature of the relationships created on the platform.

employer interest) but negatively selected on others (price, cumulative earnings), yet they do not increase the probability of a hire, suggesting they fit specific jobs worse along dimensions only the employer can assess.

The platform's algorithmic recommendations already do much of the work. A regression discontinuity design exploiting the discrete recommendation label shows that applicants just above the threshold are significantly more likely to be hired, echoing Horton (2017). The RD design reveals that recommendation has a significant positive effect on hiring when applicant pools are thin, but no detectable effect in larger pools. If employers learn their preferences through candidate comparisons (**?**), thin pools leave them least calibrated and most in need of outside guidance.

The null effect is not uniform across all job types. Customer service and the lowest expertise tier, categories where algorithmic scores plausibly capture less of worker quality, show positive hiring effects that survive multiple-hypothesis corrections.

Nevertheless, across the sample, the program cannot justify its cost. The upper bound of the 95% confidence interval on the hiring effect implies expected revenue per job post of about \$5.86, which falls dramatically short of estimated program costs with estimated net returns of between \$-1.95 and \$-7.16.

To our knowledge, this is the first field experiment testing whether hiring assistance helps employers. Active labor market programs have been studied extensively (Card et al., 2018), and job search assistance helps workers find jobs (Card et al., 2010; Belot et al., 2018; Schiprowski, 2020; Crépon et al., 2013), but no prior experiment tests whether similar help works on the employer side. Why might assistance help workers but not employers? Workers benefit from basic guidance on applications and job targeting, and intermediaries provide substantial value to workers seeking their first platform job by helping them overcome the cold-start problem of having no reputation (Stanton and Thomas, 2015). Employers are different. They can likely establish a track record more easiliy and will have to search for employees more frequently than any one employee will have to search for a job.

Search on the employer/demand side has received less attention. Black et al. (2024) document that firms actively search for talent and that outbound recruiting yields different candidates than inbound applications, and Li et al. (2025) frame hiring as an exploration problem in which firms learn about match quality over time. Audit studies document how employers respond to applicant characteristics (Bertrand and Mullainathan, 2004; Bertrand and Duflo, 2017; Kline and Walters, 2021; Kessler et al., 2019; Quadlin, 2018; Gaddis, 2015, 2018), and field experiments have tested diversity interventions in recruiting (Flory et al., 2021; Chan and Wang, 2018; Giuliano et al., 2009; Abel, 2024; Benson et al., 2021). Prior research has examined algorithmic decision aids and human recruiters as separate forms of hiring assistance. Hoffman et al. (2018) find that managers who override algorithmic test recommendations make worse hires on average, while Cowgill and Perkowski (2024) shows that third-party recruiters place weight on both employer preferences and yield management. We contribute by studying human recruiters and algorithmic hiring tools in the same market. This allows us to ask not whether algorithms or recruiters can shape hiring in isolation, but whether human intermediation adds value when employers already have access to algorithmic recommendations, platform histories, ratings, and search tools. Our null results suggest that even trained recruiters working with real candidates cannot improve on what employers armed with good algorithmic tools can do themselves.

Theory suggests that delegation is valuable when the agent holds superior information (Aghion and Tirole, 1997; Dessein, 2002; Alonso and Matouschek, 2008), a condition we argue is increasingly difficult to satisfy as algorithmic improvements erode the information advantages that human intermediaries once held. Employers could delegate to hiring consultants in several ways: advice on platform use, editing job posts, recruiting workers, reminding employers to interview, or recommending specific candidates. The form of help was left to the consultant's judgment. This design resembles real-world recruiting services, where consultants adapt their help to each situation, mirroring prior field experiments that delegate broadly to professionals (Bloom et al., 2013; Friebel et al., 2022), which generally find at least short-run gains. Unlike those settings, we find no positive effect, consistent with the information advantage that makes delegation valuable having been eroded by the platform's algorithmic infrastructure.

Prior work finds that referred workers are positively selected, hired at higher rates, and better matched to their jobs (Pallais and Sands, 2016; Brown et al., 2016; Burks et al., 2015). The leading explanation is that referrers hold private information about candidates that employers cannot directly observe. Algorithms may erode exactly this information advantage. Often, when evaluating the same applicant pool, these tools match or outperform human judgment (Kleinberg et al., 2018; Cowgill, 2019; Kuncel et al., 2013). In practice, the question is less whether algorithms replace human judgment than whether humans can complement technological assistance (**?**). We test this in the context of hiring and find that the null result persists.

We develop a model of delegated recruiting to formalize these mechanisms. A worker's value has two components. The first is captured by the platform's algorithmic score. The second reflects quality orthogonal to the score — skills like communication or reliability that do not show up in platform histories. This setting presents two challenges. First, the platform's score is noisy. Second, given that both recruiters and employers are human, they may be susceptible to the same errors, leaving their assessments correlated.

The model yields three results. First, recruiting has an ambiguous effect on the likelihood of making a hire. More candidates raise the chance of finding an acceptable match, but the employer raises her acceptance threshold in response. These effects roughly cancel when the ratio of recruited candidates to algorithm recommended candidates is small, consistent with our near-zero hiring effect. Second, when recruiter and employer signals are correlated, recruiting may harm quality. Once hired, the noise washes out and performance disappoints, explaining the lower wagebill and hours. Third, as algorithms make more of a worker's skill observable from her profile, the algorithmic pool becomes better sorted. The workers recruiters can find are increasingly those the platform correctly passed over, and algorithmic improvement narrows the domain where recruiters add value while widening the potential for harm.

The rest of the paper proceeds as follows. Section 2 describes the empirical context and experimental design. Section 3 documents that recruiters were highly active. Section 4 reports the main results on hiring: the null effect across specifications, treatment effects on match quality, propensity score evidence on recruiting versus shortlisting, and a cost-benefit analysis. Section 5 develops the theoretical model. Section 6 examines heterogeneity and additional evidence supporting the model's mechanisms, including a regression discontinuity design around the platform's recommendation threshold, time-to-hire, and subgroup effects. Section 7 concludes.

2 Setting and Experimental Design

2.1 Platform

This experiment took place on an online labor market. In online labor markets, employers search for and hire workers to complete jobs that require only a computer and an internet connection. These markets differ in their scope and focus, and platforms provide different services to employers and workers. Common services include soliciting and promoting job openings, hosting profile pages, processing payments, certifying worker skills, and maintaining a reputation system (Horton, 2010; Horton et al., 2011, 2018; Filippas et al., 2018; Roth, 2018). Information frictions are central to how these markets function (Hornstein et al., 2011; Rogerson et al., 2005). Public work histories and detailed evaluations can substantially improve workers' employment outcomes (Pallais, 2014; Agrawal et al., 2016; Benson et al., 2019), and platform design choices about what information to reveal shape both sides of the market (Fradkin, 2017; Zheng et al., 2016). Standardized information can reduce the role of informal signals (Sameen and Cornelius, 2015; Lang et al., 2011), while employer search behavior on these platforms reflects both the quality of available information and the cost of screening (Horton and Johari, 2015; Horton, 2019).

In our empirical setting, employers post job openings on the platform website with job descriptions, required skills, and scope of project. The employer then categorizes the job, for example, as "Administrative Support", "Data Entry", "Software Development", among others. Jobs can either be one-off projects called "fixed price jobs" or hourly jobs, in which case the employer gives an estimate for how many hours they expect the job to take. Our experimental sample includes both contract types. Section 4.2 shows that the cleanest match-quality reads come from the hourly subsample because hourly contracts allow the employer to terminate poor matches, so realized hours and wagebill track match quality directly.

Workers find out about job openings in three ways. They can use electronic search to seek job posts in specific categories or for job openings requiring specific skills. They can receive email notifications from the platform when someone posts a job in a particular category. And finally, they can receive invitations from employers to apply to specific jobs.

Employers find out about workers two ways. They receive *organic* applications from workers who find the job opening independently, or they search for workers themselves and invite specific workers to apply. Employers can search through worker "profiles." These profiles contain workers' history of work on the platform (jobs, hours, hourly rates, ratings) as well as their education history and skills. For both workers and employers, the platform verifies some of the information available to the other side of the market. Employers show particular interest in past experience on the platform (Pallais, 2014; Leung, 2018), and generally look for signals to overcome information asymmetries (Stanton and Thomas, 2015). Wage setting on these platforms has features of both competitive and monopsonistic markets (Dube et al., 2020; Chen and Horton, 2016).

When a worker chooses to apply to a job opening, they submit an application with a cover letter and an hourly wage bid or a total project bid for fixed price jobs. As the employer collects worker applications, they can choose to interview applicants and eventually make an offer. When workers make an offer, employers can make a counteroffer for the wage, however about 90% of hired workers accept the wage they initially proposed (Barach and Horton, 2021). To complete the work on hourly jobs, workers install custom tracking software that serves as a digital punch clock. The software records not only the time spent working, but also keystroke count and mouse movements. The software also captures images of the worker's computer screen at random intervals. This information goes to the platform's servers and becomes available to the employer for monitoring in real time.

2.2 Hiring Process

The hiring process on the platform follows the same steps as in a conventional labor market. A would-be employer posts a job opening with a job title, job description, list of desired skills, and category of work. Once the employer submits the job opening, the platform reviews it and posts it publicly to the marketplace. Once the job begins to receive applications, the employer can view all applications.

Figure A.1 in Appendix A.7 shows a stylized version of the interface. In the first tab, the employer could view the job they posted. The second tab shows where employers can search the platform for available workers and invite desirable candidates. Note that employers can see each worker's wage bid, name, self-reported skills, and a few pieces of platform-verified information, such as total hours worked and average feedback rating from previous projects (if any). The third tab shows where employers can either view all of their applications or only their "shortlisted" applications. The employer can screen applications by asking applicants questions through the messaging feature on the interface and by scheduling phone or video interviews. After screening, the employer decides whether or not to make an offer(s). In the control group 29% of job posts lead to a hire, suggesting employers need to search for good matches.

2.3 Experimental Design

In October 2021, the platform ran a six month experimental evaluation of the hiring assistance program. The platform had operated a version of this hiring assistance program for years prior, but ran this experiment to test its efficacy (marketplace experiments of this kind present design challenges discussed in Blake and Coey (2014); Horton (2017).) The sample included all high-value job openings (both hourly and fixed-price) over the experimental period requiring 30 or more hours of work per week for at least 3 months.

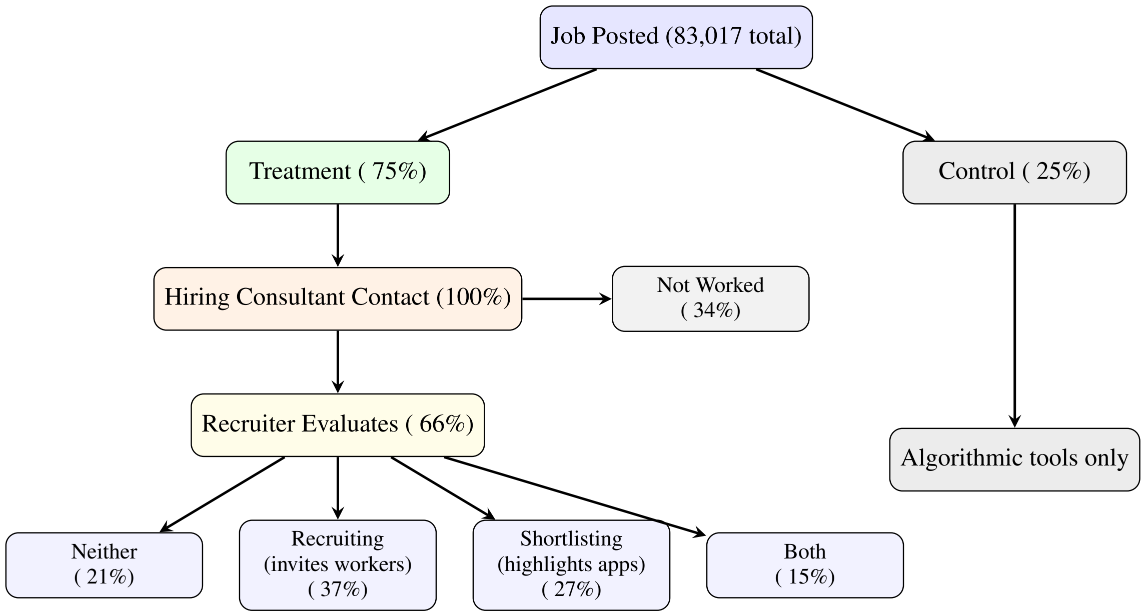
After allocation into the experiment, 75% of job posts went to the treatment group and 25% to the control group.[3] The unit of randomization was the job post. As such, employers who posted multiple jobs over the course of the experiment could have had hiring assistance for one job but not another. The sample contains 83,017 job posts. Because randomization occurred at the post level, employers can appear multiple times, which allows us to estimate within-employer specifications with employer fixed effects in some tables.

2.4 Nature of Assistance

Employers in the treatment could receive several different forms of assistance. Figure 1 illustrates the treatment process and the various forms of assistance. Treated posts receive a "Hiring Consultant" who engages directly with the employer, and a "Recruiter" who can take two actions: recruiting additional applicants by inviting workers to apply, or shortlisting top applications from the existing pool for the employer's consideration. The Hiring Consultant fields questions from the employer about things like navigating the interface and applicant quality. They also send messages reminding the employer to begin inviting and later, interviewing workers for the job. Through this messaging system the employers can also share specifics of what they seek in a hire beyond what they wrote in a job post. The Recruiter takes these into account and uses the platform search feature to invite applications from workers who have not applied for the job but whom they judge to fit well.

[3] The 75/25 split reflects the platform's prior belief that assistance was beneficial, a common feature of program evaluations where the status quo involves providing the service. The asymmetric allocation is standard in settings where denying treatment to a larger share raises ethical or business concerns.

Figure 1: Treatment process and assistance flow



Notes: This figure illustrates the flow of job posts through the hiring assistance treatment. All treated job posts received initial contact from a Hiring Consultant. Of these, a Recruiter evaluated 66% and decided what form of assistance to provide. Jobs with few quality applications typically received recruiting help (worker invitations), while jobs with sufficient quality applications received shortlisting help. Some jobs received both forms of assistance. 34% of treated posts were contacted by a Hiring Consultant but never delegated to a recruiter.

Both Hiring Consultants and Recruiters specialize in one of three categories of job posts: Web, Mobile, and Software Development; Design and Creative; or General. Hiring someone for graphic design work might take a different skill set than hiring someone to build a database, and so both the Hiring Consultant and the Recruiters assist only with job posts in one category.

When the platform designated a job post as treated, the firm received an invitation for assistance in their hiring for that job. If the firm accepts or does nothing, the platform assigns them a Hiring Consultant who sends the employer a greeting and an offer through the platform-provided messaging system and over email (see Appendix A.8 for the text of the initial message). The Hiring Consultant may edit the job post to make it clearer or more attractive to workers. They also respond to any emails or messages from the firm regarding the hiring process or navigating the interface. If the employer does not respond and has not scheduled any interviews three days after publishing the job post, the Hiring Consultant sends them a follow up email encouraging them to start interviewing applicants.

The Hiring Consultant sends the job post to a Recruiter who reads the job post and the message history between the Hiring Consultant and the employer. The Recruiter takes one of two actions depending on the quality and fit of those who applied organically and whether or not the employer actively engages in searching. If a sufficient number of high quality candidates exist, the Recruiter shortlists (at the median) three of the best fitting workers for hire. These "shortlisted" workers' applications then appear prominently in the job post's application manager page. If the Recruiter does not believe good fitting organic applications exist, they invite workers to apply for the job using the platform search feature. If the job then receives new applications, the Recruiter may follow up by shortlisting.

2.5 Both Forms of Assistance Were Prevalent

About a third of treated employers engaged with the Hiring Consultant by responding with at least one message. About a fifth of job posts received invitations to workers from the Recruiter, and a quarter of job posts received shortlisted candidates. Note that 48% of treated job posts never received either invited nor shortlisted workers. This reflects both the jobs that the Recruiter could not work in time and the job posts that the Recruiters judged would likely receive sufficient applications without help. Notably, jobs that received help from Recruiters looked systematically "worse" on observable characteristics than those that did not. They had fewer applications, fewer recommended applications, and lower projected job value (expected hours times wages) at the time of evaluation. This pattern is consistent with agents targeting assistance toward jobs that appeared to need it most (see Appendix Section A.9 for details).

Jobs with fewer applications more often received recruiting services; jobs with more applications more often received shortlisting (see Figure A.2 in Appendix A.10). We return to this selection mechanism in Section A.11, where we use it to examine whether one form of help outperformed the other.

2.6 Internal Validity

On average 307 job openings went to the treatment and 103 to the control on a given day. To assess the effectiveness of randomization, in Appendix A.12 we report the mean values and t-tests for various pre-randomization attributes of the job post, and we obtain decent balance on these covariates. The treatment group contains slightly fewer low skill jobs than the control group. While the difference in low skill jobs is statistically significant in traditional t-tests, this difference does not survive multiple hypothesis adjustments.

2.7 Estimation Framework

For each outcome we estimate

$$y_i = \beta_0 + \beta_1 \textsc{Assigned}_i + \epsilon_i,$$

where y_i is the outcome of interest for job i, $\textsc{Assigned}_i$ is an indicator for whether or not the job went to the treated group. However, this likely understates the benefits. One aspect of the assistance program is that sometimes the number of job posts assigned to treatment exceeded what staff could handle. A job post appears in a queue to Hiring Consultants and Recruiters based on the time of posting, and they must work jobs in queue order. The jobs that the Recruiters could not get to in time count as not worked, as staff neither evaluated nor helped them. As we are predominantly interested in the effect of *receiving* the treatment, in our preferred specification we use treatment assignment as an instrument for evaluation (and hence potential help). Conditional on assignment to treatment, which job posts staff evaluated depended primarily on how many job posts entered the program that week.

Staff recorded whether they evaluated a job, so we can construct a variable RECEIVED. We then estimate the following IV regression,

$$y_i = \beta_0 + \beta_1 \widehat{\text{RECEIVED}}_i + X_i'\gamma + \epsilon_i,$$

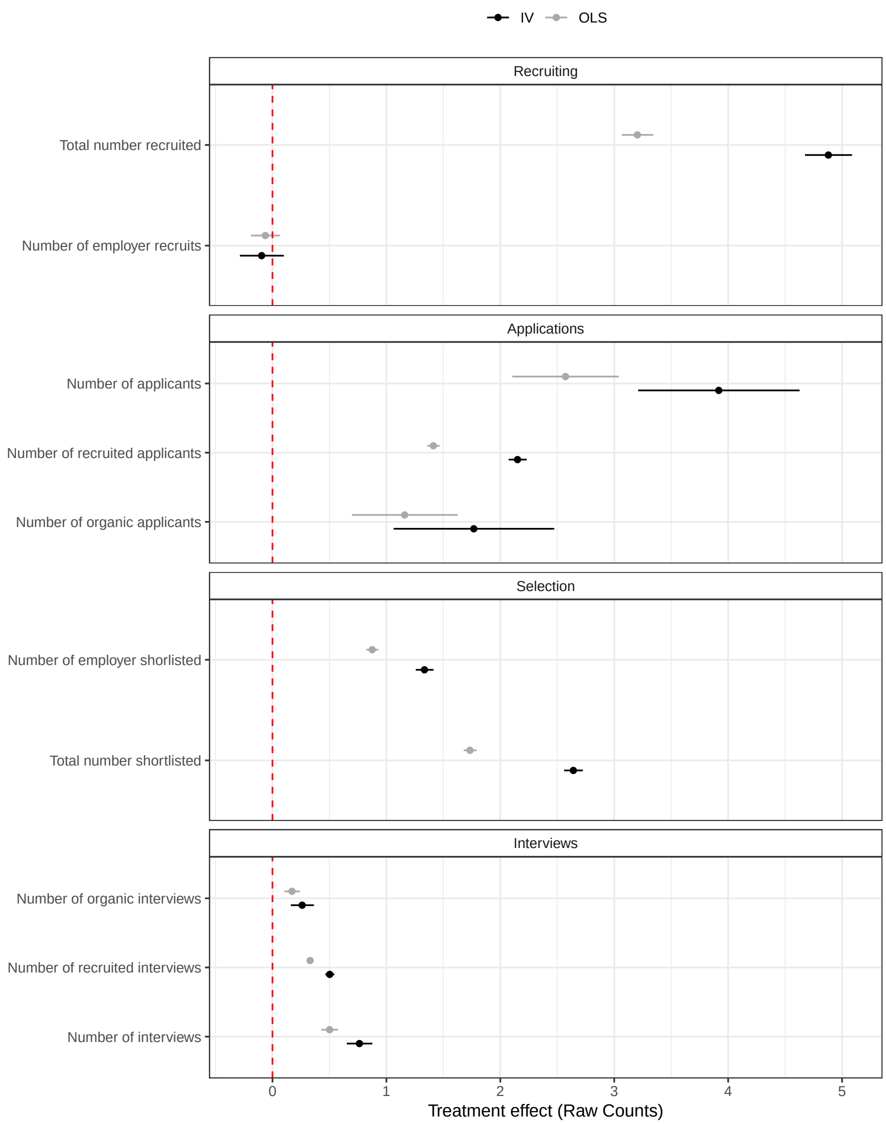
where $\widehat{\text{RECEIVED}}_i$ is instrumented by ASSIGNED_i, and X_i includes fixed effects for client expertise tier and job category. The instrument is valid because evaluation was determined by queue congestion and posting order, not by employer characteristics or job quality.

3 Recruiters Were Active

Before turning to our main result on hiring, we first establish that the recruiters were in fact actively recruiting and suggesting candidates to treated employers (Figure 2). This matters because it rules out the possibility that the null effect on hiring simply reflects a failure to deliver the treatment. The treatment substantially increased invitations: control employers sent 4 invitations, and the treatment added 3.2, though many treated posts received shortlisting instead of recruiting, so the overall difference was smaller than the 9 additional invitations sent on behalf of posts that received recruiting help. Treated posts also attracted more applicants, increasing applications by 2.6 over a control mean of 25 (about 10%), with invited applications rising by 1.4 and organic applications by 1.2, the latter possibly reflecting that consultant reminders increased employers' use of the platform's tools.[4] Shortlisting doubled: treated posts had 2.64 additional shortlisted applications relative to control, with employer shortlists increasing by 1.33 and assistants accounting for the rest. Finally, the treatment increased interviews by 0.5 over a control mean of 2.2, mostly through a 0.33 increase in interviews with invited applicants, though this increase was proportionally smaller than the rise in invited applications, indicating that employers interviewed recruited workers at a lower rate than organic applicants.

[4]The recruited and organic components are estimated from separate OLS specifications and need not sum exactly to the total.

Figure 2: Treatment effects on intermediate outcomes: Recruiters were active



Notes: This figure reports treatment effects on intermediate hiring outcomes, demonstrating that Hiring Consultants and Recruiters actively engaged in recruiting and suggesting candidates. The sample is the experimental sample of high value job posts from October 1, 2021 to April 1, 2022. OLS specification regresses outcome on treatment assignment. IV specifications use treatment assignment as an instrument for whether or not a job receives the treatment. 95% confidence intervals use heteroskedasticity-robust (Eicker-Huber-White) standard errors. Results in table form appear in Appendix Tables A.4–A.7.

4 Results

4.1 Humans Do Not Beat the Algorithm

Treated employers hired at the same rate as control employers despite this substantial activity by hiring consultants. Control job posts had access only to the platform's algorithmic tools. Job posts in the control group had a base 29% fraction that make a hire. Table 1 shows the treatment effects on hiring outcomes. The treatment effect is 0.01, or roughly a 2% increase in the likelihood of making a hire. This result is not statistically significant. The stability of this null across specifications is consistent with the broader finding that well-powered experiments often produce precisely estimated nulls that replicate (DellaVigna and Pope, 2022).

Table 1: Effects of Treatment on Hiring

	(1)	(2)	(3)	(4)
Panel A: OLS (Intent-to-Treat)				
Treatment Effect	0.006	0.006	0.006	0.006
	(0.004)	(0.007)	(0.004)	(0.007)
Panel B: IV (Treatment-on-the-Treated)				
Instrumented Treatment	0.009	0.010	0.009	0.009
	(0.006)	(0.012)	(0.006)	(0.012)
Employer FE	No	Yes	No	Yes
Expertise Tier FE	No	No	Yes	Yes
Observations	83017	83017	83017	83017

Note:
This table reports effects of treatment on hiring outcomes (whether anyone got hired). Sample is experimental sample of high value posts from October 1, 2021 to April 1, 2022. OLS specifications regress outcome on treatment assignment. IV specifications use treatment assignment as an instrument for employers receiving the treatment. Specifications include fixed effects for employer and expertise tier as indicated. All specifications use heteroskedasticity-robust (Eicker–Huber–White) standard errors. Significance indicators: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Assistance may have mainly changed which workers employers considered and hired, not whether they hired at all. Treated employers received more applications from recruited workers and more often hired from this expanded pool or from the newly highlighted candidates. The recruited workers may not have fit better. The net effect appears to have been substitution: treated employers hired recruited workers instead of organic applicants, at roughly the same overall rate. Figure **??** in Appendix A.13 shows treatment effects on variant hiring outcomes (hired recruited applicant, hired organic applicant, hired from shortlist). These are consistent with this substitution pattern.

4.2 Recruited Workers Are Worse Matches

We examine the effect of treatment on wagebill and total hours worked as proxies for match quality. Neither is a direct measure. Wagebill is a quantity-price composite, and hours on hourly contracts can fall for reasons other than mismatch: a skilled worker might finish a task faster, or an employer satisfied early might wind down the engagement. We read the pattern as match quality because the signs go the right way only on contracts where the employer can terminate freely, and the price margin barely moves. The next two paragraphs develop this argument.

Table 2 shows the effects. The treatment reduced 30-day wagebill by -63.07 (OLS, SE $= 31.52$) or -97.52 (IV, SE $= 48.73$), relative to a control mean of 1074.08. Total hours worked fall significantly under the OLS specification (coefficient $= -0.18$ on the Poisson log scale). The hourly rate effect is small (-0.54 OLS, -0.84 IV) relative to the wagebill and hours effects.

The wagebill decline could reflect several mechanisms besides match quality: treated employers might hire cheaper workers, bargain harder, or complete projects faster. Two pieces of evidence point toward mismatch rather than pure price effects. First, the hourly rate effect is small, so the wagebill decline is driven primarily by fewer hours rather than lower prices. Since hourly wages on this platform are largely set by workers' bids (Barach and Horton, 2021), and the employer size-wage premium is small in this setting (Brown and Medoff, 1989), a pure wage channel is unlikely to account for the effect.

Second, the pattern differs across contract types in a way that points to mismatch. The cleanest evidence comes from hours: in the hourly subsample, hours fall (coefficient $= -0.11$), while in the fixed-price subsample the OLS hours coefficient is positive (0.18) and the IV estimate is near zero. Hourly contracts let the employer terminate poor matches, so realized hours track match quality directly; fixed-price contracts pay a pre-agreed amount regardless. The wagebill results follow the same pattern: among hourly jobs, the treatment reduces wagebill by -76.53 (OLS) or -118.62 (Table 3), while among fixed-price jobs, the wagebill effect is small and insignificant (-22.87 OLS, -35.11 IV; Table 4). An F-test for equality of the wagebill treatment effect across contract types does not reject at conventional levels, so the contrast is suggestive rather than definitive. The hours contrast, where signs go in opposite directions, is more informative than the wagebill levels and points toward a mismatch channel that operates through the duration and intensity of the employment relationship rather than through prices. The distinction matters because wagebill confounds price and quantity; separating these channels is important for interpreting treatment effects on labor market outcomes more generally (Gruber, 1994; Holzer et al., 1991).

Controlling for advertised project scope does not change these results. Each posting includes a projected value (expected hours times wages) set at posting time. Adding $\log(\text{projected value})$ as a control leaves the wagebill and hours estimates essentially unchanged (Table 2 and the hourly and fixed-price analogs). Differences in realized spending are not driven by treated employers posting systematically smaller jobs.

The wagebill estimates above condition on hiring, so a concern is that treatment shifts who gets hired and the wagebill effect reflects this composition change rather than a true match-quality decline. We address this with a selection correction. The selection equation predicts whether a post results in exactly one hired application using treatment, category group, expertise tier, and the log number of applications received; the outcome equation models 30-day wagebill conditional on hiring with the same controls except number of applications, which serves as the exclusion restriction. We estimate the model both by Heckman two-step and by full-information maximum likelihood. The treatment effect on wagebill remains negative under both corrections (-30.63 Heckit, -47.31 MLE), and the estimated correlation between selection and outcome errors from the MLE is $--$, suggesting selection on unobservables does not drive the wagebill result (Table A.9 in Appendix A.14).

Table 2: Effects of Treatment: OLS and IV stacked (All Contracts)

	Wagebill 30d		Total Hours		Hourly Rate	
Treatment Assigned (OLS)	-63.07*	-59.97+	-0.18***	-0.17***	-0.54	-0.58
	(31.52)	(31.44)	(0.00)	(0.00)	(0.40)	(0.39)
Constant (OLS)	1074.08	790.58	12.83	10.66	29.98	36.50
Treatment Received (IV)	-97.52*	-92.71+	-0.27*	-0.25*	-0.84	-0.90
	(48.73)	(48.59)	(0.11)	(0.10)	(0.62)	(0.61)
Constant (IV)	1074.42	790.13	12.83	10.67	29.98	36.50
Projected value control	No	Yes	No	Yes	No	Yes
Observations	21,896	21,896	83,017	83,017	19,128	19,128

Notes: This table reports effects of treatment on match quality, as measured by wagebill and hours. Wagebill and hourly rate columns condition on posts where a hire was made. The total hours column (Poisson) uses all posts in the experimental sample within the relevant sample. Subgroup classification (Hourly Jobs / Project Based Jobs / mixed) is based on whether all, none, or a mix of applicants to the post have a reported hourly rate, independent of hiring. Columns (1)–(2) show 30-day wagebill in dollars. Columns (3)–(4) show total hours worked (Poisson, log scale). Columns (5)–(6) show hourly rate. Even-numbered columns add the advertised projected value (expected hours times wages) as a control. OLS specifications regress outcome on treatment assignment. IV specifications use treatment assignment as an instrument for employers receiving the treatment. All specifications use heteroskedasticity-robust (Eicker–Huber–White) standard errors. Significance indicators: $+\ p \leqslant 0.10$, $*\ p \leqslant 0.05$, $**\ p \leqslant 0.01$, $***\ p \leqslant 0.001$.

Assistants provided two conceptually distinct forms of help: recruiting (expanding the applicant pool) and shortlisting (helping employers choose from the existing pool). Recruiting addresses the possibility that employers' applicant pools are too thin, while shortlisting addresses the possibility that employers struggle to identify the best candidates from those who already applied. The type of help each job received was determined endogenously by the recruiter's assessment of the application pool, so we use propensity score methods to model this selection process and estimate the effect of each form separately and still find no effects A.11.

Table 3: Effects of Treatment: OLS and IV stacked (Hourly Jobs)

	Wagebill 30d		Total Hours		Hourly Rate	
Treatment Assigned (OLS)	-76.53*	-77.26*	-0.11***	-0.11***	-0.41	-0.40
	(36.51)	(36.28)	(0.00)	(0.00)	(0.40)	(0.39)
Constant (OLS)	1077.76	476.50	13.08	10.96	28.63	33.61
Treatment Received (IV)	-118.62*	-119.75*	-0.17	-0.17	-0.63	-0.62
	(56.56)	(56.22)	(0.12)	(0.11)	(0.61)	(0.61)
Constant (IV)	1078.27	474.89	13.08	10.96	28.63	33.61
Projected value control	No	Yes	No	Yes	No	Yes
Observations	16,334	16,334	54,367	54,367	15,856	15,856

Notes: This table reports effects of treatment on match quality, as measured by wagebill and hours. Wagebill and hourly rate columns condition on posts where a hire was made. The total hours column (Poisson) uses all posts in the experimental sample within the relevant sample. Subgroup classification (Hourly Jobs / Project Based Jobs / mixed) is based on whether all, none, or a mix of applicants to the post have a reported hourly rate, independent of hiring. Columns (1)–(2) show 30-day wagebill in dollars. Columns (3)–(4) show total hours worked (Poisson, log scale). Columns (5)–(6) show hourly rate. Even-numbered columns add the advertised projected value (expected hours times wages) as a control. OLS specifications regress outcome on treatment assignment. IV specifications use treatment assignment as an instrument for employers receiving the treatment. All specifications use heteroskedasticity-robust (Eicker–Huber–White) standard errors. Significance indicators: $+\ p \leqslant 0.10$, $*\ p \leqslant 0.05$, $**\ p \leqslant 0.01$, $***\ p \leqslant 0.001$.

4.3 Even Under Favorable Assumptions, the Program Cannot Justify Its Costs

Failing to reject the null does not mean there is no effect, but even under the most favorable assumptions, the program cannot justify its costs. We use 90-day rather than 30-day wagebill in the revenue calculation, a deliberately favorable assumption: a longer accumulation window inflates revenue per hire and thus the implied benefit of the program. The conclusion below holds despite this. We estimate program costs using Bureau of Labor Statistics wage data. In 2022, the median hourly wage for human resources workers was \$31.25. Assistants' activities include initial contact, follow-up messages, responding to questions, and occasional phone calls. They also review job posts, evaluate applications, invite workers (about 9 invitations on average), and shortlist applicants (about 3 on average). We consider three cost scenarios based on time spent per job post.

The estimated effect on hiring is 0.01 with a standard error of about 0.01. The 95% confidence interval upper bound is 0.02. We compare program costs per job post to expected revenue per job post. Since only 29% of job posts result in a hire, the 90-day wagebill per hiring job post is about \$3,011. Online labor platforms typically charge commission fees of around 10% on wagebill (Wood and Lehdonvirta, 2021), implying revenue per hire of about \$301. The expected revenue per job post is then the treatment effect on hiring multiplied by revenue per hire.

Even under the most optimistic scenario (the upper bound of the confidence interval), the expected revenue per job post (\$5.86) falls short of the cost per job post under every time assumption, with estimated

Table 4: Effects of Treatment: OLS and IV stacked (Project Based Jobs)

	Wagebill 30d		Total Hours		Hourly Rate	
Treatment Assigned (OLS)	-22.87	-24.54	0.18***	0.25***	-1.65	-1.64
	(62.56)	(62.29)	(0.00)	(0.00)	(1.31)	(1.30)
Constant (OLS)	1062.51	-62.22	10.65	8.79	36.88	28.00
Treatment Received (IV)	-35.11	-37.67	-0.31	-0.06	-2.57	-2.54
	(96.03)	(95.62)	(0.53)	(0.51)	(2.03)	(2.02)
Constant (IV)	1062.54	-63.90	10.65	8.84	36.88	27.92
Projected value control	No	Yes	No	Yes	No	Yes
Observations	5,560	5,560	26,055	26,055	3,271	3,271

Notes: This table reports effects of treatment on match quality, as measured by wagebill and hours. Wagebill and hourly rate columns condition on posts where a hire was made. The total hours column (Poisson) uses all posts in the experimental sample within the relevant sample. Subgroup classification (Hourly Jobs / Project Based Jobs / mixed) is based on whether all, none, or a mix of applicants to the post have a reported hourly rate, independent of hiring. Columns (1)–(2) show 30-day wagebill in dollars. Columns (3)–(4) show total hours worked (Poisson, log scale). Columns (5)–(6) show hourly rate. Even-numbered columns add the advertised projected value (expected hours times wages) as a control. OLS specifications regress outcome on treatment assignment. IV specifications use treatment assignment as an instrument for employers receiving the treatment. All specifications use heteroskedasticity-robust (Eicker–Huber–White) standard errors. Significance indicators: $+\ p \leqslant 0.10$, $*\ p \leqslant 0.05$, $**\ p \leqslant 0.01$, $***\ p \leqslant 0.001$.

returns of between $-1.95 and $-7.16 per post. The program breaks even only if assistants spend less than 11.3 minutes per post under the upper CI, or less than 5.0 minutes per post under the point estimate. Both fall well below plausible time per post given the activities involved: initial contact, follow-ups, evaluating posts, recruiting roughly 9 workers on average, and shortlisting roughly 3. Table 5 presents the details. These calculations suggest that effect sizes large enough to justify the program appear unlikely, though we cannot rule them out definitively.

Table 5: Estimated program costs and revenue per job post

Hiring Consultant Time	Cost per Post	Revenue per Post (Upper CI)	Net Return per Post
15 min	$7.81	$5.86	$-1.95
20 min	$10.42	$5.86	$-4.55
25 min	$13.02	$5.86	$-7.16

Notes: Costs use median HR worker wage of $31.25/hour from the Bureau of Labor Statistics. Revenue per post equals the upper bound treatment effect on hiring (0.02) multiplied by revenue per hire ($301), yielding $5.86. Revenue per hire equals average 90-day wagebill per hiring job post ($3,011) times a 10% platform commission rate (Wood and Lehdonvirta, 2021). Under the point estimate (0.01), expected revenue per post would be $2.58.

This cost structure also highlights a scaling asymmetry. The program already could not serve 48% of treated job posts because staff were overwhelmed by volume. Scaling up would require hiring proportionally more recruiters, each costing roughly the same as the last. Algorithmic tools face no such constraint: once built, they serve additional job posts at near-zero marginal cost.

5 Model

We develop a simple model to rationalize how hiring assistance from recruiters could expand applicant pools without improving hiring outcomes. We model hiring as a screening process in which employers sometimes receive help from recruiters, and both parties evaluate applicants using information available from the platform.

Canonical delegated-search models do not predict the pattern produced by our experiment(Mortensen and Pissarides, 1994; Pissarides, 2000; Wright et al., 2017). Employers given recruiters saw expanded applicant pools and ran more interviews. Hiring rates, however, were flat and match quality fell. The standard model of delegated recruiting suggests an intermediary who diversifies sampling risk or expands option value would increase hires and quality(Bull et al., 1987). Our key departures from this model are that we introduce an algorithmic screening stage and allow the noise in recruiter and employer assessments of the applicants to be correlated.

5.1 Environment

The setting is an online labor market (hereafter, "the platform.") An employer posts a job on the platform and receive applications from workers. The employer's goal is to hire the most productive applicant, but in practice due to costs of conducting interviews they hire the first applicant that they interview (ordered at random) above an endogenous cutoff.

Workers. Worker j has quality v_j, which consists of two components $v_j = \alpha_j + \varepsilon_j$. The first component, α_j, is observable to the platform's algorithm and summarizes the worker quality captured by platform data. The second component, ε_j, is orthogonal to α_j and captures traits that are not verifiable from platform data, such as communication skills, domain expertise, or fit with the employer's needs. We assume $\alpha_j \sim N(0, \sigma_\alpha^2)$ and $\varepsilon_j \sim N(0, \sigma_\varepsilon^2)$ independent. We let $\sigma_v^2 \equiv \sigma_\alpha^2 + \sigma_\varepsilon^2$.

The Algorithm. The platform has an algorithm that determines the pool of applicants that the employer sees and gets to choose among. The algorithm observes the component α_j with noise, meaning their assessment of a given worker's productivity is a score $s_j \equiv \alpha_j + \psi_j$, with $\psi_j \sim N(0, \sigma_\psi^2)$. The algorithm then selects all applicants with score s_j above a threshold $\bar{s}$ to present to the employer, including the score in the applicant's profile.

The Employer. The employer sees the pool of applicants pre-screened by the algorithm and observes their s_j. The employer then conducts interviews at random, where she learns more about the applicants. For each applicant the employer interviews, she observes the algorithm's score, s_j, alongside a noisy signal of ϵ_j, meaning their expected productivity is $q_j^E = s_j + \epsilon_j + \xi_j$, with $\xi_j \sim N(0, \sigma_\xi^2)$, or $q_j^E = \alpha_j + \psi_j + \epsilon_j + \xi_j$. At each interview she observes this noisy measure of the applicants quality q_j^E, and hires the first applicant with $q_j^E \geqslant \bar{u}$.

The Recruiter. Some employers also have access to a hiring consultant ("recruiter"). The recruiter's role is to add candidates to the pool that the employer observes that they view as promising but that the algorithm may have missed. Formally, the recruiter sees all applicants–not just those selected by the algorithm–and observes the algorithm's score and a noisy signal of ϵ_j, meaning their expected productivity is $q_j^R = s_j + \epsilon_j + \xi_j^R$, with $\xi_j^R \sim N(0, \sigma_{\xi^R}^2)$, or $q_j^R = \alpha_j + \psi_j + \epsilon_j + \xi_j^R$. The recruiter recruits m applicants with the highest q_j^R that were not already selected by the algorithm.

The model allows the recruiter and employer to be impressed by the same misleading information. The recruiter and employer may make correlated assessment errors, $\mathrm{Cov}(\xi^R, \xi^E) = \sigma_{\xi_{RE}}$ and each party's idiosyncratic assessment error may also covary with the platform's scoring error $\mathrm{Cov}(\psi, \xi^R) = \sigma_{\psi\xi_R}$ and $\mathrm{Cov}(\psi, \xi^E) = \sigma_{\psi\xi_E}$. These covariances capture whether recruiters and employers reinforce or correct the same misleading profile features that enter the algorithmic score. A positive covariance with ψ means that the party tends to make the same mistakes as the algorithm. A negative covariance means the party tends to correct the algorithmic mistake. The covariance $\sigma_{\xi_{RE}}$ captures whether the recruiter and employer make similar residual mistakes, even after conditioning on true quality and the shared profile noise.

5.2 Hiring probability

In this section we see under what conditions having access to a recruiter increases the probability that the employer makes a hire. The employer makes a hire so long as at least one selected applicant has q_j^E above their cutoff $\bar{u}$. For simplicity, we model the cutoff-determination problem in a reduced form way and let $\bar{u}$ be a strictly increasing function of the number of applicants.

Each candidate in $\mathcal{A}$ passes the employer's screen independently with probability

$$p_A = P(q^E \geqslant \bar{u} \mid s \geqslant \bar{s}), \qquad p_R = P(q^E \geqslant \bar{u} \mid s < \bar{s},\, q^R \geqslant \bar{s}). \tag{1}$$

The likelihood of making a hire is

$$P_C = 1 - (1 - p_A)^n, \qquad P_T = 1 - (1 - p_A)^n (1 - p_R)^m. \tag{2}$$

Result 1 (Ambiguous hiring effect of recruiter added candidates)**.** *If the employer keeps their hiring threshold fixed, the recruiter added candidates increases the probability the employer makes a hire.*

$$\Delta_P \equiv P_T - P_C = (1 - p_A)^n \big[1 - (1 - p_R)^m\big] \geqslant 0. \tag{3}$$

However, if the employer raises $\bar{u}$ in response to the larger applicant pool, the recruiter can have no effect on employer's hiring probability.

The idea is that a larger applicant pool raises the employer's reservation value and she is willing to hold out longer in hopes for a better fit (Appendix A.4).

5.3 Match quality

In this section we introduce the key object ω, the conditional quality gap between recruited and algorithmic hires.

Let $Q_C = E[v \mid q^E \geqslant \bar{u},\, j \in \mathcal{A}]$ and let π_R denote the probability a treatment-group hire comes from R.

Result 2 (Match quality decomposition)**.** *The treatment effect on expected match quality is*

$$\Delta_Q \equiv Q_T - Q_C = \pi_R\, \omega, \tag{4}$$

where

$$\omega \;\equiv\; E[v \mid q^E \geqslant \bar{u},\, j \in R] \;-\; E[v \mid q^E \geqslant \bar{u},\, j \in \mathcal{A}]. \tag{5}$$

The entire match-quality effect is carried by ω: the gap in expected quality between a recruited hire and an algorithmic hire, both conditional on passing the interview. The closed form (Appendix A.2) uses the regression identity for truncated normals and the Tallis (1961) formula.

5.4 Comparative statics

Recruited workers start with an α deficit: $s < \bar{s}$ means the algorithm ranked them below the cutoff, which negatively selects on α. Working against this, the recruiter's signal loads on ε, which s does not, so recruited workers carry an ε premium. The sign of ω turns on whether the premium outweighs the deficit, which depends on parameters.

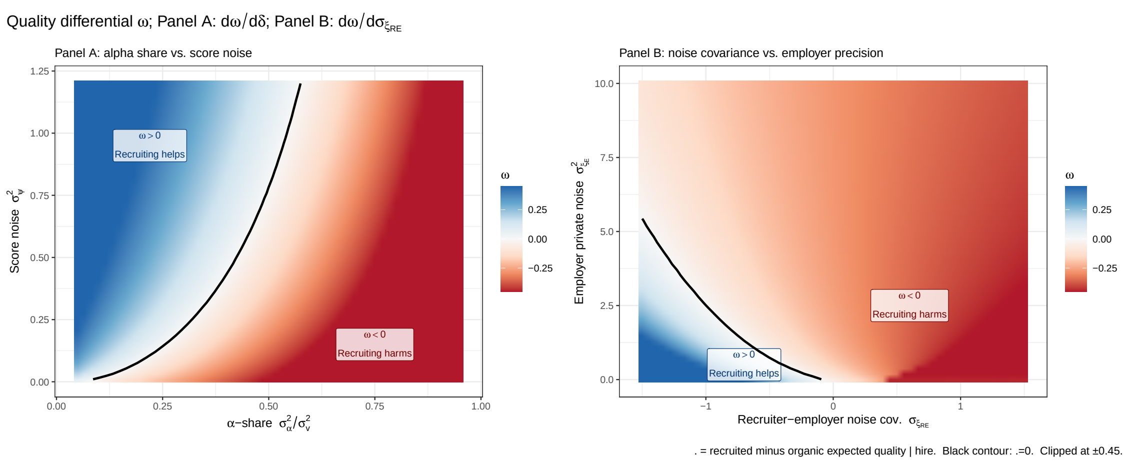
Result 3 (Comparative statics on ω)**.** *Across the parameter space in Figure 3, ω is decreasing in (i) the recruiter–employer noise covariance $\sigma_{\xi_{RE}}$ and (ii) the algorithmic share $\sigma_\alpha^2/\sigma_v^2$ at fixed σ_v^2. ω can be negative, and is negative over broad regions of the parameter space, including the full region with $\sigma_{\xi_{RE}} \geqslant 0$ in Panel B.*

Numerical evaluation of the closed-form gradient confirms both comparative statics throughout the parameter space in Figure 3; Appendix A.5 gives the gradient and documents the evaluation. We do not establish the monotonicity analytically.

The covariance $\sigma_{\xi_{RE}}$ is the central parameter. ξ^R is the recruiter's bias. When her mistakes are unrelated to the employer's, ξ^R is harmless noise the employer can see through at the interview. When the two parties make the same mistakes, $\sigma_{\xi_{RE}} > 0$ and the recruiter stops adding independent information. She adds fool's gold: candidates who look good to her for the same reasons they look good to the employer, even when they are not actually better. Both parties get taken in by the same noise. Once hired, the noise washes out and only the α deficit remains. Panel B of Figure 3 shows $\omega < 0$ throughout the region with non-negative covariance, and Figure 4 shows Δ_P is positive throughout Panel B, confirming that recruiter screening continues to raise hiring probability even when $\omega < 0$.

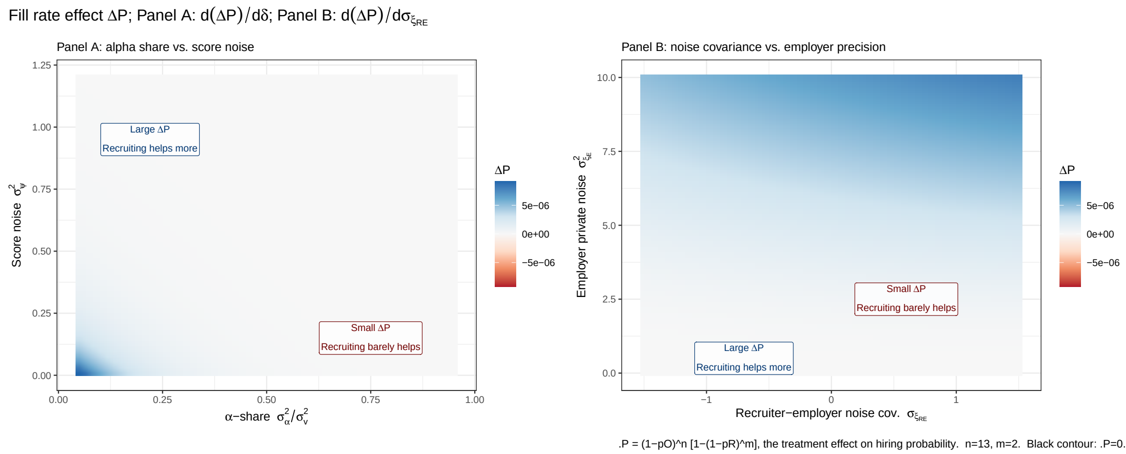
The algorithmic share $\sigma^2_\alpha/\sigma^2_v$ speaks to algorithmic improvement. A better algorithm leaves less value for the recruiter to add. As $\sigma^2_\alpha/\sigma^2_v$ rises, the score captures more of quality. The workers sitting below the cutoff are increasingly workers the algorithm correctly passed over. There is less hidden quality for the recruiter to dig up, and the region where $\omega > 0$ shrinks (Panel A).

Figure 3: Quality differential ω across the parameter space



Notes: Color shows ω, the difference between the expected quality of a recruited hire and that of an algorithmic hire, conditional on passing the employer's screen. Black contour: $\omega = 0$; red (blue) region: recruiting harms (helps) quality. Parameters held fixed across both panels: $\sigma^2_v = 1$, $\bar{s} = 0.5$, $\bar{u} = 0.3$, $\sigma_{\psi\xi_R} = \sigma_{\psi\xi_E} = 0$. Panel A additionally fixes $\sigma^2_{\xi_R} = \sigma^2_{\xi_E} = 1$ and $\sigma_{\xi_{RE}} = 0$; Panel B additionally fixes $\sigma^2_\alpha/\sigma^2_v = 0.5$, $\sigma^2_\psi = 0.3$, and $\sigma^2_{\xi_R} = 1$. Color clipped at ± 0.45.

Figure 4: Fill-rate treatment effect ΔP across the parameter space



Notes: Color shows $\Delta P = (1 - p_A)^n[1 - (1 - p_R)^m]$, the fixed-threshold treatment effect on the probability of making a hire. Blue regions indicate parameter combinations where the mechanical hiring gain from recruiting is larger; lighter regions indicate that the gain is positive but small. Black contour: $\Delta P = 0$ (absent in both panels as $\Delta P > 0$ throughout by Result 1). Parameters held fixed across both panels: $\sigma_v^2 = 1$, $\bar{s} = 0.5$, $\bar{u} = 0.3$, $\sigma_{\psi\xi_R} = \sigma_{\psi\xi_E} = 0$. Panel A additionally fixes $\sigma^2_{\xi_R} = \sigma^2_{\xi_E} = 1$ and $\sigma_{\xi_{RE}} = 0$; Panel B additionally fixes $\sigma_\alpha^2/\sigma_v^2 = 0.5$, $\sigma_\psi^2 = 0.3$, and $\sigma^2_{\xi_R} = 1$.

5.5 Mapping to the data

The three results generate four predictions we revisit in Section 6. Result 1 with threshold response rationalizes the near-zero hiring effect. Result 2 combined with $\omega < 0$ from Result 3 rationalizes the decline in wagebill and hours on hourly contracts, where match quality is observed. Result 3 also predicts that recruiter value is largest where ε's share is largest: in jobs where soft skills matter and platform histories are thin, consistent with the subgroup effects in customer service and low-tier jobs. Finally, the algorithmic-improvement comparative static in Result 3 predicts that the algorithm's marginal value is largest in thin pools, where the employer has few alternatives: this is what the RD heterogeneity shows.

6 Heterogeneity and Evidence Supporting the Model

6.1 Treated Employers Took Longer to Hire

Conditional on making a hire, treated employers also took longer to hire. Table 6 shows the treatment effect on time from posting to first hire. Treated employers searched longer before hiring, consistent with a larger applicant pool raising the employer's reservation threshold (Lemma 2). Longer search combined with worse realized outcomes suggests the additional candidates did not improve the employer's choice set.

Table 6: Effects of Treatment on Time to Hire (Hours)

	OLS		IV	
	(1)	(2)	(3)	(4)
Treatment Effect	10.674*	10.801*	17.107*	17.303*
	(5.287)	(5.285)	(8.461)	(8.455)
Expertise Tier FE	No	Yes	No	Yes
Control Mean (hours)	247.8	247.8	247.8	247.8
Observations	24704	24704	24704	24704

Note:
This table reports effects of treatment on time from posting to first hire, in hours, conditional on making a hire. Sample is experimental sample of high value posts from October 1, 2021 to April 1, 2022. OLS specifications regress outcome on treatment assignment. IV specifications use treatment assignment as an instrument for employers receiving the treatment. All specifications use heteroskedasticity-robust (Eicker–Huber–White) standard errors. Significance indicators: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

6.2 Employers Respond to Algorithmic Recommendations

The model predicts that the platform's algorithmic score carries information employers act on. We test this directly using a fuzzy regression discontinuity (RD) design around the platform's recommendation threshold. Horton (2017) shows that algorithmic recommendations can substantially increase hiring in online labor markets; we use the same platform's recommendation system to study how employers respond to algorithmic signals at the margin. Each job post has a score threshold above which applicants are flagged as "recommended." Employers cannot see the underlying recommendation score, but they observe the discrete recommendation label. We exploit the jump in recommendation status at this threshold.

The running variable is $x_{ij} = s_{j,c} - \bar{s}_{j,\text{post}}$, where $s_{j,c}$ is the within-group standardized score and $\bar{s}_{j,\text{post}}$ is the post-level recommendation threshold. We define this threshold as the midpoint between the lowest recommended score and the highest non-recommended score within each post. Appendix A.15 reports density tests, covariate balance plots, and the first-stage relationship, all of which support the validity of the design.

Recommendation increases hiring. Figure 5 shows the RD relationship for hiring. Algorithmic recommendation increases the probability of being hired. The bias-corrected RD estimate in the pooled sample is 0.012 (SE $= 0.005$, $N = 197,273$, bandwidth $= 0.094$). The effect is positive in the treated arm and imprecise in the control arm; a χ^2 test fails to reject equality of the treatment and control RD estimates. The hiring effect in the treated arm is positive across all three threshold definitions and statistically significant under the two alternatives (Tables A.16 and A.19); the wagebill RD and the control-arm hiring RD are noisy and not directionally stable across definitions, as expected given the smaller samples on those margins.

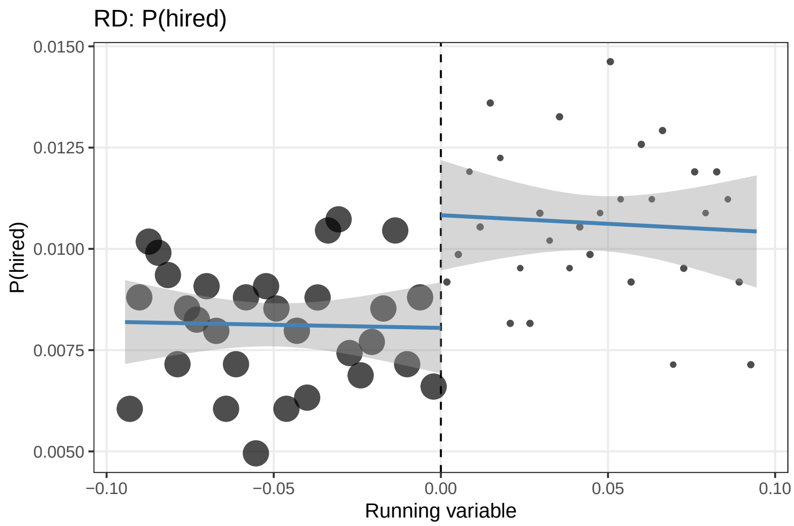


Figure 5: RD: hiring probability at the recommendation threshold

Notes: Binned scatter plot of hiring probability against the recommendation score running variable. Lines are local linear fits estimated separately on each side of the threshold. The vertical dashed line marks the recommendation threshold. Corresponding regression estimates appear in Table A.10. Additional RD results appear in Appendix A.15.

Recruiters also respond to the recommendation label. The effect of recommendation on shortlisting is positive across all three threshold definitions and statistically significant under the two alternatives (Tables A.18 and A.21); under the baseline definition the estimate is positive but imprecise (Table A.12). The significant results are driven by the recruiter: the RD estimate for recruiter shortlisting is positive and significant in the treatment arm under the alternative definitions. This is consistent with both parties conditioning on the same visible algorithmic signal, a necessary condition for the correlated signal mechanism in the model.

The recommendation effect is concentrated in thin applicant pools. We split applicants into quartiles of the number of applications received and of the recommendation threshold. Recommendation has a significant positive effect on hiring in the lowest application-count quartile and no detectable effect in the upper three (Table A.25). Wagebill estimates are directionally consistent: positive in Q1 and negative in Q3–Q4 (Table A.27). Threshold placement shows no differential hiring or wagebill effects across quartiles (Tables A.28 and A.31). Full results appear in Appendix A.16.

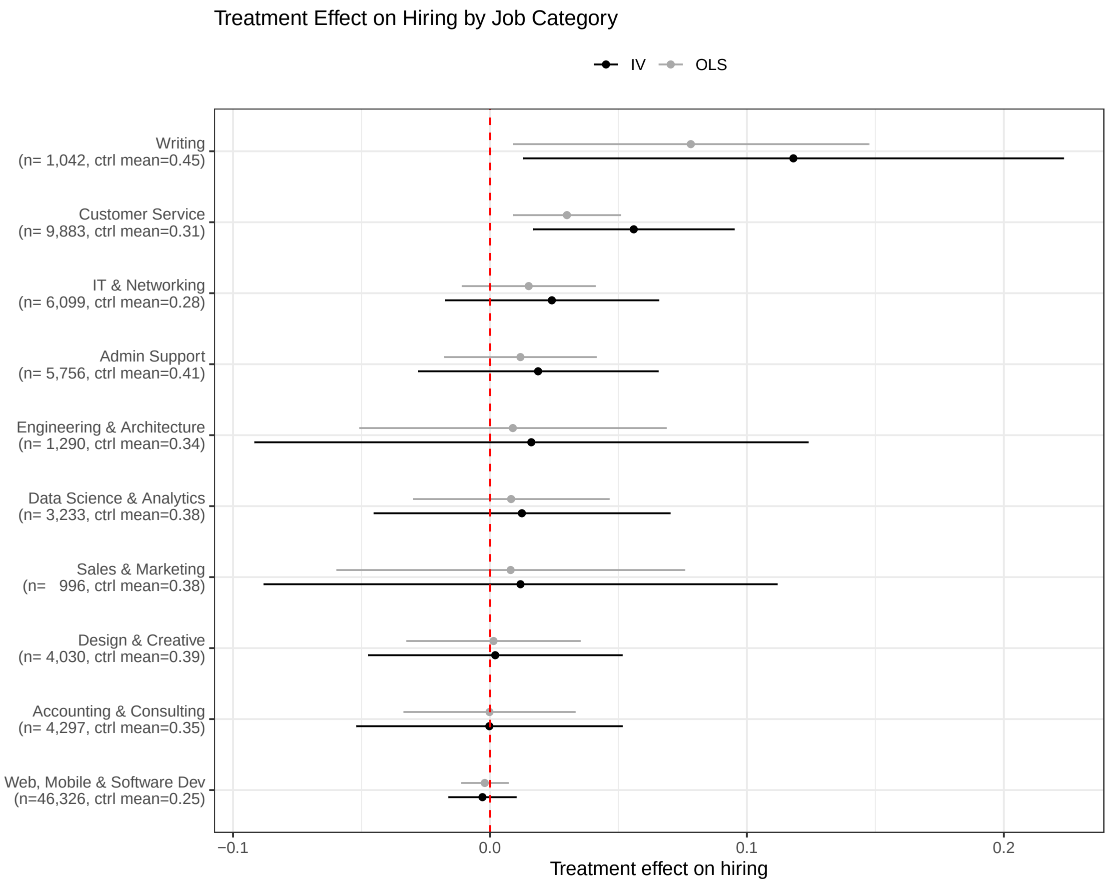
This pattern fits the model. The RD estimates the employer's marginal response to the recommendation label, holding the rest of the pool fixed. When the pool is thin, the employer has few alternatives, so moving one applicant across the threshold changes which candidate she hires. When the pool is deep, she already has many acceptable algorithmic candidates above her reservation threshold, and the marginal recommended applicant is inframarginal to the hiring decision: the employer would have hired someone comparable regardless. In the notation of Section 5, the recommendation label is informative about α, but its value to the employer is decreasing in pool size n because $p_A n$, the expected number of algorithmic candidates who pass her screen, is already large. The concentration of the effect in thin pools is therefore consistent with employers using the recommendation as one signal among many, with diminishing returns as the pool grows.

6.3 Algorithms Struggle Most with Soft Skills

The null hiring effect is consistent across most job categories and expertise tiers. We explore subgroup heterogeneity to characterize where, if anywhere, consultants came closer to helping.

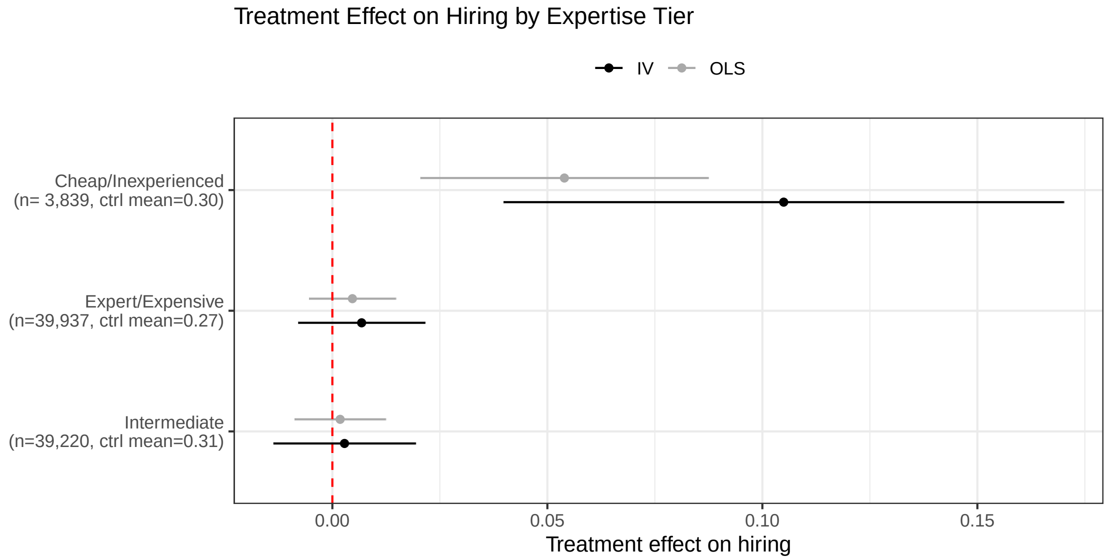
Figure 6 shows treatment effects by job category. Most categories show small, statistically insignificant effects. Writing and customer service have larger point estimates. These are categories where quality may be harder to assess from profiles alone, leaving more room for human judgment. This is consistent with the model's prediction that recruiters add value when the algorithm captures less of worker quality. Tables A.38 and A.39 in the appendix report bootstrap stepdown, Bonferroni, and FDR-adjusted p-values. Customer service remains significant after all corrections; writing does not. Among expertise tiers, the effect for the lowest tier (cheap/inexperienced) survives all corrections. Both subgroups where the treatment effect survives corrections, customer service and low-tier jobs, are settings where the algorithmic score plausibly captures less of worker quality. Customer service depends on soft skills like communication and reliability that are difficult to verify from platform histories, implying a larger share of score-orthogonal quality σ_ε^2. Low-tier workers have thinner profiles, reducing σ_α^2. In both cases, the model predicts a larger domain for recruiter value (Figure 3), and the data are consistent with this prediction.

Figure 6: Treatment effects on hiring by job category



Notes: This figure reports treatment effects on hiring (whether anyone got hired) separately by job category, excluding Legal due to small sample size. OLS specification regresses outcome on treatment assignment. IV specifications use treatment assignment as an instrument for whether or not a job receives the treatment. 95% confidence intervals use heteroskedasticity-robust (Eicker-Huber-White) standard errors. Labels show subgroup sample size and control group hiring rate.

Figure 7: Treatment effects on hiring by expertise tier



Notes: This figure reports treatment effects on hiring (whether anyone got hired) separately by expertise tier. OLS specification regresses outcome on treatment assignment. IV specifications use treatment assignment as an instrument for whether or not a job receives the treatment. 95% confidence intervals use heteroskedasticity-robust (Eicker-Huber-White) standard errors. Labels show subgroup sample size and control group hiring rate.

7 Conclusion

As algorithmic tools increasingly structure search and screening in hiring markets, the economic role of human recruiting intermediaries becomes less obvious. We study a randomized experiment that assigned hiring assistance to employers on a large online labor market. Hiring consultants expanded applicant pools, shortlisted candidates, and increased interviews, however they did not improve employers odds of hiring. Treated employers hired at the same rate as control employers who had only the platform's algorithmic tools. Treated employers had lower wagebills, suggesting worse match quality. The effect was concentrated among hourly jobs, where wagebill captures the duration and intensity of the employment relationship. Neither recruiting nor shortlisting improved outcomes when examined separately. A regression discontinuity design shows that employers respond to the algorithm's recommendations by shortlisting recommended applicants. Algorithmic recommendation raises the probability of being hired but does not improve match quality. The costs of the program are difficult to justify even under optimistic assumptions.

Our model suggests a mechanism. Worker quality has two components: quality captured by the algorithmic score and quality orthogonal to it. In principle, a large share of score-orthogonal quality leaves scope for recruiters to add value. When the algorithm captures only a small share of quality, information recruiters gain from beyond the platform profile can be useful. If recruiter and employer signals are correlated through their common reliance on the noisy score, recruited workers who pass the interview threshold do so partly because of inflated scores rather than true quality. Once hired, the noise washes out and realized surplus is lower.

Several alternative explanations are less consistent with the evidence. If recruiters were simply bad at their jobs, we would expect noise rather than a systematic quality decline, and employers would not have responded to their input. Treated employers, however, interviewed more candidates, searched longer, and shortlisted recruiter-sourced applicants. If employers were better informed than recruiters, they would have ignored the recruiter's suggestions, or at most seen no effect. They would not have shortlisted worse candidates and hired them. If the algorithm were near-perfect, leaving no room for human judgment, adding candidates should be neutral rather than harmful. The quality decline requires an explanation for why recruited hires are worse than algorithmic hires.

The model's comparative statics suggest that recruiters add value when the algorithm captures less of worker quality. In those settings, the recruited pool's selection on score-orthogonal quality ε can compensate for its lower algorithmic quality. The exploratory subgroup evidence is directionally consistent: customer service and low-tier jobs, where algorithmic scores plausibly capture less of quality, show larger treatment effects.

The model and empirical results suggest that there are still situations where human recruiters add value. As algorithms improve and capture more of worker quality from profiles, these situations become rarer. One possibility is that the industry contracts not because technology replaces it across the board, but because the profitable use cases are too few to sustain the current scale. Each additional recruiter costs roughly the same as the last, while algorithms serve additional job posts at near-zero marginal cost.

Several caveats bear on external validity, though the gap between online and conventional hiring is narrower than it may appear. Employers and recruiters evaluating candidates through online platforms is the norm rather than the exception. Still, information in our setting is more standardized than in other forms of hiring. In conventional labor markets, intermediaries such as temporary help firms provide additional services beyond matching, including training (Autor, 2001) and certification (Cullen and Farronato, 2021). The value of human intermediaries may be larger in settings with less structured information or where algorithms have less data to work with. Several settings remain open. Labor markets with stronger informal networks (Brown et al., 2016) may behave differently. So may settings with more scope for in-person assessment (Bloom et al., 2013; Ichniowski and Shaw, 1999), or where hiring decisions interact with other HR practices (Aral et al., 2012).

Our experiment also cannot speak to longer-run effects. Employers who experience poor matches from recruited workers may learn to discount recruiter recommendations over time. The platform may also improve its algorithm in response to the evidence (Leung, 2018). Our treatment combines recruiting and shortlisting assistance. Disentangling these channels more cleanly, perhaps through separate randomization of each service, would sharpen the conclusions. These limitations leave open the possibility that recruiters improve aspects of the employer experience that we do not observe, but the evidence suggests that, in this setting, human intermediation did not improve the central economic outcomes of hiring, match quality, or returns to the platform.

References

M. Abel. Do workers discriminate against female bosses? *Journal of Human Resources*, 59(2):470–501, 2024.

P. Aghion and J. Tirole. Formal and real authority in organizations. *Journal of Political Economy*, 105(1): 1–29, 1997.

A. Agrawal, N. Lacetera, and E. Lyons. Does standardized information in online markets disproportionately benefit job applicants from less developed countries? *Journal of International Economics*, 103:1–12, 2016.

R. Alonso and N. Matouschek. Optimal delegation. *Review of Economic Studies*, 75(1):259–293, 2008.

S. Aral, E. Brynjolfsson, and L. Wu. Three-way complementarities: Performance pay, human resource analytics, and information technology. *Management Science*, 58(5):913–931, 2012.

T. B. Armstrong and M. Kolesár. Simple and honest confidence intervals in nonparametric regression. *Quantitative Economics*, 11(1):1–39, 2020.

D. H. Autor. Why do temporary help firms provide free general skills training? *The Quarterly Journal of Economics*, 116(4):1409–1448, 2001.

M. A. Barach and J. J. Horton. How do employers use compensation history? evidence from a field experiment. *Journal of Labor Economics*, 39(1):193–218, 2021.

J. M. Barron, J. Bishop, and W. C. Dunkelberg. Employer search: The interviewing and hiring of new employees. *The Review of Economics and Statistics*, pages 43–52, 1985.

M. Belot, P. Kircher, and P. Muller. Providing Advice to Jobseekers at Low Cost: An Experimental Study on Online Advice. *The Review of Economic Studies*, 86(4):1411–1447, 10 2018.

Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300, 1995.

A. Benson, A. Sojourner, and A. Umyarov. Can reputation discipline the gig economy? experimental evidence from an online labor market. *Management Science*, 2019.

A. Benson, D. Li, and K. Shue. ”potential” and the gender promotion gap. Technical report, working paper, 2021.

M. Bertrand and E. Duflo. Field experiments on discrimination. In *Handbook of Economic Field Experiments*, volume 1, pages 309–393. Elsevier, 2017.

M. Bertrand and S. Mullainathan. Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. *American Economic Review*, 94(4):991–1013, 2004.

I. Black, S. Hasan, and R. Koning. Hunting for talent: Firm-driven labor market search in the united states. *Strategic Management Journal*, 45(3):429–462, 2024.

T. Blake and D. Coey. Why marketplace experimentation is harder than it seems: The role of test-control interference. In *Proceedings of the fifteenth ACM conference on Economics and computation*, pages 567–582. ACM, 2014.

N. Bloom, B. Eifert, A. Mahajan, D. McKenzie, and J. Roberts. Does management matter? evidence from india. *The Quarterly Journal of Economics*, 128(1):1–51, 2013.

C. Brown and J. Medoff. The employer size-wage effect. *Journal of Political Economy*, 97(5):1027–1059, 1989.

M. Brown, E. Setren, and G. Topa. Do informal referrals lead to better matches? evidence from a firm's employee referral system. *Journal of Labor Economics*, 34(1):161–209, 2016.

C. Bull, O. Ornati, and P. Tedeschi. Search, hiring strategies, and labor market intermediaries. *Journal of Labor Economics*, 5(4, Part 2):S1–S17, 1987.

S. V. Burks, B. Cowgill, M. Hoffman, and M. Housman. The value of hiring through employee referrals. *The Quarterly Journal of Economics*, 130(2):805–839, 2015.

D. Card, J. Kluve, and A. Weber. Active labour market policy evaluations: A meta-analysis. *The Economic Journal*, 120(548):F452–F477, 2010.

D. Card, J. Kluve, and A. Weber. What works? a meta analysis of recent active labor market program evaluations. *Journal of the European Economic Association*, 16(3):894–931, 2018.

A. P. Carnevale, T. Jayasundera, and D. Repnikov. Understanding online job ads data: A technical report. Technical report, Georgetown University Center on Education and the Workforce, April 2014. URL `https://cew.georgetown.edu/wp-content/uploads/2014/11/OCLM.Tech_.Web_.pdf`.

M. D. Cattaneo, M. Jansson, and X. Ma. Simple local polynomial density estimators. *Journal of the American Statistical Association*, 115(531):1449–1455, 2020.

J. Chan and J. Wang. Hiring preferences in online labor markets: Evidence of a female hiring bias. *Management Science*, 64(7):2973–2994, 2018.

D. L. Chen and J. J. Horton. Are online labor markets spot markets for tasks? a field experiment on the behavioral response to wage cuts. *Information Systems Research*, 27(2):403–423, 2016.

B. Cowgill. Bias and productivity in humans and algorithms: Theory and evidence from resume screening. Working Paper 19-309, W.E. Upjohn Institute for Employment Research, 2019.

B. Cowgill and P. Perkowski. Delegation in hiring: Evidence from a two-sided audit. *Journal of Political Economy Microeconomics*, 2(4):852–882, 2024.

B. Cowgill and C. E. Tucker. Algorithmic fairness and economics. *Columbia Business School Research Paper*, 2020.

B. Crépon, E. Duflo, M. Gurgand, R. Rathelot, and P. Zamora. Do Labor Market Policies have Displacement Effects? Evidence from a Clustered Randomized Experiment *. *The Quarterly Journal of Economics*, 128(2):531–580, 04 2013. ISSN 0033-5533. doi: 10.1093/qje/qjt001. URL `https://doi.org/10.1093/qje/qjt001`.

Z. Cullen and C. Farronato. Outsourcing tasks online: Matching supply and demand on peer-to-peer internet platforms. *Management Science*, 67(7):3985–4003, 2021.

S. J. Davis, R. J. Faberman, and J. C. Haltiwanger. Recruiting intensity during and after the great recession: National and industry evidence. *American Economic Review*, 102(3):584–88, 2012.

S. J. Davis, R. J. Faberman, and J. C. Haltiwanger. The establishment-level behavior of vacancies and hiring. *The Quarterly Journal of Economics*, 128(2):581–622, 2013.

S. DellaVigna and D. Pope. Stability of experimental results: Forecasts and evidence. *American Economic Journal: Microeconomics*, 14(3):889–925, 2022.

W. Dessein. Authority and communication in organizations. *Review of Economic Studies*, 69(4):811–838, 2002.

A. Dube, J. Jacobs, S. Naidu, and S. Suri. Monopsony in online labor markets. *American Economic Review: Insights*, 2(1):33–46, 2020.

Z. Eckstein and G. J. Van den Berg. Empirical labor search: A survey. *Journal of Econometrics*, 136(2): 531–564, 2007.

A. Filippas, J. J. Horton, and J. Golden. Reputation inflation. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 483–484, 2018.

J. A. Flory, A. Leibbrandt, and J. A. List. Increasing workplace diversity evidence from a recruiting experiment at a fortune 500 company. *Journal of Human Resources*, 56(1):73–92, 2021.

A. Fradkin. Search, matching, and the role of digital marketplace design in enabling trade: Evidence from airbnb. *Working paper*, 2017.

G. Friebel, M. Heinz, and N. Zubanov. Middle managers, personnel turnover, and performance: A long-term field experiment in a retail chain. *Management Science*, 68(1):211–229, 2022.

S. M. Gaddis. Discrimination in the credential society: An audit study of race and college selectivity in the labor market. *Social Forces*, 93(4):1451–1479, 2015.

S. M. Gaddis. *Audit Studies: Behind the Scenes with Theory, Method, and Nuance*. Springer, 2018.

A. Gavazza, S. Mongey, and G. L. Violante. Aggregate recruiting intensity. *American Economic Review*, 108(8):2088–2127, 2018.

L. Giuliano, D. I. Levine, and J. Leonard. Manager race and the race of new hires. *Journal of Labor Economics*, 27(4):589–631, 2009.

J. Gruber. The incidence of mandated maternity benefits. *The American Economic Review*, pages 622–641, 1994.

F. Guvenen and A. A. Smith. Inferring labor income risk and partial insurance from economic choices. *Econometrica*, 82(6):2085–2129, 2014.

M. Hoffman, L. B. Kahn, and D. Li. Discretion in hiring. *The Quarterly Journal of Economics*, 133(2): 765–800, 2018.

H. J. Holzer, L. Katz, and A. Krueger. Job queues and wages. *Quarterly Journal of Economics*, 106:739–68, 1991.

A. Hornstein, P. Krusell, and G. L. Violante. Frictional wage dispersion in search models: A quantitative assessment. *American Economic Review*, 101(7):2873–98, 2011.

J. Horton, W. R. Kerr, and C. Stanton. Digital labor markets and global talent flows. In G. H. Hanson, W. R. Kerr, and S. Turner, editors, *High-Skilled Migration to the United States and Its Economic Consequences*, pages 71–108. University of Chicago Press, 2018.

J. J. Horton. Online labor markets. *Internet and Network Economics*, pages 515–522, 2010.

J. J. Horton. The effects of algorithmic labor market recommendations: Evidence from a field experiment. *Journal of Labor Economics*, 35(2):345–385, 2017.

J. J. Horton. Buyer uncertainty about seller capacity: Causes, consequences, and a partial solution. *Management Science*, 65(8):3518–3540, 2019.

J. J. Horton and R. Johari. At what quality and what price?: Eliciting buyer preferences as a market design problem. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, pages 507–507. ACM, 2015.

J. J. Horton, D. G. Rand, and R. J. Zeckhauser. The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, 14(3):399–425, 2011.

P. Howitt and R. P. McAfee. Costly search and recruiting. *International Economic Review*, pages 89–107, 1987.

IBISWorld. Global hr & recruitment services. Technical report, 2024. Global Industry Report.

IBISWorld. Employment & recruiting agencies industry in the us. Technical report, 2025. URL `https://www.ibisworld.com/united-states/market-research-reports/employment-recruiting-agencies-industry/`. Industry Report 56131.

C. Ichniowski and K. Shaw. The effects of human resource management systems on economic performance: An international comparison of us and japanese plants. *Management Science*, 45(5):704–721, 1999.

B. Jovanovic. Job matching and the theory of turnover. *Journal of Political Economy*, 87(5):972–990, 1979.

J. B. Kessler, C. Low, and C. D. Sullivan. Incentivized resume rating: Eliciting employer preferences without deception. *American Economic Review*, 109(11):3713–3744, 2019.

J. Kleinberg, H. Lakkaraju, J. Leskovec, J. Ludwig, and S. Mullainathan. Human decisions and machine predictions. *The Quarterly Journal of Economics*, 133(1):237–293, 2018.

P. Kline and C. Walters. Reasonable doubt: Experimental detection of job-level employment discrimination. *Econometrica*, 89(2):765–792, 2021.

N. R. Kuncel, D. M. Klieger, B. S. Connelly, and D. S. Ones. Mechanical versus clinical data combination in selection and admissions decisions: A meta-analysis. *Journal of Applied Psychology*, 98(6):1060–1072, 2013.

S. Lang, S. Laumer, C. Maier, and A. Eckhardt. Drivers, challenges and consequences of e-recruiting: a literature review. In *Proceedings of the 49th SIGMIS annual conference on Computer personnel research*, pages 26–35, 2011.

M. D. Leung. Learning to hire? hiring as a dynamic experiential learning process in an online market for contract labor. *Management Science*, 64(12):5651–5668, 2018.

T. R. Lewis. A theory of delegated search for the best alternative. *The RAND Journal of Economics*, 43(3): 391–416, 2012.

D. Li, L. Raymond, and P. Bergman. Hiring as exploration. *The Review of Economic Studies*, 2025. Advance access, doi:10.1093/restud/rdaf040.

J. A. List, A. M. Shaikh, and Y. Xu. Multiple hypothesis testing in experimental economics. *Experimental Economics*, 22(4):773–793, 2019.

D. T. Mortensen and C. A. Pissarides. Job creation and job destruction in the theory of unemployment. *The Review of Economic Studies*, 61(3):397–415, 1994.

A. Pallais. Inefficient hiring in entry-level labor markets. *The American Economic Review*, 104(11):3565–3599, 2014.

A. Pallais and E. G. Sands. Why the referential treatment? evidence from field experiments on referrals. *Journal of Political Economy*, 124(6):1793–1828, 2016.

M. Pellizzari. Employers' search and the efficiency of matching. Discussion Paper 1862, IZA Institute of Labor Economics, 2005. URL `http://dx.doi.org/10.2139/ssrn.866907`.

C. A. Pissarides. *Equilibrium Unemployment Theory*. MIT Press, 2000.

N. Quadlin. The mark of a woman's record: Gender and academic performance in hiring. *American Sociological Review*, 83(2):331–360, 2018.

R. Rogerson, R. Shimer, and R. Wright. Search-theoretic models of the labor market: A survey. *Journal of Economic Literature*, 43(4):959–988, 2005.

A. E. Roth. Marketplaces, markets, and market design. *American Economic Review*, 108(7):1609–58, 2018.

S. Sameen and S. Cornelius. Social networking sites and hiring: How social media profiles influence hiring decisions. *Journal of Business Studies Quarterly*, 7(1):27, 2015.

A. Schiprowski. The role of caseworkers in unemployment insurance: Evidence from unplanned absences. *Journal of Labor Economics*, 2020.

Staffing Industry Analysts. Us staffing industry forecast: September 2022 update. Technical report, Staffing Industry Analysts, 2022.

C. T. Stanton and C. Thomas. Landing the first job: The value of intermediaries in online hiring. *The Review of Economic Studies*, 83(2):810–854, 2015.

G. M. Tallis. The moment generating function of the truncated multi-normal distribution. *Journal of the Royal Statistical Society: Series B (Methodological)*, 23(1):223–229, 1961.

Upwork Research Institute. Freelance forward 2023. Technical report, Upwork Inc., December 2023. URL https://www.upwork.com/research/freelance-forward-2023-research-report. Survey conducted by Edelman Data & Intelligence, $n = 3{,}000$.

A. J. Wood and V. Lehdonvirta. Antagonism beyond employment: how the 'subordinated agency'of labour platforms generates conflict in the remote gig economy. *Socio-Economic Review*, 19(4):1369–1396, 2021.

R. Wright, P. Kircher, B. Julien, and V. Guerrieri. Directed search: A guided tour. Technical report, National Bureau of Economic Research, 2017.

A. Zheng, Y. Hong, and P. A. Pavlou. Matching in two-sided platforms for it services: Evidence from online labor markets. *Working paper*, 2016.

Appendix

A Theory Appendix

This appendix provides formal statements and proofs of the theoretical results in Section 5. Section A.1 fixes notation and derives the joint distribution of signals. Section A.2 gives the closed-form expression for ω (Proposition 1). Section A.3 proves the fill-rate and match-quality decompositions (Results 1, 2). Section A.4 establishes threshold monotonicity in pool size (Lemma 2). Section A.5 derives the analytic gradient system used to verify the comparative statics in Result 3 numerically. Section A.6 extends the baseline to allow visibility into sourcing.

A.1 Primitives and Notation

Workers and quality. Worker j's true quality is $v_j = \alpha_j + \varepsilon_j$, with $\alpha_j \sim N(0, \sigma_\alpha^2)$ captured by the platform score, $\varepsilon_j \sim N(0, \sigma_\varepsilon^2)$ orthogonal to the score, and $\alpha_j \perp \varepsilon_j$.

Information structure. The platform produces $s_j = \alpha_j + \psi_j$ with $\psi_j \sim N(0, \sigma_\psi^2)$. Because both parties read the same profiles, ψ_j enters both evaluations. The recruiter observes $q_j^R = \alpha_j + \varepsilon_j + \psi_j + \xi_j^R$ and the employer observes $q_j^E = \alpha_j + \varepsilon_j + \psi_j + \xi_j^E$, with $\xi_j^R \sim N(0, \sigma_{\xi_R}^2)$ and $\xi_j^E \sim N(0, \sigma_{\xi_E}^2)$. We allow covariances $\text{Cov}(\xi^R, \xi^E) = \sigma_{\xi_{RE}}$, $\text{Cov}(\psi, \xi^R) = \sigma_{\psi\xi_R}$, $\text{Cov}(\psi, \xi^E) = \sigma_{\psi\xi_E}$. All noise terms are uncorrelated with α_j and ε_j.

Worker pools. The algorithmic pool is $\mathcal{A} = \{s_j \geqslant \bar{s}\}$. The recruiter screens $\{s_j < \bar{s}\}$ and invites m workers with $q_j^R \geqslant \bar{s}$, applying the same score cutoff as the algorithm. Let $n = |\mathcal{A}|$ and $N = n + m$. The employer cannot distinguish pool membership ex ante.

Employer's problem. The employer interviews sequentially at cost $c > 0$ per interview, observing signals in random order. In the baseline she applies a fixed standard $\bar{u}$, hiring the first candidate with $q_j^E \geqslant \bar{u}$. Lemma 2 shows the optimal ex-ante threshold in the homogeneous-pool dynamic program is increasing in N; we use the fixed-threshold formulation throughout for tractability and do not characterize the heterogeneous-pool threshold response.

Joint distribution. Collecting $v_j = \alpha_j + \varepsilon_j$ with (**??**)–(**??**), the vector (v_j, q_j^R, q_j^E, s_j) is jointly normal with mean zero and

$$\text{Var}\begin{pmatrix} v_j \\ q_j^R \\ q_j^E \\ s_j \end{pmatrix} = \begin{pmatrix} \sigma_v^2 & \sigma_v^2 & \sigma_v^2 & \sigma_\alpha^2 \\ \sigma_v^2 & \sigma_R^2 & C_{RE} & C_{sR} \\ \sigma_v^2 & C_{RE} & \sigma_E^2 & C_{sE} \\ \sigma_\alpha^2 & C_{sR} & C_{sE} & \sigma_s^2 \end{pmatrix}, \tag{6}$$

where $\sigma_R^2 = \sigma_v^2+\sigma_\psi^2+\sigma_{\xi_R}^2+2\sigma_{\psi\xi_R}$, $\sigma_E^2 = \sigma_v^2+\sigma_\psi^2+\sigma_{\xi_E}^2+2\sigma_{\psi\xi_E}$, $\sigma_s^2 = \sigma_\alpha^2+\sigma_\psi^2$, and the cross-covariances are $C_{RE} = \sigma_v^2 + \sigma_\psi^2 + \sigma_{\psi\xi_R} + \sigma_{\psi\xi_E} + \sigma_{\xi_{RE}}$, $C_{sR} = \sigma_\alpha^2 + \sigma_\psi^2 + \sigma_{\psi\xi_R}$, $C_{sE} = \sigma_\alpha^2 + \sigma_\psi^2 + \sigma_{\psi\xi_E}$.

Standardized thresholds. Throughout we use $z_s = \bar{s}/\sigma_s$, $z_E = \bar{u}/\sigma_E$, and, for the recruited-pool derivations, $b_1 = -\bar{s}/\sigma_s$, $b_2 = \bar{s}/\sigma_R$, $b_3 = \bar{u}/\sigma_E$. The notation z for standardized thresholds reserves α for the quality component.

Lemma 1 (Joint distribution)**.** *The covariance matrix in equation* (6) *obtains with*

$$\sigma_R^2 = \sigma_v^2 + \sigma_\psi^2 + \sigma_{\xi_R}^2 + 2\sigma_{\psi\xi_R}, \quad \sigma_E^2 = \sigma_v^2 + \sigma_\psi^2 + \sigma_{\xi_E}^2 + 2\sigma_{\psi\xi_E}, \quad \sigma_s^2 = \sigma_\alpha^2 + \sigma_\psi^2,$$

$$C_{RE} = \sigma_v^2 + \sigma_\psi^2 + \sigma_{\psi\xi_R} + \sigma_{\psi\xi_E} + \sigma_{\xi_{RE}}, \quad C_{sR} = \sigma_\alpha^2 + \sigma_\psi^2 + \sigma_{\psi\xi_R}, \quad C_{sE} = \sigma_\alpha^2 + \sigma_\psi^2 + \sigma_{\psi\xi_E}.$$

Proof. $\mathrm{Cov}(v, q^R) = \mathrm{Cov}(\alpha + \varepsilon,\, \alpha + \varepsilon + \psi + \xi^R) = \sigma_\alpha^2 + \sigma_\varepsilon^2 = \sigma_v^2$, and analogously for the employer. $\mathrm{Cov}(q^R, q^E) = \sigma_\alpha^2+\sigma_\varepsilon^2+\sigma_\psi^2+\sigma_{\psi\xi_R}+\sigma_{\psi\xi_E}+\sigma_{\xi_{RE}} = C_{RE}$. For covariances with $s = \alpha+\psi$: $\mathrm{Cov}(v, s) = \sigma_\alpha^2$, $\mathrm{Cov}(q^R, s) = C_{sR}$, $\mathrm{Cov}(q^E, s) = C_{sE}$. The signal variances follow by direct expansion. □

Remark 1. *Equal covariance* $\mathrm{Cov}(v, q^R) = \mathrm{Cov}(v, q^E) = \sigma_v^2$ *does not imply equal informativeness, since* $\mathrm{Var}(q^R) \neq \mathrm{Var}(q^E)$ *in general. The signals differ in noise composition:* ξ^R *and* ξ^E *are idiosyncratic, while* ψ *enters both.*

A.2 Closed-Form Quality Differential ω

We compute ω via the regression identity for truncated normals combined with the Tallis (1961) formula for truncated multivariate normal means.

Regression identity. Let $(v, X_1, \ldots, X_k) \sim N(0, \Sigma^*)$ with v scalar. For any rectangular truncation region $\mathcal{A} = \{X \in A\}$,

$$E[v \mid X \in \mathcal{A}] = \beta' E[X \mid X \in \mathcal{A}], \qquad \beta = \Sigma_{XX}^{-1}\Sigma_{Xv}, \tag{7}$$

where Σ_{XX} is the covariance of X and $\Sigma_{Xv} = (\mathrm{Cov}(X_1, v), \ldots, \mathrm{Cov}(X_k, v))'$.

Algorithmic pool. Selection is $s \geqslant \bar{s}$ and $q^E \geqslant \bar{u}$: bivariate lower truncation on (s, q^E). Regression coefficients are $\beta^A = (\Sigma^A)^{-1}c^A$ with

$$\Sigma^A = \begin{pmatrix} \sigma_s^2 & C_{sE} \\ C_{sE} & \sigma_E^2 \end{pmatrix}, \qquad c^A = (\sigma_\alpha^2,\, \sigma_v^2)'.$$

Adjusted thresholds are

$$h_{E|s} = \frac{z_E - \rho_{sE}\, z_s}{\sqrt{1 - \rho_{sE}^2}}, \qquad h_{s|E} = \frac{z_s - \rho_{sE}\, z_E}{\sqrt{1 - \rho_{sE}^2}}. \tag{8}$$

The truncation probability is $P_A = \Phi_2(-z_s, -z_E; \rho_{sE})$. By the bivariate Tallis formula,

$$E[X_i^A \mid \mathcal{A}] = \frac{1}{P_A}\left[\frac{\Sigma_{i1}^A}{\sigma_s}\,\phi(z_s)\,\Phi(-h_{E|s}) + \frac{\Sigma_{i2}^A}{\sigma_E}\,\phi(z_E)\,\Phi(-h_{s|E})\right] \tag{9}$$

for $i \in \{1,2\}$, and $E[v \mid \mathcal{A}] = \beta_s^A\, E[s \mid \mathcal{A}] + \beta_E^A\, E[q^E \mid \mathcal{A}]$.

Recruited pool. Selection is $s < \bar{s}$, $q^R \geqslant \bar{s}$, $q^E \geqslant \bar{u}$. Let $Y_1 = -s$, $Y_2 = q^R$, $Y_3 = q^E$, so the region is all-lower truncation $Y_j \geqslant b_j \sigma_j^Y$. The covariance of Y is $\Sigma^Y = D\Sigma D$ where $D = \mathrm{diag}(-1,1,1)$, with $\rho_{12}^Y = -\rho_{sR}$, $\rho_{13}^Y = -\rho_{sE}$, $\rho_{23}^Y = \rho_{RE}$. Partial correlations are

$$\rho_{k\ell\cdot j}^Y = \frac{\rho_{k\ell}^Y - \rho_{kj}^Y\,\rho_{\ell j}^Y}{\sqrt{(1-(\rho_{kj}^Y)^2)(1-(\rho_{\ell j}^Y)^2)}}, \tag{10}$$

and adjusted thresholds are $h_{\ell|j}^Y = (-b_\ell + \rho_{\ell j}^Y\, b_j)/\sqrt{1-(\rho_{\ell j}^Y)^2}$. The truncation probability is $P_R = \Phi_3(-b_1, -b_2, -b_3; \rho_{12}^Y, \rho_{13}^Y, \rho_{23}^Y)$, and by the trivariate Tallis formula,

$$E[Y_i \mid Y \geqslant b] = \frac{1}{P_R}\sum_{j=1}^{3} \frac{\Sigma_{ij}^Y}{\sigma_j^Y}\,\phi(b_j)\,\Phi_2\Big(h_{k|j}^Y,\, h_{\ell|j}^Y;\, \rho_{k\ell\cdot j}^Y\Big)\,, \tag{11}$$

where (k,ℓ) are the two indices other than j. Back-transforming, $E[s \mid R] = -E[Y_1 \mid Y \geqslant b]$, $E[q^R \mid R] = E[Y_2 \mid Y \geqslant b]$, $E[q^E \mid R] = E[Y_3 \mid Y \geqslant b]$, and $E[v \mid R] = (\beta^R)' E[X \mid R]$ with $\beta^R = \Sigma^{-1}(\sigma_\alpha^2,\, \sigma_v^2,\, \sigma_v^2)'$.

Proposition 1 (Closed-form ω)**.**

$$\omega = (\beta^R)' E[X \mid R] \;-\; (\beta^A)' E[X^A \mid \mathcal{A}], \tag{12}$$

with all quantities defined in equations (7)–(11).

A.3 Hire Probability and Match Quality Effects

Pass rates. Each worker independently passes the employer screen with a pool-dependent probability:

$$p_A = P(q^E \geqslant \bar{u} \mid s \geqslant \bar{s}) = \frac{\Phi_2(-z_E, -z_s; \rho_{sE})}{\Phi(-z_s)}, \tag{13}$$

$$p_R = P(q^E \geqslant \bar{u} \mid s < \bar{s},\, q^R \geqslant \bar{s}) = \frac{P_R}{\Phi_2(\bar{s}/\sigma_s,\, -\bar{s}/\sigma_R;\, -\rho_{sR})}. \tag{14}$$

Results 1 and 2 are stated in the body. We collect their proofs here.

Proof of Result 1. $P_T - P_C = [1-(1-p_A)^n(1-p_R)^m] - [1-(1-p_A)^n] = (1-p_A)^n[1-(1-p_R)^m]$. Both factors are nonnegative in $[0,1]$, and the inequality is strict when $p_R > 0$ and $p_A < 1$. $\square$

Proof of Result 2. At fixed $\bar{u}$, $Q_T = \pi_A E[v \mid q^E \geqslant \bar{u}, \mathcal{A}] + \pi_R E[v \mid q^E \geqslant \bar{u}, R]$ with $\pi_A = 1 - \pi_R$. Since $Q_C = E[v \mid q^E \geqslant \bar{u}, \mathcal{A}]$, $\Delta_Q = \pi_R \omega$. A threshold response $\bar{u} \to \bar{u}' = \bar{u} + O(m/n)$ changes $E[v \mid q^E \geqslant \bar{u}, k]$ by $O(m/n)$, since the conditional expectation is Lipschitz in $\bar{u}$ under the stated density regularity. $\square$

A.4 Threshold Effects

The following lemma shows that if the employer optimizes dynamically over a homogeneous finite pool, the optimal ex-ante threshold is increasing in pool size. We do not prove an analogous bound for the heterogeneous case.

Lemma 2 (Threshold monotonicity)**.** *Let $\mu(x) = E[v \mid q^E = x]$ be the posterior mean. Suppose $c < E[\max\{\mu(q^E), 0\}]$. Then the optimal ex-ante reservation threshold $\bar{u}(N)$ is strictly increasing in N.*

Proof. In the Gaussian model μ is strictly increasing in x, so the hiring rule $\mu(q^E_{k+1}) \geqslant V_N(k+1)$ is equivalent to a threshold rule on q^E. Let $V_N(k)$ denote the continuation value when the pool size is N and k candidates have been interviewed, with boundary $V_N(N) = 0$ (no candidates left). The Bellman equation is

$$V_N(k) = E[\max\{\mu(q^E_{k+1}), V_N(k+1)\}] - c, \qquad k = 0, 1, \dots, N-1,$$

and the optimal ex-ante threshold satisfies $\bar{u}(N) = \mu^{-1}(V_N(0))$.

We prove $V_{N+1}(k) > V_N(k)$ for all $k \leqslant N$ by backward induction on $N - k$.

Base case ($k = N$): $V_N(N) = 0$ by definition, while $V_{N+1}(N) = E[\max\{\mu(q^E_{N+1}), 0\}] - c > 0$ by the assumption on c. Hence $V_{N+1}(N) > V_N(N)$.

Inductive step: assume $V_{N+1}(k+1) > V_N(k+1)$ for some $k < N$. Then

$$\begin{aligned} V_{N+1}(k) - V_N(k) &= E\big[\max\{\mu(q^E_{k+1}), V_{N+1}(k+1)\} - \max\{\mu(q^E_{k+1}), V_N(k+1)\}\big] \\ &> 0, \end{aligned}$$

where strict inequality holds because q^E has a continuous distribution, so the event $\{V_N(k+1) < \mu(q^E_{k+1}) < V_{N+1}(k+1)\}$ has positive probability; on the complement the integrand is nonnegative. Hence $V_{N+1}(0) > V_N(0)$, and since μ is strictly increasing, $\bar{u}(N+1) > \bar{u}(N)$. $\square$

Remark 2. *In our experiment $m/n \approx 0.13$. By the envelope theorem applied to the homogeneous-pool problem above, $\bar{u}(n+m) - \bar{u}(n) = O(m/n)$. This bound holds for the homogeneous setting of Lemma 2; it is a natural benchmark for the heterogeneous case but is not a formal result there. The direction of force: pool growth raises the reservation threshold,is the same under heterogeneity, but the magnitude of the threshold response and the extent to which it offsets the mechanical fill-rate gain are not characterized here.*

A.5 Comparative Statics

The treatment effects ω, Δ_P, Δ_Q depend on θ through a layered chain:

$$\theta \xrightarrow{\text{linear}} \Sigma \xrightarrow{\text{algebraic}} (\rho,\, \beta,\, z_{\text{std}}) \xrightarrow{\Phi,\, \phi} (p_A,\, p_R,\, E[X|\cdot]) \xrightarrow{\text{linear}} (\omega,\, \Delta_P,\, \Delta_Q).$$

All quantities are closed-form functions of θ, the thresholds $(\bar{s}, \bar{u})$, and the pool sizes (n, m), requiring only univariate, bivariate, and trivariate normal CDF and PDF evaluations. We derive $\nabla_\theta(\omega, \Delta_P, \Delta_Q)$ layer by layer.

Layer 1: $\theta \to \Sigma$. The map from structural parameters to Σ, the covariance of (s, q^R, q^E), is linear:

$$\frac{\partial \Sigma}{\partial \sigma_\alpha^2} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \qquad \frac{\partial \Sigma}{\partial \sigma_\varepsilon^2} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}, \qquad \frac{\partial \Sigma}{\partial \sigma_\psi^2} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix},$$

$$\frac{\partial \Sigma}{\partial \sigma_{\xi_R}^2} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \frac{\partial \Sigma}{\partial \sigma_{\xi_E}^2} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \frac{\partial \Sigma}{\partial \sigma_{\psi \xi_R}} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$

$$\frac{\partial \Sigma}{\partial \sigma_{\psi \xi_E}} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 2 \end{pmatrix}, \qquad \frac{\partial \Sigma}{\partial \sigma_{\xi_{RE}}} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

The algorithmic submatrix Σ^A on (s, q^E) uses entries $(1,1)$, $(1,3)$, $(3,3)$ of Σ. Note that $\partial\Sigma/\partial\sigma_\alpha^2 = \partial\Sigma/\partial\sigma_\psi^2$: both variances enter Σ symmetrically through $s = \alpha + \psi$ and propagate identically to q^R and q^E. However, the two parameters are distinguished through the quality–signal covariance vectors $c^R = (\sigma_\alpha^2, \sigma_v^2, \sigma_v^2)'$ and $c^A = (\sigma_\alpha^2, \sigma_v^2)'$: quality $v = \alpha + \varepsilon$ depends on σ_α^2 but not σ_ψ^2, so $\partial c^R/\partial\sigma_\alpha^2 \neq 0$ while $\partial c^R/\partial\sigma_\psi^2 = 0$. This asymmetry is what makes algorithmic improvement (σ_α^2 up, σ_ε^2 down) different from reducing score noise (σ_ψ^2 down).

Layer 2: $\Sigma \to (\rho, \beta, z_{\text{std}})$. Standard deviations: $\sigma_i = \sqrt{\Sigma_{ii}}$, with $\partial\sigma_i/\partial\theta_k = (2\sigma_i)^{-1}\,\partial\Sigma_{ii}/\partial\theta_k$. Correlations:

$$\frac{\partial \rho_{ij}}{\partial \theta_k} = \frac{1}{\sigma_i \sigma_j}\left[\frac{\partial \Sigma_{ij}}{\partial \theta_k} - \frac{\rho_{ij}}{2}\left(\frac{\partial \Sigma_{ii}/\partial\theta_k}{\Sigma_{ii}} + \frac{\partial \Sigma_{jj}/\partial\theta_k}{\Sigma_{jj}}\right)\right]. \tag{15}$$

Partial correlations in Y-space:

$$\frac{\partial \rho^Y_{k\ell\cdot j}}{\partial \rho^Y_{k\ell}} = \frac{1}{\sqrt{(1-(\rho^Y_{kj})^2)(1-(\rho^Y_{\ell j})^2)}}, \qquad \frac{\partial \rho^Y_{k\ell\cdot j}}{\partial \rho^Y_{kj}} = \frac{-\rho^Y_{\ell j} + \rho^Y_{k\ell\cdot j}\,\rho^Y_{kj}}{1-(\rho^Y_{kj})^2}, \tag{16}$$

and symmetrically for $\partial/\partial\rho^Y_{\ell j}$. Standardized thresholds: $z_s = \bar{s}/\sigma_s$ with $\partial z_s/\partial\theta_k = -\bar{s}\,(2\sigma_s^3)^{-1}\,\partial\Sigma_{ss}/\partial\theta_k$; analogously for z_E, b_1, b_2, b_3. Adjusted thresholds:

$$\frac{\partial h^Y_{\ell|j}}{\partial b_\ell} = \frac{-1}{\sqrt{1-(\rho^Y_{\ell j})^2}}, \quad \frac{\partial h^Y_{\ell|j}}{\partial b_j} = \frac{\rho^Y_{\ell j}}{\sqrt{1-(\rho^Y_{\ell j})^2}}, \quad \frac{\partial h^Y_{\ell|j}}{\partial \rho^Y_{\ell j}} = \frac{b_j + \rho^Y_{\ell j}\, h^Y_{\ell|j}}{1-(\rho^Y_{\ell j})^2}. \tag{17}$$

Regression coefficients: $\beta^R = \Sigma^{-1}c^R$, so

$$\frac{\partial\beta^R}{\partial\theta_k} = -\Sigma^{-1}\frac{\partial\Sigma}{\partial\theta_k}\,\beta^R + \Sigma^{-1}\frac{\partial c^R}{\partial\theta_k}, \qquad \frac{\partial\beta^A}{\partial\theta_k} = -(\Sigma^A)^{-1}\frac{\partial\Sigma^A}{\partial\theta_k}\,\beta^A + (\Sigma^A)^{-1}\frac{\partial c^A}{\partial\theta_k}. \tag{18}$$

Layer 3: Pass rates and truncated means. *Algorithmic pass rate.* $p_A = \Phi_2(-z_E, -z_s; \rho_{sE})/\Phi(-z_s)$. By the quotient rule,

$$\frac{\partial p_A}{\partial\theta_k} = \frac{1}{\Phi(-z_s)}\frac{\partial\Phi_2}{\partial\theta_k} + \frac{p_A}{\Phi(-z_s)}\,\phi(z_s)\,\frac{\partial z_s}{\partial\theta_k}, \tag{19}$$

where $\partial\Phi_2/\partial\theta_k$ uses the chain rule through (z_E, z_s, ρ_{sE}) via the standard identities

$$\frac{\partial\Phi_2(a,b;\rho)}{\partial a} = \phi(a)\,\Phi\!\left(\frac{b-\rho a}{\sqrt{1-\rho^2}}\right), \qquad \frac{\partial\Phi_2(a,b;\rho)}{\partial\rho} = \phi_2(a,b;\rho). \tag{20}$$

Recruited pass rate. $p_R = P_R/D_R$ with $D_R = \Phi_2(\bar{s}/\sigma_s,\ -\bar{s}/\sigma_R;\ -\rho_{sR})$. Same quotient-rule structure, with P_R differentiated via

$$\frac{\partial\Phi_3(a_1,a_2,a_3;\rho)}{\partial a_j} = \phi(a_j)\,\Phi_2\big(h_{k|j},\ h_{\ell|j};\ \rho_{k\ell\cdot j}\big)\,, \tag{21}$$

where (k,ℓ) are the indices other than j.

Truncated means. Each term in the Tallis sum (11) is a product $(\Sigma^Y_{ij}/\sigma^Y_j)\,\phi(b_j)\,\Phi_2(h^Y_{k|j}, h^Y_{\ell|j}; \rho^Y_{k\ell\cdot j})$. Differentiating,

$$\frac{\partial}{\partial\theta_k}E[Y_i \mid Y \geqslant b] = \frac{1}{P_R}\sum_{j=1}^{3}\frac{\partial}{\partial\theta_k}\left[\frac{\Sigma^Y_{ij}}{\sigma^Y_j}\,\phi(b_j)\,\Phi_2(\cdot)\right] - \frac{E[Y_i \mid Y \geqslant b]}{P_R}\frac{\partial P_R}{\partial\theta_k}. \tag{22}$$

Each summand differentiates via the product rule into terms involving $\partial(\Sigma^Y_{ij}/\sigma^Y_j)/\partial\theta_k$, $\partial\phi(b_j)/\partial\theta_k = -b_j\,\phi(b_j)\,\partial b_j/\partial\theta_k$, and $\partial\Phi_2/\partial\theta_k$ via (20). The algorithmic truncated means differentiate analogously using (9).

Layer 4: Treatment effects. *Quality differential:*

$$\frac{\partial\omega}{\partial\theta_k} = \left(\frac{\partial\beta^R}{\partial\theta_k}\right)'\mu^R + (\beta^R)'\frac{\partial\mu^R}{\partial\theta_k} - \left(\frac{\partial\beta^A}{\partial\theta_k}\right)'\mu^A - (\beta^A)'\frac{\partial\mu^A}{\partial\theta_k}, \tag{23}$$

where $\mu^R = E[X \mid R]$ and $\mu^A = E[X^A \mid \mathcal{A}]$.

Hire probability effect:

$$\frac{\partial \Delta_P}{\partial \theta_k} = -n(1-p_A)^{n-1}[1-(1-p_R)^m]\frac{\partial p_A}{\partial \theta_k} + (1-p_A)^n\, m(1-p_R)^{m-1}\frac{\partial p_R}{\partial \theta_k}. \tag{24}$$

Match quality effect: By the law of total expectation, $\Delta_Q = \pi_R\,\omega$, where π_R is the probability that a treatment-group hire comes from the recruited pool. We do not have a closed form for π_R, but signing $\partial \Delta_Q/\partial \theta_k$ does not require one. When $\pi_R > 0$ is fixed (e.g. along a path that holds the recruited share constant), $\partial \Delta_Q/\partial \theta_k = \pi_R\,\partial\omega/\partial\theta_k$, so the sign of $\partial \Delta_Q/\partial \theta_k$ equals the sign of $\partial\omega/\partial\theta_k$. The comparative statics that follow are therefore stated in terms of ω.

Beyond these cases, the gradient (23) is a sum of four terms that can have opposite signs, so signing $\partial\omega/\partial\theta_k$ for remaining parameters requires evaluating the gradient at specific values. Numerical evaluation of (23) confirms $d\omega/d\delta < 0$ (algorithmic improvement) and $\partial\omega/\partial\sigma_{\xi_{RE}} < 0$ (rising recruiter–employer noise covariance) throughout the parameter space, supporting the claims in Section 5. The gradient is computed by central differences with step $h \sim 10^{-7}$; replication code is in the paper's supplementary materials (`check_monotonicity.py`).

A.6 Extension: Visibility into Recruited Candidates

In the baseline, interview order is random and the employer applies a uniform standard $\bar{u}$ because she cannot distinguish pool membership. We now relax both assumptions: in the *visibility regime*, the platform reveals a binary sourcing label $L_j \in \{R, A\}$ before any interviews. The label is informative about v conditional on q_j^E (assumption (A3) below makes this precise), and it lets the employer (i) set pool- and state-specific acceptance thresholds and (ii) choose interview order. The recruiter's selection rule is held fixed, so $\mathcal{A}$, R, and all joint distributions with (v, s, q^E) are identical to the baseline.

This extension produces two robust results and one ambiguous one. The robust ones: visibility weakly raises employer payoff (Proposition 2), and the optimal policy interviews algorithmic candidates first (Proposition 3). The ambiguous one: fill-rate and match-quality effects cannot be signed relative to the baseline without further restrictions (Proposition 4).

Assumptions.

(A1) *(MLRP within pools)* $\mu_k(x) \equiv E[v \mid q^E = x, j \in k]$ is strictly increasing in x for each $k \in \{A, R\}$.

(A2) *(FOSD)* $1 - F_A^E(x) \geqslant 1 - F_R^E(x)$ for all x.

(A3) *(Signal dominance)* $\mu_A(x) \geqslant \mu_R(x)$ for all x.

In the Gaussian model, (A1) holds whenever $\mathrm{Cov}(v, q^E \mid k) > 0$, which is straightforward to verify. (A2) and (A3) are parametric sufficient conditions that we expect to hold when σ_ε^2 is small relative to $\sigma_\alpha^2 + \sigma_\psi^2$, the same regime in which $\omega < 0$ is most pronounced (see Section A.5). We state these as assumptions rather than derived results.

Independence of draws within pools. Each candidate j has an independent draw of the latent vector $(\alpha_j, \varepsilon_j, \psi_j, \xi_j^E)$. Conditional on pool membership $k \in \{A, R\}$, the signals q_j^E are i.i.d. across candidates within the pool, with common CDF F_k^E. This justifies treating interviews as independent draws.

Dynamic program. The employer's problem is a finite-horizon sequential search with two classes. Within each interview: (i) she chooses a pool, (ii) pays c and observes $q_j^E = x$, and (iii) decides whether to hire. Her expected payoff from hiring is $\mu_k(x) = E[v \mid q^E = x, j \in k]$; if she continues, her payoff is the continuation value $V(\cdot)$. The optimal post-signal rule is: hire iff $\mu_k(x) \geqslant V(\cdot)$, and the acceptance threshold $u_k(i,j)$ solves $\mu_k(u_k(i,j)) = V(\cdot)$, well-defined by (A1).

The distributions F_A^E and F_R^E are known to the employer (Lemma 1), so past signals carry no information about future draws: the payoff-relevant state is (i,j), the remaining pool sizes.

The gross expected value of interviewing one candidate from pool k at continuation value v_0 is

$$H_k(v_0) \equiv \int_{-\infty}^{\infty} \max(\mu_k(x), v_0)\, dF_k^E(x). \tag{25}$$

The Bellman equation is

$$V(i,j) = \max\{H_O(V(i-1,j)) - c,\, H_R(V(i,j-1)) - c,\, 0\}, \tag{26}$$

with $V(0,0) = 0$, and H_O (H_R) available only when $i \geqslant 1$ ($j \geqslant 1$). Here H_O denotes H_A; the notation follows the convention that O stands for *organic* (algorithmic) in our empirical context.

Main results.

Proposition 2 (Visibility weakly raises employer payoff)**.** *$W^{\text{vis}} = V^{\text{vis}}(n,m) \geqslant W$, where W is the baseline payoff.*

Proof. The visibility feasible set contains the baseline strategy: at each state (i,j), draw from A with probability $i/(i+j)$ and from R with probability $j/(i+j)$, applying the common threshold $\bar{u}$. This replicates the baseline outcome distribution. Since the optimal visibility policy maximizes over a weakly larger set, $W^{\text{vis}} \geqslant W$. □

Lemma 3 (Marginal value ordering)**.** *Suppose $H_O(v) \geqslant H_R(v)$ for all $v \geqslant 0$. Then $V(i-1,j) \geqslant V(i,j-1)$ for all $i,j \geqslant 1$. Assumptions (A1)–(A3) are sufficient for $H_O(v) \geqslant H_R(v)$.*

Proof. Sufficient condition. For any $v \geqslant 0$,

$$H_O(v) = E_{F_A^E}[\max(\mu_A(q^E), v)] \geqslant E_{F_A^E}[\max(\mu_R(q^E), v)] \geqslant E_{F_R^E}[\max(\mu_R(q^E), v)] = H_R(v),$$

where the first inequality uses (A3) pointwise and the second uses (A2) with $\max(\mu_R(x), v)$ nondecreasing in x by (A1). By (A1), μ_k is strictly increasing and hence invertible on its range, so $H_k(v_0)$ is strictly increasing in v_0 (the integrand $\max(\mu_k(x), v_0)$ is strictly increasing in v_0 for $x < \mu_k^{-1}(v_0)$, a set of positive F_k^E-measure).

Induction on $i + j$. Base ($i = j = 1$): $V(1, 0) = \max(H_O(0) - c, 0) \geqslant \max(H_R(0) - c, 0) = V(0, 1)$.

Inductive step. Assume $V(a - 1, b) \geqslant V(a, b - 1)$ for all (a, b) with $a + b \leqslant K$. Consider (i, j) with $i + j = K + 1$.

Case $i, j \geqslant 2$ (interior). By the inductive hypothesis, $V(i - 2, j) \geqslant V(i - 1, j - 1) \geqslant V(i, j - 2)$. Since H_O, H_R are strictly increasing, $H_O(V(i - 2, j)) \geqslant H_O(V(i - 1, j - 1))$ and $H_R(V(i - 1, j - 1)) \geqslant H_R(V(i, j - 2))$. The maximum over the Bellman options is nondecreasing in each input, giving $V(i - 1, j) \geqslant V(i, j - 1)$.

Case $i = 1$. At $(0, j)$ only R is available: $V(0, j) = \max(H_R(V(0, j - 1)) - c, 0)$. At $(1, j - 1)$, $V(1, j - 1) \geqslant H_O(V(0, j - 1)) - c \geqslant H_R(V(0, j - 1)) - c$, so $V(1, j - 1) \geqslant V(0, j)$.

Case $j = 1$. Analogous, using only A at $(i, 0)$ and the inductive hypothesis $V(i-2, 1) \geqslant V(i-1, 0)$. □

Proposition 3 (Algorithmic-first ordering)**.** *Under (A1)–(A3), $W_A(i, j) \geqslant W_R(i, j)$ at every state (i, j) with $i, j \geqslant 1$: the employer always interviews algorithmic candidates first. The key sufficient condition is $H_O(v) \geqslant H_R(v)$ for all v; the result holds under any primitives delivering this inequality.*

Proof. $H_O(V(i - 1, j)) \geqslant H_O(V(i, j - 1)) \geqslant H_R(V(i, j - 1))$, where the first inequality uses that H_O is strictly increasing and $V(i - 1, j) \geqslant V(i, j - 1)$ (Lemma 3), and the second uses $H_O(v) \geqslant H_R(v)$. Subtracting c gives $W_A(i, j) \geqslant W_R(i, j)$. □

Proposition 4 (Ambiguity of Δ_P^{vis} and Δ_Q^{vis})**.** *Δ_P^{vis} and Δ_Q^{vis} cannot be signed relative to the baseline without further parametric restrictions.*

Proof. At fixed thresholds with independent pass events, hire probabilities are order-invariant: $P = 1 - \prod_k(1 - p_k)$. The sign of $\Delta_P^{\mathrm{vis}} - \Delta_P$ therefore depends on how pool-specific visibility thresholds differ from the common baseline threshold $\bar{u}$.

We construct two illustrative examples at $(i, j) = (1, 1)$. Pass rates are chosen to match Gaussian-model features ($p_A > p_R$ and $\mu_A > \mu_R$ when $\omega < 0$) without being derived from a specific parameter vector.

Case 1 ($\Delta_P^{\mathrm{vis}} > \Delta_P$). Baseline common threshold yields $p_A = 0.5$, $p_R = 0.3$. Under visibility, the low continuation value after interviewing A first makes marginal R candidates more attractive, lowering the R-threshold to $p_R^{\mathrm{vis}} = 0.45$. Then $P^{\mathrm{base}} = 1 - (0.5)(0.7) = 0.65 < 0.725 = 1 - (0.5)(0.55) = P^{\mathrm{vis}}$.

Case 2 ($\Delta_P^{\text{vis}} < \Delta_P$). Same signal structure, but the optimal visibility policy raises the R-threshold: the employer, anticipating interviewing A first, sets a high bar on R because R-candidates' expected quality conditional on passing is low. With $p_A^{\text{vis}} = 0.5$, $p_R^{\text{vis}} = 0.1$, $P^{\text{vis}} = 1 - (0.5)(0.9) = 0.55 < 0.65 = P^{\text{base}}$.

Both signs are feasible, so $\Delta_P^{\text{vis}} - \Delta_P$ cannot be signed. The argument for Δ_Q^{vis} is analogous: in Case 1 the lower R-threshold admits more low-quality recruited hires; in Case 2 the higher R-threshold screens them out. □

Interpretation. The baseline results $\omega < 0$ and $\Delta_P \approx 0$ are robust to relaxing random order and common thresholds. The payoff bound (Proposition 2) is unconditional. The algorithmic-first ordering (Proposition 3) holds under (A1)–(A3), conditions aligning with the regime where $\omega < 0$ is most pronounced. Fill-rate and match-quality effects remain ambiguous, with the employer trading off these margins to maximize total payoff.

A.7 Platform Interface

A.8 Initial Message

> Hi {employer first name}, I'm your [Hiring Consultant]. Chat with me about any {platform} questions or issues. I'm here to help! I'm online Monday through Friday, 8:00 AM to 4:00 PM Pacific Time. If I'm offline when you contact me, I'll reply as soon as I'm back. Freelancers will begin submitting proposals for your {job opening title} job. In the meanwhile, you can invite specific freelancers to submit proposals {at this link} and I'll start connecting you with top talent for this job. Do you have any questions as you get started?

A.9 Selection into Treatment

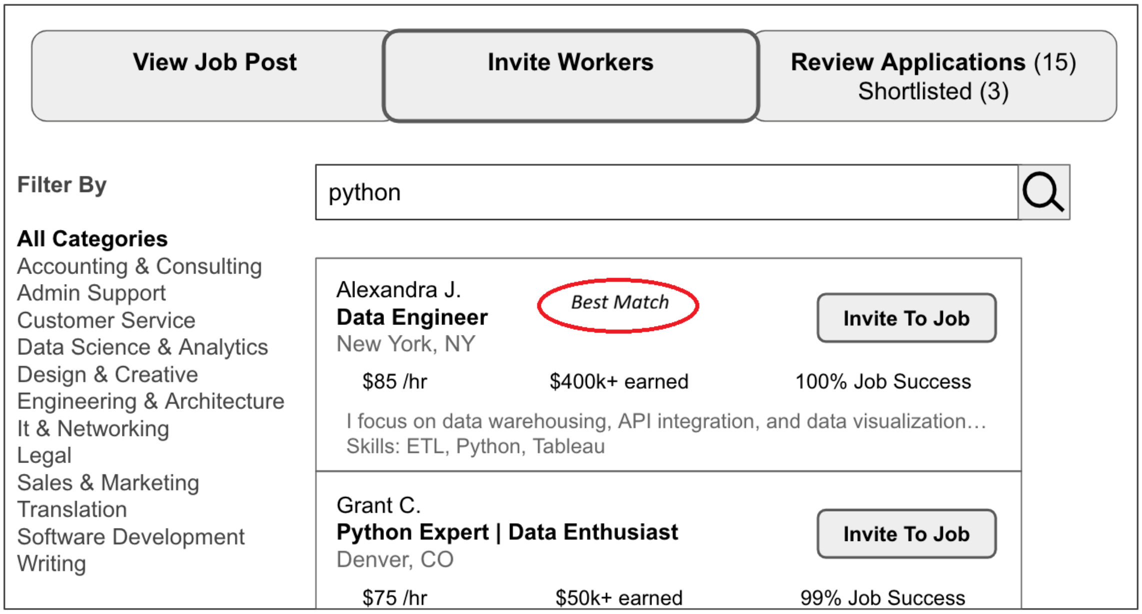
Recruiters endogenously determined whether to provide shortlisting or recruiting help to the job posts they evaluated. To characterize how assisted posts differed from unassisted ones, we restrict the sample to the 40,952 job posts that the Recruiter marked as "worked" and estimate a logistic model:

$$\text{helped}_i = \alpha + \sum_j \beta_j X_{ij} + \varepsilon_i$$

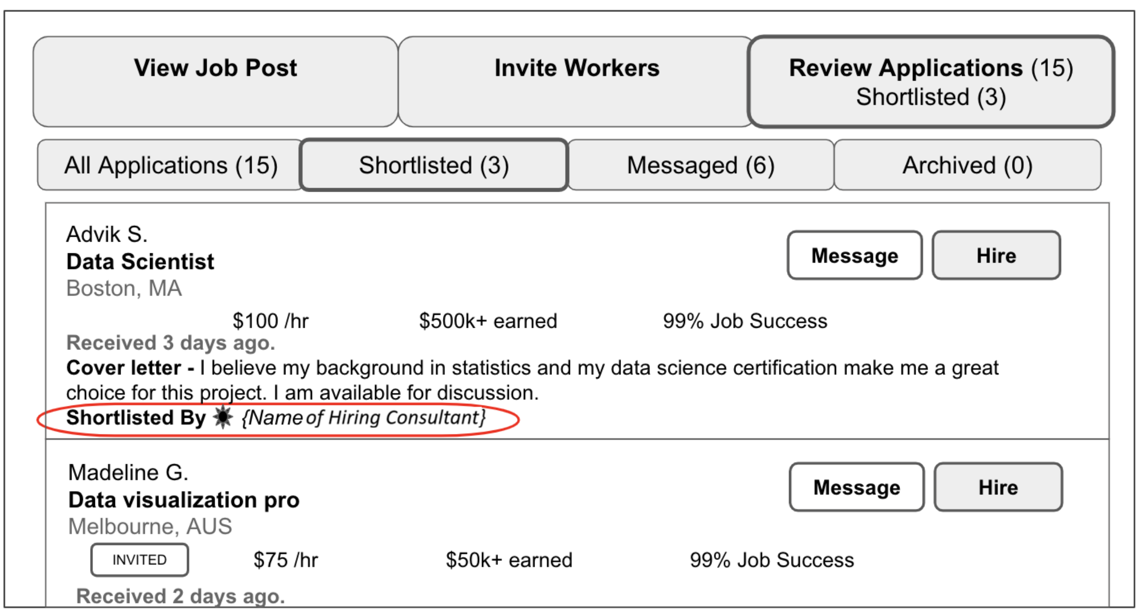
where helped_i equals 1 if the Recruiter ever shortlists or invites workers. The predictors include the number of applications and recommended applications at the time of evaluation, projected job value (expected hours times wages), employer shortlists and invites prior to evaluation, and other job-level characteristics.

Table A.1 reports the results. Jobs that received help had fewer applications and fewer shortlists at the time the Recruiter began working on them. They were also more likely to be graphic design jobs and to require more skills. This pattern is consistent with recruiters targeting assistance toward jobs that appeared to need it most.

Figure A.1: Job post manager on the platform, stylized version



(a) Inviting workers to apply for a job



(b) Proposal manager with shortlisting feature

Notes: Panel (a) shows the employer interface for inviting workers to apply. Employers can search the platform for available workers and see each worker's wage bid, name, self-reported skills, and platform-verified information such as total hours worked and average feedback rating. Panel (b) shows the proposal manager with the shortlisting feature, where employers can view all applications or filter to see only shortlisted applications.

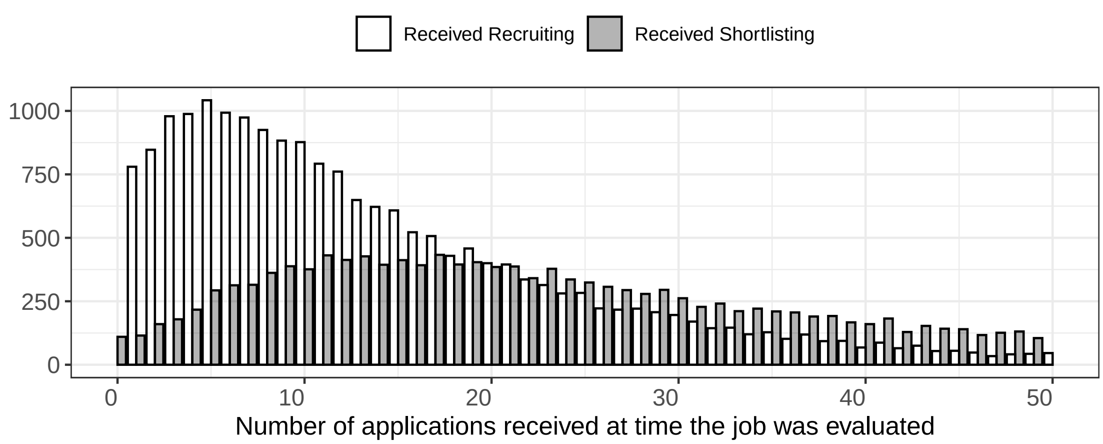
Table A.1: Which worked jobs get help from Recruiters?

	Dependent variable:
	Job post gets at least one invite or shortlist
Number of Applications	−0.006*** (0.001)
Number of Recommended Applications	0.001 (0.001)
Employer Invites	−0.200*** (0.003)
Employer Shortlists	−0.032*** (0.004)
Administrative Job	0.191*** (0.063)
Software Job	−0.123*** (0.032)
Design Job	−0.219*** (0.065)
Num Skills Req	−0.007** (0.003)
Low Skill Job	0.200** (0.081)
High Skill Job	−0.058** (0.028)
Constant	2.348*** (0.039)
Observations	40,185

Notes: This table compares job posts which get worked by Recruiters but do not receive recruiting or shortlisting help, with those that are worked and do receive at least one type of help. The sample is all treated jobs which are worked by a Recruiter. Standard errors are in parentheses. Significance indicators: $p \leqslant 0.05$: *, $p \leqslant 0.01$: **, and $p \leqslant .001$: ***.

A.10 Application Distribution by Treatment

Figure A.2: The distribution of applications received at the time the Recruiter first assessed a job post, by what action the Recruiter then took (recruiting or shortlisting)



Notes: This sample is job posts in the experimental sample assigned to treatment. The x-axis is the number of applications a given job post has at the time a Hiring Consultant first sees the job. The solid bars are the density of job posts that receive recruiting help, while the dashed bars are the density of job posts that receive shortlisting. Recruiters provided recruiting help to jobs with few applications and shortlisting help to jobs with many applications, with the decision driven primarily by application count at time of evaluation.

A.11 Neither Recruiting Nor Shortlisting Improved Hiring

We fit logit models to predict receipt of recruiting and shortlisting help using job characteristics available at the time of evaluation. The predictors include job category, post date, expertise tier, and skill count (with polynomial terms). Figure A.12 in Appendix A.17 shows that the distributions of propensity scores are nearly identical across treatment and control groups, confirming the model captures pre-treatment characteristics. The IV strategy uses treatment assignment interacted with each propensity score as instruments for the two endogenous treatments. Identification rests on randomization of treatment assignment together with a separability assumption: conditional on both propensity scores, the assignment-by-recruiting-propensity instrument affects hiring only through recruiting and not through shortlisting (and analogously for the shortlisting instrument). This is a functional-form restriction beyond what randomization delivers; it can fail if recruiting and shortlisting interact in producing hires or if the propensity model is misspecified.

Table A.2 presents the results. The treatment effect on hiring is small and statistically insignificant across all specifications. Weighting by propensity scores does not change this conclusion. IV estimates that isolate the effects of recruiting and shortlisting help separately are also small, statistically insignificant, and not economically meaningful. Neither form improved hiring outcomes.

Table A.2: Effect of Hiring Assistance on Job Fill Rate

	Recruiting	Shortlisting	Recruiting	Shortlisting
	Unweighted		Weighted	
	(1)	(2)	(3)	(4)
Panel A: OLS (Intent-to-Treat)				
Treatment Assigned	0.006	0.006	0.006	0.003
	(0.004)	(0.004)	(0.004)	(0.004)
Panel B: IV (Treatment-on-the-Treated)				
Recruiting Help [IV]	0.016		0.015	
	(0.011)		(0.011)	
Shortlisting Help [IV]		0.009		0.003
		(0.012)		(0.013)
Observations	78954	78954	78954	78954

Note:
All models control for propensity scores. Columns (1)-(2) are unweighted. Columns (3)-(4) weight by the respective propensity score. IV specifications instrument help receipt with treatment assignment interacted with the relevant propensity score. All specifications use heteroskedasticity-robust (Eicker-Huber-White) standard errors. Significance indicators: $p \leqslant 0.05$: *, $p \leqslant 0.01$: **, and $p \leqslant .001$: ***.

A.12 Internal Validity

To assess the effectiveness of randomization, in Table A.3 we report the mean values and t-tests for various pre-randomization attributes of the job post, and we obtain excellent balance on these covariates.

Table A.3: Balance Table

Variable	Control	Treatment	Difference	p-value	Adj. p (Bootstrap)	Adj. p (Bonf.)	Adj. p (FDR)
First Job Post	0.31	0.30	-0.003	0.59	1	1	0.78
Software	0.58	0.59	0.01	0.18	0.91	1	0.58
Administrative	0.06	0.06	0.004	0.09	0.66	1	0.54
Low Skill	0.05	0.04	-0.004	0.04	0.37	0.46	0.46
High Skill	0.48	0.48	-0.003	0.60	1	1	0.78
Num Skills Required	6.13	6.11	-0.01	0.77	1	1	0.83
Large Company	0.02	0.02	-0.001	0.43	1.00	1	0.78
Employer length of time on platform (days)	823.97	839.23	15.26	0.19	0.92	1	0.58
Num previous hires	3.59	3.71	0.12	0.63	1	1	0.78
Wages paid to fixed price jobs	22,538.66	23,169.44	630.79	0.88	1	1	0.88
Average hourly wage paid	16.34	16.20	-0.14	0.55	1	1	0.78
Total hours from previous hires	32,645.02	26,866.45	-5,778.57	0.55	1	1	0.78

Notes: This table reports group summary statistics for several job post attributes in the experiment, as well as a t-test comparing those means. All attributes are defined pre-experiment. Adjusted p-values control the familywise error rate following List, Shaikh, and Xu (2019) with 1,000 bootstrap replications. Bonferroni and Benjamini-Hochberg (FDR) corrections are also reported.

A.13 Intermediate Outcome Tables

Table A.4: Effects of treatment on recruiting

	Dependent variable:			
	Total number recruited		Number of employer recruits	
	(1)	(2)	(3)	(4)
Treatment Assigned	3.21^{***}		-0.06	
	(0.08)		(0.07)	
Treatment Received [IV]		4.89^{***}		-0.09
		(0.11)		(0.10)
Constant	4.22^{***}	4.21^{***}	4.16^{***}	4.16^{***}
	(0.07)	(0.07)	(0.06)	(0.06)
Observations	83,021	83,021	83,021	83,021
Adjusted R^2	0.02	0.07	-0.0000	-0.0002

Notes: This table reports effects of the treatment to recruiting. Sample is experimental sample of high value posts from October 1, 2021 to April 1, 2022. OLS specification regresses outcome on treatment assignment. IV specification uses treatment assignment as an instrument for employers receiving the treatment. Heteroskedasticity-robust (Eicker–Huber–White) standard errors are used in all specifications. Significance indicators: $p \leqslant 0.05$: *, $p \leqslant 0.01$: **, and $p \leqslant .001$: ***.

A.14 Selection-Corrected Wagebill

Table A.9 reports treatment effects on 30-day wagebill from three estimators: OLS on the sample of hired applications, Heckman two-step, and full-information maximum likelihood. The unit of observa-

Table A.5: Effects of treatment assignment to the log number of applications

	Dependent variable:					
	Total applications		Recruited applications		Organic applications	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment Assigned	0.22***		0.41***		0.14***	
	(0.01)		(0.01)		(0.01)	
Treatment Received [IV]		0.34***		0.62***		0.21***
		(0.01)		(0.01)		(0.02)
Constant	2.66***	2.66***	0.54***	0.53***	2.53***	2.53***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	83,029	83,029	83,029	83,029	83,029	83,029
Adjusted R^2	0.01	0.08	0.04	0.15	0.003	0.04

Notes: This table reports effects of treatment to log applications. Recruited applications are applications sent by workers who were recruited to the job post. Organic applications are applicants who were not recruited. Total applications sums organic applications and applications from recruited workers. All outcomes are logged. Sample is experimental sample of high value posts from October 1, 2021 to April 1, 2022. OLS specification regresses outcome on treatment assignment. IV specification uses treatment assignment as an instrument for employers receiving the treatment. Heteroskedasticity-robust (Eicker–Huber–White) standard errors are used in all specifications. Significance indicators: $p \leqslant 0.05$: *, $p \leqslant 0.01$: **, and $p \leqslant .001$: ***.

Table A.6: Effects of treatment on selection

	Dependent variable:			
	Num shortlists		Num employer shortlisted	
	(1)	(2)	(3)	(4)
Treatment Assigned	1.73***		0.88***	
	(0.03)		(0.03)	
Treatment Received [IV]		2.65***		1.34***
		(0.03)		(0.04)
Constant	1.26***	1.25***	1.26***	1.25***
	(0.02)	(0.02)	(0.02)	(0.02)
Observations	83,021	83,021	83,021	83,021
Adjusted R^2	0.03	0.15	0.01	0.07

Notes: This table reports effects of the treatment to number of shortlisted applicants. Sample is experimental sample of high value posts from October 1, 2021 to April 1, 2022. OLS specification regresses outcome on treatment assignment. IV specification uses treatment assignment as an instrument for employers receiving the treatment. Heteroskedasticity-robust (Eicker–Huber–White) standard errors are used in all specifications. Significance indicators: $p \leqslant 0.05$: *, $p \leqslant 0.01$: **, and $p \leqslant .001$: ***.

tion is the application; for posts with multiple hires, wagebill is split equally across the hired applications. The selection equation includes treatment, category group, expertise tier, and $\log(\text{number of applications})$. The outcome equation includes the same controls excluding number of applications. Identification of the selection-corrected estimates relies on $\log(\text{number of applications})$ as an exclusion restriction.

Table A.7: Effects of treatment assignment to the number of interviews conducted by the employer

	Dependent variable:					
	Total interviews		Recruited interviews		Organic interviews	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment Assigned	0.50***		0.33***		0.17***	
	(0.04)		(0.01)		(0.05)	
Treatment Received [IV]		0.76***		0.50***		0.26***
		(0.04)		(0.02)		(0.05)
Constant	2.20***	2.20***	0.39***	0.39***	1.81***	1.81***
	(0.03)	(0.03)	(0.01)	(0.01)	(0.03)	(0.03)
Observations	83,021	83,021	83,021	83,021	83,021	83,021
Adjusted R^2	0.002	0.01	0.01	0.02	0.0003	0.004

Notes: This table reports effects of treatment to employer interviews. Recruited interviews are interviews given by the employer to recruited applicants. Organic interviews are given to applicants who were not recruited. Total interviews sums organic interviews and interviews from recruited workers. Sample is experimental sample of high value posts from October 1, 2021 to April 1, 2022. OLS specification regresses outcome on treatment assignment. IV specification uses treatment assignment as an instrument for employers receiving the treatment. Heteroskedasticity-robust (Eicker–Huber–White) standard errors are used in all specifications. Significance indicators: $p \leqslant 0.05$: *, $p \leqslant 0.01$: **, and $p \leqslant .001$: ***.

A.15 Regression Discontinuity

This section reports density and balance tests, the first-stage relationship, additional regression tables, and robustness checks for the fuzzy regression discontinuity design described in Section 6.2.

Density and covariate balance. Figure A.3 reports McCrary-style density tests (Cattaneo et al., 2020) for the running variable. The density is smooth through the threshold in both the full sample and within each experimental arm, supporting the validity of the RD design.

Table A.8: Effects of treatment on whether or not a job filled for new employers

	Dependent variable:	
	Hired	
	(1)	(2)
Treatment Assigned	0.004	0.01
	(0.004)	(0.004)
First job post	−0.17***	
	(0.01)	
Treatment Assigned * First job post	0.01	
	(0.01)	
Age in weeks		0.0003
		(0.01)
Treatment Assigned * Age		−0.0000
		(0.01)
Constant	0.33***	0.26***
	(0.004)	(0.004)
Observations	83,021	83,021
Adjusted R^2	0.02	0.01

Notes: This table reports effects of treatment whether or not a job filled, by age of employer on the platform. First job post is a binary variable for whether or not it is the employers's first job post on the platform. Age is the number of weeks since the time the employer registered for the platform. Sample is experimental sample of high value posts from October 1, 2021 to April 1, 2022. OLS specification regresses outcome on treatment assignment. IV specification uses treatment assignment as an instrument for employers receiving the treatment. Heteroskedasticity-robust (Eicker–Huber–White) standard errors are used in all specifications. Significance indicators: $p \leqslant 0.05$: *, $p \leqslant 0.01$: **, and $p \leqslant .001$: ***.

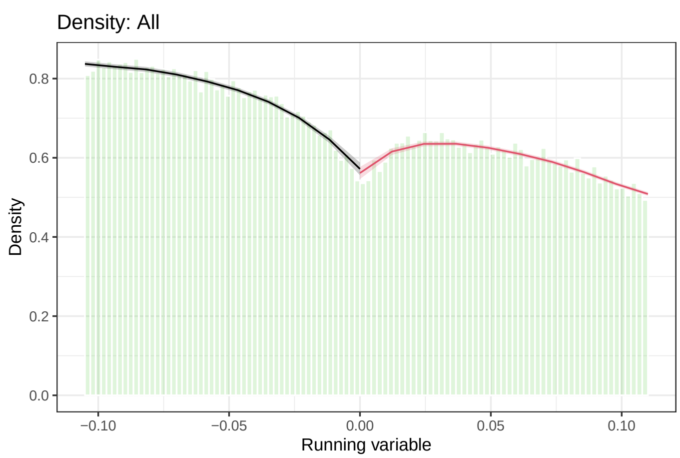


Figure A.3: Density test for the RD running variable (full sample)

Notes: McCrary-style density test (Cattaneo et al., 2020) for the recommendation score running variable. The smooth density on both sides of the threshold supports the validity of the RD design.

Figures A.4–A.5 show that predetermined covariates (log application count, hourly rate indicator) are smooth through the recommendation threshold, further supporting design validity.

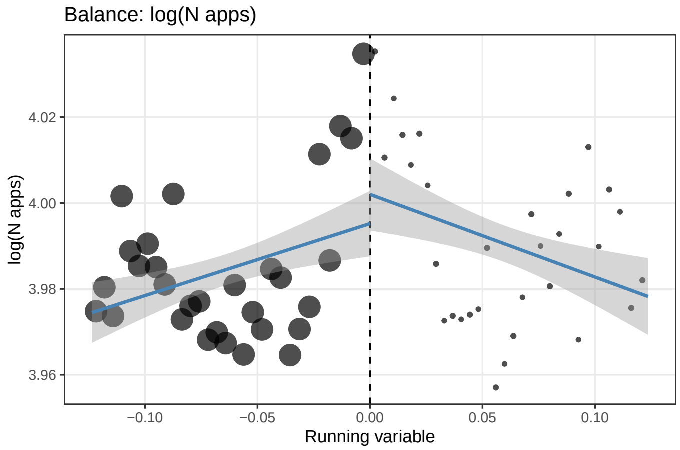


Figure A.4: RD covariate balance: log(N applications)

Notes: Binned scatter plot of log(number of applications) against the recommendation score running variable. Smoothness through the threshold supports the assumption that the running variable is not manipulated.

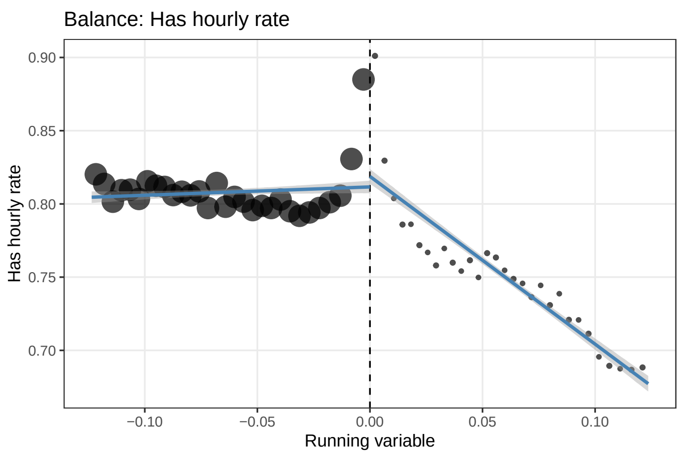


Figure A.5: RD covariate balance: has hourly rate

Notes: Binned scatter plot of the hourly rate indicator against the recommendation score running variable. Smoothness through the threshold supports the assumption that the running variable is not manipulated.

First stage. Figure A.6 shows a sharp discontinuity at the threshold: the recommendation label changes discretely at the score cutoff.

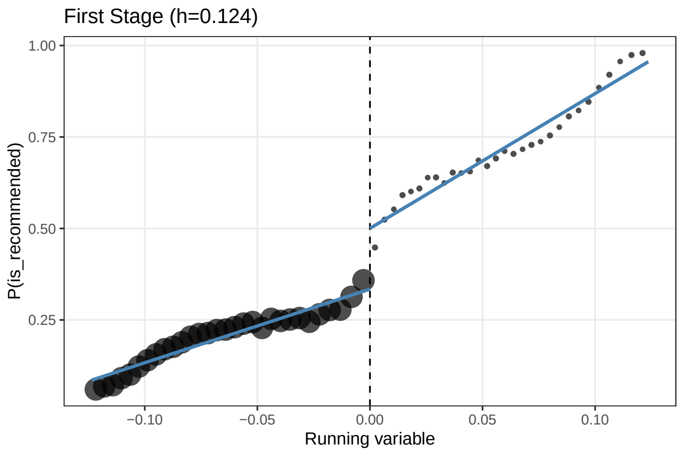


Figure A.6: First stage: recommendation probability at the score threshold

Notes: First-stage relationship between the recommendation score running variable and recommendation status. The sharp discontinuity at the threshold confirms that the recommendation label changes discretely at the score cutoff.

Wagebill. Figure A.7 shows the RD relationship for 30-day wagebill among hired workers.

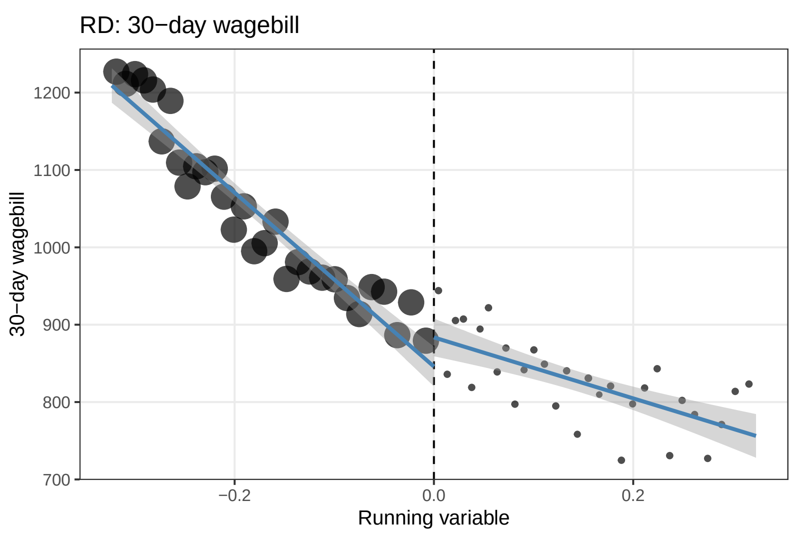


Figure A.7: RD: 30-day wagebill at the recommendation threshold

Notes: Binned scatter plot of 30-day wagebill against the recommendation score running variable among hired workers (post-level, at most one hire per post). Lines are local linear fits estimated separately on each side of the threshold. Corresponding regression estimates appear in Table A.11.

Baseline RD regression tables. Tables A.10–A.12 report fuzzy RD estimates for hiring, wagebill, and shortlisting using the baseline threshold definition. The corresponding binned scatter plot for hiring appears in Section 6.2.

Table A.10: Fuzzy RD: Hiring

	RD (bias-corrected, robust SE)			2SLS w/ shortlisting	
	(1) All	(2) Treated	(3) Control	(4) Base	(5) +Tier
Recommended	0.012*	0.010	0.017	0.011	0.011
	(0.005)	(0.005)	(0.010)	(0.007)	(0.007)
Any shortlisting				0.001	0.001
				(0.035)	(0.035)
N (in BW)	197,273	177,900	71,560	197,273	197,273
Bandwidth h	0.094	0.113	0.141	0.094	0.094

Notes: Outcome: hired (0/1). Columns (1)–(3) report bias-corrected fuzzy RD estimates with robust standard errors. Columns (4)–(5) report 2SLS estimates instrumenting recommendation and shortlisting with the threshold indicator and treatment assignment within the RD bandwidth. A χ^2 test fails to reject equality of the treatment and control RD estimates ($p = 0.502$). SEs clustered by post. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Table A.11: Fuzzy RD: Wagebill

	RD (bias-corrected, robust SE)			2SLS w/ shortlisting	
	(1) All	(2) Treated	(3) Control	(4) Base	(5) +Tier
Recommended	327.229	188.502	837.108	682.395	675.846
	(484.502)	(536.558)	(1008.389)	(739.580)	(746.348)
Any shortlisting				-1916.659	-2035.203
				(2037.353)	(2031.554)
N (in BW)	2,382	1,652	792	2,382	2,382
Bandwidth h	0.323	0.304	0.448	0.323	0.323

Notes: Outcome: 30-day wagebill. Post-level data (at most one hire per post). Columns (1)–(3) report bias-corrected fuzzy RD estimates with robust standard errors. Columns (4)–(5) report 2SLS estimates instrumenting recommendation and shortlisting with the threshold indicator and treatment assignment within the RD bandwidth. A χ^2 test fails to reject equality of the treatment and control RD estimates ($p = 0.570$). SEs clustered by post. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Table A.12: RD: Recommendation and Shortlisting

	Any shortlisting			Shortlisted by recruiter
	(1) All	(2) Treated	(3) Control	(4) Treated only
Recommended	0.012	0.011	0.011	0.008
	(0.012)	(0.014)	(0.024)	(0.009)
N (in BW)	257,015	214,917	78,876	176,898
Bandwidth h	0.124	0.136	0.154	0.118

Notes: Bias-corrected fuzzy RD estimates with robust standard errors. Column (4) restricts to treated posts and uses recruiter shortlisting as the outcome. A χ^2 test fails to reject equality of the treatment and control RD estimates ($p = 0.995$). SEs clustered by post. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

OLS and selection-corrected estimates. Tables A.13–A.14 report OLS, Heckit, and MLE estimates for the relationship between the recommendation score, recommendation status, and outcomes.

Table A.13: Effect of Recommendation: Hiring

	Score		OLS		ITT		2SLS	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Controls	Base	+Tier	Base	+Tier	Base	+Tier	Base	+Tier
s_j	0.008***	0.008***						
	(0.000)	(0.000)						
Recommended			0.012***	0.012***	0.013***	0.012***	0.012***	0.012***
			(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Shortlisted†	0.010***	0.010***	0.011***	0.011***	0.000	0.000	0.013	0.012
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.014)	(0.014)
N	1,526,791	1,526,791	1,526,791	1,526,791	1,526,791	1,526,791	1,526,791	1,526,791

Notes: Outcome: hired (0/1). Columns (1)–(2) report OLS with the continuous score s_j. Columns (3)–(4) report OLS with the recommendation indicator. Columns (5)–(6) report OLS intent-to-treat. Columns (7)–(8) instrument shortlisting with treatment assignment (2SLS). All specifications include log(applications) and has-hourly-rate controls. SEs clustered by post. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Table A.14: Wagebill: OLS, Heckit, MLE

	OLS		Heckit		MLE	
	(1)	(2)	(3)	(4)	(5)	(6)
Controls	Base	+Tier	Base	+Tier	Base	+Tier
s_j	-41.374***	-27.921*	-48.824**	-35.909*	-56.302***	-30.388
	(12.350)	(12.215)	(15.201)	(15.201)	(16.083)	(16.165)
IMR (λ)			118.658	121.816		
			(75.441)	(75.062)		
$\tanh^{-1}(\hat{\rho})$					-0.034	-0.022
					(0.045)	(0.045)
N	6,680	6,680	245,178	245,178	245,178	245,178

Notes: Outcome: 30-day wagebill. Columns (1)–(2) estimated by OLS among hired workers. Columns (3)–(4) estimated by Heckman two-step (Heckit) with an inverse Mills ratio correction for selection into hiring. Columns (5)–(6) estimated by joint maximum likelihood (MLE) of the selection and outcome equations (Lang, 2010). SEs clustered by post (OLS) or from the model. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Table A.15: Effect of Score and Recommendation: Shortlisting

	Score		Reduced Form	
	(1)	(2)	(3)	(4)
Controls	Base	+Tier	Base	+Tier
s_j	0.015***	0.016***		
	(0.000)	(0.000)		
Recommended			0.031***	0.032***
			(0.000)	(0.000)
N	1,106,091	1,106,091	1,106,091	1,106,091

Notes: Outcome: shortlisted by recruiter (treatment arm only; employer-shortlisted applications excluded). All columns estimated by OLS. s_j standardized within has-rate groups. All specifications include log(applications) and has-hourly-rate controls. SEs clustered by post. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

A.15.1 RD Robustness

We verify that the RD results are robust to alternative threshold definitions and inference procedures.

Threshold definition 2: midpoint-drop. The first alternative defines the threshold as the midpoint between the minimum recommended score and the maximum non-recommended score *below* the minimum recommended score, dropping posts where these coincide. Figures A.8–A.9 show binned scatter plots and

Tables A.16–A.18 report regression estimates. The hiring effect is positive and statistically significant in the treated arm. The shortlisting effect is statistically significant.

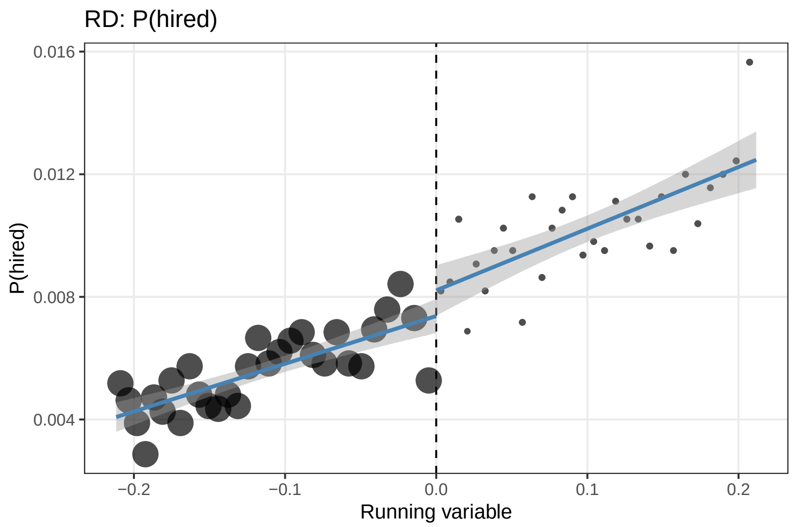


Figure A.8: RD: hiring probability (threshold definition 2)

Notes: Threshold definition 2 uses the midpoint between the minimum recommended score and the maximum non-recommended score below it, dropping posts where these coincide.

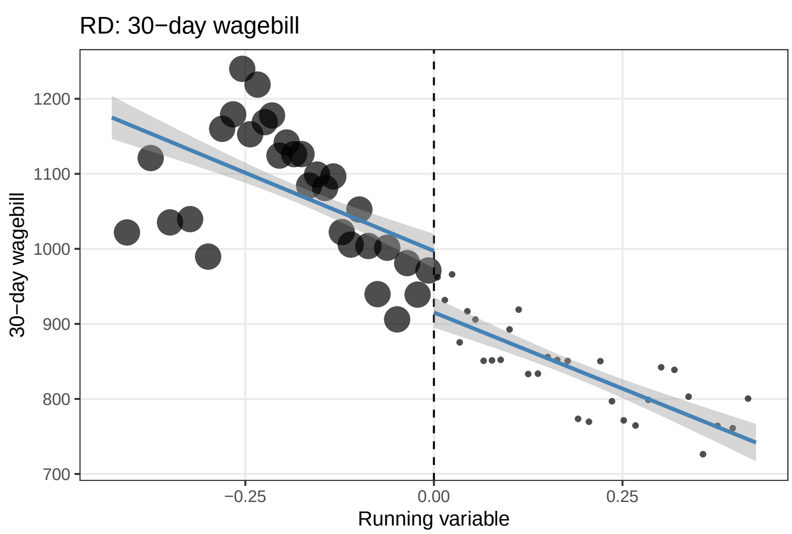


Figure A.9: RD: 30-day wagebill (threshold definition 2)

Notes: Threshold definition 2 uses the midpoint between the minimum recommended score and the maximum non-recommended score below it, dropping posts where these coincide. Post-level data (at most one hire per post).

Table A.16: Fuzzy RD: Hiring (threshold definition 2)

	RD (bias-corrected, robust SE)			2SLS w/ shortlisting	
	(1) All	(2) Treated	(3) Control	(4) Base	(5) +Tier
Recommended	0.001	0.002*	-0.003	0.001	0.001
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Any shortlisting				0.007	0.006
				(0.021)	(0.021)
N (in BW)	529,300	383,267	116,059	529,300	529,300
Bandwidth h	0.212	0.204	0.191	0.212	0.212

Notes: Outcome: hired (0/1). Columns (1)–(3) report bias-corrected fuzzy RD estimates with robust standard errors. Columns (4)–(5) report 2SLS estimates instrumenting recommendation and shortlisting with the threshold indicator and treatment assignment within the RD bandwidth. A χ^2 test rejects equality of the treatment and control RD estimates ($p = 0.023$). SEs clustered by post. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Table A.17: Fuzzy RD: Wagebill (threshold definition 2)

	RD (bias-corrected, robust SE)			2SLS w/ shortlisting	
	(1) All	(2) Treated	(3) Control	(4) Base	(5) +Tier
Recommended	131.732	-30.900	381.660	50.087	37.062
	(212.113)	(245.399)	(486.223)	(210.559)	(213.921)
Any shortlisting				-1556.873	-1700.620
				(2020.336)	(2010.043)
N (in BW)	2,774	1,965	621	2,774	2,774
Bandwidth h	0.427	0.408	0.310	0.427	0.427

Notes: Outcome: 30-day wagebill. Post-level data (at most one hire per post). Columns (1)–(3) report bias-corrected fuzzy RD estimates with robust standard errors. Columns (4)–(5) report 2SLS estimates instrumenting recommendation and shortlisting with the threshold indicator and treatment assignment within the RD bandwidth. A χ^2 test fails to reject equality of the treatment and control RD estimates ($p = 0.449$). SEs clustered by post. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Table A.18: RD: Recommendation and Shortlisting (threshold definition 2)

	Any shortlisting			Shortlisted by recruiter
	(1) All	(2) Treated	(3) Control	(4) Treated only
Recommended	0.031***	0.042***	0.007	0.033***
	(0.003)	(0.003)	(0.005)	(0.002)
N (in BW)	433,468	231,834	169,782	172,669
Bandwidth h	0.176	0.128	0.272	0.100

Notes: Bias-corrected fuzzy RD estimates with robust standard errors. Column (4) restricts to treated posts and uses recruiter shortlisting as the outcome. A χ^2 test rejects equality of the treatment and control RD estimates ($p = 0.000$). SEs clustered by post. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Threshold definition 3: minimum recommended. The second alternative uses the minimum recommended score directly as the threshold, dropping posts where the minimum recommended and maximum non-recommended scores are adjacent. Figures A.10–A.11 show binned scatter plots and Tables A.19–A.21 report regression estimates. The hiring effect is positive and statistically significant in the pooled sample. The shortlisting effect is also statistically significant.

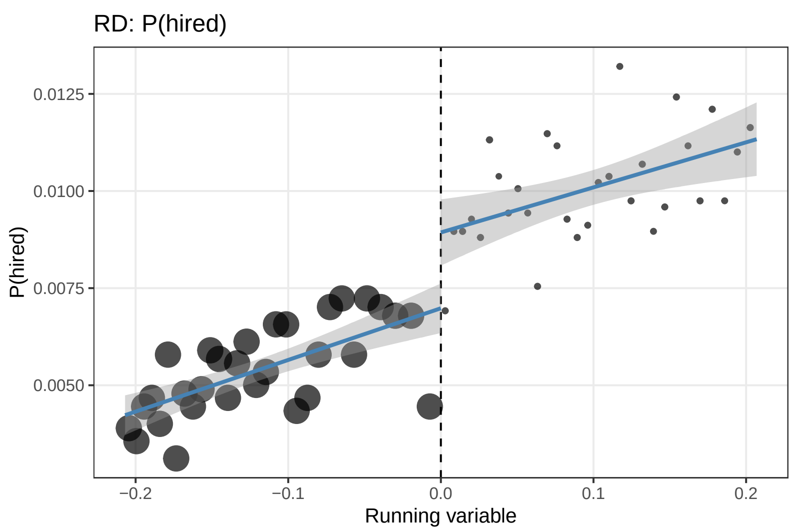


Figure A.10: RD: hiring probability (threshold definition 3)

Notes: Threshold definition 3 uses the minimum recommended score directly, dropping posts where the minimum recommended and maximum non-recommended scores are adjacent.

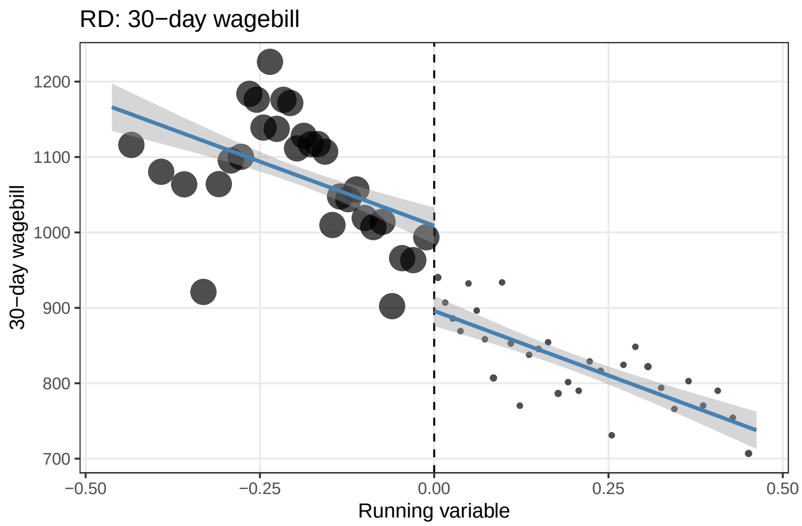


Figure A.11: RD: 30-day wagebill (threshold definition 3)

Notes: Threshold definition 3 uses the minimum recommended score directly, dropping posts where the minimum recommended and maximum non-recommended scores are adjacent. Post-level data (at most one hire per post).

Table A.19: Fuzzy RD: Hiring (threshold definition 3)

	RD (bias-corrected, robust SE)			2SLS w/ shortlisting	
	(1) All	(2) Treated	(3) Control	(4) Base	(5) +Tier
Recommended	0.003*	0.004**	-0.003	0.003*	0.003*
	(0.001)	(0.001)	(0.003)	(0.001)	(0.002)
Any shortlisting				0.002	0.001
				(0.026)	(0.025)
N (in BW)	460,330	408,697	114,119	460,330	460,330
Bandwidth h	0.207	0.239	0.208	0.207	0.207

Notes: Outcome: hired (0/1). Columns (1)–(3) report bias-corrected fuzzy RD estimates with robust standard errors. Columns (4)–(5) report 2SLS estimates instrumenting recommendation and shortlisting with the threshold indicator and treatment assignment within the RD bandwidth. A χ^2 test rejects equality of the treatment and control RD estimates ($p = 0.015$). SEs clustered by post. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Table A.20: Fuzzy RD: Wagebill (threshold definition 3)

	RD (bias-corrected, robust SE)			2SLS w/ shortlisting	
	(1) All	(2) Treated	(3) Control	(4) Base	(5) +Tier
Recommended	-98.534	-238.182	190.748	-67.646	-56.995
	(241.693)	(286.096)	(452.766)	(225.444)	(227.549)
Any shortlisting				-152.127	-256.465
				(1074.049)	(1057.018)
N (in BW)	2,692	1,901	806	2,692	2,692
Bandwidth h	0.462	0.440	0.542	0.462	0.462

Notes: Outcome: 30-day wagebill. Post-level data (at most one hire per post). Columns (1)–(3) report bias-corrected fuzzy RD estimates with robust standard errors. Columns (4)–(5) report 2SLS estimates instrumenting recommendation and shortlisting with the threshold indicator and treatment assignment within the RD bandwidth. A χ^2 test fails to reject equality of the treatment and control RD estimates ($p = 0.423$). SEs clustered by post. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Table A.21: RD: Recommendation and Shortlisting (threshold definition 3)

	Any shortlisting			Shortlisted by recruiter
	(1) All	(2) Treated	(3) Control	(4) Treated only
Recommended	0.017***	0.020***	0.003	0.019***
	(0.003)	(0.003)	(0.007)	(0.002)
N (in BW)	788,323	669,314	154,193	391,926
Bandwidth h	0.377	0.478	0.270	0.238

Notes: Bias-corrected fuzzy RD estimates with robust standard errors. Column (4) restricts to treated posts and uses recruiter shortlisting as the outcome. A χ^2 test rejects equality of the treatment and control RD estimates ($p = 0.016$). SEs clustered by post. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Bias-aware inference. Tables A.22–A.24 report results using the RDHonest estimator (Armstrong and Kolesár, 2020), which provides honest confidence intervals that account for potential bias from misspecification of the conditional expectation function. The point estimates and inference are consistent with the main rdrobust results.

Table A.22: RDHonest Fuzzy RD: Hiring

	(1) All	(2) Treated	(3) Control
Recommended	0.004***	0.004*	0.002
	(0.001)	(0.002)	(0.003)
95% CI	[0.002, 0.007]	[-0.000, 0.008]	[-0.003, 0.008]
Max. bias	0.001	0.001	0.001
Eff. N	889,780	536,310	207,941
Bandwidth h	0.592	0.380	0.513

Notes: Outcome: hired (0/1). Fuzzy RD via RDHonest (Armstrong & Kolesár, 2018). Triangular kernel, MSE-optimal bandwidth, rule-of-thumb M. EHW SEs clustered by post. CIs are honest (bias-aware). A χ^2 test fails to reject equality of the treatment and control estimates ($p = 0.626$). Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Table A.23: RDHonest Fuzzy RD: Wagebill

	(1) All	(2) Treated	(3) Control
Recommended	159.382	121.191	572.554
	(231.809)	(231.423)	(1659.336)
95% CI	[-367.920, 686.683]	[-398.353, 640.736]	[-3245.992, 4391.100]
Max. bias	141.545	133.420	1062.565
Eff. N	2,851	2,123	492
Bandwidth h	0.564	0.574	0.282

Notes: Outcome: 30-day wagebill. Post-level data (at most one hire per post). Fuzzy RD via RDHonest (Armstrong & Kolesár, 2018). Triangular kernel, MSE-optimal bandwidth, rule-of-thumb M. EHW SEs clustered by post. CIs are honest (bias-aware). A χ^2 test fails to reject equality of the treatment and control estimates ($p = 0.788$). Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Table A.24: RDHonest: Recommendation and Shortlisting

	Any shortlisting			Shortlisted by recruiter
	(1) All	(2) Treated	(3) Control	(4) Treated only
Recommended	0.007	0.007	0.008	0.004
	(0.006)	(0.008)	(0.005)	(0.005)
95% CI	[-0.006, 0.020]	[-0.010, 0.025]	[-0.002, 0.018]	[-0.007, 0.014]
Eff. N	499,386	322,486	209,694	330,954
Bandwidth h	0.269	0.236	0.524	0.250

Notes: Fuzzy RD via RDHonest (Armstrong & Kolesár, 2018). Triangular kernel, MSE-optimal bandwidth, rule-of-thumb M. Column (4) restricts to treated posts and uses recruiter shortlisting as the outcome. CIs are honest (bias-aware). A χ^2 test fails to reject equality of the treatment and control estimates ($p = 0.951$). EHW SEs clustered by post. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

A.16 RD Heterogeneity by Threshold and Pool Size

Table A.25 shows RD estimates for hiring by application-count quartile. Recommendation has a significant positive effect on hiring in the lowest quartile (0.054, $p < 0.01$) and no detectable effect in the upper three quartiles. The pattern is consistent across threshold definitions: under the alternative definition, the Q1 hiring effect remains positive and significant (0.008, $p < 0.01$; Table A.34). Table A.26 shows a parallel pattern for shortlisting, where Q1 and Q2 are both significant under the alternative definition (Table A.35). The wagebill estimates in Table A.27 are noisy but directionally consistent: positive in Q1 under the baseline definition and Q1–Q2 under the alternative, negative in Q3–Q4 under both.

RD estimates for hiring are near zero across all threshold quartiles (Tables A.28 and A.31). Shortlisting responds to recommendation at all threshold levels under the alternative definition, with a monotonically decreasing effect from Q1 to Q4 (Table A.32), but this does not translate into differential hiring or wagebill effects.

Table A.25: RD by Application Count Quartile: hired (0/1)

App-count quartile	RD estimate	N (in BW)	BW h
Q1 (low)	0.054**	39,172	0.073
	(0.020)		
Q2	0.019	40,000	0.076
	(0.013)		
Q3	-0.003	52,103	0.103
	(0.006)		
Q4 (high)	-0.000	98,515	0.188
	(0.008)		

Notes: Outcome: hired (0/1). Fuzzy RD bias-corrected estimates; robust SEs in parentheses. SEs clustered by post. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Table A.26: RD by Application Count Quartile: any shortlisting (0/1)

App-count quartile	RD estimate	N (in BW)	BW h
Q1 (low)	0.051	39,876	0.074
	(0.048)		
Q2	-0.030	41,220	0.078
	(0.034)		
Q3	-0.016	46,158	0.092
	(0.021)		
Q4 (high)	0.050*	50,073	0.098
	(0.023)		

Notes: Outcome: any shortlisting (0/1). Fuzzy RD bias-corrected estimates; robust SEs in parentheses. SEs clustered by post. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Table A.27: RD by Application Count Quartile: 30-day wagebill

App-count quartile	RD estimate	N (in BW)	BW h
Q1 (low)	1236.995**	669	0.406
	(436.021)		
Q2	130.058	834	0.532
	(748.005)		
Q3	-888.372	595	0.299
	(1009.743)		
Q4 (high)	-261.848	828	0.561
	(725.943)		

Notes: Outcome: 30-day wagebill. Fuzzy RD bias-corrected estimates; robust SEs in parentheses. SEs clustered by post. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Table A.28: RD by Threshold Quartile: hired (0/1)

Threshold quartile	RD estimate	N (in BW)	BW h
Q1 (low)	0.007	27,208	0.066
	(0.005)		
Q2	0.022	29,277	0.049
	(0.044)		
Q3	0.638	62,846	0.100
	(0.374)		
Q4 (high)	-0.000	157,278	0.376
	(0.003)		

Notes: Outcome: hired (0/1). Fuzzy RD bias-corrected estimates; robust SEs in parentheses. SEs clustered by post. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Table A.29: RD by Threshold Quartile: any shortlisting (0/1)

Threshold quartile	RD estimate	N (in BW)	BW h
Q1 (low)	-0.040	22,556	0.055
	(0.022)		
Q2	0.006	31,942	0.053
	(0.120)		
Q3	48.410	56,332	0.090
	(46.987)		
Q4 (high)	0.008	230,540	0.600
	(0.004)		

Notes: Outcome: any shortlisting (0/1). Fuzzy RD bias-corrected estimates; robust SEs in parentheses. SEs clustered by post. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Table A.30: RD by Threshold Quartile: 30-day wagebill

Threshold quartile	RD estimate	N (in BW)	BW h
Q1 (low)	-395.622	471	0.301
	(700.824)		
Q2	80.897	743	0.289
	(1381.996)		
Q3	1103.406*	712	0.379
	(443.022)		
Q4 (high)	97.667	747	0.597
	(289.940)		

Notes: Outcome: 30-day wagebill. Fuzzy RD bias-corrected estimates; robust SEs in parentheses. SEs clustered by post. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Table A.31: RD by Threshold Quartile: hired (0/1) (threshold definition 3)

Threshold quartile	RD estimate	N (in BW)	BW h
Q1 (low)	0.003 (0.004)	183,159	0.238
Q2	0.003 (0.003)	155,199	0.250
Q3	0.002 (0.002)	130,302	0.282
Q4 (high)	-0.001 (0.002)	162,740	0.374

Notes: Outcome: hired (0/1). Fuzzy RD bias-corrected estimates; robust SEs in parentheses. SEs clustered by post. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Table A.32: RD by Threshold Quartile: any shortlisting (0/1) (threshold definition 3)

Threshold quartile	RD estimate	N (in BW)	BW h
Q1 (low)	0.040*** (0.012)	146,111	0.200
Q2	0.025** (0.008)	157,998	0.254
Q3	0.027*** (0.007)	134,635	0.290
Q4 (high)	0.010*** (0.003)	245,160	0.632

Notes: Outcome: any shortlisting (0/1). Fuzzy RD bias-corrected estimates; robust SEs in parentheses. SEs clustered by post. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Table A.33: RD by Threshold Quartile: 30-day wagebill (threshold definition 3)

Threshold quartile	RD estimate	N (in BW)	BW h
Q1 (low)	-252.044 (1076.255)	694	0.483
Q2	383.476 (595.655)	604	0.315
Q3	-846.715 (672.829)	508	0.332
Q4 (high)	227.669 (205.352)	684	0.553

Notes: Outcome: 30-day wagebill. Fuzzy RD bias-corrected estimates; robust SEs in parentheses. SEs clustered by post. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Table A.34: RD by Application Count Quartile: hired (0/1) (threshold definition 3)

App-count quartile	RD estimate	N (in BW)	BW h
Q1 (low)	0.008** (0.003)	162,143	0.316
Q2	0.003 (0.002)	162,157	0.291
Q3	-0.003 (0.002)	108,680	0.196
Q4 (high)	-0.002 (0.002)	128,153	0.204

Notes: Outcome: hired (0/1). Fuzzy RD bias-corrected estimates; robust SEs in parentheses. SEs clustered by post. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Table A.35: RD by Application Count Quartile: any shortlisting (0/1) (threshold definition 3)

App-count quartile	RD estimate	N (in BW)	BW h
Q1 (low)	0.024*** (0.005)	216,304	0.464
Q2	0.025*** (0.006)	223,855	0.495
Q3	0.006 (0.006)	159,218	0.274
Q4 (high)	0.010 (0.007)	167,517	0.258

Notes: Outcome: any shortlisting (0/1). Fuzzy RD bias-corrected estimates; robust SEs in parentheses. SEs clustered by post. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

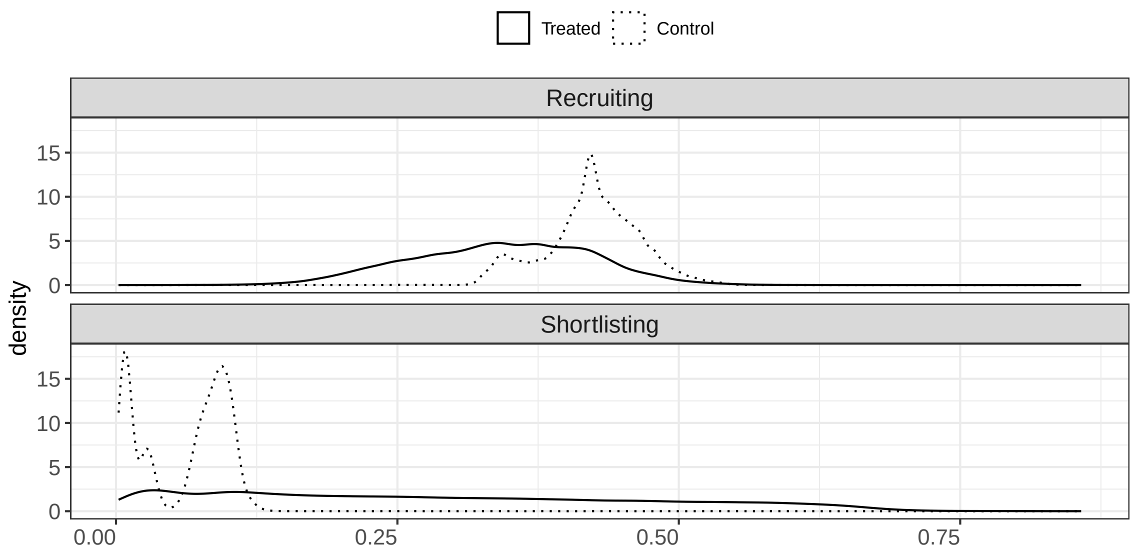
Table A.36: RD by Application Count Quartile: 30-day wagebill (threshold definition 3)

App-count quartile	RD estimate	N (in BW)	BW h
Q1 (low)	344.698 (235.765)	856	0.619
Q2	568.863 (474.773)	618	0.437
Q3	-637.942 (628.563)	635	0.414
Q4 (high)	-1103.326 (867.213)	551	0.359

Notes: Outcome: 30-day wagebill. Fuzzy RD bias-corrected estimates; robust SEs in parentheses. SEs clustered by post. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

A.17 Propensity Score Distributions

Figure A.12: Distribution of propensity scores for receiving recruiting or shortlisting, by treatment assignment



Notes: This plot shows the distribution of estimated propensity scores for recruiting and shortlisting, for both the treatment and control group. The propensity scores use job category, post date, expertise tier, and skill count as predictors. The near-identical distributions across treatment and control confirm that the propensity model captures pre-treatment characteristics.

A.18 Recruited vs. Organic Applicants

We compare Recruiter-recruited applicants to organic applicants on observable platform characteristics. Given the null treatment effect on hiring, the recruited applicants could in principle be a random draw from the distribution, negatively selected, or positively selected on dimensions that fail to translate into job-specific fit.

In Table A.37 we compare Recruiter-invited applications to all other applications to jobs in the experimental sample. Column (1) shows that recruited workers bid roughly a dollar more on the application than non-recruited workers, despite Column (2) showing that their posted profile rate is about \$3 *lower*. Columns (4) and (5) show recruited workers have substantially more hours worked on the platform and more prior employer invitations. Column (3) shows that recruited workers have *lower* cumulative platform earnings, despite working more hours.

The pattern is mixed rather than uniformly positive. Recruited workers look more active and more in-demand (more hours, more prior invites) but are cheaper on a per-job basis (lower posted profile rate, lower lifetime earnings). This is consistent with recruiters favoring engaged, available workers who have not yet been fully priced into the market.

The fact that these workers do not increase the hiring rate suggests the dimensions on which they are positively selected (engagement, availability, prior invitations) do not translate into better job-specific fit, while the dimensions on which they are negatively selected (price, cumulative earnings) may signal lower quality on margins the algorithmic score does not capture.

A.19 Multiple Hypothesis Testing

Tables A.38 and A.39 report multiple-hypothesis-corrected p-values for the subgroup hiring effects shown in Figures 6 and 7. Corrections use bootstrap stepdown (List et al., 2019), Bonferroni, and Benjamini–Hochberg FDR procedures (Benjamini and Hochberg, 1995).

Table A.9: Selection-Corrected Treatment Effects on 30-Day Wagebill

	OLS	Heckit	MLE
Panel: All Contracts			
Treatment	-59.96***	-55.31***	-37.29*
	(15.36)	(15.51)	(15.71)
IMR (λ)		-561.90***	
		(33.10)	
$\tanh^{-1}(\hat{\rho})$			-0.115***
			(0.018)
N (apps / hired)	35893	2224886 / 35893	2224886 / 35893
Panel: Hourly Jobs			
Treatment	-70.05***	-64.99***	-40.01*
	(17.18)	(17.36)	(17.53)
IMR (λ)		-556.53***	
		(38.93)	
$\tanh^{-1}(\hat{\rho})$			-0.134***
			(0.020)
N (apps / hired)	27551	1486721 / 27551	1486721 / 27551
Panel: Project Based Jobs			
Treatment	-18.71	-17.79	-10.91
	(33.75)	(33.94)	(33.76)
IMR (λ)		-551.77***	
		(61.53)	
$\tanh^{-1}(\hat{\rho})$			-0.309***
			(0.031)
N (apps / hired)	8337	738132 / 8337	738132 / 8337

Notes: Unit of observation: applications. For posts with multiple hires, wagebill is split equally across hired applications. Selection equation includes treatment, category group, expertise tier, and $\log(\text{num apps})$. Outcome equation includes the same controls except $\log(\text{num apps})$ (the exclusion restriction). Heckit is the Heckman two-step estimator with an inverse Mills ratio (IMR, λ) correction in the outcome equation. MLE is full-information maximum likelihood; $\tanh^{-1}(\hat{\rho})$ is the Fisher-transformed estimate of the selection–outcome error correlation. Standard errors in parentheses. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Table A.37: Comparison of workers recruited by Recruiter/Shortlisters to other applications of treated job posts

	Dependent variable:				
	Wage bid	Profile rate	Total earned	Hours worked	Previous invites
	(1)	(2)	(3)	(4)	(5)
Recruited worker	1.14^{***}	-2.72^{***}	$-15,270.85^{***}$	731.04^{***}	154.45^{***}
	(0.04)	(0.04)	(376.03)	(6.47)	(1.31)
Constant	28.15^{***}	34.13^{***}	$93,070.73^{***}$	$2,148.70^{***}$	330.22^{***}
	(0.02)	(0.02)	(210.98)	(2.90)	(0.59)
Observations	2,091,551	2,865,777	2,865,777	2,865,777	2,865,777
Adjusted R^2	0.0003	0.001	0.0004	0.004	0.005

Notes: The first column reports the wage bids for recruited applications, versus all other received applications. It uses all applications to jobs in the treated group of the experimental sample. Columns (2) - (5) reports the work history on the platform of recruited workers as compared to workers who applied without being recruited. Heteroskedasticity-robust (Eicker–Huber–White) standard errors are used in all specifications. Significance indicators: $p \leqslant 0.05$: *, $p \leqslant 0.01$: **, and $p \leqslant .001$: ***.

Table A.38: Multiple Hypothesis Testing: Treatment Effect on Hiring by Category Group

			p-values			
Subgroup	N	DI	Unadj.	Bootstrap	Bonf.	FDR (BH)
Data Science & Analytics	3,233	0.0082	0.6768	1.0000	1.0000	0.9914
Web, Mobile & Software Dev	46,326	-0.0020	0.6574	0.9998	1.0000	0.9914
Accounting & Consulting	4,297	-0.0002	0.9914	1.0000	1.0000	0.9914
Admin Support	5,756	0.0119	0.4268	0.9950	1.0000	0.9914
Design & Creative	4,030	0.0014	0.9396	1.0000	1.0000	0.9914
Customer Service	9,883	0.0300	0.0048***	0.0490**	0.0480**	0.0480**
IT & Networking	6,099	0.0151	0.2486	0.9236	1.0000	0.8287
Writing	1,042	0.0782	0.0266**	0.2158	0.2660	0.1330
Engineering & Architecture	1,290	0.0089	0.7726	1.0000	1.0000	0.9914
Sales & Marketing	996	0.0081	0.8202	1.0000	1.0000	0.9914

Note:
Bootstrap stepdown procedure follows List, Shaikh, and Xu (2019) Algorithm 3.1 with 1,000 bootstrap replications. DI is the difference in means (treatment minus control). *, **, *** indicate significance at 10%, 5%, and 1% levels.

Table A.39: Multiple Hypothesis Testing: Treatment Effect on Hiring by Expertise Tier

			p-values			
Subgroup	N	DI	Unadj.	Bootstrap	Bonf.	FDR (BH)
Expert/Expensive	39,937	0.0047	0.3728	0.6048	1.0000	0.5592
Intermediate	39,220	0.0018	0.7362	0.9316	1.0000	0.7362
Cheap/Inexperienced	3,839	0.0540	0.0016***	0.0054***	0.0048***	0.0048***

Note:
Bootstrap stepdown procedure follows List, Shaikh, and Xu (2019) Algorithm 3.1 with 1,000 bootstrap replications. DI is the difference in means (treatment minus control). *, **, *** indicate significance at 10%, 5%, and 1% levels.