# Steering in Online Markets:
# The Role of Platform Incentives and Credibility

Moshe A. Barach[*]     Joseph M. Golden[†]     John J. Horton[‡]

June 27, 2019

## Abstract

Platform marketplaces can potentially steer buyers to certain sellers by recommending or guaranteeing those sellers. Money-back guarantees—which create a direct financial stake for the platform in seller performance—might be particularly effective at steering, as they align buyer and platform interests in creating a good match. We report the results of an experiment in which a platform marketplace—an online labor market—guaranteed select sellers for treated buyers. The presence of a guarantee strongly steered buyers to these guaranteed sellers, but offering guarantees did not increase sales overall, suggesting financial risk was not determinative for the marginal buyer. This preference for guaranteed sellers was not the result of their lower financial risk, but rather because buyers viewed the platform's decision to guarantee as informative about relative seller quality. Indeed, a follow-up experiment showed that simply recommending the sellers that the platform would have guaranteed was equally effective at steering buyers.

---

[*]University of Minnesota

[†]Collage.com

[‡]NYU Stern. Author contact information and code are currently or will be available at http://www.john-joseph-horton.com/.

# 1 Introduction

Consumers are frequently uncertain about the quality of the product they are considering purchasing. One solution to this informational problem is a money-back guarantee. With a guarantee, the financial risk of unsatisfactory product performance is shifted to the seller (Heal, 1977) and a buyer might infer the product is of higher quality—if guaranteeing a high quality product is cheaper than guaranteeing a low quality product (Spence, 1977; Grossman, 1981). Both effects might make a purchase more likely. Perhaps unsurprisingly, these kinds of guarantees are common in conventional retail settings (McWilliams, 2012).[1] However, the rationale for a platform offering guarantees in a marketplace setting is less straightforward.

Consider a platform marketplace in which independent sellers offer goods or services. As in conventional settings, a platform money-back guarantee could inform buyers about the better sellers and reduce risk, thus making a sale more likely. It might also prevent negative consumer experiences from spilling over onto the platform as a whole (Nosko and Tadelis, 2015), either by facilitating better matches or placating a dissatisfied buyer. This spillover concern might be particularly important in online market settings where the platform has comparatively little control over which sellers participate and how they serve buyers. On the other hand, guarantees are costly to make, and a guarantee could encourage buyers to select "risky" or more expensive sellers, encouraging a kind moral hazard in their selection, or lead to hard-to-please buyers gravitating to the platform, creating a kind of adverse selection. Furthermore, although the platform does have a "bird's eye" view of the marketplace, it is not necessarily better-informed about what sellers would make a good match for a particular buyer.

---

[1]Money-back guarantees have been a central component to firm strategies at firms such as L.L. Bean, Publix, Collage.com, Trader Joe's, Stew Leonard's, Costco, Aldi, and Nordstrom. L.L. Bean's 100 percent satisfaction guarantee had been part of the firm's strategy since 1912, when the firm refunded money for a hunting boot whose poor design led to the boot's rubber bottom separating from the leather upper. However, L.L. Bean just recently limited the guarantee to just one year: `https://www.forbes.com/sites/marciaturner/2018/02/10/l-l-bean-announces-end-of-lifetime-replacement-policy-institutes-one-year-limit-on-returns/`

In this paper, we consider the effects of introducing a money-back guarantee in a platform marketplace for services. We report the results of two experiments that, together, allow us to explore (1) whether guarantees are "worth it" for the platform because of their effect on revenue, (2) whether guarantees are valued by buyers, and (3) *why* they are valued by buyers, with a particular focus on whether the steering effects of guarantees can also be obtained simply through "cheap talk" recommendations. In the first experiment, treated buyers saw that certain selected sellers were "guaranteed," meaning that if contracted-with, the platform would refund all expenses related to the contract incurred during the first two weeks of the contract, if the buyer was unsatisfied. In the second experiment, some buyers were randomized to see selected sellers as guaranteed (with the same terms as in the previous experiment), while other buyers saw those sellers as just "recommended," with no financial guarantee by the platform.

From the first experiment, we find that offering a guarantee did not increase platform revenue: the probability a contract was formed was essentially unchanged by the guarantee treatment overall, though there is perhaps some evidence it helped buyers with a high willingness to pay for quality. There is no evidence of greater expenditures in the contracts that were formed. However, treated buyers did strongly shift towards contracting with guaranteed sellers. We know this because we know which sellers applying to control buyers would have been guaranteed by the platform, had they applied to a treated buyer. Despite the possibility that treated buyers might select more expensive sellers, we find no evidence of this kind of moral hazard in selection.

The lack of an overall increase in contract-formation, as well as finding no evidence of a shift towards more expensive sellers, suggests that the direct financial implications of the guarantee were not a first-order consideration for the marginal potential buyer. However, buyers were interested in learning which sellers they *should* prefer, if they knew what the platform knew. But this finding does not imply that a same-sized shift in selection could be obtained simply by "cheap talk" recommendations: buyers might infer the platform would only be willing to guarantee those sellers unlikely to trigger

3

a refund, as those sellers are relatively cheap to guarantee. For this reason, the guarantee might still be needed for the buyer to find the platform (just as) credible.

In our second experiment, we find that the platform simply claiming a seller was "recommended"—with no guarantee—was just as effective at shifting buyers towards contracting with recommended/guaranteed sellers. It was not necessary for the platform to have a financial stake in the choice to influence selection. Despite a lack of effect on sales, a guarantee could still be worthwhile due to positive spillovers from fewer unsatisfied customers that received refunds. However, we have no strong evidence that treated buyers were any more satisfied—there was no detectable treatment effect on measures of buyer satisfaction or future business. Not unrelatedly, after the conclusion of the two experiments, the platform switched to only offering recommendations.

Our interpretation of the experimental results is that the platform had garnered sufficient trust that its cheap talk recommendations were valued. There are likely two reasons for this trust: (1) given the platform's "bird's eye view" of the marketplace and ability to collect information across the market, buyers might credibly believe that the platform has superior insights into relative seller quality, and (2) the platform already had incentives to create a good match.[2] Despite the platform's recommendations being valued, there is no evidence that overall sales increased, suggesting that whatever effect guarantees had on a buyer's anticipated match quality, it was not enough to induce the marginal non-contracting buyer to form a contract.

A main managerial implication of our findings is that guarantees were unnecessary for this marketplace, and recommendations sufficed to steer buyers. The power of these recommendations is substantial—we find that a seller with a recommendation would have to bid about 16% higher to be just as preferred as that same seller without a platform recommendation. Al-

---

[2]Although feedback provided by other users is widely-used by buyers and seems to generally be an accurate measures of latent quality (Gao et al., 2015), it also seems probable that the platform has information afforded by its ability to observe the entire marketplace. Furthermore, the platform is uniquely positioned to customize recommendations on a buyer-by-buyer basis.

though our results come from a specific empirical context, the basic features of our setting—buyers with imperfect information about the sellers that they can select from—is common in platform markets. Furthermore, our setting is one in which the economic fundamentals would, *a priori*, make guarantees seemingly attractive—unsatisfied buyers are not exceedingly rare, the stakes are high, and the service is inherently an experience "good"—and yet this conjecture was not confirmed by the evidence.

Our paper contributes to the growing literature on the management of online marketplaces (Hagiu, 2014; Eisenmann et al., 2006; Parker et al., 2016; Cusumano, 2010). Steering buyers towards select sellers is perhaps one of the most frequently practiced market interventions observed in practice, and our paper clarifies the strengths and limitations of this kind of intervention. The paper also offers some evidence in favor of the proposition that guarantees are valued more for their informational content rather than their risk-reducing effects (Grossman, 1981; Lutz, 1989; Bryant and Gerner, 1978; Garvin, 1983; Gerner and Bryant, 1981; Priest, 1981). An important caveat is that guarantees in our settings do not convey private information by the sellers, as they are offered by the platform, making them fundamentally different from a guarantee offered unilaterally by a seller.

The rest of the paper is organized as follows. Section 2 describes the empirical context and what economic conditions would tend to make guarantees effective absolutely, as well as relative to recommendations. Section 3 presents the first experiment and the results. Section 4 presents the second experiment and the results. Section 5 compares the results of two experiments. Section 6 concludes.

## 2 Empirical context

The empirical context for our study is a large online labor market. In these markets, firms contract with sellers to perform tasks that can be done remotely, such as computer programming, graphic design, data entry, and writing (Horton, 2010). The markets 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.

The online labor market that is our empirical context was founded in 2003, but did not have substantial transaction volume until 2006. As of 2018, there are about 200,000 job openings posted per month and 4 million active freelancers. As a basis of comparison, the platform has a market capitalization about 1/10 the size of eBay, though the difference in the nature of what is being bought and sold, the platform's take rate, and so on make this an imperfect measure of importance. The precise size and "real world" impact of the platform is less important than the fact that the platform essentially replicates all of the most important features of a labor market. See Agrawal et al. (2013a); Horton et al. (2017) for a more detailed description of market size and composition.

On the platform, would-be buyers write job descriptions, self-categorize the nature of the work and required skills, and then post the "requests for proposals" to the platform website. Sellers learn about requests for proposals via electronic searches or email notifications. Sellers 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, buyers can also search seller profiles and invite sellers to apply. After a seller submits an application, the buyer can interview and contract with the seller on the terms proposed by the seller or make a counteroffer, which the seller can counter, and so on. The process is not an auction and neither the buyer nor the seller are bound to accept an offer. Despite the possibility of bargaining, it is somewhat rare, and the selection process could be described as a kind of informal scoring auction.

To work on hourly contracts, sellers must install custom tracking software on their computers. The tracking software essentially serves as a digital punch clock: when working, the software records the count of keystrokes and mouse movements. The software also captures an image of the sellers' computer screen at random intervals. All of this captured data is sent to the platform's servers and then made available to the buyer for inspection, in real time. An upshot of this technology is that although it still takes some

time for buyers to learn if a match is "working out," they can learn much faster than a buyer who was limited to simply taking delivery of a completed project at some future date.

The basic design of the platform had been stable for several years at the time of the experiment, though there have been numerous improvements. One improvement was the introduction of recommendations to buyers in 2011. These recommendations were about what sellers the buyers might want to recruit for a posted opening, and not recommendations within the applicant pool (which are the kind of recommendations described in this paper). These "who to recruit" recommendations improved the fraction of jobs being filled (Horton, 2017a), though a substantial fraction of treated job openings went unfilled. The platform had a business hypothesis that the reason some openings were not filled was the financial risk in making a hire.

The value of a guarantee to buyers depends on how common disappointing buyer experiences are in the marketplace, and how large two weeks—the guarantee eligible period—is relative to the typical length of projects. On both dimensions, the data suggest that guarantees would be useful, as two weeks is a substantial period relative to the distribution of contract durations. Using historical data from the same platform that is the study for this experiment, Filippas et al. (2017) show that nearly 15% of buyers report a somewhat negative experience. As such, less-than-ideal experiences are commonplace. In terms of typical project duration, a two week guarantee would cover about 40% of projects in their entirety: when buyers post a request for proposals, they specify how long they expect the project to take, with answers ranging from "less than 1 week" to "more than 6 months." Slightly more than 30% choose less than 1 week and about 20% choose less than 1 month.

One interesting feature of this marketplace is that buyers are asked by the platform to state their "vertical" preference ("low," "medium" or "high"), which is their relative willingness to pay for quality (Horton and Johari, 2018). This feature is useful for our purposes, as buyers willing to pay higher prices for higher quality also potentially see more downside risk for a bad hire, as they will be paying higher wages. The fact that they have described

themselves as a "high tier" buyers also potentially indicates something about their risk tolerance. A high tier buyer might find a seller that the platform guarantees to be relatively more attractive than that same seller would be to a "low tier" buyer.

## 2.1 Online labor markets as a testing ground and objects of study

Online labor markets have become popular settings for research in both economics and information systems. In the economics literature, the focus has typically been on buyer preferences with respect to a number of different seller dimensions (Pallais, 2013; Chan and Wang, 2017; Stanton and Thomas, 2016; Agrawal et al., 2013b; Barach and Horton, 2017). Other papers have looked at changes in platform pricing policies (Horton, 2017b), the importance of recommendations on which sellers to recruit (Horton, 2017a), and how cross-country differences affect prices (Hong and Pavlou, 2015). The existence of a powerful third party influencing match-formation is quite new in most labor markets—the closest pre-online analogue was the labor market intermediary (Autor, 2008), which typically just verified match-relevant information (such as transcripts, certificates, criminal records, and so on).

In information systems, the literature has focused more concretely on the design and functioning of online marketplaces. Much of it focuses on the determinants of match formation as mediated by either bidding (as in the case of procurement auctions) or as mediated by marketplace reputations. For example, Snir and Hitt (2003) explore entry into the reverse auctions run by buyers and identify a market failure: excess bidding, as would-be sellers do not internalize the costs of bid evaluation. Yoganarasimhan (2013) studies IT firms bidding for projects and explores how the dynamic nature of job-filling could lead to erroneous inferences about seller reputations if analyzed as a static estimation problem. Allon et al. (2012) present a theoretical model of the platform's choice about facilitating communication among platform participants, and the effects their decision has on efficiency.

8

Goes and Lin (2012) examine the effects of a platform introducing paid certifications and, later, costly certifications, which is related to our focus here on recommendations and guarantees.

## 2.2 Nature of the good and the value of recommendations and/or guarantees

Guarantees or recommendations only potentially matter to a buyer when the good is an experience good—if the quality is known precisely *ex ante*, a guarantee is unnecessary. In online markets, arguably all goods are *de facto* experience goods because information about the good—and the quality of the seller—are imperfectly conveyed when buyers and sellers are not co-located. These informational gaps have certainly decreased over time in actual online markets—the history of online markets is a progression towards ever more goods and services being routinely sold online.[3] However, gaps clearly remain, and certainly exist in online labor markets where the "good" is a service delivered over time.

If all online good and services are *de facto* experience goods, guarantees or recommendations are potentially useful in a wide class of online markets. They are in fact used in some well-known online markets, such as eBay. And nearly all markets use recommendations. Even platform marketplaces that do not explicitly offer guarantees frequently offer *ex post* refunds for bad experiences (Cohen et al., 2018; Halperin et al., 2018), with the motivation that knowledge of future guarantees might induce future sales—similar to the intent of *ex ante* guarantees.[4]

What economic fundamentals of market make a guarantee more likely to affect a buyer's choice compared to "cheap talk" recommendations? Guaran-

---

[3]It is understandable that the initial focus of online retail was on relatively low cost goods, such as second-hand products on eBay. Amazon also famously began selling books in part because Jeff Bezos recognized that books are highly commodified. In contrast, customers now are willing to spend many thousands of dollars on accommodations on Airbnb. Even seemingly commodified goods in some markets are subject to extraneous factors affecting probability of sale (Doleac and Stein, 2013).

[4]In a market where sellers are directly matched by the platform, as in Uber and Lyft, using recommendations or guarantees to affect selection is unnecessary.

tees might steer buyers by (1) directly lowering the expected cost of buying from a guaranteed sellers through refunds and (2) being informative about the relative quality of a seller. In contrast, a recommendation only works through (2)—and only if the platform is credible. For example, recommendations might not "work" if the platform's recommendations are viewed as self-serving—say if buyers believed the platform made more money by pairing the buyer with particular sellers—such as those that had paid for prominence.[5]

In our empirical context, as the platform takes a percentage of the wage bill, it has incentives more closely aligned with sellers rather than buyers— all else equal, the platform prefers a larger wage bill. A platform that, say, imposed a lump sum fee on buyers would have an incentive, all else equal, to drive down wages to increase (extractable) buyer surplus. By taxing the wage bill, the platform might be perceived as having an incentive to steer buyers towards higher-wage sellers. However, to the extent that higher-wage sellers complete projects more quickly or employers economize on their demand for labor, it is unclear to what extent this would affect the total wage bill. It is ultimately an empirical question as to whether recommendations can be just as effective as guarantees.

It is useful to think of platforms as varying on two dimensions—(1) competence at identifying the right trading partner for a given seller and (2) disinterestedness, in the sense of the platform not being perceived as steering for self-interested reasons. If we imagine each dimension is binary, we have four possibilities:

1. Platform is competent and disinterested; both recommendations and guarantees would "work" at influencing buyers.

2. Platform is incompetent and disinterested; recommendations would not be valued and guarantees would only be valued for their financial

---

[5]This is a recurrent issue in the context of search ads and some social media—the platform benefits in the short-run if users mistake sponsored content for organic content, but at the risk of undermining trust in the long-run usefulness of the platform's organic content. The concern appears in other context as well, such as in the concerns about disc jockeys accepting side-payments, or "payola" to push certain songs (Coase, 1979).

component and not for their informational component.

3. Platform is competent and self-interested; recommendations are not valued or at least suspect, and guarantees are only valued for their financial component.

4. Platform is incompetent and self-interested; guarantees are only valued for their financial component.

In only one of the four possibilities enumerated—a competent, disinterested platform—do we expect guarantees and recommendations to both work. If both guarantees and recommendations work equally well, it would suggest the financial component of a guarantee is not worth very much.

Our plan of analysis is to first examine whether offering guarantees increases the overall probability that a contract is formed, and conditional upon a contract being formed, the attributes of that contract. We then see whether offering a guarantee for a particular seller changed the probability that that seller was contracted with. For this analysis, we switch to the level of the individual seller proposal. Our second experiment speaks directly to the question of the mechanism by which buyers are steered. Next, we examine results from the second experiment, again examining both overall effects on contract formation as well as whether the individual seller probability of being contracted with changed.

# 3 Experiment 1: Guaranteeing versus the *status quo*

In our first experiment, when buyers posted a request for proposals, they were randomized to either a treatment group in which guarantee-eligible sellers were marked as "guaranteed" or to a control group in which the guarantee-eligible sellers were not marked as special in any way. The control experience was the *status quo* on the platform prior to the experiment. For this first experiment, we first examine whether the treatment affected (1)

whether the buyer contracted with anyone at all and (2) whether it altered which specific seller the buyer contracted with.

In both the treatment and control groups, applicants were ordered in the applicant tracking system in the same manner. Buyers had the same tools for sorting applicants as they saw fit (e.g., by wage bid, time of application, experience, and so on). Figure 1a shows how a collection of seller proposals would look to a treated buyer, whereas Figure 1b shows how they would look to a buyer assigned to a control buyer.
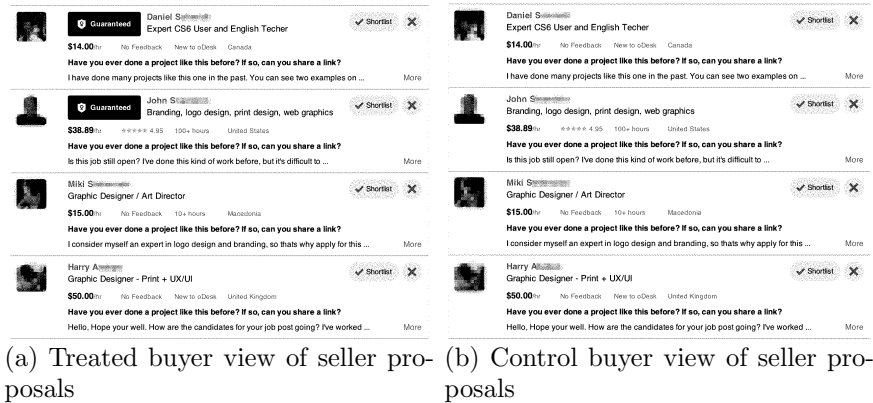
Note the guarantee "badge" for the first two proposals in the left panel of Figure 1a (we will discuss how the platform decided which sellers to guarantee in Section 3.2). Figure 1c shows more details on the badge, as well as the explanatory text shown with a mouse-over—it reads "Money-Back Guaranteed! If you are unhappy with this freelancer's first two weeks of work, [Platform] will refund your money." Although we do not know what fraction of buyers noticed the badge, as we will see, its presence strongly affected which seller the buyer contracted with. Furthermore, the plain language of the badge and its prominent display make it likely that many buyers appreciated the nature of the offer.

## 3.1 Sample definition and internal validity

The experiment began on 2013-11-15 and ended on 2014-08-05. A relatively small fraction of all buyer requests for proposals were assigned to the experiment (less than 5%) to mitigate financial risk to the platform. This small allocation also reduces concerns about validity-threatening movements of the market (Blake and Coey, 2014). After forming a contract with a guaranteed-seller, buyers had two weeks to request a refund. During the course of the experiment, the platform refunded approximately $110,000 to about 600 users resulting from the money-back guarantee. The minimum refund was 56 cents and the maximum was $2,700.

The experimental sample is composed of 36,264 requests for proposals, or "openings," which collectively received 1,051,778 proposals, or "applications," from 186,564 distinct sellers. Our primary sample is all the buyer

12

Figure 1: Comparison of buyer interface in the guarantee treatment group and control group, as well the details on the guarantee presented in the interface



(a) Treated buyer view of seller proposals

(b) Control buyer view of seller proposals



(c) Close-up view of the guarantee badge with explanatory text

*Notes:* This figure shows samples of the interfaces presented to buyers. The top left panel shows the interface presented to a buyer in the treated group, whereas the right panel shows the interface for the same buyer had they been assigned to the control group. The bottom panel shows the zoomed-in view that a treated buyer would see if they hovered their mouse pointer over the "guaranteed" badge. The actual name of the platform has been replaced with the word "Platform."

openings, or requests for proposals, which is also the unit of randomization.[6] To be included in the experiment, a request for proposals had to be "public" in the sense that any seller could apply to it. The buyer also had to specify

---

[6]Buyers can and do post multiple requests for proposals. These subsequent requests received the same allocation as the original request (to prevent buyers from seeking out their preferred cell). However, we restrict the sample to the first request for proposals by a buyer following their allocation to the experiment, as subsequent observations could be influenced by the treatment assignment.

an hourly contract structure, where the sellers bid a wage. So-called "fixed price" (i.e., not hourly) contracts were not eligible for the experiment and no guarantee was offered in these cases. The buyer also had to receive at least one proposal from a seller that would be eligible for a guarantee, had they been assigned to the guarantee treatment. All of these restrictions left 85.7% of all the requests for proposals that otherwise met the criteria for inclusion. The length of the experiment was determined by an ex ante power calculation conducted by the platform.[7] As expected, the sample is well-balanced with respect to buyer and seller characteristics. See Appendix B for balance tables.

During the experiment, as noted above, guarantees were offered to a relatively small number of buyers, and sellers were not made aware of the intervention. As such, sellers did not alter their bids in response to the intervention. However, in an equilibrium in which guarantees were offered, we might expect sellers to "claw back" some of this implicit subsidy through higher wage bids. This claw back matters, as it would somewhat offset the benefit of a guarantee to a buyer. However, looking at whether buyers were more likely to contract when gaining the "full" subsidy of the guarantee is a good empirical starting point: if treated buyers could not be induced to transact in the experiment, they would be even less likely to do so in equilibrium.

## 3.2  The platform's selection of sellers to guarantee

The platform had to decide, in real-time, whether to guarantee a seller. Given the financial cost associated with a guarantee, a profit-maximizing platform should, in theory, prefer those sellers with the highest probability of being able to complete a project successfully. The platform more or less followed this logic: it assigned each proposal a score and then used a score cut off for the binary guaranteeing decision (if the proposal was

---

[7]The intent was to have an experiment large enough to have sufficient power to detect a 5 percentage point change in the probability a contract was formed, at 90% power. The experiment ran longer than required for this level of power, as making a quick business decision was not essential; the "realized" power was vanishingly close to 100% for a 5 percentage point effect.

submitted to a treated buyer). This score was generated by a predictive model trained on historical platform data. The inputs to the model included the seller characteristics, such as their experience on the platform, their appropriateness for the project given the skills required by the project and the seller's skills, and also their wage bid. A higher wage bid resulted in a lower score, at least on average.

The score is normalized to always fall within $[0, 1]$. Those sellers with a score above a certain threshold, 0.5, were eligible to be marked as guaranteed if they applied to a treated buyer. This score was also computed if the seller applied to a control buyer, though in this case, guarantee-eligible sellers were not marked in any way. A small number of applicant sellers with scores below the threshold were also eligible for the guarantee based on a separate model that attempted to predict promising new entrants who otherwise would not have been guaranteed.

Table 1 compares the characteristics of sellers who were eligible for the guarantee to those who were not, reporting the level differences and the difference in percentage terms. We can see that guarantee-eligible sellers had substantially more experience. Unsurprisingly, they also charged more for their services compared to non-eligible sellers. As we will see, the score was highly correlated with the applicant being contracted with, even in the control group where the badge was not observed by the buyer. This suggests that despite higher bids, these sellers were still perceived as offering more surplus.

The treatment assignment associated with a buyer was not observable by sellers when they applied, and so we would expect seller applicant pools to be balanced with respect to seller characteristics. To compare pools, in Table 2 we compare the mean attributes of sellers with a score above 0.5, by the treatment assignment of the applied-to buyer. As expected, the table shows that there was no appreciable difference in the two groups. Two of the seller characteristic differences are marginally significant at the 10% level, though given the number of characteristics examined, having this many marginally significant differences (or more) would be expected about 15% of the time,

Table 1: Mean characteristics of sellers, by guarantee eligibility

|  | Mean (SCORE ≤ 0.5) | Mean (SCORE > 0.5) | Difference in Means | % Δ | |
|---|---|---|---|---|---|
| *Seller Attributes* | | | | | |
| Hours Worked to Date | 609.13 (3.68) | 984.71 (6.13) | 375.58 (5.86) | 61.7 | ** |
| Num Past Jobs Worked | 11.98 (0.05) | 26.10 (0.15) | 14.12 (0.14) | 117.9 | ** |
| Past Hourly Earnings | 5,271.98 (38.65) | 11,634.06 (83.83) | 6,362.08 (76.75) | 120.7 | ** |
| Num Prior Relationships | 9.80 (0.04) | 19.84 (0.11) | 10.04 (0.10) | 102.4 | ** |
| Wage Bid $/hour | 9.89 (0.05) | 14.41 (0.08) | 4.52 (0.07) | 45.6 | ** |
| Profile Wage $/hour | 9.81 (0.05) | 14.03 (0.07) | 4.23 (0.06) | 43.1 | ** |

*Notes:* This table reports means and standard errors for a number of seller characteristics at the time of application, by whether those sellers were eligible for a guarantee. Sellers with a score greater than 0.5 were guaranteed if they applied to a treated buyer. A small fraction of sellers with scores below 0.5 were also guarantee-eligible, on the basis of a separate predictive model trying to identify promising sellers new to the market. Standard errors are clustered at the level of the buyer. Significance indicators: $p \leq 0.10 : \dagger$, $p \leq 0.05 : *$ and $p \leq .01 : **$

even if all attributes were independent of each other.[8]

## 3.3 Effects of offering guarantees on whether the buyer contracted with any seller

Our first outcome of interest is whether the buyer contracted with any of the applying sellers. In Table 3, Column (1), we report an ordinary least squares estimate of

$$\text{CONTRACTED}_j = \beta_0 + \beta_1 \text{MBG}_j + \epsilon, \tag{1}$$

where $\text{CONTRACTED}_j$ is an indicator for whether the buyer $j$ spent some amount of money on one or more contracted-with sellers and $\text{MBG}_j$ is an indicator for whether buyer $j$ was assigned to the guarantee treatment

---

[8]This figure is calculated with 1,000,000 simulations under the null of a uniformly distributed p-value. This is a conservative estimate, as we would expect a higher fraction if these measures are correlated, which they are.

Table 2: Mean characteristics of guarantee-eligible applicants by the treatment assignment of the applied-to buyer

|  | Control Mean | Treatment Mean | Difference In Means | p-value | |
|---|---|---|---|---|---|
| *Seller Attributes* | | | | | |
| Hours Worked to Date | 979.11 (8.53) | 990.43 (8.81) | 11.32 (12.26) | 0.36 | |
| Num Past Jobs Worked | 26.09 (0.21) | 26.12 (0.21) | 0.03 (0.30) | 0.91 | |
| Past Hourly Earnings | 11,488.29 (115.27) | 11,783.07 (121.84) | 294.78 (167.72) | 0.08 | † |
| Num Prior Relationships | 19.81 (0.15) | 19.87 (0.15) | 0.05 (0.22) | 0.81 | |
| Wage Bid $/hour | 14.30 (0.11) | 14.52 (0.11) | 0.22 (0.16) | 0.16 | |
| Profile Wage $/hour | 13.91 (0.10) | 14.16 (0.10) | 0.25 (0.14) | 0.07 | † |

*Notes:* This table reports means and standard errors across experimental groups for characteristics of applicants who were eligible for the money-back guarantee. Standard errors are clustered at the buyer level. Significance indicators: $p \leq 0.10 : †$, $p \leq 0.05 : *$ and $p \leq .01 : **$.

group.[9] We can see that $\hat{\beta}_1$, the treatment effect, is close to zero and far from conventionally significant. It is also a precise estimate—the 95% CI for the effect is $[-0.0039, 0.0164]$.

As different buyers might value the guarantee differently, in Column (2), we interact the treatment indicator with the buyer's vertical preference indicator. We can see that among high-tier buyers, the treatment increases the probability of contracting by about 11% from the base contracting rate for these high-tier buyers. Of course, by analyzing interaction effects, we increase the probability of falsely rejecting the null. However, a likelihood ratio test comparing the Column (2) specification to one that includes the vertical preference tier but not the interaction terms has a p-value of 0.012.

Assuming the high tier "effect" is not due to sampling variation, one possible interpretation is that the sellers that high-tier buyers are interested in tend to be higher wage and hence higher risk. For these buyers, a money back guarantee is simply worth more to them. Another possibil-

---

[9]Although our outcome is binary, because we have a true experiment with a binary treatment, the marginal effects from the logit or probit are identical to those obtained from the ordinary least squares.

Table 3: Effect of offering a gurantee on one or more applying sellers on the probability the buyer forms a contract with any seller

| | Dependent variable: | | |
|---|---|---|---|
| | Contract formed | | Log wage bill |
| | (1) | (2) | (3) |
| Money-back guarantee offered on select sellers, MBG | 0.006 | −0.007 | −0.003 |
| | (0.005) | (0.010) | (0.062) |
| MedTier | | −0.022* | 0.651** |
| | | (0.009) | (0.054) |
| HighTier | | −0.090** | 1.146** |
| | | (0.010) | (0.069) |
| MBG × MedTier | | 0.008 | 0.018 |
| | | (0.012) | (0.077) |
| MBG × HighTier | | 0.041** | 0.008 |
| | | (0.015) | (0.097) |
| Constant | 0.362** | 0.394** | 4.554** |
| | (0.004) | (0.007) | (0.044) |
| Sample | All | All | Filled |
| Observations | 36,264 | 36,264 | 13,249 |

*Notes:* The table reports regressions where the dependent variable is an indicator for whether the buyer formed a contract i.e., at least one seller was paid some amount of money. The estimation method is OLS. The sample consists of the requests for proposals by all buyers allocated to the first experiment where at least one applicant was guarantee-eligible. If the buyer contracted with a guaranteed seller, the platform would reimburse the buyer for the first two weeks of any contract if requested. In Column (2), the treatment indicator is interacted with the buyer's vertical preference tier i.e., whether the buyer stated they were interested in hiring low experience, low price sellers (the omitted category), high experience, high price sellers, HighTier $= 1$ or somewhere in between, MedTier $= 1$. These buyer selections are made ex ante, before randomization. Significance indicators: $p \leq 0.10 : \dagger$, $p \leq 0.05 : *$ and $p \leq .01 : **$.

ity is that high-tier buyers require greater expertise, which the buyer might have difficulty assessing. As such, the buyer might be more interested in the platform's judgment about the "best" applicant. This would be consistent with Horton (2017a), which found—using an online labor market—that employers with "technical" job openings requiring greater skill that received recruiting recommendations were far more likely to form a contract. We will return to this issue in our follow-on experiment.

Although offering guarantees had no overall effect on the probability that a contract was formed, as we will see, the guarantee strongly affected which seller the buyer selected. Because of this composition effect, the total billings might be affected, say because buyers now tend to select more expensive sellers. This in turn could affect platform profit. We test this hypothesis in Column (3), in which the outcome is the log total wage bill, conditional upon a contract being formed. We can see that although the total wage bill is increasing in the vertical preference of the buyer, there is no evidence that the guarantee raised the wage bill (the coefficient on MBG) in the low tier, nor that it raised it in the other tiers (the near-zeros on the interaction term coefficients).

## 3.4 Effects of offering the guarantee on selection

Despite having no overall effect on the probability a contract was formed— except perhaps in the high-tier group—offering guarantees could have altered *which* sellers were contracted with. This change in preferences could be due to the financial effects or the informational effects of a guarantee. For now, we set aside this question of why guarantees were valued, and simply explore the buyer selection among seller (we will return to the "why" question in our second experiment).

To illustrate the effects of a guarantee on selection, in Figure 2, we plot the application mean success rate by score "band," by the treatment status of the applied-to buyer, pooling over all requests for proposals. We can see that for low score bands, the treatment and control have similar success rates, though the control is everywhere above the guarantee. As we near

Figure 2: Mean application contracting probability by seller score and the treatment assignment of the applied-to seller
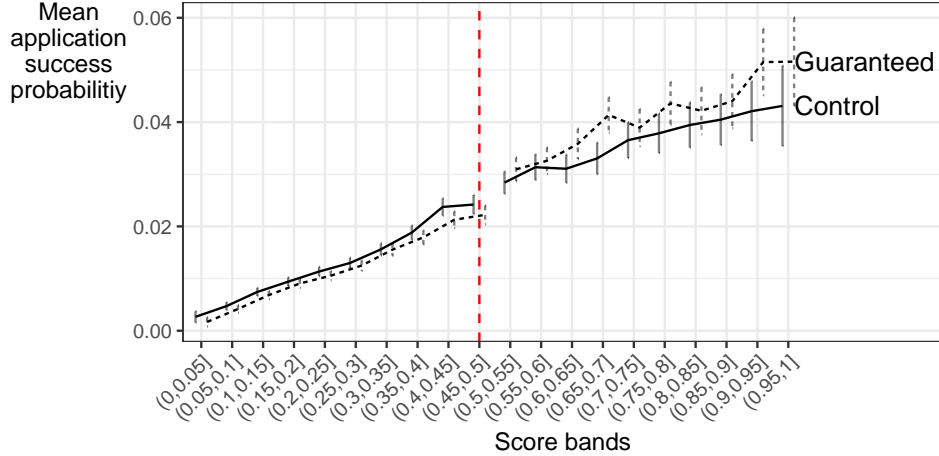
*Notes:* This figure shows the mean seller success rate by score "band," by the treatment status of the applied-to buyer. The threshold is indicated by a dashed vertical line. For each band, a 95% CI is shown around the mean.

the threshold (but are still below it), sellers applying to buyers with the guarantee treatment do slightly worse, suggesting crowd-out effects. Above the threshold, we can see that sellers applying to treated buyers do better. For example, in the $[0.95, 1]$ band, the effect in levels is about 0.01, which is about 20% higher.

As a more direct way to examine how guarantees affected buyer selection, we can exploit the fact that sellers apply to multiple buyers, and so we have within-seller variation in whether or not they receive a guarantee for a particular application. With this application data, we can switch our analysis from the level of the buyer to the level of the individual seller applicant. As sellers cannot condition on the treatment assignment of the buyer, guarantee eligibility can be treated as exogenous when a seller-specific fixed effect is included. We estimate

$$
\begin{aligned}
\text{CONTRACTED}_{ij} = \alpha_i &+ \beta_1 \cdot \mathbf{1}\{\text{SCORE}_{ij} > 0.5\} + \beta_2 \text{MBG}_j \\
&+ \beta_3 \left(\text{MBG}_j \times \mathbf{1}\{\text{SCORE}_{ij} > 0.5\}\right) + \epsilon_i, \qquad (2)
\end{aligned}
$$

where $\textsc{Contracted}_{ij}$ is an indicator for whether seller $i$ was contracted with by buyer $j$, $\alpha_i$ is a seller-specific fixed effect, $\mathbf{1}\{\textsc{Score}_{ij} > 0.5\}$ is an indicator whether the applying seller has a score higher than the cut-off to receive the guarantee, and $\textsc{MBG}_j$ is an indicator for whether the applied-to opening was assigned to the treatment (and hence the applicant would receive the guarantee if their score was above the 0.5 threshold). We cluster standard errors at the level of the individual seller.

Figure 3 illustrates the effects of the score and guarantee the probability a seller is selected, by showing predictions from an estimation of Equation 2. The x-axis is treatment status of the applied-to buyer. The y-axis is the predicted change in contracting probability. The predictions are split between those sellers above and below the threshold. We also plot 30 lines connecting point estimates from a bootstrap, sampling at the level of the individual buyer, with replacement.

The first panel from the left plots the predictions for all observations in the sample. In the control, we can see that when a seller with a score above the threshold applies to a buyer, he or she is substantially more likely to be contracted with—the point estimate is about 0.005. However, being above the threshold makes a much larger difference in contracting probability when the seller is applying to a guarantee-treated buyer—the contracting probability is now nearly 0.009. For sellers below the threshold, applying to a guaranteed buyer perhaps somewhat lowers contracting probability through a crowd-out effect.

The three panels labeled "Low" "Medium" and "High" illustrate effects by the buyer's vertical preference tier. Looking across tiers, we can see that as the vertical preference increases, contracting probability declines. This is expected, assuming bids do not change much: keeping the identity of the seller fixed, the greater the vertical preference of the buyer, the less likely the seller exceeds the buyer's reservation value. For sellers above the score cut-off, the contracting probability is much higher even when applying to a control buyer. However, as we saw with the "Pooled" panel, being above the threshold helps much more when applying to a treated buyer. The treatment effect—the slope of the line in the above group—appears to be about the

21

Figure 3: Application-level estimates of effects of being above the threshold on being contracted with, by experimental group and tier
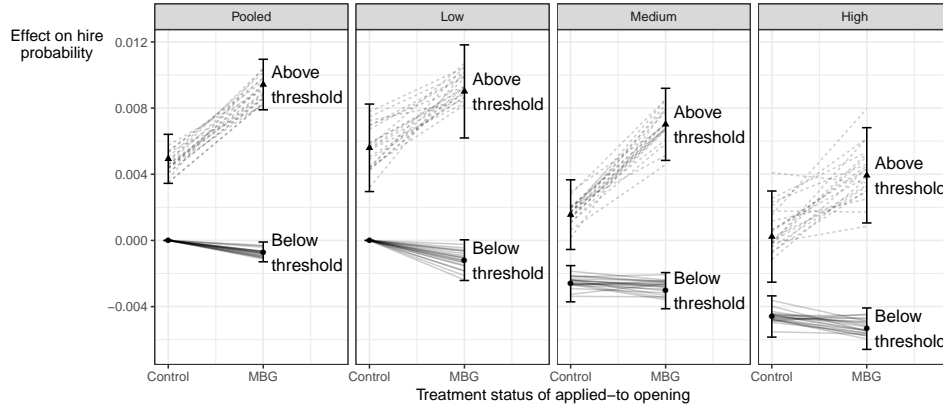


*Notes:* This figure plots marginal effects based on Equation 2. The gray lines indicate block-bootstrap estimates of the effects, sampling freelancers with replacement.

same, regardless of the tier. If anything, it is slightly less steep in the high tier.

In our experiment, the treatment could lead to better sellers getting hired and "taken off the market." This could lead to fewer good sellers being available for buyers later in the experiment, causing a decline in average quality. However, this would affect both treatment and control buyers. Despite this possibility, we find no evidence of an overall decline over time, nor of any difference in applicant quality between treatment and control. See Appendix C for this analysis.

### 3.5 Moral hazard in selection

A concern for the platform in offering a guarantee is that buyers might show less care in selecting a seller, exhibiting a kind of moral hazard. In particular, they might be less price sensitive or careful when selecting among guaranteed sellers. To test for this behavior, we estimate a regression

$$\text{CONTRACTED}_{ij} = \alpha_i + \gamma_j + \beta_1 \log w_{ij} + \beta_2 \left( \log w_{ij} \times \text{MBG}_j \right) + \epsilon, \quad (3)$$

22

where $w_{ij}$ is seller $i$'s wage bid to buyer $j$, and $\alpha_i$ and $\gamma_j$ are seller and buyer fixed effects, respectively. We expect that with a higher wage bid, all else equal, the buyer is less likely to make a hire. For this reason, we expect $\hat{\beta}_1 < 0$. However, if buyers become less price sensitive because of the guarantee—with some probability, the platform will pay that charge—we would expect that $\hat{\beta}_2 > 0$.

Column (1) of Table 4 reports a regression similar to Equation 3, but with the interaction term omitted. The sample is restricted to only applications where the seller had a score above the guarantee threshold. As expected, the higher the wage bid, the less likely the buyer is to hire that seller. As we include a seller-specific fixed effect, the coefficient on the hourly wage bid lets us price how much a recommendation is worth in terms of hire probability—the implied effect is that a seller with a recommendation would have to bid about 16% higher to be just as preferred to that same seller without a platform recommendation (given the demand elasticity estimate).

In Column (2), we add the interaction term to the Column (1) regression. Contrary to our moral hazard conjecture, the coefficient, $\hat{\beta}_2$ is close to zero (and the "wrong" sign), Furthermore, this a precisely estimated zero, with a 95% CI of $[-0.007, 0.0011]$. In short, there is no evidence of moral hazard in selection, at least as measured by the price sensitivity of buyers.

To illustrate the importance of the within-seller approach to identifying the effects of price on selection, in Column (3), we use the full sample and remove the seller and buyer fixed effects. Now, the coefficient on the wage bid is positive, suggesting a seller with a higher wage bid is *more* likely to be contracted with, which is clearly not a causal effect.

## 3.6 Project outcomes

The primary goal of the experiment was to increase buyer willingness to form a contract. As the platform guaranteed "better" (albeit more expensive) sellers, there was little concern that the experiment would lead to worse contractual outcomes. If anything, given that treatment altered which sellers were contracted with (for the better), we might anticipate some positive

Table 4: Effect of seller wage bid on the probability of the buyer contracting with that seller, by treatment assignment

| | *Dependent variable:* | | |
| | Contract formed? | | |
| | (1) | (2) | (3) |
|---|---|---|---|
| Log wage bid | −0.031*** | −0.029*** | 0.002*** |
| | (0.002) | (0.003) | (0.0002) |
| Log wage bid × MBG | | −0.003 | |
| | | (0.002) | |
| Constant | | | 0.015*** |
| | | | (0.0004) |
| Seller SCORE | > 0.5 | > 0.5 | Any |
| Buyer FE | Y | Y | N |
| Seller FE | Y | Y | N |
| Observations | 232,972 | 232,972 | 1,029,366 |

*Notes:* This table reports application-level regressions where the dependent variable is an indicator for whether the buyer contracted with the applying seller. The key independent variable is the proposed hourly wage of the applying seller. Note that in Columns (1) and (2), both seller-specific and buyer-specific fixed effects are included. In Column (3), the sample is all contracts and no fixed effects are included. Standard errors are clustered at the level of the individual seller. Significance indicators: $p \leq 0.10 : \dagger$, $p \leq 0.05 : *$ and $p \leq .01 : **$.

changes in project outcomes. In Table 5, we report regressions where the outcomes are the "public" feedback a seller received, the "private" feedback (which is often more candid—see Filippas et al. (2017)), and finally, a revealed preference measure, which is whether the buyer re-contracted with the seller at a later date. In all three regressions, the sample is restricted to the requests for proposals in which (1) a contract was formed, and (2) the associated outcome measure is available. As not all buyers leave feedback (and not all leave both kinds), the sample sizes are different across regressions.

Table 5: Effects of guaranteeing sellers on buyer contract outcome measures

|  | *Dependent variable:* | | |
|  | Public Feedback | Private Feedback | Rehired? |
|  | (1) | (2) | (3) |
| MBG | 0.002 | 0.036 | 0.003 |
|  | (0.016) | (0.049) | (0.005) |
| Constant | 4.727*** | 8.876*** | 0.140*** |
|  | (0.012) | (0.036) | (0.003) |
| Observations | 13,011 | 13,990 | 21,794 |

*Notes:* The sample is restricted to buyers who formed contracts with a seller and who had at least one guarantee-eligible seller in their application pool. If the buyer contracted with a guaranteed seller, the platform would reimburse the buyer for the first two weeks of any contract. The estimation method is OLS. Significance indicators: $p \leq 0.10 : \dagger$, $p \leq 0.05 : *$ and $p \leq .01 : **$.

For the two feedback measures—Columns (1) and (2)—the effect of being offered the guarantee appears to be positive, albeit close to zero and far from significant. At least with respect to feedback, there is no evidence that the matches formed in the treatment group were different than the control. This lack of effects is perhaps not too surprising, given that most feedback outcomes are conditioned on the price paid—even if the treatment caused buyers to hire better sellers, they are also more expensive.

Another measure of buyer satisfaction is whether they contracted with the same seller again in the future. In Column (3), we can see that there was a small increase in the probability a buyer in the treatment contracted with

the same seller again in the future—in percentage terms, this is a 0.46% increase, but the standard error for that measure is 4%, making it quite likely that the increase was due to sampling variation. Presumably those buyers that received refunds from the platform were more satisfied than they would have been in the control, but this evidently did not show up in the overall satisfaction metrics available to us.

# 4    Experiment 2: Guaranteeing versus Recommending Experiment

Although the first experiment showed no evidence that the quantity of transactions appreciably increased with a guarantee, it still could be useful to shift buyers towards platform-preferred sellers. Steering could be useful both for Nosko and Tadelis (2015) platform spillover reasons, as well overcoming inefficient over-reliance on experienced sellers (Pallais, 2013; Horton, Forthcoming). But the natural question is whether steering could be done as effectively without money. Following the first experiment, a second experiment was conducted to answer this question.

The second experiment began on 2014-07-09 and ended on 2014-08-26. There is some temporal overlap in the two experiments, but each experiment has its own independent treatment and control groups.[10] Buyers were randomized into two groups: those who saw guarantee-eligible sellers as guaranteed (with the same terms as the previous guarantee), and those who saw them as "recommended," with no further information about what this recommendation meant. Figure 4 shows the messages seen by buyers in the "guarantee" and "recommends" groups, respectively. In terms of how would-be buyers interpreted a recommendation, given that applicants are already self-selected with respect to the nature of the job opening, it is unlikely they viewed it as "horizontal" and about the buyer's idiosyncratic taste, but rather viewed the recommendation as being about the seller's "vertical" attributes relative to the price being proposed.

---

[10]The reason for the overlap is that the first experiment continued to run for a few weeks after the initial analysis of the results, which motivated the second experiment.

The rules for determining an applicant's eligibility for being guaranteed remained the same. The empirical context for the second experiment was largely the same, except for one important difference: in the initial experiment, buyers in the treatment group saw both guaranteed and non-guaranteed sellers in the default view of the applicant pool, whereas in the second experiment, buyers only saw guarantee-eligible sellers in the initial view of the user interface, but could view all sellers by selecting a different "view" in the interface.[11] As we will show later, this change in interface design reduced the probability of sellers below the threshold in the second experiment getting hired, but the effects were fairly modest.

Figure 4: Comparison of the messaging presented about sellers in the (a) "guarantee" and (b) "recommend" experimental groups



(a) "Guaranteed" messaging          (b) "Recommends" messaging
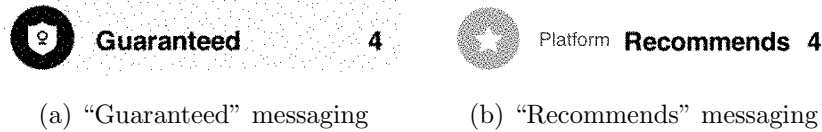
*Notes:* This figure shows the two "badges" used in the second experiment. The left panel shows the guaranteed messaging, whereas the right panel shows the recommended messaging (with the actual name of the platform removed). The criteria for this badge were the same as in the first experiment. The "4" next to the badge is the number of candidates in this particular applicant pool that were being recommended or guaranteed.

## 4.1    Sample definition and internal validity

As with the first experiment, the sample is restricted to the first request for proposals by a buyer following their allocation to the experiment. The sample consists of a total of 14,232 requests for proposals, which collectively received 427,516 proposals from 90,818 distinct sellers. As expected, pre-randomization attributes are well balanced, as are the characteristics of applicants. See Appendix B for this analysis.

---

[11]The decision to add this change was not ours, but reflected a business decision to try to more rapidly shift buyers to using higher-score sellers.

## 4.2 Effects of offering guarantees on whether the buyer contracted with anyone

As with the first experiment, we examine whether there was a difference in the probability that a contract was formed, by experimental group. In Table 6, the dependent variable is an indicator for whether the buyer formed a contract. The model is fit using ordinary least squares, with the single independent variable being the treatment indicator. The omitted category is the recommended group i.e., MBG = 0.

Starting in Column (1), we can see that there is no evidence of a large difference in the probability a contract was formed. In Column (2), we test whether guarantee versus recommendation worked differently for different kinds of buyers. Recall from Table 3 that buyers in the high-tier seemed to be more likely to form a contract when offered guaranteed sellers. In the second experiment, we find no appreciable difference between guarantees and recommendations, suggesting that any "lift" in the high tier in the first experiment was due the information effects of the recommendation rather than the financial effects (assuming not sampling variation).

## 4.3 Effects of offering the guarantee on buyer contracting, relative to recommending

Now we turn to the effects of the treatment on which sellers a buyer contracts with. We are primarily interested in whether there was any difference in effectiveness between offering a guarantee for a seller and simply recommending that seller. As we did for the first experiment, in Figure 5, we can simply plot the mean application success probability by score band, by treatment assignment. Figure 5 shows that, as before, sellers with higher scores are more likely to be selected. However, in contrast to the first experiment, there is no evidence that guaranteed sellers are more likely to be contracted with relative to those sellers in the same score band but who applied to "recommended"-only buyers. If anything, the "recommended" contracting probability seems to be above the guarantee in the above-threshold values of the score. We can also, by comparing Figure 5 and Figure 2 that those

Table 6: Effect of offering a guarantee versus simply recommending an applicant on the probability a buyer forms a contract

| | Dependent variable: | |
| --- | --- | --- |
| | Contract formed | |
| | (1) | (2) |
| Money-back guarantee offered on select sellers, MBG | −0.005 | −0.006 |
| | (0.009) | (0.016) |
| MEDTIER | | −0.021 |
| | | (0.014) |
| HIGHTIER | | −0.044*** |
| | | (0.017) |
| MBG × MEDTIER | | 0.017 |
| | | (0.020) |
| MBG × HIGHTIER | | −0.030 |
| | | (0.024) |
| Constant | 0.395*** | 0.415*** |
| | (0.006) | (0.011) |
| Observations | 12,929 | 12,929 |

*Notes:* The table reports regressions where the dependent variable is an indicator for whether the job opening was filled i.e., at least one seller was contracted with and paid some amount of money. The estimation method is OLS. The sample consists of all job openings allocated to the experiment where at least one applicant could have been recommended, i.e., had a SCORE > 0.05. The key dependent variable is whether the job opening was assigned to the treatment, in which case those sellers with a score above the threshold were guaranteed by the platform, or in the control, where they were not. If the buyer contracted with a guaranteed seller, the platform would reimburse the buyer for the first two weeks of any contract. In Column (2), the treatment indicator is interacted with the buyer's vertical preference tier i.e., whether they are interested in hiring low experience, low price sellers (the omitted category), high experience, high price sellers, HIGHTIER = 1 or somewhere in between, MEDTIER = 1. These buyer selections are made ex ante, before randomization. Significance indicators: $p \leq 0.10 : \dagger$, $p \leq 0.05 : *$ and $p \leq .01 : **$.

Figure 5: Mean application success probability by applicant score and the treatment assignment of the applied-to seller, from the second experiment
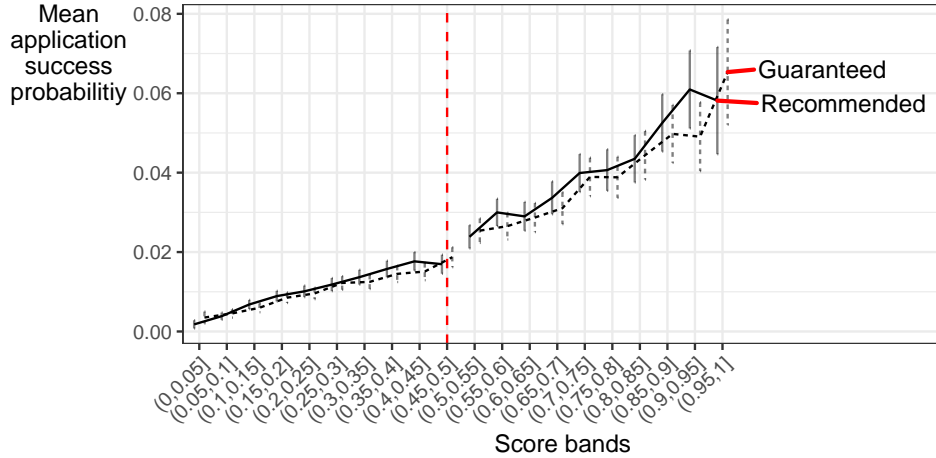


*Notes:* This figure shows the mean seller success rate by score "band," by the treatment status of the applied-to buyer. The threshold is indicated by a dashed vertical line. For each band, a 95% CI is shown around the mean.

sellers below the threshold in the second experiment had an overall lower hire probability, as we would expect given the interface change (though the effects are fairly modest).

Moving to a regression framework, in Table 7, Column (1), we report an application-level regression where the outcome is whether the applicant was contracted with. This regression is identical to the one based on Equation 2, except for the interpretation: now an applicant with a score above the threshold was marked as "recommended" whereas in the first experiment, in the control group, they had no special indicator. We can see that having a score above the threshold but not being guaranteed, $\mathbf{1}\{\text{SCORE} > 0.5\}$, increases hire-probability by 0.013. In contrast, in the first experiment, this effect was just 0.005. Presumably this more than doubling in probability is due to both the change in the interface to only show high score applicants by default and by the labeling of those candidates as "recommended."

The most important result from Column (1) is the precisely estimated near-zero coefficient on the $\mathbf{1}\{\text{SCORE} > 0.5\} \times \text{MBG}$ interaction term. This

Table 7: Effects of the platform guaranteeing versus recommending on contracting probability and the characteristics of contracted sellers

|  | *Dependent variable:* | | |
|---|---|---|---|
|  | Contracted? | **1**{Score > 0.5} | Score |
|  | (1) | (2) | (3) |
| **1**{Score > 0.5} | 0.013*** | | |
|  | (0.001) | | |
| MBG of the applied-to opening | −0.001 | 0.00004 | 0.001 |
|  | (0.0005) | (0.015) | (0.007) |
| **1**{Score > 0.5} ×MBG | −0.0005 | | |
|  | (0.002) | | |
| Constant | | 0.586*** | 0.542*** |
|  | | (0.010) | (0.005) |
| Sample | All | Contracted only | Contracted only |
| Seller FE | Y | N | N |
| Observations | 417,695 | 7,188 | 7,188 |

*Notes:* The table reports regressions where the unit of analysis is the proposal sent by a seller. In Column (1), the dependent variable is an indicator for whether that particular seller applicant was contracted with. The sample consists of all proposals to all buyers assigned to the experiment. **1**{Score > 0.5} is an indicator for whether the applying seller had a platform-provided quality score greater than 0.5, hence making them eligible to be either guaranteed or recommended, depending on which buyer they applied to. If the buyer hired a guaranteed seller in the treatment, the platform would reimburse the buyer for the first two weeks of any contract if the buyer requested such a refund. In Columns (2), the outcome is an indicator for the contracted with seller exceeding the 0.5 threshold. In Column (3), the outcome is the score of the contracted with seller, the sample restricted to only contracted with sellers. Standard errors are clustered at the level of the individual seller in Column (1). Standard errors are clustered at the level of buyer for Columns (2) and (3), as buyers can make more than one hire. Significance indicators: $p \leq 0.10 : \dagger$, $p \leq 0.05 : *$ and $p \leq .01 : **$.

means that from a seller's perspective, a recommendation and a guarantee have essentially the same effect on contracting probability. Similarly, the precisely estimated near-zero on the MBG indicator implies that the crowd-out effects—whatever they were—were no different in the two experimental cells for below-threshold applicants.

In Column (2), we restrict the sample to only those sellers who were contracted with, and use whether their score was above the threshold as the outcome. In Column (3), we use the same sample, but then use the score itself as the outcome. There is no evidence that in the treatment, the average score of the hired seller increased. This result stands in sharp contrast with the first experiment, where it clearly favored above threshold sellers.

# 5    Comparing experiments

The analysis of Experiment 2 suggests that recommendations worked just as well guarantees at steering buyers towards select sellers. In this section, we directly compare the two experiments, testing whether the guarantee worked better for some types of buyers. We also examine directly whether buyers of different types were more or less likely to act on the platform's recommendations/guarantees by using as an outcome an indicator for contracting with an above-threshold seller.

## 5.1    The effects of offering the guarantee on buyer selection, by experiment

Figure 6 illustrates the effects of the score and guarantee on individual seller selection, by showing predicted effects from an estimate of Equation 2 for both experiments and by buyer vertical preference type.[12] In the top panel, the observations are pooled across all openings. In the bottom three panels effects are shown for low, medium, and high type buyers, respectively. The

---

[12]As our predictors are all binary indicators, the marginal effects are simply the coefficients appropriately added up.

x-axis is the treatment assignment of the buyer. For each set of estimates, 30 bootstrap estimates are plotted.

We can see from the figure that in Experiment 1, being above the threshold and applying to an MBG-treated buyer helped in forming a contract. This just recapitulates what we saw in Figure 6. In Experiment 2, in the right column, we can see that although being above the threshold helped—typically more so than in Experiment 1—there is no evidence of a treatment effect when moving from recommended to guaranteed. This holds across experimental vertical preference tiers.

## 5.2   Treatment effects by buyer experience with the platform

We might expect that buyers differ in their response to the guarantee, particularly with respect to their experience with the platform. We explore this possibility in Figure 7, which reports the marginal effects of guarantee on different sub-populations of buyers and across the two experiments.

For each buyer, we divide them by whether they have prior experience with the platform, meaning they have made at least one hire before the start of the experiment. We label those with experience "Experienced" and those without experience "Inexperienced." For each vertical preference tier and prior experience level, we compute the effects of the guarantee on the probability that the buyer formed a contract. We also include an estimate "Pooled" that is both experienced and inexperienced buyers pooled together.

There is little evidence that buyer experience with the platform matters—moving from inexperienced to experienced, we see no obvious jump in treatment effects for either experiment. Figure 7 shows point estimates that are close to zero for the low and medium tiers, for both experience levels of buyers, for both experiments. In the high tier, we can see that those buyers offered the guarantee in the first experiment were more likely to form a contract (which recapitulates Table 3). However, in the second experiment, we can see that the guarantee under-performs relative to simply offering a recommendation. This is evidence that it was not the financial component of the money-back guarantee that matters so much as the informational

33

Figure 6: Application-level estimates of effects of being above the threshold by experimental group, tier and experiment
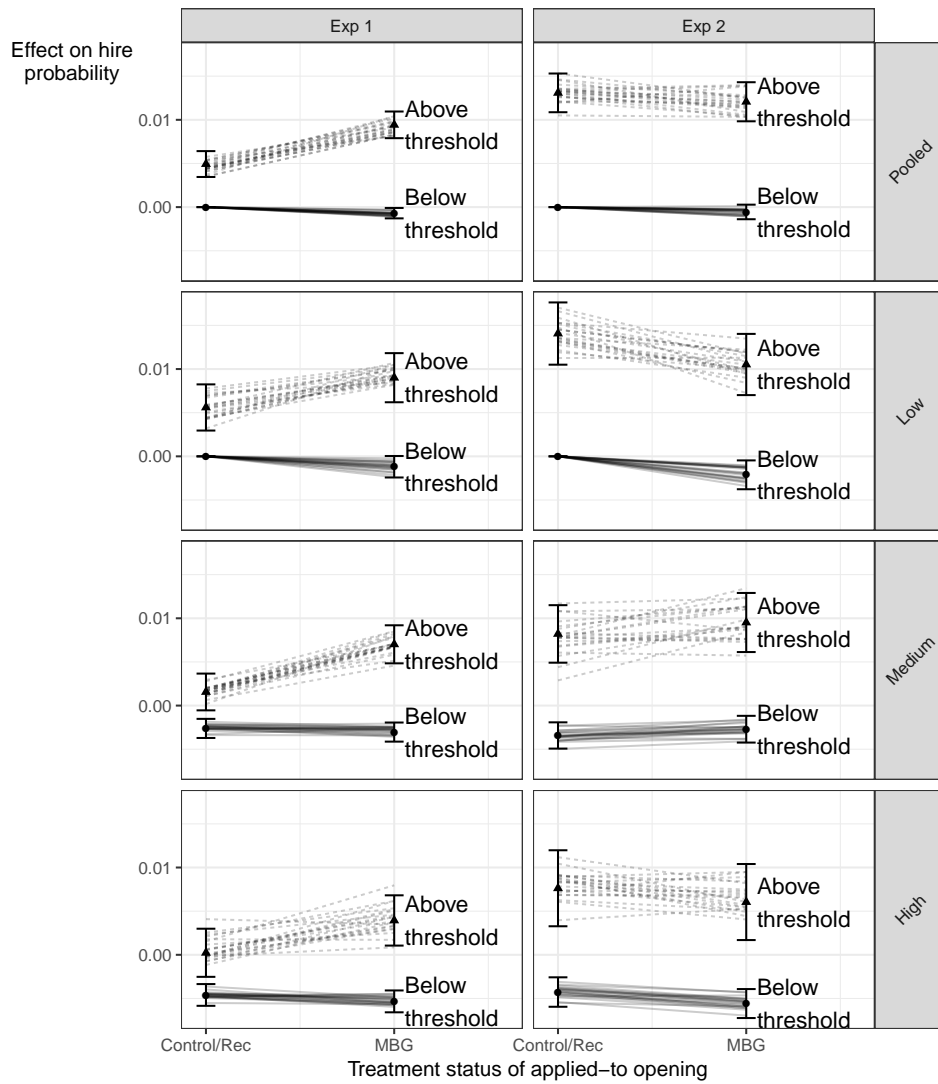


*Notes:* This figure plots marginal effects based on Equation 2. The gray lines indicate block-bootstrap estimates of the effects, sampling freelancers with replacement.

Figure 7: Effects of offering a guarantee on whether a contract is formed, by buyer vertical preference and platform experience, for both experiments
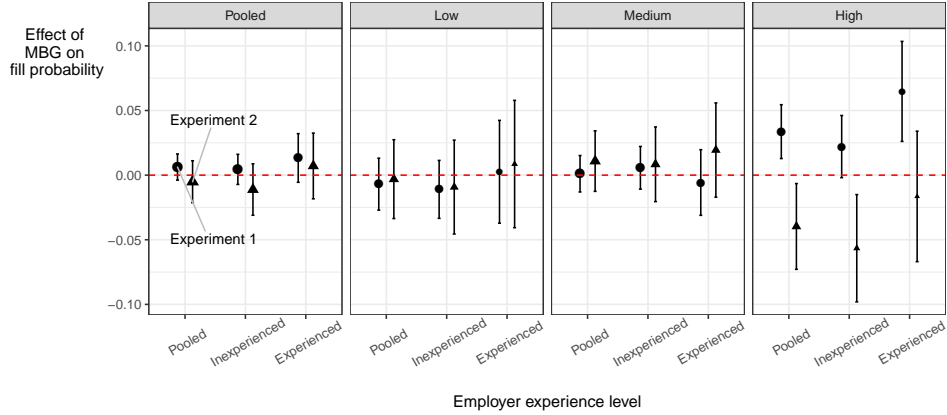


*Notes:* The figure reports regression estimates of the effects of offering a guarantee on whether a contract is formed. Effects are calculated by buyer platform experience, as well as their vertical preference over sellers. It includes estimates for both experiments.

content. Note that the standard errors in the high tier are larger, as the fraction of buyers selecting this tier is relatively small.

## 5.3 Effects by project duration

One feature of the guarantee and its time-limited window means that usefulness could vary by duration. If the project is substantially shorter than two weeks, the financial stakes are not large and whatever fixed cost is associated with claiming a refund might swamp the financial benefit. However, for a longer project, the returns to forming a good match could be higher, and so we might see the informational role of guarantees mattering more for longer projects.

In Figure 8, we explore this hypothesis by interacting the guarantee indicator with the expected project duration and report the effects on whether a contract was formed. This duration is chosen by buyers *ex ante* when describing their project. There are five possible durations, ranging from "Less than 1 week" to "More than 6 months." The various duration possibilities

Figure 8: Effects of the money-back guarantee on contract formation, by project duration, by experiment
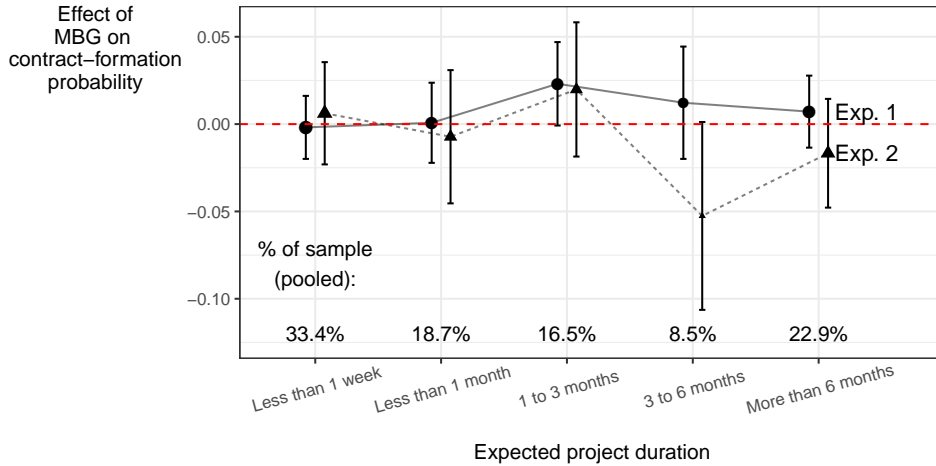


*Notes:* The figure reports regression estimates of the effects of offering a guarantee on whether a contract is formed. Effects are calculated by buyer expected project duration. It includes estimates for both experiments.

are shown on the x-axis of Figure 8. Above each duration label, the percentage of buyers choosing that label is shown, with observations pooled across the two experiments.

From the first experiment, there is some limited evidence of a greater probability of contract formation with longer durations. However, the evidence is exceedingly modest, in that the 95% CI for each point estimate comfortably includes zero. In the second experiment, the effects of a guarantee are also all fairly closer to zero. For the longer duration projects, the point estimate changes signs relative to the first experiment, undercutting the notion that buyers with longer duration projects find the guarantee more useful. The patterns of data are most consistent with the hypothesis that there is no substantial difference in the effectiveness of guarantees with respect to project duration. This is again consistent with the view that guarantees were valued for their informational content rather than their direct financial implications.

## 5.4 Comparison of buyer uptake of recommendations/guarantees in the two experiments, by buyer attributes

Although we have no evidence that the effects of the guarantee or recommendations on contract formation differed by buyer type, we can also look at differences in the "uptake" of recommendations, or whether the buyer contracted with a recommended/guaranteed seller. Several main dimensions on which buyers/job openings differ that might affect recommendation uptake include: (1) number of applicants received, (2) the number of applicants recruited, (3) the hourly wage of applicants, (4) number of prior openings, and (5) tenure on the platform. The last two attributes in particular might be correlated with trust in the platform.

One complication with using many different proxies is that they are on different scales, making it difficult to compare effects—or to show any non-linearities in effects. As a solution, for each proxy, we break it up into quartiles and then interact those group indicators with the treatment. For some measures that have substantial point mass at zero (greater than 25%), quartiles are not possible, and so for those, the largest number of cuts that can be included are shown.

In Figure 9, we plot the probability the buyer contracted with a seller above the threshold along with the standard error for that prediction, by quartile. We restrict the sample to only those openings where a contract was formed. Several things are readily apparent: In Experiment 1, treated buyers were systematically more likely to hire a seller above the threshold—MBG effects are above the control effects. Second, there is not much evidence of difference in treatment effects by buyer attributes—the gap seems relatively stable. Comparing across experiments, we see generally higher levels of uptake in Experiment 2 and the same "shape" of effects, but no evidence of a difference between MBG and simply recommending. Taken together, there is simply not much of evidence of heterogeneous treatment effects.

Figure 9: Uptake of recommendations
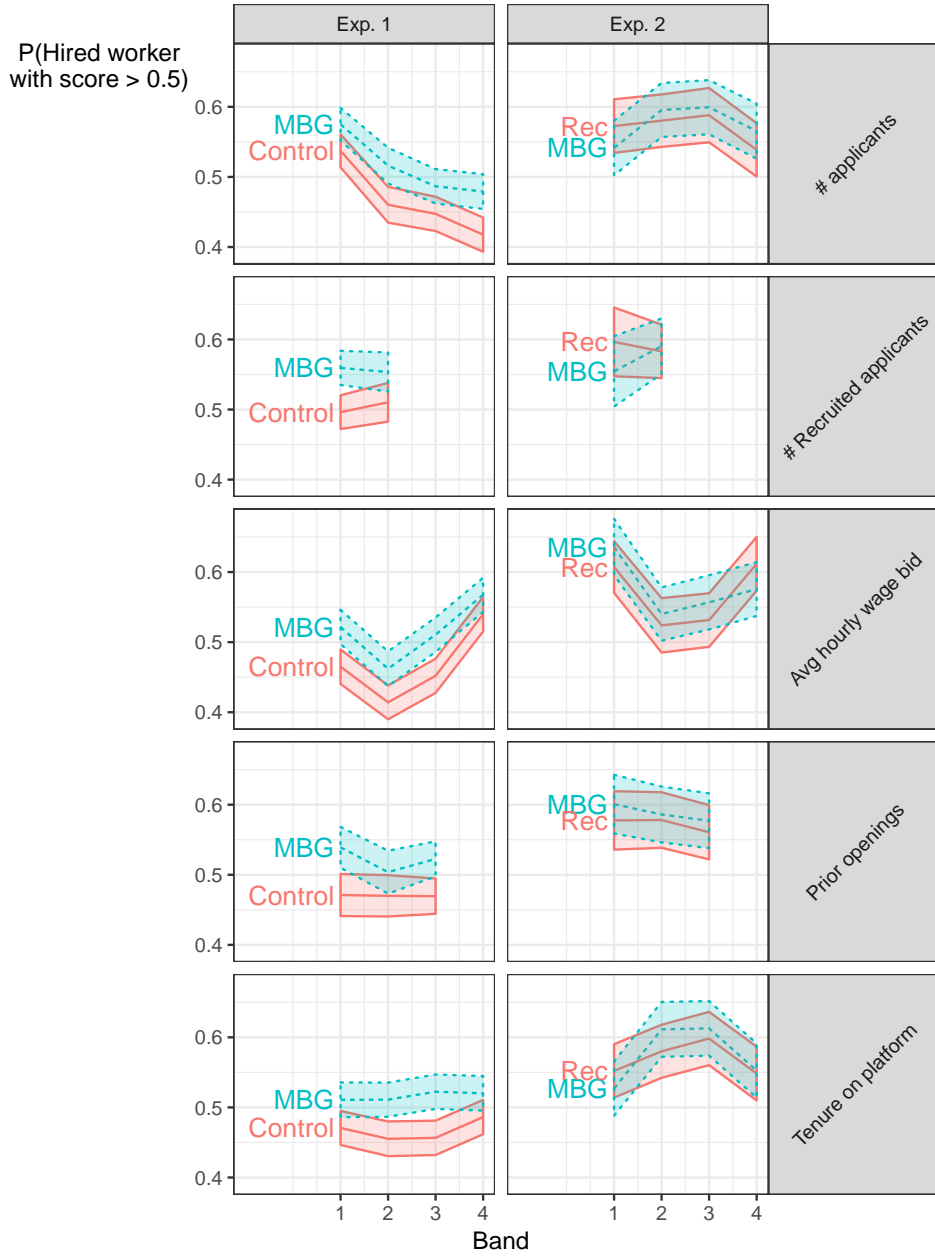
*Notes:* This figure shows the probability the buyer contracted with a seller above some threshold along with the standard error for that predicted mean, by quartile for a number of buyer attributes.

# 6 Conclusion

The main finding of our experiments is that buyers are more likely to contract with guaranteed sellers, but not because of the direct financial effect of the guarantee, but rather because they view the guarantee as informative about latent seller quality. Despite affecting which seller is selected, guarantees do not seem to have much effect on whether the marginal buyer forms a contract at all, at least overall. It is possible that guarantees increase contract formation for "high tier" buyers, but this is likely because these buyers value the information conveyed by the guarantee rather than the guarantee *per se*, as we saw no evidence of a difference between guarantees and recommendations in our second experiment.

Although a large literature documents the usefulness of algorithmic recommendations (Adomavicius and Tuzhilin, 2005), it is surprising that backing recommendations with money (in the form of the guarantee) did no better than recommendations not backed with money. This matches the findings of Panniello et al. (2017), who find that adapting recommendations to reflect the seller's profit margins for different products had no detectable negative effects.

Despite the lack of effects on quantity of transactions, it is beyond the scope of this paper to answer definitively whether the guarantee was a good idea, as the critical question is whatever spillover effects those presumably happier refunded buyers had on the platform. However, we find no evidence that buyers were any more satisfied or more likely to use the platform in the future. Further, to the extent refund-seeking buyers are hard to please, it is unclear how desirable they would be as long-term customers.

One reason the platform's recommendations might be seen as credible is that it is a relatively disinterested party—it makes money regardless of the seller selected. This stands in sharp contrast to non-platform retailers, who gain no direct benefit from a sale that instead goes to a competitor. This feature raises an interesting question as to whether alternative configurations that create seller-specific incentives—say by the platform owning some sellers and/or competing with some sellers directly (as in Zhu and

Liu (2016))—reduces recommendation credibility. Hagiu and Wright (2016) present a model in which a platform can endogenously have a mixture of relationships with sellers, some of which offer more revenue to the platform than others, which would presumably create an incentive to tilt business towards preferred sellers. This platform credibility issue also arises in search engines. Although search engines such as Google have generally been careful to avoid favoring their own products in search, they are not always successful—consider Google's recent $2.7 billion fine by the EU for favoring Google's own shopping platform in search results.[13]

Another interesting question for future work is the economics of requiring sellers to offer guarantees. In this study, the platform paid out the refunds, but it could compel sellers to share some of the cost of making refunds.[14] Having sellers partially pay for refunds would help overcome the challenging economics of platform-provided guarantees. It is also an interesting question as to whether would-be buyers would view seller-provided guarantees differently. Seller-backed guarantees might signal more information about latent seller quality if the seller is the one choosing. However, as these guarantees would not be based on the platform's presumably superior information about what seller is actually best for the buyer, they may be seen as less informative. On the other hand, a platform that compels seller-provided guarantees might be attractive to the best sellers, so there might be some platform competition benefits to requiring seller-provided guarantees.

Another potential avenue for future work is understanding the potential customer selection of effects of guarantees. Although we found that among existing buyers, the financial component of the guarantee had little effect, it is not difficult to imagine that particularly risk-averse buyers are deterred by the financial risk. This hypothesis could be relatively easy to test with marketing experiments that altered access (or salience) of a guarantee at the "top of the funnel" when buyers are first considering the platform and

---

[13]https://www.cnbc.com/2017/06/27/eu-hits-google-with-a-record-antitrust-fine-of-2-point-7-billion.html

[14]eBay requires sellers to compensate buyers requesting a refund, though there are extensive conditions that make the guarantee program not as buyer-friendly as it might seem.

then measuring uptake. Of course, whether risk-averse customers are worth acquiring—this could easily be an adversely selected population—is a separate question. A related question is whether a guarantee could be offered to *some* sellers who have a high willingness to pay for it. Although adverse selection issues might be particularly important for this sub-population, if there are highly financially risk-averse populations, the platform might be able to profitably offer guarantees, particularly given the platform's birds eye view and ability to price risk.

Perhaps the most promising direction for future work is greater exploration of how the platforms can use their substantial steering power. Whatever trust or goodwill is needed for these recommendations to be followed could be squandered if the platform began using that market-shaping power to serve its own ends rather than that of the buyers. But it is not necessarily black or white, and on the margin, the platform can use the recommendation power to pursue platform goals. For example, there are likely substantial benefits to a new, inexperienced sellers "breaking in" to the market (Pallais, 2013) or the platform helping buyers avoid over-subscribed sellers (Horton, Forthcoming). If the platform can given some extra "weight" to a new but otherwise promising and qualified applicant, it would likely improve platform efficiency.

# References

**Adomavicius, Gediminas and Alexander Tuzhilin**, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Transactions on Knowledge and Data Engineering*, 2005, *17* (6), 734–749.

**Agrawal, Ajay K, John Horton, Nico Lacetera, and Elizabeth Lyons**, "Digitization and the Contract Labor Market: A Research Agenda," in "Economics of Digitization," University of Chicago Press, 2013.

_ , **Nicola Lacetera, and Elizabeth Lyons**, "Does information help or hinder job applicants from less developed countries in online markets?," January 2013, (NBER Working Paper 18720).

**Allon, Gad, Achal Bassamboo, and Eren B. Çil**, "Large-scale service marketplaces: The role of the moderating firm," *Management Science*, 2012, *58* (10), 1854–1872.

**Autor, David**, "The economics of labor market intermediation: an analytic framework," Technical Report, National Bureau of Economic Research 2008.

**Barach, Moshe A. and John Horton**, "How do Employers User Compensation History? Evidence from a Field Experiment," *Working Paper*, 2017.

**Blake, Thomas and Dominic Coey**, "Why marketplace experimentation is harder than it seems: The role of test-control interference," in "Proceedings of the fifteenth ACM conference on Economics and computation" ACM 2014, pp. 567–582.

**Bryant, W Keith and Jennifer L Gerner**, "The price of a warranty: The case for refrigerators," *Journal of Consumer Affairs*, 1978, *12* (1), 30–47.

**Chan, Jason and Jing Wang**, "Hiring preferences in online labor markets: Evidence of a female hiring bias," *Management Science*, 2017.

**Coase, Ronald H**, "Payola in radio and television broadcasting," *The Journal of Law and Economics*, 1979, *22* (2), 269–328.

**Cohen, Maxime, Michael-David Fiszer, and Baek Jung Kim**, "Frustration-Based Promotions: Field Experiments in Ride-Sharing," *Working paper*, 2018.

**Cusumano, Michael**, "Technology strategy and management: The evolution of platform thinking," *Communications of the ACM*, 2010, *53* (1), 32–34.

**Doleac, Jennifer L and Luke CD Stein**, "The visible hand: Race and online market outcomes," *The Economic Journal*, 2013, *123* (572), F469–F492.

**Eisenmann, Thomas, Geoffrey Parker, and Marshall W Van Alstyne**, "Strategies for two-sided markets," *Harvard Business Review*, 2006, *84* (10), 92.

**Filippas, Apostolos, John Horton, and Joe Golden**, "Reputation Inflation," *Working Paper*, 2017.

**Gao, Guodong, Brad N Greenwood, Ritu Agarwal, and Jeffrey S McCullough**, "Vocal minority and silent majority: How do online ratings reflect population perceptions of quality," *MIS Quarterly*, 2015, *39* (3), 565–590.

**Garvin, David A**, "Quality on the line," *Harvard Business Review*, 1983, *61* (5), 64–75.

**Gerner, Jennifer L and W Keith Bryant**, "Appliance warranties as a market signal?," *Journal of Consumer Affairs*, 1981, *15* (1), 75–86.

**Goes, Paulo and Mingfeng Lin**, "Does information really 'unravel'? Understanding factors that motivate sellers to seek third-party certifications in an online labor market," *Working Paper*, 2012.

**Grossman, Sanford J**, "The informational role of warranties and private disclosure about product quality," *The Journal of Law & Economics*, 1981, *24* (3), 461–483.

**Hagiu, Andrei**, "Strategic decisions for multisided platforms," *MIT Sloan Management Review*, 2014, *55* (2), 71.

_ **and Julian Wright**, "Controlling versus enabling," *Working Paper*, 2016.

**Halperin, Basil, Benjamin Ho, John List, and Ian Muir**, "Toward an understanding of the economics of apologies: evidence from a large-scale natural field experiment," 2018.

**Heal, Geoffrey**, "Guarantees and risk-sharing," *The Review of Economic Studies*, 1977, *44* (3), 549–560.

**Hong, Yili and Paul A Pavlou**, "Is the world truly "flat"? Empirical evidence from online labor markets," *Working Paper*, 2015.

**Horton, John**, "Online labor markets," *Internet and Network Economics*, 2010, pp. 515–522.

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

_ , "Price floors and employer preferences: Evidence from a minimum wage experiment," *Working paper*, 2017.

**Horton, John J**, "Buyer Uncertainty about Seller Capacity: Causes, Consequences, and a Partial Solution," *Management Science*, Forthcoming.

**Horton, John J. and Ramesh Johari**, "Buyer Signaling Improves Matching: Evidence from a Field Experiment," *Working Paper*, 2018.

**Horton, John, William R Kerr, and Christopher Stanton**, "Digital labor markets and global talent flows," Technical Report, National Bureau of Economic Research 2017.

**Knoepfle, Dan, Jonathan V. Hall, and John J. Horton**, "Pricing Efficiently in Designed Markets: Evidence from Uber," *Working Paper*, 2018.

**Lutz, Nancy A**, "Warranties as signals under consumer moral hazard," *The RAND journal of economics*, 1989, pp. 239–255.

**McWilliams, Bruce**, "Money-back guarantees: Helping the low-quality retailer," *Management Science*, 2012, *58* (8), 1521–1524.

**Nosko, Chris and Steven Tadelis**, "The limits of reputation in platform markets: An empirical analysis and field experiment," *Working Paper*, 2015.

**Pallais, Amanda**, "Inefficient hiring in entry-level labor markets," *American Economic Review*, March 2013, (18917).

**Panniello, Umberto, Michele Gorgoglione, Shawndra Hill, and Kartik Hosanagar**, "Incorporating profit margins into recommender systems: A randomized field experiment of purchasing behavior and consumer trust," *Working Paper*, 2017.

**Parker, Geoffrey G, Marshall W Van Alstyne, and Sangeet Paul Choudary**, *Platform revolution: How networked markets are transforming the economy–and how to make them work for you*, WW Norton & Company, 2016.

**Priest, George L**, "A theory of the consumer product warranty," *The Yale Law Journal*, 1981, *90* (6), 1297–1352.

**Snir, Eli M and Lorin M Hitt**, "Costly bidding in online markets for IT services," *Management Science*, 2003, *49* (11), 1504–1520.

**Spence, Michael**, "Consumer misperceptions, product failure and producer liability," *The Review of Economic Studies*, 1977, *44* (3), 561–572.

**Stanton, Christopher T. and Catherine Thomas**, "Landing the first job: The value of intermediaries in online hiring," *The Review of Economic Studies*, 2016, *83* (2), 810–854.

**Yoganarasimhan, Hema**, "The value of reputation in an online freelance marketplace," *Marketing Science*, 2013, *32* (6), 860–891.

**Zhu, Feng and Qihong Liu**, "Competing with complementors: An empirical look at Amazon. com," 2016.

# A    Conceptual framework

To illuminate the implications of a guarantee and motivate our empirical analyses, we formalize the platform's guaranteeing problem. First, we take the platform's view and characterize when offering a universal seller guarantee would be directly profitable, given its effects on revenue per transaction and the number of transactions. Next, we present a model of buyers and sellers in a market and characterize the equilibrium effects of a guarantee. This allows us to described under what conditions a guarantee can be profitable for a platform. Finally, we consider how a guarantee would change a buyer's "micro" seller selection problem. We focus on what a Bayesian buyer would infer from the platform's offer of a guarantee.

## A.1    The platform's decision problem

Consider a platform offering a guarantee to dissatisfied buyers. Such a guarantee program could be profit-maximizing if the incremental sales obtained offset the presumably lower per-transaction revenue (due refunds to dissatisfied buyers).[15] A guarantee could increase the number of transactions on

---

[15]While it seems likely that offering refunds would simply lower per-transaction platform expected revenue, if a guarantee leads to better matches, it is possible that per-transaction revenue could increase.

the platform because the platform has increased the expected surplus of buyers per-transaction directly (particularly if buyers are risk-averse), but also because of repeat business, better word-of-mouth, and so on.

Let $R$ be the average platform revenue per transaction, and let $Q$ be the total number of transactions. Let $\Delta R$ be the corresponding change in revenue from offering a guarantee, and let $\Delta Q$ be the change in the number of transactions. The platform is just indifferent to offering a guarantee if

$$(R - \Delta R)(Q + \Delta Q) - RQ = 0$$
$$\left(\frac{R - \Delta R}{R}\right)\left(\frac{Q + \Delta Q}{Q}\right) = 1$$
$$|\Delta R\%| + |\Delta Q\%| \approx 0. \tag{4}$$

**Remark 1.** *The platform finds it profitable to offer guarantees if the percentage increase in transactions is greater than the percentage decrease in per-transaction revenue.*

## A.2    Marketplace perspective

The actual effects of a guarantee depend on how the guarantee affects the marketplace—specifically $Q$ and $R$—in equilibrium. We now present a model of buyers and sellers that will allow us to characterize the equilibrium effects of a guarantee. Buyers have some project they would like completed. All sellers have the same probability, $p \in (0, 1)$, of being able to complete the project "successfully." A successfully completed project is worth $y = 1$ to the buyer. There is a cost to not having the project completed, which we can think of as the cost of delay given that the buyer can always return to the marketplace. This delay cost is $c \geq 0$. As such, an unsuccessful project gives a payoff of $y = -c$.

A seller proposes a total price $w$ for *attempting* to complete the project. If a contract is formed, the buyer has to pay the seller even if the project is not successful. The buyer's expected payoff is

$$\pi = p - w - (1 - p)c. \tag{5}$$

If sellers are paid their expected product, then in equilibrium

$$w = p - (1-p)c. \tag{6}$$

Note that for the market to exist, $p \geq \frac{c}{1+c}$. Now suppose that whether the output is produced is verifiable by the platform, and the buyer is reimbursed $w$ if the project is not completed successfully. The buyer is not compensated for the delay cost $c$.

The platform guarantee is essentially a market subsidy. The incidence of this subsidy—i.e., the equilibrium effect on $w$—would depend on the relative supply and demand elasticities. For example, suppose sellers are completely inelastic, but buyers are elastic with respect to the surplus. Offering a guarantee increases demand from buyers, but the supply of sellers does not increase, and so wages rise as the new buyers compete for sellers, causing those sellers to capture the subsidy with higher wages, which the per-transaction revenue depends upon.

To consider the platform's revenue explicitly, we introduce an ad valorem charge, $\tau$, that platform imposes. With this charge, when the agreed-upon price is $w$, $(1-\tau)w$ is paid to the seller, and $w\tau$ goes to the platform. Note that this $w$ would reflect the incidence of the platform's charge. With the guarantee, the platform's revenue changes to $\tau w - (1-p)w$, before any equilibrium adjustment in $w$, and so

$$\Delta R\% = \frac{1-p}{\tau}.$$

For the buyer, with the guarantee, the price faced changes to $(1-\tau)(w + (1-p))$, and so

$$\Delta w\% = 1 - p$$
$$= \tau \Delta R\%. \tag{7}$$

**Remark 2.** *As the platform takes only some fraction of the transaction i.e., $\tau < 1$, offering a guarantee always has a larger effect on revenue, in percentage terms, for the platform than it does for buyers and sellers (in*

*the absence of some "clawback" from the sellers who contracted with buyers seeking a refund).*

Now we consider how the market adjusts due to the introduction of a guarantee, which we assume is a subsidy small enough that the typical linearization assumptions of comparative statics hold. Assume that buyers collectively have a demand elasticity of $\epsilon_w^D$ and sellers have a supply elasticity of $\epsilon_w^S$. For buyers and sellers to have finite (or non-zero) elasticities, there would have to be some idiosyncratic components to Equation 5 and Equation 6. The subsidy is $\tau \Delta R\%$, in percentage terms, of which sellers get a fraction $x$ and buyers get $1 - x$. The total change in the quantity of transactions is

$$\Delta Q\% = \frac{1}{2}\tau \Delta R\% \left( x|\epsilon_w^S| + (1-x)|\epsilon_w^D| \right).$$

As marketing clearing requires that $x|\epsilon_w^S| = (1-x)|\epsilon_w^D|$, we have that

$$\Delta Q\% = \tau \Delta R\% \left( \frac{1}{|\epsilon_w^S|} + \frac{1}{|\epsilon_w^D|} \right)^{-1}.$$

This condition, combined with Equation 4, implies offering a guarantee is profitable for the platform if

$$\left( \frac{1}{|\epsilon_w^S|} + \frac{1}{|\epsilon_w^D|} \right)^{-1} > \frac{1}{\tau}. \tag{8}$$

The condition in Equation 8 is fairly intuitive—a guarantee "works" if the quantity of *market* transactions is collectively highly elastic—when this is the case, even a small reduction in platform revenue-per-transaction leads to a large increase in the number of transactions.

We can see from Equation 8 that the smaller the platform ad valorem charge, the harder it is for a guarantee to be profitable for the platform because it requires market transactions to be exceptionally elastic. For platform charges we see in practice in platform markets—10% to 30% is typical—the market transaction elasticity has to be quite high—3.3 in the case of a 30% charge, and 10 in the case of a 10% charge for a guarantee to

be profitable. Although the implied values of $\epsilon_w^D$ and $\epsilon_w^S$ would be very large to create a market elasticity of 10 (e.g., 20 each if symmetric), recall that these are elasticities with respect to the *platform*, which could be quite high if switching costs are low. As a case in point, Knoepfle et al. (2018) finds that drivers on Uber have a market labor supply elasticity indistinguishable from infinity, at least in the "long-run" of about 8 weeks.

High platform elasticities are likely found in practice, though at least one side of the market has to be somewhat inelastic, as the ability of the platform to impose a charge $\tau > 0$ depends on it. However, this highlights the difficulty of the platform's problem for a guarantee to be profitable: market transaction elasticities have to be large, but if this is the case, the ad valorem charge has to be small (otherwise these elastic buyers and sellers would switch to other platforms), which in turn implies the market has to be *very* elastic.

## A.3 Effects of a guarantee on selection

In the model sketched above, we have assumed that all sellers have the same $p$. Furthermore, we assumed all buyers and sellers are price-takers. In reality, sellers differ, and we will now add seller heterogeneity in $p$ and no longer assume that this $p$ pins down the wage. These changes allow us to explore how the guarantee affects the buyer's selection of a seller when he or she has multiple sellers to choose from. To keep things simple, we will also assume that $c = 0$.

A seller $i$ has a probability $p_i \in (0,1)$ of being able to complete the project "successfully." Without a $c$, a successfully completed project is worth $y = 1$ to the buyer, whereas an unsuccessful project is now $y = 0$. A seller proposes a total price $w_i$ for *attempting* to complete the project. If a contract is formed, the buyer has to pay the seller even if the project is not successful. The buyer's expected payoff from selecting seller $i$ is $\pi_i = p_i - w_i$. If the platform offers a guarantee, the payoff from contracting with

a guaranteed seller is

$$\pi_i^{\mathrm{MBG}} = p_i - p_i w_i.$$

**Remark 3.** *As $p_i < 1$ and $w_i > 0$, a guaranteed seller offers a higher expected payoff to the buyer.*

The guarantee also affects buyer price sensitivity, in the sense that a small increase in $w_i$ has different implications for the payoff obtained from that seller, depending on whether a guarantee is offered.

**Remark 4.** *The marginal effect on the payoff from the proposed change has a smaller magnitude for the guaranteed seller than for the non-guaranteed seller, as $\partial \pi^G / \partial w = -1 + p$, whereas for a non-guaranteed seller, $\partial \pi / \partial w = -1$.*

If we imagine the buyer as selecting among several sellers, the above remark implies, all else equal, a guaranteed seller could raise his or her price more without causing the buyer to switch to some other seller, compared to a non-guaranteed seller.

If sellers differ in $p$, the platform will find guaranteeing some sellers cheaper, though because $p_i < 1$, the platform guaranteeing a seller always faces some expected cost.

**Remark 5.** *All else equal, sellers with the highest probability of success are the least expensive for the platform to guarantee, as the expected costs to guaranteeing a seller are $(1 - p_i)w$.*

An implication of the above remark is that a buyer who is uncertain about a seller's $p$ might view the platform's decision to guarantee as informative—i.e., the platform is more likely to guarantee a seller it is confident will complete the project successfully. Note that as $w$ is common knowledge, the guaranteeing decision would specifically be informative about $p$.

Suppose buyers know the distribution of seller success probabilities which forms their prior, $h(p)$. The platform receives a signal, $p + \epsilon$ where $\epsilon \sim N(0, \sigma)$ where $p$ is the seller's true success probability. Let $f(\cdot)$ and $F(\cdot)$ be

the partial and cumulative density functions of $\epsilon$, respectively. The buyer does not observe the platform's signal, but does observe whether the platform offers a guarantee in response the the *platform's* posterior on the seller's success probability. The platform's optimal guarantee would be a wage-conditioned cutoff rule, guaranteeing any seller with a $p > \underline{p}|w$.

**Remark 6.** *So long as the platform does not guarantee all sellers and guarantees on the basis of an informative private signal, the offer of a guarantee can only revise upwards the buyer's beliefs about the probability the seller can complete the project successfully.*

Let MBG $= 1$ indicate that a given seller came with a money-back guarantee. If the platform offered a guarantee, it implies that the signal it received was above its threshold, or

$$Pr\{\text{MBG} = 1\} = Pr\{p + \epsilon > \underline{p}\}$$
$$= F(p - \underline{p}).$$

From the Bayesian buyer's perspective, who had prior $h(p)$ about a seller, observing a guarantee gives him or her a posterior

$$h(p|\text{MBG} = 1) \propto h(p)F(p - \underline{p}).$$

The monotone likelihood ratio property (MLRP) holds in $p$ for the prior and the posterior, as

$$\frac{\partial}{\partial p}\left(\frac{h(p|\text{MBG} = 1)}{h(p)}\right) = \frac{f(p - \underline{p})}{\int_0^1 h(x)F(x - \underline{p})\,dx} > 0,$$

and since the MLRP implies first order stochastic dominance, i.e., $H(p|\text{MBG} = 1)$ is below $H(p)$ for all $p$, then we have that $\mathbf{E}[p|\text{MBG} = 1] > \mathbf{E}[p]$. By the same reasoning, a buyer infers the a seller that is not guaranteed has a lower probability of completing the project compared to that buyer's prior.

Figure 10: Mean applicant quality score over time
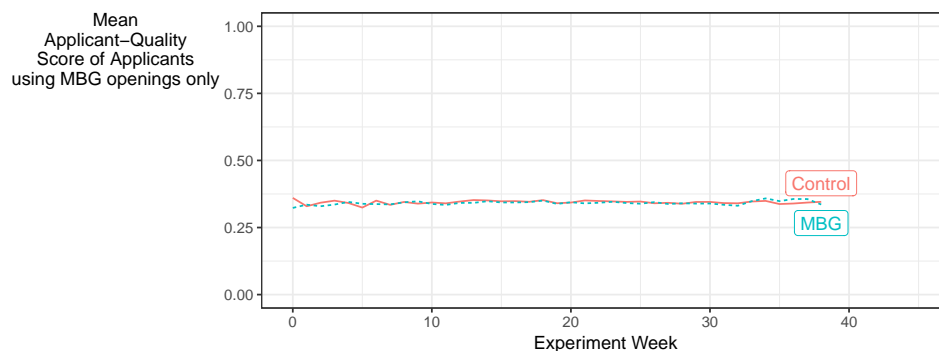


*Notes:* This figure plots the by-week average applicant score, with scores computed on a seller-specific basis, pooled over all applications.

# B  Internal validity

Table 8 shows the balance table for the first experiment; Table 9 shows the balance table for the second experiment. In both cases, the randomization was effective, and the treatment and control groups are well balanced.

# C  SUTVA violations

One concern with our experiment—and any marketplace experiment—is interference across experimental cells (Blake and Coey, 2014). In our experiment, the treatment could lead to better sellers getting selected and "taken off the market." If sellers are inelastic, this could lead to fewer good sellers being available for buyers later in the experiment. Although this could cause a decline in average quality, it would affect both treatment and control buyers. Despite this possibility, we find no evidence that this is the case—for each seller, we average his or her score over all job applications to *MBG-treated sellers.*

In Figure 10, we plot the by-week average score to all applicants. We find no evidence of a difference between treatment and control, nor any trend.

Table 8: Balance for Experiment I

| | Control Mean: $\bar{X}_{CTL}$ | Treatment Mean: $\bar{X}_{TRT}$ | Difference In Means | p-value | |
|---|---|---|---|---|---|
| *Employer Attributes* | | | | | |
| Prior Job Postings | 7.56 (0.15) | 7.51 (0.15) | -0.05 (0.21) | 0.82 | |
| Prior Billed Jobs | 3.25 (0.09) | 3.20 (0.08) | -0.05 (0.12) | 0.68 | |
| Prior Spend by Employers | 2,867.08 (172.77) | 2,970.86 (177.94) | 103.78 (247.97) | 0.68 | |
| Num Prior Contractors | 3.28 (0.08) | 3.26 (0.09) | -0.02 (0.12) | 0.87 | |
| Avg Feedback Score of Employer | 4.80 (0.01) | 4.79 (0.01) | -0.01 (0.01) | 0.40 | |
| Num of Reviews of Employer | 2.34 (0.07) | 2.34 (0.07) | -0.01 (0.10) | 0.95 | |
| *Job Posting Attributes* | | | | | |
| Number non-invited Applicants | 25.22 (0.23) | 25.43 (0.24) | 0.21 (0.33) | 0.52 | |
| Avg Best Match Score | 0.36 (0.00) | 0.36 (0.00) | 0.00 (0.00) | 0.08 | † |
| Avg Bid | 13.09 (0.08) | 13.22 (0.09) | 0.13 (0.11) | 0.26 | |
| Prefered Experiance in Hours | 31.53 (0.82) | 30.03 (0.82) | -1.50 (1.16) | 0.19 | |
| Estimated Job Duration in Weeks | 15.52 (0.14) | 15.38 (0.14) | -0.14 (0.19) | 0.48 | |
| *Applicant Attributes* | | | | | |
| Hours Worked to Date | 682.85 (4.61) | 692.59 (4.85) | 9.74 (6.69) | 0.15 | |
| Num Past Jobs Worked | 15.03 (0.08) | 14.99 (0.08) | -0.04 (0.12) | 0.73 | |
| Past Hourly Earnings | 6,545.21 (55.54) | 6,671.21 (58.99) | 126.00 (81.02) | 0.12 | |
| Num Prior Employers | 11.97 (0.06) | 11.95 (0.06) | -0.01 (0.09) | 0.88 | |
| Min Feedback Rating | 0.33 (0.00) | 0.33 (0.00) | 0.00 (0.00) | 0.59 | |
| Wage Bid | 10.80 (0.07) | 10.91 (0.08) | 0.12 (0.10) | 0.26 | |
| Profile Wage | 10.66 (0.07) | 10.73 (0.07) | 0.07 (0.10) | 0.50 | |

*Notes:* This table reports means and standard errors across experimental groups of various attributes. The top panel reports characteristics of buyers allocated to treatment and control. The middle panel reports characteristics of requests for proposals by treatment and control groups for the first request for proposals submitted by that buyer after allocation to the experiment, for each buyer. The bottom panel reports characteristics of buyers at the time they were allocated to treatment or control groups. Reported p-values are the for two-sided t-tests of the null hypothesis of no difference in means across groups. In the bottom panel, standard errors are clustered at the buyer level. Significance indicators: $p \leq 0.10 : \dagger$, $p \leq 0.05 : *$ and $p \leq .01 : **$

Table 9: Balance for Experiment 2: Guaranteed vs. Recommended

| | Control Mean: $\bar{X}_{CTL}$ | Treatment Mean: $\bar{X}_{TRT}$ | Difference In Means | p-value | |
|---|---|---|---|---|---|
| *Employer Attributes* | | | | | |
| Prior Job Postings | 16.61 (0.49) | 16.48 (0.51) | -0.13 (0.71) | 0.85 | |
| Prior Billed Jobs | 7.73 (0.26) | 7.63 (0.28) | -0.10 (0.38) | 0.80 | |
| Prior Spend by Employers | 6,045.07 (408.75) | 6,239.66 (556.60) | 194.59 (688.11) | 0.78 | |
| Num Prior Contractors | 7.89 (0.27) | 7.61 (0.27) | -0.28 (0.38) | 0.47 | |
| Avg Feedback Score of Employer | 4.81 (0.01) | 4.81 (0.01) | 0.00 (0.01) | 0.84 | |
| Num of Reviews of Employer | 5.64 (0.21) | 5.58 (0.21) | -0.05 (0.30) | 0.86 | |
| *Job Posting Attributes* | | | | | |
| Number non-invited Applicants | 29.92 (0.44) | 29.46 (0.44) | -0.46 (0.62) | 0.46 | |
| Avg Best Match Score | 0.40 (0.00) | 0.40 (0.00) | -0.00 (0.00) | 0.97 | |
| Avg Bid | 13.16 (0.14) | 13.11 (0.12) | -0.04 (0.19) | 0.82 | |
| Prefered Experiance in Hours | 28.13 (1.33) | 30.93 (1.47) | 2.80 (1.98) | 0.16 | |
| Estimated Job Duration in Weeks | 16.53 (0.24) | 17.11 (0.25) | 0.58 (0.35) | 0.10 | † |
| *Applicant Attributes* | | | | | |
| Hours Worked to Date | 726.97 (7.95) | 741.54 (8.62) | 14.57 (11.73) | 0.21 | |
| Num Past Jobs Worked | 15.81 (0.14) | 15.97 (0.16) | 0.16 (0.21) | 0.46 | |
| Past Hourly Earnings | 7,154.92 (99.31) | 7,207.41 (103.18) | 52.48 (143.20) | 0.71 | |
| Num Prior Employers | 12.54 (0.11) | 12.68 (0.12) | 0.14 (0.16) | 0.37 | |
| Min Feedback Rating | 0.37 (0.00) | 0.37 (0.00) | 0.00 (0.00) | 0.34 | |
| Wage Bid | 10.98 (0.12) | 10.89 (0.12) | -0.08 (0.17) | 0.64 | |
| Profile Wage | 10.93 (0.11) | 10.87 (0.11) | -0.06 (0.15) | 0.70 | |

*Notes:* This table reports means and standard errors across experimental groups of various attributes. The top panel reports characteristics of buyers allocated to treatment and control. The middle panel reports characteristics of requests for proposals by treatment and control groups for the first request for proposals submitted by that buyer after allocation to the experiment, for each buyer. The bottom panel reports characteristics of buyers at the time they were allocated to treatment or control groups. Reported p-values are the for two-sided t-tests of the null hypothesis of no difference in means across groups. In the bottom panel, standard errors are clustered at the buyer level. Significance indicators: $p \leq 0.10 : \dagger$, $p \leq 0.05 : *$ and $p \leq .01 : **$.