

“Skin-in-the-Game” and Platform Credibility: Evidence from Field Experiments

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

November 3, 2017

Abstract

A platform marketplace guaranteed certain sellers for a randomly selected pool of would-be buyers. Offering a guarantee did not increase sales overall, but it did cause buyers to preferentially contract with guaranteed sellers, at the expense of those not guaranteed. However, other evidence from the experiment suggests buyers shifted to guaranteed sellers not because they offered 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” select sellers, with no offer of a guarantee, was equally effective at shifting buyers towards those selected sellers.

*Georgetown University

†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-backed guarantee. With a guarantee, the financial risk of unsatisfactory product performance is shifted to the seller (Heal, 1977), ostensibly making a purchase more likely. Furthermore, this guarantee can signal high quality, if guaranteeing a high quality product is cheaper than guaranteeing a low quality product (Spence, 1977; Grossman, 1981). Perhaps unsurprisingly, these kinds of guarantees are exceedingly common in conventional retail settings (McWilliams, 2012).¹ However, the rationale for offering guarantees in a platform marketplace setting is less straightforward.

Consider a platform marketplace in which independent sellers offer goods that are somewhat substitutable for each other. On the one hand, with many small sellers, a would-be buyer is likely less knowledgeable about sellers and their relative quality, making a guarantee more useful. Furthermore, as Nosko and Tadelis (2015) show, in a platform marketplace setting, negative consumer experiences spill-over on to the platform as a whole, potentially making bad buyer experiences more costly to a “horizontal” platform compared to a single retailer not expecting much repeat business. On the other hand, if the platform is also uncertain about the quality of sellers—and has limited tools to control quality—guarantees could be expensive to offer, even if only some sellers are guaranteed. A guarantee could enable buyers to select “risky” sellers, encouraging a kind moral hazard in their selection, or lead to hard-to-please sellers gravitating to the platform, creating a kind of adverse selection. Finally, the economics of a guarantee could be challenging, in that to increase revenue, they have to stimulate a large enough increase in the

¹Money-back guarantees have been a central component to firm strategies at firms such as L.L. Bean, Publix, Trader Joe’s, Stew Leonard’s, Costco, Aldi, and Nordstrom. L.L. Bean’s 100 percent satisfaction guarantee has been part of the firm’s strategy since 1912, when the firm refunded money for a hunting boot whose poor design lead to the boot’s rubber bottom separating from the leather upper.

number of transactions to offset the additional costs of providing refunds.

Even if guarantees do not increase the number of transactions, if the platform can tilt the buyers towards transacting with “better” sellers, then offering a selective guarantee might be useful for the spill-over benefits [Nosko and Tadelis \(2015\)](#). However, a natural, important and previously unexplored question is whether a financial guarantee is necessary for tilting buyers. Simply recommending certain sellers might be sufficient, as platform recommendations have been shown to work well in other product settings ([Adomavicius and Tuzhilin, 2005](#)). However, guarantees might be viewed as a far more credible—and hence more effective—kind of recommendation, as the platform making the recommendation has some “skin in the game,” or a financial stake in a good outcome.

In this paper, we consider the effects of introducing a money-backed guarantee in a platform marketplace for services. We report the results of two experiments that, together, allow us to explore (1) whether guarantees are valued by buyers and (2) *why* they are valued by buyers. In the first experiment, treated buyers saw that certain candidate sellers were “guaranteed,” meaning that if contracted-with, the first two weeks of their work were guaranteed by the platform. In the second experiment, some buyers were randomized to see select 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, as the probability a contract was formed was unchanged by the guarantee treatment. However, treated buyers did strongly shift towards contracting with guaranteed sellers. We know this because we know which sellers the platform *would have* guaranteed, 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 no evidence of a shift towards more expensive sellers, suggests that the direct economic implications of the guarantee were not a first-order consideration for would-be buyers. Instead, buyers were interested in learning which sellers they *should* prefer, if they knew what the platform knew. Buyers might have thought that the platform would only be willing to guarantee those sellers least likely to fail to complete the project and trigger a refund, as those sellers are relatively cheap to guarantee. As such, the guarantee might still be needed for the buyer to find the platform credible. However, in our second experiment, we find that simply claiming a seller was “recommended”—with no guarantee—was just as effective at shifting buyers towards contracting with recommended/guaranteed sellers. In short, “skin [directly] in the game” was not necessary.

Our interpretation of the experimental results is that the platform had garnered sufficient trust that its recommendations—even not when “backed” by money—were valued (Horton, 2017a). 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 that platform has superior insights into relative seller quality, and (2) given that the platform has no strong preference over which seller is pursued, it has no incentive to give bad i.e., self-dealing advice.²

The main managerial implication of our findings is that guarantees were unnecessary for this marketplace, and that for the marginal buyers choosing not to purchase, the direct financial risk of contracting was not determinative. Despite a lack of effect on sales, a guarantee could still be worthwhile due to positive spill-overs from less unsatisfied customers that received refunds. However, we have no strong evidence that treated buyers seemed any more satisfied. If there were positive spill-overs, we could not detect them, as

²Although feedback provided by other other users is widely-used by buyers and seems to generally be 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.

they did not show up in our measures of buyer satisfaction or in more future business. Not un-relatedly, after the conclusion of the two experiments, the platform switched to only offering recommendations.

Our paper contributes to the growing literature on the management of online marketplaces. We clarify the strengths and limitations of an important “tool” platform designers have to overcome informational problems (Hagi, 2014; Eisenmann et al., 2006; Parker et al., 2016; Cusumano, 2010). The paper also offers some evidence for 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 follow. Section 2 discusses the empirical context. Section 3 presents a simple model of guarantees and recommendations. Section 4 describes the first experiment of offering a guarantee and present the results. Section 5 describes the second experiment comparing offering guarantees to describing a seller as recommended and presents the results. Section 6 concludes.

2 Empirical context

The empirical context for our study is a large online labor market. In these markets, firms hire 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. On the platform, would-be buyers write job descriptions,

self-categorize the nature of the work and required skills, and then post the vacancies to the platform website. Sellers learn about vacancies via electronic searches or email notifications.

Sellers submit applications, 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.

One interesting feature of this marketplace is that buyers are asked by the platform to state their “vertical” preference, or their relative willingness to pay for quality. 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” buyer 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.

These online labor markets have become popular settings for research. Several papers have explored buyer preferences with respect to a number of different seller dimensions (Pallais, 2013; Chan and Wang, 2017; Stanton and Thomas, 2016; Agrawal et al., 2013; Barach and Horton, 2017). Other papers have looked at changes in policy 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).

There is a growing literature in information systems 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 procure-

ment 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. [Hong and Pavlou \(2015\)](#) provide a detailed look at how differences in time-zone, language and cultural factors affect prices in online labor markets. This paper fits this pattern of focusing on some aspect of the market, but it also takes a more design-based view. Also in this “design” vein, [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. [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.

3 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 describe 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.

3.1 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).³ A guarantee could increase the number of transactions on 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 ΔR be the corresponding change in revenue from offering a guarantee, and let ΔQ be the change in the number of transactions. The platform is just indifferent to offering a guarantee if

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

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.*

3.2 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

³While 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.

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

$$\pi = p - w - (1 - p)c. \tag{2}$$

If sellers are paid their expected product, then in equilibrium

$$w = p - (1 - p)c. \tag{3}$$

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, τ , 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 $(1 - \tau)(w + (1 - p)w)$, and so

$$\begin{aligned} \Delta w\% &= 1 - p \\ &= \tau \Delta R\%. \end{aligned} \tag{4}$$

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 ϵ_w^D and sellers have a supply elasticity of ϵ_w^S . For buyers and sellers to have finite (or non-zero) elasticities, there would have to be some idiosyncratic components to Equation 2 and Equation 3. 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

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

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 1, implies offering a guarantee is profitable for the platform if

$$\left(\frac{1}{|\epsilon_w^S|} + \frac{1}{|\epsilon_w^D|} \right)^{-1} > \frac{1}{\tau}. \quad (5)$$

The condition in Equation 5 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 5 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 ϵ_w^D and ϵ_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. \(2017\)](#) 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.

3.3 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

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

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 ϵ , respectively. The buyer does not observe the platform's signal, but does observe whether the platform offers a guarantee in response to 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 $\text{MBG} = 1$ indicate that a given seller came with a money-backed guarantee. If the platform offered a guarantee, it implies that the signal it received was above its threshold, or

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

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.

4 Guaranteeing Experiment

The design of our first experiment was simple: 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. We will discuss how the platform decided which sellers to guarantee below. The control experience was the status quo on the platform prior to

the experiment.

Figure 1a shows how a collection of seller proposals look to a treated buyer, whereas Figure 1b shows how they would have looked to a buyer assigned to the control. Note the guarantee “badge” for the first two proposals in the left panel. 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.”

One implication of the experimental design is that we did not expect any equilibrium effects on w . However, as the profitability criterion for the guarantee requires some increase the quantity of transactions, looking at whether a particular buyer was more likely to contract when gaining the “full” subsidy of the guarantee is a necessary but not sufficient condition. In short, if treated buyers could not be induced to transact in the experiment, they would be even less likely to do so in equilibrium.

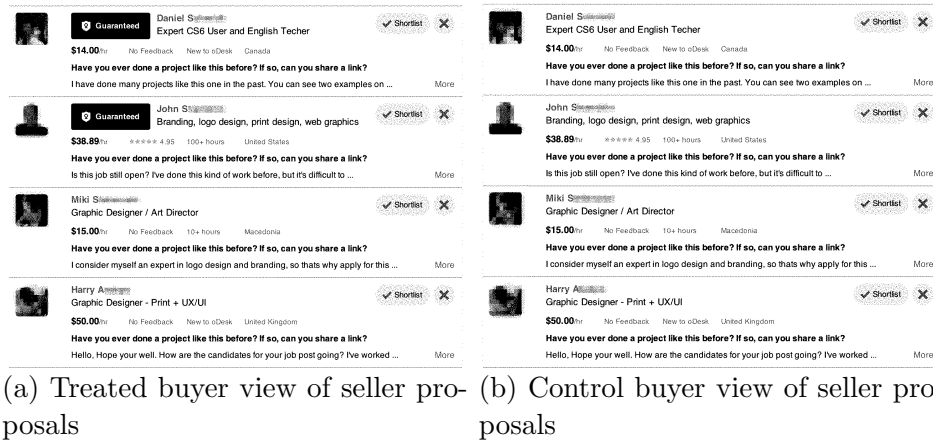
4.1 Sample definition and internal validity

The experiment began in November 2013 and ended in August 2014. A relatively small fraction of all buyer postings were assigned to the experiment to mitigate financial risk. This small allocation also reduces concerns about validity-threatening market movements (Blake and Coey, 2014). After making a hire of a guaranteed-seller, buyers had two weeks to request a refund. Information on the actual guarantee payouts is proprietary and not reported.

Our primary sample is all the buyer requests for proposals, which is also the unit of randomization.⁴ Each buyer typically received multiple proposals, and so for some research questions, we also conduct our analysis at the level

⁴Buyers 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.

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



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 figure 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.”

of the proposal. The sample is composed of 36,264 requests for proposals, which collectively received 1,051,778 proposals from 186,564 distinct sellers. To be included in the experiment, a request for proposals had to “public” in the sense that any seller could apply to it. The buyer also had to specify an hourly contract structure, where the sellers bid a wage. 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 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.⁵ As expected, the sample is well-balanced with respect to buyer and seller characteristics (see Appendix A).

4.2 The platform’s selection of sellers to guarantee

The platform had to decide, in real-time, whether to guarantee a seller. Recall Remark 5, which implied that, all else equal, the platform would 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 applied a score cut off for the guaranteeing decision (if the proposal was 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 “fit” 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 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

⁵The 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.

otherwise would not have been guaranteed.

Table 1 compares the characteristics of sellers who were eligible for the guarantee to those who were not. We can see that guarantee-eligible sellers had substantially more experience and better, more extensive, feedback. Unsurprisingly, they also charge more for their services compared to non-eligible sellers. It is clear that the algorithm was selecting the “better,” albeit more expensive, sellers.

Table 1: Mean characteristics of sellers, by guarantee eligibility

	Mean (SCORE \leq 0.5)	Mean (SCORE $>$ 0.5)	Difference in Means	p-val.
<i>Seller Attributes</i>				
Hours Worked to Date	609.13 (3.68)	984.71 (6.13)	375.58 (5.86)	<0.001 **
Num Past Jobs Worked	11.98 (0.05)	26.10 (0.15)	14.12 (0.14)	<0.001 **
Past Hourly Earnings	5,271.98 (38.65)	11,634.06 (83.83)	6,362.08 (76.75)	<0.001 **
Num Prior Relationships	9.80 (0.04)	19.84 (0.11)	10.04 (0.10)	<0.001 **
Wage Bid \$/hour	9.89 (0.05)	14.41 (0.08)	4.52 (0.07)	<0.001 **
Profile Wage \$/hour	9.81 (0.05)	14.03 (0.07)	4.23 (0.06)	<0.001 **

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: $p \leq 0.10$: †, $p \leq 0.05$: * and $p \leq .01$: **.

As we noted, the platform recorded which sellers would have been guaranteed, had they proposed to a treated buyer. 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. It is important to note that the treatment assignment associated with a buyer was not observable by sellers, and so we would expect seller applicant pools to be balanced. 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, even if all attributes were independent of each other.⁶

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	
<i>Seller Attributes</i>					
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 $p \leq .01$: **.

4.3 Effects of offering guarantees on whether the buyer contracted with anyone

Our first outcome of interest is whether the buyer contracted with any of the applying sellers. Recall from Remark 1 that for the guarantee to be worthwhile for the platform, there must be an increase in the number of

⁶This 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.

transactions. In Table 3, Column (1), we report an OLS estimate of

$$\text{CONTRACTED}_j = \beta_0 + \beta_1 \text{MBG}_j + \epsilon, \quad (6)$$

where CONTRACTED_j is an indicator for whether the buyer j spent some amount of money on one or more contracted-with sellers and MBG_j is the treatment assignment for buyer j . 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 $[-0.004, 0.016]$.

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 base contracting rate for these high-tier buyers. However, this estimate is not very precise. Assuming the “effect” is not due to sampling variation, one interpretation is that given that the sellers that high-tier buyers are interested in tend to be higher wage—and hence higher risk—a money back guarantee is simply worth more to them. Another possibility 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 judgement about the “best” applicant.

4.4 Effects of offering the guarantee on buyer selection

Despite having no overall effect on the probability a hire was made—except perhaps in the high-tier group—offering guarantees could have altered *which* sellers were contracted with because of the direct incentive effects—recall Remark 3 which stated that, all else equal, a guaranteed seller offers a higher payoff. However, recall that Remark 6 implied the same directional effect, in that a buyer might prefer a guaranteed seller because of the latent information about quality signaled by the platform’s decision to guarantee. We can not disentangle the two causes in this first experiment (it is the focus of our

Table 3: Effect of offering a guarantee on the probability the buyer forms a contract

	<i>Dependent variable:</i>	
	Job opening is filled	
	(1)	(2)
Money-back guarantee offered on select sellers, MBG	0.006 (0.005)	-0.007 (0.010)
MEDTIER		-0.022** (0.009)
HIGHTIER		-0.090*** (0.010)
MBG \times MEDTIER		0.008 (0.012)
MBG \times HIGHTIER		0.041*** (0.015)
Constant	0.362*** (0.004)	0.394*** (0.007)
Observations	36,264	36,264

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 contracting with 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: $p \leq 0.10$: †, $p \leq 0.05$: * and $p \leq .01$: **.

second experiment). For now, we set aside this question and simply explore the buyer selection among sellers.

We start with the question of whether being guaranteed helped a seller be selected by the buyer for a contract. We exploit the fact that sellers apply to multiple requests for proposals, and so we have within-seller variation in whether or not they receive a guarantee for a particular application. The ability to do this highlights an advantage of our empirical setting in which the whole buyer consideration set is observed, in that we can switch our analysis from the level of the buyer to the level of the individual seller applicant.

To begin, in Figure 2, we plot the application mean success rate by score “band,” by the treatment status of the applied-to buyer. We can see that for low score bands, the treatment and control have similar rates. As we near the threshold (but are still below it), in the treatment, sellers do slightly worse. As these sellers are more likely to be substitutes to slightly higher score sellers, we can see some evidence of a crowd-out effects. Above the threshold, we can see that sellers in the treatment do better. For example, in the 0.9 and above threshold, the effect in levels is about 0.01, which is about 25% better.

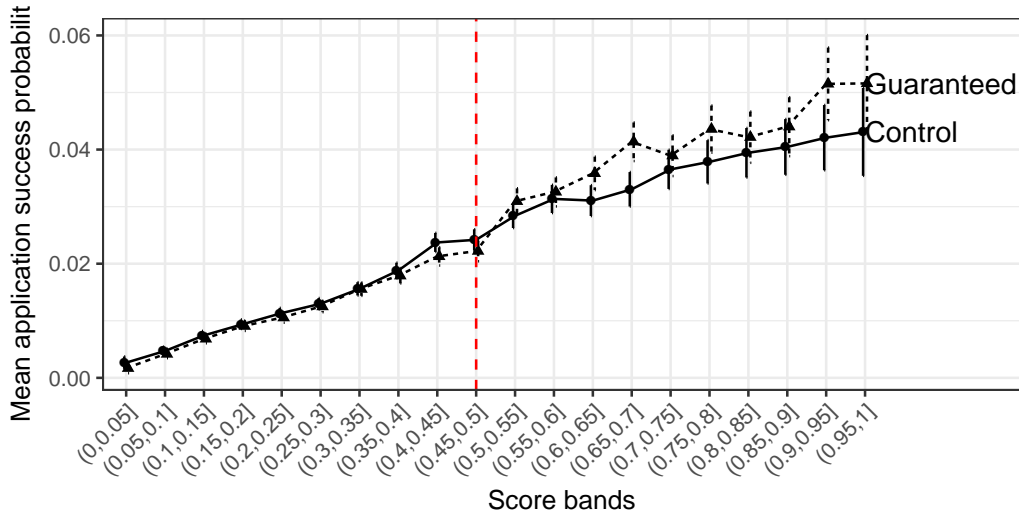
Now we switch to a regression framework. As sellers cannot condition on the treatment assignment of the buyer, guarantee eligibility can be treated as exogenous when a seller-specific fixed effect is included. In Column (1) of Table 4, we report an OLS⁷ estimate of

$$\begin{aligned} \text{CONTRACTED}_{ij} = & \alpha_i + \beta_1 \cdot \mathbf{1}\{\text{SCORE}_{ij} > 0.5\} + \beta_2 \text{MBG}_j \\ & + \beta_3 (\text{MBG}_j \times \mathbf{1}\{\text{SCORE}_{ij} > 0.5\}) + \epsilon_i, \end{aligned} \quad (7)$$

where CONTRACTED_{ij} is an indicator for whether seller i was contracted with by buyer j , α_i is a seller-specific fixed effect, $\mathbf{1}\{\text{SCORE}_{ij} > 0.5\}$ is an indicator whether the applying seller has a score higher than the cut-off to

⁷We use the linear probability model because our interest is primarily in marginal effects.

Figure 2: Mean application success probability by applicant 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.

receive the guarantee, and 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.

Starting in Column (1) of Table 4, we can see that having a score above the 0.5 threshold helps, even in the control, as $\hat{\beta}_1$ —the coefficient on the above-the-threshold indicator—is positive and highly significant. Given the baseline probability of selection, the coefficient implies that being above the threshold raises a seller’s probability of forming a contract by 27%, though given that sellers with higher scores have a higher success probability, this percentage estimate is somewhat of an over-estimate. It might seem surprising that the score would matter with the inclusion of a seller-specific fixed effect, but recall that the score is computed on a per-opening basis and reflects the seller’s “fit” for that particular buyer, as well as the proposed wage. As such,

Table 4: Effects of offering a guarantee on the buyer selection process

	<i>Dependent variable:</i>			
	Contracted? (1)	$\mathbf{1}\{\text{SCORE} > 0.5\}$ (2)	$\mathbf{1}\{\text{SCORE} > 0.5\}$ (3)	SCORE (4)
$\mathbf{1}\{\text{SCORE} > 0.5\}$	0.005** (0.001)			
MBG of Applied-to buyer	-0.001* (0.0003)	0.049** (0.009)	0.043** (0.016)	0.018** (0.004)
$\mathbf{1}\{\text{SCORE} > 0.5\} \times \text{MBG}$	0.005** (0.001)			
MEDTIER			0.035** (0.013)	
HIGHTIER			0.078** (0.023)	
MBG \times MEDTIER			0.004 (0.020)	
MBG \times HIGHTIER			0.018 (0.029)	
Constant		0.444** (0.006)	0.412** (0.011)	0.463** (0.003)
Applicant sample	All	Contracted	Contracted	Contracted
Seller FE	Y	N	N	N
Observations	1,051,508	19,000	19,000	19,000

Notes: The table reports regressions where unit of analysis is the proposal submitted by sellers. In Column (1), the dependent variable is an indicator for whether the buyer contracted with that particular seller for that application. The sample consists of all applications submitted to buyers assigned to the experiment. $\mathbf{1}\{\text{SCORE} > 0.5\}$ is an indicator for whether the applying seller had a score greater than 0.5, hence making them eligible for a guarantee if applying to a treated buyer. If the buyer contracted with a guaranteed seller in the treatment, the platform would reimburse the buyer for the first two weeks of any contract if the buyer requested a refund. In Columns (2) and (3), the sample is only those sellers who were hired, and the outcome is an indicator for whether that seller had a score above the 0.5 threshold. In Column (3), 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. In Column (4), the sample is the same as in (3) and (4), but the outcome is simply the score. Significance indicators: $p \leq 0.10$: †, $p \leq 0.05$: * and $p \leq .01$: **.

a higher score should be correlated with being more likely to be contracted with, even if getting over this threshold is not visible to buyers.

Our main coefficient of interest is the interaction term, β_3 , which measures whether being guaranteed helps. We can see that receiving the guarantee-badge when applying to treated buyers enhances the boost in contracting probability to exceeding the guarantee threshold—the marginal effect nearly doubles.

Not all sellers applying are above the threshold, and we might imagine that these sellers suffer from the comparison to their badged fellow applicants. The coefficient on MBG, β_2 , measures this crowd-out effect—a common concern in hiring settings (Crépon et al., 2013). The negative sign of $\hat{\beta}_2$ implies that sellers submitting a proposal to a buyer where a guarantee was available for some sellers, but they themselves were not guaranteed had a lower chance of being selected. However, the effect is not that large—given the baseline success probability for a proposal (pooled over all observations), this effect amounts to about a 4% reduction on a per-application basis.

Given that the treatment clearly helped sellers with higher scores get selected, a natural question is whether this effect manifested itself in the composition of contracted-with sellers. For the rest of the regressions in Table 4, the sample is restricted only to those contracted-with-sellers. We cluster at the level of the individual buyer (as some buyers do make multiple hires per request), instead of the level of the seller. While this is a selected sample, recall that there was little evidence that the treatment changed the probability a contract was formed.

In Column (2), the outcome is whether a contracted-with-seller had a score above the 0.5 threshold, with the sample restricted to only those who formed a contract. We can see that for buyers assigned to the treatment, sellers with a score above the threshold are over-represented—there are about 10% more of them. This is just a manifestation of the selection effect observed in Column (1).

As different buyers might be more or less interested in the guaranteed sellers, in Column (3) we interact the treatment indicator with the vertical preference tiers. From the coefficients on the tier indicators, we can see that high-tier buyers were more likely to hire sellers with higher scores. The interaction effects, however, are quite small and not conventionally significant, implying that the treatment did not have heterogeneous effects as measured by buyer tier.

In Column (4), we use the actual score as the outcome rather than whether the cutoff is exceeded. We can see, as expected, that sellers contracted with by treated buyers had higher scores. Together, these results imply that guaranteed sellers were favored by buyers and this resulted in treated employers disproportionately selecting seller with higher scores.

4.5 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—recall Remark 4. In particular, they might be less price sensitive when considering guaranteed sellers. Consider a regression of the form

$$\text{CONTRACTED}_{ij} = \alpha_i + \gamma_j + \beta_1 \log w_{ij} + \beta_2 (\log w_{ij} \times \text{MBG}_j) + \epsilon, \quad (8)$$

where w_{ij} is seller i 's wage bid to buyer j , and α_i and γ_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 5 reports a regression similar to Equation 8, 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. In Column (2), we add the interaction term. Somewhat surprisingly, the coefficient is close to zero (and the “wrong” sign), contrary to our moral hazard conjecture. Furthermore, this a precisely estimated zero, with a 95% CI of $[-0.006, 0.001]$. In short, there is no evidence of moral hazard in selection, at least as measured by the price sensitivity of buyers.

Table 5: Effect of seller wage bid on the probability of the buyer contracting with that seller, by treatment assignment

	<i>Dependent variable:</i>		
	Contract formed?		
	(1)	(2)	(3)
Log wage bid	-0.031*** (0.002)	-0.029*** (0.003)	0.002*** (0.0002)
Log wage bid \times 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: $p \leq 0.10$: †, $p \leq 0.05$: * and $p \leq .01$: **.

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.

4.6 Project outcomes

The primary goal of the experiment was to increase buyer willingness to form a contract—or to increase Q in the language of the model. As the platform guaranteed “better” (albeit more expensive) applicants, 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 changes in project outcomes. In Table 6, 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. 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.

Across the various outcome measures, the effect of being offered the guarantee appears to be a fairly precisely estimated zero. 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, it is unclear that buyer surplus would be any higher, given that these better sellers are more expensive.

Another measure of direct interest to the platform is whether the buyer posted a subsequent request for proposals—another way to increase Q . There was a small increase in the probability a buyer in the treatment posted a subsequent request—0.46%, but the standard error for that measure is 4%, making it quite likely that the increase was due to sampling variation.

Table 6: Effects of guaranteeing sellers on buyer contract outcome measures

	<i>Dependent variable:</i>		
	Public Feedback	Private Feedback	Rehired?
	(1)	(2)	(3)
MBG	0.002 (0.016)	0.036 (0.049)	0.003 (0.005)
Constant	4.727*** (0.012)	8.876*** (0.036)	0.140*** (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: $p \leq 0.10$: †, $p \leq 0.05$: * and $p \leq .01$: **.

5 Guaranteeing versus Recommending Experiment

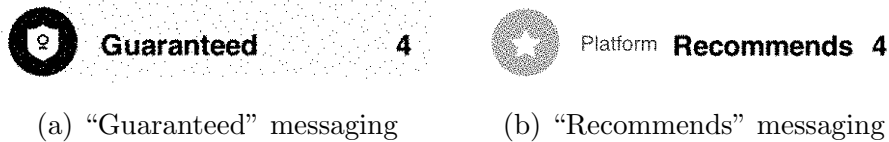
Although the first experiment showed no evidence that the quantity of transactions could be appreciably increased with a guarantee, it still could be useful to shift buyers towards platform-preferred sellers. It could be useful both for [Nosko and Tadelis \(2015\)](#) platform spill-over reasons, as well as a way to overcome over-reliance on experienced sellers ([Pallais, 2013](#); [Horton, Forthcoming](#)).

Following the first experiment, a second experiment was conducted to see if the same shifting of buyer attention could be obtained without an explicit financial guarantee. 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 3](#) shows the messages seen by buyers in the “guarantee” and “recommends” groups, respectively. 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.⁸

Figure 3: Comparison of the messaging presented about sellers in the (a) “guarantee” and (b) “recommend” experimental groups



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 badging was 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.

5.1 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 applications from 90,818 distinct sellers. As expected, pre-randomization attributes are well balanced, as are the characteristics of applicants (see Appendix A).

⁸This 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.

5.2 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 7, the dependent variable is an indicator for whether the buyer formed a contract. The model is fit using OLS, 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 difference in the probability a contract was formed. In Column (2), we test whether guarantee versus the recommendation worked differently for different kinds of buyers. Recall from Table 3 that buyers in the high-tier seemed to be more likely to hire when offered guaranteed sellers. Here, we find no appreciable difference, indicating that even in the high tier, guaranteeing sellers seemed to have the same effect on a contract being formed as simply recommending sellers (which is none).

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

Now we turn to the effects of the treatment on buyer selection. 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, we can simply plot the mean application success probability by score band, by treatment assignment. Figure 4 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.

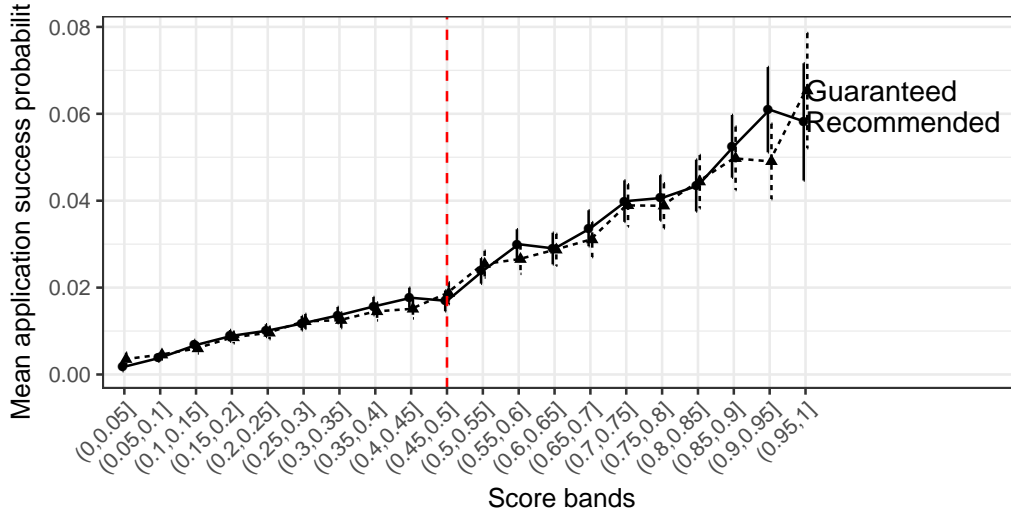
Moving to a regression framework, in Table 8, Column (1), we report

Table 7: Effect of offering a guarantee versus simply recommending an applicant on the probability a buyer forms a contract

	<i>Dependent variable:</i>	
	Contract formed	
	(1)	(2)
Money-back guarantee offered on select sellers, MBG	-0.005 (0.009)	-0.006 (0.016)
MEDTIER		-0.021 (0.014)
HIGHTIER		-0.044*** (0.017)
MBG \times MEDTIER		0.017 (0.020)
MBG \times HIGHTIER		-0.030 (0.024)
Constant	0.395*** (0.006)	0.415*** (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: $p \leq 0.10$: †, $p \leq 0.05$: * and $p \leq .01$: **.

Figure 4: 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.

an application-level regression where the outcome is whether the applicant was contracted with. This regression is identical to the one based on Equation 7, except for the interpretation. Now an applicant with a score above the threshold was marked as “recommended” whereas in the first experiment, they had no special indicator.

We can see that having a score above the threshold, $\mathbf{1}\{\text{SCORE} > 0.5\}$, increases hire-probability by 0.029. This effect is now substantially larger relative to the first experiment—3x larger. Presumably this is due to both the “recommended” badge that is now shown for such candidates, and the new interface to only show high score applicants by default.

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. The 95% CI is $[-0.003, 0.001]$, which means that from a seller’s perspective, in terms of being contracted-with, a recommendation and a guarantee have

Table 8: Effects of the platform guaranteeing versus recommending on contracting probability and the characteristics of contracted sellers

	<i>Dependent variable:</i>		
	Contracted?	$\mathbf{1}\{\text{SCORE} > 0.5\}$	SCORE
	(1)	(2)	(3)
$\mathbf{1}\{\text{SCORE} > 0.5\}$	0.013*** (0.001)		
MBG of the applied-to opening	-0.001 (0.0005)	0.00004 (0.015)	0.001 (0.007)
$\mathbf{1}\{\text{SCORE} > 0.5\} \times \text{MBG}$	-0.0005 (0.002)		
Constant		0.586*** (0.010)	0.542*** (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. $\mathbf{1}\{\text{SCORE} > 0.5\}$ 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: $p \leq 0.10$: †, $p \leq 0.05$: * and $p \leq .01$: **.

essentially the same effect. 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.

The total lift that guaranteed applicants above the threshold got in the first experiment was $\hat{\beta}_1 + \hat{\beta}_2 \approx 0.014$. This is still lower than the lift found in the second experiment. The difference is likely attributable to the new default view in the second experiment, which was to only show high score, guarantee-eligible applicants initially. The new design already benefitted these applicants, so the marginal effects had to be smaller, given that there was not overall increase in the probability a contract was formed.

In Column (2), we add a seller-specific fixed effect, which we can do because of the fact that sellers submit multiple proposals. There is no appreciable change in the results, except that the effect of being above the threshold declines, which is unsurprising if higher-score sellers are simply more likely to be contracted with. In Column (3), we restrict the sample to only those sellers who were contracted with, and use their score as the outcome. Unsurprisingly given the lack of evidence for differential selection in Columns (1) and (2), we see no evidence of a difference in the scores of contracted with sellers. This stands in sharp contrast with the first experiment.

6 Discussion and conclusion

The main finding of the experiment 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. Consistent with the financial component not mattering, (1) there is no overall increase in sales, and (2) there is no evidence that buyers use the guarantee to hire more expensive sellers i.e., there is no evidence of a moral hazard in selection. The second experiment confirms this “informational” view of the guarantee, showing that merely recommending a seller had essentially the

same effect as guaranteeing. 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. In short, skin in the game did not matter. Interestingly, this matches the findings of [Panniello et al. \(2017\)](#), who find that adapting recommendations to reflect the margins for different products—in a sense, skin-in-the-game, but in the “wrong” direction from the buyer’s perspective—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 spill-over 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 generally 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 and whether this hit is “worth” it. [Hagi 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.

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.⁹ Having sellers partially pay for refunds would help overcome the challenging economics of guarantees highlighted by Equation 5. It is also an interesting question as to whether would-be buyers would view seller-provided guarantees differently. Seller-backed guarantees might signal 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.

⁹eBay 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.

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, 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.
- 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.
- Crépon, Bruno, Esther Duflo, Marc Gurgand, Roland Rathelot, and Philippe Zamora**, “Do labor market policies have displacement effects? Evidence from a clustered randomized experiment,” *The Quarterly Journal of Economics*, 2013, *128* (2), 531–580.

- Cusumano, Michael**, “Technology strategy and management: The evolution of platform thinking,” *Communications of the ACM*, 2010, *53* (1), 32–34.
- 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 in the Long-Run,” *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.

- 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.
- Knoepfle, Dan, Jonathan V. Hall, and John J. Horton**, “Labor Market Equilibration: Evidence from Uber,” *Working Paper*, 2017.
- 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 Internal validity

Table 9 shows the balance table for the first experiment; Table 10 shows the balance table for the second experiment. In both cases, the randomization was effective, and the treatment and control groups are well balanced.

Table 9: Balance for Experiment I

	Control Mean: \bar{X}_{CTL}	Treatment Mean: \bar{X}_{TRT}	Difference In Means	p-value
<i>Employer Attributes</i>				
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
<i>Job Posting Attributes</i>				
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 Experience 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
<i>Applicant Attributes</i>				
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: $p \leq 0.10$: †, $p \leq 0.05$: * and $p \leq .01$: **

Table 10: Balance for Experiment 2: Guaranteed vs. Recommended

	Control Mean: \bar{X}_{CTL}	Treatment Mean: \bar{X}_{TRT}	Difference In Means	p-value
<i>Employer Attributes</i>				
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
<i>Job Posting Attributes</i>				
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
Preferred Experience 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
<i>Applicant Attributes</i>				
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 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: $p \leq 0.10$: †, $p \leq 0.05$: * and $p \leq .01$: **.