Small numbers bargaining in the age of big data:

Evidence from a two-sided labor matching platform

Moshe A. Barach	Aseem Kaul	Ming Leung	Sibo Lu
University of Minnesota	University of Minnesota	UC Riverside	Upwork.com

In this study, we examine how firms engage with the big data capabilities of two-sided matching platforms. While such platforms can use artificial intelligence (AI) and big data techniques to help firms find better transaction partners, the effectiveness of these technologies depends on the concentration of transactions undertaken on the platform. These big data technologies that enable two-sided matching platforms to create value also give rise to a small numbers bargaining problem, as users become increasingly dependent on the platform to find transaction partners. We argue that firms deal with this appropriation challenge through strategies of partial reliance: they make use of the platform's AI driven recommendations to identify an initial set of generally acceptable partners, then rely on their own internal capabilities to select the best firm-specific match. We test this argument in the context of an online labor platform, using a regression discontinuity design to causally identify the effect of the platform's recommendations on the hiring choices of firms at every stage of the recruiting process. Consistent with our theory, we find that firms rely heavily on the platform's recommendations when screening job applicants to view, but ignore these recommendations when deciding whom to hire from among those interviewed, with the reliance on the platform's recommendations being weaker for more specialized jobs and for more experienced employers. The study contributes to our understanding of how firms use big data technologies, highlighting the challenges these technologies pose for organizational governance and scope choice, and providing a bridge between research on AI and big data and work on two-sided markets.

Introduction

The potential for artificial intelligence (AI) and big data technologies to improve business productivity and performance is a topic of growing interest to researchers (McAfee and Brynjolfsson, 2012; Autor, 2015; Byrnjolfsson and Mitchell, 2017; Furman and Seamans, 2018). While most of this work focuses on the use of AI and big data technologies within firms, one important application of these technologies is in two-sided matching platforms (Rysman, 2009; Gawer, 2014) that use sophisticated algorithms and big data techniques to identify and recommend high quality matches between buyers and suppliers, thus adding value to market transactions (Dawes et al, 1989; Grove et al., 2000). The strategy literature has grown increasingly interested in such platforms (Eisenmann, Parker, and Van Alstyne, 2006; Cennamo and Santalo, 2013; Teece, 2018)—with scholars examining the entry and investment decisions of platform businesses (Eisenmann, 2006; Zhu and Iansiti, 2012), their pricing and content choices (Hagiu, 2009; Rysman, 2009; Seamans and Zhu, 2014; 2017; Kretschmer and Claussen, 2016), and the ways in which other firms successfully engage with platforms (Huang et al., 2013; Zhu and Liu, 2018; Agarwal and Kapoor, 2018)—yet the role of AI and big data technologies in the context of platform strategies has remained largely unexplored.

In this study, we start to address that lacuna, bridging research on platform strategies and research on the use of AI within organizations. We argue that the big data technologies that enable two-sided matching platforms to create value also give rise to a small numbers bargaining problem. Because these platforms rely on big data technologies to offer better matches to their users, their ability to create value is directly proportional to the number of transactions that are routed through the platform (Varian, 2018). The operation of this network effect in the platform's favor results in a fundamental transformation (Williamson, 1975; 1985), however, replacing a multitude of potential exchange partners with a single platform, and consequently exposing the platform's users to the risk of opportunistic hold-up. Importantly, the small numbers bargaining problem resulting from the use of big data is most pronounced for transactions that are general in nature, because it is precisely in ensuring better matches for such general transactions that big data is most valuable. In this sense it is distinct from the hold-up problem in traditional transaction cost theory (Williamson, 1975; Klein et al., 1978)—or the risk of appropriation resulting from the platform-specific investments of complementors discussed in recent work on technology platforms (Gawer and Henderson, 2007; Eisenmann, 2008; Hagiu, 2009; Kapoor and Agarwal, 2017)—which increases with the firm-specificity of the transaction.

We contend that one way for firms to deal with this small numbers bargaining problem is through strategies of partial reliance on the platform's AI. Specifically, we contend that firms will take advantage of the big data capabilities of two-sided matching platforms to identify transaction partners who meet general criteria, but will complement their use of the platform's recommendations with their own internal expertise to make a firm-specific partner choice, thus limiting their overall reliance on the platform. Further, firms will place less reliance on the platform's recommendations when undertaking relatively specialized tasks, and will decrease their reliance as they grow in size and experience and develop their own internal expertise in identifying valuable exchange partners. In this way, firms will limit their reliance on the big data capabilities of two-sided matching platforms, retaining their competitive advantage through the development and maintenance of internal firm-specific capabilities, while still benefiting from the platform's use of AI.

We test our arguments in the context of a large online labor market platform that matches employers (firms) to freelance employees. We leverage two characteristics of our setting: First, the labor platform uses an algorithm to generate a continuous score for all applicants in terms of their fit for every job, but only tells the potential employer whether the applicant is above a 0.5 cut-off on this rating. This allows us to implement a regression-discontinuity design (RDD) to causally identify firms' use of the recommendation provided by the platform. Second, we are able to observe firms' choices at each stage of the hiring process—initial view, interview, and hiring—so we can pinpoint how much firms rely on the platform's recommendation at each stage of the process.

Consistent with our theoretical arguments, our empirical results show that firms do rely on AI generated recommendations from the platform in their hiring process, but primarily for screening which applicants to consider (a relatively general step); when it comes to determining whom among the interviewed candidates they hire, firms pay essentially no attention to the recommendations from the platform, choosing to rely on their own internal expertise instead. Moreover, firms' reliance on the platform's recommendations is lower when hiring for more specialized jobs, and when they have more experience. Supplementary analyses suggest that the platform's AI is most useful for identifying and screening out the truly weak applicants, further supporting a strategy of partial reliance.

Our study makes several contributions to the existing literature. First, by discussing the small numbers bargaining problem associated with big data it highlights a fundamental organizational governance trade-off between the potential for value creation from channeling all transactions (and their associated information) through a single platform and the value appropriation benefits from keeping transactions firm-specific (Ellison and Fudenburg, 2003). Second, we advance research on the use of AI and big data (Brynjolfsson and McAfee, 2014; Autor, 2015), examining precisely when and how firms use third party big data capabilities, especially in the context of online labor markets (Autor and Scarborough, 2008; Chan and Wang, 2018; Horton, 2017). Third, our study contributes to research on two-sided platforms (Parker and Van Alstyne, 2005; Rochet and Tirole, 2006; Eisenmann, 2008; Hagiu, 2009; Rysman, 2009), emphasizing the key role that big data technologies play in enabling the success of two-sided matching, and shifting the focus from strategies to compete as platforms (Eisenmann, 2006; Zhu and Iansiti, 2012) to strategies to engage with platforms (Huang et al., 2013; Zhu and Liu, 2018). In doing so, we also shift away from the field's traditional focus on technology platforms, which require substantial platform-specific investments from complementors (Gawer and Henderson, 2007; Hagiu and Wright, 2015; Kapoor and Agarwal, 2017), to examine matching platforms, where such investments are minimal (Rysman, 2009; Zhu and Liu, 2018), thus shedding new light on a business model that lies at the heart of the 'gig economy'.

Big data, small numbers, and partial reliance

Big data and two-sided matching platforms

Artificial intelligence (AI) technologies are widely acknowledged as having the potential to fundamentally transform business (McAfee and Brynjolfsson, 2012; Brynjolfsson and Mitchell, 2017; Furman and Seamans, 2018). In particular, big data and machine learning techniques have been hailed as a significant general purpose technology (Brynjolfsson and McAfee, 2014), with the potential to speed up innovation (Cockburn, Henderson, and Stern, 2018), boost labor productivity (Brynjolfsson, Rock, and Syverson, 2018; Choudhury, Starr, and Agarwal, 2018), and enable data driven decision making (Brynjolfsson and Mcelheran, 2016). Among the myriad of exciting uses for these technologies, one key application is in the context of two-sided matching platforms (Rysman, 2009), i.e., platforms that match parties to a transaction, lowering search costs for their users through the use of big data algorithms that help predict and recommend high quality matches. As already mentioned, such platforms find wide application in a variety of different contexts, including online marketplaces (Amazon Marketplace, e-Bay, Craigslist), dating (Match.com, Tinder), and a range of 'gig-economy' businesses (Uber, AirBnB, Handy).

Two-sided platforms create value by leveraging network effects (Katz and Shapiro, 1985); indeed, scholars have argued that the presence of such network effects is the defining characteristic that makes a market two-sided (Rochet and Tirole, 2006; Hagiu, 2009). And while many factors may give rise to network effects—including superior coordination between buyers and suppliers to solve the 'chicken and egg' problem (Caillaud and Jullien, 2003; Rochet and Tirole, 2003; Rysman, 2009) and the enabling of investments in complementary technologies (Gawer and Henderson, 2007; Hagiu and Wright, 2015; Kapoor and Agarwal, 2017)—one key source of such network effects is the superior matching of users, whereby "participating in a mediated marketplace improves the odds of finding suitable trading partners" (Eisenmann, 2006; p. 1193). We can thus distinguish between two-sided matching platforms, where the platform essentially serves as an enabling intermediary (Rysman, 2009), and the technology platforms that have been the focus of much of the prior research in strategy (Gawer, 2014; Kapoor and Agarwal, 2017; Teece, 2018). The key difference is that technology platforms generally require platform-specific investments (Gawer and Henderson, 2007; Hagiu and Wright, 2015)—often requiring users to develop entirely new offerings for the platform, for instance—and feature complex interdependencies between users on at least one side of the platform (Kapoor and Agarwal, 2017; Helfat and Raubitschek, 2018; Agarwal and Kapoor, 2018). In contrast, what we call two-sided matching platforms require little or no platform-specific investments and involve few technical interdependencies between users[1]. So, for instance, the firms that sell products on Amazon Marketplace are not generally developing these products exclusively for sale on Amazon, nor are their products technologically linked to the platform or to other products sold on it (Zhu and Liu, 2018). It is this latter type of platforms that are the focus of our paper.

Big data and machine learning algorithms play a key role in enabling value creation by such two-sided matching platforms, allowing the platform to use data from prior transactions to predict the pairs of users that are most likely to prove a good match. Indeed, in the absence of such technologies to streamline the search process, increasing aggregation of users on a single platform may well lead to increased confusion and congestion, as users struggle to find their best matches in an ever-growing sea of alternatives. Of course, the accuracy of these machine learning algorithms is itself proportional to the number and range of prior transactions used to develop their predictions (Varian, 2018). It is this that leads to the network advantage from the use of machine learning algorithms in two-sided matching platforms—the more transactions that go through the platform, the better its AI, and the more value it can add to subsequent transactions—leading to increased pressure for the concentration of all transactions on a single platform. Thus its access to data, rather than the quality of its algorithms, may be the primary or only source of a platform's competitive

[1] This distinction is, to some extent, a matter of degree. Users on a two-sided matching platform may certainly make some platform-specific investments in making their offering appear more appealing. This is still quite different, however, from developing an entire offering or technology that can only be used on the platform.

advantage[2] (McAfee and Brynjolfsson, 2012; Khan, 2017; 2018; Cockburn, Henderson, and Stern, 2018; Furman and Seamans, 2018).

The small numbers bargaining problem

The network effects generated by the use of big data not only enable platforms to create value by lowering search costs for users in the marketplace, they also allow the platform to appropriate this value. In particular, the use of these techniques may result in a fundamental transformation in the marketplace (Williamson, 1975; 1985): as increased use of the platform makes it increasingly effective at offering the best match, users may find it increasingly difficult to find what they are looking for outside the platform, with the fit of trading partners they find outside the platform being substantially lower than those on the platform. As a result, "what was a large numbers bidding condition at the outset is effectively transformed into one of bilateral supply" (Williamson, 1985; p. 61), with users finding themselves increasingly dependent on the platform to find transaction partners. This, in turn, leaves them vulnerable to opportunistic hold-up by the platform, effectively forced to accept the terms set by the platform, for want of viable alternatives[3]. In essence, the successful use of big data techniques, by rewarding the routing of all transactions (and their accompanying information) through a single platform, results in the concentration of economic power (Zhu and Iansiti, 2012). As a result, successful platforms with well-functioning AI may enjoy both higher prices and dominant market share (Brynjolfsson and Smith, 2000).

Note that one key distinction between the small numbers bargaining problem resulting from the use of big data techniques and that emphasized in traditional transaction cost theory (Williamson, 1975; 1985; Klein et al., 1978), as well as in prior work that has discussed the problem of hold-up in

[2] As Google's Peter Norvig puts it, "We don't have better algorithms. We just have more data." (quoted in McAfee and Brynjolfsson, 2012, p.5)

[3] This hold-up problem is presumably made worse in so far as both sides of the market see substantially better matches on the platform. The two-sidedness of the hold-up may thus lead to a reverse chicken and egg problem—though both sides could reduce their dependence on the platform if they left it simultaneously, such a mutual exit would require precisely the kind of coordination that made the platform necessary in the first place. Absent coordination, a user who left the platform may find slim pickings outside the platform and would therefore be motivated to go back.

technology platforms (Gawer and Henderson, 2007; Eisenmann, 2008; Hagiu and Wright, 2015; Kapoor and Agarwal, 2017) is that, unlike in these cases, the hold-up we discuss here is not the result of specialized investments. On the contrary, to the extent that the predictive accuracy of big data techniques is greater the more general the transaction, the potential for opportunistic hold-up by the platform may be greatest for transactions that are entirely generic. Users with highly specialized needs may gain little from relying on the platform's AI; having few relevant priors from which to generate a prediction, it may do no better than a random draw at finding a match. The more typical or standard the user's requirements[4], the greater the ability of big data techniques to predict and identify a strong match, and therefore the more severe the potential for hold-up.

This reliance on big data also means that internalization may not be an effective solution for users to overcome the threat of opportunistic hold-up by the platform. While a firm could invest in developing its own AI, the accuracy of its predictions may be compromised by the fact that they are based only on the sample of transactions undertaken by the firm itself. This may be especially limiting for new or small firms that would have little prior data on which to train their internal algorithms, and may therefore realize the greatest advantage from relying on the platform's recommendations when searching for transaction partners. But so long as the marginal increase in predictive accuracy from running an additional transaction through a common AI is positive, any attempt by a firm to leave the shared platform and set up its own internal platform would prove value destroying. Firms dealing with a two-sided matching platform thus face a fundamental trade-off (Ellison and Fudenburg, 2003): using a common platform creates value by allowing them to find better transaction partners (or, equivalently, comparable partners at a lower cost), but remaining independent of the platform allows them to appropriate more of the value from the transactions.

[4] To be clear, a user's requirement being typical does not mean that all users have the same requirements; only that the user's requirements are the same as the requirements of other users of the same (potentially unobservable) type and therefore relatively predictable using big data. So, for instance, the books that strategy professors buy may be very different from the books other people buy, but the more an individual strategy professor's tastes are similar to that of other strategy professors, the more she may benefit from the recommendations she gets on Amazon.

Strategies of partial reliance

How might a firm deal with the risk of opportunistic hold-up when transacting through a two-sided matching platform? While several recent studies have examined platform strategies (Eisenmann et al., 2006; Teece, 2018), their focus has generally been on understanding how platforms compete with each other, examining such choices as entry timing (Zhu and Iansiti, 2012), investments (Eisenmann, 2006), exclusivity (Cennamo and Santalo, 2013), compatibility (Kretschmer and Claussen, 2016), complexity (Kapoor and Agarwal, 2017), pricing strategies (Hagiu, 2009; Seamans and Zhu, 2014), and product mix decisions (Seamans and Zhu, 2017), as well as the capabilities that enable firms to orchestrate these decisions (Helfat and Raubitschek, 2018). We thus know comparatively little about the strategies firms dealing with a two-sided matching platform may use to protect themselves against value appropriation by the platform (Huang et al., 2013; Agarwal and Kapoor, 2018), especially in the context of two-sided matching platforms (Zhu and Liu, 2018).

Our contention in this study is that one way in which firms may deal with the threat of opportunistic hold-up by a two-sided matching platform, especially in so far as it relates to the small numbers problem arising from the use of big data, is through strategies of partial reliance[5]. Under such strategies, firms still make use of the platform, but rather than relying exclusively on the recommendations of the platform's AI to identify transaction partners, they develop internal, firm-specific capabilities to supplement the information provided by the platform. So, for instance, a firm might use the platform's recommendations as an initial screen to identify potential partners, but then use an additional set of firm-specific criteria to select an eventual partner. Or it may use its own internal capabilities to search on the platform, looking beyond the recommendations offered by the platform's AI to make its own choices. Or it may use the recommendations provided by the platform

[5] These strategies are not the only way for firms to deal with platform hold-up. Another option is multi-homing (Choi, 2010; Park, Seamans, and Zhu, 2017), which limits the firm's reliance on any one platform. Such strategies may be less successful, however, where the use of big data techniques makes the largest platform substantially more effective than the second largest at identifying matches. And even if the top two or three platforms were relatively comparable in their performance, the need for concentration driven by the use of AI may produce an oligopoly that would leave firms susceptible to coordination among platforms.

to compare and validate its own (independently generated) choices. A firm may also vary its reliance on the platform's recommendations across transactions, relying heavily on the platform's AI when choosing partners for relatively general transactions, but effectively ignoring the platform's recommendations for more firm-specific transactions.

While the specific form of partial reliance may vary from platform to platform and setting to setting, the common thread in all these strategies is that the recommendations generated by the platform's AI serve as a complement to, rather than a substitute for, the firm's own internal judgment. On one hand, the firm takes advantage of the platform's superior big data capabilities; by transacting through the platform, it gains access to the platform's recommendations, and benefits from this information when identifying transaction partners that meet its general requirements. On the other, the firm retains its competitive advantage by developing and maintaining its own independent, firm-specific expertise, thus allowing it to find transaction partners that are better suited to its unique needs, while also bolstering its bargaining power relative to the platform (Huang et al., 2013). In this sense, partial reliance is conceptually similar to the concept of a firm maintaining its absorptive capacity (Cohen and Levinthal, 1990)—a firm may draw heavily on externally generated knowledge, but rely on its own internal capabilities to do so more effectively, recombining external knowledge with its internal expertise to give it a firm-specific advantage.

Hypotheses development

Online labor hiring

In order to study the use of partial reliance strategies empirically, we translate the arguments above into a set of testable hypotheses. We ground these hypotheses in the specific context of a firm looking to hire employees in an online labor marketplace. This setting has several advantages. First, online labor marketplaces are an important and growing sector (Leung, 2014; Agrawal et al., 2015), with firms increasingly relying on such marketplaces to hire temporary or part-time workers (Gartside et al., 2013; Katz and Kreuger, 2016) and scholarship becoming increasingly interested in the so-called 'gig economy' (Barley and Kunda, 2004; Edelman, Luca, and Svirsky, 2017). Second, the role of AI and big data techniques as a screening mechanism in these markets has been widely recognized, with a number of studies highlighting ways in which the use of algorithmic techniques may help overcome subjective human biases (Kahnemann, 2011; Kuncel et al., 2013; Chan and Wang, 2018) and identify more deserving and better matched candidates (Hoffman, Kahn, and Li, 2017; Horton, 2017; Cowgill, 2018). Third, by focusing on a factor market setting we are able to examine how firms engage with two-sided matching platforms as buyers, thus complementing existing work that focuses on the strategies of firms as suppliers of complementary products to a platform (Huang et al., 2013; Zhu and Liu, 2018; Kapoor and Agarwal, 2018).

A final useful feature of using online labor markets as a setting is that the overall hiring process may be easily deconstructed into three distinct steps—viewing an application, interviewing, and hiring (see Figure 1)—and we can observe the outcomes of each step separately. This means that we can observe the firm's reliance on the platform's recommendations not only in the overall choice of whom to hire, but at each stage of the decision process. The online hiring process thus provides an ideal setting to test our theory of partial reliance, especially since, as discussed later, we expect the three stages of the process to vary in their firm-specificity.

Partial reliance on platform recommendations

Building on our argument for partial reliance, our baseline hypothesis is that firms will take the recommendations of the platform's AI into account when deciding whom to hire on the platform. As already discussed, the platform's big data capability will give it an advantage in identifying potential employees for the firm, and the firm may therefore benefit from incorporating this information into its hiring decisions—indeed, access to these recommendations is part of the value that the firm realizes from being on the platform in the first place[6].

---

[6] All our hypotheses focus on whom the firm chooses to hire, conditional on choosing to hire on the platform. While firms could and do hire from other sources, we are unable to observe and therefore do not study firms' hiring outside the platform. Our study thus examines the intensive margin of reliance on the platform's AI rather than the extensive margin of platform use, consistent with our theoretical focus on the small numbers problem arising from the use of big data, rather than on the bargaining power of platforms more generally.

We thus predict:

> *H1 (baseline): Being recommended for a job by the platform's algorithm increases an applicant's probability of being hired*

This baseline hypothesis speaks to *whether* the firm will make use of the platform's recommendations. Our primary interest, however, is in understanding *how* the firm uses these recommendations. In particular, we are interested in examining how a firm's reliance on the platform's recommendations varies by the stage of the hiring process. This is because we expect the hiring process to become increasingly firm specific at each subsequent step. At one end, the initial choice of which applicants to consider may be relatively general. A potential employer will want to determine whether the applicant has the basic qualifications, skills, and experience relevant to the job, and may screen out ineligible or undesirable candidates based on fairly basic criteria. At the other end, the final decision on whom to hire is likely to be relatively firm-specific, with firms evaluating not only the applicant's general qualifications, but her match with the specialized needs of the position, as well as with the firm's culture and people. At this stage, the best candidate may not be the one with the strongest stand-alone resume, but the one who best fits with the rest of the organization. So, for instance, a law firm looking to hire a new associate may initially screen on legal credentials and practice area, but will eventually hire the person they believe will best complement their existing team. Similarly, an academic department may arrive at a shortlist of applicants based on their publication records, but may look to their collegiality and their compatibility with others in the department when deciding whom to eventually hire. The intermediate step—choosing whom to interview out of the applicants viewed—is likely to lie in between these two extremes in terms of its level of firm-specificity, as Figure 1 shows.

Insert Figure 1 about here

The idea that firms will first screen potential employees on their general characteristics, and then select from among the generally suitable candidates based on firm-specific criteria[7], has implications for the extent to which the firm may rely on the platform's recommendations at each stage. As discussed, a strategy of partial reliance means that the firm relies heavily on the input from the platform's AI for sub-tasks that are more general (thus benefiting from the platform's big data capabilities where they are the most useful) but pays little attention to these recommendations for sub-tasks that are firm-specific (thus maintaining its unique advantage relative to other platform users). It follows that, in the context of hiring decisions, firms will rely heavily on the platform's AI when deciding which applicants to view, may make some use of the platform's AI when choosing whom to interview, and are likely to choose the applicant they finally hire independently, largely ignoring the platform's recommendations at this final stage. In essence, a firm adopting a strategy of partial reliance may use the platform's AI to prepare a shortlist of generally acceptable applicants, then make a firm-specific choice among them, based on its own expertise. Thus:

H2a: The positive effect of a recommendation from the platform's algorithm will be stronger when selecting which applicants to view than which of the viewed applicants to interview

H2b: The positive effect of a recommendation from the platform's algorithm will be stronger when selecting which of the viewed applicants to interview than which of the interviewed applicants to hire.

Moderating effect of job and employer characteristics

The hypotheses above predict whether and how firms will make use of the recommendations from the AI's platform in making hiring decisions. However, a strategy of partial reliance also suggests that employers will vary in where and when they rely on the platform's big data capabilities. First, firms may be less likely to make use of the platform's recommendations when hiring for relatively specialized tasks. On one hand, the benefits of big data may decline as jobs get

[7] We do not think that such an approach is limited to hiring processes. It seems reasonable to expect that when looking for any transaction partner, the firm will first identify those that meet some basic general criteria, then select those that offer a strong firm-specific fit. Given that firm-specific criteria are, by definition, more limiting than general criteria, it is less likely to make sense to do it the other way around.

more specialized. Since the predictive accuracy of the platform's AI derives from having data on relevant prior transactions, the platform may be less accurate when matching for jobs with specialized or unusual features; in fact, the more atypical the job the less accurate the AI's prediction, and therefore the less useful the recommendations generated by the platform. On the other hand, more specialized tasks may also be more firm-specific, making it more useful for the employer to rely on its own internal capabilities to identify matching applicants. We therefore predict[8]:

H3: The positive effect of a recommendation from the platform's algorithm on an applicant's probability of being hired will be weaker, the more specialized the job

Second, the value of the platform's big data capabilities may vary with the level of the firm's own experience. A firm with no prior experience in the marketplace may have little ability to identify the best applicants for its positions, and may therefore rationally choose to rely more heavily on the platform's AI. As the firm learns from its own experience, however, it may get better at identifying high quality applicants on its own. In the extreme, as mentioned earlier, the firm may develop its own AI application, and use that to screen applicants, potentially triangulating its choices against the platform's recommendations. Even if the firm continues to rely on human judgment, however, its expertise is likely to grow as its number of previous hires increases. Of course, no individual employer's experience will ever equal that of the platform, so that the platform's big data capabilities may always exceed the firm's internal expertise. But, assuming diminishing returns to scale (Varian, 2018), the relative disadvantage of the firm may decline with experience, especially since the firm's internal expertise, though more limited in general, may have the advantage of being tailored to the firm's specific needs. Since the fundamental trade-off underlying a strategy of partial reliance is between benefiting from the superior (big data enabled) capabilities of the platform and maintaining one's own (firm-specific) expertise, that trade-off is likely to shift in favor of less reliance on the platform as the firm moves along its own learning curve.

[8] We expect these moderators to work in the same direction across all three stages of the hiring process, and therefore do not define separate hypotheses for each stage. We do, however, examine the effect of these moderators separately at each stage in our empirical analyses.

Thus:

H4: The positive effect of a recommendation from the platform's algorithm on an applicant's probability of being hired will be weaker, the more experienced the potential employer

Data and Methods

Empirical Setting

As previously mentioned, we test our hypothesis in the context of a large online labor platform. Such markets are an increasingly important part of the hiring landscape (Autor and Scarborough, 2008; Leung, 2014; Agrawal et al., 2015; Horton, 2017), and allow firms to contract with freelance employees to perform tasks that can be done remotely, such as computer programming, graphic design, data entry, and writing (Horton, 2010). Online labor platforms differ in their scope and focus, but common services provided by the platforms include maintaining job listings, hosting user profile pages, arbitrating disputes, certifying seller skills, and maintaining reputation systems. Such online labor platforms are a good fit with our notion of a two-sided matching platform: they have two sets of users—employers (firms) and employees—and neither side develops its offerings primarily or exclusively in ways that are specific to the platform; nor are there substantial interdependencies between users. Some examples of such platforms include oDesk, Upwork, and Freelancer.com.

The specific platform that constitutes our setting is similar to these platforms[9]. In 2015, employers spent over $ 1 billion on wages through the platform, and millions of employers and freelancers had created on-line profiles on this platform by the end of that year. As such, the platform has truly accumulated a big-data capability, which gives it an advantage in identifying potential employees for firms. Note that employers on this platform are mostly small firms—the two largest categories being individuals (42%) or small businesses consisting of <10 employees (43%)—which makes the platform's big-data capabilities all the more valuable for them.

[9] Non-disclosure agreements prevent us from revealing the name of the platform.

On the platform, would-be employers (firms) write job descriptions, self-categorize the nature of the work and required skills, and then post the requests for proposals to the platform website. Potential employees learn about requests for proposals via electronic searches or email notifications. These individuals submit applications, or "proposals", which generally include a wage bid (for hourly jobs) or a total project bid (for fixed-price jobs) and a cover letter. In addition to seller-initiated applications, employers can also search individual's profiles and invite them to apply.

After an individual applies, the platform employs a machine learning algorithm to assess each applicant and score him or her on a zero to one scale, in terms of his or her fitness for that specific job. While the exact score of the algorithm is never made public, all applicants who score a 0.50 or above, at the time of their application, are automatically "recommended" by the platform, resulting in a "recommended" flag being placed above their picture. Additionally, applicants are automatically sorted by algorithm score, placing the highest scoring applicant at the top of the list that employers view[10]. Applicants are never made aware of their recommendation status nor their exact algorithm score[11].

The employer then screens through the applicants on the platform's interface, selects some applicants to interview, and finally chooses one or more applicants with whom it contracts for the job, often on the terms proposed by the applicant. Note that the process is not an auction and neither the buyer nor the seller is bound to accept an offer. While bargaining is a possibility—the employer can make a counteroffer, which the applicant can counter, and so on—it is somewhat rare. Applicants turning down job offers is also fairly rare. In 91% of cases where an employer makes an offer, the applicant eventually accepts the job and charges the firm for work completed.

[10] Appendix A contains detailed screen shots and explanations of the hiring process. As discussed later, we control for the effect of viewing order in all our regressions.
[11] From the frequent discussions on the platform's message board, it is clear that the applicants are aware of the recommendation algorithm but do not know enough detail to guess the algorithm.

Data and Variables

We study a random sample of job postings that were submitted to the platform between March 3, 2016 and May 30, 2016. We limit our sample to only non-invited applicants. We additionally winsorize our sample by dropping job postings which are above the 99th percentile, with respect to either number of applications or number of applicants hired per job. We further drop outlier applications which have bids above the 99^{th} percentile ($99). The remaining sample consists of 1,500,603 applications from 204,068 unique freelancers, submitted to 125,300 unique job postings posted by 71,321 unique employers. Table 1 presents summary statistics for all applicants to the jobs in our complete sample.

Insert Table 1 about here

Empirical Strategy

As already discussed, we seek to understand how firms make use of the recommendations generated by the platform's algorithm in their hiring decisions. Specifically, we examine the effect of being recommended by the algorithm on the likelihood of an applicant moving forward at each step along the hiring funnel shown in Figure 1. Thus, we begin by analyzing the employer's overall hiring choice, then break that choice into its various stages, studying the firm's decision to first view the applicant, then to interview the applicant through an online messaging system (conditional on being viewed), and finally examine the applicant's likelihood of being hired (conditional on being interviewed).

In order to assess the causal effect of being recommended, we cannot simply compare the selection rates of recommended and non-recommended applicants, as these applicants may vary in quality in ways that are not easily observable, leading to potential omitted variable bias. To eliminate such problems, we take advantage of the fact that we have access to the precise scores for each applicant for each job generated by the algorithm; scores that are not revealed to either the applicant or the potential employer. Since these scores are unavailable to potential employers, they cannot factor into their hiring decisions (except in so far as they impact viewing order, which we control for below). Moreover, job applicants are unable to precisely manipulate their algorithm score, since they are unaware of both their exact scores and the algorithm used to generate them[12]. As such, the variation in treatment—i.e., being recommended by the platform—near the recommendation threshold can be considered nearly randomized. In other words, those who just make the threshold to be recommended and those who just fail to meet it may be considered comparable in quality, with the only difference between them being that the former receive the platform's recommendation and the latter do not. Taking advantage of this institutional detail allows us to use a regression discontinuity design (RDD) to control for the heterogeneous quality of applicants (Hahn, Todd, and Van der Klaauw, 2001; Imbens and Kalyanraman, 2012). To get a better idea of the amount of variation that exists close to the recommendation discontinuity, Table 2 presents summary statistics for what we call our discontinuity sample, which is limited to job postings with at least one recommended applicant and applications which are close to the algorithmic threshold (between .45 and .55). Notice that while the number of job postings is reduced to 53,134, the summary statistics including average applicant bid is quite similar to Table 1. We base all our main analyses on this discontinuity sample.

Insert Table 2 about here

We show the results of our RDD in two ways. We begin with a graphical analysis of the effects of an applicant's being recommended by the platform on the employer's likelihood of selecting that applicant at various stages in the hiring funnel (Lee and Lemieux 2010). Figure 2a plots the average probability of an applicant's being hired by the application's algorithm score. To further unpack the hiring process, Figure 2b plots the average probability of an applicant's being viewed, viewed applicants being interviewed, and interviewed applicants being hired by the application's algorithm score. The average value of each outcome was calculated for bins of width of .002 on either side of the .5 recommendation discontinuity for algorithm scores ranging between .45 and .55.

[12] See Appendix C for tests which show that applicants are unable to manipulate whether or not they are recommended.

The bin width was chosen to clearly demonstrate the effect of the recommendation discontinuity without over-smoothing the data. Results with alternative bin widths are similar.

In addition to these graphs, we report regression results which estimate the average treatment effect of the recommendation on employers' decisions for applicants with hiring algorithm scores close to the recommendation discontinuity. These results allow us to precisely measure the impact of the algorithmic recommendation on the employers' hiring decisions at every stage of the process. Specifically, our preferred regression model is a local linear regression that allows for slopes to vary on either side of the cutoff (Hahn, Todd, and van der Klaauw 2001)[13]. We calculate the bandwidth used in each model using the Imbens-Kalyanaraman Optimal Bandwidth Calculation procedure (Imbens and Kalyanaraman, 2012). Our main models thus take the form:

$$Y_{ij} = \beta_1 * Recommeded\ (Rec)_{ij} + \beta_2 * I(Rec = 0) * BM\ score_{ij} + \beta_3 * I(Rec = 1) * BM\ score_{ij} + X_{ij} + \gamma_j + \theta_j + \epsilon_{ij} \quad (1)$$

These regressions use only data within a small bandwidth around the cut-point. In Table 3 Model (1), the outcome of interest, Y_{ij} is an indicator that applicant i, to job opening j, was hired by the employer. In Model (2) the outcome of interest is the applicant being viewed, in Model (3) it's being interviewed conditional on being viewed, and in Model (4) it's being hired conditional on being interviewed.

Insert Figure 2 and Table 3 about here

In all models, Recommended (Rec) is an indicator which is equal to 1 if the applicant was recommended, BM score is the applicant's undisclosed algorithm score, and X_{ij} is a vector of controls which include: the applicant's hourly wage, the employer's on platform tenure and the number of previous jobs the employer has filled. γ_j is a job posting fixed effect, and θ_j is a default sort order fixed effect. The job posting fixed effect ensures that we only make use of variation between applicants within a job posting, and that our results are not driven by heterogeneous

[13] Table B1 in Appendix B shows that these results are robust to both a variety of model specifications and bandwidth choices.

employers or heterogeneous job postings. The default sort order fixed effect controls for mechanical effects related to how the algorithmic information is presented to employers. We report linear models, which allow for differential slope effects of the algorithm score on either side of the discontinuity, but all our empirical results are robust to logit models (available from the authors upon request).

Results

Main findings

Table 3 displays the results of our main analysis. Hypothesis 1 proposes that an applicant being recommended by the platform will positively impact an applicant's likelihood of being hired. Figure 2a shows that there is a significant positive relationship between the algorithm score and the applicant's likelihood of being hired. Moreover, there is a large discontinuity in the likelihood of hiring an applicant who is right at the recommendation threshold. This indicates that employers are more likely to hire recommended applicants of similar quality than non-recommended applicants. These results are confirmed by Table 3 Model (1), which indicates that applicants who are marked as recommended by the hiring algorithm are about .9 percentage points more likely to be hired, relative to a baseline probability of being hired of about 1.5% just to the left of the discontinuity. This implies that recommended applicants are about 60% more likely to be hired than comparable applicants who are not recommended by the platform. Hypothesis 1 is thus supported.

Hypothesis 2a and hypothesis 2b state that the positive effect of being recommended will decrease across stages of the hiring process, being strongest for the initial (relatively general) choice of which applicants to view, and weakest for the final (relatively firm-specific) choice of which interviewed candidates to hire. We begin by considering the firm's decision to view a candidate's detailed application. The left panel of Figure 2b shows that there is a large discontinuity in the probability of viewing an applicant across the recommendation threshold. This indicates that an employer is much more likely to view the detailed application of an applicant who is recommended by the platform, than that of a similarly skilled applicant who is not recommended. The regression results in Table 3 (Model 2) confirm these findings and show that for applicants who are close to the threshold, being marked as recommended by the hiring algorithm makes an applicant 10 percentage points more likely to be viewed by an employer, implying a 38% increase in likelihood of an applicant being viewed from a baseline of 26%.

Next, we consider the probability of a viewed applicant being chosen to be interviewed. In Model (3) of Table 3, we limit the sample to only applicants who were viewed by the employer. After conditioning on applicants who have been viewed by the employer, the effect of the recommendation is substantially smaller, but still positive and significant. The recommendation increases an employer's likelihood of interviewing an applicant they have already viewed, by 1.7 percentage points from a baseline probability of about 24% at the threshold. This translates to an increase in the likelihood of interviewing a viewed applicant of about 5%. Comparing this 5% increase, in the probably of interviewing viewed applicants, to the increase in the probably of viewing any applicants (which is about 38%), it is clear that the positive effects of a recommendation by the platform's algorithm is stronger when selecting applicants to view than when selecting which viewed applicants to interview. This is further confirmed by Figure 2b, where the discontinuity in the left panel (probability of view) is far larger than the discontinuity in the center panel (probability of interview | view). Hypothesis 2a is thus supported.

Finally, Model (4) of Table 3 examines the probability of being hired conditional on being interviewed, and shows that the effect of being recommended is not statistically different from zero at this stage. This result is visually confirmed in the right panel of Figure 2b (probability of hire | interview), where there is no visual discontinuity across the recommendation threshold. Thus, while firms pay close attention to the platform's recommendations when choosing whom to view, and some attention to it when choosing whom among the viewed candidates to interview, by the time they get to choosing among the interviewed applicants they no longer take the platform's recommendation into consideration. This result supports hypothesis 2b, and is strongly consistent with a strategy of partial reliance.

Heterogeneous Effects by job specialization

Hypothesis 3 proposes that the positive effects of being recommended will be lower for jobs with more specialized requirements. To investigate this differential effect, we separate job postings into two groups: those job postings on which the employer requires applicants to have at least one job-specific skill such as WordPress, graphic design, PHP, etc.; and those job postings which do not request any job-specific skills from applicants[14]. From Table 1 we can see that 56% of job postings in our sample requested at least 1 job-specific skill.

Table 4 reports results of models that incorporate an interaction between our main predictor (Rec) and an indicator for whether a job-specific skill was requested. These models are similar to those used in Table 3, except that we are no longer able to include job posting fixed effects since our indicator for job-specific skill does not vary within jobs. To compensate for this, we add a number of additional job and employer level controls including: the applicant's hourly wage, the job category, the number of applicants, the number of applicants squared, the number of recommended applicants, job value indicators, indicators for the contract type (hourly or fixed price), the employer's tenure and previous experience on the platform, as well as an employer fixed effect.

From Table 4 we can see that the coefficient on the interaction term $Rec_{ij} * I(Request\ Skill = 1)_j$ is negative and significant for Model (2), where the outcome of interest is whether the applicant was viewed. This indicates that when hiring for jobs that require job-specific skills, employers rely less on the platform's recommendations when deciding whom to view (the stage where the effect of the platform's recommendations is otherwise most pronounced). Hypothesis 3 is thus supported.

[14] Robustness checks that use the specific number of unique skills requested as opposed to a binary measure are available from the authors, and show similar results.

We do also see a negative coefficient for our interaction term when predicting the probability of being interviewed conditional on being viewed (Model 3), though the effect is not statistically significant at conventional levels. Given that the effect of the platform's recommendations at this stage is marginal in any case (as shown in Table 3b), it may be that we're unable to estimate a moderating effect with sufficient precision at this stage. In any case, the negative and significant moderating effect of job specificity on probability of viewing recommended applicants provides support for hypothesis 3.

Insert Tables 4 and 5 about here

Heterogeneous Effects by employer experience

Hypothesis 4 predicts that the positive effect of a recommendation from the platform's algorithm on an applicant's probability of being hired will be weaker, the more experienced the employer. To investigate this hypothesis, we run a regression which is structurally similar to that in Table 4, but investigates differential use of the algorithm by a firm's experience on the platform. Specifically, we use the number of previous jobs the employer has filled on the platform as our measure of experience, dividing firms into quartiles of prior experience. We choose to focus on prior filled jobs, as it more accurately represents expertise in transacting via the platform than something like tenure. Some firms go months between transactions, or wait a substantial period of time prior to successfully posting their first job. The 25th percentile of prior job postings is 0, the 50th percentile is 2 prior successful job postings, the 75th percentile of prior job postings is 14, and the maximum number of previous job postings in our sample is 323. Furthermore, the regression models presented in Table 5 do not include firm specific fixed effects as this fixed effect would remove from the analysis all firms with no prior experience on the platform, which is the group of firms we would expect to make greatest use of the platform's recommendations.

In Model (2) of Table 5, the outcome of interest is whether an applicant is viewed. Compared to the baseline group which includes firms with no prior platform experience very experienced firms (greater than 15 previous jobs filled) are 5 percentage points less likely to view recommended applicants. Thus, these very experienced firms rely on the platform's recommendations about half as much as firms with no experience. Similarly, firms with moderate experience (between 3 to 14 prior jobs) are about 3 percentage points less likely to view recommended applicants than inexperience firms, implying about a 25% reduction in reliance. These results clearly show that firms' reliance on the platform's recommendations is lower, the greater a firm's experience, supporting Hypothesis 4.

While we do not see a significant interaction effect in Models (3) and (4), suggesting that, once again, the dampening effect of experience is limited to the first stage of the hiring process (where the main effect of platform recommendations was most salient), Table 5 does show two other interesting patterns, both consistent with a strategy of partial reliance. First, we see a positive and significant main effect of experience on the probability of being interviewed conditional on being viewed (Model 3), and the probability of being hired, conditional on being interviewed (Model 4). This suggests that as firms gain experience they may become better at screening for candidates who will be a good fit for them, consistent with firms developing stronger internal expertise over time. Second, Model (4) shows a positive and significant main effect of the platform's recommendation on the conditional probability of being hired, once we account for the effects of experience. This indicates that the subset of firms with no previous platform experience rely on the algorithm not just in the early stages of hiring, but even in the final stage of choosing which applicant among those interviewed to hire. Together, these results further support our claim in Hypothesis 4 that firms will increasingly rely on their own internal expertise rather than the recommendations of the platform as they gain experience, and are consistent with a strategy of partial reliance more generally.

Supplementary analyses

While our main results provide strong support for the idea that firms adopt strategies of partial reliance, they leave open the question of how well the platform's AI matches firms' preferences, and whether the firm would therefore benefit from using the platform's AI differently? While we cannot fully answer that question, we can get some insight into it by looking at the correlation between the score for each applicant generated by the platform's AI and the firms' choices. Since these scores are never disclosed to the firms, they do not factor into their choices; we may therefore think of the correlation as measuring how closely the firm's independent internal assessment matches that of the platform's AI.

Figure 3a and Figure 3b shows these correlation plots for two different sub-samples of applicants. Figure 3a plots the correlation between platform score and firm choice for jobs where none of the applicants were found to be above the 0.5 threshold and the firm therefore received no recommendations. We may think of these choices as reflecting the efficacy of the firm's screening process unassisted by the platform's AI. Figure 3a shows a positive correlation between the platform's score and the firm's choices, with the positive slope growing steeper as we move from initial stage to final stage. This suggests that not only would the firm have benefited from access to the platform's scores in these cases, but that the lack of input from the platform's AI had pass through effects into the later stages of the hiring process, with the firm viewing and even interviewing candidates that the platform's AI could have weeded out for them. While the firms in this case did not have the option of taking advantage of the platform's AI (since the platform does not share the scores from its algorithm), the slope of the lines in Figure 3a imply that firms do, on average, benefit from at least partially relying on the platform by taking the recommendations from its AI into account when making hiring decisions, if only because the AI identifies and screens out the weakest candidates.

Insert Figures 3a and 3b about here

Figure 3b shows the correlation between the score of the recommended applicants (in the cases where the platform did offer recommendations) and the firms' choices. Here we see the opposite pattern from the one we saw in Figure 3a[15]: the platform's scores are strongly predictive of

[15] To be clear, Figures 3a and 3b are based on different sub-samples—jobs that did not and did receive recommendations, respectively, so we cannot meaningfully compare the slopes between Figure 3a and 3b. What we can do is compare how the slope of each set of lines changes across the different stage of the hiring process within each figure.

whom the firm views, but largely uncorrelated with the firm's subsequent choices, with the slope of the line in the right panel of Figure 3b being essentially flat. We see this pattern as being entirely consistent with the logic behind a strategy of partial reliance. In the initial stage of deciding which applicant to view, the AI's score matches closely with the firm's own internal preferences, so it makes sense for the firm to make us of the platform's recommendation at this stage (and it may have been even better off if the platform had shared the detailed scores). In the final decision of whom to hire, however, employers rely on their own firm-specific expertise, which is why their choices and the ratings of the platform's AI are essentially uncorrelated.

Overall, the results of these supplementary analyses suggest that the platform's AI is most useful for weeding out weak and unacceptable candidates. Firms are thus well-served by paying attention to the platform's recommendations in the first stage (deciding whom to consider), and are worse off when they get no input from the platform's AI; once these weak candidates have been weeded out, however, the choice between the remaining applicants must be firm-specific, which is why the firm chooses to develop and rely on its own internal expertise rather than depending on the platform's AI. As such, these supplementary analyses offer further support for our theoretical arguments, especially for firms' use of a strategy of partial reliance.

We further investigate the appropriateness of firms' use of the platform's AI by looking at the job performance of the hired freelancers. If firms were mistakenly ignoring the platform's recommendations in their hiring, then the recommended applicants would outperform the non-recommended applicants, even conditional on being hired (Hoffman et al., 2017); conversely, if firms were placing too much reliance on the platform's recommendations, then the non-recommended applicants would outperform the recommended applicants. Figure D1 in Appendix D shows that there is no measurable effect of being recommended on the applicant's work product (measured as employer feedback ratings) once we account for applicant quality (using the RDD). This is as we would expect, and suggests that firms seem to be using the platform's AI capabilities appropriately.

Conclusion and Discussion

Overall, our empirical findings are strongly consistent with firms adopting a strategy of partial reliance when dealing with an AI enabled platform. Not only do firms rely heavily on the platform's recommendations to screen for generally acceptable transaction partners while paying no attention to these recommendations when making their eventual (firm-specific) choice, their reliance on platform recommendations is greater for more general tasks, and decreases as they develop their own internal expertise. Further, supplementary analyses suggest that the primary benefit from the platform's AI is in weeding out the weakest applicants, further supporting partial reliance.

In presenting these findings, we contribute to the existing literature in several ways. To begin with, we draw attention to the potential small numbers bargaining problem associated with the use of big data techniques. While the value creation potential of AI and big data technologies has been widely acknowledged (McAfee and Brynjolfsson, 2012; Brynjolfsson and Mitchell, 2017; Furman and Seamans, 2018; Varian, 2018), relatively less attention has been paid to the value appropriation challenges posed by these technologies, and their implications for organizational strategies and governance. Our study highlights the fundamental trade-off firms face between the need for firm-specificity to maintain a competitive advantage (Mahoney, 2001; Arygres and Zenger, 2012; Kaul, 2013), and the need to pool information to achieve the gains from big data (Ellison and Fudenburg, 2003). In doing so, we not only connect with prior literature that has examined the relationship between platform dominance and differentiation of complementors (Augereau, Greenstein, and Rysman, 2006; Huang et al., 2013)—suggesting that increased investments by complementors in differentiation may be a consequence of being faced with a winner take all platform—we also speak to recent work discussing the anti-trust consequences of AI and the need for more stringent regulation of big data access and use (O'Neil, 2016; Khan, 2017; 2018; Cockburn et al., 2018). In addition, our study also sheds new light on work that has sought to explain the gap between AI's revolutionary potential and the relatively modest productivity gains realized thus far (Brynjolfsson et al., 2017); while this work blames this gap on organizational inertia, our study offers a more structural explanation, suggesting that firms may rationally choose to limit their reliance on AI and big data technologies in order to maintain their firm-specific competitive advantages.

In addition to highlighting the small numbers problem with big data, our study also contributes to our understanding of the strategies firms use to cope with this problem. Specifically, we argue and show that firms may respond to this problem by adopting strategies of partial reliance: using the recommendations generated by the platform's AI to screen for potential transaction partners, but complementing them with their own internal expertise when making a final choice. As such, our study builds on and extends the idea of absorptive capacity (Cohen and Levinthal, 1990) to the context of AI and big data, suggesting that to realize and appropriate the value of a platform's big data algorithms, firms need to invest in developing and maintaining their own internal specialized capabilities. By showing that firms use third-party AI to substitute for general tasks but complement firm-specific tasks, we also provide fresh insight into the specific ways in which firms make use of big data techniques, thus contributing to prior research on the use of computers and automation as complements or substitutes more generally (Autor, Levy, and Murnane, 2003; Autor, 2015; Furman and Seamans, 2018).

Finally, our study also contributes to the growing literature on strategies for two-sided platforms. While much of this literature focuses on strategies for platform owners—including choices around pricing (Hagiu, 2009; Rysman, 2009; Seamans and Zhu, 2013), entry and investment timing (Eisenmann, 2006; Zhu and Iansiti, 2012), openness (Cennamo and Santalo, 2013; Kretschmer and Claussen, 2016), complexity (Kapoor and Agarwal, 2017), and product mix (Seamans and Zhu, 2017)—we focus on the strategies firms use to engage with platforms (Huang et al., 2013; Zhu and Liu, 2018; Kapoor and Agarwal, 2018), specifically, the ways in which they deal with platforms as buyers. As already mentioned, our study is also novel in that it focuses on two-sided matching platforms where users make little or no platform-specific investments (Zhu and Liu, 2018), in contrast to the bulk of the prior strategy literature that focuses on technology platforms where such platform-specific investments and interdependencies between users are critical (Gawer, 2014; Hagiu and Wright, 2015; Helfat and Raubitschek, 2018). Our study thus advances our understanding of a distinct type of two-sided platform, one that lies at the heart of the so-called 'gig economy' (Barley and Kunda, 2004; Edelman et al., 2017). In particular, it sheds new light on the strategies firms use when hiring from online labor markets, which are an increasingly important source of human capital for firms (Leung, 2014; Horton, 2010; 2017; Agrawal et al., 2015; Hoffman et al., 2017; Chan and Wang, 2018). Moreover, we highlight the strategic implications of the use of AI and big data technologies in the context of such platform strategies, thus serving as a bridge between the platform strategy literature and the work on AI and automation discussed earlier.

As with any study, ours has several limitations. As mentioned, we are only able to observe the transactions firms undertake on our focal platform, and can therefore only examine the intensive margin of AI use on the platform. While firms may use other platforms or sources to search for and hire employees—and our theory would suggest that they may do so to avoid giving our focal platform too much bargaining power—we are unable to observe such activities, and therefore cannot really speak to firms' choice to hire on the platform in the first place. Future work could certainly explore the extensive margin of platform use further; in particular, it could build on our theory to examine the point where a firm switches from relying on the platform's AI to building its own. In addition, our focus on a single platform also potentially limits the generalizability of our findings. While the specific design of the focal platform—the fact that it only tells a potential employer whether an applicant is above the 50% threshold or not—is helpful for achieving stronger causal identification, it is certainly possible that firms may use AI recommendations from other platforms differently. More work is thus needed to replicate our findings across other platforms and contexts. Future work could also look at how the strategies of partial reliance we discuss here play out in the context of technology platforms (Huang et al., 2013; Gawer, 2014; Agarwal and Kapoor, 2018).

To conclude, we highlight the small numbers problem associated with big data, with the increasing concentration of transactions on a single platform serving to realize big data's value creation potential but also giving rise to value appropriation hazards for firms relying on the platform to find transaction partners. We argue that firms will respond to this problem by adopting strategies of partial reliance, whereby they take advantage of the platform's big data capabilities to screen for transaction partners that meet their general criteria, but then use their internal expertise to choose that partner that offers the best firm-specific match. Evidence from an online labor marketplace is consistent with this prediction, with firms relying on the platform's recommendations when choosing which applicants to consider, but ignoring these recommendations when deciding whom to hire from among those they interview, and this effect being weaker for more specialized jobs and more experienced employers. Our study thus contributes to research on the use of AI and big data technologies—highlighting the value appropriation consequences of these technologies and the strategies firms use to cope with them—as well as to our understanding of how firms maintain their competitive advantage while engaging with two-sided matching platforms.

References

Agarwal, S. & Kapoor, R. (2018). Two faces of value creation in business ecosystems: Leveraging complementarities and managing interdependencies. Working Paper, Wharton School.

Agrawal, A., Horton, J., Lacetera, N., & Lyons, E. (2015). Digitization and the contract labor market: A research agenda. In *Economic analysis of the digital economy* (pp. 219-250). University of Chicago Press.

Argyres, N. S., & Zenger, T. R. (2012). Capabilities, transaction costs, and firm boundaries. *Organization Science*, *23*(6), 1643-1657.

Augereau, A., Greenstein, S., & Rysman, M. (2006). Coordination versus differentiation in a standards war: 56k modems. *The RAND Journal of Economics*, *37*(4), 887-909.

Autor, D. H. (2015). Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives*, *29*(3), 3-30.

Autor, D. H., & Scarborough, D. (2008). Does job testing harm minority workers? Evidence from retail establishments. *The Quarterly Journal of Economics*, *123*(1), 219-277.

Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly journal of economics*, *118*(4), 1279-1333.

Barley, S. R., & Kunda, G. (2004). *Gurus, hired guns, and warm bodies: Itinerant experts in a knowledge economy*. Princeton University Press. Princeton: NJ

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company.

Brynjolfsson, E., & McElheran, K. (2016). The rapid adoption of data-driven decision-making. *American Economic Review*, *106*(5), 133-39.

Brynjolfsson, E., & Mitchell, T. (2017). What can machine learning do? Workforce implications. *Science*, *358*(6370), 1530-1534.

Brynjolfsson, E., & Smith, M. D. (2000). Frictionless commerce? A comparison of Internet and conventional retailers. *Management science*, *46*(4), 563-585.

Brynjolfsson, E., Rock, D., & Syverson, C. (2017). Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics. In *The Economics of Artificial Intelligence: An Agenda* (Agarwal, A.K., Gans, J., and Goldfarb, A. eds.). University of Chicago Press forthcoming.

Caillaud, B., & Jullien, B. (2003). Chicken & egg: Competition among intermediation service providers. *RAND journal of Economics*, 34(2), 309-328.

Cennamo, C., & Santalo, J. (2013). Platform competition: Strategic trade‐offs in platform markets. *Strategic management journal*, *34*(11), 1331-1350.

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

Choi, J. P. (2010). Tying in two‐sided markets with multi‐homing. *The Journal of Industrial Economics*, *58*(3), 607-626.

Choudhury, P., Starr, E., & Agarwal, R. (2018). Machine Learning and Human Capital: Experimental Evidence on Productivity Complementarities. Available at SSRN: https://ssrn.com/abstract=3185022

Cockburn, I. M., Henderson, R., & Stern, S. (2018). *The Impact of Artificial Intelligence on Innovation* (No. w24449). National Bureau of Economic Research.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive Capacity: A New Perspective On Learning And Innovation. *Administrative Science Quarterly*, *35*(1), 1281-52

Cowgill, B. (2018). Bias and Productivity in Humans and Algorithms: Theory and Evidence from Resume Screening. *Columbia Business School, Columbia University*, *29*.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, 243, 1668-1674.

Edelman, B.G., Luca, M., Svirsky, D. 2017. Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment. *American Economic Journal: Applied Economics*. 9(2) 1-22.

Eisenmann T, Parker G, Van Alstyne MW (2006) Strategies for two-sided markets. Harvard Business Review 84(10):92.

Eisenmann, T. R. (2006). Internet companies' growth strategies: determinants of investment intensity and long-term performance. *Strategic Management Journal*, *27*(12), 1183-1204.

Eisenmann, T. R. (2008). Managing proprietary and shared platforms. *California Management Review*, *50*(4), 31-53.

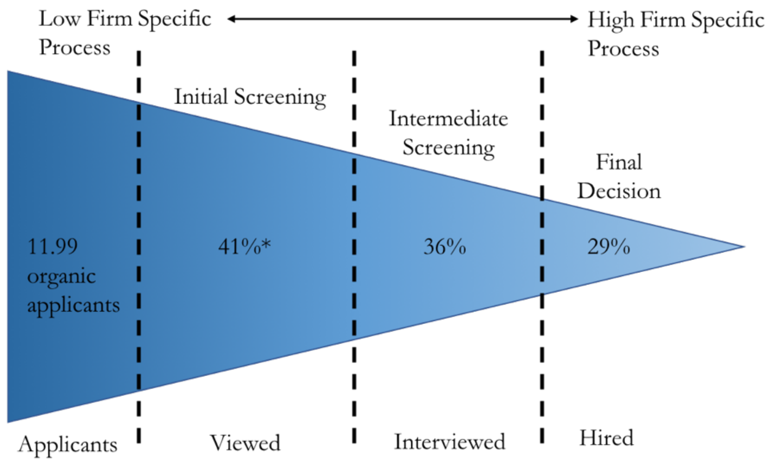
Ellison, G., & Fudenberg, D. (2003). Knife-edge or plateau: When do market models tip?. *The Quarterly Journal of Economics*, *118*(4), 1249-1278.

Furman, J., & Seamans, R. (2018). *AI and the Economy* (No. w24689). National Bureau of Economic Research.

Gartside, D., Yang, C., Sloman, C., & Cantrell, S. M. (2013). Trends reshaping the future of HR. *Accenture: Accenture Institute for High Performance*, 1-16.
Gawer, A. (2014). Bridging differing perspectives on technological platforms: Toward an integrative framework. *Research policy*, *43*(7), 1239-1249.
Gawer, A., & Henderson, R. (2007). Platform owner entry and innovation in complementary markets: Evidence from Intel. *Journal of Economics & Management Strategy*, *16*(1), 1-34.
Gomez-Uribe, C. A., & Hunt, N. (2016). The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)*, *6*(4), 13.
Greenwood, B.N., Wattal, S. 2017. Show Me the Way to Go Home: An Empirical Investigation of Ride Sharing and Alcohol Related Motor Vehicle Fatalities. *MIS Quarterly*. 41(1) 163-187
Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: a meta-analysis. *Psychological assessment*, *12*(1), 19.
Hagiu, A. (2009). Two-sided platforms: product variety and pricing structures. *Journal of Economics & Management Strategy*, *18*(4), 1011-1043.
Hagiu, A., & Wright, J. (2015). Multi-sided platforms. *International Journal of Industrial Organization*, *43*, 162-174.
Hahn, J., Todd, P., & Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, *69*(1), 201-209.
Helfat, C. E., & Raubitschek, R. S. (2018). Dynamic and integrative capabilities for profiting from innovation in digital platform-based ecosystems. *Research Policy*, *47*(8), 1391-1399.
Hoffman, M., Kahn, L. B., & Li, D. (2017). Discretion in hiring. *The Quarterly Journal of Economics*, *133*(2), 765-800.
Horton, J. J. (2010, December). Online labor markets. In *International Workshop on Internet and Network Economics* (pp. 515-522). Springer, Berlin, Heidelberg.
Horton, J. J. (2017). The effects of algorithmic labor market recommendations: Evidence from a field experiment. *Journal of Labor Economics*, *35*(2), 345-385.
Huang, P., Ceccagnoli, M., Forman, C., & Wu, D. J. (2013). Appropriability mechanisms and the platform partnership decision: Evidence from enterprise software. *Management Science*, *59*(1), 102-121.
Imbens, G., & Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of economic studies*, *79*(3), 933-959.
Kahnemann, D. (2011). *Thinking. Fast and Slow*. Farrar, Straus and Giroux, New York USA.
Kapoor, R., & Agarwal, S. (2017). Sustaining superior performance in business ecosystems: Evidence from application software developers in the iOS and Android smartphone ecosystems. *Organization Science*, *28*(3), 531-551.
Katz, L. F., & Krueger, A. B. (2016). *The rise and nature of alternative work arrangements in the United States, 1995-2015* (No. w22667). National Bureau of Economic Research.
Katz, M. L., & Shapiro, C. (1985). Network externalities, competition, and compatibility. *The American economic review*, *75*(3), 424-440.
Kaul, A. (2013). Entrepreneurial action, unique assets, and appropriation risk: Firms as a means of appropriating profit from capability creation. *Organization Science*, *24*(6), 1765-1781.
Khan, L. (2018). The Separation of Platforms and Commerce. *Columbia Law Review*, Vol. 119, Forthcoming. Available at SSRN: https://ssrn.com/abstract=3180174
Khan, L.M. (2017). Amazon's antitrust paradox. *Yale Law Journal* 126, 710-805
Klein, B., Crawford, R. G., & Alchian, A. A. (1978). Vertical integration, appropriable rents, and the competitive contracting process. *The journal of Law and Economics*, *21*(2), 297-326.
Kretschmer, T., & Claussen, J. (2016). Generational transitions in platform markets—The role of backward compatibility. *Strategy Science*, *1*(2), 90-104.
Kuncel, N. R., Klieger, D. M., Connelly, B. S., & Ones, D. S. (2013). Mechanical versus clinical data combination in selection and admissions decisions: A meta-analysis. *Journal of Applied Psychology*, *98*(6), 1060.
Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of economic literature*, *48*(2), 281-355.
Leung, M. D. (2014). Dilettante or renaissance person? How the order of job experiences affects hiring in an external labor market. *American Sociological Review*, *79*(1), 136-158.
Mahoney, J. T. (2001). A resource-based theory of sustainable rents. *Journal of Management*, *27*(6), 651–660.
McAfee, A., Brynjolfsson, E. (2012). Big data: the management revolution. *Harvard business review*, *90*(10), 60-68.
Mullainathan, S., & Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87-106.

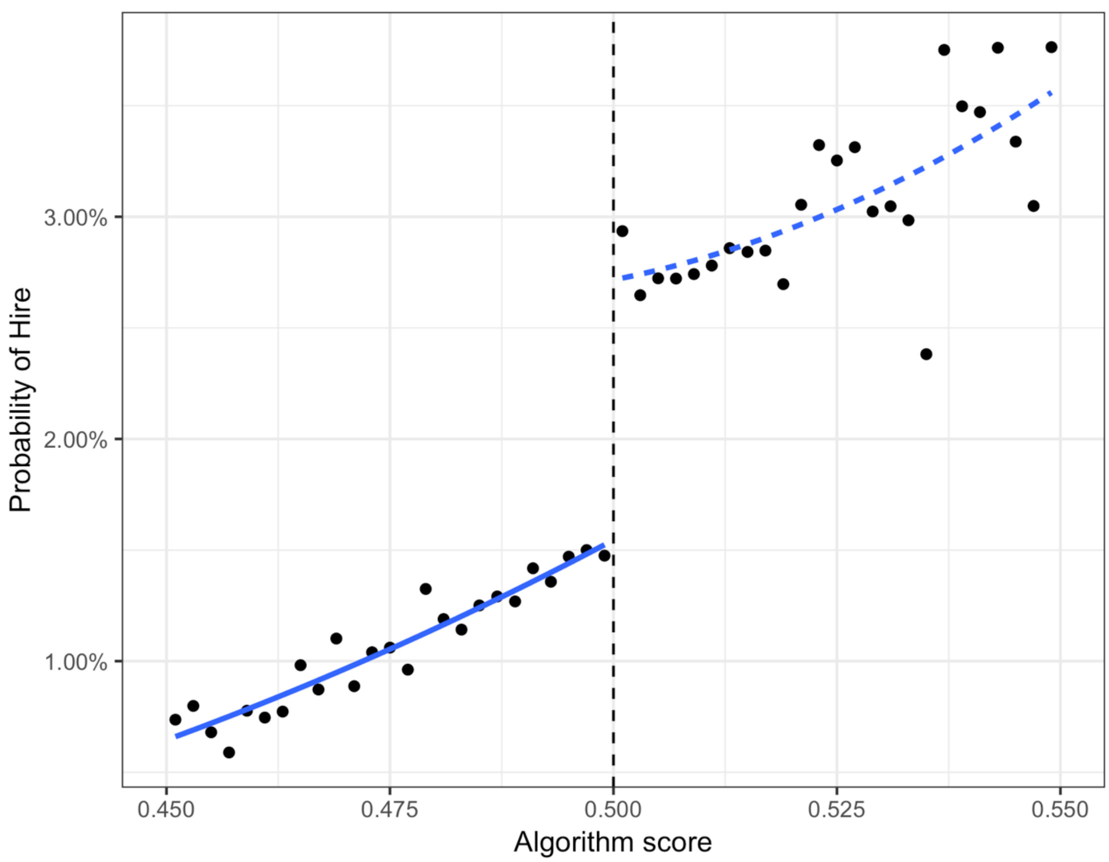
O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books.
Park, K. F., Seamans, R., & Zhu, F. (2017). Multi-homing and Platform Strategies: Historical Evidence from the US Newspaper Industry. Available at SSRN: https://ssrn.com/abstract=3048348
Parker, G. G., & Van Alstyne, M. W. (2005). Two-sided network effects: A theory of information product design. *Management science*, *51*(10), 1494-1504.
Rochet, J. C., & Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the european economic association*, *1*(4), 990-1029.
Rochet, J. C., & Tirole, J. (2006). Two-sided markets: a progress report. *The RAND journal of economics*, *37*(3), 645-667.
Rysman, M. (2009). The economics of two-sided markets. *Journal of economic perspectives*, *23*(3), 125-43.
Seamans, R., & Zhu, F. (2014). Responses to entry in multi-sided markets: The impact of Craigslist on local newspapers. *Management Science*, *60*(2), 476-493.
Seamans, R., & Zhu, F. (2017). Repositioning and cost-cutting: The impact of competition on platform strategies. *Strategy Science*, *2*(2), 83-99.
Teece, D. J. (2018). Profiting from innovation in the digital economy: Enabling technologies, standards, and licensing models in the wireless world. *Research Policy*. 47(8), 1367-1387
Tilcsik, A. (2011). Pride and prejudice: Employment discrimination against openly gay men in the United States. *American Journal of Sociology*, *117*(2), 586-626.
Varian, H. (2018). Artifical intelligence, economics, and industrial organization. In *The Economics of Artificial Intelligence: An Agenda* (Agarwal, A.K., Gans, J., and Goldfarb, A. eds.). University of Chicago Press forthcoming.
Williamson OE. 1975. *Markets and Hierarchies*. New York, NY: Free Press.
Williamson OE. 1985. *The Economic Institutions of Capitalism.* New York, NY: Free Press.
Zervas, G., Proserpio, D., Byers, J. 2015. The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry. Boston U. School of Management Research Paper(2013-16).
Zhu, F., & Iansiti, M. (2012). Entry into platform-based markets. *Strategic Management Journal*, *33*(1), 88-106.
Zhu, F., Liu Q. (2018). Competing with complementors: An empirical look at Amazon.com. *Strategic Management Journal.* 39, 2618-2642

Figure 1: The Hiring Process



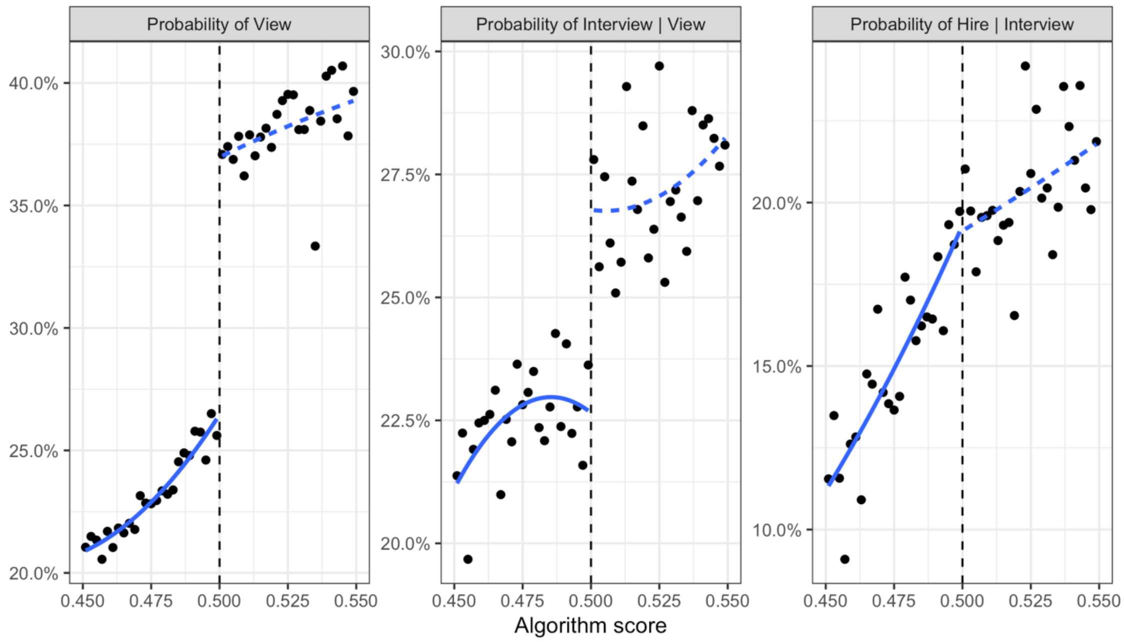
*Percentage of prior stage applicants selected on average. Note 3% of interviewed applicants are not viewed, and 29% of hired applicants are not interviewed. Fewer than 1% of applicants are hired without the firm even viewing the applicants CV.

Figure 2a: Regression Discontinuity Plot of Probability of Hire



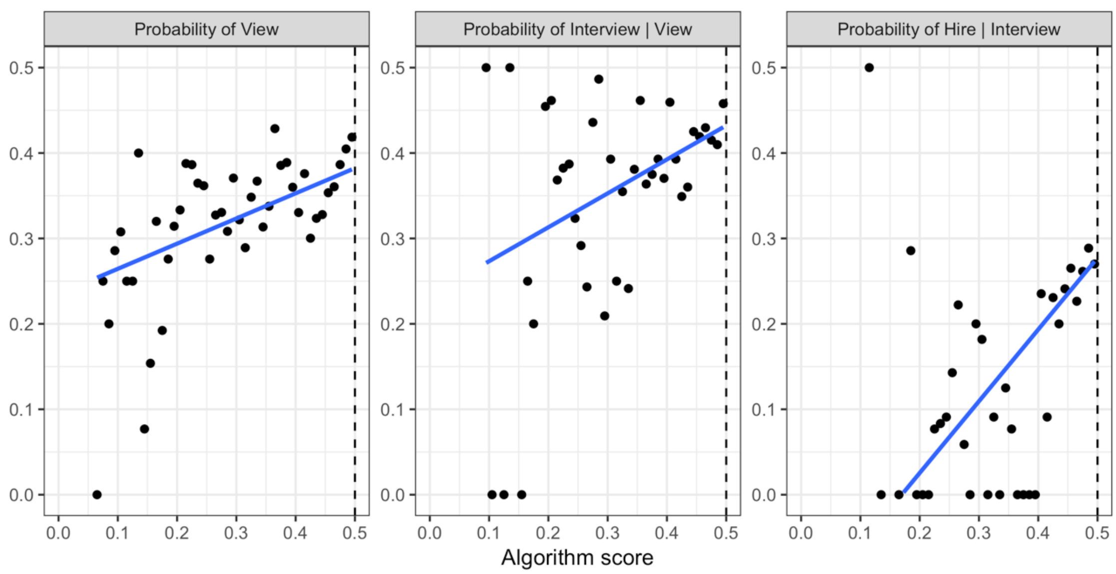
Notes: Each point equals the mean probability of hire for bins of size 0.002.
Fitted line is from separate quadratic fits on either side of recommendation threshold (.05).

Figure 2b: Regression Discontinuity Plot of Hiring Process



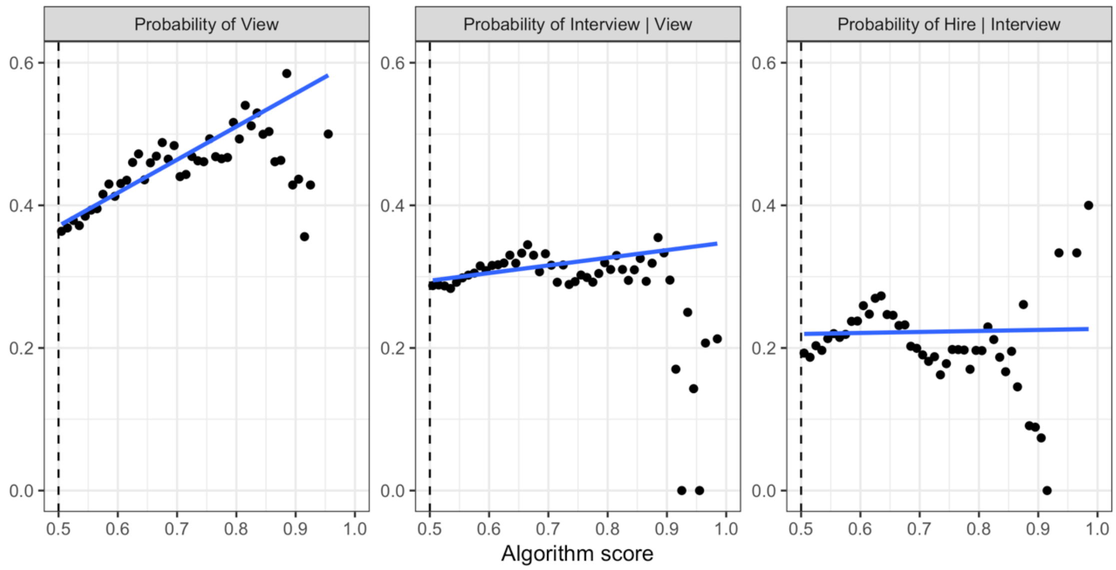
Notes: Each point equals the mean probability of view, probability of interviewing viewed applicants, or probability of hiring interviewed applicants for bins of size 0.002.
Fitted line is from separate quadratic fits on either side of recommendation threshold (.05).

Figure 3a: Supplementary Analysis of platform AI – Non-recommended applicants



Notes: Each point represents the mean probability for bins of size 0.01. Bins with only 1 observation are dropped.
Fitted line is from weighted linear fits. Weights are added for the number of observations in each bin.
Data is from jobs with no recommended applicants.

Figure 3b: Supplementary Analysis of platform AI – Recommended applicants



Notes: Each point represents means for bins of size 0.01. Bins with only 1 observation are dropped.
Fitted line is from weighted linear fits. Weights are added for the number of observations in each bin.
Data is from recommended applicants to jobs with greater than 1 recommended applicant.

Table 1: Opening-level Summary Statistics, Organic Applications Only (n=125,300)

Statistic	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Job Specific Skills Requested (YN)	0.56	0.50	0	0	1	1	1
Avg. Hr. Bid ($)	22.08	9.97	3.00	15.50	20.53	27.17	99.99
Avg. Algorithm Score	0.52	0.05	0.002	0.49	0.52	0.54	0.90
Num. Employers prior filled jobs	16.85	38.22	0	0	2	13	323
Num. Applicants	12.12	12.92	1	4	8	16	112
Num. Recommended	6.37	7.19	0	2	4	8	86
Num. Viewed	4.02	5.96	0	1	2	5	96
Num. Interviewed	1.18	2.12	0	0	0	2	69
Num. Hired	0.32	0.53	0	0	0	1	4

Notes: This table reports summary statistics on the non-invited applicant pool. All reports are on a per-opening basis. An application is "recommended" if the applicant is marked by the ML algorithm as a recommended applicant. An applicant is "interviewed" if the employer sent a message to the applicant. Top 1% of applications by hr. bid and employers prior filled jobs are trimmed.

Table 2: Opening-level Summary Stats, Organic Applications
with algorithm scores in [0.45,0.55] (n= 53,134)

Statistic	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Job Specific Skills Requested (YN)	0.56	0.50	0	0	1	1	1
Avg. Hr. Bid ($)	22.09	10.89	3.00	15.00	20.00	27.50	99.99
Avg. Algorithm Score	0.49	0.02	0.45	0.48	0.49	0.50	0.55
Num. Employers prior filled jobs	7.20	22.36	0	0	0	4	320
Num. Applicants	6.38	6.58	1	2	4	8	73
Num. Recommended	2.51	2.87	0	1	2	3	28
Num. Viewed	1.85	2.88	0	0	1	2	50
Num. Interviewed	0.48	1.06	0	0	0	1	38
Num. Hired	0.12	0.34	0	0	0	0	4

Notes: This table reports summary statistics on the non-invited applicant pool. Applications with algorithm scores greater than 0.55 and less than 0.45 are removed. Jobs with no recommended applicants have been removed from the sample. All reports are on a per-opening basis. An application is "recommended" if the applicant is marked by the ML algorithm as a recommended applicant. An applicant is "interviewed" if the employer sent a message to the applicant. Top 1% of applications by hr. bid and employers prior filled jobs are trimmed.

Table 3: Effect of AI recommendation on hiring process

	Dependent variable:			
	Applicant Hired	Applicant Viewed	Applicant Interviewed|View	Applicant Hired|Interview
	(1)	(2)	(3)	(4)
Recommended	0.009***	0.102***	0.019**	0.007
	(0.001)	(0.003)	(0.008)	(0.013)
Norm. Score (< 0)	0.121***	0.924***	0.636***	−0.095
	(0.020)	(0.064)	(0.212)	(0.232)
Norm. Score (>= 0)	0.209***	0.642***	0.559**	0.074
	(0.035)	(0.086)	(0.217)	(0.193)
Sample	All Applicants	All Applicants	Viewed Applicants	Interviewed Applicants
Job-level FE	Yes	Yes	Yes	Yes
I-K Optimal Bandwidth	0.049	0.049	0.043	0.08
Dep. Var. Mean	0.018	0.289	0.249	0.186
Baseline	0.015	0.256	0.236	0.197
Observations	330,235	332,444	83,806	37,219

Notes: The bandwidth used in each model was calculated using the Imbens-Kalyanaraman Optimal Bandwidth Calculation. The estimations are from a local-linear model which allows for slope differences on either side of the cutoff. Controls include applicant's bid on the job posting, default sort rank fixed effects, and job opening fixed effects.Baseline is dependent variable mean for applications within [-0.02, 0) in normalized algorithm score. Model (1) the outcome is an indicator if the applicant is hired. Model (2) the outcome is an indicator if the applicant is viewed. Model (3) the outcome is an indicator if the applicant is interviewed conditional on the applicant being viewed. Model (4) the outcome is an indicator if the applicant is hired conditional on the applicant being interviewed. Heteroskedasticity-robust standard errors clustered at the job posting level are reported. Significance indicators: $p \leq 0.10 : *$, $p \leq 0.05 : **$, and $p \leq .01 : ***$

Table 4: The effect of AI recommendation by firm requested job-specific skills

	Dependent variable:			
	Applicant Hired	Applicant Viewed	Applicant Interviewed|View	Applicant Hired|interview
	(1)	(2)	(3)	(4)
Recommended (rec)	0.010***	0.109***	0.025**	0.004
	(0.002)	(0.005)	(0.012)	(0.020)
Job-specific skill requested?	0.003	−0.001	0.056	0.080
	(0.006)	(0.025)	(0.051)	(0.057)
rec x skill requested?	−0.001	−0.013**	−0.010	0.004
	(0.002)	(0.006)	(0.015)	(0.026)
Sample	All Applicants	All Applicants	Viewed Applicants	Interviewed Applicants
Job-level FE	No	No	No	No
Employer FE	Yes	Yes	Yes	Yes
Sort order FE	Yes	Yes	Yes	Yes
I-K Optimal Bandwidth	0.049	0.049	0.043	0.08
Dep. Var. Mean	0.018	0.289	0.249	0.186
Baseline	0.015	0.256	0.236	0.197
Observations	330,235	332,444	83,806	37,219

Notes: The bandwidth used in each model was calculated using the Imbens-Kalyanaraman Optimal Bandwidth Calculation. The estimations are from a local-linear model which allows for slope differences on either side of the cutoff.The omitted category are jobs which requested no job-specific skills.Controls include applicant's bid on the job posting, category indicators, number of applicants, number of applicants sq., job value group indicators, hourly or fp indicators, num. previous jobs posted by employer, the number of recommended organic applicants, employer FE and application default sorting FE. Baseline is dependent variable mean for applications within [-0.02, 0) in algorithm score. Model (1) the outcome is an indicator if the applicant is hired. Model (2) the outcome is an indicator if the applicant is viewed. Model (3) the outcome is an indicator if the applicant is interviewed conditional on the applicant being viewed. Model (4) the outcome is an indicator if the applicant is hired conditional on the applicant being interviewed. Significance indicators: $p \leq 0.10 : *$, $p \leq 0.05 : **$, and $p \leq .01 : ***$

Table 5: The effect of AI recommendation by employer's prior experience

	Dependent variable:			
	Applicant Hired	Applicant Viewed	Applicant Interviewed|View	Applicant Hired|interview
	(1)	(2)	(3)	(4)
Recommended (rec)	0.014***	0.120***	0.040***	0.036***
	(0.002)	(0.005)	(0.008)	(0.010)
1-2 prior filled jobs	0.006***	−0.023***	0.016	0.037*
	(0.002)	(0.008)	(0.014)	(0.019)
3-14 prior filled jobs	0.007***	−0.027***	0.036***	0.046***
	(0.002)	(0.007)	(0.013)	(0.017)
15+ filled jobs	0.010***	−0.006	0.061***	0.062***
	(0.002)	(0.009)	(0.015)	(0.019)
rec x 1-2 prior jobs	−0.003	0.006	−0.007	−0.011
	(0.003)	(0.010)	(0.018)	(0.024)
rec x 3-14 prior jobs	−0.003	−0.026***	−0.010	−0.017
	(0.003)	(0.009)	(0.017)	(0.021)
rec x 15+	−0.005	−0.051***	−0.029	−0.026
	(0.004)	(0.011)	(0.019)	(0.024)
Sample	All Applicants	All Applicants	Viewed Applicants	Interviewed Applicants
Job-level FE	No	No	No	No
Employer FE	No	No	No	No
Sort order FE	Yes	Yes	Yes	Yes
I-K Optimal Bandwidth	0.047	0.049	0.044	0.077
Dep. Var. Mean	0.018	0.289	0.25	0.184
Baseline	0.015	0.255	0.236	0.196
Observations	288,004	299,146	81,170	33,552

Notes: The bandwidth used in each model was calculated using the Imbens-Kalyanaraman Optimal Bandwidth Calculation. The estimations are from a local-linear model which allows for slope differences on either side of the cutoff. The omitted category are employers who have never completed an on-platform job. Controls include applicant's bid on the job posting, category indicators, number of applicants, number of applicants sq., job value group indicators, hourly or fp indicators, job-skill request indicator, the number of recommended organic applicants, and application sort order FE. Baseline is dependent variable mean for applications within [-0.02, 0) in algorithm score. Model (1) the outcome is an indicator if the applicant is hired. Model (2) the outcome is an indicator if the applicant is viewed. Model (3) the outcome is an indicator if the applicant is interviewed conditional on the applicant being viewed. Model (4) the outcome is an indicator if the applicant is hired conditional on the applicant being interviewed. Significance indicators: $p \leq 0.10 : *$, $p \leq 0.05 : **$, and $p \leq .01 : ***$.

Appendix A: Platform Hiring Details

Figure A1 shows how a typical job posting might appear, once it is posted to the platform.

Insert Figure A1 about here

Applying to a job

Applicants can apply to any public job posting on the platform. When they apply, they include a bid (the amount they are willing to work for), and a cover letter, which consists of a paragraph of text meant to convince the employer that the applicant is "right" for the job. After applying, the applicant immediately appears in the employer's "applicant tracking system" or ATS, which is the dashboard the employer sees after logging into their account and clicking on an open job posting. Each application in the ATS shows the applicant's' name, picture, bid, self-reported skills, country and a few pieces of platform-verified information, such as the total number of hours-worked and the average feedback rating from previous projects (if any). Figure A2 shows the employers' view of the "applicant tracking system" (ATS).

Insert Figure A2 about here

Viewing Applications and Interviewing Applicants

An employer is able to learn more about a specific applicant and view that applicant's detailed application by clicking on any of the job proposals which are listed in the ATS. Viewing an applicant's detailed application is similar to reading a conventional applicant resume. While it is possible for an applicant to be hired without the employer reviewing the detailed application, this is extremely rare. In our data only 1.8% of hired applicants are hired without being viewed by the employer. An employer who views a detailed application sees more information about that applicant, including their self-written overview, as well as their past work history. Figure A3 shows how a detailed application would look to an employer on the platform. The applicant self-categorizes as a “Photographer and Digital Retoucher”, and states that he is skilled in: “Adobe Photoshop”, “Photo Editing,” and a few extra skills. Employers can also review the applicants work history and feedback on this page. This work history contains an item for each job the applicant has completed on the platform and contains both the review and feedback left by the applicant’s previous employer concerning that job.

Insert Figure A3 about here

After viewing an application, the employer can either message the applicant to conduct an “interview” or directly hire the applicant. The website encourages employers to interview their applicants and about 70% of hired applicants are interviewed. Figure 4 shows the prompt through which employers can initiate the interview process via messages.

Insert Figure A4 about here

Hiring Applicants

An employer is free to hire whomever they wish. The employer hires the worker on the terms proposed by the worker or makes a counter offer, which the worker can accept, reject or negotiate. On average, only 43% of job postings are filled. If an employer chooses to hire anyone, 90% of the time they hire only one freelancer, although they have the possibility to hire more. Once a freelancer is hired, employer and employee exchange job details and the job is completed virtually. Payment is conducted through the website.

Figure A1: Job Post for Java Backend Developer

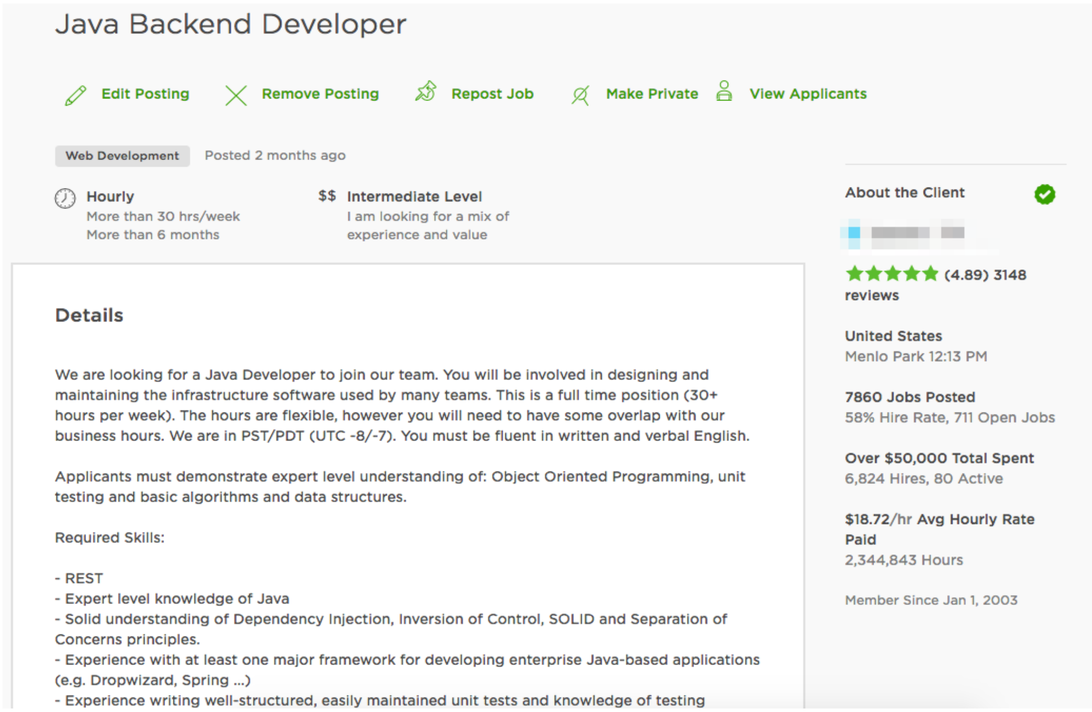


Figure A2: The Application Tracking System (ATS)

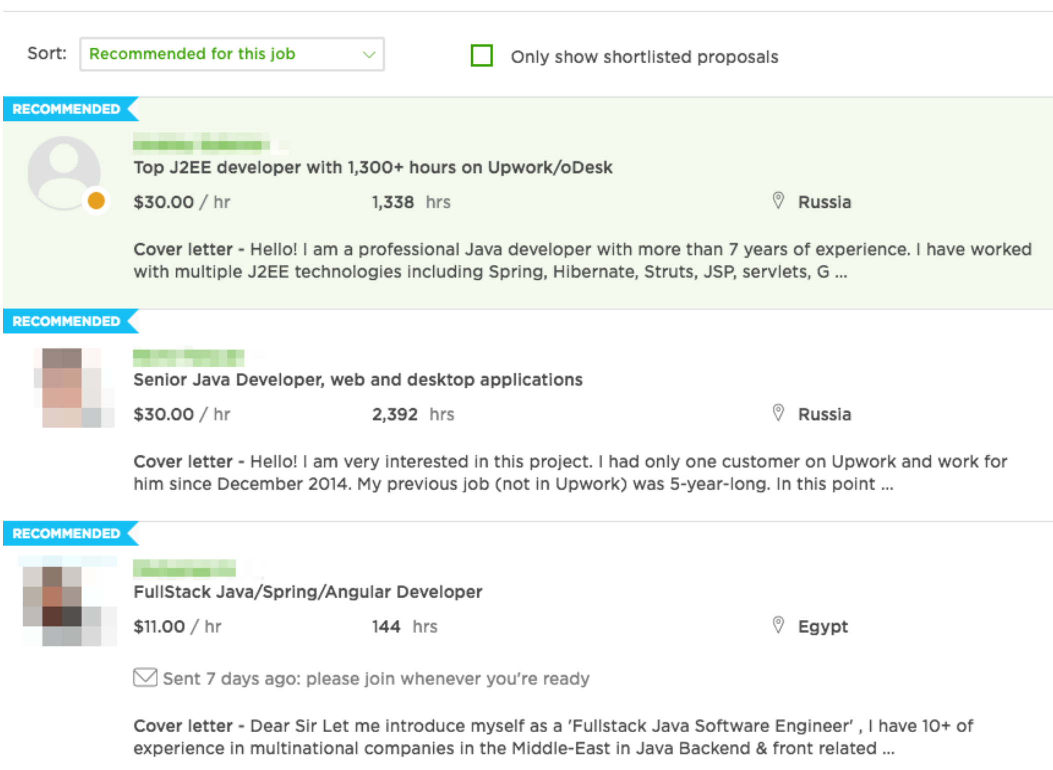


Figure A3: Viewed Application

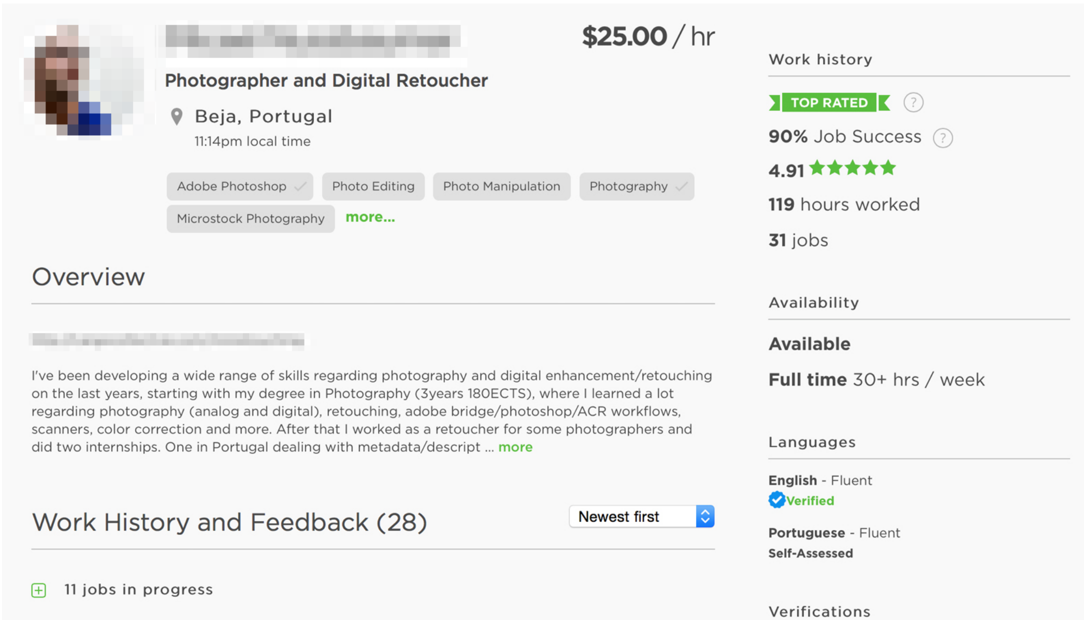
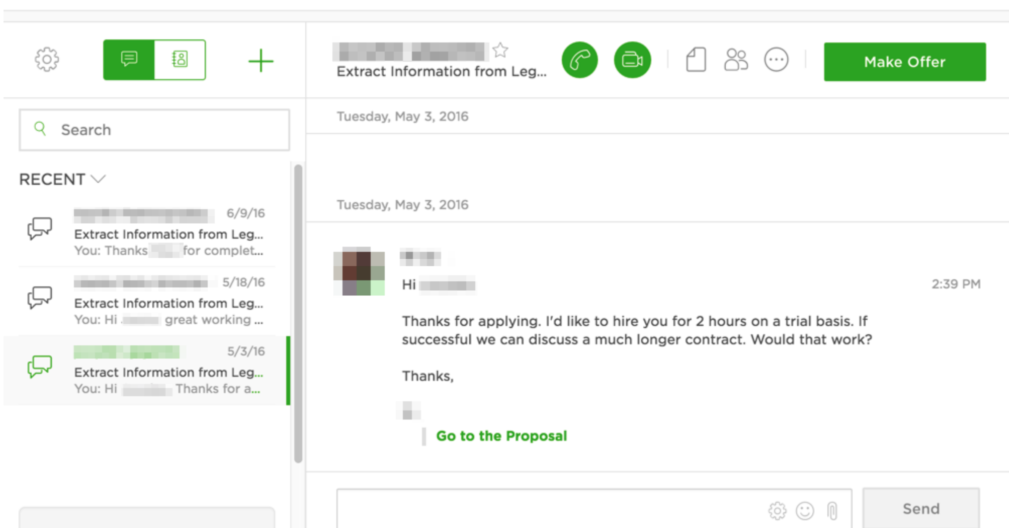


Figure A4: Applicant messaging (interview) screen



Appendix B: Bandwidth robustness checks

Table B1 shows that the results of Table 3 are robust to both bandwidth choices and model specifications. The results do not substantively differ from those presented in table 3 indicating that our main specification is robust to different choices of both bandwidth or functional form.

Table B1

Regression Results Sensitivity Analysis

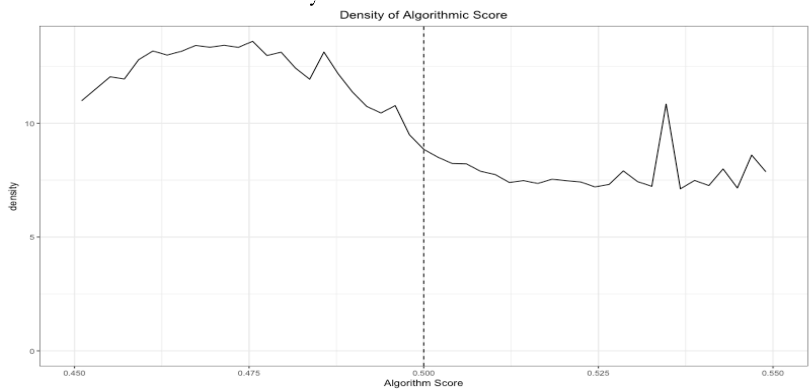
Bandwidth	View	Interview	Interview|View	Hired|Interview	Hired
Linear Model					
0.02	0.096 (0.006)	0.029 (0.004)	0.014 (0.014)	0.003 (0.05)	0.007 (0.002)
0.03	0.096 (0.004)	0.032 (0.003)	0.024 (0.01)	-0.018 (0.031)	0.008 (0.002)
0.04	0.101 (0.004)	0.034 (0.002)	0.019 (0.008)	-0.015 (0.022)	0.009 (0.001)
0.05	0.101 (0.003)	0.033 (0.002)	0.017 (0.007)	-0.011 (0.018)	0.009 (0.001)
0.06	0.101 (0.003)	0.034 (0.002)	0.018 (0.006)	-0.008 (0.015)	0.01 (0.001)
0.07	0.103 (0.003)	0.035 (0.002)	0.023 (0.005)	0.003 (0.014)	0.012 (0.001)
0.08	0.103 (0.003)	0.036 (0.002)	0.027 (0.005)	0.006 (0.013)	0.012 (0.001)
Quadratic Model					
0.02	0.099 (0.009)	0.028 (0.006)	0.012 (0.022)	0.022 (0.076)	0.008 (0.003)
0.03	0.098 (0.006)	0.032 (0.004)	0.017 (0.015)	-0.01 (0.046)	0.009 (0.002)
0.04	0.097 (0.005)	0.033 (0.003)	0.029 (0.012)	-0.027 (0.034)	0.009 (0.002)
0.05	0.1 (0.005)	0.035 (0.003)	0.029 (0.01)	-0.003 (0.028)	0.009 (0.002)
0.06	0.102 (0.004)	0.034 (0.003)	0.021 (0.009)	-0.019 (0.024)	0.009 (0.002)
0.07	0.099 (0.004)	0.032 (0.002)	0.017 (0.008)	-0.022 (0.021)	0.009 (0.001)

Notes: The results are from local linear and local quadratic regressions allowing for slope differences across the threshold. The bandwith was varied between .02 and .08 indicating that the results of table 3 are robust to bandwith and model specifications. Covariates used include: job opening fixed effects, default sort order FE, and the applicant's bid. Heteroskedasticity-robust standard errors clustered at the job posting level are reported.

Appendix C: Balance Checks

Figure C1

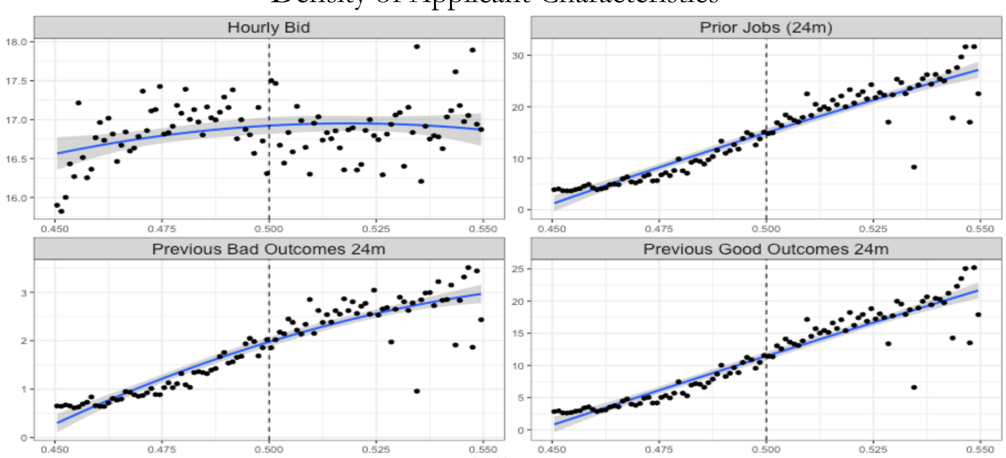
Density of Best Match scores



Notes: Figure C1 shows the density of best match scores across applications in our sample. The lack of bunching above 0.5[16] indicates that applicants at the margin are unable to manipulate whether or not they are recommended. Since the algorithm is a black box to its users, this is not surprising. Moreover, the algorithm relies on signals that are hard for users to manipulate, such as verified work history and past feedback.

Figure C2

Density of Applicant Characteristics



Notes: Figure C2 shows the distribution of applicant's characteristics including the applicant's' bid (top left), the applicant's number of previous outcomes (top right), the applicants' number of previous bad outcomes (lower left) and the applicants' number of previous good outcomes (lower right) by the applicant's algorithm score. The lack of bunching around the discontinuity indicates that applicants are not able to manipulate their algorithmic score, and that close to the discontinuity, there is no measurable difference in the previous outcomes results of applicants above and below the threshold.

---

[16] McCrary (2008). "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test". Journal of Econometrics. 142 (2): 698–714.

Appendix D: Job Outcomes

As the algorithm has such large effects on the hiring behavior of employers, it is logical to question if the machine learning algorithm leads to differential job outcomes. As we are comparing algorithm recommended applicants to equivalent quality non-algorithm recommended applicants we do not expect there to be any measurable effect on the quality of the work product. However, if expectations are lower for non-recommended applicants than recommended applicants it is possible that we might detect differences in job satisfaction between employers who hired recommended applicants who are just above the threshold, and employers who hired non-recommended applicants who are just below the threshold. Figure D1 shows that while there is a large positive relationship between algorithm score and job satisfaction overall, we find no discontinuity across the recommendation threshold in job outcomes, as measured from employer feedback ratings.

Figure D1: Effect of the recommendation on job outcomes