Matching in Online Marketplaces when Talent Is Difficult to Discern*

Elena Belavina, Karan Girotra
SC Johnson College of Business, Cornell University, Ithaca, NY 14850 and New York, NY 10044
elena@belavina.com, karan@girotra.com

Ken Moon, Jiding Zhang
The Wharton School, University of Pennsylvania, Philadelphia, PA 19104
{kenmoon, jiding}@wharton.upenn.edu

We study the problem of assigning workers to short-term jobs in online marketplaces. In settings where workers' most relevant skills and attributes are readily collected at scale (e.g., drivers' ratings and locations on Uber), the marketplace benefits by prioritizing the matches of workers supplying the best quality and most compatible attributes. However, in many important settings, workers are distinguished by skills and attributes that market participants learn privately upon interacting. In collaboration with a major online labor platform matching freelancers into millions of transactions per year, we structurally estimate market-level demand preferences for publicly and privately observed worker attributes. We propose a novel, choice-based estimation methodology accommodating an asymptotically large number of uncertain quality freelancers mixing into hiring consideration sets that are each essentially unique rather than one of large N instances. That is, no single pair of freelancers is directly compared for hire asymptotically many times, and the rate of learning about quality lags behind the arrival rate of freelancers new to the platform. Studying 1.2 million job postings receiving 29 million applications in two highly active marketplaces on our partner's platform, we find that the majority of variation in hiring choices attributes to buyers' privately sourced information rather than observable skills compatibilities or ratings reputation. To enhance buyer value, the platform optimizes how markets balance hiring flexibility (exploration of new matches) against repeat-dealing relationships (exploitation of known matches), and we study how the intermediary should encourage market participants who generate and consume private information to alternately explore and exploit.

Key words: Empirical operations management; Freelance labor; Market intermediaries; Marketplace design; Matching with costly inspection; People operations

---

* We gratefully acknowledge support from our data sponsor and from the Wharton School, University of Pennsylvania (Fishman-Davidson Center). Seminar participants at the 2017-2019 MSOM conferences, 2017-2019 INFORMS Annual Meetings, 2018 INFORMS Revenue Management and Pricing conference, 2019 5^{th} Marketplace Innovation Workshop, 2019 Wharton Empirical OM Workshop, 2019 Cornell ORIE Young Researchers Workshop, and Wharton OID seminar provided valuable feedback.

1. Introduction

Spot labor markets operated by online platforms as intermediaries have grown drastically over the past decade, with market leader Upwork's portal alone in 2014 exceeding USD $940M (2.8M jobs posted) in global transactional volume and USD $600M (1.5M jobs posted) in the US.[1] Rather than crowding out existing hiring, the online economy's reduced costs of search and recruiting have been credited with filling positions that would not have been filled otherwise, yielding substantial welfare benefits (Horton (2017)). In these increasingly important marketplaces, managers are tasked with designing matching systems that sort and allocate the right workers, in terms of skills, quality, and fit, to the right jobs.

However, matching workers well is often complex. Workers specialize and thereby accrue skills and knowledge determining their compatible jobs. Search frictions impede matches in such horizontally differentiated markets (Bakos (1997), Kanoria and Saban (2017), Bimpikis et al. (2018), Cullen and Farronato (2018)) and can even outweigh the benefits of offering more abundant options as a marketplace grows (Li and Netessine (2018)). Consequently, substantial search-related effort and activity are observed in many online markets (Fradkin (2015), Arnosti et al. (2018), Li and Netessine (2018)).[2] Additionally, workers similarly compatible with given jobs may vary in their performance and the quality of work they produce. As a result, participants in online marketplaces respond significantly to information about counterparty quality and reputation (Moreno and Terwiesch (2014), Pallais (2014), Nosko and Tadelis (2015), Hui et al. (2016), Tadelis (2016), Garg and Johari (2018a), Cui et al. (2019)).

The ease of search and of facilitating search importantly depends on how easily information about worker compatibility and quality is ascertained and by whom. At one extreme are ride-hailing platforms like Uber: because the platform readily collects drivers' most relevant skills and attributes (real-time ratings and locations) at scale using mobile technologies, the intermediary itself expeditiously directs market participants into matched transactions. Existing literature closely studies such settings (Banerjee et al. (2015), Afeche et al. (2018), Besbes et al. (2018), Chen and Hu (2018), Bimpikis et al. (2019)).

Instead, on the platform this paper studies, workers are distinguished by skills and attributes that market participants learn as private information upon interacting. Two stylized facts are important to our study. First, significant information about freelancers remains private among buyers, because reputation systems fail to fully elicit market participants' valuable information about their counterparties' skills and attributes (Pallais (2014), Filippas et al. (2017)). Notably, with

[1] Upwork (2017), "Online Work Report," available at: `http://elance-odesk.com/online-work-report-global`.

[2] E.g., on Airbnb in 2014, travelers transmitting a booking request spent a median 58 minutes viewing 73 unique listings, yet prospective hosts rejected 42% of such contacts (Fradkin (2015)).

implications for platform matching, marketplace intermediaries cannot readily observe important determinants of match quality and fit. Second, knowledge about job applicants is produced at cost by the buyers themselves, who expend time and effort gathering information using interviews (*by screening*) and from prior dealings with previously hired applicants (*by transacting*). Thus, whereas the platform might encourage or incentivize buyers to create or exploit private information, it can neither easily obtain nor handle the involved knowledge directly.

Intermediaries operating such markets encounter distinctive challenges of academic interest. First, when information required to productively match workers to jobs is missing from visible covariates, including ratings, it must be inferred from buyers' revealed preferences, i.e., whom they hire rather than ostensibly why. Corroboratingly, our industry partner recently began de-emphasizing skills tests and instead screens freelancers for continued platform access based on how frequently they are hired after applying for jobs. Second, because identifying suitable matches entails costly inspection and direct experience, matches can become sub-optimally sticky. For example, in the annually $400-800 billion market for labor-intensive home services (e.g., home cleaning or childcare), customers flatly resist flexibly substituting across on-platform providers.[3] Such reactions can derail the platform's economies of scale that could otherwise leverage a deeper supplier pool to more closely fit buyers' needs (e.g., at specified times or on demand). More broadly, the efficacy of search and match significantly impacts the platform's revenues, job fill rates, and effective availability of supply (Kanoria and Saban (2017), Bimpikis et al. (2018), Cullen and Farronato (2018), Li and Netessine (2018)).

Thus, this paper addresses the question of how to match workers to jobs in online marketplaces, when workers possess difficult-to-verify skills and competencies. Leveraging a dataset tracking two large online markets' 1.2 million freelance job postings receiving 29 million applications and resulting in 744 thousand hires, we study two questions: (1) Can scalably data-driven methods evaluate the impact of buyers' private information about prospective hires' quality and fit on a market's hiring outcomes? (2) When buyers source and retain such information privately, what are the prescriptive implications for the platform's matching of marketplace participants? Should platforms encourage buyer-freelancer relationships, and if so, in which markets and against what trade-off with intermediary-enabled hiring flexibility?

Our methodological contribution develops a transparent, data-driven framework to recover demand preferences for publicly and privately observed worker attributes from large-scale data on

[3] See, e.g., Madden, Sam (2015), "Why Homejoy Failed . . . and the Future of the On-demand Economy", TechCrunch, July 31, available at: https://techcrunch.com/2015/07/31/why-homejoy-failed-and-the-future-of-the-on-demand-economy/: "when [clients] find those trusted, quality pros, the last thing [they] want to do is lose them. These client-pro relationships can last decades." See also Stout, Hillary (2015), "Amazon, Google and More Are Drawn to Home Services Market", New York Times, April 12, available at: https://www.nytimes.com/2015/04/13/technology/amazon-google-and-more-are-drawn-to-home-services-market.html.

hiring decisions. We apply this novel methodology to estimate demand preferences in two highly active marketplaces coordinating freelancers seeking data entry and web development jobs, respectively. To control for horizontal differentiation in each market, estimation accounts for over 2,000 category-specific skills comprehensively identified and tagged for both freelancers and jobs and also accommodates workers offering idiosyncratic fit for specific jobs. For vertical differentiation, special attention is required when buyers observe freelancer quality signals not captured by the platform's data. In standard settings, such unobserved quality can be captured through freelancer fixed effects when the data's number of observed hiring choices N grows while the assortment remains reasonably stable. In contrast, our markets' numbers of freelancers grows asymptotically along with the data size, and they sort into hiring consideration sets offering essentially unique assortments ($N = 1$). Consequently, no single pair of freelancers is directly compared for hire asymptotically many times. Moreover, the rate at which new freelancers enter the market asymptotically dominates the platform's rate of learning freelancer quality by observing who is hired over others, resulting in significantly biased, conventional fixed-effect estimators from panel econometrics.

Thus, our choice-based estimation methodology novelly accommodates an asymptotically large number of horizontally and vertically differentiated freelancers mixing into hiring consideration sets that are each essentially unique. Instead of relying on observing large N hiring choices made from a small number of stable assortments, we introduce a method that depends on observing a market's asymptotically many hiring switches, where a switch is a pair of freelancers being hired over one another in separate hiring instances. By aggregating switches to explain why they happen, we consistently bound buyer preferences using random set inference (Chernozhukov et al. (2007), Romano and Shaikh (2008, 2010), Bugni (2010), Ho and Pakes (2014), Pakes et al. (2015)). Using moment inequalities to bound random sets of empirically valid preferences, the computationally scalable methods we develop could be used by practitioners as data-driven, diagnostic tools.

1.1. Related Literature

Our work relates to several streams of literature at the intersection of operations management and marketplace analytics. First, it contributes to the study of preferences and incentives in online markets and platforms (e.g., Cachon et al. (2017) and Hu et al. (2018)) and the sizeable empirical literature applying structural estimation to marketplace analytics (e.g., Allon et al. (2011), Kim et al. (2014), Bimpikis et al. (2018), Feldman et al. (2018), Moon et al. (2018), Zheng et al. (2018), Ata et al. (2019), Buchholz (2019), Kabra et al. (2019)). To our knowledge, our empirical work is among the first to study the implications for market design when decentralized marketplace participants collect private information about online workers for hire. Moreover, we propose structural methods novel to this space.[4]

Our work also relates to studies of information in online markets, including the prior work regarding online search and reputation systems described above. Relevantly, accreting evidence (e.g., Pallais (2014), Filippas et al. (2017)) indicates that online spot labor markets' reputational systems omit economically valuable information about freelance workers that is learned by the marketplace buyers who hire them. Intervening in a spot marketplace for data entry labor where only 17% of ratings fall below four stars, Pallais (2014)'s field experiment demonstrates that the platform's reputational reviews are significantly less informative than straightforward reviews created from job performance (e.g., quantiles in speed, accuracy, and timeliness). Well-performing freelancers randomly assigned to receive such reviews averaged additional six-month earnings of $60 over their matched peers receiving typical reviews as a control group. Focusing on new marketplace participants joining the platform, Stanton and Thomas (2016) and Johari et al. (2018) focus on learning their unknown attributes. See also Altonji and Pierret (2001). In contrast, our study focuses on markets where information about workers is persistently private over their tenures, and indeed Pallais (2014) finds that the field experiment's intervention has a larger treatment effect on *experienced* freelancers.

More generally, a literature stemming from Autor (2001) argues that while technology has steadily diminished the costs of gathering information about workers' skills and attributes, many employers value traits that remain difficult to measure accurately at scale. Relevant empirical evidence (Autor et al. (2003), Bartel et al. (2007)) indicates that decades-long technological shifts affecting nearly all industries have increased the share of workers' non-routine tasks demanding softer skills such as adaptivity and effective communication. See also Cappelli (1995) and Cappelli and Neumark (2004). Online markets for products face similar matching challenges, where uncertain fit and quality lead to product returns (e.g., Anderson et al. (2009), Bandi et al. (2017), Gallino and Moreno (2018)) and reliance on branding (e.g., Degeratu et al. (2000)).

Finally, our work ties to previous research regarding the role of intermediaries in marketplaces (Allon et al. (2012, 2017), Stanton and Thomas (2016)) and operations management (Taylor and Plambeck (2007a,b), Belavina and Girotra (2012)). Whereas intermediaries valuably allow buyers flexibly match with platform service providers by enabling communication and policing reputations, buyers possessing costly private information may benefit from relationships with trusted

[4] Hagedorn et al. (2017) and Garg and Johari (2018b) similarly utilize revealed preference comparisons, in the context of inferring underlying preference rankings. Although focused on vertically differentiated rankings, Hagedorn et al. (2017)'s identification strategy bears similarities to our own.

providers. We focus on the intermediary's role in influencing how marketplace participants explore new partners and exploit trusted relationships. Exploration and exploitation trade-offs have a long history in the operations literature (Lai and Robbins (1985), Graves and Lai (1997), Agrawal and Goyal (2013), Russo and Van Roy (2014), Bastani and Bayati (2019)), with many applications closer to marketplaces (Feng et al. (2018), Zhang and Jasin (2018), Chen et al. (2019)). In contrast to this literature, we examine a setting in which decentralized market participants must actively collect their own private information through costly but possibly incentivized inspection.

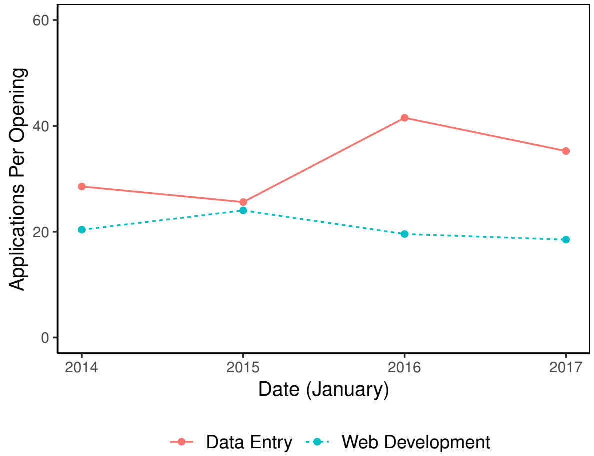
2. Background and Data

2.1. Data Sponsor and Platform

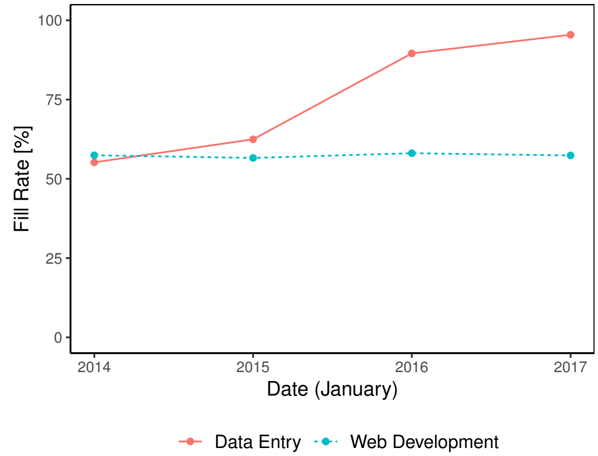
Our dataset is obtained through collaboration with an intermediary that owns and operates a global platform marketplace for freelance labor. By organizing and operating online markets for nearly ninety different categories of professional services, such as web development, graphical design, and data entry, the firm matches buyers with freelancers that provide these services.

The platform's largest categories are Web Development (356K openings in 2016 alone), Graphical Design (184K), Article & Blog Writing (104K), Translation (95K), and Data Entry (70K). In this study, we focus our analysis on the Web Development and Data Entry categories for two reasons. First, both categories offer substantial transaction volumes for our analysis. Second, based on the informal intuitions expressed by the managers with whom we collaborated, Data Entry and Web Development jobs present two extremes in terms of the effectiveness of matching workers using observable skills. Data Entry jobs consist of simple, well-defined tasks, and the basic skills required (e.g., familiarity with Excel software) are readily verified and documented at scale. On the other hand, Web Development jobs can entail substantially complex planning, adaptive resourcefulness, and motivation. Thus, its jobs require not only a multitude of technical skills (e.g., facility with different programming frameworks), but also softer skills, such as understanding the buyers' needs, recognizing brand values, and working collaboratively, that are difficult to measure without effort and attention. From these categories, our analysis intends to draw high-level, prescriptive insights applicable to other service categories that lie between these on the spectrum.

Our data sponsor granted us access to 2.5 years of transaction records spanning from January 1, 2014 to June 30, 2017. The marketplace relies on two basic activities: freelancers' applications and buyers' hires. In Figure 2, we plot the key metrics of the number of applications per opening and the fill rates at which jobs are assigned hires. In the Web Development category, these metrics hold steady over the study's duration. The Data Entry category exhibits concurrent increases in applications per opening and fill rates, which are explained by a decline in the number of jobs posted in the category. Our empirical method will naturally control for variations, like this one, in



(a) number of applications per opening



(b) fill rate

Our sample covers all records from January 1, 2014 to June 30, 2017. Over this 2.5-year period, the Data Entry (Web Development) category records 260K (978K) openings, with 9.3M (19.8M) applications and 184K (560K) hires. The plot shows the fill rates and numbers of applications per opening in January 2014, 2015, 2016 and 2017.

Figure 1 Marketplace Activity

relative marketplace supply and demand. In level, Data Entry jobs' fill rates are higher, and their applicants face fiercer competition in number.

Our empirical analysis and data variables focus on the hiring decisions that buyers make from their available consideration sets of freelancers. For background, these consideration sets are formed as follows. After registering on the marketplace, buyers post jobs, which include a description, requirements in terms of time, skills, and location, and other details. Freelancers are then invited to apply to the job openings. In addition to the applying freelancers' bids for their hourly wages, the buyer sees their credentials, including their self-reported skills, work experience, past ratings, and personal information (location, biography). After interviewing candidates and negotiating wage bids, each buyer chooses a freelancer from the consideration set of applicants or leaves the position unfilled. Upon the job's completion, the freelancer is paid, and the buyer rates her satisfaction with the freelancer's job performance. The market operator receives a commission based on the hired freelancer's total earnings from the job.

2.2. Theoretical Choice Framework

The online labor marketplace is differentiated both horizontally (capturing the *fit* between a job opening and job applicants) and vertically (capturing job applicants' *quality*). Both aspects drive buyers' choices under our model of hiring decisions.

For horizontally differentiated fit, our modeling framework accommodates information about the compatibility of freelancers' skills supplied to different jobs. Freelancers report a list of skills they possess, and buyers view these lists on applicant profiles. Similarly, when posting openings buyers list their jobs' desired skills using the same category-specific skills catalogue. In addition to workers offering compatible skills, our model allows them to offer idiosyncratic fit for different jobs. While buyers can observe such idiosyncratic fit, neither the platform nor we the researchers can.

Buyers learn about freelancers' vertically differentiated quality through several avenues. First, on the platform we study, buyers intensively inspect and screen job applicants by administering interviews and tests. Second, buyers may learn freelancer quality by transacting, making their knowledge from prior dealings valuable when previous hires apply to new positions. Lastly, through the intermediary's reputation system, buyers may gather additional information about applicants' quality gleaned from *other* buyers' prior dealings with them. While freelancers' ratings and past dealings are available in our data, we lack direct data regarding the knowledge buyers generate by inspecting and screening applicants to their jobs. Because such buyer-applicant interactions transpire mainly through third-party applications (e.g., Skype-hosted interviews), such data are also difficult for the platform to secure.

In what follows, we discuss the information observed in our data set and construct the variables used later in our hiring choice model.

2.3. Data Variables

2.3.1. Horizontal differentiation. As the primary determinant of horizontally differentiated fit between workers and jobs, we observe skills in the marketplace data.

Skills. The main function of the online marketplace and its intermediary is to provide matches, and a key advantage of operating large-scale markets online is that it allows for closer matches between the skills required by jobs and those possessed by freelancers. Buyers hire flexibly from freelancers' diverse pool of skills, and we model the resulting compatibility of fit between the workers and jobs.

For a sense of the value generated by matching freelancers to jobs with compatible skills, we characterize the empirical distributions of the variety in skills demanded by jobs and those supplied by freelancers. In the Data Entry (Web Development) category, a total of 728 (1,260) distinct skills are demanded, and 2,061 (2,283) distinct skills are supplied. Moreover, individual buyers continually shift the variety of skills demanded across their posted jobs, such that consecutively posted jobs average only 80.7% (64.1%) of skills kept in common. Such variety suggests that substantial value is generated by matching freelancers with skills compatible to the jobs they consider, and the efficiency of matching on such skills may importantly drive the volume of marketplace hiring. The skills most popularly demanded and supplied in our focal categories are listed in Table 1.

Demand		Supply	
Skill Name	Proportion	Skill Name	Proportion
Data Entry	76.6%	Data Entry	60.7%
Microsoft Excel	40.1%	Microsoft Excel	44.8%
Internet Research	22.7%	Internet Research	42.5%
Data Mining	8.1%	Microsoft Word	29.8%
Administrative Support	7.4%	Virtual Assistant	23.3%
PHP	38.1%	PHP	52.6%
JavaScript	29.9%	WordPress	48.5%
WordPress	28.0%	JavaScript	40.1%
CSS	23.6%	jQuery	36.0%
HTML5	23.3%	HTML5	35.6%

Unshaded lines show skills in the Data Entry category. Shaded lines show skills in the Web Development category. Demand means a skill requested by an opening. Supply means a skill is provided by a freelancer. Proportion denotes the percentage of openings requiring a skill, or the percentage of freelancers who claim to have that skill.

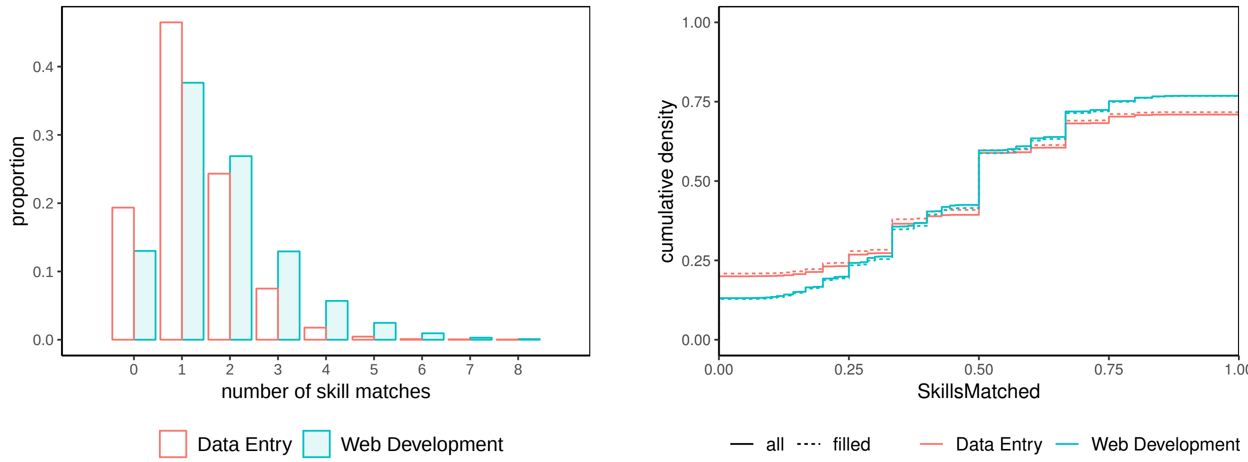
Table 1 Popular Skills

To capture an applicant's skills compatibility for a prospective job, we define the variable SkillsMatched. For any job-freelancer pair (concretely, an applicant), SkillsMatched represents the proportion of the job's required skills fulfilled by the freelancer:

$$\text{SkillsMatched} = \frac{\#\text{ of skill matches between the buyer's requirement and freelancer's expertise}}{\#\text{ of skill requirements in the job description}}. \tag{1}$$

For the applications in our data sample, Figures 2a and 2b report the distributions of the SkillsMatched variable and of the corresponding number of skills compatibly matched. In general, Data Entry jobs require fewer skills (2.73 per opening in Data Entry, as opposed to 3.81 in Web Development) and hence tally fewer skills actually matched in Figure 2a. The Web Development market exhibits greater variability in both the skills required by jobs and those supplied by freelancers. In terms of applicants' fractions of demanded skills matched (i.e., SkillsMatched), Data Entry jobs exhibit substantial probability masses at each of 0 and 1, because required skills are more concentrated in a small focal set that many workers either satisfy or miss completely (Figure 2b). In the Web Development category, the bimodal concentration is noticeably weaker, because more diversely variable skills are demanded and supplied.

The freelancer who satisfies the greatest proportion of a job's required skills is not necessarily the one hired. On one hand, the typical job receives applications that vary considerably in the proportion of required skills matched (the within-job coefficient of variation is 0.52 in Data Entry and 0.46 in Web Development). On the other hand, average SkillsMatched for hires does not significantly exceed that for applicants (Figure 2b). Thus it seems that factors other than skills compatibility significantly sway buyers' hiring decisions.



(a) histogram of the number of skill matches (b) cumulative density of SkillsMatched

In Panel 2a, the horizontal axis shows the number of required skills matched by each applicant. The vertical axis shows the proportion of applications achieving that number.

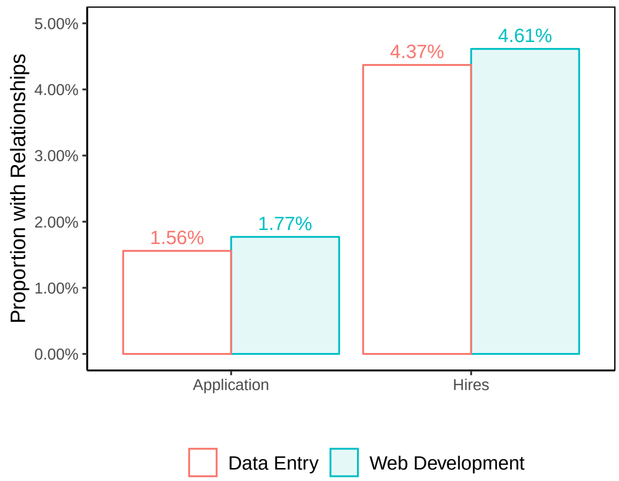
In Panel 2b, the horizontal axis shows SkillsMatched, the proportion (between 0 and 1) of required skills matched by the applicant. The vertical axis shows the cumulative density of applications reaching that proportion matched. The solid line represents the entire sample of applications, and the dotted line represents resulting hires.

Figure 2 Distributions of Jobs' Required Skills Matched by Applicants

2.3.2. Vertical differentiation. Our data include market participants' prior dealings and ratings. As a limitation in studying vertical differentiation, the data omit the signals of freelancer quality that buyers acquire by inspection, such as by interviewing applicants.

Prior dealings. Approximately 7% of our sample's job openings receive at least one application from freelancers who previously transacted with the posting buyer. In such cases, the buyer may have learned the freelancer's quality more precisely than she could have from inspection alone. However, buyers may also engage in repeat dealings with freelancers to avoid the setup and transaction costs incurred when initiating new working relationships, especially for jobs entailing complex tasks. More explicitly, one of two motivations (or both) can justify marketplace buyers' repeat dealing: (A) to avoid incurring initial transaction and setup costs (this rationale applies even when re-hiring freelancers of lesser quality); and (B) to beneficially re-engage trusted freelancers after ascertaining their high quality either by inspecting or by transacting.

While empirically distinguishing between buyers' dual motivations is important, we reserve this task for the empirical model. Here, we merely outline how the data elucidate prior dealings. Our data include buyers' hiring histories over all unique freelancers and usefully indicate every repeat hiring instance to which benefit (A) attaches. In contrast, benefit (B) leverages buyers' knowledge of trusted freelancers from *satisfactory* past dealings, which allow them to rationally anticipate an elevated likelihood of quality job performance from their trusted freelancers. Our data include the positively or negatively rated job performance in each prior dealing, so that we can re-construct buyers' expectations of being satisfied by known freelancers. We empirically model buyers' such expectations and allow the expectations to affect hiring decisions.



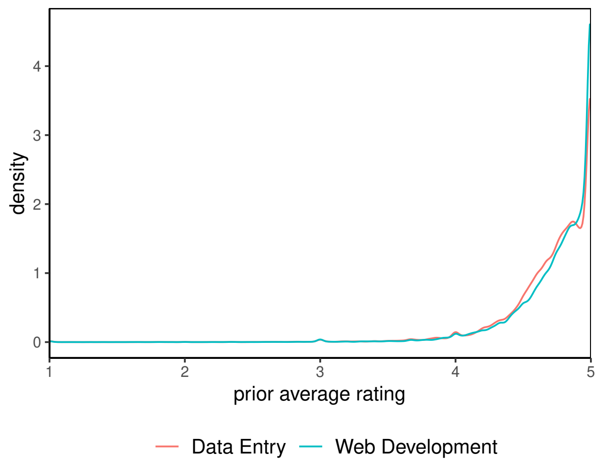
The bars show the proportion of applications/hires that are associated with at least one prior transaction (relationship) between the buyer and the applicant/hire at the time of application.

Figure 3 Shares of Applications and Hires Made with Past Dealings

Figure 3 shows the fraction of applications made by freelancers having prior dealings with posting buyers, as well as the fraction of hires with prior dealings. Buyers tend to hire applicants with past dealings: in both categories, the fraction of hires known from prior dealings substantially exceeds the fraction of known applicants. The fractions of prior dealing among applicants and hires are nominally higher in Web Development.

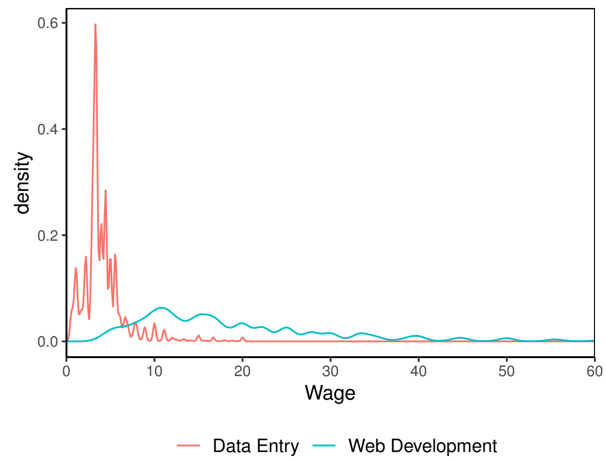
Ratings and predicted performance. Through the platform's reputational system, buyers provide public feedback regarding freelancers' job performance. By aggregating many buyers' experiences, public reputational ratings plausibly reveal freelancer quality beyond what buyers can privately learn.

After a job is completed, the buyer rates the freelancer on a 1-5 scale in each of the five areas of quality, communications, skills, cooperation, and ability to meet the given deadline. Figure 4 displays the distribution of applicants' scores averaged over all past jobs (when applying) and weighted equally over the five performance areas. We exclude the 8.0% (7.6%) of Data Entry (Web Development) applications lacking prior ratings from the respective plot. Most applicants are highly rated, such that the median area-averaged score among applicants is 4.77 in Data Entry and 4.81 in Web Development.



The horizontal axis shows applicants' average prior ratings at application. The vertical axis shows the density of such ratings among all applications.

Figure 4 Probability Density of Prior Ratings



The horizontal axis shows applications' final standing hourly wage bids. The vertical axis shows the corresponding density among all applications.

Figure 5 Probability Density of Wage

Public ratings that are informative and influential must accurately signal some quality attributes underlying freelancers' past job performance. In turn, stronger ratings revealing the signaled quality attributes should bolster a freelancer's hiring prospects. Accordingly, we model buyers' expectations of being satisfied in these ratings-captured attributes of freelancer quality. Relying on an extensive empirical literature on ratings in online markets (e.g., Bhattacharjee and Goel (2005), Forman et al. (2008), Dang et al. (2010), Mudambi and Schuff (2010), Fu et al. (2013), Ho-Dac et al. (2013), Mudambi et al. (2014)), we designate four stars as the threshold rating for positive feedback, which is also the threshold used by Pallais (2014). We define the indicator variable, $\text{SatisfiedRating}_f^c$, to represent the ex post event that the buyer rates freelancer f's performance with four or five stars upon completing job c. In analyzing buyers' hiring choices, we are interested in the $\text{SatisfiedRating}_f^c$ buyers *expect* (i.e., the perceived probability of a satisfied rating) when deciding which freelancer f to hire.

We model the buyer's expected satisfaction from freelancer f's job c performance using the logistic regression, where $l(\cdot)$ denotes log odds:

$$l\Big(\text{SatisfiedRating}_f^c\Big) = \alpha_0 + \alpha_1 PastRatings_f^c + \alpha_2 PastRelationship_f^c, \tag{2}$$

where the predictive covariates are applicant f's vector of average past ratings in the five performance dimensions, which we denote as $(PastRatings_f^c)$, whether she has had prior dealings with the posting buyer, and whether the buyer's ratings in such prior dealings express satisfaction with the job(s) completed. For brevity, we use ($PastRelationship^c_f$) to denote the covariates for both past dealing and its reputational interactions. Thus, we allow buyers to form rational expectations of being satisfied by a freelancer using the combined knowledge gleaned from prior dealings and from public ratings supplied by other platform buyers. We find neither an improved fit nor a statistically significant coefficient from adding to specification (2) a linear dependence on skills compatibility.

After training the model on the data to derive the estimates, $\{\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2\}$, we use these estimates to project the buyers' expected SatisfiedRating for every application in our sample.

2.4. Compensation and Platform Commissions

Buyers and freelancers negotiate jobs' hourly wage rates. Freelancers initiate the process with wage bids in their applications. Buyers can accept bids or make counteroffers to which freelancers can respond in kind. Our data records each applying freelancer's final standing wage bid (i.e., the last wage she offered or accepted), which for hired freelancers equals their transacted wage rates.

Figure 5 plots the distributions of final standing wage bids in the two job categories' marketplaces. Averaging \$4.13 per hour, over 50% of Data Entry final bids fall between \$3 and \$5. Substantially higher Web Development final bids average \$18.62 per hour. The categories' coefficients of variation are similar at 0.61 in Data Entry and 0.58 in Web Development.

2.4.1. Natural experiment arising from a change in commission structure. Empirical studies of marketplaces have long dealt with equilibrium prices that endogenously correlate with unobserved quality and fit. Our empirical methodology (Section 3) directly controls for freelancers' unobserved, vertically differentiated quality and for the prior literature's primary concerns about confounders including time-varying demand shocks. However, a natural experiment conveniently allows us to handle additionally minor concerns about price endogeneity.

During our study, a one-time change in the platform's commission structure provides exogenous variation in prices (i.e., negotiated wages). Before the change occurring on June 21, 2016, freelancers paid the platform a flat 10% commission on all earnings. From June 21, 2016 onward, the platform collected tiered commissions. Under the new system, freelancers who work more extensively for a single buyer receive quantity-based (dollar amount in earnings) discounts on their platform commissions. On her work for any one buyer, the freelancer pays the platform 20% of her first \$500 earned, 10% on her next \$9,500 in earnings (i.e., up to \$10,000 earned from the buyer), and 5% on all subsequent earnings. The platform-wide change affected all job categories.

Moreover, introducing tiered commissions differentially impacted freelancer-buyer dyads depending on whether they already had prior dealings. Thus, the switch provides useful empirical variation to separately identify wage effects from the value buyers attach to prior dealings (to the extent

We plot the daily proportion of job openings (by posting date) receiving at least one applicants having prior dealings with the posting buyer. The black vertical line marks the date the commission scheme changed.

Figure 6 Tiered Commission's Effect on Rates of Receiving Applications with Prior Dealings

buyers' cost savings from re-engaging a known freelancer may be "priced in"). As Figure 6 shows, tiered commissions boosted applications from freelancers with prior dealings. The share of openings receiving applicants with prior dealings sharply increases June 21, 2016.

3. Empirical Model and Estimation

Our methodology recovers buyer preferences for horizontally and vertically differentiated freelancers from online marketplace hiring data. To this end, the method accounts for channels through which buyers acquire information about freelancer quality and fit that are unobservable in the data; instead such information is inferred from hiring outcomes.

This section is organized as follows. Section 3.1 introduces the buyer utility model of hiring choice, and Section 3.2 overviews the empirical methodology. Section 3.3 derives a central methodological building block, the revealed preference moment inequality, which novelly bounds demand preferences using data observing hiring switches. Section 3.4 concludes by laying out an estimation procedure that bounds the set of preference parameters consistent with the data.

3.1. Model

We model buyers' hiring choices made in a two-sided marketplace for spot labor. After posting jobs to the marketplace platform, buyers hire from the consideration sets of freelancers who apply. We focus on buyers' hiring choices made from consideration sets already formed by applying freelancers, and we do not assume how freelancers choose where to apply. Instead, we directly control for freelancers' own quality and that of their competition. Consistent with the way the platform operates, we estimate demand separately for Data Entry and Web Development as distinct markets.

The hiring model breaks down into two parts describing preferences and choices, respectively.

3.1.1. Demand preferences. For each market, we model a representative buyer whose utility benefit from hiring freelancers is explained by their horizontally differentiated fit for jobs' requirements and by their vertically differentiated quality in job performance.

Horizontally differentiated fit. A freelancer's fit for a job depends foremost on how compatibly her skills match those demanded by the job. As defined in Section 2.3.1, the variable $\text{SkillsMatched}_{c,f}$ denotes the fraction of skills required by job c that are compatibly matched by freelancer f. In addition to skills, we allow f's fit for job c to include an additively separable, idiosyncratic component, ε_f^c, that the buyer perceives. In our model, ε_f^c is drawn i.i.d.

Therefore, by hiring freelancer f, the buyer receives her base value v_c of hiring for job c adjusted to reflect f's horizontally differentiated fit to job c:

$$v_c + \beta_{Skill} \cdot \text{SkillsMatched}_{c,f} + \varepsilon_f^c. \tag{3}$$

Vertically differentiated quality. The buyer interviews and tests applicant freelancers to closely assess quality. As a result of these efforts, the buyer collects a noisy signal, $q_f + \xi_c^f$, of freelancer f's quality. Here, q_f is f's true quality, and ξ_c^f represents i.i.d. noise in the buyer's signal. In terms of buyer utility, the perceived quality $q_f + \xi_c^f$ is the quality-based premium that the buyer receives on top of her fit-based value (3).

In addition, the buyer recalls knowledge of freelancer f's quality from their past dealings, if any, and learns from the platform's reputational ratings. As Section 2.3.2 explained, we project the effect of such knowledge on the variable $\text{SatisfiedRating}_f^c$, which is the buyer's rationally expected probability of being satisfied with f's job performance on c.

Therefore, the buyer's expected utility from freelancer f's vertically differentiated quality of performance on job c is:

$$q_f + \beta_{Rep} \cdot \text{SatisfiedRating}_f^c + \xi_f^c, \tag{4}$$

where we move the noise component of the quality signal ξ_f^c for later notational convenience.

Transaction costs and price. The buyer incurs setup costs to establish a working relationship (e.g., from training, establishing modes of communication, and imparting knowledge of the buyer's brand, needs, or working environment) that can be avoided by re-hiring a known freelancer. We define the indicator variable, R_f^c, for whether the buyer has previously hired freelancer f, and use it to estimate the setup cost, $cost_{Setup}$, as the coefficient on term $1 - R_f^c$. Lastly, the buyer pays the applicant's hourly wage p_f^c.

Utility from hiring. We bring together (3), (4), and the transaction costs. By hiring freelancer f for job c, the buyer collects utility, Π_f^c, in per-hour dollar terms:

$$\begin{aligned}\Pi_f^c = \ & v_c + q_f + \beta_{Rep} \cdot \text{SatisfiedRating}_f^c + \beta_{Skill} \cdot \text{SkillsMatched}_{c,f} \\ & - cost_{Setup} \cdot (1 - R_f^c) - p_f^c + \underbrace{\xi_f^c + \varepsilon_f^c}_{\epsilon_f^c},\end{aligned} \tag{5}$$

where we denote the i.i.d. sum of quality signal noise and idiosyncratic fit as ϵ_f^c.

For each category of jobs, we estimate the representative buyer's preference parameters, $\theta := \{\beta_{Rep}, \beta_{Skill}, cost_{Setup}\}$. When appropriate for exposition, we write the buyer's utility (5) as $\Pi_f^c(X_f, p_f^c, \epsilon_f^c; \theta)$ or $\Pi_f^c(\theta)$ to explicitly recognize its functional dependence on θ and freelancer f's attributes, $X_f := \{q_f, R_f^c, \text{Skills}_f\}$, at the time f applies.

3.1.2. Hiring choices. Buyer demand preferences are revealed through hiring decisions. When the buyer hires freelancer f for job c, we infer that hiring f maximizes the buyer's payoff from the set of applicants $\mathcal{F}_c$:

$$f \in \arg\max_{f' \in \mathcal{F}_c} \Pi_{f'}^c. \tag{6}$$

The buyer's revealed preference (6) implies the revealed preference inequalities:

$$\Pi_f^c \geq \Pi_{f'}^c \text{ for all applying freelancers } f' \in \mathcal{F}_c. \tag{7}$$

The revealed preference inequalities figure importantly in our empirical methodology.

3.2. Overview of Empirical Methodology

Our methodological contribution addresses a common and increasingly important type of short-term labor marketplace. These decentralized online marketplaces are characterized by a fast-churning base of workers supplying services for buyers who fill jobs by choosing hires from unique sets of self-selected applicants. When buyers *privately* observe a significant portion of applicants' quality, this setting poses an econometric challenge.

Given that quality information is privately learned, standard econometric methods for panel data would still permit us to discern freelancers' individual quality levels as freelancer fixed effects *if* our data were to observe large samples of hiring decisions made repeatedly from stable assortments. Intuitively, we could discern freelancers' quality differences by observing who is consistently hired over whom. However, in the setting we address, assortments are far from stable, as buyers hire from consideration sets entered into by very small slivers of the platform's hundreds of thousands of freelancers. Thus each buyer's choice assortment is an essentially unique grouping of freelancers, and no pair of freelancers is directly compared for hire asymptotically many times. Even if assortments were somehow made stable, the market's number of freelancers is asymptotically large, leading to the well-known incidental parameters problem causing inconsistency for conventional, panel-based estimates using fixed effects. (Section 4.2 later compares our results against biased and inconsistent panel estimates.)

This section introduces a choice-based estimation methodology to validly accommodate an asymptotically large number of horizontally and vertically differentiated freelancers, despite privately observed attributes and unstable choice assortments. Unlike traditional panel-based methods, our approach depends on observing asymptotically many hiring switches, in which pairs of freelancers being hired over one another in separate hiring instances. Section 3.2.1 describes how hiring switches imply informative bounds on buyer demand preferences. Section 3.3 then formally derives moment inequalities that bound preferences using observational data, and Section 3.4 introduces random set inference.

Lastly, our approach handles several, additional empirical concerns familiar to marketplace studies, including:

1. *Market demand shocks or seasonality.* Empirical studies of marketplaces have long controlled for time-varying market demand shocks (e.g., Berry et al. (1995)). Our estimates remain identical after accommodating time-varying demand fixed effects in specification (5).
2. *Bargaining selection bias.* The data contain applicants' final standing bids for wages, instead of fully negotiated wages. Buyers will rationally concentrate their bargaining efforts towards better applicants' bids. Despite selective bargaining, the revealed preference inequalities (7) still hold: the transacted offer is preferred over the rejected standing offers.

3.2.1. Counterfactual intuition. To build basic intuition as to why hiring switches informatively bound buyer preferences, consider the following counterfactual thought experiment.

Set a baseline hiring outcome in which worker 'A' is hired over worker 'B'. The thought experiment contemplates that we freely alter the attributes of workers 'A' and 'B' and then observe the buyer's resulting, counterfactual hiring decision. (We henceforth drop the single quotes placed around the worker labels 'A' and 'B'.)

Suppose we experiment counterfactually as follows. In the first experiment, we raise A's SkillsMatched by 10% while leaving all other attributes unchanged. Worker A's increase in her job-compatible skills enhances the value she delivers to the buyer, and the buyer consequently maintains her choice of A over B. In our next experiment, we still raise A's SkillsMatched by the same 10% but additionally decrease worker B's hourly wage bid by \$5. The revised manipulation results in the buyer choosing B over A, i.e., a hiring switch.

The switch from A to B reveals demand preferences. Specifically, we infer an upper bound on the buyer's willingness to pay for skills compatibility, β_{Skill}: to rationalize the buyer's choice in switching to B, increasing SkillsMatched by 10% cannot be worth more to the buyer than \$5 in hourly wage. More generally, B's value must have increased more than A, owing to changes in their time-varying attributes.

3.3. Bounding Revealed Preferences Using Moment Inequalities

From the data's hiring switches, we construct preference bounds in the form of moment inequalities.

3.3.1. Formal treatment of hiring switches. Take a hiring switch observed from the data. Without loss of generality, workers A and B are respectively hired over one another for jobs c and job c', where c' is the later job. Both workers apply for both jobs: $\{\mathrm{A},\mathrm{B}\} \subset \mathcal{F}_c \cap \mathcal{F}_{c'}$.

The corresponding revealed preference inequalities (7) are:

$$\Pi_A^c \geq \Pi_B^c \text{ and } \Pi_B^{c'} \geq \Pi_A^{c'}. \tag{8}$$

Together, they imply the following inequality, which serves as our methodology's building block:

$$\Pi_A^c - \Pi_B^c + \Pi_B^{c'} - \Pi_A^{c'} \geq 0, \tag{9}$$

We clarify intuition by re-arranging (9) as:

$$\Pi_B^{c'} - \Pi_B^c \geq \Pi_A^{c'} - \Pi_A^c. \tag{10}$$

Recalling Section 3.2.1's rationalization of hiring switches, we can readily interpret inequality (10) as positing that B's improvement in buyer value, $\Pi_B^{c'} - \Pi_B^c$, exceeds A's, $\Pi_A^{c'} - \Pi_A^c$. This conclusion is required to rationalize the observed switch in buyer choice.

While the revealed preference inequalities (9) necessarily follow from observed choices, they fail to translate into consistent and informative bounds on buyer preferences unless we apply three additional steps. The next procedure undertakes these steps to construct moment inequalities.

3.3.2. Moment inequalities. To validly bound demand preferences from observational data, moment inequalities are constructed as follows.

Step 1: Average away idiosyncratic variation using sample moments. First, carefully constructed sample moments "average away" the idiosyncratic (non-preference) influences on hiring choices in observational data. The results are sample moment inequality bounds on preferences.

Specifically, worker f's perceived value to the buyer, Π_f^c, includes a stochastic component, ϵ_f^c, consisting of her idiosyncratic fit for job c and the noise within the buyer's signal about her quality:

$$\Pi_f^c = v_c + q_f + \beta_{Rep} \cdot \mathrm{SatisfiedRating}_f^c + \beta_{Skill} \cdot \mathrm{SkillsMatched}_{c,f} - cost_{Setup} \cdot (1 - R_f^c) - p_f^c + \epsilon_f^c.$$

By holding ϵ_f^c fixed, our thought experiments validly attribute their counterfactual hiring switches exclusively to buyer preferences. In this way, we were able to place an upper bound on preference parameter β_{Skill} in Section 3.2.1. In observational data, the hiring switch inequality (9) still holds, yet buyer preferences are not the exclusive explanation for hiring switches. Instead, they occur for cases where $\epsilon_A^c - \epsilon_B^c + \epsilon_B^{c'} - \epsilon_A^{c'}$ is large. In other words, the buyer could switch its hiring due to an exceptionally strong idiosyncratic fit or an especially noisy signal of worker quality.

Even when the inequalities (9) are individually uncertain, we can distill down to reliable bounds on preferences by using aggregation to "average away" the effects of idiosyncratic fit and noise. We construct the basic moment inequality as follows.

Let $n = (c, c', f, f')$ index the hiring switches observed in the data, and use N to denote the set of observations n. The basic moment inequality $m(\theta)$ and its sample analogue $\hat{m}(\theta)$ are defined as:

$$m(\theta) := \mathbb{E}\left[\Pi_f^c(\theta) - \Pi_{f'}^c(\theta) + \Pi_{f'}^{c'}(\theta) - \Pi_f^{c'}(\theta)\right] \tag{11}$$

$$\hat{m}(\theta) := \frac{1}{|N|}\sum_{n\in N}\left(\hat{\Pi}_f^c(\theta) - \hat{\Pi}_{f'}^c(\theta) + \hat{\Pi}_{f'}^{c'}(\theta) - \hat{\Pi}_f^{c'}(\theta)\right) \tag{12}$$

$$\text{where } \hat{\Pi}_f^c(\theta) := \beta_{Rep}\cdot \text{SatisfiedRating}_f^c + \beta_{Skill}\cdot \text{SkillsMatched}_{c,f} - cost_{Setup}\cdot(1 - R_f^c) - p_f^c. \tag{13}$$

In particular, the moment inequality $m(\theta) \geq 0$ represents the buyer's ideal revealed preference bound and derives from inequality (9) when all idiosyncratic horizontal fit and perception errors have been averaged away. Indeed, by such averaging, the sample analogue $\hat{m}(\theta) \geq 0$ bounds buyers' preferences with increasing precision as the data sample increases in size:

$$\hat{m}(\theta) \rightarrow_p m(\theta) \text{ as } |N| \rightarrow \infty. \tag{14}$$

Thus the sample moments approach the true moments, just as in GMM (Generalized Method of Moments, Hansen (1982)) estimation. While GMM works with moment equalities, we work with moment inequalities, $m(\theta) \geq 0$, to bound preferences (instead of adding modeling assumptions that restrictively pin down point estimates). At the true preference values, θ_0, the hiring switch inequality (9) implies that:

$$\hat{m}(\theta_0) \rightarrow_p m(\theta_0) \geq 0. \tag{15}$$

For this reason, we are interested in using sample moment inequalities to suss out the preference parameter values, θ, rationalizing observed hiring patterns. Like GMM for moment equalities, we next use instrumental variables to generate additional moment inequalities.

Step 2: Generate informative moment inequalities. As Section 3.2.1's example illustrates in deriving a preference upper bound, differently chosen sets of counterfactual observations differently bound buyer preferences. For proper identification, we clearly require at least two bounds, an upper and a lower, on each model parameter. To obtain these, we must generate a sufficient number of informative moment inequalities.

Mirroring GMM, we construct additional moment inequalities using instrumental variables. Since each new moment $m_h(\theta)$ is defined by an instrument variable h, the problem amounts to selecting an informative set of instrumental variables, H.

For the instrumental variable $h \in H$, we define the associated moment, $m_h(\theta)$, and its sample analogue, $\hat{m}_h(\theta)$:

$$m_h(\theta) := \mathbb{E}\left[h_n \times \left(\Pi^c_f(\theta) - \Pi^c_{f'}(\theta) + \Pi^{c'}_{f'}(\theta) - \Pi^{c'}_f(\theta)\right)\right] \tag{16}$$

$$\hat{m}_h(\theta) := \frac{1}{|N|}\sum_{n\in N} h_n \times \left(\hat{\Pi}^c_f(\theta) - \hat{\Pi}^c_{f'}(\theta) + \hat{\Pi}^{c'}_{f'}(\theta) - \hat{\Pi}^{c'}_f(\theta)\right). \tag{17}$$

We provide sufficient conditions assuring that an instrument h associates with a valid moment inequality, $m_h(\theta) \geq 0$. We impose the following exclusion restriction.

Assumption 1 **(Exclusion Restriction)**. *Let random variable $Z_{c,f}$ be indexed for jobs c and freelancers f. The exclusion restriction holds when ex ante for $n \in N$:*

$$\mathbb{E}\left(\epsilon_{\tilde{c},\tilde{f}} | Z_{c,f}, Z_{c,f'}, Z_{c',f'}, Z_{c',f}\right) = 0 \textit{ for all } \tilde{c} \in \{c, c'\}, \tilde{f} \in \{f, f'\}. \tag{18}$$

Such random variables $Z_{c,f}$ (when also relevant) can be thought of as akin to conventional instrumental variables used in GMM. Notably, the observed covariates in $\hat{\Pi}^c_f$ qualify as conventional GMM instruments satisfying the exclusion restriction vis-à-vis the idiosyncratic $\epsilon_{c,f}$. (We later re-visit this assumption in the case of the worker's wage bid, p^c_f.)

However, for our method to work like GMM, each conventional GMM instrument should supply *two* moment inequalities, corresponding to upper and lower bounds identifying a parameter. Moreover, unlike GMM, for valid moment inequalities $m_h(\theta) \geq 0$, we must carefully choose h_n to preserve the underlying choice-based inequality (9) (e.g., avoid flipping its sign).

The following Proposition 1 assures that non-negative functions of conventional GMM-style instruments generate valid moment inequalities. Appendix A provides all proofs.

Proposition 1 **(Moment Inequality Validity)**. *Suppose Assumption 1 holds for the random variable $Z_{c,f}$ with support D_Z. Let h be the random variable defined by any given non-negative function $g : D_Z \times D_Z \times D_Z \times D_Z \to \mathbb{R}_+$:*

$$h_n := g(Z_{c,f}, Z_{c,f'}, Z_{c',f'}, Z_{c',f}) \geq 0. \tag{19}$$

Then moment condition $m_h(\theta) \geq 0$ holds.

Drawing from Ho and Pakes (2014)'s practical experience applying moment inequalities, we use as g the non-negative functions given by the (1) absolute values, (2) positive parts, and (3) negative parts of applicants' differences in the covariate Z:

1. $g_1^1(Z_{c,f}, Z_{c,f'}, Z_{c',f}, Z_{c',f'}) = (Z_{c,f} - Z_{c,f'})^+$, $g_1^2(Z_{c,f}, Z_{c,f'}, Z_{c',f}, Z_{c',f'}) = (Z_{c',f} - Z_{c',f'})^+$
2. $g_2^1(Z_{c,f}, Z_{c,f'}, Z_{c',f}, Z_{c',f'}) = (Z_{c,f} - Z_{c,f'})^-$, $g_2^2(Z_{c,f}, Z_{c,f'}, Z_{c',f}, Z_{c',f'}) = (Z_{c',f} - Z_{c',f'})^-$
3. $g_3^1(Z_{c,f}, Z_{c,f'}, Z_{c',f}, Z_{c',f'}) = |Z_{c,f} - Z_{c,f'}|$, $g_3^2(Z_{c,f}, Z_{c,f'}, Z_{c',f}, Z_{c',f'}) = |Z_{c',f} - Z_{c',f'}|$
4. $g_4(Z_{c,f}, Z_{c,f'}, Z_{c',f'}, Z_{c',f}) = |Z_{c,f} - Z_{c,f'} + Z_{c',f'} - Z_{c',f}|$

When g is defined as the positive (negative) part, it places positive weight onto observations where the applicants' differences in $Z_{c,f}$ are positive (negative). The moment inequalities resulting from applying positive and negative parts typically provide upper and lower bounds on the buyer's preference for $Z_{c,f}$. We illustrate with an example.

Consider the conventional GMM instrumental variable, $\text{SkillsMatched}_{c,f}$, which is the covariate for skills compatibility between worker f and job c. Because $\epsilon_{c,f}$ represents truly idiosyncratic fit and quality perception error, the $\text{SkillsMatched}_{c,f}$ variable satisfies the exclusion restriction. Without loss of generality, let f be hired for job c and f' be hired for the later job c'.

We illustrate how taking g to be, alternately, the positive and negative part functions serves to identify β_{Skill}, the buyer's preference for skills compatibility. First, define the instrument by taking the positive part:

$$h_n := \left(\text{SkillsMatched}_{c',f} - \text{SkillsMatched}_{c',f'}\right)^+. \tag{20}$$

Then h_n is large precisely when worker f offers skills that are significantly more compatible with job c' than the skills offered by f'. Yet the buyer rejects worker f to hire her skills-deficient competitor f'. On average over many observations positively weighted by such h, it stands to reason that the buyer chooses worker f' for job c' because the buyer values her non-skill attributes. (On average over the same observations, worker f' must have improved her non-skill attributes since job c. Otherwise, if they were just as good then, she should have been hired on average.) Her improvement in such attributes is used to upper bound the value of skills, β_{Skill}. The underlying intuition conveniently echoes Section 3.2.1's thought experiment.

Next, consider the instrument taking the negative part:

$$h_n := (\text{SkillsMatched}_{c',f} - \text{SkillsMatched}_{c',f'})^- = (\text{SkillsMatched}_{c',f'} - \text{SkillsMatched}_{c',f})^+, \tag{21}$$

which instead weights observations in which hired worker f''s compatible skills for job c' significantly dominate the rejected worker f's. Despite f being rejected for job c', such h places increased weight onto instances where f improves her buyer value between jobs c and c'. This is because when her competitor f' significantly gains in skills compatibility from c to c', she is more likely to be rejected for c' even after she substantially enhances her own value. Thus, compared to the unconditional average ignoring whether f gains a skills advantage for c', we will tend to see higher average improvements from the rejected worker f'. Such average improvements lower bound the value of skills, since f is still hired.

Step 3: Price endogeneity concerns. Marketplace studies prevalently concern themselves with equilibrium prices endogenously correlating with unobserved quality. Our methodology handles this concern by controlling directly for freelancer fixed effects, q_f. By construction, both the hiring switch preference inequality (9) and the moment inequalities $m_h(\theta) \geq 0$ neatly difference away such fixed effects. By eliminating the fixed effects, we are able to consistently estimate the parameters θ. We later recover the distribution of freelancer fixed effects.

While fixed effects resolve our main concern about endogenous prices, we can also handle a second, more subtle concern. The wage bid p_f^c may also correlate endogenously with $\epsilon_{c,f}$, a worker's idiosyncratic fit for the job or the buyer's noise in perceiving her quality. For example, worker f might plausibly recognize the buyer's realized $\epsilon_{c,f}$ and "price in" some of the value. The wage bid p_f^c then fails the exclusion restriction, making it unsuitable for deriving moment inequalities (Proposition 1).

As Section 2.4 explains, we exploit the exogenous change in prices (i.e., wage bids) that results from the introduction of tiered platform commissions. The introduction event (and an interaction with relationship variable R_f^c) are used as conventional instrumental variables to generate moment inequalities in place of prices. Appendix E presents statistical support for the relevance of these instruments.

In practice, price-quality endogeneity appears fully addressed by the model's fixed effects. We do not find that our estimates change significantly when we forego replacing the wage bid with the commission change instruments.

3.4. Estimation of the Identified Set

To recap, in Section 3.3, we aggregate choice-based inequalities of the form (9) derived from observed hiring. By weighting such observations using carefully chosen instrumental variables, the resulting moment inequalities informatively bound the multiple dimensions of buyer demand preferences. By averaging away idiosyncratic horizontal fit and perception errors, these inequalities bound the preference parameters of interest with increasing precision as the data sample increases in size.

Together, such moment inequalities bound an *identified set* Θ_0 of parameter values:

$$\text{Identified set } \Theta_0 := \left\{\theta \in \mathbb{R}^3 : m_h(\theta) \geq 0 \text{ for all } h \in H\right\}. \tag{22}$$

Geometrically, each moment inequality constraint $m_h(\theta) \geq 0$ represents a half-space delineated by its moment's hyperplane in the space of parameters θ. All parameter values consistent with revealed preference lie to one side of each hyperplane. Representing the intersection of all such half-spaces, Θ_0 contains the parameter values consistent with all moment inequalities $m_h(\theta) \geq 0$.

The estimation procedure we describe is conservative and robust in two ways. First, the object we aim to estimate is the identified set Θ_0, rather than the true parameter values it contains. Thus, we forego making additional modeling assumptions and rely on conservative bounds inferred from revealed preference inequalities alone. Second, we carry out random set inference. Concretely, we estimate a coverage set (Chernozhukov et al. 2007, Bugni 2010) that covers the *entire* identified set Θ_0 with an appropriate coverage probability (e.g., 90%). For economic interpretation, the confidence interval for any single parameter is then constructed by projecting down the coverage set's extreme points, which is a method proposed by Pakes et al. (2015).

While the mechanics of random set inference are quite different, inference using moment inequalities shares core intuitions in common with inference under the econometric workhorse of GMM. With more data, the estimated sample moment inequalities $\hat{m}_h(\theta) \geq 0$ become more precise, asymptotically approaching the true moment inequalities $m_h(\theta) \geq 0$ bounding the identified set. As done for GMM's sample moments $\hat{m}_h(\theta) = 0$ which asymptotically approximate GMM moment conditions $m_h(\theta) = 0$, we derive large-sample confidence bounds based on scaling the observed variability of $\hat{m}_h(\theta)$. Since the coverage set must cover Θ_0 entirely with a prescribed confidence level, it depends on the joint variability of the sample moments, $\hat{m}_h(\theta)$ for $h \in H$. For parameter values in the interior of Θ_0, violations of the sample moment inequality conditions are $o_p(1)$.

We define the set estimator, discuss its consistency and strategies for inference.

Definition 1 (Set Estimator). Let D be a diagonal matrix, with each diagonal element equal to the variance of each moment evaluated at the true parameter:

$$D_{ii} \triangleq \mathbb{V}(m_i(\theta)). \tag{23}$$

Also let $\hat{D}$ be a consistent estimator of D.[5] The set estimator $\hat{\Theta}$ is defined as follows:

$$\hat{\Theta} \triangleq \arg\min_{\theta} \left\| \hat{D}^{-1/2} \{\hat{m}_h(\theta)_-, h \in H\} \right\|, \tag{24}$$

where $\hat{m}_h(\theta)_-$ denotes the negative part of $\hat{m}_h(\theta)$, i.e., the extent to which $\hat{m}_h(\theta) \geq 0$ is violated, and $\{\hat{m}_h(\theta)_-, h \in H\}$ is organized as a vector.[6]

Proposition 2 formally establishes the consistency of the set estimator $\hat{\Theta}$ in that the Hausdorff distance between $\hat{\Theta}$ and the identified set Θ_0 approaches zero asymptotically as the number of observations increases.

[5] In practice, $\hat{D}$ could be obtained by a two-step approach, analogous to the optimal weighting matrix used in GMM.

[6] The set estimator can be a point singleton. In practice, it is not uncommon for moment inequalities to "cross", such that no θ values set (24)'s minimand to zero (see Pakes et al. (2015)). That is, the minimand's 0-sublevel set is empty, without parameter values satisfying all sample moment inequalities $\hat{m}_h(\theta) \geq 0$. See Ho and Pakes (2014) for details regarding the procedure we use to construct a coverage set in this case. Others including Bugni (2010) instead utilize asymptotic slack to prevent crossing and directly derive coverage sets.

PROPOSITION 2 **(Consistency of the set estimator)**. *Define $d_H(\cdot,\cdot)$ the Hausdorff distance. Then if the parameter space is compact and $\hat{D}$ is a consistent estimator of D, we have $d_H(\hat{\Theta},\Theta_0) = o_p(1)$.*

We are particularly interested in the range in which each true parameter lies. Hence we project the set estimator $\hat{\Theta}$ onto each dimension of θ, i.e., $cost_{Setup}$, β_{Rep}, and β_{Skill}, respectively. For any $\theta_i \in \theta$, the boundaries of projections, i.e., $\arg\min_{\theta\in\hat{\Theta}} \theta_i$ and $\arg\max_{\theta\in\hat{\Theta}} \theta_i$, are by definition conservative bounds of θ_i. In fact, such bounds are consistent, demonstrated in the following proposition. The proof follows directly from Proposition 1 in the Appendix of Pakes et al. (2015).

PROPOSITION 3 **(Consistency of bounds in each dimension of the parameter space)**. *For each $\theta_i \in \theta$, define $\hat{\bar{\theta}}_i \triangleq \arg\max_{\theta\in\hat{\Theta}} \theta_i$, and $\hat{\underline{\theta}}_i \triangleq \arg\min_{\theta\in\hat{\Theta}} \theta_i$, where $\hat{\Theta}$ is the set estimator defined in (24). Analogously define $\bar{\theta}_i \triangleq \arg\max_{\theta\in\Theta_0} \theta_i$, and $\underline{\theta}_i \triangleq \arg\min_{\theta\in\Theta_0} \theta_i$, where Θ_0 is the identified set defined in (22). Then, $\hat{\bar{\theta}}_i \xrightarrow{p} \bar{\theta}_i$, and $\hat{\underline{\theta}}_i \xrightarrow{p} \underline{\theta}_i$.*

This projection method is proposed by Pakes et al. (2015) and is particularly useful in applied settings, and we closely follow their steps in estimation. Details are in Appendix B.

4. Empirical Findings

In this section, we report our empirical findings for the Data Entry and Web Development categories. To preview, we find that buyer-sourced, private information about workers' quality consistently explains the majority of variation in buyers' hiring decisions. In contrast, the value of skills compatibility varies by category, being found valuable in Data Entry.

In contrast, we find no evidence that setup costs play a role in Data Entry hiring decisions. Intuitively, because Data Entry tasks are relatively simple, little setup is required. Instead, it is arguably surprising (although notably consistent with Pallais (2014)) that buyers rely so strongly on privately vetting their prospective hires, even for relatively unsophisticated tasks. Setup costs are estimated to be substantial for the more complex projects found in Web Development.

We further compare results obtained using our moment inequality approach against those from alternative discrete-choice methods. Our method is shown to outperform the alternatives, which are affected by expected empirical biases and known computational issues.

4.1. Estimation Results

Estimation results for the Data Entry and Web Development categories are displayed in Table 2. Using the consistent estimates obtained using moment inequalities, we additionally recover the distribution of freelancers' quality premia, q_f. Outlined in Appendix G, the procedure allows for freelancers' wage bids to be correlated with q_f.

Given the sample's variation in worker covariates, we find that freelancers' quality q_f explains buyers' choices to a greater extent than the observable information from skills and ratings. Freelancers' inspected quality accounts for 86.7% of hiring variation in the Data Entry category and 87.6% in Web Development.[7] Practically speaking, the result rationalizes the fact that buyers significantly vet freelancers privately in practice.

	Data Entry			Web Development		
	Our Method	Logit	Logit	Our Method	Logit	Logit
β_{Rep}	0.42	1.32	40.80**	1.94**	−0.00	206.05**
	$[-0.38, 1.18]$	(10.42)	(6.72)	$[0.66, 5.83]$	(0.07)	(29.08)
$cost_{Setup}$	0.09	1.71	65.37**	1.16**	0.23	356.28**
	$[-0.25, 0.22]$	(42.14)	(7.79)	$[0.26, 2.11]$	(6.65)	(41.48)
β_{Skill}	0.92**	−32.18	−4.54**	1.89	−4.60	9.09**
	$[0.56, 3.51]$	(272.58)	(0.92)	$[-1.54, 8.96]$	(62.68)	(3.16)
Quality	Y	Y	N	Y	Y	N
N	full sample	subsample of 1500	full sample	full sample	subsample of 1500	full sample
McFadden R^2	79.34%	92.58%	1.53%	85.13%	90.15%	1.51%
McKelvey R^2	98.89%	95.72%	0.60%	98.70%	95.04%	0.59%
Count R^2	99.7%	99.4%	62.5%	99.0%	99.4%	53.5%

In Logit models, standard errors are reported in rounded parentheses. In our method, confidence sets are reported using squared brackets. ** shows statistical significance at the confidence level of 0.05.
McFadden $R^2 = 1 - LL_{full}/LL_{null}$, where the null model is the restricted model where all q_f's and demand estimators are zero except an intercept. McKelvey & Zavoina $R^2 = \mathbb{V}(\hat{y}^*)/(\mathbb{V}(\hat{y}^*) + \mathbb{V}(\hat{\epsilon}))$. Count R^2 = the number of correct predictions/total count.

Table 2 Demand Estimates

4.1.1. Data Entry category. In the Data Entry category, only β_{Skill} is statistically significant (Table 2, column 1 under Data Entry). We fail to reject that the setup cost is zero, nor that ratings-based reputations fail to valuably influence hiring.

β_{Skill}'s significance implies that the market values horizontal differentiation and flexibility: buyers' payoffs are enhanced when the platform attracts freelancers who match their skill requirements, holding all else equal. The intuition grounds in the nature of Data Entry jobs: many tasks in such jobs are routine and decompose into well-defined procedures (Autor et al. 2003). Required skills are hence easily defined by looking at the needs for each substep. For example, "Microsoft Word" is a concrete skill summarizing the capacity to complete tasks involving input and formatting in that software. Empirically, skills compatibility accounts for just under \$1 in hourly wage value, in a market where the average wage is just over \$4, meaning that it determines up to 22% of the value in hourly wages. (This figure ranges from 14% to 85% of the wage over our conservative bounds.)

[7] Explanatory fit is computed through $1 - LL_{full}/LL_{restricted}$, where in the restricted model, we restrict all $q_f = 0$.

We do not estimate statistically significant β_{Rep}. Echoing Pallais (2014), ratings do not appear significantly informative in Data Entry, and we fail to infer that they are valued much by marketplace buyers. We also do not estimate significant $cost_{Setup}$ in the Data Entry category. Because Data Entry jobs are relatively simple, buyers do not avoid significant setup costs by re-hiring the same freelancers.

4.1.2. Web Development category. In the Web Development category, we observe the opposite pattern, wherein β_{Rep} and $cost_{Setup}$ are statistically significant, but not β_{Skill} (Table 2, column 1 under Web Development).

Given the complexity of Web Development jobs, they entail high setup costs, including familiarizing with the architecture of client webpages and understanding existing code and development processes. Hence, it is useful to make repeated hires and prefer the freelancer who has already acclimated. Based on our estimates, this preference amounts to \$1.16 per hour of inconvenience saved by continuing to work with a known freelancer in a market with an \$18.62 average wage (thus about 6%).

On the other hand, contrary to Data Entry jobs, Web Development jobs often involve non-routine and non-modular tasks. Concrete, easily verified skills may offer limited descriptors for fit. For example, "CSS" far from fully describes traits that suffice to build a well-designed website, and fit may be substantially idiosyncratic. Other valuable traits such as adaptability and resourcefulness are instead reflected in the quality premium. Therefore we do not find significant β_{Skill}.

Suggesting further that quality matters, we estimate significant β_{Rep} for the Web Development category. Buyers respond to freelancers' expected job performances based on the platform's reputational ratings. A 10% (50%) increase in the buyer's probability of being satisfied ratings-wise is estimated to be worth \$0.19 per hour (\$0.97 per hour), or 1% (5%) of the hourly wage. Thus our findings validate an apparently modest influence of public ratings on hiring decisions; instead, buyers' privately gleaned information about freelancers' quality dominates decision-making.

4.2. Discrete-choice Method Comparison

We compare our method against alternative discrete-choice methods in the form of logit models estimated with (Table 2, column 2 for each category) and without freelancer fixed effects (column 3). The results corroborate that both approaches fail to resolve empirical issues addressed by our moment inequalities approach.

First consider the consequences of not including the freelancer fixed effects q_f. The classic endogeneity concern affecting marketplaces is that prices are biased by unobserved quality. In our setting, higher quality applicants demand higher wages, and ignoring that we observe less quality-relevant information than buyers do would result in under-estimating their price elasticities. Hence, not including q_f results in an endogeneity bias attentuating price elasticity.

We find evidence of exactly such a bias in the estimation without fixed effects (Table 2, column 3 for each category). Because utility is normalized to dollar terms, an attenuated price elasticity manifests as overly inflated coefficients for the other attributes. In both categories, the other attributes' coefficients are indeed implausibly large and significant. For example, the estimated cost incurred from working with an unfamiliar Data Entry freelancer is \$65.37 hourly, which dwarfs the category's average hourly wage of \$4.13. Likewise, the method estimates that a 10% increase in the buyer's probability of being satisfied ratings-wise is worth an implausibly high \$4.08 per hour. While Web Development wages average \$18.62 hourly, the buyer would pay an estimated \$356.28 hourly to avoid working with an unfamiliar freelancer and \$20.61 hourly for a 10% increase in the probability of being satisfied. Adding to the implausible findings for Data Entry, the value of compatible skills is found to be significantly *negative*. This type of mis-attribution can occur when high quality applicants are hired over their skilled counterparts but such quality is privately known to buyers and not controlled for.

The logit model with fixed effects (Table 2, column 2 for each category) attempts to control for freelancer quality, but suffers from the incidental parameters problem. Consequently, the estimated coefficients are biased. Additionally, while our moment inequality methods are robust to selective bargaining of wages, the logit ignores that certain recorded bids are more thoroughly negotiated by buyers than others. While the exact effects of each bias are difficult to predict, the end result is stark. The logit with fixed effects is unable to discern any significant coefficients of interest in either market, making it rather unsuitable for our prescriptive marketplace analyses. The failure to identify any significant coefficients may stem from the fixed effects overfitting individuals in the absence of a long panel. We additionally note that unlike our eminently scalable moment inequalities approach, the logit with fixed effects is computationally prohibitive for large-scale data such as ours. We report estimates from tractably limited subsamples, which may contribute to the method's lack of statistical power.

In sum, both discrete-choice alternatives suffer significant biases affecting the reliability of their estimates. At the same time, these issues corroborate that controlling for buyers' private information about their available choices is critical for reliably estimating marketplace demand preferences.

5. Implications

In this section, we study the intermediary's problem of matching workers to jobs so as to increase buyer utility and satisfaction. More specifically, we compare policies that differently allocate freelance workers into buyers' choice sets by prioritizing chosen freelancer attributes, such as compatible skills or the existence of prior dealings. We evaluate the performance of these policies using dynamic simulations that incorporate our empirically estimated buyer demand preferences.

Our objective is to recommend simple rules to diagnose and exploit the mix of worker information needed to assign better matches. Our results show that, compared to the most straightforward policy that prioritizes compatible skills, prioritizing repeat-dealing relationships improves buyer welfare. Critically, the appropriate policy will depend on the category's job complexity, which entails higher setup costs. In categories where setup costs are high, the platform should prioritize initial matches and avoid resets, whereas in markets where setup costs are low, it can encourage both repeat-dealing relationships and exploration to improve the value of matches.

5.1. Overview

The major value that an online labor marketplace delivers is to match buyers with freelancers that fit their needs, by allocating freelancers to buyers' choice sets.[8] The counterfactual experiment is hence designed to evaluate how different policies of allocation affect the long-run buyer welfare. Given a specific platform's policy, the simulation uses observed data and estimated buyers' utility function to mimic the real job application and hiring process, as well as dynamics in the marketplace. Our model accommodates both horizontal and vertical differentiation in the same way as described in Section 4. Freelancers differ in their qualities vertically, through their q_f. The difference between buyers is captured by different needs in skills, the extent of prior dealings, and for any given job the idiosyncratic fit offered by a freelancer.

The problem that the platform faces is how to design choice sets for buyers in a dynamic setting. We discuss the unique features of this problem as follows.

1. Given the unobservability of freelancers' quality premium, there is an exploration-exploitation trade-off. To be more specific, from the buyers' perspective, they might want to explore more freelancers and obtain their quality in order to hire a better applicant. On the other hand, there is also value to exploitation, where freelancers who are proved to be good in quality should be encouraged to be hired more. The platform's dynamic allocation strategy has to take this into account: buyers' exploration/exploitation has to be aided by the platform: The platform's strategic allocations of freelancers affects the set of freelancers the buyer can potentially explore or exploit.
2. The outcome of the exploration-exploitation trade-off depends on buyers' decisions and hence is also related to information only observable to buyers (e.g., idiosyncratic perception error), and such information is carried over time through prior dealings.

In practice, oftentimes the platform conducts match based on transparent information like skills, instead of prioritizing high-quality freelancers. We, instead, design and compare allocation policies that capture each feature of the market highlighted in the discussion above. Our candidate policies

[8] In practice, this is done through displaying search results and recommender systems.

including prioritizing horizontal differentiation, and focusing more on information local to agents: the relationships. Our results highlight the value of relationships through creating and conveying information. Not only it serves as a good proxy for the unobserved quality, it also helps exploitation. We recommend policies that balance the value of exploration and exploitation.

In what follows we describe the assumptions in the counterfactual model.

The Marketplace. We consider all job openings posted from January 1, 2015 to January 7, 2015, and job applications submitted during the same period of time. We view each different application corresponding to a different freelancer (applicant) in the simulation, and each job opening corresponding to one buyer. Specifically, we look at the Data Entry category, where there are 559 freelancers and 103 buyers, and the Web Development category, where there are 585 freelancers and 166 buyers.

Buyers. We assume each buyer only hires one freelancer within her choice set, based on the utility model (5) with coefficients estimated in section 4.[9] Specifically, in the Data Entry market,

$$\Pi_f^c = v_c + 0.92 \times \text{SkillsMatched}_f^c + q_f - p_f^c + \epsilon_f^c. \tag{25}$$

In the Web Development market,

$$\Pi_f^c = v_c + 1.94 \times \text{SatisfiedRating}_f^c - 1.16 \times (1 - R_f^c) + q_f - p_f^c + \epsilon_f. \tag{26}$$

Sequence of Events. In each period of our dynamic allocation model, the platform moves first to allocate freelancers to buyers' choice sets, then each buyer makes the hiring decision within the choice set, and each freelancer's R_c^f is updated after the hires. To mimic the dynamics on the platform, including inflows and outflows of freelancers and the platform growth, we also embed freelancers' churning behavior into the model.

In each period, the following events happen sequentially:

1. Churning: At the beginning of each period, a random subset of freelancers are replaced by new freelancers. The new freelancer has zero relationships with all buyers, and her characteristics (skill matches, ratings, etc.) are drawn from the empirical distribution. The probability of churns is set to the empirical churning rate in each market (0.023 in the Data Entry market and 0.029 in the Web Development market).
2. Allocation: The platform allocates freelancers to each employees' choice sets, according to policies to be discussed in the next subsection.
3. Hiring: Buyers hire freelancers from their choice sets.

[9] If a coefficient is estimated to be not statistically significant, it does not enter the buyers' utility function. For example, we find in Section 3 that $cost_{Setup}$ is not significant in the Data Entry market. Hence $cost_{Setup}$ is not included in equation (25), implying that buyers do not take R_f^c into account when they decide whom to hire.

4. Updating: If a freelancer is hired, relationship increases by one. For pairs where no hiring happens, relationship does not change. Skills and ratings are assumed to not change over periods.

5.2. Candidate Policies and Evaluation.

We consider the following candidate policies, with mild constraints on the size of choice sets and freelancers' capacity.

Benchmarks.

• *"Full Information Benchmark"*. For each buyer, the "Full Information Benchmark" policy finds freelancers with highest payoff according to model (5),[10] and allocates them to the buyer's choice set. Note, under this policy it is assumed that the platform has all information, including each freelancer's quality. This is typically unrealistic, and this policy only serves as a benchmark and gives us an upper bound for comparison.

• *"Random"*. The "Random" policy allocates freelancers in a purely random manner, and does not use any information. Clients can still select the most appealing hire from among the freelancers present in their consideration sets.

Policies Focusing on Horizontal and Vertical Fit.

• *"Skills"*. The "Skills" policy only uses skills information. The platform ranks each freelancer according to skill matches with the buyer, and allocates freelancers to buyers' choice sets according to the rank. The "Skills" policy aims at maximizing horizontal fit and flexibility of the platform.

• *"Relationship priority + random"*. Under the "Relationship priority + random" policy, the platform first tries to fill each buyer's choice set with freelancers that have been hired previously by that buyer. The remaining slots, if any, are filled randomly. The "Relationship priority + random" policy uses only the information of relationships, and focuses on vertical differentiation.

Policies Based on Multiple Observables.

• "*Weighted Average*". The "Weighted Average" policy uses information on both skills and relationships. The platform gives weights to skill matches and prior hiring records (i.e., ability to avoid the setup cost), where the coefficients are obtained from the demand estimation. Specifically,

$$WeightedAverage = \begin{cases} -0.09 \times (1 - R_f^c) + 0.92 \times \text{SkillsMatched}_f^c & \text{for Data Entry} \\ -1.16 \times (1 - R_f^c) + 1.89 \times \text{SkillsMatched}_f^c & \text{for Web Development} \end{cases} \tag{27}$$

Then the platform allocates freelancers according to the rank of such scores. Note, the platform does not utilize information on freelancers' quality, or bids.

[10] In computing this score, we leave the noisy signal out, and plug in demand estimates for coefficients.

- *"Relationship priority + skills"*. The platform first prioritize freelancers who have been hired previously by the buyer. If some slots remain empty, the platform fills the remaining slots with freelancers who have the most skills matched. This policy does not use demand estimates, and does not require the platform to know freelancers' quality either. Compared to "Skills", this policy demonstrates the impact of setup costs, and compared to "Relationship priority + random", this policy demonstrates the value of skills and exploration.

Evaluation. We take the same (cross section of) market, and simulate forward for 100 periods. The performance is measured by per-period total buyer welfare, where in each period, total buyer welfare is computed by the sum of buyers' payoffs from hiring the chosen freelancer. We take 10 simulation runs, and record the mean across all these simulations.

In implementation, we assume the platform offers a choice set of 5 freelancers to each buyer, and each freelancer can be allocated to no more than 2 choice sets. These capacity constraints are enforced by using a greedy type rule: If there are any conflicts, then the platform allocates the next "most prioritized" available freelancer, depending on the prioritization policy in each policy.

5.3. Results

	Data Entry	Web Development
Full Information Benchmark	1352.743	7854.623
Random	831.494	5481.757
Weighted Average	1026.337	6602.046
Relationship priority + skills	1017.237	6583.643
Relationship priority + random	1122.681	6782.193
Skills	869.885	5707.243

Table 3 Per-Period Total Buyer Welfare in the Dynamic Setting

Our results show that policies prioritizing relationships perform better than skill-based policies. Especially, pairing such prioritization with exploration generates even higher buyer welfare. The magnitude of such improvement depends on the complexity of jobs.

Simulation results show that total buyer welfare is mostly consistent with the ladder of information. The "Full Information Benchmark" policy, requiring the most information (including the unobserved freelancers' quality), achieves the highest total buyer welfare. The "Random" policy, not using any, has the worst performance. "Weighted Average" results in similar buyer welfare to "Relationship priority + skills", by requiring a similar level of information, but using them in a slightly different way. Further, by comparing the "Skills" policy to "Random" benchmark, we quantify the value of horizontal fit and flexibility of the platform, as well as the value added from quickly being matched on skills and avoiding repeat setups.

The "Relationship priority + skills" policy performs better than "Skills" in both markets. This is not surprising in markets like Web Development, since the saving from avoiding setup costs is substantial. Especially, the effect of the stickiness in relationships amplifies in the long run, because the long run facilitates the formation of such relationships. It is noteworthy, however, to observe that even in markets where the setup cost is low, like the Data Entry market, "Relationship priority + skills" still works better. The reason lies in the informational value of relationships on freelancers' quality. High-ability freelancers are more likely to be hired by the buyer before. Hence, by giving priorities to freelancers with relationships, the platform is able to pick high-ability freelancers without knowing their precise information on their qualities.

Another striking result is that "Relationship priority + random" performs even better than "Relationship priority + skills". This reverses the ladder of information. In fact, by using "Relationship priority + random" instead of "Relationship priority + skills", the platform is able to achieve a 18.9% increase in total revenue (measured using wage bills) on the platform in the Data Entry market, and is able to increase total revenue by 8.7% in the Web Development market.

To understand why, recall that freelancers' quality explains the most variation in buyers' choices. They wish to find and hire high-quality freelancers. A random allocation allows different freelancers to enter a buyer's choice set in different periods, and hence the buyer has the opportunity to vet each one, and explore on their quality. On the other hand, the "Relationship priority + skills" tends to allocate the same set of freelancers, who have high SkillsMatched, to the buyer, leading to less room for exploration.

Further, the randomness in allocation and the consequent exploration only helps when relationship is prioritized. With a prioritization in relationships, once a freelancer is hired, her quality information is preserved, since she enters the buyer's choice set in every later period. Hence, the buyer has the option to hire the same freelancer and obtain the same payoff, and at the same time keep exploring on other freelancers' quality premium, seeking an opportunity to find a better one and gain an even higher welfare. In other words, while the randomness allows the buyers to explore, the prioritization of relationship aids in buyer's exploitation in what she has explored.

Finally, it is worth noting that in market where setup cost is low, "Relationship priority + random" policy is particularly good, because the value from repeated transaction is less, but rather exploration through relationships is more important.

6. Concluding Remarks

The online environment has allowed markets and market participants to readily create, disseminate, collect, and evaluate troves of observable information to drive search and match. Yet, at the same time, online markets for short-term labor face growing complexity in tasks and the heady intricacy of ever-richer differentiation. Such developments call for better understanding how these markets value and assess difficult-to-measure worker attributes (demand side) and deliver closely tailored value efficiently (intermediaries and supply side). To this end, our data-driven study seeks to measure the marketplace value of information that buyers privately collect regarding worker ability and fit and to identify the most valuable information in each market we study.

We utilize an extensive dataset covering 1.2 million job postings (29 million applications and 744 thousand hires) and develop a novel estimation framework to deal with private information actively collected and held by participants throughout a decentralized marketplace. We find that the majority of hiring decision variation is in fact explained by marketplace participants' privately gleaned information, rather than by observable skills compatibilities or publicly available reputational ratings, in two of the most highly active (but quite different) marketplaces on our data sponsor's online platform.

Our methodology is sufficiently general that it could be applied to marketplaces which generate private information regarding heterogeneous choices. It relies on minimal structural assumptions, yet circumvents the econometric difficulties brought by substantial information being unobservable to the platform (and researchers) and consistently estimates market preferences.

We also discover heterogeneity in buyers' preferences across markets. In the Data Entry market, buyers value matches that flexibly accommodate needs in required skills, which is a central advantage delivered by two-sided matching platforms. The Web Development market, on the other hand, exhibits substantial setup costs, and hence repeat-dealing relationships are preferred.

References

Afeche, Philipp, Zhe Liu, Costis Maglaras. 2018. Ride-hailing networks with strategic drivers: The impact of platform control capabilities on performance. *Columbia Business School Research Paper* (18-19) 18–19.

Agrawal, Shipra, Navin Goyal. 2013. Thompson sampling for contextual bandits with linear payoffs. *Internat. Conf. Machine Learning*. 127–135.

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

Allon, Gad, Achal Bassamboo, Eren B Çil. 2017. Skill management in large-scale service marketplaces. *Prod. Oper. Management* **26**(11) 2050–2070.

Allon, Gad, Awi Federgruen, Margaret Pierson. 2011. How much is a reduction of your customers' wait worth? An empirical study of the fast-food drive-thru industry based on structural estimation methods. *Manufacturing & Service Oper. Management* **13**(4) 489–507.

Altonji, Joseph G, Charles R Pierret. 2001. Employer learning and statistical discrimination. *Quart. J. Econom.* **116**(1) 313–350.

Anderson, Eric T, Karsten Hansen, Duncan Simester. 2009. The option value of returns: Theory and empirical evidence. *Marketing Sci.* **28**(3) 405–423.

Arnosti, Nick, Ramesh Johari, Yash Kanoria. 2018. Managing congestion in matching markets. *Available at SSRN 2427960* .

Ata, Baris, Nasser Barjesteh, Sunil Kumar. 2019. Spatial pricing: An empirical analysis of taxi rides in new york city. Tech. rep., Working Paper.

Autor, David H. 2001. Wiring the labor market. *J. Econom. Perspectives* **15**(1) 25–40.

Autor, David H, Frank Levy, Richard J Murnane. 2003. The skill content of recent technological change: An empirical exploration. *Quart. J. Econom.* **118**(4) 1279–1333.

Bakos, J Yannis. 1997. Reducing buyer search costs: Implications for electronic marketplaces. *Management Sci.* **43**(12) 1676–1692.

Bandi, Chaithanya, Antonio Moreno, Zhiji Xu. 2017. The hidden costs of dynamic pricing: Empirical evidence from online retailing in emerging markets. *Working paper* .

Banerjee, Siddhartha, Ramesh Johari, Carlos Riquelme. 2015. Pricing in ride-sharing platforms: A queueing-theoretic approach. *Proceedings of the Sixteenth ACM Conference on Economics and Computation*. ACM, 639–639.

Bartel, Ann, Casey Ichniowski, Kathryn Shaw. 2007. How does information technology affect productivity? Plant-level comparisons of product innovation, process improvement, and worker skills. *Quart. J. Econom.* **122**(4) 1721–1758.

Bastani, Hamsa, Mohsen Bayati. 2019. Online decision-making with high-dimensional covariates. *Operations Research, forthcoming* .

Belavina, Elena, Karan Girotra. 2012. The relational advantages of intermediation. *Management Sci.* **58**(9) 1614–1631.

Berry, Steven, James Levinsohn, Ariel Pakes. 1995. Automobile prices in market equilibrium. *Econometrica* 841–890.

Besbes, Omar, Francisco Castro, Ilan Lobel. 2018. Surge pricing and its spatial supply response. *Columbia Business School Research Paper* (18-25).

Bhattacharjee, Rajat, Ashish Goel. 2005. Avoiding ballot stuffing in eBay-like reputation systems. *Proc. 2005 ACM SIGCOMM Workshop on Econom. Peer-to-peer Sys.*. ACM, 133–137.

Bimpikis, Kostas, Ozan Candogan, Daniela Saban. 2019. Spatial pricing in ride-sharing networks. *Operations Research* .

Bimpikis, Kostas, Wedad J Elmaghraby, Ken Moon, Wenchang Zhang. 2018. Managing market thickness in online B2B markets. *Working Paper* .

Buchholz, Nicholas. 2019. Spatial equilibrium, search frictions and efficient regulation in the taxi industry. *Working paper* .

Bugni, Federico A. 2010. Bootstrap inference in partially identified models defined by moment inequalities: Coverage of the identified set. *Econometrica* **78**(2) 735–753.

Cachon, Gerard P, Kaitlin M Daniels, Ruben Lobel. 2017. The role of surge pricing on a service platform with self-scheduling capacity. *Manufacturing & Service Oper. Management* **19**(3) 368–384.

Cappelli, Peter. 1995. Rethinking employment. *British J. Indust. Relations* **33**(4) 563–602.

Cappelli, Peter, David Neumark. 2004. External churning and internal flexibility: Evidence on the functional flexibility and core-periphery hypotheses. *Industrial Relations* **43**(1) 148–182.

Chen, Qi, Stefanus Jasin, Izak Duenyas. 2019. Nonparametric self-adjusting control for joint learning and optimization of multiproduct pricing with finite resource capacity. *Mathematics of Operations Research* **44**(2) 601–631.

Chen, Yiwei, Ming Hu. 2018. Pricing and matching with forward-looking buyers and sellers. *Rotman School of Management Working Paper* (2859864).

Chernozhukov, Victor, Han Hong, Elie Tamer. 2007. Estimation and confidence regions for parameter sets in econometric models 1. *Econometrica* **75**(5) 1243–1284.

Cui, Ruomeng, Jun Li, Dennis J Zhang. 2019. Reducing discrimination with reviews in the sharing economy: Evidence from field experiments on airbnb. *Management Sci.* (forthcoming).

Cullen, Zoë, Chiara Farronato. 2018. Outsourcing tasks online: Matching supply and demand on peer-to-peer Internet platforms. *Working Paper* .

Dang, Yan, Yulei Zhang, Hsinchun Chen. 2010. A lexicon-enhanced method for sentiment classification: An experiment on online product reviews. *IEEE Intelligent Systems* **25**(4) 46–53.

Degeratu, Alexandru M, Arvind Rangaswamy, Jianan Wu. 2000. Consumer choice behavior in online and traditional supermarkets: The effects of brand name, price, and other search attributes. *Internat. J. Res. Marketing* **17**(1) 55–78.

Feldman, Pnina, Jun Li, Hsin-Tien Tsai. 2018. Welfare implications of congestion pricing: Evidence from sfpark. *Ross School of Business Paper* (1330).

Feng, Yifan, Rene Caldentey, Christopher Thomas Ryan. 2018. Learning customer preferences from personalized assortments .

Filippas, Apostolos, John J Horton, Joseph Golden. 2017. Reputation in the long-run. *CESifo Working Paper Series* .

Forman, Chris, Anindya Ghose, Batia Wiesenfeld. 2008. Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Inform. Sys. Res.* **19**(3) 291–313.

Fradkin, Andrey. 2015. Search frictions and the design of online marketplaces. *MIT Working Paper* .

Fu, Bin, Jialiu Lin, Lei Li, Christos Faloutsos, Jason Hong, Norman Sadeh. 2013. Why people hate your app: Making sense of user feedback in a mobile app store. *Proc. 19th ACM SIGKDD Internat. Conf. Knowledge Discovery and Data Mining*. ACM, 1276–1284.

Gallino, Santiago, Antonio Moreno. 2018. The value of fit information in online retail: Evidence from a randomized field experiment. *Manufacturing & Service Oper. Management* **20**(4) 767–787.

Garg, Nikhil, Ramesh Johari. 2018a. Designing informative rating systems for online platforms: Evidence from two experiments Submitted.

Garg, Nikhil, Ramesh Johari. 2018b. Designing optimal binary rating systems. *arXiv preprint arXiv:1806.06908* .

Graves, Todd L, Tze Leung Lai. 1997. Asymptotically efficient adaptive choice of control laws in controlled Markov chains. *SIAM J. Control Optim.* **35**(3) 715–743.

Hagedorn, Marcus, Tzuo Hann Law, Iourii Manovskii. 2017. Identifying equilibrium models of labor market sorting. *Econometrica* **85**(1) 29–65.

Hansen, Lars Peter. 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 1029–1054.

Ho, Kate, Ariel Pakes. 2014. Hospital choices, hospital prices, and financial incentives to physicians. *Amer. Econom. Rev.* **104**(12) 3841–84.

Ho-Dac, Nga N, Stephen J Carson, William L Moore. 2013. The effects of positive and negative online customer reviews: do brand strength and category maturity matter? *J. Marketing* **77**(6) 37–53.

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

Hu, Bin, Ming Hu, Han Zhu. 2018. Surge pricing and two-sided temporal responses in ride-hailing. *Available at SSRN 3278023* .

Hui, Xiang, Maryam Saeedi, Zeqian Shen, Neel Sundaresan. 2016. Reputation and regulations: Evidence from eBay. *Management Sci.* **62**(12) 3604–3616.

Johari, Ramesh, Vijay Kamble, Yash Kanoria. 2018. Matching while learning. *arXiv preprint arXiv:1603.04549* .

Kabra, Ashish, Elena Belavina, Karan Girotra. 2019. Bike-share systems: Accessibility and availability. *Management Sci.* (forthcoming).

Kanoria, Yash, Daniela Saban. 2017. Facilitating the search for partners on matching platforms: Restricting agents' actions. *Stanford University Graduate School of Business Research Paper* (17-53).

Kim, Sang Won, Marcelo Olivares, Gabriel Y Weintraub. 2014. Measuring the performance of large-scale combinatorial auctions: A structural estimation approach. *Management Sci.* **60**(5) 1180–1201.

Lai, Tze Leung, Herbert Robbins. 1985. Asymptotically efficient adaptive allocation rules. *Adv. Appl. Math.* **6**(1) 4–22.

Li, Jun, Serguei Netessine. 2018. Higher market thickness reduces matching rate in online platforms: Evidence from a quasi-experiment. *Management Sci.* (forthcoming).

Moon, Ken, Kostas Bimpikis, Haim Mendelson. 2018. Randomized markdowns and online monitoring. *Management Sci.* **64**(3) 1271–1290.

Moreno, Antonio, Christian Terwiesch. 2014. Doing business with strangers: Reputation in online service marketplaces. *Inform. Systems Res.* **25**(4) 865–886.

Mudambi, Susan M, David Schuff. 2010. What makes a helpful review? A study of customer reviews on Amazon.com. *MIS Quart.* **34**(1) 185–200.

Mudambi, Susan M, David Schuff, Zhewei Zhang. 2014. Why aren't the stars aligned? an analysis of online review content and star ratings. *2014 47th Hawaii Internat. Conf. Sys. Sci.*. IEEE, 3139–3147.

Nosko, Chris, Steven Tadelis. 2015. The limits of reputation in platform markets: An empirical analysis and field experiment. Tech. rep., National Bureau of Economic Research.

Pakes, Ariel, Jack Porter, Kate Ho, Joy Ishii. 2015. Moment inequalities and their application. *Econometrica* **83**(1) 315–334.

Pallais, Amanda. 2014. Inefficient hiring in entry-level labor markets. *Amer. Econom. Rev.* **104**(11) 3565–99.

Romano, Joseph P, Azeem M Shaikh. 2008. Inference for identifiable parameters in partially identified econometric models. *J. Statist. Planning and Inference* **138**(9) 2786–2807.

Romano, Joseph P, Azeem M Shaikh. 2010. Inference for the identified set in partially identified econometric models. *Econometrica* **78**(1) 169–211.

Russo, Daniel, Benjamin Van Roy. 2014. Learning to optimize via posterior sampling. *Math. Oper. Res.* **39**(4) 1221–1243.

Stanton, Christopher T, Catherine Thomas. 2016. Landing the first job: The value of intermediaries in online hiring. *Rev. Econom. Stud.* **83**(2) 810–854.

Tadelis, Steven. 2016. Reputation and feedback systems in online platform markets. *Annual Rev. Econom.* **8** 321–340.

Taylor, Terry A, Erica L Plambeck. 2007a. Simple relational contracts to motivate capacity investment: Price only vs. price and quantity. *Manufacturing & Service Oper. Management* **9**(1) 94–113.

Taylor, Terry A, Erica L Plambeck. 2007b. Supply chain relationships and contracts: The impact of repeated interaction on capacity investment and procurement. *Management Sci.* **53**(10) 1577–1593.

Zhang, Huanan, Stefanus Jasin. 2018. Online learning and optimization of (some) cyclic pricing policies in the presence of patient customers. *Available at SSRN 3144018* .

Zheng, Fanyin, Pu He, Elena Belavina, Karan Girotra. 2018. Customer preference and station network in the London bike share system. *Columbia Business School Research Paper* (18-20).

Electronic Companion

Appendix A: Proofs

For notational convenience, we define the operator $M^{(c,c',f,f')}$ on characteristic x, that calculates the following sum of differences of x:

$$M^{(c,c',f,f')}x \triangleq (x_f^c - x_{f'}^c) + (x_{f'}^{c'} - x_f^{c'}). \tag{EC.1}$$

In each pair, reflecting revealed preference, the operator subtracts the x of the freelancer not chosen, from the chosen freelancer's x. In the rest of the electronic companion, we drop the superscript for brevity.

Using M, we can express inequality (9) for each pair $n \in N$ as:

$$M\Pi(\theta) \geq 0. \tag{EC.2}$$

A.1. Proof of Proposition 1

From (9) and (19), we have

$$H \times \Big(\beta_{Rep} \cdot M\text{PredictSuccess} + cost_{Setup} \cdot M(1-R) + \beta_{Skill} \cdot M\text{SkillsMatched} - Mp + M\epsilon\Big) \geq 0. \tag{EC.3}$$

Taking the expectation,

$$\mathbb{E}^{(c,c',f,f')}\left(H \times \Big(\beta_{Rep} \cdot M\text{PredictSuccess} + cost_{Setup} \cdot M(1-R) + \beta_{Skill} \cdot M\text{SkillsMatched} - Mp + M\epsilon\Big)\right) \geq 0. \tag{EC.4}$$

Further, from Assumption 1 and the law of iterated expectations, $\mathbb{E}(H\epsilon_{c,f}) = 0$, and thus $\mathbb{E}(HM\epsilon) = 0$. So

$$\mathbb{E}^{(c,c',f,f')}\left(H \times \Big[\beta_{Rep} \cdot M\text{PredictSuccess} + cost_{Setup} \cdot M(1-R) + \beta_{Skill} \cdot M\text{SkillsMatched} - Mp\Big]\right) \geq 0,$$

which completes the proof. $\square$

A.2. Proof of Proposition 2

Define the distance $d(b, A) = \inf_{a\in A} ||b - a||$ and the ϵ-expansion of a set A as $A^\epsilon = \{a \in A : d(a, A) \leq \epsilon\}$.

We first present a few lemmas before proceeding to the proof of the main result.

Lemma EC.1. *Let Θ be the parameter space which is compact. For any $\theta \in \Theta$, $||(\hat{m}(\theta))_- - (m(\theta))_-|| \xrightarrow{a.s.} 0$.*

Proof. Since the parameter space Θ is compact, it follows the Uniform Law of Large Numbers (ULLN) that $||(\hat{m}(\theta)) - (m(\theta))|| \xrightarrow{a.s.} 0$. And since $f(x) = x_- = \max\{-x, 0\}$ is a continuous function, the result holds. $\square$

Lemma EC.2. *For any $\epsilon > 0$, there exists $\delta > 0$ such that $\inf_{\theta\in(\Theta_0^\epsilon)^c} ||(m(\theta))_-|| > \delta$.*

Proof. Given the linearity of m, the identified set, Θ_0, is a polyhedron which is compact. Also, $||m_-||$ is a continuous function. Consider set $B = \{b \in \Theta : d(b, \Theta_0) \geq \epsilon/2\}$. It is immediate that B is compact and $(\Theta_0^\epsilon)^c = \{a \in \Theta : d(a, \Theta_0) > \epsilon\} \subset B$. By Weierstrass Theorem, $\min_{\theta\in B} ||(m(\theta))_-||$ is attained on B, and it is strictly positive (Otherwise, $\arg\min_{\theta\in B} ||(m(\theta))_-||$ is in Θ_0, since $(m(\theta))_- = 0$ if and only if $\theta \in \Theta_0$. This is contradictory to the definition of B). Let $\delta = \min_{\theta\in B} ||(m(\theta))_-||$, so $||(m(\theta))_-|| \geq \delta$ for all $\theta \in B$ and hence for all $\theta \in (\Theta_0^\epsilon)^c$. $\square$

LEMMA EC.3. *For any $\epsilon > 0$, there exists $\delta > 0$ such that $\inf_{\theta \in (\hat{\Theta}^\epsilon)^c} ||(\hat{m}(\theta))_-|| > \delta$.*

Proof. If $\min_{\theta\in\Theta} ||(\hat{m}(\theta))_-|| > 0$, the result holds. If $\min_{\theta\in\Theta} ||(\hat{m}(\theta))_-|| = 0$, $\hat{\Theta}$ is a compact set. Also notice that $||\hat{m}_-||$ is continuous. Hence the logic follows the proof in the previous lemma. □

Now we return to the main proposition. The proof is decomposed into two steps: (A) $\hat{\Theta} \subset \Theta_0^\epsilon$, and (B) $\Theta_0 \subset \hat{\Theta}^\epsilon$. If both statements are true, then by the definition of the Hausdorff distance, $d_H(\hat{\Theta}, \Theta_0) = \max\{\sup_{\theta\in\hat{\Theta}} d(\theta, \Theta_0), \sup_{\theta\in\Theta_0} d(\theta, \hat{\Theta})\} \le \epsilon$. Since ϵ is arbitrary, the result is proven.

Step (A): We first show that $\hat{\Theta} \subset \Theta_0^\epsilon$ for any ϵ, i.e., $\hat{\Theta} \cap (\Theta_0^\epsilon)^c = \emptyset$ with probability 1. So we only need to show: $\sup_{\hat{\Theta}} ||\hat{D}^{-1/2}\hat{m}(\theta)_-|| < \inf_{(\Theta_0^\epsilon)^c} ||\hat{D}^{-1/2}\hat{m}(\theta)_-||$.

Note, for $\theta \in (\Theta_0^\epsilon)^c$,

$$||\hat{D}^{-1/2}\hat{m}(\theta)_-|| = ||D^{-1/2}m(\theta)_- - (D^{-1/2}m(\theta)_- - \hat{D}^{-1/2}\hat{m}(\theta)_-)|| \tag{EC.5}$$

$$\ge ||D^{-1/2}m(\theta)_-|| - ||D^{-1/2}m(\theta)_- - \hat{D}^{-1/2}\hat{m}(\theta)_-||. \tag{EC.6}$$

By Lemma EC.2, for $\theta \in (\Theta_0^\epsilon)^c$, there exists some δ' such that $||D^{-1/2}m(\theta)_-|| > \delta'$. Also $||D^{-1/2}m(\theta)_- - \hat{D}^{-1/2}\hat{m}(\theta)_-|| \to 0$ because $\hat{D}$ is a consistent estimator of D and Lemma EC.1. Hence, for any ϵ, there exists δ'', if J is big enough, $\inf_{(\Theta_0^\epsilon)^c} ||\hat{D}^{-1/2}\hat{m}(\theta)_-|| \ge \delta''$.

On the other hand, $\sup_{\hat{\Theta}} ||\hat{D}^{-1/2}\hat{m}(\theta)_-|| \to 0$ as J grows. Pick $\theta' \in \Theta_0$, $\hat{m}(\theta')_- \to m(\theta')_- = 0$. So $\sup_{\hat{\Theta}} ||\hat{D}^{-1/2}\hat{m}(\theta)_-|| = \min_\Theta ||\hat{D}^{-1/2}\hat{m}(\theta)_-|| \le ||\hat{D}^{-1/2}\hat{m}(\theta')_-|| \to 0$.

Hence $\sup_{\hat{\Theta}} ||\hat{D}^{-1/2}\hat{m}(\theta)_-|| < \inf_{(\Theta_0^\epsilon)^c} ||\hat{D}^{-1/2}\hat{m}(\theta)_-||$. This finishes Step (A).

Step (B): We show that $\Theta_0 \subset \hat{\Theta}^\epsilon$ for any ϵ, i.e., $\Theta_0 \cap (\hat{\Theta}^\epsilon)^c = \emptyset$ with probability 1. We only need to show $0 = \sup_{\Theta_0} ||D^{-1/2}m(\theta)_-|| < \inf_{(\hat{\Theta}^\epsilon)^c} ||D^{-1/2}m(\theta)_-||$. The equality sign holds by definition. To proceed, we observe that for any $\epsilon > 0$, for J big enough, there exists δ'', such that

$$||D^{-1/2}m(\theta)_-|| = ||\hat{D}^{-1/2}\hat{m}(\theta)_- - (\hat{D}^{-1/2}\hat{m}(\theta)_- - D^{-1/2}m(\theta)_-)|| \tag{EC.7}$$

$$\ge ||\hat{D}^{-1/2}\hat{m}(\theta)_-|| - ||\hat{D}^{-1/2}\hat{m}(\theta)_- - D^{-1/2}m(\theta)_-||. \tag{EC.8}$$

By Lemma EC.3, for $\theta \in (\hat{\Theta}^\epsilon)^c$, there exists some δ' such that $||\hat{D}^{-1/2}\hat{m}(\theta)_-|| > \delta'$. Also $||\hat{D}^{-1/2}\hat{m}(\theta)_- - D^{-1/2}m(\theta)_-|| \to 0$ because $\hat{D}$ is a consistent estimator of D and Lemma EC.1. Hence, for any ϵ, there exists δ'', if J is big enough, $\inf_{(\hat{\Theta}^\epsilon)^c} ||D^{-1/2}m(\theta)_-|| \ge \delta''$.

This finishes Step (B). □

Appendix B: Procedures for Deriving Inference in Section 3

We layout steps of the estimation procedure. Without loss of generality, we discuss the lower bound of the confidence interval for the extreme value in the first dimension. For more details in econometric theory of this approach, see Pakes et al. (2015).

For the set estimator, we solve (24), where $\hat{D}$ is obtained by a two-step procedure: In the first step, we obtain a consistent estimator of θ (denoted $\tilde{\theta}$) by doing Minimum Distance without weighing (i.e., use the identity matrix as $\hat{D}$). Then we evaluate the sample variance of moments at the consistent estimator: $\tilde{D} = diag(\mathbb{V}(m(\tilde{\theta})))$. In the second step, we plug in $\tilde{D}$ (as $\hat{D}$) in (24), and solve the optimization problem.

For the estimation of confidence intervals, take the lower bound as an example.

1. Denote $\Omega(\underline{\theta})$ as the correlation matrix of moments (evaluated at the true lower bound), and $\Gamma(\underline{\theta})$ as the Jacobian matrix. Similar to the estimation of $\hat{\Theta}$, we obtain consistent estimators $\hat{\Gamma}$ ($\hat{\Omega}$) of $\Gamma(\underline{\theta})$ ($\Omega(\underline{\theta})$) using a two-step approach.

2. Simulate $Z^* \sim N(0, \hat{\Omega}_J)$.

3. For each simulated Z^*, we try to find τ that satisfies the following system of inequalities:

$$0 \leq \hat{D}^{-1/2}\hat{\Gamma}\tau + Z^* + r_J\left(\hat{D}^{-1/2}m(\underline{\hat{\theta}})\right)_+ \tag{EC.9}$$

where $r_J = o(\sqrt{J}/\sqrt{2\ln\ln J})$.

   - If a solution for τ exists, then minimize τ_1 (remember, this is an illustration for the first dimension).
   - If not, eliminate moments in the order of $\hat{D}^{-1/2}m_j(\underline{\hat{\theta}})$ starting from the largest value, until a solution exists.

4. Solve the stochastic LP with the remaining moments (indices denoted by s):

$$\underline{\tau}_1^* = \min\left\{\tau_1 : 0 \leq \hat{D}_s^{-1/2}\hat{\Gamma}_s\tau + Z_s^* + r_J\left(\hat{D}_s^{-1/2}m_s(\underline{\hat{\theta}})\right)_+\right\}. \tag{EC.10}$$

5. Repeat steps 2-4, and denote $\underline{\tau}_1^*(95)$ as the 95-th percentile of $\underline{\tau}_1^*$. Our estimate of the lower bound of the confidence interval is

$$\underline{\hat{\theta}}_1 - \frac{1}{\sqrt{J}}\underline{\tau}_1^*(95).$$

Appendix C: First-Stage Regression on SatisfiedRating

Appendix D: Test of Model Mis-specification vs. Sampling Error

We follow Ho and Pakes (2014). Denote $\Sigma(\hat{\theta})$ the variance-covariance matrix of $m(\hat{\theta})$, evaluated at the point estimates. We take simulation draws $Z_i \sim N(0, \Sigma(\hat{\theta}))$, and further calculate $||(Z_i)_-||$ and $||(m(\hat{\theta}))_-||$. The model is accepted if $||(m(\hat{\theta}))_-||$ is less than the 95-th percentile of $||(Z_i)_-||$.

Appendix E: Instrument Relevance

To show our instruments are relevant for wage bids, we regress wage bids on instruments (see "unrestricted" panels in Table EC.2). We observe sufficiently large F statistics (against the null hypothesis that the model is fitted by the mean wage), providing evidence of valid instruments.

Additionally, the change in commission structure provides orthogonal variation across different individuals. To do so, we set the coefficients of variables provided by the commission structure change, AfterChange×Relationship, and AfterChange×$\mathbb{I}$(Relationship), to zero, as restricted models. We further compute the F statistics of the unrestricted model against the restricted models. In Web Development, the F statistic is 21.49, and in Data Entry, the F statistic is 43.88. They are both large enough to support that the commission structure change indeed provides enough orthogonal variation for identification.

Appendix F: Robustness Checks for Demand Estimation

In this section, we present robustness checks for estimating equation (5) using moment inequalities. We consider the following different specifications:

1. Instead of using all relationships (R), we decompose all relationships into relationships with positive feedback ($PosR$) and relationships without positive feedback ($NonPosR$).

Table EC.1 Predicting SatisfiedRating

	Dependent variable:	
	SatisfiedRating	
	Data Entry	Web Development
score_avg	1.470***	0.914**
	(0.530)	(0.423)
score_avg_communication	−0.126	−0.056
	(0.120)	(0.101)
score_avg_cooperation	−0.380***	−0.208**
	(0.126)	(0.099)
score_avg_deadlines	−0.211	0.022
	(0.133)	(0.104)
score_avg_quality	−0.188	0.024
	(0.130)	(0.116)
score_avg_skills	−0.174	−0.188
	(0.138)	(0.120)
$\mathbb{I}$(relationship)	−2.426***	−2.498***
	(0.143)	(0.112)
$\mathbb{I}$(positive history)	4.242***	4.173***
	(0.157)	(0.122)
$\mathbb{I}$(negative history)	0.385	−0.314
	(0.411)	(0.416)
SkillsMatched	0.039***	−0.024***
	(0.012)	(0.007)
Constant	−1.509***	−1.882***
	(0.174)	(0.149)
Observations	32,249	60,185
R^2	0.075	0.067
χ^2 (df = 10)	1,835.781***	3,026.262***
Note:		*p<0.1; **p<0.05; ***p<0.01

2. Instead of an indicator of relationships (R), we use the numerical value of relationships ($numR$)

	Dependent variable:			
	Wage			
	Web Development		Data Entry	
	unrestricted	restricted	unrestricted	restricted
# Relationship	0.214**	0.227**	0.091	0.124**
	(0.106)	(0.097)	(0.087)	(0.060)
$\mathbb{I}$(Positive History)	−13.369***	−12.438***	−1.103***	−0.811***
	(0.518)	(0.490)	(0.172)	(0.152)
$\mathbb{I}$(Non-Positive History)	16.259***	16.823***	0.931***	1.105***
	(0.724)	(0.715)	(0.203)	(0.190)
AfterChange×Relationship	0.237		−0.048	
	(0.249)		(0.111)	
AfterChange×$\mathbb{I}$(Relationship)	3.079***		1.702***	
	(0.682)		(0.239)	
SkillsMatched	−0.375***	−0.378***	−0.892***	−0.891***
	(0.104)	(0.104)	(0.025)	(0.025)
# Skill Matches	0.149***	0.149***	0.415***	0.415***
	(0.027)	(0.027)	(0.011)	(0.011)
SatisfiedRating	40.308***	40.257***	2.988***	2.975***
	(1.217)	(1.218)	(0.315)	(0.315)
Constant	−5.484***	−5.451***	2.595***	2.601***
	(0.733)	(0.733)	(0.178)	(0.178)
Observations	113,201	113,201	87,197	87,197
R^2	0.010	0.010	0.025	0.024
Adjusted R^2	0.010	0.010	0.025	0.024
F Statistic	145.643***	186.961***	281.345***	360.149***
Note:				*p<0.1; **p<0.05; ***p<0.01

Table EC.2 Relevance of Instruments

3. Instead of using a percentage of skill matches (SkillsMatched, defined in (1)), we use *num*SkillsMatched, which is defined as

$$num\text{SkillsMatched} = \#\text{ of skill matches between the employer's requirement and freelancer's expertise.} \tag{EC.11}$$

We estimate the model using the aforementioned specifications, and results are shown in Table EC.3 and EC.4. In these specifications, the insight of the original models hold: In the Data Entry market, skill matches are significant, but not relationships. In the Web Development market, relationships are significant, but not skill matches.

Table EC.3 Robustness Checks: the Data Entry Market

	Original Model	Model 1	Model 2	Model 3
SatisfiedRating	0.42 $[-0.38, 1.18]$	0.4651 $[-1.0304, 3.6408]$	0.5072 $[-0.2270, 1.4171]$	0.3583 $[-0.4936, 0.9275]$
R	0.09 $[-0.25, 0.22]$			0.0660 $[-0.3666, 0.1427]$
$numR$			0.0553 $[-0.2198, 0.1172]$	
$PosR$		0.0650 $[-0.5973, 0.2870]$		
$NonPosR$		0.0313 $[-0.6482, 1.0395]$		
SkillsMatched	0.92^{**} $[0.56, 3.51]$	0.9070^{**} $[0.5990, 4.1016]$	0.9064^{**} $[0.5198, 4.0981]$	
numSkillsMatched				0.2010^{**} $[0.0454, 1.2934]$
Fixed Effect	Y	Y	Y	Y
N	87197	87197	87197	87197

Table EC.4 Robustness Checks: the Web Development Market

	Original Model	Model 1	Model 2	Model 3
SatisfiedRating	1.94^{**} $[0.66, 5.83]$	0.4156 $[-0.2823, 6.9466]$	2.0410^{**} $[0.2914, 5.2494]$	1.8030^{**} $[0.5261, 6.3545]$
R	1.16^{**} $[0.26, 2.11]$			1.1088^{**} $[0.2509, 2.2475]$
$numR$			0.3039^{**} $[0.1473, 0.5474]$	
$PosR$		0.3828^{**} $[0.0374, 0.5703]$		
$NonPosR$		-0.3497 $[-0.7853, 1.4075]$		
SkillsMatched	1.89 $[-1.54, 8.96]$	0.5168 $[-5.3504, 5.2221]$	0.8400 $[-4.3170, 6.3165]$	
numSkillsMatched				-0.6175 $[-2.3062, 0.0717]$
Fixed Effect	Y	Y	Y	Y
N	113201	113201	113201	113201

Appendix G: Estimating Freelancers' Quality

The estimation of freelancers' quality is built upon the demand estimation in previous sections, and relies on consistent estimators of demand elasticities obtained in the main model. It imposes more structural assumptions on the perception error, yet enabling us to infer the unobserved quality given the observed choices. The model is designed to allow for correlation between freelancer quality and bids.

G.1. Estimation Procedures

We first identify "high-type freelancers" and "low-type freelancers", by comparing the average bid for a freelancer to the median of freelancers' average bids. A freelancer is characterized as a "high-type freelancer" if the average bid he asks for exceeds the median of freelancers' average bids. Otherwise, he is characterized as a "low-type freelancer".

To allow for correlations between freelancers' quality and their bids, we assume high type freelancers' quality and low type freelancers' quality are drawn from two different distributions, with the same variance but different means. Formally, the quality of a high type freelancers is denoted as $\eta_{n_H} \sim N(\lambda_H, \rho^2)$, and the quality a low-type freelancers is denoted as $\eta_{n_L} \sim N(\lambda_L, \rho^2)$.

In this section, instead of using pairs (c, c', f, f') as is in Section 3, we construct pairs of freelancers for each employer' decision instance, where one of the freelancers in the pair is hired, and the other is not. For pair k, the employer could compute the difference in utilities Z_k of hiring freelancer 1 vs. hiring freelancer 2. Let f_{ki} denote freelancer i in pair k and c_k denote the employer in pair k, we have:

$$Z_k = X_{k1} + \eta_{f_{k1}} - (X_{k2} + \eta_{f_{k2}}) + \epsilon_k, \tag{EC.12}$$

where

$$X_{ki} = \beta_{Rep}\text{PredictSuccess}_{f_{ki}}^{c_k} - cost_{Setup}(1 - R_{f_{ki}}^{c_k}) + \beta_{Skill}\text{SkillsMatched}_{f_{ki}}^{c_k} \tag{EC.13}$$

for $i \in \{1, 2\}$. As before, in the model there is an idiosyncratic error in the utility function. We denote in pair k the idiosyncratic error is $\epsilon_k \sim N(0, \sigma^2)$.

Observing the actual hiring decision in each pair, we are able to form a likelihood function using the model described above. We also complement the analysis in pairs with observations that: (1) For openings in which no freelancer is hired, the utility of hiring each applicant to this opening must all be smaller than the outside value; (2) The utility of hiring a chosen freelancer must be greater than the outside value. These facts are also modeled in the likelihood function.

The parameters of interest, $(\lambda_H, \lambda_L, \rho, \epsilon)$ could be estimated using the Maximizing Likelihood Estimation (MLE). In practice the estimation is done through a Monte-Carlo Expectation-Maximization (MCEM) algorithm. After parameters are estimated, one can obtain a posterior distribution of quality for each freelancer, given all hiring decisions made in pairs that he is involved.

We now summarize the setup.

- Complete data D:

1. Fixed effects: high-type freelancers: $\eta_{n_H} \sim N(\lambda_H, \rho^2)$, low-type freelancers: $\eta_{n_L} \sim N(\lambda_L, \rho^2)$
2. The actual difference in each pair: $Z_k = X_{k1} + \eta_{j_{k1}} - (X_{k2} + \eta_{j_{k2}}) + \epsilon_k$ where $\epsilon_k \sim N(0, \sigma^2)$

3. Actual utility for freelancers (in openings where nobody is hired): $U^1_{m_1} = X_{m_1} + \eta_{j_{m_1}} + \epsilon_{m_1}$ where $\epsilon_{m_1} \sim N(0, \frac{1}{2}\sigma^2)$
4. Actual utility for freelancers hired: $U^2_{m_2} = X_{m_2} + \eta_{j_{m_2}} + \epsilon_{m_2}$ where $\epsilon_{m_2} \sim N(0, \frac{1}{2}\sigma^2)$

- Observed data:

1. Y_k: whether the first freelancer in a pair is hired:

$$Y_k = \mathbb{I}\Big(X_{k1} + \eta_{j_{k1}} - (X_{k2} + \eta_{j_{k2}}) + \epsilon_k \geq 0\Big), \tag{EC.14}$$

where X is known

2. Openings where no freelancer is hired: $\{m_1 : U^1_{m_1} < 0\}$
3. Decision instances where a freelancer is hired: $\{m_2 : U^2_{m_2} \geq 0\}$

- Parameters to be estimated: $\theta = (\lambda_H, \lambda_L, \rho, \sigma)$
- Complete data likelihood (suppose latent variables are known):

$$\begin{aligned}\mathcal{L}_Y(\theta) = p(\eta, Z, U^1, U^2; \theta) = &\prod_{n=1}^{N_H} \frac{1}{\sqrt{2\pi}\rho} \exp\left(-\frac{1}{2\rho^2}\Big(\eta_{n_H} - \lambda_H\Big)^2\right)\\ &\times \prod_{n=1}^{N_L} \frac{1}{\sqrt{2\pi}\rho} \exp\left(-\frac{1}{2\rho^2}\Big(\eta_{n_L} - \lambda_L\Big)^2\right)\\ &\times \prod_{k=1}^{K} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}\Big(Z_k - \big((X_{k1} + \eta_{j_{k1}}) - (X_{k2} + \eta_{j_{k2}})\big)\Big)^2\right)\\ &\times \prod_{m_1=1}^{M_1} \frac{1}{\sqrt{2\pi}\frac{\sqrt{2}}{2}\sigma} \exp\left(-\frac{1}{2\left(\frac{\sqrt{2}}{2}\sigma\right)^2}\Big(U^1_{m_1} - X_{m_1} - \eta_{j_{m_1}}\Big)^2\right)\\ &\times \prod_{m_2=1}^{M_2} \frac{1}{\sqrt{2\pi}\frac{\sqrt{2}}{2}\sigma} \exp\left(-\frac{1}{2\left(\frac{\sqrt{2}}{2}\sigma\right)^2}\Big(U^2_{m_2} - X_{m_2} - \eta_{j_{m_2}}\Big)^2\right)\end{aligned} \tag{EC.15}$$

G.2. Estimation Results

We report estimates of freelancers' quality. Table EC.5 displays estimators of parameters in the model.

	Data Entry	Web Development
$\hat{\lambda}_H$	6.6486	22.6080
$\hat{\lambda}_L$	3.5148	9.4115
$\hat{\rho}$	8.4816	35.8761
$\hat{\sigma}$	1.3463	5.0098

Table EC.5 Estimates of freelancers' quality

In our model, we allow freelancers who ask for higher prices (high-type freelancers) and freelancers who ask for lower prices (low-type freelancers) to differ in their quality. Without imposing any assumptions on the difference, we estimate the mean of high type freelancers' quality, λ_H, and the mean of low type freelancers' quality λ_L. Any differential in $\hat{\lambda}_H$ and $\hat{\lambda}_L$ provides evidence of correlation between prices asked and freelancers' quality.

We find from estimation results, that a positive correlation between prices and quality exists. In both markets, $\hat{\lambda}_H > \hat{\lambda}_L$, implying that a freelancer asking for higher bids is associated with a higher freelancer quality in general. There could be a number of reasons for the observed phenomenon. It is possible that freelancers signal their high quality by asking for higher prices. Or, high quality freelancers pay higher efforts for the same job, and hence they ask for higher prices for compensation. Although the exact reason is not identifies, the estimation results strongly supports the quality differential between the two groups of freelancers.

$\hat{\rho}$ reports the estimator of the standard deviation of the distribution of freelancers' quality. A higher value of $\hat{\rho}$ indicates a more dispersed distribution of freelancer quality. We find that in the Web Development market, quality is more dispersed. The dispersion also reassures the value of obtaining quality information to employers, since it could be very different across different freelancers.

$\hat{\sigma}$ reports the variation in the idiosyncratic error in the hiring decision marking process. Recall, if a freelancer with good ratings, skill matches and relationships is not hired, the reason must lie in the low quality and/or the idiosyncratic error. Which factor contributes more depends on their relative magnitudes. Estimation results show that, the freelancer's quality is the dominating factor to explain such inconsistency between actual choices and observed characteristics. We find $\hat{\sigma}$ to be much greater than $\hat{\rho}$ in both markets, implying a higher variability in freelancer's quality than in idiosyncratic error.

Appendix H: Performance of Allocation Policies in the Static Setting

	Data Entry	Web Development
Full Information Benchmark	1352.847	7606.141
Random	828.727	5453.069
Weighted Average	888.480	5680.312
Relationship priority + skills	927.166	5706.584
Relationship priority + random	906.790	5569.320
Skills	868.648	5507.059

Table EC.6 Per-Period Total Employer Welfare in the Static Setting

We conduct simulation analysis in a static setting. In the static scheme, given the (cross section of) market, we simulateone period ahead. The simulation is done for 100 times. We measure the performance by the mean of the total employer welfare over these 100 simulations. Results are shown in Table EC.6.

Similar to results in the dynamic scheme, here in the static setting, the ladder of information still roughly holds. Also, "Relationship + Random" still outperforms "Skills", even in markets where setup cost is low. Again, this is because of the correlation between high-bandwidth information and relationships. However, "Relationship + Random" is not as good as "Relationship + skills", given the limited ability of exploration in the static setting. This result again highlights the importance of exploration in such online marketplaces.