The Diminishing Returns to Human Recruiting in Online Labor Markets

Emma Wiles* Boston University
John Horton† MIT & NBER
John Fallon ‡ Brown University

September 10, 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 employers. The recruiters increased the search and screening intensity of treated job posts: treated job posts had more recruiting and more applicant screening done on their behalf, resulting in 5% more applicants. Employers of treated job posts redirected their own screening efforts toward recruiter-sourced candidates and away from applicants who found the post on their own. Despite the increase in search intensity, treated employers were no more likely to hire than employers of control posts with access only to algorithmic tools. Match quality appears, at best, no better for the hires that resulted from the treated job posts, and treated employers were less likely to return to the platform to post a second vacancy after the experiment. We develop a model of delegated recruiting in which recruiters and employers rely on a common noisy signal; it predicts that recruiting adds value when the recruiter brings information the algorithm lacks and can subtract value when the two parties' assessments are correlated. Consistent with the model's screening 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 seminar participants at University of Minnesota's Carlson School. We would like to thank Ayush Gupta, Kevin Lang, Johannes Schmieder, and Edward Wiles for helpful feedback.

1 Introduction

Hiring is costly. Firms invest substantially in interviewing and screening (Barron et al., 1985), and 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 (Bull et al., 1987), as recruiters can re-use information from prior searches (Howitt and McAfee, 1987). Intermediaries offer an immediate increase in hiring intensity (Gavazza et al., 2018), and vacancies with more intense searches are more likely to result in a hire (Davis et al., 2012, 2013). Black et al. (2024) document that firms actively search for talent and that outbound recruiting yields different candidates than inbound applications. Yet there is little causal evidence on whether intermediaries improve hiring outcomes for employers or merely increase search activity.

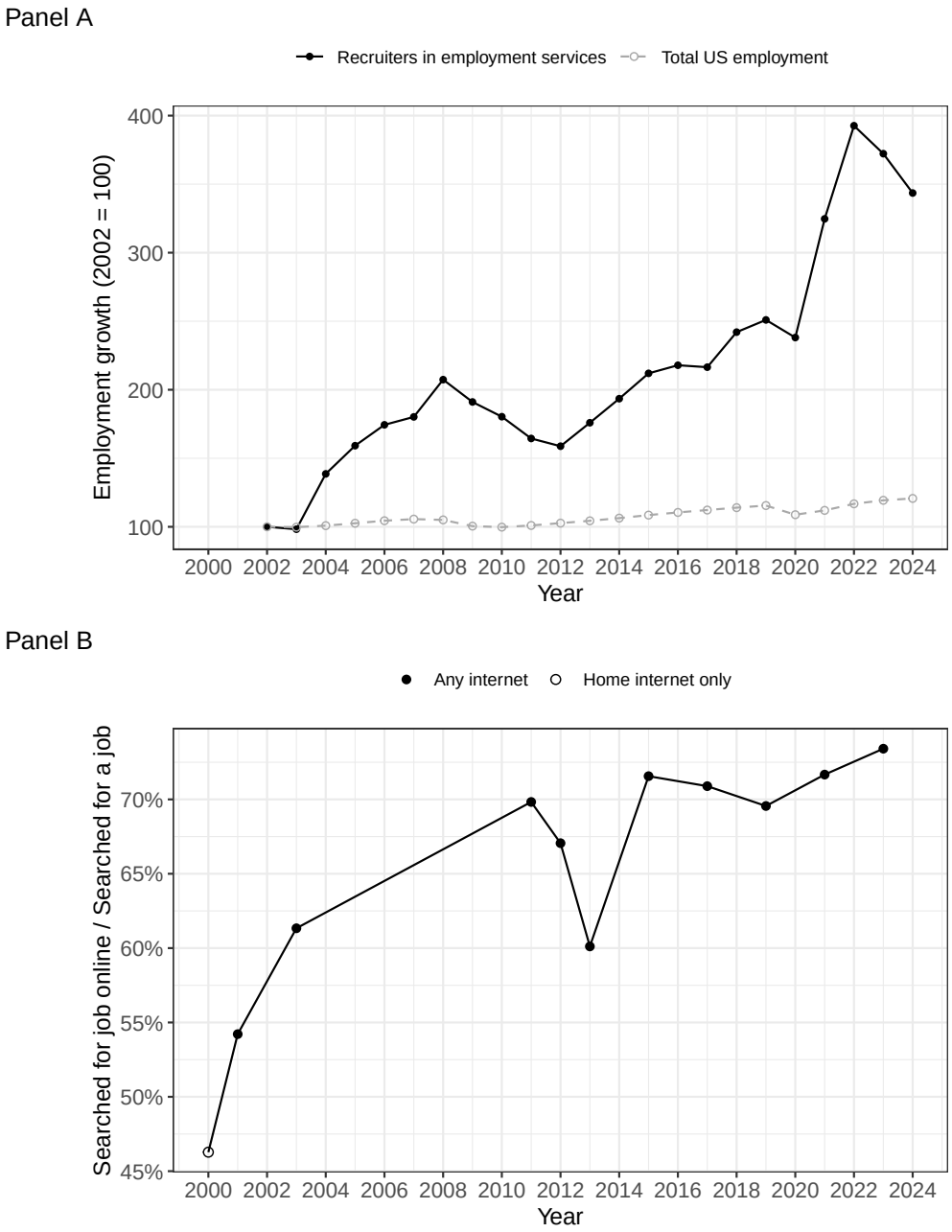
For decades, the recruiting industry grew faster than the rest of the labor market (Figure 1, Panel A). Over the same period, an increasing majority of jobseekers searched for work online (Figure 1, Panel B), and much of the hiring process was mediated by job boards (Carnevale et al., 2014). In addition, applicant tracking systems routinely perform search and screening functions (ranking applicants, recommending matches) once central to human intermediaries. In recent years, the recruiting industry has begun contracting, and AI powered recruiting tools have taken a more central role in the matching process. In this paper, we study whether and under what conditions hiring intermediaries add value to employers who already have access to algorithmic tools.

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 work histories and ratings (Pallais, 2014; Stanton and Thomas, 2016).[2] The platform randomly assigned human hiring consultants to 83,017 job posts, with 25% serving as a control. Employers of control posts had access to the platform's existing tools: worker search, algorithmic recommendations, detailed profiles with ratings and work history, and standard matching features. Because employers could appear in the sample more than once, our preferred specification is on the sample of only each employer's first job post which appears in the experiment, giving us 52,471 job posts.

The intervention generated a strong first stage. The hiring consultants recruited additional applicants and screened the applicant pool, highlighting candidates they considered to be a good fit. Employers of treated job posts received 5% more applications under intent-to-treat (7.5% instrumented by actual treatment receipt). The increase in applications was driven entirely by the hiring consultant's recruiting activity; employers of treated posts did not increase their own recruiting efforts. However, employers did increase their screening effort: treated employers shortlisted more applicants themselves and conducted more interviews. It also redirected employer screening effort away from applicants who applied for the job organically and towards applicants found by the recruiter.

[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.

Figure 1: Human recruiting expanded relative to the rest of the economy even as job search moved online



Notes: Panel A compares growth for recruiters in employment services alongside total US nonfarm employment using data from the OES and indexed to 2002. The recruiter series is defined as "human resource specialists" occupation (SOC code 13-1071, reported under transitional code 13-1078 in 2010 and 2011), whose primary task is to "recruit, screen, interview, or place individuals within an organization." To focus the series only on outsourced and not in house recruiters, we narrow this to only workers in the employment services industry (NAICS code 56-1300). Panel A extends a figure in Cowgill and Perkowski (2024). Panel B plots the share of unemployed job seekers who report using the internet to search for work, from IPUMS CPS Computer and Internet Use supplements; the 1998 supplement measured home internet use only and is shown with an open circle.

Despite this increased search and screening intensity, employers assigned hiring consultants were no more likely to make a hire. About 29% of job posts led to a hire in both groups, with an insignificant treatment effect of 0.006. While it is possible the hiring consultants could have had no impact on hiring probability but still found a better candidate for the employer to hire, we find no evidence in favor of this. Hours worked and wage rates were similar in both groups, and the ratings the workers and employers give each other at the end of each contract were slightly lower in the treatment group, although imprecisely estimated. In some specifications, treated employers spent less on their hires in the first 30 days after posting the job but this disappears if you look at wagebill over the first 90 days. The most compelling evidence that the program did not improve the employers' experience hiring is that after the experiment treated employers are significantly less likely to return to the platform to post another vacancy. While match quality did not improve, this is not because recruiters were recruiting low quality applicants. Recruited applicants are positively selected on most observable platform characteristics (engagement, prior employer interest) but negatively selected on wage; and the workers treated employers actually hired were observably identical to control hires on job-success score, earnings, tenure, and experience. This suggests that recruiters were able to identify applicants with desirable observable characteristics, but not applicants who were particularly well suited to the employer's specific needs.

The platform's algorithmic recommendations already do much of the work of recommending workers and applicants. We use a regression discontinuity design exploiting the discrete recommendation label the algorithm shows to employers for applicants just above a threshold to show that recommended applicants are significantly more likely to be hired, echoing Horton (2017). This response is driven by the employers least able to evaluate candidates on their own: small firms, first-time and infrequent hirers, and newly registered accounts, as well as job posts with thin applicant pools. Where the employer has the experience or the candidates to screen well herself, the recommendation moves nothing, consistent with employers leaning on the algorithm precisely when their own judgment is weakest.

The null effect may not be uniform across all job types. Hiring rises for the lowest expertise tier (cheap, inexperienced workers), the only subgroup anywhere in our heterogeneity analysis that is significant before adjusting for multiple testing, though it does not survive that adjustment (Section 6). It is where the model predicts the most room to help: workers with thin platform histories are precisely those the algorithmic score reads least well. We report it as a hypothesis the design was not powered to settle rather than as a finding, and we treat the category-level splits, none of which is significant even unadjusted, the same way.

This program could not justify its cost. The upper bound of the 95% confidence interval on the hiring effect implies expected revenue per job post of about $6.26, which falls dramatically short of estimated program costs with estimated net loss per post of between $1.56 and $6.76. Some firms may prefer to not manage hiring internally and would be willing to pay for the convenience of an outside recruiter, even if it comes at some cost to match quality. However, as algorithms reduce the effort required to hire well, fewer firms will likely find this tradeoff worthwhile.

To our knowledge, this is the first field experiment testing whether hiring assistance helps employers.

Active labor market programs to assist jobseekers have been studied extensively (Card et al., 2018), and job search assistance helps workers find jobs through caseworkers (Schiprowski, 2020), online advice (Belot et al., 2019), and government programs (Card et al., 2010), even when spillovers to other workers are taken into account (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 (Pallais, 2014; Stanton and Thomas, 2016). On the other hand, employers post many vacancies and may develop a reputation faster or may learn how to hire faster.

Prior research has examined algorithmic decision aids and human recruiters as separate forms of hiring assistance. Often, when evaluating the same applicant pool, these tools match or outperform human judgment in hiring decisions (Cowgill, 2018), judicial settings (Kleinberg et al., 2018), and across many selection contexts (Kuncel et al., 2013). Hoffman et al. (2018) find that managers who override algorithmic test recommendations make worse hires on average. 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.

Hiring using an outside recruiter is a form of delegation. Theory suggests that delegation works best when the agent knows more than the principal (Dessein, 2002; Alonso and Matouschek, 2008), which also determines who holds real versus formal authority in firms (Aghion and Tirole, 1997), a condition we argue is increasingly difficult to satisfy as algorithmic improvements erode the information advantages that human intermediaries once held. In our case, 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, as in real-world recruiting services, where consultants adapt to each situation. Outside of recruiting, prior field experiments that encourage firms to delegate to outside professionals generally find at least short-run gains, including in management consulting (Bloom et al., 2013) and retail supervision (Friebel et al., 2022). 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.

Employers do not only receive applications from workers who find their vacancy organically. Outside of formal intermediaries like recruiters (Cowgill and Perkowski, 2024), employers can also hire workers from within their organization or hire workers from internal referrals. Prior work has found that internally referred workers are positively selected (Burks et al., 2015), hired at higher rates (Pallais and Sands, 2016), and better matched to their jobs (Brown et al., 2016). The leading explanation is that referrers hold private information about candidates that employers cannot directly observe: referrers know the candidates personally and may

have information that does not appear on the candidate's resume. External recruiters who evaluate the same profiles as the employer may have less independent information about candidates to complement the employer's information.

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, such as communication or reliability that do not show up in platform histories, and is not observed by either party ex ante (Jovanovic, 1979). This setting presents two challenges. First, the platform's score is noisy. Second, given that both recruiters and employers are responding to the same platform information, they may be susceptible to the same errors, leaving their assessments correlated.

The model yields three results. First, recruiting has a technically ambiguous, yet typically weakly positive effect on the likelihood of making a hire. More candidates raise the chance of finding an acceptable match, but the employer's acceptance threshold can adjust in response. At a calibration matching the control group's fill rate, the gain is 1.72 percentage points per post on which the recruiter adds candidates, within the confidence interval of our experimental estimate, and the threshold response is negligible. Second, when recruiter and employer signals are correlated, recruiting may harm quality. Candidates who clear the interview do so partly because of shared signal noise; once hired, the noise washes out and performance disappoints. How strong this force is depends on how correlated the two parties' assessments are. Third, as algorithms make more of a worker's skill observable from their 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 increased search and screening intensity. Section 4 reports the main results on hiring outcomes and a cost-benefit analysis of the program. 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 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 posting job openings (Horton, 2010), hosting worker profiles and handling payments (Horton, 2010), tracking reputation (Filippas et al., 2018; Horton et al., 2018), and designing the marketplace itself (Roth, 2018). Information frictions drive wage dispersion (Hornstein et al., 2011) and are central to

how these markets work; Rogerson et al. (2005) survey this literature. Visible work histories can improve workers' outcomes (Pallais, 2014), especially for workers from less developed countries (Agrawal et al., 2016), and platform reputation systems shape hiring as well (Benson et al., 2020). Platform design choices about what information to show shape both sides of the market (Fradkin, 2017). Even with this information, employers must decide which workers to consider and expend effort searching and screening. Platform interventions that direct employer attention toward particular workers can substantially affect recruiting and hiring (Horton, 2017).

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. Then the employer flags a desired "expertise tier", for workers who are either Cheap/Inexperienced, Intermediate, or Expert/Expensive. 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. Hourly contracts allow the employer to terminate or extend contracts at any time, so realized hours and wagebill track match quality directly for hourly jobs.

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. 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 pay close attention to past platform experience (Pallais, 2014) and refine their hiring over time (Leung, 2018), and generally look for signals to overcome information asymmetries (Stanton and Thomas, 2016). Wage setting on these platforms has features of both competitive (Chen and Horton, 2016) and monopsonistic markets (Dube et al., 2020).

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 employers make an offer, workers 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. At the end of each contract, the worker and the employer both provide feedback

and numerical ratings of their experience working together.

2.2 Hiring process

The hiring process on the platform follows the same steps as it does 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.1 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). They also see which applicants the platform's recommender system suggests is the best match for the job. 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. 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 September 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 vacancies (both hourly and fixed-price) over the experimental period requiring 30 or more hours of work per week for at least 3 months.

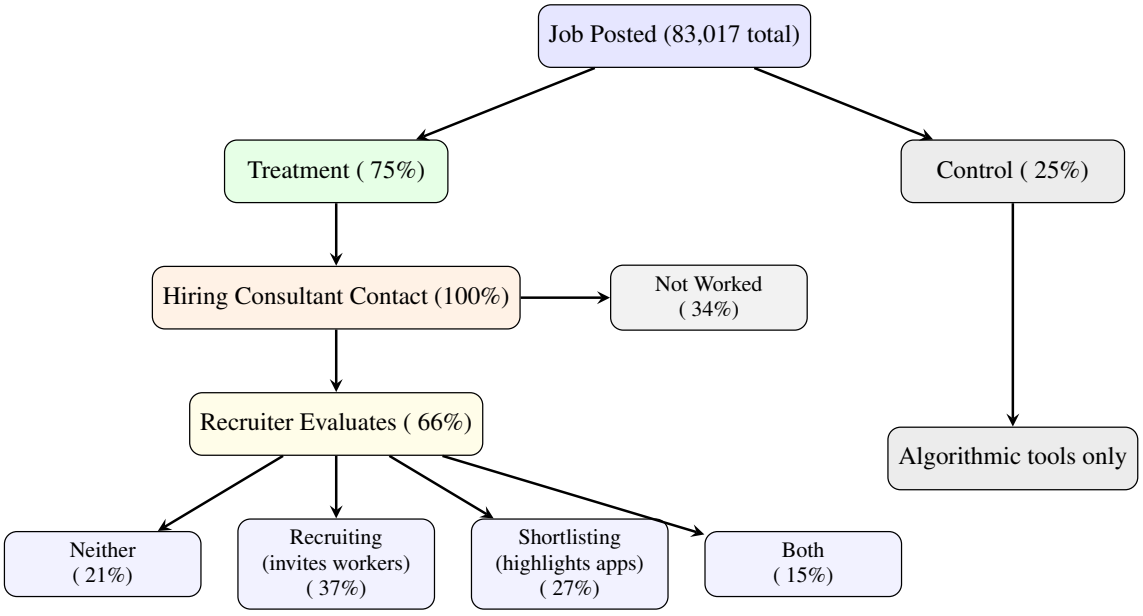
After the platform flagged a vacancy for the experiment, the resulting sample was approximately 75% treatment and 25% control.[3] The unit of randomization was the job post. As such, employers who posted multiple jobs over the course of the experiment could have had hiring assistance for one job but not another. The full sample contains 83,017 job posts. Because an employer's experience with its first experimental vacancy could affect whether it returns to the platform, our primary sample includes only each employer's first job post allocated into the experiment.

[3]This 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.

2.4 Nature of Assistance

Employers in the treatment could receive several different forms of assistance. Figure 2 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.

Figure 2: 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 the offer 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.2 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 Hiring consultants and recruiters were active

For about a third of treated job posts, the employer engaged with the Hiring Consultant by responding with at least one message. 66% of vacancies were "worked" by the recruiter, with the recruiter choosing to recruit jobseekers in 37% of cases, shortlist applicants in 27% of cases, or do both in 15% of cases. The remaining fifth of cases where the recruiter chooses to take no action result from vacancies which already received enough high quality applications that the employer was engaging with, so the recruiter judged the employer to need no help unless asked. The vacancies that received recruiting help looked systematically "worse" on observable characteristics than those that received shortlisting help. 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 B.1 for details).

Jobs with fewer applications more often received recruiting services; jobs with more applications more often received shortlisting (see Figure B.1 in Appendix B.3). We return to this relationship in Appendix Section B.5, where we use it to examine whether one form of help outperformed the other.

While all treated vacancies received contact from the Hiring Consultant, only 66% received assistance from a Recruiter, because the number of treated vacancies exceeded the Recruiter team's capacity.

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.3 we report the mean values and t-tests for various pre-randomization attributes of the job post, and we obtain good balance on these covariates. The treatment group contains slightly fewer low skill jobs than the control group, but this difference does not survive multiple hypothesis adjustments.

2.7 Estimation framework

For each outcome we estimate

$$y_i = \beta_0 + \beta_1 \textsc{Assigned}_i + \delta_t + \gamma_c + \eta_e + \varepsilon_i \tag{1}$$

where y_i is the outcome of interest for job i and $\textsc{Assigned}_i$ indicates that the job went to the treated group. All specifications include fixed effects for the day the vacancy was posted δ_t. For robustness, we add a second specification which includes fixed effects for category of work γ_c and level of expertise required from a worker η_e. Because assignment is randomized, β_1 is the intent-to-treat (ITT) effect of being enrolled in the program. It averages over the treated posts that staff worked and those they never got to, so it understates the effect of assistance actually delivered. Because our central results are nulls, that attenuation works in our favor, so we report the IV alongside every ITT rather than resting on the ITT alone.

To recover the effect of assistance *received*, we also report an instrumental-variables specification. Staff recorded whether they evaluated a job, giving a take-up indicator $\textsc{Received}_i$, and we estimate

$$y_i = \beta_0 + \beta_1 \widehat{\textsc{Received}}_i + \delta_t + \gamma_c + \eta_e + \varepsilon_i \tag{2}$$

with $\textsc{Received}_i$ instrumented by $\textsc{Assigned}_i$, with the same fixed effects as in Equation 1. Recall, recruiters worked posts in queue order as volume allowed, with 66% of the vacancies receiving attention from the recruiters. We define $\textsc{Received}_i$ as 1 if the vacancy received any attention from the recruiter, regardless of whether or not the recruiter determined the employer to benefit from assistance and then providing recruiting or shortlisting services, and 0 if the vacancy was never worked at all. Random assignment still makes the IV consistent for a local average treatment effect, the effect of assistance on the posts worked *because* they were assigned, but that complier population is not the whole treated group, which is one reason we lead with the assignment (ITT) effect. The exclusion restriction requires assignment to move outcomes only through take-up. A treated post's initial consultant contact precedes the sourcer's decision to work it, so if that early outreach independently shifted employer behavior, the IV would load it onto RECEIVED; the ITT is immune to this concern because it does not separate the channels. In practice ITT and IV agree in sign on every outcome and differ only in the expected scaling, so the choice does not affect the direction of any estimates. We use the IV (treatment-on-the-treated) estimate only in the cost-benefit calculation of

Section 4.4, where matching benefits to the costs that accrue on worked posts is the economically correct comparison; it is also the larger of the two estimates, so the program's benefits are evaluated at the more favorable one.

Inference. Treatment is assigned independently at the level of the job post. For post-level outcomes each post is its own assignment, so heteroskedasticity-robust standard errors are the design-justified choice (Abadie et al., 2023). For outcomes measured at the application level, where all of a post's applications share that post's single assignment, we cluster on the job post. On the appendix sample where an employer contributes several job posts, we conservatively cluster standard errors at the employer level. The selection-corrected models cluster on the employer throughout.

Selection-corrected estimates (FIML). Match-quality outcomes (wagebill, hours, the hired worker's wage, and feedback) are observed only when a post results in a hire. Although treatment is randomly assigned, comparisons conditional on hiring need not preserve randomization if treatment affects which posts result in a hire. We therefore complement each estimate conditional on a hire with a Heckman-style selection model (Heckman, 1974, 1979).

The selection equation models whether a post results in a hire, while the outcome equation models match quality conditional on hiring. Both equations include treatment and the baseline controls; the selection equation additionally includes the log number of applications as an exclusion restriction. Application volume strongly predicts whether a post results in a hire but, conditional on job observables, is assumed not to directly affect the quality of the resulting match. The model allows the unobservables determining hiring and match quality to be correlated. We report the treatment coefficient from the outcome equation as the selection-corrected estimate, with standard errors clustered at the employer level.

Samples and within-firm spillovers. Because randomization is at the post level, a firm that posts several jobs during the experiment can hold both treated and control posts, and help given to one post could plausibly spill over to its siblings, through the shared applicant pool, the employer's attention, or what the employer learns about the service. Our main specification therefore uses each firm's *first* experimental post (52,471 posts, one per firm): a first post has no earlier siblings whose treatment could contaminate it, and whether a post is a firm's first is determined before that post's assignment, so the restriction conditions on nothing realized after randomization. The appendix reports every result on both the first post only sample as well as the *full* experimental sample (all posts, immune to any posting response but admitting within-firm spillover). The hiring, first-stage, match-quality, and heterogeneity results are materially the same across both samples; the one exception is the 30-day wagebill, significantly negative in the full sample but attenuated and insignificant on the first-post main specification, which we flag where we present it (Section 4.3).

3 The first stage: recruiters provided recruiting and screening

Before turning to our main result on hiring, we first establish that the recruiters were in fact actively recruiting and suggesting candidates to treated job posts (Figure 3). This matters because it rules out the possibility that the null effect on hiring simply reflects a failure to deliver the treatment. Throughout this section we will report the ITT first, with the IV estimate following in parentheses.

The treatment substantially increased the number of recruiting invitations sent. Employers of control posts sent 4 invitations, and on average treated posts received an added 3.1 (IV: 4.59) recruiting invitations sent. This average treatment effect obscures that many treated employers received shortlisting instead of recruiting help: treated job posts which received any recruiting help at all received an average of 9 recruiting invitations sent by the recruiter.

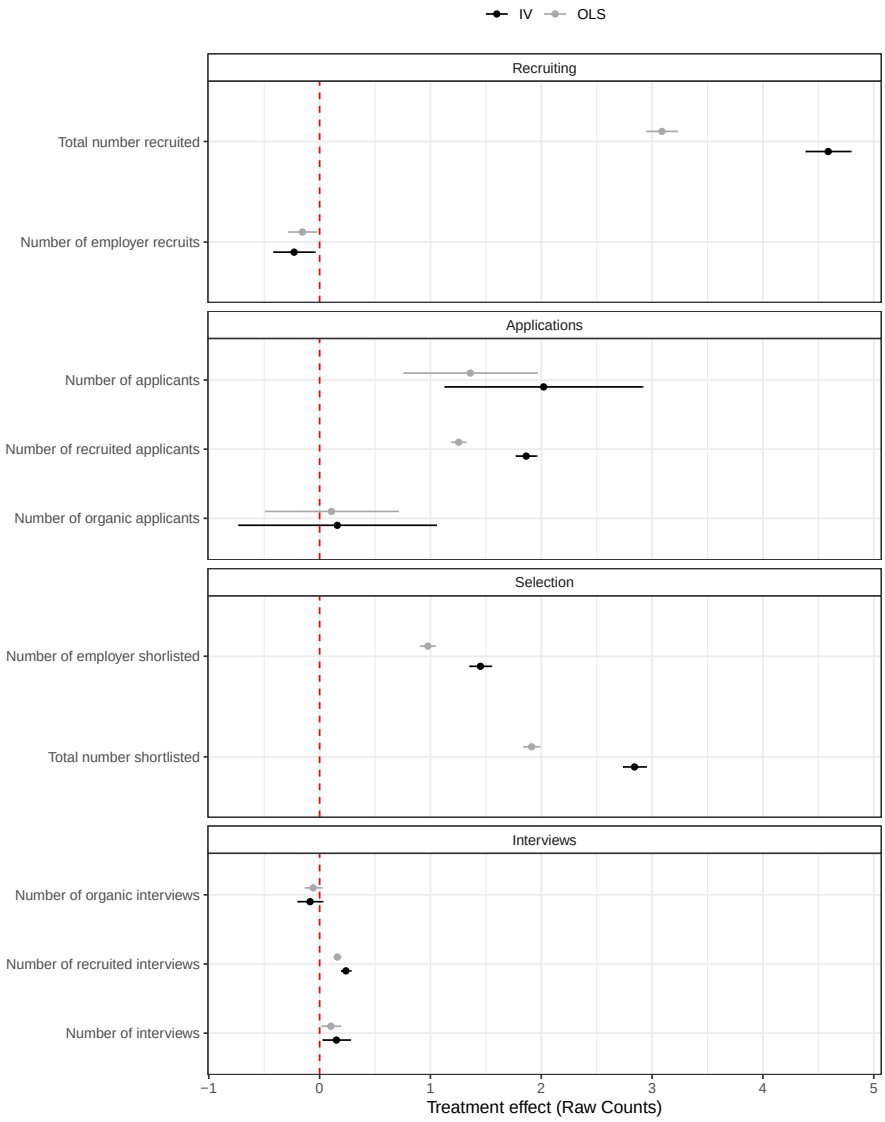
Treated posts received 5% (7.8%) more applications overall, over the control mean of 26. Applications come in two kinds: those from jobseekers who found the post and applied on their own ("organic" applications) and those that came from jobseekers who were invited to apply (either by the employer or a recruiter). The treatment increased invited applications by 1.3 (1.86) and left organic applications essentially unchanged (0.11 (0.16) against a control mean of 23), so the pool grew entirely through the recruiter's channel.[4]

Treated job posts received additional screening from the recruiter. On job posts where the recruiters choose to shortlist applicants, recruiters shortlisted an average of 3 applicants. The average effect of being assigned treatment was 1.91 (2.84) additional shortlisted applications. Employers' own shortlists rose by 0.98 (1.45) in the treatment group, so on this margin the consultants complemented rather than displaced employer effort.

Total interviews rose by 0.1 (0.15) in the primary sample, against a control mean of about 2 interviews, a small but statistically significant increase. The composition of interviews shifted more than the total expanded: interviews with recruiter-sourced applicants rose by 0.16 while interviews with organic applicants edged down by -0.06 (-0.09), though this decline is not statistically significant on the first-post sample. There is nonetheless some evidence for the substitution in the full experimental sample, where the organic decline is larger and significant, a fall of -0.09 (-0.14), and the organic interview rate falls from 12.0 to 11.3%. Employers reallocated their interviewing toward recruiter-sourced candidates rather than expanding it. This is not a labeling artifact in which workers who would have applied anyway were recruited and then classified as "invited": the number of organic *applications* was essentially unchanged. The substitution therefore appears to occur primarily at the employer's interview decision rather than at the application stage. This substitution is a first sign of why the strong first stage need not translate into a higher hiring rate: the intervention expanded the set of candidates employers considered, but primarily shifted their interviewing toward recruiter-sourced candidates.

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

Figure 3: Effects of the treatment on search and screening outcomes



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 02, 2021 to April 21, 2022, narrowed to employer's first vacancy. OLS specification regresses outcome on treatment assignment. IV specifications use treatment assignment as an instrument for whether or not a job receives the treatment. All specification include day fixed effects. 95% confidence intervals use heteroskedasticity-robust standard errors. Results in table form appear in Appendix Tables B.2–B.5.

4 Results

4.1 The treatment did not increase employer's hiring probability

Employers of treated job posts hired at the same rate as employers of control posts despite this substantial recruiting and screening activity. Recall that employers of control job posts had access only to the platform's algorithmic recommendation tools. Job posts in the control group had a base hiring rate of 29%. Table 1 shows the treatment effects on hiring outcomes. The treatment effect is 0.006, a 2% relative increase on that base, and is not statistically significant. Table 1 runs on the first-post main specification with EHW-robust standard errors; the null is equally present on the full experimental sample (Appendix Table B.9).

Table 1: Effects of Treatment on Hiring (first-post main specification)

	(1)	(2)	(3)	(4)
Panel A: OLS (Intent-to-Treat)				
Treatment Effect	0.006	0.006	0.006	0.005
	(0.005)	(0.005)	(0.005)	(0.005)
Panel B: IV (Treatment-on-the-Treated)				
Instrumented Treatment	0.009	0.008	0.009	0.008
	(0.007)	(0.007)	(0.007)	(0.007)
Posting-Day FE	Yes	Yes	Yes	Yes
Category Group FE	No	Yes	No	Yes
Expertise Tier FE	No	No	Yes	Yes
Control Mean	0.286	0.286	0.286	0.286
N (posts)	52471	52471	52471	52471

Notes: This table reports effects of treatment on hiring outcomes (whether anyone got hired). Sample is each employer's first post in the experiment. OLS specifications regress outcome on treatment assignment. IV specifications use treatment assignment as an instrument for employers receiving the treatment. All specifications include posting-day fixed effects (the treated share varied over the experimental period) and add category-group and expertise-tier fixed effects as indicated; standard errors are EHW-robust. Significance indicators: $+\ p \leqslant 0.10$, $*\ p \leqslant 0.05$, $**\ p \leqslant 0.01$, $***\ p \leqslant 0.001$.

Recruiters 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 on hiring probability and still find no effects (see Appendix Section B.5).

4.2 Hired workers were more likely to be recruited, and less likely to be recommended by the platform's reputation system

Although the assistance did not change whether an employer hired, it changed the pool of applications employers considered, and the channel through which they hired. Appendix Table B.6 reports the hiring effect by applicant source; Appendix Table B.7 reports the same on the full experimental sample.[5] While in the control group 10% of employers hired an applicant who had been recruited, employers of treated job posts were 1.5 (2.2) percentage points more likely to do so. Treated employer were 0.6 points more likely to hire a shortlisted applicant.

Although the treatment changed *who* was hired, it did not improve the observable quality of the workers employers ultimately selected. Narrowing the sample to only job posts which result in a hire, Table 2 compares the observable characteristics of those hired worker across arms. On every measure of the worker's observable quality the two arms are statistically indistinguishable: the worker's average prior ratings, prior earnings, profile wage, platform tenure, prior hours worked, and likelihood of having had no prior platform experience are all null. Treated hires are 3.4 (5.2) percentage points more likely to have come from a recruited candidate and 2.0 (3.1) percentage points less likely to have been an algorithm recommended applicant. Appendix Table B.11 shows these effects are similar in the full sample. Overall, the recruiter redirected employer attention toward candidates who looked the same on the platform's observables and were preferred by the recruiter, and away from candidates who found the job posting on their own or were recommended by the platform's reputation system.

[5] The channels are overlapping indicators rather than a partition of hires: a post can hire through more than one, and shortlisted hires come almost entirely from the organic pool, so the components need not sum to the overall effect.

Table 2: Effect of the treatment on characteristics of hired workers (first-post main specification)

	Observable quality						Channel of the hire	
	Platform rating	Newcomer	Log rate	Log earn.	Log hours	Tenure	Via recruiter	Recommended
Panel A: OLS (Intent-to-Treat)								
Treatment effect	0.002	−0.011	−0.021	0.004	0.025	0.038	0.034***	−0.020*
	(0.002)	(0.008)	(0.017)	(0.047)	(0.052)	(0.067)	(0.010)	(0.010)
Panel B: IV (Treatment-on-the-Treated)								
Instrumented treatment	0.003	−0.017	−0.032	0.006	0.038	0.058	0.052***	−0.031*
	(0.003)	(0.013)	(0.026)	(0.073)	(0.079)	(0.103)	(0.016)	(0.016)
Control mean	0.955	0.209	3.483	9.939	6.156	4.056	0.384	0.403
N	11,160	13,883	13,878	13,883	13,883	13,883	13,883	13,883

Notes: Each column is a characteristic of the hired worker. Panel A (OLS) is the intent-to-treat effect from regressing the characteristic on treatment assignment; Panel B (IV) instruments actual sourcing take-up with treatment assignment. Each panel reports the treatment effect with the heteroskedastic robust standard error beneath. Estimates run over each employer's first experimental post that made a hire (one post per firm, so no clustering is needed), with posting-day fixed effects. Column headings: Platform rating = platform's score based on work history; Newcomer = indicator (no platform history to rate); Log rate = log profile hourly rate; Log earn. = log prior earnings; Log hours = log prior hours worked; Tenure = platform tenure in years; Via recruiter = the hire came through a recruiter invitation; Recommended = the hired worker carried the algorithmic recommendation. The first six columns describe observable quality (joined from platform profiles); the last two describe the channel of the hire.+ $p \leqslant 0.10$, $*$ $p \leqslant 0.05$, $**$ $p \leqslant 0.01$, $***$ $p \leqslant 0.001$.

4.3 The treatment, at best, did not improve match quality

Our primary measures of match quality are hours worked on hourly contracts, the hourly wage rate, total wagebill, and employer and worker feedback ratings, the platform's direct measure of how the engagement went. We also look at whether the treatment impacts employers probability of returning to the platform after the experiment. No single measure is decisive: hours can fall for benign reasons, such as a skilled worker finishing faster, and the wagebill is a price–quantity composite, so we report them together. Table 3 collects the treatment effects on all six outcomes under one set of conventions: our main specification of each firm's first experimental post (Section 2.7).

Each outcome downstream of hiring is estimated three ways: OLS on treatment assignment (intent-to-treat); IV using assignment as an instrument for receiving help (treatment-on-the-treated); and a full-information maximum likelihood (FIML) selection model. The FIML model uses hiring as the selection stage and the log number of applications as the exclusion restriction: conditional on the other controls, application counts affect whether an employer makes a hire, but not directly the quality of the eventual match. This allows treatment to affect the composition of observed hires, rather than assuming that treatment changes only match quality conditional on hiring.

Hours, hourly wages, and wagebill. Employers posted jobs that were either hourly or fixed price. For hourly hires, we observe the number of hours worked for each hire and the hourly wage paid. For fixed price jobs, we only observe one lump sum payment. Wagebill includes both lump sum payments for fixed price

Table 3: Treatment effects on match quality (first-post main specification)

	Hours	Hourly Wages	High Emp Rating	High Worker Rating	Wagebill 30d	Wagebill 90d
Panel A: Post level						
OLS	4.67	−0.60	−0.0155^{+}	−0.0051	−27.35	31.85
	(7.90)	(0.67)	(0.0081)	(0.0111)	(36.37)	(92.00)
IV	7.06	−0.91	−0.0241^{+}	−0.0080	−41.91	48.81
	(11.95)	(1.03)	(0.0127)	(0.0174)	(55.73)	(140.97)
Control mean	197.42	34.79	0.9281	0.8655	998.66	2,751.95
N (posts)	7,698	9,521	5,976	5,502	13,864	13,864
Panel B: FIML selection-corrected, application level						
FIML	1.31	0.08	−0.0137^{+}	−0.0045	−21.34	10.77
	(9.60)	(0.66)	(0.0079)	(0.0106)	(28.88)	(71.74)
$\hat{\rho}$	0.39	−0.44	0.05	0.06	−0.28	−0.11
Control mean	205.05	31.56	0.9259	0.8704	791.46	2,180.68
N (hires)	10,225	13,026	7,052	6,768	17,724	17,724

Notes: The table has two panels estimated on two different samples. *Panel A* is at the post level (one row per post): hours, wages, and ratings are averaged over the post's hires and wagebill is the post's total spend (hourly-wage columns use hourly posts; feedback columns use posts with a rated hire). OLS is the intent-to-treat effect of treatment assignment and IV instruments recruiter take-up with assignment (treatment-on-the-treated). *Panel B* is at the application level, on the full applicant pool: FIML is the Heckman selection model estimated by full-information maximum likelihood, with hiring as the selection stage, the log number of applications as the exclusion restriction, and $\hat{\rho}$ the estimated selection–outcome error correlation. Panel B's unit is the individual hire rather than the post, so its estimand, control mean (over control-group hires), and N (number of hires) are not directly comparable to Panel A (over posts). Sample: each employer's first experimental post. Every specification includes posting-day fixed effects and category-group and expertise-tier controls. Standard errors are EHW robust. Significance: $+\ p \leqslant 0.10$, $*\ p \leqslant 0.05$, $**\ p \leqslant 0.01$, $***\ p \leqslant 0.001$.

jobs and wages paid for hourly jobs.

Conditional on hiring, we find no evidence that employers of treated jobs demanded more hours of work or that they paid higher wages, and the FIML selection correction specification finds no evidence that a change in composition is masking an effect. This null for hours is robust to counting hours as zero for the full sample and reporting hours in logs. The hired worker's hourly rate is a precise null, a point estimate of $\$-0.60$ on a $\$34.79$ control mean, indistinguishable from zero under all three estimators. Because fixed price jobs do not appear in hours or wages, we turn to wagebill for a more full picture of how the treatment affected employer demand.

Thirty days after posting, treated posts that hired had spent -27.35 (-41.91) less than control posts that hired on the first-post main specification, against a $998.66 control mean, negative but not statistically significant. The FIML selection corrected estimate is somewhat smaller in magnitude ($\$-21.34$), and none of these estimates are statistically significant in this sample. On the full sample of all posts the OLS estimate is more than twice as large ($\$-72.27$) and statistically significant. Widening the window to 90 days leaves the point estimate indistinguishable from zero. However, the 90-day estimate is substantially less precise, with standard errors roughly two and a half times as large, and therefore cannot rule out a spending reduction

of the magnitude estimated at 30 days.

Several mechanisms fit the pattern, and we are agnostic among them: a modest decline in realized match quality that disappears as engagements mature; front-loaded coordination or onboarding frictions that resolve over the first month; or simply slower hiring by treated employers, who face larger applicant pools and may take marginally longer to commit, mechanically truncating early spending. We therefore report the 30-day reduction as a fact about the experimental population rather than as strong evidence on match quality, and the cost-benefit calculation (Section 4.4) uses the more conservative 90-day window. We interpret these as generally null effects to hours, wages, and wagebill.

Feedback ratings. At the close of a job, each side of the match rates the other on a five-point scale: the employer rates the hired worker, and the worker rates the employer. We recode each rating as an indicator for 4 or above, so that a negative treatment effect indicates worse match quality. "High Emp Rating" is the score the worker gives the employer, and "High Worker Rating" is the rating the employer gives the worker. Estimates are negative in all specifications for both ratings: the rating hired workers receive falls by 0.5 percentage points against a control rate of 87%, small and insignificant in this sample. The treated employers were 1.5 (2.4) less likely to receive high ratings, on a base of 93%, and this is significant at the 10% level. Whether the employer or worker ratings are significant flips in the full sample, and all treatment effects are under 2%, so we read the feedback evidence as a small negative lean that never clears $p = 0.05$ significance on any single margin. Selection correction changes little ($\hat{\rho}$ is small for each feedback outcome), suggesting that differential selection into hiring is not driving the estimates. We revisit heterogeneity by tier and category under these same conventions in Section 6.

Returning to the platform. After posting a vacancy that is in the experiment, employers were free to post another at a later date. In the control group, 22.8% of employers return within 3 months. Treatment reduced the probability that an employer ever posted another vacancy by 4.0% (6.0%), or by 0.91 (1.35) percentage points. The decline in reposting does not appear to reflect treated firms simply returning later: treatment effects are similar when reposting is measured within 30, 60, 90, or 120 days.

Because some jobs on this platform are short term, it is possible for the employer to rehire the same worker for a future vacancy, although this only happens in 0.17% of the control group. The treatment did not appear to affect this outcome, although the low baseline rate creates a potential floor effect.

In summary: results are broadly similar in the first-post and full samples (Appendix B.4), with estimates that are generally negative but statistically insignificant. The results in the full sample agree on significance on all but 30-day wagebill, which is more negative and statistically significant. Overall, we find no evidence that the treatment improved match quality, and treated employers are significantly less likely to return to the platform.

Table 4: Treatment effect on returning to the platform and rehiring (first-post main specification)

	Any subsequent post	Rehire same freelancer
Panel A: OLS (Intent-to-Treat)		
Treatment effect	−0.91*	0.05
	(0.43)	(0.05)
Panel B: IV (Treatment-on-the-Treated)		
Instrumented treatment	−1.35*	0.07
	(0.64)	(0.07)
Control mean (pp)	22.77	0.17
N (posts)	52,471	52,471

Notes: One observation per firm (the 52,471 employers whose first experimental post is not tied with another; 0 tied first posts excluded). Both outcomes are indicators in percentage points: whether the firm makes any later experimental post, and whether it hires, on a later post, a freelancer it had hired on its first post (coded zero for firms whose first post did not hire). Panel A (OLS) is the intent-to-treat effect of the firm's randomly assigned *first*-post treatment; Panel B (IV) instruments whether the post received the treatment with that assignment (treatment-on-the-treated). Both include first-post posting-day fixed effects; standard errors (in parentheses) are heteroskedasticity-robust. The control mean is the mean outcome among control-group posts, in percentage points. Significance: $+\ p \leqslant 0.10$, $*\ p \leqslant 0.05$, $**\ p \leqslant 0.01$, $***\ p \leqslant 0.001$.

4.4 Cost-benefit analysis

Even under assumptions deliberately chosen to favor the program, the benefits do not cover 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 calculation uses the IV (treatment-on-the-treated) estimate, matching the cost side, which accrues only on posts staff actually work, from the same first-post specification as Table 1: the estimated effect of receiving the treatment on hiring is 0.009 with a standard error of about 0.007. The 95% confidence interval upper bound is 0.023. 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 \$2,776. Online labor platforms typically charge commission fees of around 10% on wagebill (Wood and Lehdonvirta, 2021), implying revenue per hire of about \$278. 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 (\$6.26) falls short of the cost per job post under every time assumption, with estimated losses between \$1.56 and \$6.76 per post. The program breaks even only if assistants spend less than 12.0

minutes per post under the upper CI, or less than 4.7 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	\$6.26	\$-1.56
20 min	\$10.42	\$6.26	\$-4.16
25 min	\$13.02	\$6.26	\$-6.76

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 (0.023) multiplied by revenue per hire (\$278), yielding \$6.26. Revenue per hire equals average 90-day wagebill per hiring job post (\$2,776) times a 10% platform commission rate (Wood and Lehdonvirta, 2021). Under the IV point estimate (0.009), expected revenue per post would be \$2.42.

This cost structure also highlights a scaling asymmetry. The program already could not serve 34% 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.

Standard search-and-matching models (Mortensen and Pissarides, 1994; Pissarides, 2000) and directed search frameworks (Wright et al., 2021) do not predict the pattern we find. Employers given recruiters saw expanded applicant pools and ran more interviews. Hiring rates, however, were flat and match quality, if anything, 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 receives applications from workers. The employer's goal is to hire the most productive applicant, but due to costs of conducting interviews she hires the first applicant that she interviews above an endogenous cutoff she fixes after observing the size of the applicant pool.

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 its 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. These candidates are classified as the algorithmic pool ($\mathcal{A}$).

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 the remaining true quality component ε_j, and a noisy signal $\xi_j^E \sim N(0, \sigma_\xi^2)$. Therefore, the applicant's productivity is $q_j^E = s_j + \varepsilon_j + \xi_j^E$, or $q_j^E = \alpha_j + \psi_j + \varepsilon_j + \xi_j^E$. At each interview she observes this noisy measure of the applicant's 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"), we call these employers "treated." The recruiter's role is to add candidates to the pool that the employer observes that he views 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, the remaining true quality component ε_j, and a noisy signal $\xi_j^R \sim N(0, \sigma_{\xi^R}^2)$. The recruiter's signal of each candidates productivity is $q_j^R = s_j + \varepsilon_j + \xi_j^R$, with $\xi_j^R \sim N(0, \sigma_{\xi^R}^2)$, or $q_j^R = \alpha_j + \psi_j + \varepsilon_j + \xi_j^R$. The recruiter is not aware of the employer's assessment of the candidates. The recruiter recruits m applicants $q_j^R \geqslant \bar{s}$ that were not already selected by the algorithm. These candidates are classified as the recruiter pool ($\mathcal{R}$).

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

The employer makes a hire so long as at least one selected applicant has q_j^E above her cutoff $\bar{u}$, which she fixes after she observes the size of the applicant pool. She does not observe which applicants were selected by the algorithm versus the recruiter. Each candidate in $\mathcal{A}, \mathcal{R}$ passes the employer's screen independently with probability

$$p_{\mathcal{A}} = P(q^E \geqslant \bar{u} \mid s \geqslant \bar{s}), \qquad p_{\mathcal{R}} = P(q^E \geqslant \bar{u} \mid s < \bar{s},\, q^R \geqslant \bar{s}). \tag{3}$$

We assume that candidates are presented to the employer in random order with uniform probability across and within pools. The likelihood of making a hire for the treated employers using recruiters is versus the control employers without recruiters is

$$P_T = 1 - (1 - p_{\mathcal{A}})^n (1 - p_{\mathcal{R}})^m, \qquad P_C = 1 - (1 - p_{\mathcal{A}})^n. \tag{4}$$

Result 1 (Hiring effect of recruiter-added candidates is weakly positive)**.** *If the employer keeps her hiring threshold fixed, the recruiter-added candidates increase the probability the employer makes a hire.*

$$\Delta_P \equiv P_T - P_C = (1 - p_{\mathcal{A}})^n \left[1 - (1 - p_{\mathcal{R}})^m\right] \geqslant 0. \tag{5}$$

However, if the treated employer raises $\bar{u}$ in response to the larger applicant pool, the recruiter's effect on employer's hiring probability will shrink.

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

5.3 Match quality

The employer is interested in hiring better quality candidates, so it matters how the quality of the hired candidate changes in the presence of the recruiter. Recall that in the control group, employers only consider applicants from $\mathcal{A}$ and in the treated group they receive applicants from both $\mathcal{A}$ and $\mathcal{R}$, therefore they will pick different cutoffs of u, $\bar{u}_C$ and $\bar{u}_T$.

Let a worker's true value be denoted v. A candidate hired to a job in the control group will have expected quality Q_C

$Q_C = E[v \mid q^E \geqslant \bar{u}_C,\, j \in \mathcal{A}]$.

Let π_R[6] be the probability that a treated employer hires a candidate from the recruited pool, $\mathcal{R}$. For a worker hired in the treatment group, their expected quality will be a weighted average of the probability that the candidate comes from the algorithmic or recruited pool.

$Q_T = \pi_R E[v \mid q^E \geqslant \bar{u}_T,\, j \in \mathcal{R}] + (1 - \pi_R) E[v \mid q^E \geqslant \bar{u}_T,\, j \in \mathcal{A}]$.

The treatment effect on expected match quality is

$$\Delta_Q \equiv Q_T - Q_C. \tag{6}$$

Although the experiment identifies Δ_Q for observed match quality, the model clarifies which margin drives it.

Definition. Let $\omega(u_T)$ be the difference between average quality of workers above the employer's quality threshold in the recruited and algorithmic pool, conditional on being above the employer's hiring quality threshold. This difference comes from two sources. First, that the quality of the candidates in $\mathcal{R}$ and $\mathcal{A}$ are different. Second, this difference is also driven by the strength of the threshold response by the treated employers.

$$\omega(u_T) \;\equiv\; E[v \mid q^E \geqslant \bar{u}_T,\, j \in \mathcal{R}] \;-\; E[v \mid q^E \geqslant \bar{u}_T,\, j \in \mathcal{A}]. \tag{7}$$

The treatment effect on match quality therefore decomposes to:

$$\Delta_Q = \underbrace{\pi_{\mathcal{R}}(\bar{u}_T)\,\omega(\bar{u}_T)}_{\text{source composition}} \;+\; \underbrace{E[v \mid q^E \geqslant \bar{u}_T,\, \mathcal{A}] - E[v \mid q^E \geqslant \bar{u}_C,\, \mathcal{A}]}_{\text{threshold response}}. \tag{8}$$

The first component of the match quality effect is the probability that the employer hires a recruited applicant times the difference in quality between the recruited and algorithmic applicant pools. Without a threshold response, a negative ω means the hired workers drawn from the recruited pool have lower expected true quality than the hired workers from the algorithmic pool. The second component is the difference in match quality driven by how much the treated employers move the threshold u_T. When the treated employer responds to the additional candidates by raising her hiring threshold, this shrinks the negative match quality effect driven by negative ω.

5.4 When does recruiter screening improve match quality?

Recruiters might add value by improving the quality of the workers hired, but this is not a given. Recruiters can only pick among candidates the algorithm passed over, so recruited applicants always look worse on the score. The score, however, is imperfect. It captures α_j only with noise ψ_j and misses ε_j entirely, so

[6] We derive π_R in Appendix C.3

recruiting helps if the recruiter finds candidates who are strong enough in ε_j to make up for being below the threshold in the algorithmic score.

Proposition 1 (Under a fixed threshold, correlated errors raise the recruited share of hires)**.** *Hold the thresholds and every other primitive fixed and raise the covariance $\sigma_{\xi_{RE}}$ between the recruiter's and the employer's assessment errors.*

(i) Recruited candidates clear the employer's screen more often, so the hiring gain Δ_P and the recruited share of hires $\pi_{\mathcal{R}}$ both rise, while nothing about algorithmic hires changes.

(ii) At a common threshold $\Delta_Q = \pi_{\mathcal{R}}\,\omega$, so wherever recruited hires are worse than algorithmic ones, correlated errors lower match quality by raising the share of hires that are recruited.

Intuitively, the recruiter invites an applicant because they look good to him. When his errors are correlated with the employer's, that same applicant looks good to her for the same reason, so more recruits clear her screen (Appendix C.5).

Proposition 2 (Recruiting helps when the score is uninformative)**.** *If the score carries none of true quality ($\sigma^2_\alpha = 0$), the assessment errors are independent of the score's noise, and $0 \leqslant \sigma_{\xi_{RE}} \leqslant \min(\sigma^2_{\xi_R}, \sigma^2_{\xi_E})$, recruited hires are better than algorithmic ones, $\omega > 0$.*

Since Δ_Q has the sign of ω at a common threshold, recruiting then raises match quality. Intuitively, a pure-noise score says nothing about a worker while the recruiter's screen does, so even the worst recruit who clears both the recruiter's and the employer's screens beats the average algorithmic hire. The size of the gain, not its sign, depends on how much the assessments know; with very noisy or nearly identical assessments it vanishes.

Outside of simple cases, the sign depends on various parameters, so we calibrate the model to match observed statistics in the experiment (Appendix C.6). The share of candidates recommended, shortlisted and hired pin the recommendation bar, the recruiter's bar and the employer's threshold; three within-post score gradients pin the correlations among the score and the two assessments; the pool sizes are observed counts; and the two quantities the cross-section cannot pin, how informative a human assessment is relative to the score and how informative the score is, are set to benchmarks. At the calibration recruited candidates clear the employer's screen at the same rate as organic ones, so her threshold barely moves. The hiring gain is 1.72 percentage points per recruiter-acted post and the match-quality effect is -0.027 stars, each inside the confidence interval of the corresponding experimental estimate.

Result 2 (Correlated recruiter–employer errors lower match quality)**.** *Around the calibration (the domain of Figure C.1), Δ_Q is negative and decreasing in the correlation $\rho_{\xi_{RE}}$ between the recruiter's and the employer's assessment errors.*

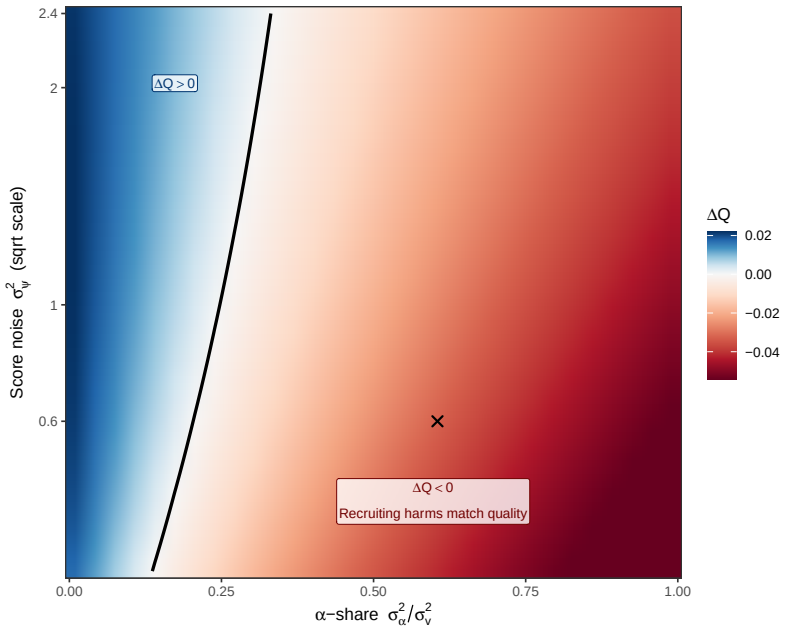
When the recruiter's residual assessment is unrelated to the employer's, his mistakes rarely survive her screen. When the two assessments are positively correlated, candidates who look good to him for idiosyncratic reasons look good to her for the same reasons: he adds "fool's gold", and once such a candidate

is hired, the noise washes out and their true quality shows. Correlated errors also let more recruits through (Proposition 1), so recruited hires become both worse and more common, and both channels push Δ_Q down.

Result 3 (A better algorithm lowers the value of recruiting)**.** *Around the calibration (the domain of Figure 4), Δ_Q is decreasing in the algorithmic share $\sigma_\alpha^2/\sigma_v^2$ at fixed σ_v^2, and positive only below a threshold share that rises with the noise in the score, σ_ψ^2.*

As the algorithm captures more of true quality, the workers below its cutoff are increasingly workers the algorithm was right to pass over, so there is less hidden quality for the recruiter to find. Figure 4 varies the algorithmic share on the x-axis and the score's noise on the y-axis: recruiting raises match quality only to the left of the black contour, and the contour shifts right as the score gets noisier.

Figure 4: Comparative statics of the match-quality effect of recruiting as the algorithm improves



Notes: The figure spans the $(\sigma_\alpha^2/\sigma_v^2, \sigma_\psi^2)$ plane, algorithmic share in $(0, 1)$ and $\sigma_\psi^2 \in [0.24, 2.4]$, and plots Δ_Q, the treatment effect on expected match quality per recruiter-acted post under the reservation-signal threshold rule, both pieces of equation (8) included (Appendix C.6). The black contour denotes $\Delta_Q = 0$; red (blue) regions indicate that recruiting lowers (raises) expected match quality. Values are clipped at $[-0.054, 0.022]$. The three bars are re-pinned at every point to the recommendation share, the recruiter's selectivity, and the control fill rate; n, m and the noise covariances not on the axes are held at their calibrated values, so $\mathrm{corr}(v, s)$ and k vary across the plane with the axes. The cross marks the calibration point (Appendix C.6).

5.5 Mapping to the data

The model generates four predictions we revisit in Section 6. Result 1 rationalizes the near-zero hiring effect: at the calibration matching the control fill rate, the hiring-probability gain from the recruited pool is

1.72 percentage points per recruiter-acted post, within the confidence interval of the experimental estimate, and the employer's threshold response to it is negligible (Appendix C.6). Result 2 rationalizes a null to negative impact on match quality. Proposition 2 and Result 3 suggest that recruiter value is largest where ε's share is largest, that is, in jobs where platform histories are thin. This matches the direction of the low-tier hiring gain, the one subgroup result significant before multiple-testing adjustment but not after. Finally, the model predicts that the algorithm's marginal value is largest when the employer has the least ability to find a good match on her own. In the model, firms with a particularly noisy signal gain more from the algorithm and potentially, a recruiter. In the data the recommendation moves hiring for small, inexperienced, and newly registered employers and in thin pools, the settings where employers are less likely to have developed a keen sense for identifying a good fit.

6 Heterogeneity and Evidence Supporting the Model

6.1 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. For each applicant–vacancy pair, the platform constructs an algorithmic score measuring the applicant's predicted fit for the vacancy. Each job post has a threshold above which applicants are labeled "recommended." Employers do not observe the underlying score, but they do observe this discrete recommendation label. We exploit the discontinuous change in recommendation status at the 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. Because the platform's threshold is not observed directly and must be reconstructed from the pattern of recommendation labels, we repeat every estimate under two alternative reconstruction rules, defined in Appendix B.8.1; we refer to the three as the baseline and alternative threshold definitions below. In Figure 5 we show the first stage. Appendix B.8 reports density tests, and the covariate balance plots, which support the validity of the design.

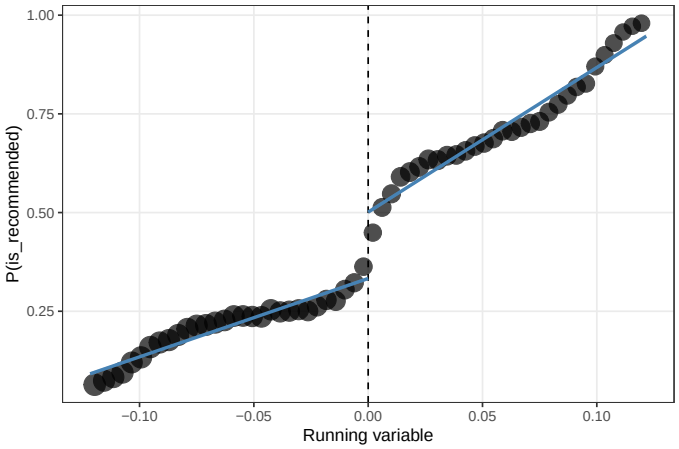


Figure 5: First stage: recommendation probability at the score threshold

Notes: First-stage relationship between the recommendation score running variable and recommendation status. 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. The outcome is the probability a jobseeker is recommended for the vacancy.

Recommendation increases hiring. Figure 6 shows how being recommended for a vacancy increases the probability a given jobseeker is hired. Any given applicant is not very likely to be hired for a vacancy, among applicants just below the threshold only 0.8% get hired. Algorithmic recommendation increases the probability of being hired. The bias-corrected RD estimate is 0.012 (SE $= 0.005$, $N = 197,665$, bandwidth $= 0.095$). Therefore, being recommended by the platform more than doubles a given applicant's chance of being hired.

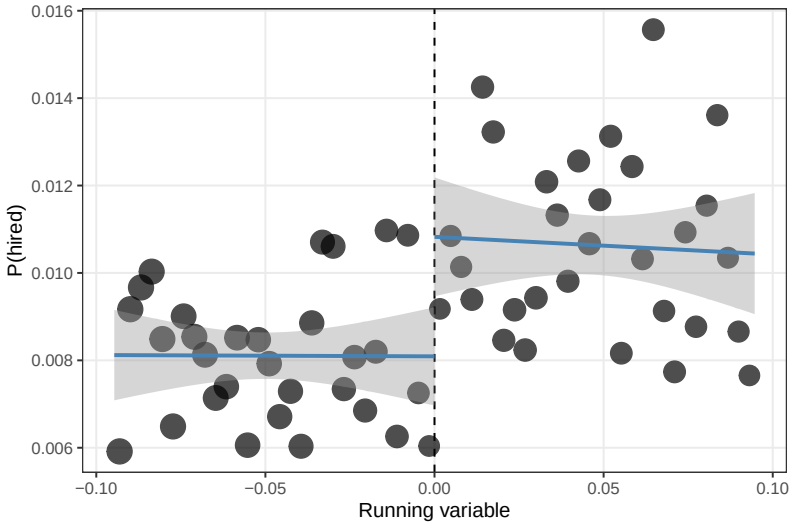


Figure 6: RD: hiring probability at the recommendation threshold

Notes: The unit of observation is the individual application: the plot bins applications by their recommendation-score running variable and plots the share hired within each bin, so the vertical axis is the probability that a *given applicant* is hired, distinct from the post-level hiring rate in Table 1. 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 B.21. Additional RD results appear in Appendix B.8.

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 B.29 and B.32); under the baseline definition the estimate is positive but imprecise (Table B.23). 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 driven by employers who rely most on the algorithm. The hiring response to recommendation is concentrated among the newest and least experienced employers. Splitting the discontinuity by employer characteristics, the effect is positive and significant for firms with no prior hires, for the smallest size tier (1–10 employees), and for the newest accounts, and is undetectable for larger, more experienced, or longer-tenured firms (Appendix B.9). The same pattern appears in pool size: the effect concentrates in the lowest application-count quartile and is absent in the upper three (Table B.36), with wagebill directionally consistent (positive in Q1, negative in Q3–Q4; Table B.38). Threshold placement itself produces no differential effects (Table B.39). Because these splits test many cells at once, we adjust for multiple hypotheses across all 24 hiring cells in the five splits: 5 clear 5% unadjusted and all 5 survive

a Benjamini-Hochberg correction (largest surviving $q = 0.037$), so the pattern is not an artifact of testing many subgroups. The RD estimates the employer's marginal response to the recommendation label, and that response is largest when she has the weakest basis for choosing on her own, consistent with the model.

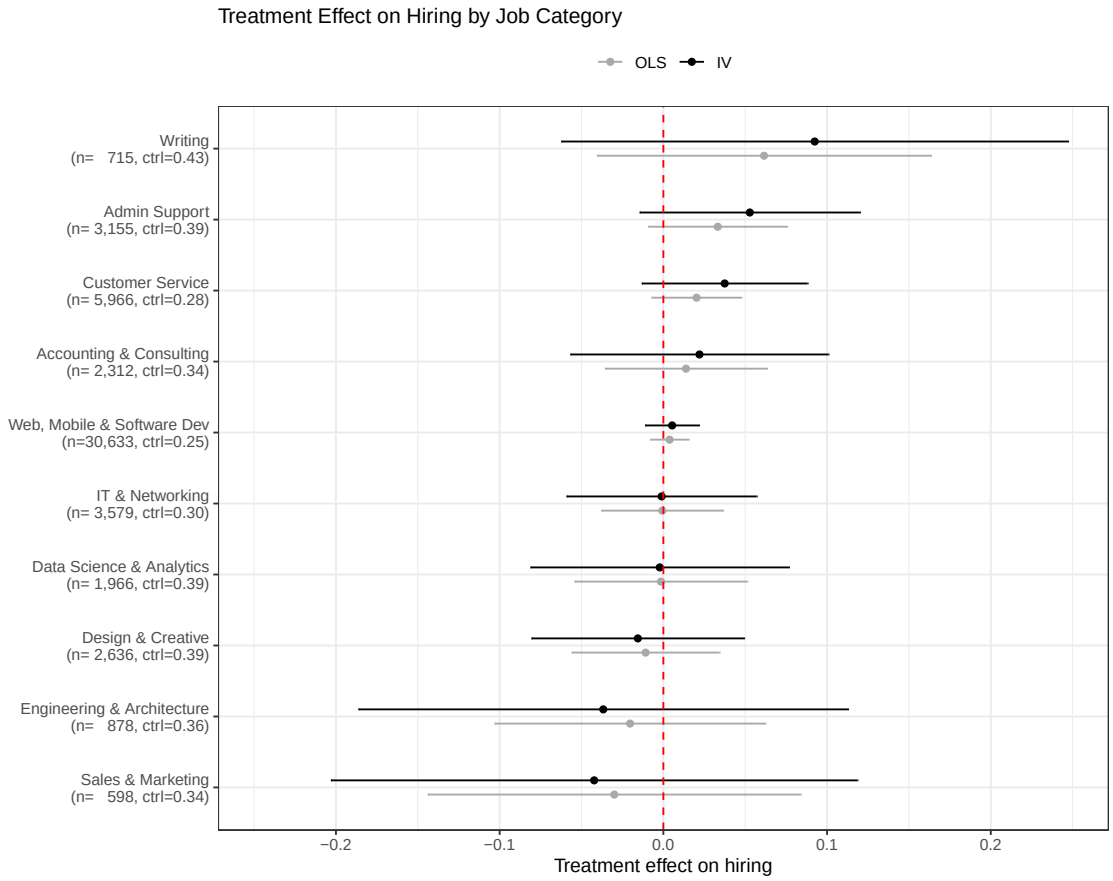
6.2 Treatment effect heterogeneity: where the hiring consultants helped

We investigate if the (lack of a) hiring treatment effect differed across job category and job types. We explore subgroup heterogeneity to characterize where, if anywhere, consultants came closer to helping. When posting a vacancy, employers categorize it into one of twelve categories, and flag a desired "expertise tier", for workers who are either Cheap/Inexperienced, Intermediate, or Expert/Expensive. [7]

The treatment effect on hiring is not significant for any job category, even before multiple hypothesis testing. We show these results in Figure 7 and Table B.51). Administrative Support, Writing, and Customer Service jobs have the largest point estimates. Writing has the largest point estimate, but includes fewer than a thousand posts and its interval is correspondingly wide. With ten categories tested jointly, all stepdown-adjusted p-values exceed 0.7. This provides no statistically distinguishable evidence of heterogeneity, although the relatively wide confidence intervals leave substantial room for the possibility of category-specific effects.

[7] Because we test many subgroups at once, we adjust for multiple hypotheses. Appendix Tables B.51 and B.52 report the unadjusted, stepdown, Bonferroni, and FDR-adjusted p-values. The test statistic is the studentized within-subgroup treatment effect, and the null distribution comes from an employer-level pairs bootstrap drawn jointly across subgroups; the stepdown procedure of List et al. (2019) controls the familywise error rate while exploiting the dependence across subgroups.

Figure 7: Treatment effects on hiring by job category. Sample: each firm's first experimental post (first-post main specification).



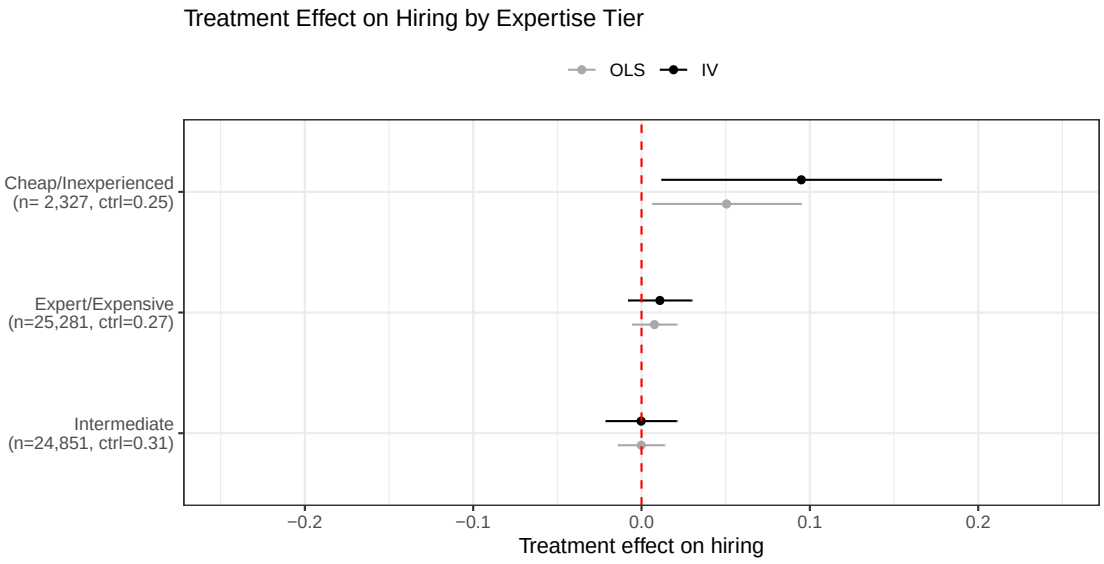
Notes: This figure reports treatment effects on hiring (whether anyone got hired) separately by job category, excluding Legal due to small sample size. OLS specification regresses outcome on treatment assignment. IV specifications use treatment assignment as an instrument for whether or not a job receives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

The expertise-tier split is the only place a signal appears at all, although it does not survive multiple hypothesis testing correction (Figure 8, Table B.52). The lowest skill tier (cheap, inexperienced workers) shows a treatment effect of 0.051 on a 24.7% control hiring rate, a 20% relative increase, and it is the one subgroup anywhere in the heterogeneity analysis that clears 5% significance unadjusted. Correcting for the three tiers tested moves it to 0.080 under the stepdown procedure and 0.063 under Bonferroni, both short of conventional levels. The intermediate and expert tiers are null. We therefore treat the lowest tier as the single suggestive result in the heterogeneity analysis rather than an established one. It is worth recording because it points where the model predicts the most room to help: workers with thin platform histories have a smaller

algorithmic-quality variance σ_α^2, so the score captures less of true quality and a recruiter's independent look can add information (Figure 4). On a sample of 2,327 first posts in that tier, the experiment cannot separate that reading from chance.

The match-quality outcomes, examined by tier on the full experimental sample with category-group and posting-day fixed effects and standard errors clustered on the employer, are individually noisy (Appendix Table B.16). The 30-day wagebill decline is largest and significant at the 10% level in the expert tier where the treatment group had 8% lower wagebill (with a treatment effect of $\$-103$) and the treatment effect is null for cheaper tiers, and the feedback leans negative the same way among expert-tier hires. Expert workers have rich, score-legible profiles, so the recruiter's signal is most redundant with the algorithm's there, the regime in which the model predicts screening can subtract value. Category-level quality splits do not provide any evidence that match quality is significantly different between the treatment and control group (Appendix Table B.17). Taken together, the tier evidence traces the two edges of the model: recruiters help where the algorithm sees least (the lowest tier's hiring gain) and, if anything, hurt where it already sees most (the expert tier's wagebill and feedback).

Figure 8: Treatment effects on hiring by expertise tier. Sample: each firm's first experimental post (first-post main specification).



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 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. Employers of treated job posts hired at the same rate as employers of control posts who had only the platform's algorithmic tools. Across these measures, we find no evidence that match quality improves; ratings and wagebill lean negative, and treated employers are significantly less likely to return to the platform. The costs of the program are difficult to justify even under optimistic assumptions.

The platform's recommender system nevertheless plays a meaningful role in employer decisions. A regression discontinuity design shows that employers respond to its recommendations by shortlisting and hiring recommended applicants, although recommendation does not improve realized match quality. The experiment therefore takes place in a setting where algorithmic information is both consequential and incomplete, leaving potential scope for human recruiters to add value, but we find little evidence that they do so.

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 predicted to be lower, which is consistent with the negative lean in feedback ratings we observe and lower probability of employers returning to the platform.

Several alternative explanations are less consistent with the evidence. If recruiters were simply ineffective, we would expect noise rather than a consistent, if individually insignificant, tilt toward lower feedback, and employers would not have responded to their input. Employers of treated job posts, 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 candidates who went on to receive lower ratings. If the algorithm were near-perfect, leaving no room for human judgment, adding candidates should be neutral rather than harmful. The marginal quality tilt we observe requires an explanation for why recruited hires, if anything, fare slightly 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: the lowest expertise tier, and to a lesser extent customer service, 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. Evaluating candidates using online tools and platforms is the norm rather than the exception offline as well as online. Still, information in our setting is more standardized than in other forms of hiring. Human intermediaries may therefore add more value in settings where relevant worker attributes are less codified, where algorithms have thinner data, or where recruiters have access to information that is not already reflected in the platform's signals. Labor markets that rely more heavily on informal networks, for example, may operate differently (Brown et al., 2016). Moreover, conventional intermediaries such as temporary-help firms often provide services beyond search and matching, including training (Autor, 2001), and human-resource interventions may be complementary with other organizational practices (Aral et al., 2012). Our experiment does not speak to these broader organizational roles of HR outside of hiring.

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, as employers have been shown to learn from their hiring experiences (Leung, 2018). Our treatment combines search and screening 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

A. Abadie, S. Athey, G. W. Imbens, and J. M. Wooldridge. When should you adjust standard errors for clustering? *Quarterly Journal of Economics*, 138(1):1–35, 2023.

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, 2019.

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*, 66(5):1802–1825, 2020.

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.

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.

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 résumé screening. Working paper, Columbia Business School, 2018.

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. 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.

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.

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.

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.

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.

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

J. J. Heckman. Shadow prices, market wages, and labor supply. *Econometrica*, 42(4):679–694, 1974.

J. J. Heckman. Sample selection bias as a specification error. *Econometrica*, 47(1):153–161, 1979.

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

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. In *International workshop on internet and network economics*, pages 515–522. Springer, 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.

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.

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

S. Karlin and Y. Rinott. Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *Journal of Multivariate Analysis*, 10(4):467–498, 1980.

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.

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.

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.

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.

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

R. L. Plackett. A reduction formula for normal multivariate integrals. *Biometrika*, 41(3/4):351–360, 1954.

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.

A. Schiprowski. The role of caseworkers in unemployment insurance: Evidence from unplanned absences. *Journal of Labor Economics*, 38(4):1189–1225, 2020.

D. Slepian. The one-sided barrier problem for Gaussian noise. *Bell System Technical Journal*, 41(2):463–501, 1962.

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, 2016.

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.

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 and competitive search equilibrium: A guided tour. *Journal of Economic Literature*, 59(1):90–148, 2021.

Appendix

A Experiment details

A.1 Platform Interface

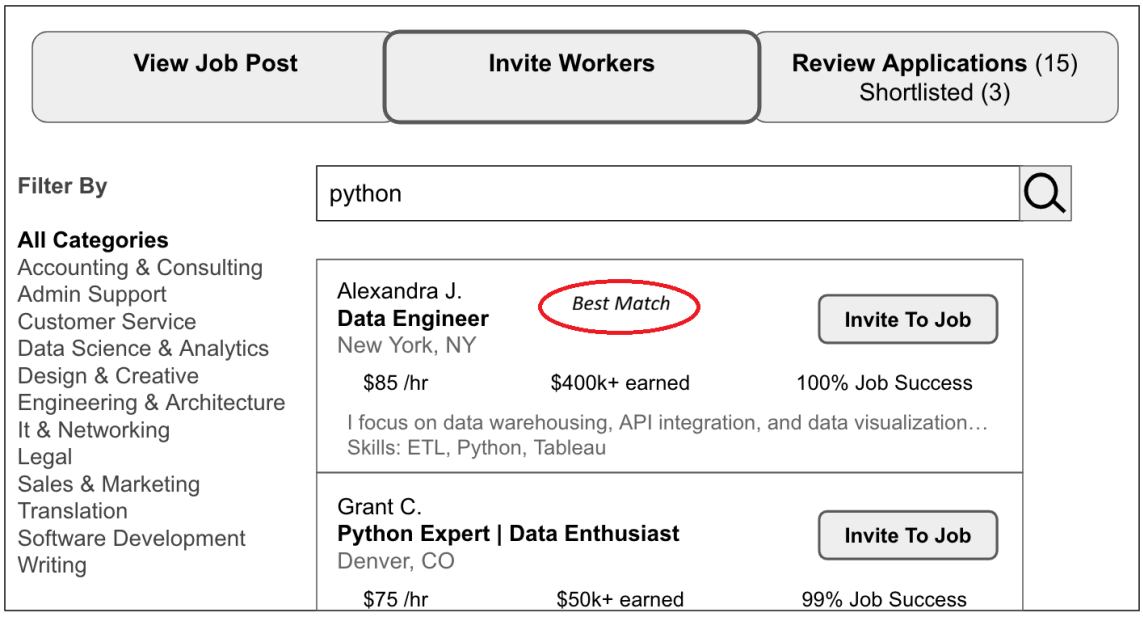
Supplements Section 2: the employer-facing platform interface.

A.2 Initial Message

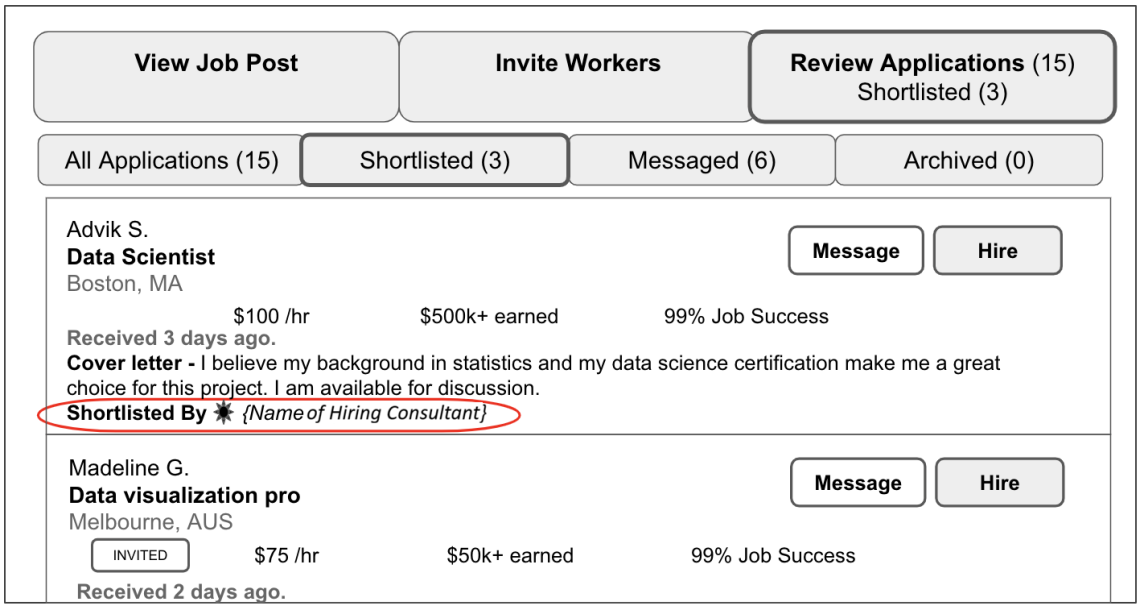
Supplements Section 2.3: the template message treated employers received.

> 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?

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.

A.3 Internal Validity

Supports the Estimation Framework (Section 2.7). Covariates are pre-randomization, measured on each firm's first experimental post; balance is a design check, not an outcome.

To assess the effectiveness of randomization, in Table A.1 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.1: Balance Table

Variable	Control	Treatment	Difference	p-value	Adj. p (Bootstrap)	Adj. p (Bonf.)	Adj. p (FDR)
First Job Post	0.30	0.30	0.0002	0.97	1	1	0.97
Software	0.58	0.58	0.002	0.70	1	1	0.94
Administrative	0.06	0.06	0.004	0.12	0.79	1	0.75
Low Skill	0.05	0.04	-0.004	0.10	0.71	1	0.75
High Skill	0.48	0.48	-0.001	0.87	1	1	0.94
Num Skills Required	6.14	6.11	-0.02	0.57	1	1	0.94
Large Company	0.02	0.02	-0.0003	0.82	1	1	0.94
Employer length of time on platform (days)	825.17	838.83	13.66	0.24	0.96	1	0.94
Num previous hires	3.57	3.71	0.14	0.57	1	1	0.94
Wages paid to fixed price jobs	22,124.41	23,308.16	1,183.75	0.77	1	1	0.94
Average hourly wage paid	16.36	16.19	-0.17	0.46	1.00	1	0.94
Total hours from previous hires	30,282.10	27,657.72	-2,624.38	0.79	1	1	0.94

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. Because the share of posts assigned to treatment varied over the experimental period, attributes are demeaned by posting day (with the overall mean added back), so the comparison reflects balance conditional on posting date — the level at which assignment was randomized. Adjusted p-values control the familywise error rate following List, Shaikh, and Xu (2019) with 5,000 bootstrap replications. Bonferroni and Benjamini-Hochberg (FDR) corrections are also reported.

B Additional tables and figures

B.1 Recruiter selection into each type of treatment among worked jobs

Among posts that Recruiters evaluated, the decision to provide recruiting or shortlisting assistance was endogenous. We characterize this selection to understand how sourcers allocated their effort across posts. To characterize how assisted posts differed from unassisted ones, we restrict the sample to the 40,989 job posts that the Recruiter marked as "worked." Recall that which vacancies are "worked" is random, but which action the recruiter takes conditional on "working" a vacancy is endogenous. We 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 B.1 reports the results. Jobs that received help had fewer applications and fewer shortlists at the time the Recruiter began working on them. They also had fewer employer recruiting invitations and shortlists. This pattern is consistent with recruiters targeting assistance toward jobs that appeared to need it most.

Table B.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	-0.0002 (0.002)
Employer Invites	-0.216^{***} (0.004)
Employer Shortlists	-0.031^{***} (0.006)
Administrative Job	0.302^{***} (0.087)
Software Job	-0.148^{***} (0.041)
Design Job	-0.313^{***} (0.081)
Num Skills Req	-0.007^{*} (0.004)
Low Skill Job	0.219^{**} (0.105)
High Skill Job	-0.059^{*} (0.035)
Constant	2.448^{***} (0.050)
Observations	26,172

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, restricted to each firm's first experimental post (one post per firm). Standard errors are in parentheses. Significance indicators: p $\leqslant$ 0.05 : *, p $\leqslant$0.01 : **, and p $\leqslant$ .001 : ***.

B.2 Intermediate Outcome Tables

Table form of the first-stage results in Section 3, on the first-post main sample.

Table B.2: Effects of treatment on recruiting (first-post sample)

	Total number recruited		Number of employer recruits	
	(1)	(2)	(3)	(4)
Treatment Assigned	3.09^{***}		-0.16^{*}	
	(0.07)		(0.06)	
Treatment Received [IV]		4.59^{***}		-0.23^{*}
		(0.10)		(0.09)
Posting-day FE	Yes	Yes	Yes	Yes
Control mean	3.80	3.80	3.78	3.78
Observations	52,471	52,471	52,471	52,471

Notes: Total number recruited is invitations to apply; employer recruits are invitations sent by the employer (excluding recruiter invitations). Sample is each employer's first post in the experiment. OLS columns regress the outcome on treatment assignment (ITT); IV columns instrument recruiter take-up with treatment assignment (TOT). All specifications include posting-day fixed effects; EHW-robust standard errors in parentheses. $+\ p \leqslant 0.10$, $*\ p \leqslant 0.05$, $**\ p \leqslant 0.01$, $***\ p \leqslant 0.001$.

Table B.3: Effects of treatment assignment on the number of applications (first-post sample)

	Total applications		Recruited applications		Organic applications	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment Assigned	1.36^{***}		1.25^{***}		0.11	
	(0.31)		(0.03)		(0.31)	
Treatment Received [IV]		2.02^{***}		1.86^{***}		0.16
		(0.46)		(0.05)		(0.45)
Posting-day FE	Yes	Yes	Yes	Yes	Yes	Yes
Control mean	25.09	25.09	1.45	1.45	23.61	23.61
Observations	52,471	52,471	52,471	52,471	52,471	52,471

Notes: Recruited applications are applications from workers recruited to the post; organic applications are from workers who were not recruited; total sums the two (raw counts). Sample is each employer's first post in the experiment. OLS columns regress the outcome on treatment assignment (ITT); IV columns instrument recruiter take-up with treatment assignment (TOT). All specifications include posting-day fixed effects; EHW-robust standard errors in parentheses. $+\ p \leqslant 0.10$, $*\ p \leqslant 0.05$, $**\ p \leqslant 0.01$, $***\ p \leqslant 0.001$.

Table B.4: Effects of treatment on selection (first-post sample)

	Num shortlists		Num employer shortlisted	
	(1)	(2)	(3)	(4)
Treatment Assigned	1.91^{***}		0.98^{***}	
	(0.04)		(0.03)	
Treatment Received [IV]		2.84^{***}		1.45^{***}
		(0.05)		(0.05)
Posting-day FE	Yes	Yes	Yes	Yes
Control mean	1.26	1.26	1.26	1.26
Observations	52,471	52,471	52,471	52,471

Notes: Num shortlists is the number of shortlisted applicants; num employer shortlisted is those shortlisted by the employer (excluding recruiter shortlists). Sample is each employer's first post in the experiment. OLS columns regress the outcome on treatment assignment (ITT); IV columns instrument recruiter take-up with treatment assignment (TOT). All specifications include posting-day fixed effects; EHW-robust standard errors in parentheses. $+\ p \leqslant 0.10$, $*\ p \leqslant 0.05$, $**\ p \leqslant 0.01$, $***\ p \leqslant 0.001$.

Table B.5: Effects of treatment assignment on the number of interviews conducted by the employer (first-post sample)

	Total interviews		Recruited interviews		Organic interviews	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment Assigned	0.10^{*}		0.16^{***}		-0.06	
	(0.04)		(0.01)		(0.04)	
Treatment Received [IV]		0.15^{*}		0.24^{***}		-0.09
		(0.06)		(0.02)		(0.06)
Posting-day FE	Yes	Yes	Yes	Yes	Yes	Yes
Control mean	2.13	2.13	0.44	0.44	1.70	1.70
Observations	52,471	52,471	52,471	52,471	52,471	52,471

Notes: Recruited interviews are interviews given by the employer to recruited applicants; organic interviews are given to applicants who were not recruited; total interviews sums the two. Sample is each employer's first post in the experiment. OLS columns regress the outcome on treatment assignment (ITT); IV columns instrument recruiter take-up with treatment assignment (TOT). All specifications include posting-day fixed effects; EHW-robust standard errors in parentheses. $+\ p \leqslant 0.10$, $*\ p \leqslant 0.05$, $**\ p \leqslant 0.01$, $***\ p \leqslant 0.001$.

Table B.6: Effects of treatment on whether the employer made a hire, by channel (first-post main specification)

	Hired anyone		Hired recruited		Hired organic		Hired shortlisted	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment Assigned	0.006		0.015***		−0.003		0.006*	
	(0.005)		(0.003)		(0.004)		(0.003)	
Treatment Received [IV]		0.009		0.022***		−0.004		0.009*
		(0.007)		(0.005)		(0.006)		(0.004)
Posting-day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control mean	0.286	0.286	0.100	0.100	0.159	0.159	0.076	0.076
Observations	52,471	52,471	52,471	52,471	52,471	52,471	52,471	52,471

Notes: Outcomes are indicators for making any hire and for hiring through each channel (recruited, organic, shortlisted). Sample is each employer's first post in the experiment. OLS columns regress the outcome on treatment assignment (ITT); IV columns instrument recruiter take-up with treatment assignment (TOT). All specifications include posting-day fixed effects; EHW-robust standard errors in parentheses. $+\ p \leqslant 0.10$, $*\ p \leqslant 0.05$, $**\ p \leqslant 0.01$, $***\ p \leqslant 0.001$.

Table B.7: Effects of treatment on whether the employer made a hire, by channel (full experimental sample)

	Hired anyone		Hired recruited		Hired organic		Hired shortlisted	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment Assigned	0.004		0.013***		−0.004		0.005*	
	(0.004)		(0.003)		(0.003)		(0.002)	
Treatment Received [IV]		0.006		0.020***		−0.007		0.007*
		(0.006)		(0.004)		(0.005)		(0.003)
Posting-day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control mean	0.293	0.293	0.098	0.098	0.162	0.162	0.074	0.074
Observations	83,017	83,017	83,017	83,017	83,017	83,017	83,017	83,017

Notes: Outcomes are indicators for making any hire and for hiring through each channel (recruited, organic, shortlisted). Sample is the experimental sample of high value posts from October 1, 2021 to April 1, 2022. OLS columns regress the outcome on treatment assignment (ITT); IV columns instrument recruiter take-up with treatment assignment (TOT). All specifications include posting-day fixed effects; standard errors clustered on the employer in parentheses. $+\ p \leqslant 0.10$, $*\ p \leqslant 0.05$, $**\ p \leqslant 0.01$, $***\ p \leqslant 0.001$.

Table B.8: Effects of treatment on whether a job filled, by employer age and first-post status

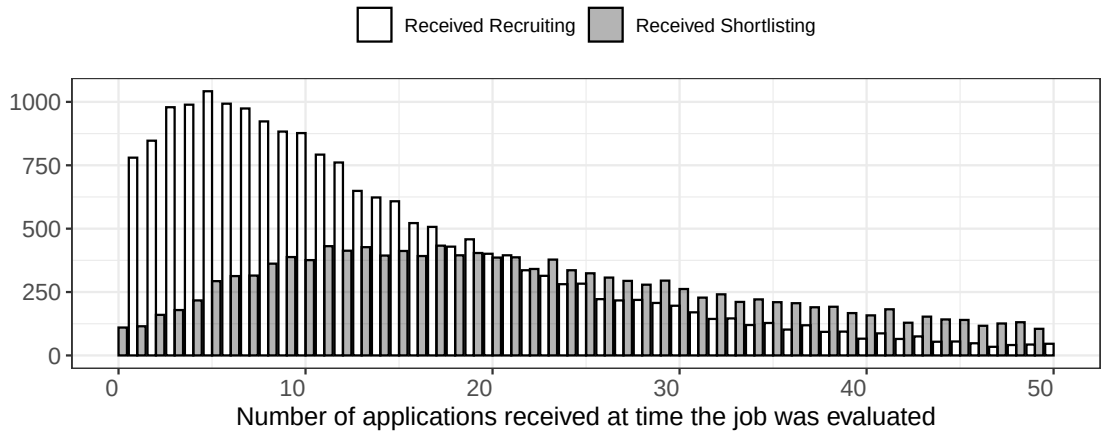
	Hired	
	(1)	(2)
Treatment Assigned	0.004	0.004
	(0.006)	(0.005)
First job post	−0.181***	
	(0.008)	
Treatment × First job post	0.007	
	(0.009)	
Age (weeks)		0.000***
		(0.000)
Treatment × Age		0.000
		(0.000)
Posting-day FE	Yes	Yes
Observations	52,471	52,471

Notes: This table reports the treatment effect on whether a job filled, interacted with employer characteristics. First job post is an indicator for the employer's first job post on the platform; age is weeks since the employer registered. Sample is each firm's first experimental post among the high value posts from October 1, 2021 to April 1, 2022 (one post per firm), which conditions on nothing realized after randomization. Both columns include posting-day fixed effects; heteroskedasticity-robust standard errors in parentheses. $* \, p \leqslant 0.05$, $** \, p \leqslant 0.01$, $*** \, p \leqslant 0.001$.

B.3 Application Distribution by Treatment

Supports Section 3: how the application pool at evaluation predicted recruiting versus shortlisting. Sample: treated posts.

Figure B.1: 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.

B.4 Robustness: Alternative Samples

The main text estimates every result on each firm's first experimental post (Section 2.7). This appendix reproduces them on the *full* experimental sample of all posts, which conditions on nothing realized after randomization but allows for possible spillovers between employer's multiple vacancies (Appendix B.4.1).

B.4.1 Full experimental sample

All experimental posts, standard errors clustered on the employer. This sample conditions on nothing post-randomization; its only vulnerability is spillover between a firm's differently-treated sibling posts. Hiring is a precise null (Table B.9) and the match-quality battery matches the main text except that the 30-day wagebill decline is larger and statistically significant here (Table B.10).

Table B.9: Effects of Treatment on Hiring (full experimental sample)

	(1)	(2)	(3)	(4)
Panel A: OLS (Intent-to-Treat)				
Treatment Effect	0.004	0.004	0.004	0.004
	(0.004)	(0.004)	(0.004)	(0.004)
Panel B: IV (Treatment-on-the-Treated)				
Instrumented Treatment	0.006	0.006	0.007	0.006
	(0.006)	(0.006)	(0.006)	(0.006)
Posting-Day FE	Yes	Yes	Yes	Yes
Category Group FE	No	Yes	No	Yes
Expertise Tier FE	No	No	Yes	Yes
Control Mean	0.293	0.293	0.293	0.293
N (posts)	83017	83017	83017	83017

Notes: 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. All specifications include posting-day fixed effects (the treated share varied over the experimental period) and add category-group and expertise-tier fixed effects as indicated; standard errors are clustered on the employer. Significance indicators: $+\ p \leqslant 0.10$, $*\ p \leqslant 0.05$, $**\ p \leqslant 0.01$, $***\ p \leqslant 0.001$.

Table B.10: Treatment effects on match quality (full experimental sample)

	Hours	Hourly Wages	High Emp Rating	High Worker Rating	Wagebill 30d	Wagebill 90d
Panel A: Post level						
OLS	1.37	−0.33	−0.0069	−0.0139	−72.27*	−51.24
	(6.08)	(0.49)	(0.0070)	(0.0091)	(30.45)	(80.86)
IV	2.12	−0.52	−0.0109	−0.0219	−112.50*	−79.77
	(9.39)	(0.76)	(0.0110)	(0.0143)	(47.40)	(125.88)
Control mean	212.77	32.06	0.9165	0.8593	1,074.29	3,033.29
N (posts)	13,436	16,329	9,198	8,879	21,859	21,859
Panel B: FIML selection-corrected, application level						
FIML	−2.60†	0.08	−0.0057	−0.0166+	−79.37***	−57.12
	(7.45)	(0.48)	(0.0069)	(0.0091)	(23.97)	(58.67)
$\hat{\rho}$	0.38†	−0.40	0.00	0.06	−0.32	−0.18
Control mean	221.21	28.82	0.9155	0.8681	814.73	2,300.12
N (hires)	18,610	23,334	11,071	11,322	29,333	29,333

Notes: The table has two panels estimated on two different samples. *Panel A* is at the post level (one row per post): hours, wages, and ratings are averaged over the post's hires and wagebill is the post's total spend (hourly-wage columns use hourly posts; feedback columns use posts with a rated hire). OLS is the intent-to-treat effect of treatment assignment and IV instruments recruiter take-up with assignment (treatment-on-the-treated). *Panel B* is at the application level, on the full applicant pool: FIML is the Heckman selection model estimated by full-information maximum likelihood, with hiring as the selection stage, the log number of applications as the exclusion restriction, and $\hat{\rho}$ the estimated selection–outcome error correlation. Panel B's unit is the individual hire rather than the post, so its estimand, control mean (over control-group hires), and N (number of hires) are not directly comparable to Panel A (over posts). Sample: the full experimental sample of all high-value posts. Every specification includes posting-day fixed effects and category-group and expertise-tier controls. Standard errors (in parentheses) cluster on the employer. Significance: $+\ p \leqslant 0.10$, $*\ p \leqslant 0.05$, $**\ p \leqslant 0.01$, $***\ p \leqslant 0.001$. †: the FIML hours estimate (-2.60, standard error 7.45, $\hat{\rho} = 0.38$) is the frontier-winsorized value; without winsorization the likelihood rails at the $\hat{\rho} \to 1$ boundary (a tobit-2 degeneracy under hours totals still accumulating at the data pull). The winsorization caps each hire's hours five hours below the second-highest total among feedback-rated engagements starting the same day or later (5.0% of hires capped), and the resulting estimate is a null.

Table B.11: Effect of the treatment on characteristics of hired workers (full sample)

	Observable quality						Channel of the hire	
	Platform rating	Newcomer	Log rate	Log earn.	Log hours	Tenure	Via recruiter	Recommended
Panel A: OLS (Intent-to-Treat)								
Treatment effect	0.002	−0.010	−0.011	−0.001	0.045	0.026	0.035***	−0.018*
	(0.002)	(0.007)	(0.014)	(0.038)	(0.041)	(0.054)	(0.008)	(0.008)
Panel B: IV (Treatment-on-the-Treated)								
Instrumented treatment	0.003	−0.015	−0.017	−0.001	0.069	0.041	0.055***	−0.028*
	(0.003)	(0.011)	(0.021)	(0.059)	(0.063)	(0.084)	(0.013)	(0.013)
Control mean	0.953	0.234	3.397	9.791	6.168	3.960	0.369	0.407
N	17,013	21,891	21,877	21,891	21,891	21,891	21,891	21,891

Notes: Each column is a characteristic of the hired worker. Panel A (OLS) is the intent-to-treat effect from regressing the characteristic on treatment assignment; Panel B (IV) instruments actual sourcing take-up with treatment assignment. Each panel reports the treatment effect with the employer-clustered standard error beneath. Estimates run over the full experimental sample of posts that made a hire (one row per post, the first hire), with posting-day fixed effects. Column headings: Platform rating = platform's score based on work history; Newcomer = indicator (no platform history to rate); Log rate = log profile hourly rate; Log earn. = log prior earnings; Log hours = log prior hours worked; Tenure = platform tenure in years; Via recruiter = the hire came through a recruiter invitation; Recommended = the hired worker carried the algorithmic recommendation. The first six columns describe observable quality (joined from platform profiles); the last two describe the channel of the hire.+ $p \leqslant 0.10$, $*$ $p \leqslant 0.05$, $**$ $p \leqslant 0.01$, $***$ $p \leqslant 0.001$.

Table B.12: Effects of treatment on recruiting (full experimental sample)

	Total number recruited		Number of employer recruits	
	(1)	(2)	(3)	(4)
Treatment Assigned	3.14***		−0.13*	
	(0.07)		(0.06)	
Treatment Received [IV]		4.80***		−0.19*
		(0.10)		(0.09)
Posting-day FE	Yes	Yes	Yes	Yes
Control mean	4.22	4.22	4.16	4.16
Observations	83,017	83,017	83,017	83,017

Notes: Total number recruited is invitations to apply; employer recruits are invitations sent by the employer (excluding recruiter invitations). Sample is experimental sample of high value posts from October 1, 2021 to April 1, 2022. OLS columns regress the outcome on treatment assignment (ITT); IV columns instrument recruiter take-up with treatment assignment (TOT). All specifications include posting-day fixed effects; standard errors clustered on the employer in parentheses. $+$ $p \leqslant 0.10$, $*$ $p \leqslant 0.05$, $**$ $p \leqslant 0.01$, $***$ $p \leqslant 0.001$.

Table B.13: Effects of treatment assignment on the number of applications (full experimental sample)

	Total applications		Recruited applications		Organic applications	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment Assigned	1.27***		1.24***		0.03	
	(0.25)		(0.03)		(0.25)	
Treatment Received [IV]		1.94***		1.90***		0.04
		(0.38)		(0.04)		(0.38)
Posting-day FE	Yes	Yes	Yes	Yes	Yes	Yes
Control mean	24.76	24.76	1.48	1.48	23.24	23.24
Observations	83,017	83,017	83,017	83,017	83,017	83,017

Notes: Recruited applications are applications from workers recruited to the post; organic applications are from workers who were not recruited; total sums the two (raw counts). Sample is experimental sample of high value posts from October 1, 2021 to April 1, 2022. OLS columns regress the outcome on treatment assignment (ITT); IV columns instrument recruiter take-up with treatment assignment (TOT). All specifications include posting-day fixed effects; standard errors clustered on the employer in parentheses. $+\ p \leqslant 0.10$, $*\ p \leqslant 0.05$, $**\ p \leqslant 0.01$, $***\ p \leqslant 0.001$.

Table B.14: Effects of treatment on selection (full experimental sample)

	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***
		(0.04)		(0.04)
Posting-day FE	Yes	Yes	Yes	Yes
Control mean	1.26	1.26	1.26	1.26
Observations	83,017	83,017	83,017	83,017

Notes: Num shortlists is the number of shortlisted applicants; num employer shortlisted is those shortlisted by the employer (excluding recruiter shortlists). Sample is experimental sample of high value posts from October 1, 2021 to April 1, 2022. OLS columns regress the outcome on treatment assignment (ITT); IV columns instrument recruiter take-up with treatment assignment (TOT). All specifications include posting-day fixed effects; standard errors clustered on the employer in parentheses. $+\ p \leqslant 0.10$, $*\ p \leqslant 0.05$, $**\ p \leqslant 0.01$, $***\ p \leqslant 0.001$.

Table B.15: Effects of treatment assignment on the number of interviews conducted by the employer (full experimental sample)

	Total interviews		Recruited interviews		Organic interviews	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment Assigned	0.06		0.15***		−0.09**	
	(0.04)		(0.01)		(0.03)	
Treatment Received [IV]		0.08		0.22***		−0.14**
		(0.06)		(0.02)		(0.05)
Posting-day FE	Yes	Yes	Yes	Yes	Yes	Yes
Control mean	2.20	2.20	0.39	0.39	1.81	1.81
Observations	83,017	83,017	83,017	83,017	83,017	83,017

Notes: Recruited interviews are interviews given by the employer to recruited applicants; organic interviews are given to applicants who were not recruited; total interviews sums the two. Sample is experimental sample of high value posts from October 1, 2021 to April 1, 2022. OLS columns regress the outcome on treatment assignment (ITT); IV columns instrument recruiter take-up with treatment assignment (TOT). All specifications include posting-day fixed effects; standard errors clustered on the employer in parentheses. $+\ p \leqslant 0.10$, $*\ p \leqslant 0.05$, $**\ p \leqslant 0.01$, $***\ p \leqslant 0.001$.

Table B.16: Treatment effect on match quality by expertise tier (full experimental sample)

	Cheap/Inexperienced	Intermediate	Expert/Expensive
30-day wagebill			
Treatment effect	97.55	−50.89	−103.02^{+}
	(166.93)	(34.61)	(54.45)
Control mean	774.88	899.90	1,302.39
N (posts)	1,171	10,937	9,783
High Emp Rating			
Treatment effect	−0.0729^{+}	−0.0001	−0.0102
	(0.0406)	(0.0099)	(0.0109)
Control mean	0.9290	0.9158	0.9132
N (hires)	728	5,456	4,887
High Worker Rating			
Treatment effect	−0.0561	−0.0075	−0.0197
	(0.0399)	(0.0135)	(0.0134)
Control mean	0.8873	0.8553	0.8801
N (hires)	919	5,759	4,644

Notes: Each block is a separate match-quality outcome, estimated within expertise tier on the full experimental sample. The 30-day wagebill is measured over posts that hired (one row per hired post); "High Emp Rating" is an indicator that the score the worker gives the employer is 4 or above and "High Worker Rating" that the rating the employer gives the worker is 4 or above, each over the hires carrying that rating. Each cell regresses the outcome on treatment assignment (intent-to-treat) with category-group and posting-day fixed effects; standard errors clustered on the employer in parentheses. The control mean is the mean outcome among control-group observations in that tier. Significance: $+\ p \leqslant 0.10$, $*\ p \leqslant 0.05$, $**\ p \leqslant 0.01$, $***\ p \leqslant 0.001$.

Table B.17: Treatment effect on match quality by category group (full experimental sample)

	30-day wagebill		High Emp Rating		High Worker Rating	
Category	N	TE (SE)	N	TE (SE)	N	TE (SE)
Web, Mobile & Software Dev	10,118	-69.45 (44.14)	5,209	-0.0074 (0.0092)	4,759	-0.0037 (0.0123)
Customer Service	2,866	-99.94 (73.15)	1,484	-0.0073 (0.0271)	1,999	-0.0345 (0.0322)
Admin Support	2,155	3.95 (70.03)	1,044	0.0061 (0.0277)	1,248	-0.0303 (0.0362)
IT & Networking	1,586	-79.90 (166.88)	602	0.0163 (0.0370)	563	0.0236 (0.0432)
Design & Creative	1,461	-84.10 (56.79)	969	0.0135 (0.0281)	883	-0.0176 (0.0288)
Accounting & Consulting	1,358	-44.64 (153.21)	601	-0.0021 (0.0407)	669	-0.0679^{+} (0.0371)
Data Science & Analytics	1,108	-240.17 (212.06)	513	0.0177 (0.0415)	489	0.0625 (0.0536)
Writing	480	-198.15 (303.25)	280	-0.0676 (0.0424)	315	-0.0909 (0.0644)
Engineering & Architecture	404	97.55 (276.31)	172	-0.0684 (0.0705)	186	0.1840 (0.1609)
Sales & Marketing	323	164.24 (567.29)	180	0.0418 (0.0693)	195	-0.0425 (0.0988)

Notes: Each row is a category group; each outcome reports the number of observations (N), the treatment effect (TE), and its standard error (SE, in parentheses). The 30-day wagebill is measured over posts that hired, so its N is hired posts; the ratings are measured over the hires carrying that rating, so their N is smaller. "High Emp Rating" is an indicator that the score the worker gives the employer is 4 or above and "High Worker Rating" that the rating the employer gives the worker is 4 or above, each over the hires carrying that rating. Each cell regresses the outcome on treatment assignment (intent-to-treat) with posting-day fixed effects on the full experimental sample; standard errors clustered on the employer. Rows are ordered by the number of hired posts in the category. Category groups with fewer than 100 hired posts, Translation and Legal, are removed. Significance: $+$ $p \leqslant 0.10$, $*$ $p \leqslant 0.05$, $**$ $p \leqslant 0.01$, $***$ $p \leqslant 0.001$.

B.5 Neither Recruiting Nor Shortlisting Improved Hiring

Full propensity-score analysis behind the recruiting-versus-shortlisting null (Section 4); first-post main sample.

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 B.2 in Appendix B.6 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 B.18 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 B.18: Effect of Hiring Assistance on Job Fill Rate

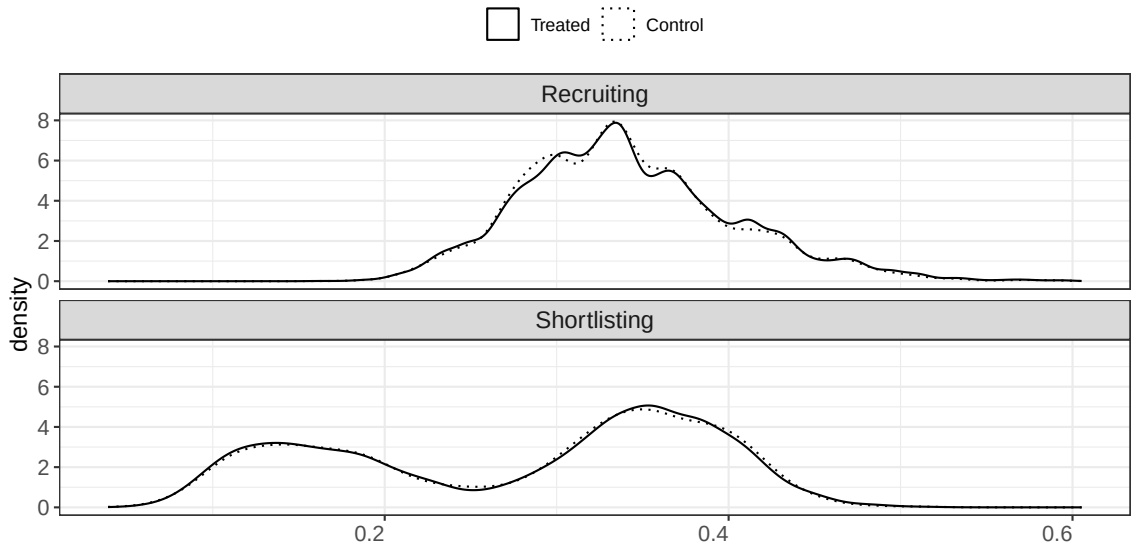
	Unweighted		Weighted	
	Recruiting (1)	Shortlisting (2)	Recruiting (3)	Shortlisting (4)
Panel A: OLS (Intent-to-Treat)				
Treatment Assigned	0.005	0.006	0.005	0.004
	(0.005)	(0.005)	(0.005)	(0.005)
Panel B: IV (Treatment-on-the-Treated)				
Recruiting Help [IV]	0.015		0.014	
	(0.014)		(0.015)	
Shortlisting Help [IV]		0.006		0.005
		(0.015)		(0.015)
Posting-day FE	Yes	Yes	Yes	Yes
Observations	50114	50114	50114	50114

Notes: 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 standard errors. Significance indicators: $+\ p \leqslant 0.10$, $*\ p \leqslant 0.05$, $**\ p \leqslant 0.01$, $***\ p \leqslant 0.001$. Sample restricts to each firm's first experimental post (one post per firm), which conditions on nothing realized after randomization.

B.6 Propensity Score Distributions

Supports the propensity-score design in Appendix B.5.

Figure B.2: 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.

B.7 Recruited vs. Organic Applicants

Supports Section 3: observable characteristics of recruited versus organic applicants. Sample: applications to treated posts, first-post main sample.

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 B.19 we compare Recruiter-invited applications to organic (uninvited) applications to jobs in the experimental sample; employer-invited applications that were not recruiter-sourced are excluded so the baseline is genuinely organic. Column (1) shows that recruited workers bid roughly three dollars more on the application than organic (uninvited) applicants, despite Column (2) showing that their posted profile rate is about two dollars *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.

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.

Table B.19: Comparison of recruiter-sourced applicants to other applicants on treated job posts

	Wage bid	Profile rate	Total earned	Hours worked	Previous invites
Recruited worker	2.90^{***}	-2.00^{***}	-14818.00^{***}	810.14^{***}	216.67^{***}
	(0.07)	(0.07)	(612.33)	(9.29)	(2.05)
Control mean (organic)	28.72	35.90	103010.20	2049.86	332.25
Posting-day FE	Yes	Yes	Yes	Yes	Yes
Observations	952,181	1,534,361	1,534,361	1,534,361	1,534,361

Notes: Each column regresses a pre-hire observable of the applicant on an indicator for being recruiter-sourced (*Recruited worker*), comparing recruiter-sourced applications to organic (uninvited) applications on treated posts; employer-invited applications that were not recruiter-sourced are excluded so the baseline is genuinely organic. Column (1) is the wage bid; columns (2)–(5) are the applicant's platform history. It uses all applications to jobs in the treated group of the experimental sample, restricted to each firm's first experimental post (one post per firm). All columns include posting-day fixed effects; the control mean is the organic-applicant mean. Heteroskedasticity-robust standard errors in parentheses. $+\ p \leqslant 0.10$, $*\ p \leqslant 0.05$, $**\ p \leqslant 0.01$, $***\ p \leqslant 0.001$.

B.8 Regression Discontinuity

Supports Section 6.1 (heterogeneity). All RD estimates use the full experimental sample and the 30-day wagebill. The RD identifies off the recommendation label crossing a threshold within a post's applicant pool and does not use the experimental assignment, so the first-post restriction that defines the main specification elsewhere in the paper is not required here; imposing it discards most of the applications belonging to experienced, repeat-posting firms and leaves the subgroup estimates too imprecise to read. The first-post variant is reported in the replication materials.

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 6.1.

Density and covariate balance. Figure B.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.

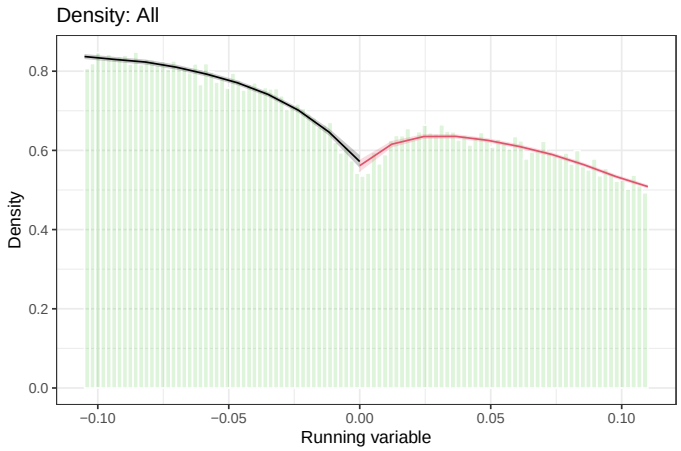


Figure B.3: Density test for the RD running variable (full sample)

Notes: McCrary-style density test (Cattaneo et al., 2020) for the recommendation score running variable. The density exhibits no discontinuity at the recommendation threshold, providing no evidence of sorting around the cutoff.

Figures B.4 shows that log application count is smooth through the recommendation threshold, further supporting design validity.

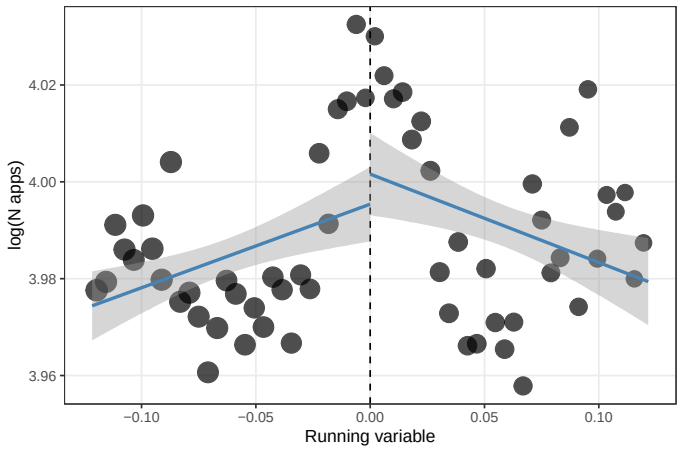


Figure B.4: RD covariate balance: log(N applications)

Notes: Binned scatter plot of log(number of applications) against the recommendation score running variable. There is no discontinuity at the recommendation threshold, providing no evidence of sorting around the cutoff.

First stage. Figure B.5 shows a sharp discontinuity at the threshold: the recommendation label changes discretely at the score cutoff.

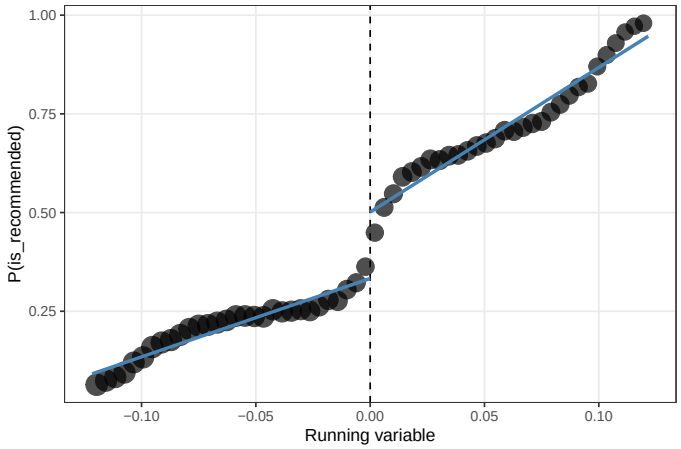


Figure B.5: First stage: recommendation probability at the score threshold

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 B.6 shows the RD relationship for 30-day wagebill among hired workers.

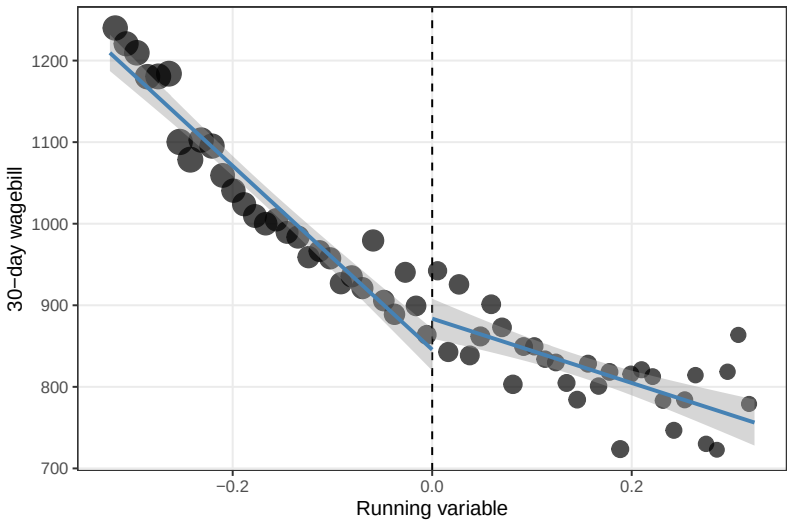


Figure B.6: RD: 30-day wagebill at the recommendation threshold

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 B.22.

Hourly rate at the threshold. Table B.20 reports fuzzy RD estimates for the applicant's hourly rate at the recommendation threshold (defined only for hourly jobs). If the recommendation label were correlated with unobserved applicant quality, we would expect a discontinuity in posted rates. The absence of a significant jump provides additional support for the validity of the design.

Table B.20: Fuzzy RD: Hourly Rate (hourly)

	(1) All	(2) Treated	(3) Control
Panel A: rdrobust (bias-corrected, robust SE)			
Recommended	1.047	0.247	-4.451
	(7.316)	(7.036)	(8.447)
N (in BW)	85,765	74,543	35,155
Bandwidth h	0.053	0.061	0.086
χ^2 p-value		0.669	
Panel B: RDHonest (Armstrong & Kolesár, 2020)			
Recommended	1.178	2.722	1.913
	(5.291)	(5.736)	(5.243)
95% CI	[-10.564, 12.920]	[-10.154, 15.598]	[-9.823, 13.648]
Eff. N	206,626	158,649	85,433
Bandwidth h	0.153	0.156	0.237
χ^2 p-value		0.917	

Notes: Outcome: hourly rate ($). Sample restricted to applications with a posted hourly rate. Panel A reports bias-corrected fuzzy RD estimates with robust standard errors. Panel B reports fuzzy RD via RDHonest; CIs are honest (bias-aware). SEs clustered by post. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Baseline RD regression tables. Tables B.21–B.23 report fuzzy RD estimates for hiring, wagebill, and shortlisting using the baseline threshold definition. The corresponding binned scatter plot for hiring appears in Section 6.1.

Table B.21: 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)	(0.006)	(0.006)
Any shortlisting				0.001	0.001
				(0.035)	(0.035)
N (in BW)	197,665	178,301	71,406	197,665	197,665
Bandwidth h	0.095	0.113	0.141	0.095	0.095

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 B.22: Fuzzy RD: Wagebill

	RD (bias-corrected, robust SE)			2SLS w/ shortlisting	
	(1) All	(2) Treated	(3) Control	(4) Base	(5) +Tier
Recommended	327.340	188.337	836.972	683.125	676.594
	(484.696)	(537.157)	(1008.329)	(740.459)	(747.243)
Any shortlisting				-1917.959	-2036.563
				(2038.514)	(2032.721)
N (in BW)	2,382	1,651	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 B.23: 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	0.007
	(0.012)	(0.014)	(0.024)	(0.009)
N (in BW)	252,982	208,021	78,719	179,946
Bandwidth h	0.122	0.132	0.154	0.120

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 ($p = 0.979$). SEs clustered by post. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

OLS and selection-corrected estimates. Tables B.24–B.25 report OLS, Heckit, and MLE estimates for the relationship between the recommendation score, recommendation status, and outcomes.

Table B.24: 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	1,526,615	1,526,615	1,526,615	1,526,615	1,526,615	1,526,615	1,526,615	1,526,615

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 B.25: Wagebill: OLS, Heckit, MLE

	OLS		Heckit		MLE	
	(1)	(2)	(3)	(4)	(5)	(6)
Controls	Base	+Tier	Base	+Tier	Base	+Tier
s_j	-41.375***	-27.922*	-48.825**	-35.910*	-56.304***	-30.389
	(12.350)	(12.215)	(15.202)	(15.202)	(16.085)	(16.065)
IMR (λ)			118.658	121.816		
			(75.441)	(75.062)		
$\tanh^{-1}(\hat{\rho})$					-0.034	-0.022
					(0.045)	(0.044)
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 (Heckman, 1979). Columns (5)–(6) estimated by joint maximum likelihood (MLE) of the selection and outcome equations (Heckman, 1974). SEs clustered by post (OLS) or from the model. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Table B.26: 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	1,105,935	1,105,935	1,105,935	1,105,935

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: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

B.8.1 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 B.7–B.8 show binned scatter plots and

Tables B.27–B.29 report regression estimates. The hiring effect is positive and statistically significant in the treated arm. The shortlisting effect is statistically significant.

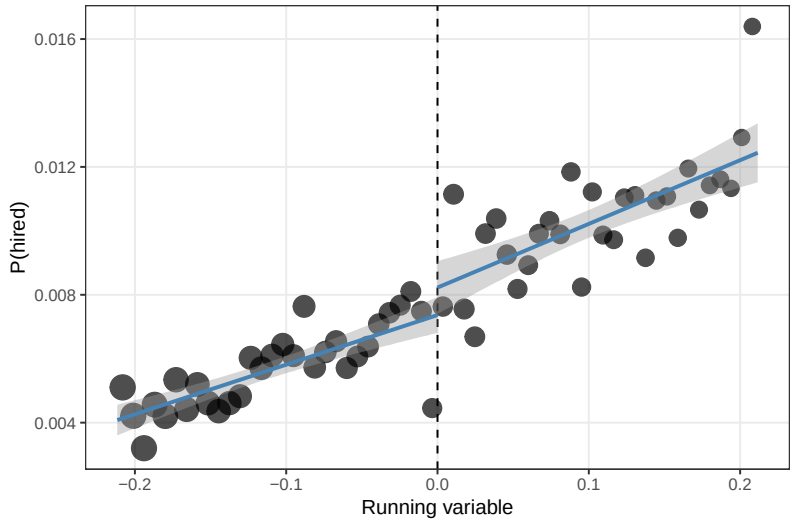


Figure B.7: RD: hiring probability (threshold definition 2)

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.

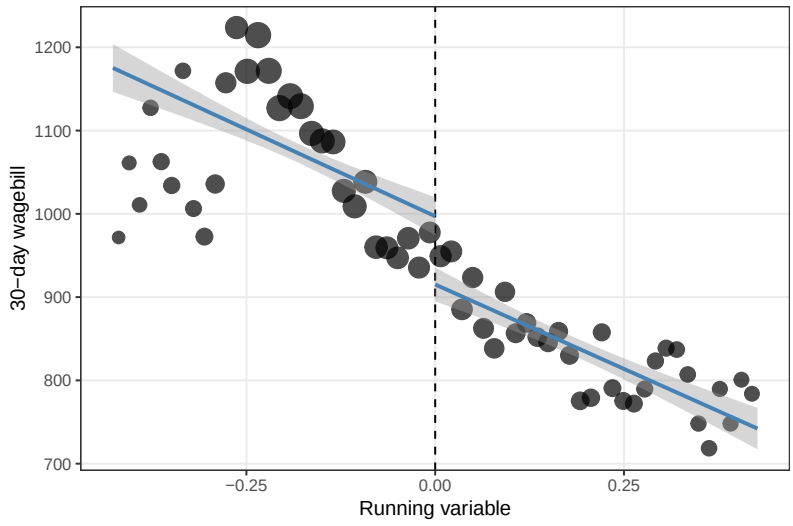


Figure B.8: RD: 30-day wagebill (threshold definition 2)

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 B.27: 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				0.006	0.006
				(0.021)	(0.021)
N (in BW)	529,036	383,040	115,955	529,036	529,036
Bandwidth h	0.212	0.204	0.190	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 B.28: 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	131.812	-30.915	381.795	50.101	37.065
	(212.071)	(245.398)	(486.296)	(210.640)	(214.002)
Any shortlisting				-1557.633	-1701.416
				(2021.371)	(2011.082)
N (in BW)	2,773	1,965	621	2,773	2,773
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 B.29: RD: Recommendation and Shortlisting (threshold definition 2)

	Any shortlisting			Shortlisted by recruiter
	(1) All	(2) Treated	(3) Control	(4) Treated only
Recommended	0.031***	0.041***	0.007	0.033***
	(0.003)	(0.003)	(0.005)	(0.002)
N (in BW)	434,859	232,443	169,761	172,822
Bandwidth h	0.177	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 B.9–B.10 show binned scatter plots and Tables B.30–B.32 report regression estimates. The hiring effect is positive and statistically significant in the pooled sample. The shortlisting effect is also statistically significant.

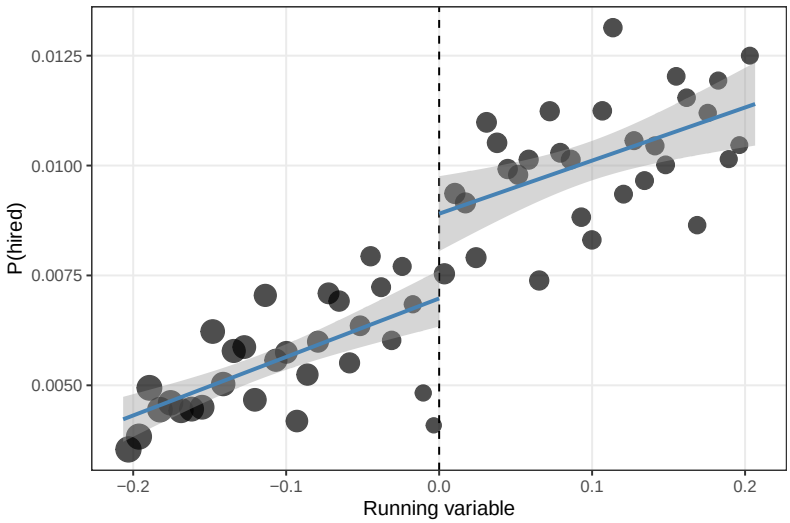


Figure B.9: RD: hiring probability (threshold definition 3)

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

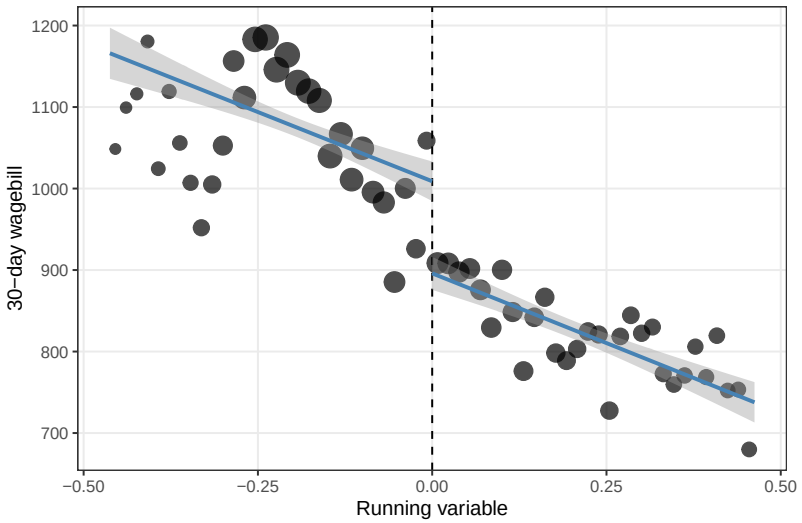


Figure B.10: RD: 30-day wagebill (threshold definition 3)

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 B.30: 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				0.001	0.001
				(0.026)	(0.025)
N (in BW)	459,304	407,810	113,938	459,304	459,304
Bandwidth h	0.207	0.238	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 B.31: 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	-98.526	-238.181	190.468	-67.642	-56.991
	(241.672)	(286.098)	(452.802)	(225.437)	(227.541)
Any shortlisting				-152.149	-256.478
				(1074.028)	(1056.998)
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 ($p = 0.424$). SEs clustered by post. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Table B.32: RD: Recommendation and Shortlisting (threshold definition 3)

	Any shortlisting			Shortlisted by recruiter
	(1) All	(2) Treated	(3) Control	(4) Treated only
Recommended	0.017***	0.021***	0.003	0.019***
	(0.003)	(0.003)	(0.007)	(0.002)
N (in BW)	788,420	669,524	154,184	392,808
Bandwidth h	0.377	0.479	0.270	0.239

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 B.33–B.35 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 B.33: 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	889,616	536,382	207,939
Bandwidth h	0.592	0.380	0.513

Notes: Outcome: hired (0/1). Fuzzy RD via RDHonest (Armstrong & Kolesár, 2020). 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 ($p = 0.624$). Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Table B.34: RDHonest Fuzzy RD: Wagebill

	(1) All	(2) Treated	(3) Control
Recommended	159.389	121.205	572.670
	(231.801)	(231.414)	(1659.293)
95% CI	[-367.895, 686.673]	[-398.321, 640.731]	[-3245.764, 4391.105]
Max. bias	141.542	133.417	1062.524
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, 2020). 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 B.35: 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	499,199	322,382	209,692	330,864
Bandwidth h	0.269	0.235	0.524	0.250

Notes: Fuzzy RD via RDHonest (Armstrong & Kolesár, 2020). 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 ($p = 0.931$). EHW SEs clustered by post. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

B.9 RD Heterogeneity by Threshold, Pool Size, and Employer Characteristics

Extends Appendix B.8; first-post sample.

Table B.36 shows RD estimates for hiring by application-count quartile. Recommendation has a significant positive effect on hiring in the lowest quartile (0.054, se 0.020, $p = 0.008$, FDR $q = 0.037$) and no detectable effect in the upper three quartiles. Table B.37 shows a parallel pattern for shortlisting. The wagebill estimates in Table B.38 are noisy but directionally consistent: positive in Q1 and negative in Q3–Q4.

RD estimates for hiring are near zero across all threshold quartiles (Table B.39), and threshold placement produces no differential hiring or wagebill effects (Tables B.40 and B.41).

We further split the recommendation discontinuity along three employer-side dimensions that proxy for how much an employer relies on the platform's algorithm rather than her own judgment: prior hiring experience (the number of hires completed before the experiment), company size tier, and platform tenure (time since account registration). Hiring, shortlisting, and wagebill estimates appear in Tables B.42–B.50. The positive hiring response to recommendation is driven by smaller and less experienced employers: it is positive and significant for firms with no prior hires and for the smallest size tier (1–10), and undetectable for larger or more experienced firms. The same pattern holds for platform tenure, where only the newest employers show a response. As with the pool-size split, this describes where the effect concentrates rather than a test that the subgroup coefficients differ, which the data are not powered to establish. It is consistent with the thin-pool mechanism: employers with the least in-house capacity to evaluate candidates lean most on the algorithmic label.

Table B.36: RD by Application Count Quartile: hired (0/1)

App-count quartile	RD estimate	FDR q	N (in BW)	BW h
Q1 (low)	0.054**	0.037	39,190	0.073
	(0.020)			
Q2	0.019	0.372	40,016	0.076
	(0.013)			
Q3	-0.003	0.771	52,163	0.103
	(0.006)			
Q4 (high)	-0.000	0.970	98,666	0.188
	(0.008)			

Notes: Outcome: hired (0/1). Fuzzy RD bias-corrected estimates; robust SEs in parentheses. SEs clustered by post. FDR q is the Benjamini-Hochberg adjusted p-value, computed jointly across every hiring cell in all five heterogeneity splits reported in this appendix. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Table B.37: RD by Application Count Quartile: any shortlisting (0/1)

App-count quartile	RD estimate	N (in BW)	BW h
Q1 (low)	0.051	39,904	0.074
	(0.048)		
Q2	-0.030	41,257	0.078
	(0.034)		
Q3	-0.018	46,255	0.092
	(0.021)		
Q4 (high)	0.050*	50,070	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 B.38: RD by Application Count Quartile: 30-day wagebill

App-count quartile	RD estimate	N (in BW)	BW h
Q1 (low)	1236.903**	669	0.406
	(436.042)		
Q2	130.077	834	0.532
	(748.043)		
Q3	-888.363	595	0.299
	(1009.669)		
Q4 (high)	-261.778	828	0.561
	(726.003)		

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 B.39: RD by Threshold Quartile: hired (0/1)

Threshold quartile	RD estimate	FDR q	N (in BW)	BW h
Q1 (low)	0.007	0.372	27,156	0.066
	(0.005)			
Q2	0.028	0.708	28,755	0.048
	(0.047)			
Q3	0.779	0.282	62,714	0.100
	(0.431)			
Q4 (high)	-0.000	0.922	157,109	0.375
	(0.003)			

Notes: Outcome: hired (0/1). Fuzzy RD bias-corrected estimates; robust SEs in parentheses. SEs clustered by post. FDR q is the Benjamini-Hochberg adjusted p-value, computed jointly across every hiring cell in all five heterogeneity splits reported in this appendix. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Table B.40: RD by Threshold Quartile: any shortlisting (0/1)

Threshold quartile	RD estimate	N (in BW)	BW h
Q1 (low)	-0.040 (0.022)	22,565	0.055
Q2	0.001 (0.099)	36,976	0.061
Q3	16.990 (18.030)	57,798	0.092
Q4 (high)	0.008 (0.004)	230,271	0.599

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 B.41: RD by Threshold Quartile: 30-day wagebill

Threshold quartile	RD estimate	N (in BW)	BW h
Q1 (low)	-395.590 (700.840)	471	0.301
Q2	80.919 (1381.974)	743	0.289
Q3	1103.382* (443.017)	712	0.379
Q4 (high)	97.678 (289.935)	747	0.597

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 B.42: RD by Employer Prior-Hire Experience: hired (0/1)

Prior hires (before experiment)	RD estimate	FDR q	N (in BW)	BW h
0	0.014**	0.037	83,399	0.095
	(0.005)			
1-2	0.049**	0.029	32,517	0.106
	(0.017)			
3+	-0.003	0.708	285,775	0.294
	(0.005)			

Notes: Outcome: hired (0/1). Fuzzy RD bias-corrected estimates; robust SEs in parentheses. SEs clustered by post. FDR q is the Benjamini-Hochberg adjusted p-value, computed jointly across every hiring cell in all five heterogeneity splits reported in this appendix. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Table B.43: RD by Employer Prior-Hire Experience: any shortlisting (0/1)

Prior hires (before experiment)	RD estimate	N (in BW)	BW h
0	0.004	100,613	0.114
	(0.013)		
1-2	0.091*	34,679	0.114
	(0.046)		
3+	-0.000	126,658	0.141
	(0.029)		

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 B.44: RD by Employer Prior-Hire Experience: 30-day wagebill

Prior hires (before experiment)	RD estimate	N (in BW)	BW h
0	661.515	815	0.331
	(621.745)		
1-2	86.668	440	0.382
	(730.154)		
3+	60.272	1,395	0.395
	(770.815)		

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 B.45: RD by Company Size Tier: hired (0/1)

Company size tier	RD estimate	FDR q	N (in BW)	BW h
	0.017	0.532	32,202	0.155
	(0.014)			
Tier 1: 1 - 10	0.015***	0.017	149,847	0.122
	(0.005)			
Tier 2: 11 - 50	-0.011	0.708	70,634	0.160
	(0.017)			
Tier 3: 51 - 200	0.023	0.686	31,122	0.217
	(0.026)			
Tier 4: 201 - 500	-0.068	0.686	4,008	0.119
	(0.077)			
Tier 5: 501 - 1000	0.035	0.694	3,635	0.219
	(0.046)			
Tier 6: 1001 - 5000	-0.053	0.372	3,649	0.269
	(0.038)			
Tier 7: 5001 - 10000	0.022	0.694	268	0.177
	(0.028)			
Tier 8: 10001 +	0.077	0.372	1,264	0.303
	(0.050)			

Notes: Outcome: hired (0/1). Fuzzy RD bias-corrected estimates; robust SEs in parentheses. SEs clustered by post. FDR q is the Benjamini-Hochberg adjusted p-value, computed jointly across every hiring cell in all five heterogeneity splits reported in this appendix. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Table B.46: RD by Company Size Tier: any shortlisting (0/1)

Company size tier	RD estimate	N (in BW)	BW h
	0.015	22,841	0.109
	(0.042)		
Tier 1: 1 - 10	0.014	166,212	0.135
	(0.013)		
Tier 2: 11 - 50	-0.023	49,238	0.113
	(0.041)		
Tier 3: 51 - 200	0.078	19,382	0.142
	(0.108)		
Tier 4: 201 - 500	-0.017	5,000	0.147
	(0.217)		
Tier 5: 501 - 1000	0.104	2,556	0.160
	(0.172)		
Tier 6: 1001 - 5000	-0.027	2,509	0.187
	(0.132)		
Tier 7: 5001 - 10000	0.332	283	0.192
	(0.269)		
Tier 8: 10001 +	-0.208	1,314	0.314
	(0.160)		

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 B.47: RD by Company Size Tier: 30-day wagebill

Company size tier	RD estimate	N (in BW)	BW h
	-528.959 (1550.841)	239	0.348
Tier 1: 1 - 10	-366.036 (374.473)	1,827	0.538
Tier 2: 11 - 50	1364.105 (982.862)	652	0.424
Tier 3: 51 - 200	954.749 (739.689)	232	0.501
Tier 4: 201 - 500	-1183.694 (3515.964)	59	0.311
Tier 5: 501 - 1000	889.168 (2414.712)	26	0.368
Tier 6: 1001 - 5000	3331.564 (4367.779)	23	0.541
Tier 7: 5001 - 10000	–	–	–
Tier 8: 10001 +	–	–	–

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 B.48: RD by Employer Platform Tenure Quartile: hired (0/1)

Employer platform tenure (quartile)	RD estimate	FDR q	N (in BW)	BW h
Q1 (low)	0.019** (0.006)	0.017	42,058	0.087
Q2	-0.001 (0.004)	0.852	197,391	0.351
Q3	0.010 (0.013)	0.694	53,087	0.100
Q4 (high)	-0.002 (0.011)	0.922	62,771	0.115

Notes: Outcome: hired (0/1). Fuzzy RD bias-corrected estimates; robust SEs in parentheses. SEs clustered by post. FDR q is the Benjamini-Hochberg adjusted p-value, computed jointly across every hiring cell in all five heterogeneity splits reported in this appendix. Significance: * $p \leqslant 0.05$, ** $p \leqslant 0.01$, *** $p \leqslant 0.001$.

Table B.49: RD by Employer Platform Tenure Quartile: any shortlisting (0/1)

Employer platform tenure (quartile)	RD estimate	N (in BW)	BW h
Q1 (low)	-0.001 (0.017)	41,226	0.085
Q2	-0.001 (0.032)	66,190	0.128
Q3	-0.016 (0.039)	44,655	0.084
Q4 (high)	0.060 (0.034)	50,604	0.092

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 B.50: RD by Employer Platform Tenure Quartile: 30-day wagebill

Employer platform tenure (quartile)	RD estimate	N (in BW)	BW h
Q1 (low)	400.261 (450.804)	664	0.406
Q2	399.307 (565.345)	737	0.457
Q3	-2434.609* (955.901)	763	0.414
Q4 (high)	2211.963 (1197.902)	656	0.380

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

B.10 Multiple Hypothesis Testing

Supports Section 6: multiple-hypothesis correction for the subgroup hiring effects; first-post main sample.

Tables B.51 and B.52 report multiple-hypothesis-corrected p-values for the subgroup hiring effects shown in Figures 7 and 8. Corrections use bootstrap stepdown (List et al., 2019), Bonferroni, and Benjamini–Hochberg FDR procedures (Benjamini and Hochberg, 1995).

Table B.51: Multiple Hypothesis Testing: Treatment Effect on Hiring by Category Group (first-post main specification)

				p-values			
Subgroup	N	TE	(SE)	Unadj.	Stepdown	Bonf.	FDR (BH)
Admin Support	3,155	0.033	(0.022)	0.1199	0.7183	1.0000	0.7892
Customer Service	5,966	0.020	(0.014)	0.1578	0.7293	1.0000	0.7892
Writing	715	0.062	(0.052)	0.2398	0.8701	1.0000	0.7917
Web, Mobile & Software Dev	30,633	0.004	(0.006)	0.5285	0.9930	1.0000	0.7917
Accounting & Consulting	2,312	0.014	(0.025)	0.5724	0.9960	1.0000	0.7917
IT & Networking	3,579	−0.001	(0.019)	0.9690	0.9960	1.0000	0.9690
Data Science & Analytics	1,966	−0.001	(0.027)	0.9471	0.9960	1.0000	0.9690
Design & Creative	2,636	−0.011	(0.023)	0.6334	0.9960	1.0000	0.7917
Engineering & Architecture	878	−0.020	(0.042)	0.6164	0.9960	1.0000	0.7917
Sales & Marketing	598	−0.030	(0.058)	0.5854	0.9960	1.0000	0.7917

Notes: Sample is each firm's first experimental post with posting-day fixed effects, matching the main hiring specification; TE is the within-subgroup regression treatment effect and SE its employer-clustered standard error. p-values are bootstrap-t: the test statistic is the studentized within-subgroup estimate, and the null distribution comes from 1,000 employer-level pairs-bootstrap replications recentered at the sample estimate, drawn jointly across subgroups. Stepdown is the Romano–Wolf / List-Shaikh-Xu max-t procedure, which controls the familywise error rate while exploiting dependence across subgroups; Bonferroni and Benjamini–Hochberg (FDR) adjustments of the unadjusted bootstrap p-values are shown for comparison. $+$, $*$, $**$, $***$ indicate significance at the 10%, 5%, 1%, and 0.1% levels.

C Theory Appendix

This appendix provides formal statements and proofs of the theoretical results in Section 5. Section C.1 fixes notation and derives the joint distribution of signals. Section C.2 gives the closed-form expression for ω (Proposition 3). Section C.3 proves the hiring-probability result (Result 1) and the match-quality decomposition (equation 8). Section C.4 sets out the employer's threshold rule. Section C.5 proves Propositions 1 and 2, which hold for every parameter value. Section C.6 calibrates the model and checks Results 2 and 3 numerically.

Table B.52: Multiple Hypothesis Testing: Treatment Effect on Hiring by Expertise Tier (first-post main specification)

				p-values			
Subgroup	N	TE	(SE)	Unadj.	Stepdown	Bonf.	FDR (BH)
Cheap/Inexperienced	2,327	0.051	(0.023)	0.0210*	0.0799+	0.0629+	0.0629+
Expert/Expensive	25,281	0.008	(0.007)	0.2468	0.4466	0.7403	0.3701
Intermediate	24,851	−0.000	(0.007)	0.9790	0.9790	1.0000	0.9790

Notes: Sample is each firm's first experimental post with posting-day fixed effects, matching the main hiring specification; TE is the within-subgroup regression treatment effect and SE its employer-clustered standard error. p-values are bootstrap-t: the test statistic is the studentized within-subgroup estimate, and the null distribution comes from 1,000 employer-level pairs-bootstrap replications recentered at the sample estimate, drawn jointly across subgroups. Stepdown is the Romano–Wolf / List-Shaikh-Xu max-t procedure, which controls the familywise error rate while exploiting dependence across subgroups; Bonferroni and Benjamini–Hochberg (FDR) adjustments of the unadjusted bootstrap p-values are shown for comparison. $+$, $*$, $**$, $***$ indicate significance at the 10%, 5%, 1%, and 0.1% levels.

C.1 Primitives and Notation

Worker j's quality is $v_j = \alpha_j + \varepsilon_j$. The platform score captures $\alpha_j \sim N(0, \sigma_\alpha^2)$ and misses $\varepsilon_j \sim N(0, \sigma_\varepsilon^2)$, with $\alpha_j \perp \varepsilon_j$. The score is $s_j = \alpha_j + \psi_j$, $\psi_j \sim N(0, \sigma_\psi^2)$. The recruiter and the employer each observe one composite impression, $q_j^R = v_j + \psi_j + \xi_j^R$ and $q_j^E = v_j + \psi_j + \xi_j^E$, with $\xi_j^R \sim N(0, \sigma_{\xi_R}^2)$, $\xi_j^E \sim N(0, \sigma_{\xi_E}^2)$, $\mathrm{Cov}(\xi^R, \xi^E) = \sigma_{\xi_{RE}}$, $\mathrm{Cov}(\psi, \xi^R) = \sigma_{\psi\xi_R}$ and $\mathrm{Cov}(\psi, \xi^E) = \sigma_{\psi\xi_E}$. The noise terms are uncorrelated with α_j and ε_j, primitives are independent across candidates, and neither party observes the components of its composite separately.

Pools and thresholds. In Section 5.1 the algorithmic pool is $\mathcal{A} = \{s_j \geqslant \bar{s}\}$; the calibration takes $\mathcal{A}$ to be every organic applicant, and Appendices C.2 and C.3 give both. The recruiter screens $\{s_j < \bar{s}\}$ and invites m workers with $q_j^R \geqslant \bar{r}$, drawn at random from those clearing the cutoff, so invitees are independent draws from the recruited-pool distribution. Let $n = |\mathcal{A}|$. The employer reviews applications in random order at a positive cost per review (Appendix C.4) and hires the first with $q_j^E \geqslant \bar{u}$; a review is a reading of the proposal and profile, not the platform's recorded interview. Her threshold is endogenous and Appendix C.6 solves it, but the closed forms below hold for any given $\bar{u}$. Standardized thresholds are $z_s = \bar{s}/\sigma_s$, $z_{\mathcal{R}} = \bar{r}/\sigma_R$ and $z_E = \bar{u}/\sigma_E$.

Lemma 1 (Joint distribution)**.** (v_j, q_j^R, q_j^E, s_j) *is jointly normal with mean zero and*

$$\mathrm{Var}\begin{pmatrix} v_j \\ q_j^R \\ q_j^E \\ s_j \end{pmatrix} = \begin{pmatrix} \sigma_v^2 & \sigma_v^2 & \sigma_v^2 & \sigma_\alpha^2 \\ \sigma_v^2 & \sigma_R^2 & C_{RE} & C_{sR} \\ \sigma_v^2 & C_{RE} & \sigma_E^2 & C_{sE} \\ \sigma_\alpha^2 & C_{sR} & C_{sE} & \sigma_s^2 \end{pmatrix}, \tag{9}$$

where $\sigma_v^2 = \sigma_\alpha^2 + \sigma_\varepsilon^2$, $\sigma_s^2 = \sigma_\alpha^2 + \sigma_\psi^2$, $\sigma_R^2 = \sigma_v^2 + \sigma_\psi^2 + \sigma_{\xi_R}^2 + 2\sigma_{\psi\xi_R}$, $\sigma_E^2 = \sigma_v^2 + \sigma_\psi^2 + \sigma_{\xi_E}^2 + 2\sigma_{\psi\xi_E}$, $C_{RE} = \sigma_v^2 + \sigma_\psi^2 + \sigma_{\psi\xi_R} + \sigma_{\psi\xi_E} + \sigma_{\xi_{RE}}$, $C_{sR} = \sigma_\alpha^2 + \sigma_\psi^2 + \sigma_{\psi\xi_R}$ *and* $C_{sE} = \sigma_\alpha^2 + \sigma_\psi^2 + \sigma_{\psi\xi_E}$. *Correlations are* $\rho_{RE} = C_{RE}/(\sigma_R\sigma_E)$, $\rho_{sR} = C_{sR}/(\sigma_R\sigma_s)$ *and* $\rho_{sE} = C_{sE}/(\sigma_E\sigma_s)$.

Proof. Each entry is the variance of the components the two variables share plus the covariances allowed among their noise terms, and the variances follow by expansion. □

C.2 Closed-Form Quality Differential ω

Proposition 3 (Closed-form ω)**.** *For any threshold* $\bar{u}$ *and nonsingular* Σ_{XX}, $\omega(\bar{u}) = E[v \mid \mathcal{R}, q^E \geqslant \bar{u}] - E[v \mid \mathcal{A}, q^E \geqslant \bar{u}]$*, and each term has the form* $\beta' E[X \mid \text{cell}]$*:* X *is the vector of signals that defines the cell,* $\beta = \Sigma_{XX}^{-1}\Sigma_{Xv}$ *is the coefficient vector of the Gaussian regression of* v *on* X*, and* $E[X \mid \text{cell}]$ *is the mean of a truncated multivariate normal, given in closed form by Tallis (1961). The recruited cell* $\{s < \bar{s},\, q^R \geqslant \bar{r},\, q^E \geqslant \bar{u}\}$ *is a trivariate orthant of* $(-s, q^R, q^E)$*. The algorithmic cell is the bivariate orthant* $\{s \geqslant \bar{s},\, q^E \geqslant \bar{u}\}$ *for the recommended pool and, for the pool of every organic applicant, the univariate event* $\{q^E \geqslant \bar{u}\}$*, for which* $P_{\mathcal{A}} = \Phi(-z_E)$ *and* $E[v \mid q^E \geqslant \bar{u}] = (\sigma_v^2/\sigma_E)\,\lambda(z_E)$ *with* $\lambda = \phi/(1-\Phi)$.

Proof. Under joint normality $E[v \mid X] = \beta' X$, so by iterated expectations $E[v \mid X \in T] = \beta' E[X \mid X \in T]$ for any rectangle T. Each cell is a rectangle once the sign of s is flipped in the recruited cell, and the mean of a normal vector truncated to a rectangle is Tallis's formula. The covariances entering β and the truncation probabilities are those of Lemma 1; $\bar{u}$ enters only through z_E. □

C.3 Hire Probability and Match Quality Effects

Pass rates. Each candidate independently clears the employer's screen with a pool-specific probability:

$$p_{\mathcal{A}} = P(q^E \geqslant \bar{u} \mid s \geqslant \bar{s}) = \frac{\Phi_2(-z_E, -z_s; \rho_{sE})}{\Phi(-z_s)}, \tag{10}$$

$$p_{\mathcal{R}} = P(q^E \geqslant \bar{u} \mid s < \bar{s},\, q^R \geqslant \bar{r}) = \frac{\Phi_3(z_s,\, -z_{\mathcal{R}},\, -z_E;\, -\rho_{sR},\, -\rho_{sE},\, \rho_{RE})}{\Phi_2(z_s,\, -z_{\mathcal{R}};\, -\rho_{sR})}, \tag{11}$$

where the numerator of $p_{\mathcal{R}}$ is the probability of the recruited hire cell. Equation (10) is for the recommended pool. For the pool of every organic applicant the score truncation drops out and $p_{\mathcal{A}} = \Phi(-z_E)$.

Proof of Result 1. $P_T - P_C = [1-(1-p_{\mathcal{A}})^n(1-p_{\mathcal{R}})^m] - [1-(1-p_{\mathcal{A}})^n] = (1-p_{\mathcal{A}})^n[1-(1-p_{\mathcal{R}})^m]$. Both factors lie in $[0,1]$, and the product is strictly positive when $m > 0$, $p_{\mathcal{R}} > 0$ and $p_{\mathcal{A}} < 1$. □

Proof of the decomposition (8). Fix a threshold $\bar{u}$. Let $N_{\mathcal{A}}$ and $N_{\mathcal{R}}$ be the numbers of candidates in each pool who would clear it, independent and distributed $\text{Binomial}(n, p_{\mathcal{A}}(\bar{u}))$ and $\text{Binomial}(m, p_{\mathcal{R}}(\bar{u}))$. Random interview order makes each clearing candidate equally likely to be reached first, so the hired candidate

is uniform on the clearing set and the probability that a treated employer's hire is recruited is

$$\pi_{\mathcal{R}}(\bar{u}) = E\left[\frac{N_{\mathcal{R}}}{N_{\mathcal{A}} + N_{\mathcal{R}}} \;\middle|\; N_{\mathcal{A}} + N_{\mathcal{R}} \geqslant 1\right]. \tag{12}$$

Conditional on its source, the hired candidate is a uniformly random member of that pool's clearing set, whose members are independent draws from the pool's conditional distribution, so its expected quality is $E[v \mid q^E \geqslant \bar{u}, k]$. Hence $Q_C = E[v \mid q^E \geqslant \bar{u}_C, \mathcal{A}]$ and $Q_T = \pi_{\mathcal{R}}(\bar{u}_T)\, E[v \mid q^E \geqslant \bar{u}_T, \mathcal{R}] + (1 - \pi_{\mathcal{R}}(\bar{u}_T))\, E[v \mid q^E \geqslant \bar{u}_T, \mathcal{A}]$; subtracting, and adding and removing $E[v \mid q^E \geqslant \bar{u}_T, \mathcal{A}]$, gives (8). At a common threshold the second term vanishes and $\Delta_Q = \pi_{\mathcal{R}}\, \omega$. □

The threshold-response term takes the sign of the response, which Appendix C.5 signs in the error covariance, together with the following lemma.

Lemma 2 (Threshold and algorithmic-hire quality)**.** *$Q_{\mathcal{A}}(\bar{u}) \equiv E[v \mid q^E \geqslant \bar{u}, \mathcal{A}]$ is nondecreasing in $\bar{u}$: unconditionally for the pool of every organic applicant, and for the recommended pool whenever $\sigma_\alpha^2\, \sigma_{\psi\xi_E} \leqslant \sigma_\varepsilon^2\, (\sigma_\alpha^2 + \sigma_\psi^2)$.*

Proof. For the pool of every organic applicant, $Q_{\mathcal{A}}(\bar{u}) = (\sigma_v^2/\sigma_E)\, \lambda(\bar{u}/\sigma_E)$ with $\lambda = \phi/(1-\Phi)$ increasing. For the recommended pool write $\mu_{\mathcal{A}}(x) = E[v \mid q^E = x, s \geqslant \bar{s}]$, so that $Q_{\mathcal{A}}(\bar{u}) = E[\mu_{\mathcal{A}}(q^E) \mid q^E \geqslant \bar{u}, s \geqslant \bar{s}]$; raising $\bar{u}$ shifts the conditioning distribution up in the first-order sense, so it suffices that $\mu_{\mathcal{A}}$ is nondecreasing. Conditioning on $q^E = x$ and then on $s \geqslant \bar{s}$ gives $\mu_{\mathcal{A}}(x) = \beta_1 x + \beta_s\, \sigma_{s|E}\, \lambda\big((\bar{s} - m_s(x))/\sigma_{s|E}\big)$, where $\beta_1 = \sigma_v^2/\sigma_E^2$ is the coefficient of v on q^E alone, β_s is the coefficient on s in the regression of v on (q^E, s), $m_s(x) = (C_{sE}/\sigma_E^2)\, x$, and $\sigma_{s|E}^2$ is the residual variance of s given q^E. Differentiating, $\mu'_{\mathcal{A}}(x) = (1 - w)\, \beta_1 + w\, \beta_E$ with $w = \lambda'(\cdot) \in (0, 1)$, where $\beta_E = \beta_1 - \beta_s C_{sE}/\sigma_E^2$ is the coefficient on q^E in the regression of v on (q^E, s). Since $\beta_1 > 0$, $\mu'_{\mathcal{A}} \geqslant 0$ whenever $\beta_E \geqslant 0$, and $\beta_E \propto \sigma_v^2 \sigma_s^2 - \sigma_\alpha^2 C_{sE} = \sigma_\varepsilon^2(\sigma_\alpha^2 + \sigma_\psi^2) - \sigma_\alpha^2\, \sigma_{\psi\xi_E}$. □

C.4 Threshold Effects

The employer's threshold. The employer's problem in Section 5.2 is a sequential search without recall. Applications arrive at rate ζ, holding the vacancy open costs c_f per unit time, and she hires the first whose q^E clears her reservation value V. Without discounting, $\dot{V} = c_f - \zeta\, E(q^E - V)^+$. Only the cost per application, c_f/ζ, is identified, and the control employer's problem pins it: her bar is the value at which one more application is just worth reviewing, $E_{\mathcal{A}}[(q^E - \bar{u}_C)^+] = c_f/\zeta$.

The treated employer searches the same way over a pool the recruiter has enlarged by m candidates, so her reservation value moves as the pool runs down. It solves

$$\dot{V} = c_f - \zeta\, E_{\text{mix}}\big[(q^E - V)^+\big],$$

backward from $V(T) = \bar{u}_C$ at the horizon $T = (n+m)/\zeta$ at which the pool runs out, with E_{mix} the expectation over the treated pool's mixture of algorithmic and recruited candidates. Her bar is the average of V over $[0, T]$. This mechanism justifies the employer's hiring threshold in the model.

C.5 General Comparative Statics

By Lemma 1, $\sigma_{\xi_{RE}}$ enters the joint law only through C_{RE}: the law of (v, s, q^E), hence everything about the algorithmic pool at a given threshold, is unchanged, and so is $P(\mathcal{R})$ for the recruited pool $\mathcal{R} = \{s < \bar{s},\, q^R \geqslant \bar{r}\}$, which involves (s, q^R) only. For any cutoff t, $\mathcal{R} \cap \{q^E \geqslant t\}$ is an upper orthant of the centered Gaussian vector $(-s, q^R, q^E)$ in which only the correlation of q^R and q^E rises, so its probability is nondecreasing by Slepian's inequality (Slepian, 1962). By Plackett's identity (Plackett, 1954) its derivative in that correlation is the bivariate density of (q^R, q^E) at the boundary times a conditional probability, so the rise is strict at finite thresholds. Hence $P(q^E \geqslant t \mid \mathcal{R})$ rises for every t, and in particular $p_{\mathcal{R}}$ does.

Proof of Proposition 1. (i) Δ_P of (5) is increasing in $p_{\mathcal{R}}$ for $m > 0$, so it rises. Write $\pi_{\mathcal{R}}$ of (12) as $1 - E[N_{\mathcal{A}}/(N_{\mathcal{A}} + N_{\mathcal{R}})]/P(N_{\mathcal{A}} + N_{\mathcal{R}} \geqslant 1)$, with the ratio set to zero when both counts are zero. A binomial rises in the usual stochastic order with its success probability, so the recruited counts at the two values of $p_{\mathcal{R}}$ can be coupled with $N_{\mathcal{R}} \leqslant N'_{\mathcal{R}}$ pointwise, with the independent $N_{\mathcal{A}}$ held fixed. The count share $N_{\mathcal{A}}/(N_{\mathcal{A}} + N_{\mathcal{R}})$ falls and the hire indicator rises in $N_{\mathcal{R}}$, so the quotient falls and $\pi_{\mathcal{R}}$ rises. The fractional pools of Appendix C.6 are fixed-weight mixtures over integer sizes, so both monotonicities carry over. (ii) At a common threshold (8) reads $\Delta_Q = \pi_{\mathcal{R}}\, \omega$, so the rise in $\pi_{\mathcal{R}}$ lowers Δ_Q through this channel wherever $\omega < 0$. □

Lemma 3 (Threshold response to correlated errors)**.** *Under the threshold rule of Appendix C.4, the treated employer's threshold is nondecreasing in $\sigma_{\xi_{RE}}$. Relative to freezing her threshold as $\sigma_{\xi_{RE}}$ rises, the response dampens the rise in the hiring gain and, under the condition of Lemma 2, raises the expected quality of algorithmic hires. Its effect on the recruited share of hires is not signed.*

Proof. The employer's option value in the treated pool, $E_{\text{mix}}[(q^E - u)^+]$, averages an increasing function of q^E over a fixed-weight mixture of the two pools, so by the shift above it rises at every u, while the cost c_f, from the control problem, does not. The bar averages the path of $\dot{V} = c_f - \zeta\, E_{\text{mix}}(q^E - V)^+$ backward from $V(T) = \bar{u}_C$, with $\bar{u}_C$, c_f, ζ and T unchanged; the drift is Lipschitz in V and smaller at every V, so by the comparison lemma the path and its average are higher. The higher bar lowers her hire probability, while the control employer's is unchanged, so Δ_P rises by less than at a frozen threshold. It raises algorithmic-hire quality by Lemma 2. The sign of the change in $\pi_{\mathcal{R}}$ is not determined: each pool's pass rate falls at its own hazard at $\bar{u}$, and the two hazards are not ordered in general. □

Proof of Proposition 2. With $\sigma_\alpha^2 = 0$, $v = \varepsilon$ and $s = \psi$; the pools are those of Section 5.1. Write $\xi^k = b_k\, \psi + \eta_k$ with $b_k = \sigma_{\psi\xi_k}/\sigma_\psi^2$, so that (η_R, η_E) is independent of (ε, ψ), with variances $\nu_k = \sigma_{\xi_k}^2 - \sigma_{\psi\xi_k}^2/\sigma_\psi^2$

and covariance $\nu_{RE} = \sigma_{\xi_{RE}} - \sigma_{\psi\xi_R}\sigma_{\psi\xi_E}/\sigma^2_\psi$. Let $X_R = \varepsilon + \eta_R$ and $X_E = \varepsilon + \eta_E$. Given $\psi = p$, a recruited hire has $X_R \geqslant \bar{r} - (1 + b_R)p$, $X_E \geqslant \bar{u} - (1 + b_E)p$ and $p < \bar{s}$; an algorithmic hire has only the second, with $p \geqslant \bar{s}$. Two conditions, both implied by the hypothesis, suffice. (a) $1 + b_k > 0$ for both parties, so both cutoffs fall in p. (b) $0 \leqslant \nu_{RE} \leqslant \min(\nu_R, \nu_E)$. Under (b) the density of (ε, X_R, X_E) is multivariate totally positive of order two (MTP2): with $D = \nu_R\nu_E - \nu^2_{RE}$, the mixed second derivatives of its log density are ν_{RE}/D, $(\nu_E - \nu_{RE})/D$ and $(\nu_R - \nu_{RE})/D$, all nonnegative. For an MTP2 density, conditional means of nondecreasing functions on rectangles $\prod_i [a_i, \infty)$ are nondecreasing in every a_i (Karlin and Rinott, 1980). Hence, by (b), the means $m_{\mathcal{R}}(p)$ and $m_{\mathcal{A}}(p)$ of ε on the two cells are nonincreasing in p, since (a) makes both cutoffs fall in p, and $m_{\mathcal{R}}(p) \geqslant m_{\mathcal{A}}(p)$, the recruiter's screen adding a lower bound. Averaging $m_{\mathcal{R}}$ over $\psi < \bar{s}$ and $m_{\mathcal{A}}$ over $\psi \geqslant \bar{s}$ gives $\omega \geqslant m_{\mathcal{R}}(\bar{s}) - m_{\mathcal{A}}(\bar{s}) \geqslant 0$, strictly since the recruiter's bound binds with positive probability. □

C.6 Calibration

Results 2 and 3 are numerical. We compute Δ_Q at every point of the grids plotted below, both pieces of (8), with ω from Proposition 3, and check that it falls from one point to the next.

Calibration of the pools and thresholds. All quantities below are pinned to population moments on recruiter-acted treated first posts (26,172 posts; the recruiter acted on 68% of treated first posts). The share of applications the algorithm recommends (37.4%) pins $z_s = \bar{s}/\sigma_s = 0.32$. The recruiter's selectivity among the scored applicants the algorithm did not recommend (1.8%) pins his bar, $z_{\mathcal{R}} = \bar{r}/\sigma_R = 1.91$. The control-group fill rate pins the employer's bar.

Three within-post score gradients pin the three signal correlations. The semi-elasticity of the recruiter's contact decision in the algorithmic score, among the scored applicants the algorithm did not recommend (0.347, se 0.015), pins $\rho_{sR} = 0.244$. The semi-elasticity of hiring in the score, among recommended applicants on control posts (0.411, se 0.032), pins $\rho_{sE} = 0.254$. The effect of recruiter contact on hiring, among the scored applicants the algorithm did not recommend, relative to the control hiring rate among recommended applicants (0.579, se 0.130), pins $\rho_{RE} = 0.124$.

The pool sizes are direct counts per acted post: $n = 31.6$ organic applicants, of which 12.7 are recommended and 18.9 are not, and $m = 2.3$ recruiter-added candidates. These are not integers, so the hired-share formula (12) and the fill probabilities below use $\text{Binomial}(\lfloor\mu\rfloor, p)$ plus one further candidate, who exists with probability $\mu - \lfloor\mu\rfloor$ and then passes with probability p. This has mean μp.

Table C.1: Calibration: targeted moments, in the order they are solved

Moment	Value	Model expression	Pins
Share of applications recommended	37.4%	$\Phi(-z_s)$	$z_s = 0.32$
Recruiter contact gradient in the score	0.347 (0.015)	$\Phi_2(z_s, -z_{\mathcal{R}}; -\rho_{sR})/\Phi(z_s)$	$\rho_{sR} = 0.24$
Recruiter selectivity, non-recommended	1.8%	$\Phi_2(z_s, -z_{\mathcal{R}}; -\rho_{sR})/\Phi(z_s)$	$z_{\mathcal{R}} = 1.91$
Control fill rate	28.6%	$1 - (1 - \Phi(-z_E))^n$	$z_E = 2.30$
Hiring gradient in the score, control	0.411 (0.032)	$\Phi_2(-z_s, -z_E; \rho_{sE})/\Phi(-z_s)$	$\rho_{sE} = 0.25$
Effect of recruiter contact on hiring	0.579 (0.130)	$\Phi_3(z_s, -z_{\mathcal{R}}, -z_E; -\rho_{sR}, -\rho_{sE}, \rho_{RE})$	$\rho_{RE} = 0.12$
Organic applicants per acted post	31.6	direct count	n
of which recommended	12.7	direct count	
of which not	18.9	direct count	
Recruiter-added candidates per acted post	2.3	direct count	m

Notes: Moments are population values on the 26,172 recruiter-acted treated first posts, 68% of treated first posts. They are solved by nested univariate roots in the order shown. Selectivity and the two contact moments are measured among the scored applicants the algorithm did not recommend, and the hiring gradient among recommended applicants on control posts. The two gradients are semi-elasticities in the standardised score of the probability shown beside them. The contact gradient and the recruiter's bar are solved jointly, since the bar is re-solved inside the gradient's root. Each moment is matched exactly, so the model column gives the expression inverted rather than a fitted value.

The employer's pool $\mathcal{A}$ is every organic applicant. Appendices C.2 and C.3 give the closed forms for both pools; for this one the score truncation on the $\mathcal{A}$ margin is dropped. The recommendation threshold $\bar{s}$ then enters only through the recruiter's truncation and through ρ_{sE}'s identifying moment.

The employer's threshold. The rule of Appendix C.4 is solved at the calibrated primitives, with the control employer's bar $z_E = 2.30$. At the calibration recruited candidates clear the employer's screen at the same rate as the pool they join ($p_{\mathcal{R}}/p_{\mathcal{A}} = 1.00$, against 1.69 in the data).

The quality scale. Two statistics are not identified by the moments above. The first is k, how informative a human impression is relative to the algorithmic score, defined by $k \equiv \mathrm{corr}(v, q^{\mathcal{R}})/\mathrm{corr}(v, s)$ and equal to $\mathrm{corr}(v, q^E)/\mathrm{corr}(v, s)$. The second is $\mathrm{corr}(v, s)$ itself. We set $k = 0.25$ and $\mathrm{corr}(v, s) = 0.55$ as benchmarks, because the cross-section does not pin them.

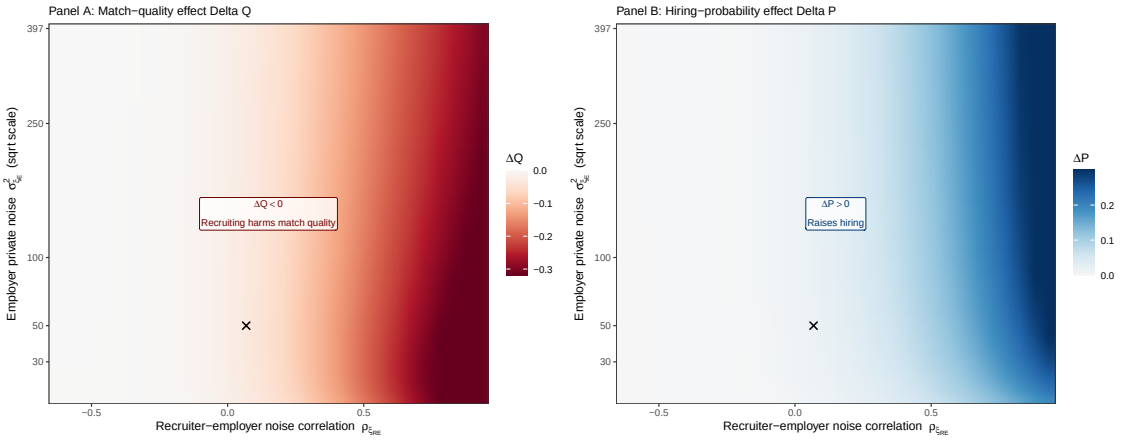
The threshold across the planes. The three bars are re-pinned at every plotted point, so the control employer's pass rate is the same everywhere, while the treated employer's bar moves and her pass rate on the algorithmic pool moves with it. In levels the threshold-selection piece of (8) is small beside the composition piece everywhere except near the low-$\rho_{\xi_{RE}}$ edge, where Δ_Q is itself near zero.

Result 2 (Correlated recruiter–employer errors lower match quality). *Over the plotted range of Figure C.1, $\rho_{\xi_{RE}} \in [-0.63, 0.93]$ and $\sigma^2_{\xi_E} \in [15, 397]$, Δ_Q is negative throughout and falls in $\rho_{\xi_{RE}}$ at every plotted value of $\sigma^2_{\xi_E}$ (41 of 41).* The quality gap ω alone does the same on 39 of 41 rows. On the remaining 2, the smallest $\sigma^2_{\xi_E}$ plotted, it flattens near the top of the $\rho_{\xi_{RE}}$ range and rises from its minimum by at most 0.0015, under 0.9% of its fall; the rising recruited share of hires (Proposition 1) more than offsets it.

Panel B of Figure C.1 shows the hiring gain. A higher $\rho_{\xi_{RE}}$ makes recruited candidates clear the employer's screen more often and makes them worse. The gain rises in $\rho_{\xi_{RE}}$ at every plotted $\sigma^2_{\xi_E}$ (41 of 41) and stays above 76% of its fixed-bar value everywhere. Lemma 3 signs the threshold response that works against it; here it does not reverse the gain.

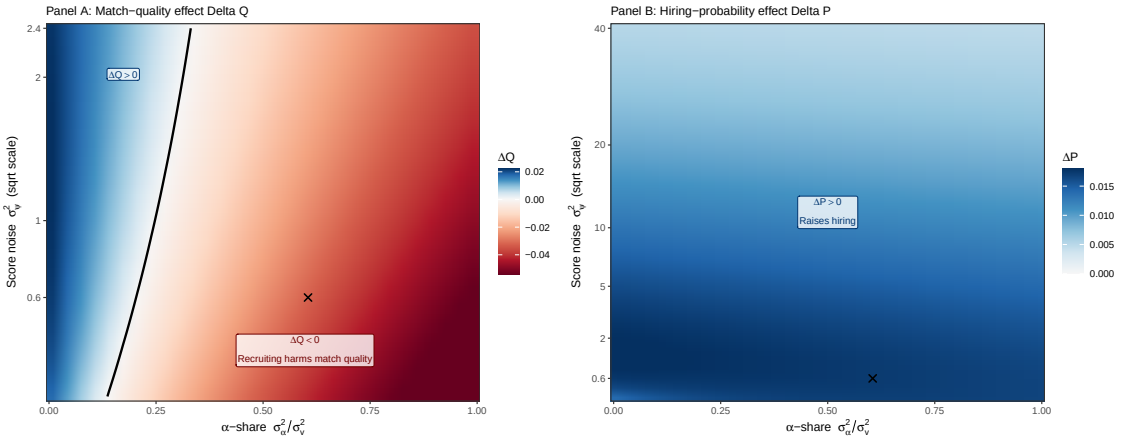
Result 3 (A better algorithm lowers the value of recruiting). *Over the plotted range of Panel A of Figure C.2, the algorithmic share $\sigma^2_\alpha/\sigma^2_v$ at fixed σ^2_v and $\sigma^2_\psi \in [0.24, 2.4]$, Δ_Q falls in the share at every plotted value of σ^2_ψ (41 of 41) and changes sign once, from positive to negative, at a share rising from 0.14 to 0.33.* The quality gap ω falls in the share at every plotted σ^2_ψ as well (41 of 41).

Figure C.1: Comparative statics in the recruiter–employer noise correlation $\rho_{\xi_{RE}}$



Notes: Both panels span the $(\rho_{\xi_{RE}}, \sigma^2_{\xi_E})$ plane, $\rho_{\xi_{RE}} \in [-0.63, 0.93]$ and $\sigma^2_{\xi_E} \in [15, 397]$, on which the threshold solves at every plotted point. n, m, $\mathrm{corr}(v, s)$ and the noise covariances not on the axes are at their calibrated values, and the three bars are re-pinned at every point. Panel A shows the match-quality effect Δ_Q per recruiter-acted post, both pieces of (8) included; negative throughout, clipped at -0.32. Panel B shows the effect on the probability of at least one hire; positive throughout, clipped at 0.30. The cross marks the calibration.

Figure C.2: Comparative statics in the algorithmic share $\sigma_\alpha^2/\sigma_v^2$



Notes: Both panels span the $(\sigma_\alpha^2/\sigma_v^2, \sigma_\psi^2)$ plane at fixed σ_v^2, with n, m and the noise covariances at their calibrated values, so $\mathrm{corr}(v, s)$ and k vary with the axes; the three bars are re-pinned at every point as in Figure C.1. Panel A: match-quality effect Δ_Q, both pieces of (8), $\sigma_\psi^2 \in [0.24, 2.4]$; black contour $\Delta_Q = 0$, red (blue) region: recruiting harms (helps) quality, clipped at $[-0.054, 0.022]$. Panel B: effect on the probability of at least one hire, on its own taller range $\sigma_\psi^2 \in [0.24, 40]$, since Δ_P only falls once the score is far noisier than at the calibration; positive throughout (0.5 to 1.8 percentage points), clipped at 0.018. The cross marks the calibration.