Estimation of Beauty Contest Auctions

Hema Yoganarasimhan *

University of California, Davis

November 7, 2013

Abstract

Beauty contests are auction mechanisms used to buy or sell differentiated products. Beauty contests are widely used in procuring welfare-to-work projects, freelance services, in selling online ads, in real-estate transactions, and in hiring, dating/marriage decisions. Unlike price-only auctions, beauty contests have no closed-form bidding strategies and suffer from non-multiplicatively separable unobserved auction heterogeneity, which makes their estimation challenging. To address these challenges, we formulate beauty contests as incomplete information games and present a two-step method to estimate them. A key contribution of our method is its ability to account for common-knowledge auction-specific unobservables using finite unobserved types. We show that unobserved auction types and distributions of bids are nonparametrically identified and recoverable in the first step using a nonparametric EM-like algorithm, and this can then be used in the second step to recover cost distributions. We present an application of our method in the online freelancing context. We find that seller margins/marketpower in this marketplace are around 15%; and that not accounting for unobserved heterogeneity can significantly bias estimates of costs in this setting. Based on our estimates, we run counterfactual simulations and provide guidelines to managers of freelance firms.

1 Introduction

1.1 Beauty Contest Auctions

Beauty contest auction is a procurement mechanism where the auctioneer does not specify an allocation rule. Beauty contests are typically used in differentiated product settings, where considerations other than price are of importance to the buyer. In beauty contests, sellers submit multidimensional bids (*e.g.*, price, reputation, speed of delivery), and the buyer awards the contract to a seller of his choice. Beauty contests are distinct from scoring auctions, where a deterministic scoring rule, specified by buyers and known to sellers, is used to decide the outcome. This paper focuses on beauty contests where a seller is endowed with a set of attributes and then chooses an optimal bid given her own attributes and her beliefs on the buyer's allocation rule.

The moniker 'beauty contest' comes from beauty pageants, where the winner is chosen by a selection committee that does not announce any scoring or selection rule *a priori*. While beauty is an important attribute, contestants realize that other attributes matter too, *e.g.*, compassion and general knowledge. Auction mechanisms that shares these features of beauty pageants are usually referred to as beauty contests. See Klemperer (2000), Janssen (2002), and Klemperer (2002) for detailed descriptions of beauty contests and discussions on the relative merits of beauty contests *vs.* traditional auctions.[1]

Beauty contests are extensively used in both private and public sector procurement. Governments use them to procure welfare-to-work projects (Bruttel, 2004), sell 3G licenses (*e.g.*, Spain, Sweden, Bangladesh), and in military contracting. In the private sector, beauty contests are used to procure television franchises (Cabizza and Fraja, 1998), and are used by small businesses to procure freelance programmers (Yoganarasimhan, 2013). In fact, the popular Google Adwords auction and Facebook ads auctions, which generate over $40 billion dollars of revenues (Google, 2013; SEC, 2012), can also be interpreted and modeled as beauty contests since these are multi-attribute auctions where the allocation rule of the auctioneer is not observable to the seller or researcher. Apart from these obvious examples, any setting in which an agent invites bids from a set of discrete alternatives and picks an optimal option, without using pre-specified allocation rules, can be interpreted as a beauty contest, *e.g.*, real-estate bidding, hiring or employment transactions, and marriage/dating decisions.

1.2 Online Freelance Marketplaces

A popular and growing application of beauty contest auctions is in online freelancing, which is the focus of our work. Online freelance marketplaces are websites that match buyers of electronically deliverable services with sellers or freelancers. Buyers with procurement needs invite bids from freelancers. Then the buyer chooses a winner based on her discretion, and in doing so she may trade off sellers' reputations, bid prices, and other attributes; *i.e.*, the lowest priced bidder is not the default winner. The site generates revenues through percentage commissions from winning sellers. The most popular freelance marketplaces are Elance, Guru, vWorker, ODesk, and Freelancer, and the most popular categories of jobs are web development, programming, writing,

---

[1]Procurers often prefer beauty contests over scoring auctions because of two reasons. First, it is costly in time and effort for buyers to spell out optimal scoring rules, especially when there are a large number of bid dimensions. Second, beauty contests do not bound buyers to a scoring rule. This insures them against changes in own preferences (*e.g.*, modifications in project specifications) and environmental conditions (*e.g.*, changes in economic conditions) in the time between the announcement of the scoring rule and the final decision. On the flip side, because beauty contests lack transparency, they can suffer from agency problems if the procurer is a middle-man, whose objective function includes factors irrelevant to the maximization of profits from the procurement. Klemperer (2002) argues that beauty contests are plagued by the perception, if not reality, of corruption and favoritism.

translation, design, and multimedia (Kozierok, 2011).

Online freelancing has grown tremendously in the last few years, with industry revenues now topping $360 million dollars (Morgan, 2011). This surge can be attributed to two main factors. First, technological innovations such as electronic deliverability of jobs and fast Internet connections have increased the supply of jobs that can be performed by freelancers. Second, online freelance marketplaces offer a low cost way for geographically distant players to trade, especially since there is an abundance of unemployed skilled workers in emerging economies (*e.g.*, Indian subcontinent, Eastern Europe), and a healthy demand for skilled workers in developed countries, where local labor is expensive. In fact, the majority of auctions in our data (over 82%) are posted by buyers from developed countries, while the majority of bids (over 54%) belong to developing countries (see Table 2).

Because online freelancing is playing an increasingly important role in the global labor market, understanding this market is of paramount importance to many players – researchers are interested in market primitives such as seller costs, managers are looking to optimize their procurement mechanisms and improve profits, and policy-makers want to quantify the impact of policy changes such as taxing offshore procuring.

Broadly speaking, there are three central questions of interest to researchers and managers in this area – (1) What is the equilibrium bidding strategy of sellers and the equilibrium allocation rule used by buyers? (2) What is the underlying distribution of seller costs, and can we estimate it without making any parametric assumptions on sellers' types or buyers' unobserved preferences? Information on seller costs is fundamental to understanding the marketplace and to answer important questions such as – how much marketpower do sellers have and how competitive is the market? (3) What are the managerial implications of modifying various aspects of the market? For example, who is more valuable in this two-sided market – buyers or sellers? Should a freelance site invest in attracting sellers from developing countries or should it focus on local sellers? In this paper, we present a structural framework to estimate beauty contest auctions and answer the above questions for a specific freelance marketplace.

1.3 Modeling and Estimation Challenges

There are three primary challenges in modeling and estimating beauty contest auctions. First, the unobservability of buyers' allocation rule is problematic because sellers' equilibrium strategy is a function of their beliefs on buyer behavior. Hence, in addition to modeling the seller side, we need to infer buyer behavior from the data. Second, we don't have a closed form solution for sellers' bidding strategy. The multidimensionality of bids and the unobservability of buyers' allocation rule implies that the first order condition of seller profit is no longer a simple analytical relationship between the distribution of observed bids and seller's unobserved cost. Hence, information on the distribution of bid prices is not sufficient to infer seller costs, as is common in the auctions literature (Guerre et al., 2000). Third, there may exist auction specific variables that are visible to sellers, but not to the researcher, *i.e.*, auction specific unobserved heterogeneity. Not accounting for such auction specific unobservables can lead to biased estimates of seller costs. For example, a freelance job with high unobserved difficulty will invite high bids; a model without unobserved heterogeneity will mistakenly attribute the high bids to high underlying seller costs, when in fact, it should be attributed to the unobserved difficulty of the job.

The problem of accommodating for auction specific unobservables is similar to the well-known and challenging problem of accommodating market-specific unobservables in discrete games. See Aguirregabiria and Mira (2007); Ellickson and Misra (2008); Arcidiacono and Miller (2011); Pakes et al. (2011); Ellickson et al.

(2013) for discussions on the recent developments in this area. An important difference between continuous choice games such as beauty contests and discrete games, is that common-knowledge unobservables cannot be assumed to enter agents' utility functions in additively/multiplicatively separable formats, making estimation somewhat more challenging, an issue we discuss in detail in §5.4.

1.4 Our Approach

In this paper, we formulate beauty contest auctions as two stage games of strategic interaction with incomplete information, and present a two step estimation method that can address the three challenges and recover the underlying distributions of seller costs. Our method not only builds on the literature on nonparametric estimation of auctions (Guerre et al., 2000; Athey and Haile, 2007), but also leverages the large and growing literature on two step estimation of games (Aguirregabiria and Mira, 2007; Bajari et al., 2007; Pesendorfer and Schmidt-Dengler, 2008; Bajari et al., 2010). While two step estimators are computationally light and provide a clean solution to the multiple equilibria problem common in game settings, they are also known to suffer from a significant drawback – they cannot allow for common knowledge unobservables due to their limited information structure and reliance on a nonparametric first stage.

However, we show that it is possible to accommodate common knowledge unobservables through finite unobserved types within two step methods in auction settings. In sealed bid auctions with independent bidders, we show that joint distributions of seller attributes and bid prices for unobserved auction types are nonparametrically identified. We propose an estimation method similar in spirit to Arcidiacono and Miller (2011) to accommodate common knowledge unobservables for continuous state space problems. We employ a kernel-smoothed nonparametric EM-like algorithm (similar to that in Benaglia et al. (2009a)) to recover nonparametric estimates of underlying bid distributions as functions of both observable attributes and unobservable auction types, as well as the population distribution of unobserved types in the first step itself. At this stage, we also estimate the buyers' allocation rule. In the second step, we obtain a numerical estimate of the expected probability of winning the auction for a given seller using the first stage estimates. These are then plugged into the first order condition of seller's maximization problem to infer the seller's private cost. Armed with the inferred private cost for each seller, we then estimate the nonparametric distribution of seller costs as functions of both observed and unobserved state variables.

Our method has many advantages. First, it can be used to estimate auctions that don't have pre-specified allocation rules. Second, it is computationally simple and does not require us to solve for equilibrium bidding strategies, which is a challenging task in this complex setting. Third, it does not impose any optimality assumptions on the buyer or third-party site which conducts the auction. This allows us to use the estimated costs to run counterfactuals and offer recommendations to improve site profits. Fourth, it does not require any parametric assumptions on seller types, seller attributes, or bid distributions. In auctions with sufficiently small state spaces, the entire estimation procedure can be nonparametric.

Importantly, our method provides an elegant solution to the problem of auction specific unobserved heterogeneity. In general, allowing for auction specific unobservables is difficult even in first price auctions. The key challenge lies in separately identifying the impact of two unobservables in a bid – the unobservable cost of the seller and the unobservable auction heterogeneity. While Krasnokutskaya (2011) has recently proposed a deconvolution method to account for auction specific unobservables in first price auctions, her method requires

a multiplicative separability assumption, which is not valid in our setting (see §5.4 for details). Moreover deconvolution methods are difficult to apply in many settings. On the other hand, our nonparametric EM-like algorithm is relatively simple to implement, does not require the multiplicative separability assumption, and can be used to accommodate unobserved heterogeneity in a vast range of auction settings, including the commonly studied first price auction.

1.5 Findings and Contribution

We estimate the distribution of seller costs in a prominent freelance marketplace and present the following findings. First, we infer seller costs for auctions with a non-binding maximum bid of $500. We show that cost differences across freelancers can be explained by heterogeneity in geographic location, past experience on the site, previous interactions with the buyer, and by unobserved auction heterogeneity. Specifically, we show that there are three unobserved types of auctions, which we call Low, Medium, and High, respectively, with the following distribution: Low type – 17.07%, Medium type – 46.47%, and High type – 36.45%. Low type auctions have low unobserved costs and the inferred cost distribution for these auctions is first order stochastically dominated (FOSD) by that of Medium type auctions that have moderate costs, which in turn is FOSD by the distribution of costs for High type auctions. The median cost for the three types of auctions are: Low type – $195.55, Medium type – $322.49, and High type – $365.14. The dollar amount of the margins in this marketplace is not very high, with average percentage margins around 15%, *i.e.*, the marketplace is quite competitive, and consequently sellers do not enjoy much marketpower.

Second, we find that not accounting for unobserved heterogeneity can significantly bias the estimates of seller costs. The Relative Horizontal Distance between inferred CDFs of cost distributions for the three types with and without unobserved heterogeneity are: Low – 90.83%, Medium – 7.94%, and High – 17.79%. Thus, without unobserved heterogeneity, there is significant overprediction of costs for Low type auctions, a good amount of underprediction of costs for High type auctions, and some overprediction for Medium type auctions. Overall, these results affirm the value of controlling for auction specific unobserved heterogeneity.

Third, using our seller costs estimates, we model endogenous buyer entry and infer the drivers of buyers' entry decision and their entry probabilities. We find that entry probabilities are a little over 0.5 and are significantly different across auction types. For example, buyers with Low type auction-jobs are 19% more likely to enter an auction compared to those with High type jobs.

Fourth, we conduct three counterfactuals to examine the impact of policy changes on site revenues. In the first experiment, we address the most important question that managers of two-sided markets face – who is more valuable to the marketplace, buyers or sellers? While an increase in the supply of either buyers or sellers always has a positive impact on site profits, we find that for the same percentage increase, buyers are preferable to sellers. For example, a 40% increase in seller supply only leads to 5.5% revenue increase, whereas in the case of buyers it leads to nearly five times as much increase in revenue (25%). Hence, when growing their business, managers of freelance sites should focus on the buyer side. In the second experiment, we examine the relative importance of the three unobserved auction types. Since there is a significant difference between High and Low type auctions (over $150), an important question that a manager faces is the following: should the site encourage high value auctions and discourage low and medium value auctions to improve commission revenues? Interestingly, we find that all three auction types are almost equally valuable. While High type

auctions invite higher bids and bring in larger commissions, they also clear at lower rates. Thus, their overall contribution is not much higher than Low/Medium type auctions, which clear at lower prices, but do so with higher frequency. So the site should not promote one type of auction at the expense of others. In the third experiment, we infer the relative value of sellers from different geographic regions. This is a key question for freelance sites that position themselves as intermediaries that match buyers from developed countries with sellers from developing countries. Interestingly, we find that sellers from the Indian sub-continent are the least valuable and those from developed English-speaking countries are most valuable to the site. Thus even sites that focus on offshoring may derive value from attracting local sellers.

In sum, our paper makes three key contributions to literature. First, from a methodological perspective, we provide an empirical framework to model and estimate beauty contest auctions. Our framework is fairly general and can be adapted to suit a large class of auction problems that lack observable buyer allocation rules and closed form solutions to seller strategies. Importantly, our method can handle non-multiplicatively separable auction specific unobservables. Second, from a substantive perspective, we derive the cost distributions of sellers in a prominent freelance marketplace and show that freelancing is a very competitive industry with low seller margins. We also show that not accounting for unobservables in this market can significantly bias estimates of seller costs. Third, from a normative perspective, our work offers guidelines to managers of freelance sites. Our framework can be used as a tool to evaluate the impact of policy changes, such as the value of changing the auction design or attracting a different distribution of sellers. Finally, we note that online freelancing is a large and growing industry that contributes significantly to offshore outsourcing of jobs, thereby putting it at the center of the raging debate on the impact of outsourcing on local economies (Mankiw and Swagel, 2006; Lacity and Rottman, 2008). While a complete analysis of the costs and benefits of offshoring is outside the scope of this paper, our estimates can serve as inputs in the larger cost-benefit analysis that policy-makers must undertake to settle this debate.

2   Related Literature

Our paper relates and contributes to many broad streams of literature in marketing and economics.

First, our paper relates to the theoretical literature on procurement of differentiated products using auction mechanisms. Starting with Che (1993), many researchers have considered multidimensional scoring auctions (Branco, 1997; Asker and Cantillon, 2008, 2010), where sellers submit bids on both quality and price. The focus of this literature is mechanism design, *i.e.*, it seeks to identify the auction mechanism that maximizes buyers' expected profits. While our setting is similar to those used in the above papers, we cannot import the closed-form solutions from them into our empirical analysis for two reasons. First, in our case, quality (and other pay-off relevant attributes) cannot be modified by sellers, *i.e.*, sellers can only optimize their bid price. Second, our setting is a beauty contest, not a scoring auction, and buyers' preferences are not perfectly observable to us.

Second, our paper contributes to the literature on nonparametric estimation and identification of auction models that aim to infer the distributions of sellers' private costs from observed outcome data. This approach was developed by Guerre et al. (2000) in the context of first price auctions with Independent Private Values (IPV) symmetric bidders. It utilizes the relationship between the equilibrium distributions of observed bids and a seller's private cost and her bid, in order to back out the cost of each seller. The original method has been augmented in many directions; for example, Li et al. (2002) extend it to Affiliated Private Values (APV) setting, Li et al. (2003) allow for models with conditionally IPV, Campo et al. (2003) consider asymmetric bidders

and APV, and Hong and Shum (2002), Haile et al. (2003), Guerre et al. (2009), and Krasnokutskaya (2011) discuss methods to control for unobserved auction heterogeneity. See Athey and Haile (2007) for a detailed overview of the past work on nonparametric approaches to auctions.

Third, our paper relates to the literature on the estimation of strategic games in marketing – supermarket entry models (Singh and Zhu, 2008; Datta and Sudhir, 2011; Orhun, 2013), demand estimation with social interactions and joint consumption (Hartmann, 2010; Narayanan, 2013), product introductions (Draganska et al., 2009), network effects (Shriver, 2011), and pricing strategies (Ellickson and Misra, 2008; Ellickson et al., 2012). See Ellickson and Misra (2011) for an excellent survey of this literature.

Fourth, it relates to the broader literature on Finite Mixture Models and Hidden Markov Models. Starting with Dempster et al. (1977), researchers in a variety of fields have employed FMMs to accommodate latent unobserved heterogeneity. See Dempster et al. (1977); McLachlan and Peel (2004) for general discussions and Kamakura and Russell (1989); Chintagunta et al. (1991); Allenby and Rossi (1998); Liechty et al. (2003); Moon et al. (2007); Netzer et al. (2008); Ansari et al. (2012) for marketing applications.

Fifth, our paper relates to the literature on online auctions. Bajari and Hortacsu (2004) provide a detailed discussion of the recent advances in this area and the remaining challenges. More recently, the following marketing papers have also made significant contributions to this literature. Zeithammer (2006, 2007) examine the dynamics of optimal bidding strategies and buyer behavior with forward-looking sellers. Zeithammer and Adams (2010) investigate the validity of the ubiquitous assumption that online sealed bid auctions are strategically equivalent to second-price auctions. Yao and Mela (2008) consider a model of both seller and buyer behavior and estimate the impact of varying commission rates and the value of sellers to the marketplace. Of course, all these papers pertain to price-only auctions, not beauty contests.

Finally, our paper relates to a small, but growing literature on procurement auctions without pre-specified allocation rules (Jap and Haruvy, 2008; Haruvy and Jap, 2013). Our paper closely relates to Yoganarasimhan (2013), who presents a dynamic structural framework to quantify buyers' valuations of seller reputations. There are three key differences between the two papers. First, the focus there is on building a partial equilibrium model of buyer's optimization problem taking the sellers' side as given. Here, we aim to uncover the distribution of private costs by modeling both sellers' strategic bidding behavior and buyers' choice decisions in a static setting. Thus, here we forgo dynamics to solve a full-equilibrium model. Second, the technical challenges that the two papers address are significantly different – in the former the key issue is controlling for dynamic selection within auctions, whereas here the main challenge is controlling for unobserved auction heterogeneity in sellers' bidding behavior. Third, because we have a full equilibrium model here, we can answer a broader set of questions that Yoganarasimhan (2013) cannot, *e.g.*, what is the value of growing the buyers' side of the market relative to growing the sellers' side.

3 Setting and Data

Our data comes from one of the leading online freelance firms in the 2006–2010 time frame (Morgan, 2011). Site membership is free and there are no fees for either posting auctions or for bidding. The site is mainly technology-oriented and the majority (over 80%) of auctions fall under the Information Technology (IT) services category. While the site hosts many types of auctions, all of them are sealed-bid, *i.e.*, sellers have no information on other bids received by the buyer. Our data comprises of 4002 auctions posted from January

1st 2006 to December 31st 2010 that have a non-binding maximum bid of $500.

The site employs a feedback-based reputation mechanism. After each transaction, buyers and sellers are allowed to rate each other on a symmetric numeric rating scale of 1-10 (and optional text comments). A rating of 1 stands for very bad and 10 for excellent. The site has taken some good measures to ensure the robustness of its reputation system. First, after a trade, both the buyer and seller are given a fixed time period to rate each other, after which they lose the right to rate. The ratings are revealed publicly only after both parties have rated each other. If one of the parties fails to turn in its feedback, then the other party's rating is revealed only after the fixed time period, at which point the delinquent party cannot retaliate. Second, if there is a dispute following the trade and the case goes into arbitration, both parties lose the right to rate each other, though the neutral arbiter may rate either or both players. The site enjoys a high feedback rate compared to other auction sites such as eBay, where feedback rate is barely 50% (Resnick and Zeckhauser, 2002). Over 90% of buyers have rated their sellers, and more than 70% of sellers have rated their buyers.

The site charges a 15% commission on the transaction amount in return for its services, which is paid by the winning bidder. For example, if a seller with a bid of $100 wins a project, the buyer escrows $100 with the freelance site, and after the project is completed, the site releases $85 to the winning seller.

We now describe the timeline of the procurement auctions in our dataset in detail.

- **Stage 1:** A buyer with a procurement need initiates an auction, by specifying a project title and a short description of the project. The description usually consists of information on the deliverables and the programming skills necessary to perform the job. If the buyer wants to provide more information, then he may also include a project-attachment that describes the project in greater detail. He can also specify a deadline for project delivery, *i.e.*, the number of days given to the winner to complete the job.

- **Stage 2:** After confirming that the project does not involve illegal activities, the site posts the auction on its public forum, which can be browsed by all its members. Sellers can also obtain up-to-date information on new auctions by subscribing to newsletters from the freelance site. The auction posting contains information provided by the buyer (*e.g.*, project description, auction start date) as well as information on the buyer himself (*e.g.*, his past ratings and his geographic location) through a link to his homepage.

- **Stage 3:** Sellers submit sealed bids, which are visible only to the buyer. Each bid contains a link to the respective seller's homepage, where the seller's attributes are visible, *e.g.*, her past average rating and geographic location. At the point of submitting the bid, sellers can see the exact number of bids that the auction has received so far. In our analysis, we approximate this process and treat the number of bids received as a variable that is perfectly observable to sellers. In a subsequent extension in §9.2.3, we allow sellers to have uncertainty on the total number of bids the auction will receive and condition their beliefs on the number of bids it has received so far.

- **Stage 4:** The buyer makes his decision by either picking one of the bidders as the winner or canceling the project. Cancellation can be interpreted as the outside option, since the buyer may take a canceled job elsewhere (locally or to another freelance site) or do it himself.

For each auction in our sample, we have the following information:

- The following auction attributes:

  - Number of bids received by the auction.

- Indicator for whether the buyer has posted an attachment describing the project.
- Deadline for project delivery (in days).
- The following buyer attributes:
  - Geographic region of the buyer – region codes are shown in Table 1.
  - Total number of past auctions initiated by the buyer.
  - Success ratio – fraction of past auctions in which the buyer chose a bid. A buyer who has initiated 10 auctions and canceled three auctions has a success ratio of 0.7; by default, it is zero for buyers with zero past auctions. Success ratio is indicative of the buyer's inherent choosiness and/or the quality of his outside options.
  - Number of past ratings and the sum of all past ratings.
  - Mean rating, defined as the 'sum of all past ratings/total number of past ratings' if the buyer has at least one rating, and zero otherwise.
  - Tenure on the site – number of days since the buyer signed up.
- The following seller attributes for all bids received:
  - Bid price and seller's geographic region (see Table 1).
  - Number of her past ratings, sum of her past ratings, and her mean rating.
  - Indicator of whether the seller has worked for the buyer in the past on this site.

Only 20% of the auctions end with the buyer picking a bid, while the rest are canceled. This is mostly because the site doesn't charge any monetary fees for posting and/or canceling auctions. Hence, buyers often err towards posting, even when they have good outside options. Table 3 provides an overview of the auctions in our data by buyer and auction attributes.[2] There is considerable heterogeneity across auctions in the number of bids received. An average auction receives about 11 bids, with the median being 7. Auctions in which the buyer picks a bid (as opposed to canceling) tend to receive slightly higher numbers of bids. An average auction has a deadline of 19 days, *i.e.*, the winning seller has approximately three weeks to deliver the job. About 18% of buyers post project attachments.

A big chunk of buyers (52.5%) have no past ratings, and while a small set of them (6.5%) have a mean rating of 10, with 10 or more ratings. Among uncanceled auctions, the percentage of buyers without past ratings is lower at 44.2%, and the percentage of buyers with very good reputation (mean rating of 10, with 10 or more ratings) is higher at 12.35%. The average rating for buyers who have been rated at least once in the past is 9.73, and this number is higher at 9.85 for uncanceled auctions. Most buyers in the data have previous experience on the site; the median buyer has posted one successful and one canceled auction. The median buyer who picks a bid has posted two successful auctions and one canceled auction. Overall, we find that buyers with past experience on the site and those with good reputation are more likely to pick a seller and not cancel the auction. Finally, as shown in Table 2, the majority (82.68%) of buyers belong to developed English-speaking countries, *i.e.*, Region 2.

Table 4 shows the summary statistics of seller and bid attributes for two sets of bids – all the bids received and accepted bids. A large percentage (36.53%) of bidders have no ratings, but 25% have more than 16 past ratings. The average seller who has been rated has a rating of 8.99. About 0.52% of the bidders have interacted

---

[2]In order to preserve the privacy of buyers and the freelance site, summary statistics for buyer tenure are not shown.

with the buyer in the past. The average price quoted by sellers is \$434.83. While the majority of bidders quote the MaxBid as their price (54.54%), many of them also quote much lower prices, with \$85 being the lowest observed quote. Note that the MaxBid is not binding and a small fraction of sellers actually quote prices higher than \$500, with \$5000 being the maximum quote observed. Unlike buyers, majority of the bidders (54.91%) belong to the Indian sub-continent (Table 2). The other three regions are about equally represented ($\approx$ 13-17%). The distribution of the seller's geographic region also varies with buyer's region.

There are systematic differences in the attributes of winning bidders compared to the full distribution of bidders. On average, they quote lower prices (Table 4), have significantly better reputations, are more likely to belong to developed countries, and have a higher likelihood (9.63%) of past interaction with the buyer.

4 Preliminary Analysis

We now present structure-free measures of the market outcomes and equilibrium bids. We start with simple metrics to quantify the beauty contest aspect of these auctions and bid dispersion in the marketplace.

- Chosen Bid Gap

$$\text{ChosenBidGap}_i = \frac{\text{Chosen bid in auction } i - \text{Minimum bid in auction } i}{\text{Minimum bid in auction } i} \times 100 \qquad (1)$$

It is the percentage difference between the chosen bid and the lowest bid. It is zero when the lowest bidder is also the winner, or when the winner quotes the same price as the lowest bidder. It is a measure of the beauty contest aspect of the auction – if buyers place significant value on other seller attributes (apart from price), then this metric is high. In price-only auctions, Chosen Bid Gap is always zero.

- Bid Range

$$\text{BidRange}_i = \frac{\text{Maximum bid in auction } i - \text{Minimum bid in auction } i}{\text{Minimum bid in auction } i} \times 100 \qquad (2)$$

The percentage difference between the maximum and minimum bids in an auction. If all bids are the same, then Bid Range is zero. It thus characterizes the underlying dispersion in bids within an auction.

- Relative Dispersion

$$\text{RelativeDispersion}_i = \frac{\text{Standard deviation of bids in auction } i}{\text{Mean of bids in auction } i} \times 100 \qquad (3)$$

The coefficient of variation in bids within an auction, in percentage. A high value means that the standard deviation of bids in auction i is high compared to the mean, or that there is a considerable amount of dispersion in the bids received within auction i. If all sellers quote the same price, then Relative Dispersion is zero.

First, we present the CDF of ln(ChosenBidGap+1) in Figure 1 for auctions in our dataset that have received two or more bids (3511 auctions). In over 50% of the auctions, Chosen Bid Gap is zero, *i.e.*, the lowest bid and the chosen bid have the same price. This does not necessarily imply that price is the only driver of choice in these auctions, because there may be multiple bids with the same price, with different seller attributes. In fact, in nearly 30% of the auctions there is no variation in the prices of bids received, and Chosen Bid Gap is zero for these auctions, by default. Importantly, note that in nearly 40% of the auctions, the Chosen Bid Gap is more than 50%

– buyers are picking sellers with prices 50% greater than the minimum price. This suggests that factors other than price play a significant role in these buyers' decisions, *i.e.*, the beauty contest aspect plays an important role here.

Next, we present the CDFs of the ln of Bid Range and Relative Dispersion in Figures 2 and 3. We see significant correlation in bid prices *within* auctions – over 35% of auctions have zero Bid Range, *i.e.*, there is no variation in the prices of bids received. Further, the median Relative Dispersion within auctions is 30%, whereas the Relative Dispersion for all the bids in the dataset is 55%. Since sellers are drawn independently in sealed bid auctions, the high correlation of bids within an auction (compared to across auctions) suggests that auction-specific variables, both observable and unobservable have a significant influence on bids.

To further explore this issue, we regress ln(price) on a slew of observable auction, buyer, and seller specific state variables. We present these results in Model M1 in Table 5. First, as expected, competition, *i.e.*, the number of bids, has a negative impact on bid price. On average, auctions with an attachment and those with shorter deadlines receive lower prices. Buyer and seller specific observables also have a significant impact on bid prices. Buyers from the developed countries (Region 2) command the lowest prices, followed by those from the Indian sub-continent and Eastern Europe (Regions 1 and 3). Further, buyers with higher past reputation (high mean ratings and number of ratings) receive lower prices. So do buyers who have posted many auctions on the site and those with high success ratio. Sellers from developed countries charge the highest prices, followed by those in the Indian sub-continent. Many seller reputation variables are also significant, suggesting that a seller's past experience and rating on the site has a significant impact on her bid. In spite of several significant effects, the overall explanatory power of the model is quite low (based on the adjusted R-squared), implying that only a small amount of the variation in bids is explained by auction, buyer, and seller specific observables.

The unexplained variance can stem from three factors – (1) the exogenous variation in sellers' costs of doing the jobs, (2) unobserved auction heterogeneity, and (3) the strategic behavior of sellers which is highly non-linear in both observables and unobservables (see Equation (11)), and hence not captured in this simple regression model. Since both the second and third factors are affected by unobserved auction heterogeneity, we need to examine whether it is a significant concern in this setting. Hence, we modify Model M1 to include auction dummies and present the results in Model M2. Due to space constraints, we do not present the auction dummies in the table. However, we note that 32.31% of the buyer dummies are significant at the 10% confidence level. Moreover, the R-squared of the model with buyer dummies (Model M2) jumps to 0.3548 compared to 0.0278 in Model M1. These findings suggest that unobserved auction heterogeneity has a significant impact on bids.

Overall, our preliminary findings suggest that there exist significant variations in bid prices; and that such variations are driven, at least in part, by auction specific unobservables, apart from variations in sellers' costs and observable buyer, auction, and seller attributes. Armed with these structure-free insights, we now specify and estimate a structural model of seller behavior in freelance auctions.

5   Empirical Framework

We now present an empirical framework to model beauty contest auctions.

5.1   Set-up

Throughout, we use a nomenclature consistent with procurement auctions, *i.e.*, the auctioneer is the buyer and bidders are sellers. All the players are risk-neutral. Further, we use script letters to denote functions, capital

letters to denote sets of variables, and small letters to denote a single variable. There are n buyers, indexed by i, who each seek to buy a single and indivisible product, and conduct one auction each, referred to as auction i, with some abuse of notation.

Observed Auction Specific Variables: Auctions are allowed to be heterogeneous through a set of observed state variables O_i, which is common knowledge, *i.e.*, known to the buyer and all sellers. O_i includes buyer specific variables that are constant for the duration of the auction, properties of the auctioned object, and a non-binding maximum bid (MaxBid).

In practice, auctions are usually conducted by a third-party who coordinates the entire process. In such cases, the third-party charges a commission on successful transactions, either paid by the winning seller or the buyer. In this analysis, we assume that the winning seller pays a percentage commission, r_i, on her bid amount. Modifying the model so that the commission is paid by the buyer is straightforward.

Auction i receives q_i bids, and this number is known (or observed) by all potential sellers.[3] Henceforth, we refer to $\{O_i, q_i, r_i\}$ as the complete set of observed auction specific variables, A_i, because they constitute all the attributes of the auction which are invariant across sellers and observed by the researcher.

Unobserved Auction Specific Variable: In many settings, factors that affect sellers' costs may be common knowledge among *all* bidders, but invisible to the researcher. For example, in the freelancing context, all sellers may perceive a project to be more difficult or challenging based on the project description and condition their bids on this information. However, it is difficult for the researcher to infer project difficulty from project descriptions. It is well-known that not accounting for such common knowledge unobservables can bias estimates of sellers' private costs. Hence we include an auction specific unobservable v_i, which is independent of the observables O_i and is drawn from a set of finite types $v_i \in \{v^1, v^2, ..., v^K\}$. We denote the population probability that an auction is of unobserved type k as π_k. Please see §5.4 for a detailed discussion on auction specific unobservables.

Seller-specific Variables: Sellers are indexed by j, and for auction i, $j \in \{1, 2, ..., q_i\}$. Let $\{c_{ji}, X_{ji}\}$ be the state variables that seller j is endowed with at the beginning of auction i. X_{ji} is the set of observable seller attributes that are relevant to the auction, *i.e.*, affect buyer i's probability of choosing seller j. X_{ji} is assumed to be fixed for the duration of auction i, that is, seller j cannot optimize on it.[4] In a freelance setting, X_{ji} may include variables such as the rating/reputation of the seller and her geographic location. While the seller has control over these factors in the long run, she is unlikely to be able to change these attributes in the short run. c_{ji} is seller j's private cost of completing the project specified in auction i. It can be expressed as:

$$c_{ji} \equiv c_{ji}(O_i, v_i, X_{ji}, \tilde{c}_{ji}). \tag{4}$$

Intuitively, a seller's cost of providing a service can vary with her own attributes, with the auction's attributes (both observed and unobserved), as well as a private shock $\tilde{c}_{ji}$ (Athey and Haile, 2007). For notational simplicity, we henceforth denote $c_{ji}(O_i, v_i, X_{ji}, \tilde{c}_{ji})$ as c_{ji}. Finally, we denote the seller's decision variable or bid as b_{ji}.

Timeline of the game: We consider a sealed bid beauty contest auction with three stages. At stage 1, buyer i

---

[3]Extending the model to allow for scenarios where sellers don't know the number of bidders is straightforward. In such cases, we would assume that sellers know the conditional distribution of the number of bids, $\mathcal{H}(q_i | O_i, v_i, r_i)$.

[4]X_{ji} is also assumed to be verifiable by the buyer and cannot be falsified by the seller.

posts the auction and the auction specific variables, $\{O_i, v_i, r_i\}$, become visible to the sellers. At stage 2, sellers, $j \in \{1,...,q_i\}$, submit a bid b_{ji} to the buyer, at which point the buyer observes $\{b_{ji}, X_{ji}\}$ for all submitted bids. At stage 3, the buyer makes a decision d_i, and chooses one of the submitted bidders as the winner ($d_i = j$) or chooses the outside option ($d_i = 0$, *i.e.*, rejects all bids).

Buyers' Information Set: Buyers know the observable and unobservable auction attributes and commission rate, $\{O_i, v_i, r_i\}$, number of bids received q_i, and the bid price and seller attributes for each bid $\{b_{ji}, X_{ji}\}$.

Sellers' Information Set: Seller j knows own attributes $\{c_{ji}, X_{ji}\}$, auction attributes $\{O_i, v_i, r_i\}$, the number of bids that the auction receives q_i, and the joint distribution of seller costs and types, c_{ji} and X_{ji}, given $\{A_i, v_i\}$, where $A_i = \{O_i, q_i, r_i\}$. They do not observe the costs, bids, or attributes of the other sellers.

5.2 Sellers' Optimization

Let $X_{-ji} = \{X_{1i}, ..., X_{(j-1)i}, X_{(j+1)i}, ..., X_{q_ii}\}$ and $b_{-ji} = \{b_{1i}, ..., b_{(j-1)i}, b_{(j+1)i}, ..., b_{q_ii}\}$ be sets of the attributes and prices of all the bidders in auction i, except j. The expected utility of seller j in auction i, from choosing bid b_{ji}, is:

$$u(A_i, v_i, X_{ji}, c_{ji}, b_{ji}) = [b_{ji}(1-r_i) - c_{ji}] \cdot \mathcal{S}_{ji}(X_{ji}, b_{ji} | A_i, v_i) \tag{5}$$

where $\mathcal{S}_{ji}(X_{ji}, b_{ji} | A_i, v_i)$ is j's expected probability of winning in equilibrium given observed and unobserved auction attributes $\{A_i, v_i\}$. We work with expected probabilities because seller j doesn't observe the attributes or prices of other bidders in sealed-bid auctions. $\mathcal{S}_{ji}(\cdot)$ can be further expanded as follows:

$$
\begin{aligned}
\mathcal{S}_{ji}(X_{ji}, b_{ji} | A_i, v_i) &= E[\mathcal{P}_{ji}(X_{ji}, b_{ji}, X_{-ji}, b_{-ji} | A_i, v_i)] \\
&= \int \mathcal{P}_{ji}(X_{ji}, b_{ji}, X_{-ji}, b_{-ji} | A_i, v_i) \mathcal{G}(X_{-ji}, b_{-ji} | A_i, v_i) d(X_{-ji}, b_{-ji} | A_i, v_i)
\end{aligned}
\tag{6}
$$

where $\mathcal{P}_{ji}(X_{ji}, b_{ji}, X_{-ji}, b_{-ji} | A_i, v_i)$ is buyer i's equilibrium probability of choosing seller j given observed and unobserved auction attributes $\{A_i, v_i\}$ (which includes the number of bidders q_i), seller j's own attributes and price $\{X_{ji}, b_{ji}\}$, and the attributes and prices of all other bidders in the auction, $\{X_{-ji}, b_{-ji}\}$.[5] $\mathcal{P}_{ji}(\cdot)$ can be interpreted as the observed probabilistic outcome of some unobserved decision rule used by the buyer. Hence, it can be denoted or interpreted as the buyer's Conditional Choice Probability (CCP). $\mathcal{G}(X_{-ji}, b_{-ji} | A_i, v_i)$ is the equilibrium probability distribution of seller attributes and bid quotes of the other $q_i - 1$ sellers in the auction.

When the buyer-specified maximum bid is non-binding (as it is in our setting), we have an unconstrained maximization problem,[6] and the seller's optimization problem can be expressed as:

$$\max_{b_{ji}} [b_{ji}(1-r_i) - c_{ji}] \cdot \mathcal{S}_{ji}(X_{ji}, b_{ji} | A_i, v_i) \tag{7}$$

The First Order Condition (FOC) of this problem is:

$$[b_{ji}(1-r_i) - c_{ji}] \cdot \frac{\partial \mathcal{S}_{ji}(X_{ji}, b_{ji} | A_i, v_i)}{\partial b_{ji}} + (1-r_i) \cdot \mathcal{S}_{ji}(X_{ji}, b_{ji} | A_i, v_i) = 0 \tag{8}$$

---

[5]Note that $\mathcal{P}_{ji}(\cdot)$ need not be assumed to depend on the commission rate r_i, because the commission is paid by the winning seller, and hence not relevant to the buyer.

[6]In settings with non-binding maximum bids, the estimated cost distributions are uncensored (Athey and Haile, 2007).

This, in turn, can be rearranged to obtain the seller cost, c_{ji}, as:

$$c_{ji} = \xi(A_i, v_i, X_{ji}, b_{ji}, \mathcal{G}(\cdot), \mathcal{P}(\cdot)) = (1-r_i) \cdot \left[b_{ji} + \left(\frac{\partial \mathcal{S}_{ji}(X_{ji}, b_{ji} | A_i, v_i)}{\partial b_{ji}} \right)^{-1} \mathcal{S}_{ji}(X_{ji}, b_{ji} | A_i, v_i) \right] \qquad (9)$$

5.3 Assumptions

We make the following assumptions. We note that these assumptions are standard in the auctions literature, and that whenever possible, we either empirically test them or provide robustness checks. We also present detailed discussions on their implications on identification, and the trade-offs involved in relaxing them. Please see §5.5 and §9.2 for details.

Assumption 1. Conditional Independence: In sealed-bid auctions, conditional independence follows naturally, *i.e.*, sellers are drawn independently from a joint distribution $\mathcal{F}(\tilde{c}_{ji}, X_{ji} | A_i, v_i)$. That is, conditional on auction attributes, sellers' types $\{\tilde{c}_{ji}, X_{ji}\}$ are not correlated within an auction. This assumption is analogous to the independent private values assumption in standard auction models (Guerre et al., 2000). Note that this does not rule out correlation within X_{ji}, between X_{ji} and c_{ji}, and between $\{c_{ji}, X_{ji}\}$ and A_i. For example, it allows for a scenario where sellers with good reputations also have lower costs, and one where buyers are more likely to attract sellers from the same geographic region as themselves.

Assumption 2. Continuity and Monotonicity: We assume that, $\mathcal{P}_{ji}(X_{ji}, b_{ji}, X_{-ji}, b_{-ji} | A_i, v_i)$, the probability of winning, is continuous and twice differentiable in b_{ji} and b_{-ji}. It is assumed to be strictly decreasing in b_{ji} and strictly increasing in b_{-ji}, *i.e.*, keeping everything else constant, an increase in own bid leads to a lower probability of being chosen, and an increase in an opponent's bid price increases own probability of winning. This assumption is similar in spirit to condition $C2$ of Theorem 1 in Guerre et al. (2000).

Assumption 3. Second Order Condition: As is standard in the literature, the SOC of seller's maximization problem is assumed to be satisfied.

$$[b_{ji}(1-r_i) - c_{ji}] \frac{\partial^2 \mathcal{S}_{ji}(X_{ji}, b_{ji} | A_i, v_i)}{\partial^2 b_{ji}} + 2(1-r_i) \frac{\partial \mathcal{S}_{ji}(X_{ji}, b_{ji} | A_i, v_i)}{\partial b_{ji}} < 0 \qquad (10)$$

Note that $\frac{\partial \mathcal{S}_{ji}(X_{ji}, b_{ji} | A_i, v_i)}{\partial b_{ji}} < 0$, $1 - r_i > 0$, and under individual rationality constraints, $[b_{ji}(1-r_i) - c_{ji}] > 0$. For condition (10) to be satisfied, we require that $\frac{\partial^2 \mathcal{S}_{ji}(X_{ji}, b_{ji} | A_i, v_i)}{\partial^2 b_{ji}}$ is not too positive, *i.e.*, we assume that the function $\mathcal{S}_{ji}(X_{ji}, b_{ji} | A_i, v_i)$ is not too convex.

Assumption 4. Symmetry and Private Information: All auctions, buyers, and sellers are assumed to be symmetric after accounting for $\{A_i, v_i\}$ and $\{X_{ji}, b_{ji}\}$, which implies that $\mathcal{P}_{ji}(\cdot) = \mathcal{P}(\cdot) \forall j, i$. Given a draw of $\{A_i, v_i, X_{-ji}, b_{-ji}\}$, seller j's probability of winning auction i is the same as that of another seller k if her attributes and bid price are the same as k's, *i.e.*, $\{X_{ji}, b_{ji}\} \equiv \{X_{ki}, b_{ki}\}$. Symmetry imposes restrictions on the extent to which unobservables (to the researcher) are common knowledge in the system; it assumes that there exists no seller specific variable that is visible to both the buyer and seller, but is unobservable to the researcher. Note that it does not rule out buyer-specific unobservable preferences; it only assumes that sellers are not privy to realizations of such unobserved tastes. We discuss potential relaxations of this assumption in §11.1.

Equilibrium: A symmetric Bayesian Nash equilibrium of this incomplete information game consists of a bidding strategy for sellers, $b_{ji} = \beta(A_i, v_i, c_{ji}, X_{ji}, \mathcal{G}(\cdot), \mathcal{P}(\cdot))$, and a bid selection strategy for buyers, $\mathcal{P}(X_{ji}, b_{ji}, X_{-ji}, b_{-ji} | A_i, v_i)$. The existence of pure strategy Bayesian Nash equilibrium can be established following the arguments in Athey (2001); Athey and Levin (2001); McAdams (2003). Moreover, with the assumption of unique equilibrium in data and Assumptions 1–4, for a given buyer decision rule $\mathcal{P}(\cdot)$, there exists a unique best response function $\beta(A_i, v_i, c_{ji}, X_{ji}, \mathcal{G}(\cdot), \mathcal{P}(\cdot))$.

5.4 Discussion of Unobserved Auction Heterogeneity

We now present a detailed discussion on the importance of accounting for auction specific unobservables.

In general, allowing for auction specific unobservables is difficult even in simple first price auctions. The challenge lies in separately identifying the impact of two unobservables on a bid – the unobservable cost of the seller and the unobservable auction heterogeneity. For example, in a freelancing context, a seller might bid high either because she has high programming costs or because the job is inherently difficult even for skilled programmers. In the context of first price auctions, a recent paper by Krasnokutskaya (2011) shows that the distribution of auction specific unobservables and the distribution of seller costs are both separately identified if auction specific unobservables enter valuations multiplicatively.[7] She then uses a deconvolution method to recover both these unobservables. However, Krasnokutskaya's identification (and estimation) rests firmly on the multiplicative separability assumption: $c_{ji} = v_i \cdot \tilde{c}_{ji}$, where v_i is the unobserved type of the auction. For first price auctions, it is easy to show that multiplicative separability in costs translates to multiplicative separability in bidding strategies $\Rightarrow \beta(c_{ji}) = v_i \cdot \beta(\tilde{c}_{ji})$. This ensures that seller margins are multiplicatively separable in v_i, and thereby allows her to employ a deconvolution estimator.

Unfortunately, in our setting, multiplicative separability of v_i in costs does not translate to multiplicative separability of v_i in bids. In fact, the existence of an outside option completely rules it out – if everyone raises their bid by a factor v_i, then the probability of being chosen certainly does not remain the same; it decreases by an unspecified amount because the buyer can simply choose not to buy. Hence, the deconvolution estimator presented by Krasnokutskaya (2011) does not work for our setting.

Two solutions from the literature can be imported to resolve this problem. First, if we are willing to assume that unobserved heterogeneity affects some outcomes, but not all, then estimation is straightforward. For example, Campo et al. (2003) and Haile et al. (2003) assume that auction specific unobservables affect seller's decision to enter the auction, *i.e.*, the number of bidders in the market, but not the bids. In such settings, unbiased estimates of costs can be obtained by simply conditioning the bid distributions and choice probabilities on the number of bidders in the auction (as we already do). While this is a relatively simple fix, the assumptions necessary to implement are not valid in our setting.[8] Second, if the unobserved auction heterogeneity can be recast as unobserved buyer heterogeneity (*e.g.*, some buyers are more difficult to work for), and if we have a long panel of data with repeat observations on the same buyer, then the first stage can be estimated at the buyer level. Alternately, if there are different auction markets and the auctions within a given market are

---

[7]Multiplicative separability is conceptually no different from additive separability since the monotonic *ln* transformation of a multiplicatively separable function gives us a additively separable function.

[8]An example of a case where this would work is this: Suppose there is randomly induced heterogeneity in how well an auction is advertised to sellers. Then, bids would not be affected by the advertisement, though the number of bidders would be. However, in most realistic settings, if auction specific unobservables influence sellers' entry, then they are also likely to affect bid prices.

homogeneous, then the first stage estimation can be at the market level. This approach has been used in the games context by Misra and Nair (2011) and Ellickson and Misra (2012). While this method is ideal if the researcher has access to datasets with long panels of buyers and the only source of unobserved heterogeneity is from the buyer (rather than auction), such datasets are seldom available in auction settings.

Hence, in this paper we present a novel solution to accommodate auction specific unobservables through a set of finite types. Our method is similar in spirit to that proposed by Arcidiacono and Miller (2011) for discrete games. We provide a constructive proof of identification of nonparametric mixtures of bid distributions based on the literature on nonparametric identification of mixtures with i.i.d draws (Hettmansperger and Thomas, 2000; Elmore et al., 2004) and propose a two step estimator that includes a smoothed nonparametric EM-like algorithm in the first step to account for finite types of unobservables (Benaglia et al., 2009a).

5.5   Identification

First consider the simple case where there are no auction specific unobservables. Then, with Assumptions 1–4, it is easy to see that $\mathcal{P}(\cdot)$ and $\mathcal{G}(\cdot)$ are nonparametrically identified. That is, given a set of state variables, the probability that a buyer will choose a given bid and the joint distributions of bids and seller attributes can be nonparametrically inferred from data. This combined with the FOC ensures that the conditional cost distribution, $\mathcal{F}(c_{ji}|X_{ji},A_i)$, is identified. The key idea is that, after controlling for the strategic behavior of sellers, the variation in bids is assumed to purely stem from the variations in sellers' private costs.

However, with auction specific unobservables, the situation becomes tricky. Now $\mathcal{G}(X_{-ji},b_{-ji}|A_i,v_i)$ and $\mathcal{P}_{ji}(X_{ji},b_{ji},X_{-ji},b_{-ji}|A_i,v_i)$ are not directly available from data because they are functions of the unobservable state variable v_i. Moreover, the mixture probabilities π_ks also need to be identified. Since v_i is drawn from a set of finite types, we are essentially looking to identify a nonparametric mixture model.

We start by considering the identification of component distributions $\mathcal{G}(X_{-ji},b_{-ji}|A_i,v_i)$s and mixture probabilities π_ks. Recall that we have multiple bids per auction and that these bids are i.i.d because of the sealed bid setting. Thus the identification problem devolves to one of identification of nonparametric mixtures with i.i.d draws. The basic results in this setting were established by Hettmansperger and Thomas (2000) and Elmore et al. (2004). They show that the identification problem can be recast as the identification of binomial mixture models using a simple discretization theorem.[9] Our identification proof, presented in Web Appendix A.1, follows along the same lines as theirs. We note that the identification result holds as long as the number of bids per auction is greater than or equal to $2K-1$. For instance, with two unobserved types, we need at least three observed bids per auction.

We now explain this identification constraint using a very simple example. Consider a setting with two unobserved auctions types with population probabilities π and $1-\pi$. Suppose that each auction gets two i.i.d bids that can each take on one of two values – High (H) and Low (L). Let the probability of L and H bids for a Type 1 auction be a and $1-a$; and the corresponding probabilities for a Type 2 auction be b and $1-b$. We thus have three parameters, $\{\pi,a,b\}$, to identify. In terms of data, we have three types of

---

[9]Note that this identification problem is distinct from the identification of nonparametric multivariate mixture models, where the draws are independent but not identical. Hall and Zhou (2003) were the first to examine this general problem and Allman et al. (2009) provide the complete solution to it. Identification is actually easier without the assumption that draws are identical and is achieved as soon as the number of draws per unit is at least three, irrespective of the number of unobserved states. See Theorem 8 in Allman et al. (2009) and the subsequent discussion for details.

observed outcomes, $\{HH, LL, LH\}$, and three corresponding observed probabilities $Pr(HH)$, $Pr(LL)$, and $Pr(LH)$. Since $Pr(LH)$ is just $1 - Pr(HH) - Pr(LL)$, we essentially have two equations to work with, which are: $Pr(LL) = \pi \cdot a^2 + (1-\pi) \cdot b^2$ and $Pr(HH) = \pi \cdot (1-a)^2 + (1-\pi) \cdot (1-b)^2$. With two equations and three parameters, the system is not identified. On the other hand, if we had three i.i.d bids per auction, the number of equations would increase to three (because we would have four possible outcomes $\{HHH, LLL, LLH, HHL\}$); and the system is just identified. Thus, for a given number of unobserved types, we need a minimum number of bids to obtain nonparametric identification. Intuitively, the identification arguments treats an auction as a unit and the bids within an auction as independent draws from the same unit. Thus, variations in bids within an auction are treated as stochastic errors, while variations across auctions are attributed to unobserved heterogeneity; but to be able to reliably separate the stochasticity of bids and unobserved types in a nonparametric sense, we need enough draws of bids within an auction.

Next, consider $\mathcal{P}(X_{ji}, b_{ji}, X_{-ji}, b_{-ji} | A_i, v_i)$, which is the buyer's CCP. Unfortunately, this is not non-parametrically identified because we only observe one decision choice per auction. Note that this is in stark contrast to the identification of bid distributions discussed above, where multiple bids per auction are observed. Without dynamics or repeat observations within the same auction, there is no way for us to separately identify K underlying CCPs, $\mathcal{P}(X_{ji}, b_{ji}, X_{-ji}, b_{-ji} | A_i, v_i)$s, from the single observed probability distribution $\mathcal{P}(X_{ji}, b_{ji}, X_{-ji}, b_{-ji} | A_i)$. The intuition behind this non-identification result is as follows – when we see two buyers with the same observed state variables make different decisions, we do not know whether the difference in the decisions is due to the inherent stochasticity of the decisions or due to the buyer-auction specific unobservable v_i. See Kasahara and Shimotsu (2009) for a detailed non-identification proof. Therefore, to proceed with estimation, we need to make the following assumption:

Assumption 5. Buyer i's equilibrium decision rule does not depend on the auction specific unobservable v_i. That is, $\mathcal{P}(X_{ji}, b_{ji}, X_{-ji}, b_{-ji} | A_i, v_i) \equiv \mathcal{P}(X_{ji}, b_{ji}, X_{-ji}, b_{-ji} | A_i) \ \forall \ v_i$.

Assumption 5 implies that common shocks to sellers' costs can influence buyers' decisions only through bid prices. For example, in freelance auctions, the unobserved difficulty of the job is allowed to act as a common shock on all sellers' costs, *i.e.*, all sellers may quote higher prices for a difficult job, and these prices obviously affect buyers' decisions. However, the shock in and of itself is assumed to be irrelevant to the buyer's decision. To the extent that the buyer only pays the actual bid, this is reasonable.

Nevertheless, this assumption can be restrictive in certain circumstances. For example, if the buyer's outside option upon canceling a job is to do it himself, then the unobserved job difficulty is likely to influence his decision to choose a winner. Moreover, the auction specific unobservable v_i may include variables that do not necessarily affect sellers' costs, but influence buyers' decisions. For example, sellers may know that some buyers are more price sensitive or have better outside options. In such settings, assuming that buyers' decisions are independent of v_i can bias estimates of sellers costs. Unfortunately, $\mathcal{P}(\cdot)$ is not nonparametrically identified, and there are no clean solutions to this issue. If the researcher believes that: (a) v_i has a significant impact on buyers' decisions, even after accounting for its impact through the bid price, and (b) she is willing to forgo nonparametric identification, then she may parametrize the buyers' decision rule. Then it is possible to simultaneously estimate nonparametric mixtures of bid distributions and a parametric mixture model of buyers' decision in the first step of estimation. We present this extension in §9.2.1. Alternately, if the researcher is willing to think of v_i as simply

buyer (and not auction) specific unobservable, and has long panels on the same buyer, then she may estimate the decision rule at the buyer level. Since we don't have long panels of buyers in our data, we don't attempt this.

Finally, note that once $\mathcal{G}(X_{-ji},b_{-ji}|A_i,v_i)$ and $\mathcal{P}(X_{ji},b_{ji},X_{-ji},b_{-ji}|A_i)$ are identified and retrieved, then the FOC implies that the conditional cost distribution, $\mathcal{F}(c_{ji}|X_{ji},A_i,v_i)$, is naturally identified because there is a one-to-one relationship between bids and costs.

6 Estimation

We now outline our two-step estimation strategy in detail. The key equation of interest is the rearranged FOC of the sellers' optimization problem:

$$c_{ji}=\xi(A_i,v_i,X_{ji},b_{ji},\mathcal{G}(\cdot),\mathcal{P}(\cdot))=(1-r_i)\left[b_{ji}+\left(\frac{\partial\widehat{\mathcal{S}}(X_{ji},b_{ji}|A_i,v_i)}{\partial b_{ji}}\right)^{-1}\widehat{\mathcal{S}}(X_{ji},b_{ji}|A_i,v_i)\right] \qquad (11)$$

where $\left(\frac{\partial\widehat{\mathcal{S}}(X_{ji},b_{ji}|A_i,v_i)}{\partial b_{ji}}\right)^{-1}$ and $\widehat{\mathcal{S}}(X_{ji},b_{ji}|A_i,v_i)$ are the numerical estimates of $\left(\frac{\partial\mathcal{S}(X_{ji},b_{ji}|A_i,v_i)}{\partial b_{ji}}\right)^{-1}$ and $\mathcal{S}(X_{ji},b_{ji}|A_i,v_i)$, respectively. We can write out $\widehat{\mathcal{S}}(X_{ji},b_{ji}|A_i,v_i)$ as:

$$\widehat{\mathcal{S}}(X_{ji},b_{ji}|A_i,v_i)=\int\widehat{\mathcal{P}}(X_{ji},b_{ji},X_{-ji},b_{-ji}|A_i)\widehat{\mathcal{G}}(X_{-ji},b_{-ji}|A_i,v_i)d(X_{-ji},b_{-ji}|A_i,v_i) \qquad (12)$$

Hence, to obtain numerical estimates of $\widehat{\mathcal{S}}(\cdot)$ and $\left(\frac{\partial\widehat{\mathcal{S}}(\cdot)}{\partial b_{ji}}\right)^{-1}$, we need to estimate $\mathcal{P}(X_{ji},b_{ji},X_{-ji},b_{-ji}|A_i)$ and $\mathcal{G}(X_{-ji},b_{-ji}|A_i,v_i)$, which represent sellers' beliefs on the probability of being chosen given a set of state variables and the equilibrium distribution of bids.

First, we briefly explain the estimation procedure in a simpler setting, where there is no auction specific unobservable v_i; and then present the estimation for the full model. Without unobserved heterogeneity, the estimation is relatively simple. In the first step, the buyer's decision rule and its derivative, and the equilibrium distributions of bids and seller attributes are nonparametrically estimated. Under the assumption of rational expectations, unique equilibrium in data, and no common-knowledge unobservables (*i.e.*, in the absence of v_i), these distributions can be directly estimated from the data. Then in the second step, the expected probability of winning and its derivative are simulated using the distributions estimated in the first step, and plugged back into the FOC to obtain the pseudo costs for each seller observed in the data. These pseudo costs are then used to generate the nonparametric distribution of seller costs. Please see Web Appendix A.2.2 for step-by-step instructions on estimating beauty contest models without unobserved heterogeneity.

In a model with auction specific unobservables, estimation involves a significant additional challenge – we now need to estimate bid distributions, $\mathcal{G}(X_{-ji},b_{-ji}|A_i,v_i)$, as functions of unobservables v_i. Specifically, we have the following problem: we only observe bid distributions as functions of observables, *i.e.*, $\mathcal{G}(X_{-ji},b_{-ji}|A_i)$s, but we need to recover K underlying distributions, $\mathcal{G}(X_{-ji},b_{-ji}|A_i,v^k)$s, as well as the population distribution of the K unobserved types. Moreover, this entire exercise needs to be nonparametric.

In this paper, we address this challenge by employing nonparametric EM-like algorithm to estimate the components $\mathcal{G}(X_{-ji},b_{-ji}|A_i,v^k)$ and π_ks in the first step and then using these mixture distributions in the second step estimation of pseudo costs. We discuss the estimation steps in detail below.

6.1 First Step Estimation

In the first step, we estimate the buyer's decision rule, its derivative, and the equilibrium distributions of bids and seller attributes. We now describe each of them.

6.1.1 Buyers' Decision Rule $\mathcal{P}(X_{ji}, b_{ji}, X_{-ji}, b_{-ji} | A_i)$

This is the probability that seller j will win auction i given state variables $\{A_i, X_{ji}, b_{ji}, X_{-ji}, b_{-ji}\}$. Given the assumption that common knowledge unobservables don't affect CCPs, these probabilities are directly available from data following Hotz and Miller (1993). They can be estimated without making functional form assumptions using either sieve estimators or kernel densities when the researcher has large samples and a relatively small state space. However, with large state spaces, finite samples may not be amenable to purely nonparametric methods. In such cases, the researcher may employ semi-parametric or parametric methods, with the goal of maximizing the fit or the predictive ability of these first stage models. Finally, the derivative of the CCPs, $\frac{\partial \mathcal{P}(X_{ji}, b_{ji}, X_{-ji}, b_{-ji} | A_i)}{\partial b_{ji}}$ are easily computed by taking the numerical derivative of $\mathcal{P}(\cdot)$.

6.1.2 EM-like Algorithm to Estimate Nonparametric Mixtures of Bid Distributions

We now present a nonparametric EM-like algorithm to estimate the individual components of mixture distributions. Our algorithm is similar in spirit to the one proposed by Benaglia et al. (2009a), which has been used to estimate nonparametric mixture models in other settings.

Note that we are careful to call the algorithm described below EM-like (and not EM). Traditional EM algorithms for parametric models have theoretical properties such as non-decreasing likelihood at each step of the EM (Dempster et al., 1977). However, nonparametric models are not estimated using Maximum likelihood and the standard results from the EM literature do not extend to this setting. Nevertheless, EM-like algorithms are now increasingly used to estimate nonparametric mixtures and are generally well-behaved (Benaglia et al., 2009b). Of course, in complex settings where convergence is difficult, it is possible to use instead a nonparametric MM algorithm which has the desirable descent property, like traditional EMs, and guarantees convergence. See Levine et al. (2011) for details.

Basic Setup

Consider a setting where seller attributes, X_{ji}, are a set of continuous state variables. Recall that the unobserved type v_i comes from a finite set, $\{v^1, ..., v^K\}$. So there are $K-1$ population probabilities (π_ks) to be estimated. Let λ_{ik} denote the posterior probability that auction i belongs to unobserved type k, given observed bids. Next, let A_i take H possible levels, $A_i \in \{A^1, A^2, ..., A^H\}$. Then all the bids in the data can be partitioned into H groups based on observed auction attributes A_i. For example, auctions with $A_i = A^1$ go into group 1, those from auctions with $A_i = A^2$ go into group 2, and so on. It is essential that the total number of groups be small enough so that each group has sufficient data. This can be challenging in finite samples, especially if A_i has continuous variables or if the size of the state space is large. A simple solution is to lump together groups for which the joint distributions of X_{ji} and b_{ji} look similar and make coarser partitions with more data in each group. Instead of eye-balling the extent of similarity across groups, it is possible to do this scientifically using Kolmogorov-Smirnov tests.

We now present an iterative algorithm where each iteration consists of three steps – the KDE step, E-step,

and the M-step. Let the initial guess of both the population and posterior probabilities be $\pi_k^0 = \lambda_{ik}^0 = 1/K \forall k$. The superscript denotes the iteration number, which for the initial guess is zero.

KDE-Step

Let the dimensionality of $\{X_{ji}, b_{ji}\}$ be D, and let Z denote a point in this D-dimensional space. Further, the bids in each group are indexed by $m \in \{1, 2, ..., n_h\}$, where n_h is the total number of bids in group h where $h \in \{1, ..., H\}$, . Then $\mathcal{G}_{h,k}^t(Z|A^h, v^k)$ denotes the multivariate kernel density function at observed state variables A^h and unobserved type v^k in iteration t. For each group h, we have K probability density functions. So overall, we need to estimate $H \times K$ joint distributions of seller attributes and bid prices. Let λ_{mhk}^t be the posterior probability that bid m in group h is drawn from an auction of unobserved type k, where $\lambda_{mhk}^t = \lambda_{ik}^t$ because auction i is the parent auction of bid m. Since the KDE step precedes the E-step in iteration t, the posterior probabilities from the last step (λ_{mhk}^{t-1}s) are used. We now define $\mathcal{G}_{h,k}^t(Z|A^h, v^k)$ as:

$$\mathcal{G}_{h,k}^t(Z|A^h, v^k) = \frac{1}{\left(\mu_h^t\right)^D \left[\sum_{m=1}^{n_h} \lambda_{mhk}^{t-1}\right]} \sum_{m=1}^{n_h} \lambda_{mhk}^{t-1} \cdot \mathcal{K}\left(\frac{Z - Z_m}{\mu_h^t}\right) \forall h, k \tag{13}$$

where μ_h^t is the bandwidth window for group h in iteration t, $\mathcal{K}(\cdot)$ is the D dimensional kernel function satisfying the property $\int_{R^D} \mathcal{K}(Z) d(Z) = 1$, and λ_{mhk}^{t-1} is the weight attached to each point m.

E-Step

Recall that λ_{ik}^t is the probability that auction i belongs to unobserved type k. In the Expectation step, we update the posterior probabilities λ_{ik}^ts, for each auction, for this iteration, as follows:

$$\lambda_{ik}^t = \frac{\pi_k^{t-1} \prod_{j=1}^{q_i} \mathcal{G}_{h,k}^t\left(X_{ji}, b_{ji}|A_i = A^h, v^k\right)}{\sum_{k=1}^{K} \pi_k^{t-1} \left[\prod_{j=1}^{q_i} \mathcal{G}_{h,k}^t\left(X_{ji}, b_{ji}|A = A^h, v^k\right)\right]} \forall k \tag{14}$$

where π_k^{t-1} is the population probability of unobserved type k from the previous iteration.

M-Step

Finally, in the Maximization step, we update the population probabilities for this iteration as follows:

$$\pi_k^t = \frac{\sum_{i=1}^{n} \lambda_{ik}^t}{n} \forall k \tag{15}$$

We iteratively perform the three steps till the population probabilities (π_ks) converge, at which point, we have consistent estimates of the population probabilities of unobserved types, posterior probability of an auction belonging to a given unobserved type, and the $H \times K$ joint probability density functions of seller attributes and bid prices, $\mathcal{G}_{h,k}\left(X_{ji}, b_{ji}|A^h, v^k\right)$.

6.2 Second Step Estimation

We now discuss the estimation of seller costs. Recall that we have K unobserved types of auctions. Hence, we need to estimate K cost distributions. To obtain these distributions, we require numerical estimates of $\widehat{\mathcal{S}}(\cdot)$ and $\left(\frac{\partial \widehat{\mathcal{S}}(\cdot)}{\partial b_{ji}}\right)^{-1}$ for each seller j, in each auction i, for each unobserved state v^k, which are then used to obtain the cost distributions. Below, we describe the steps in detail.

1. For each type v^k belonging to $\{v^1,...,v^K\}$, *i.e.*, iterate over all unobserved types:

    (a) For each bid j in each auction i, *i.e.*, iterate over all bids in all auctions:

    i. Make $(q_i - 1)$ draws of equilibrium seller attributes and bids from $\widehat{\mathcal{G}}_{h,v^k}(X_{ji}, b_{ji}|A_i = A^h, v^k)$. Denote these draws as: $\tilde{X}_{-ji} = \{\tilde{X}_{1i},...,\tilde{X}_{(j-1)i}, \tilde{X}_{(j+1)i},...,\tilde{X}_{q_i i}\}$ and $\tilde{b}_{-ji} = \{\tilde{b}_{1i},...,\tilde{b}_{(j-1)i}, \tilde{b}_{(j+1)i},...,\tilde{b}_{q_i i}\}$. Together with j's own attributes and bid, this constitutes one simulation of auction i for $v_i = v^k$.

    ii. Using the $q_i - 1$ simulated draws from Step i and j's own attributes and bid, obtain the probability of being chosen and its derivative, $\tilde{\mathcal{P}}\left(X_{ji}, b_{ji}, \tilde{X}_{-ji}, \tilde{b}_{-ji}|A_i\right)$ and $\frac{\partial \tilde{\mathcal{P}}\left(X_{ji}, b_{ji}, \tilde{X}_{-ji}, \tilde{b}_{-ji}|A_i\right)}{\partial b_{ji}}$.

    iii. Repeat Steps (i) and (ii) a large number of times and take the averages to obtain:

$$\widehat{\mathcal{S}}(X_{ji}, b_{ji}|A_i, v^k) = \frac{\sum_{l=1}^{L} \tilde{\mathcal{P}}\left(X_{ji}, b_{ji}, \tilde{X}^l_{-ji}, \tilde{b}^l_{-ji}|A_i\right)}{L} \tag{16}$$

$$\frac{\partial \widehat{\mathcal{S}}(X_{ji}, b_{ji}|A_i, v^k)}{\partial b_{ji}} = \frac{\sum_{l=1}^{L} \frac{\partial \tilde{\mathcal{P}}\left(X_{ji}, b_{ji}, \tilde{X}^l_{-ji}, \tilde{b}^l_{-ji}|A_i\right)}{\partial b_{ji}}}{L} \tag{17}$$

    where $\{\tilde{X}^l_{-ji}, \tilde{b}^l_{-ji}\}$ is the lth set drawn (from Step ii).

    iv. Using the above estimates of $\widehat{\mathcal{S}}(X_{ji}, b_{ji}|A_i, v^k)$ and $\frac{\partial \widehat{\mathcal{S}}(X_{ji}, b_{ji}|A_i, v^k)}{\partial b_{ji}}$ and Equation (11), obtain pseudo costs c_{ji}.[10]

    (b) Perform Steps i-iv to obtain c_{ji}s for each bid j in each auction i.

2. Generate the nonparametric distribution of costs for unobserved type v^k, using c_{ji}s obtained above as the datapoints and the posterior probabilities λ_{ik}s as the weights.

6.3 Implementation with Discrete or High Dimensional Seller Attributes

There are two implementation issues with the algorithm described above. First, it does not allow for discrete seller attributes (all the X_{ji}s are continuous). Second, it has very high data requirements because it relies on multi-dimensional KDEs, especially if there are a large number of seller attributes. In general, finite sample estimation of high dimensional KDEs is difficult even without unobserved states because of the curse of dimensionality. With multi-modalities due to unobserved states, this problem is exacerbated because there are clearly areas in the space where the density of the distribution is low. Without multi-modalities, it is possible to ease the data burden by employing large bandwidths. However, large bandwidths will smooth out multi-modalities, making the recovery of the underlying KDEs difficult. To address this issue, we suggest the following assumption (which we also implement in our estimation).

---

[10]The inferred costs are also referred to as 'pseudo costs' because they are derived by the researcher, and not observed in data.

Assumption 6. The joint distribution of seller attributes and bid prices is multiplicatively separable in seller attributes and bid prices as follows:

$$\mathcal{G}(X_{ji}, b_{ji}|A_i, v_i) = \mathcal{G}^X(X_{ji}|A_i) \times \mathcal{G}^b(b_{ji}|X_{ji}, A_i, v_i) \tag{18}$$

According to Assumption 6, v_i may affect the number of sellers entering the auction and the prices charged by those sellers, but not the attributes of those sellers. This is reasonable to the extent that sellers have complete control over price – certain types of sellers will not systematically avoid entering the market since they can always adjust their prices to account for v_i, upon entering. This is true, as long as the maximum bid is not binding. With a binding maximum bid, this may no longer be true because some types of sellers may prefer dropping out of the auction rather than bidding bounded prices. In our setting, the maximum bid is non-binding, hence the unobserved difficulty of the freelance job is unlikely to systematically deter high or low reputation sellers from bidding since both types of sellers can adjust their bids to account for the unobserved difficulty.

With some small modifications, the estimation algorithm described earlier can be made to accommodate Assumption 6. It would involve a more sequential first step estimation, as follows: within the first-step, first estimate $\mathcal{G}^X(X_{ji}|A_i)$, and then employ an EM-like algorithm to estimate $\mathcal{G}^b(b_{ji}|X_{ji}, A_i, v_i)$. Please see Web Appendix A.3 for detailed step-by-step instructions on the modified estimation procedure.

Two points of note regarding Assumption 6. First, it is an assumption of convenience, and it need not be made if the researcher is dealing with only a small number of seller attributes and has access to large datasets. Second, and more importantly, it is testable. In §9.2.2, we present empirical tests to confirm the validity of this assumption in our setting.

6.4   Discussion on Estimation Methodology

We now discuss how our estimation method relates to: (a) the estimation of auctions, and (b) the literature on accommodation of common-knowledge or persistent unobservables within two step methods.

Our estimation method builds on the nonparametric estimation of first price auctions pioneered by Guerre et al. (2000). However, there are significant differences between our method and theirs because of the nature of the beauty contests. In first price auctions, the expected probability of winning and its derivative are simply the cumulative density and probability density of the bid distribution at the observed bid. Thus, information on observed bids is sufficient to back out the distribution of seller costs. In contrast, in the beauty contest setting, we estimate both the joint distribution of bids and seller attributes, as well as the buyers' allocation rule. Moreover, expected probabilities of winning and their derivatives are not directly available from the first step estimates; rather they are obtained using simulations.

In terms of controlling for auction specific unobservables, compared to Krasnokutskaya's deconvolution estimator, our method has the advantage that it does not require the multiplicative separability assumption. Hence, it can be used to estimate auctions with outside options and auctions where multiplicative separability in costs/values does not translate to multiplicative separability in bids. It is also relatively simple to implement, even in complex settings such as beauty contests. However, unlike her method, which allows for a fully nonparametric distribution of unobservables, our method relies on the finite mixture assumption, that is, it only allows for a finite set of unobserved types.

More broadly, our method closely relates to the recently proposed CCP-based estimator by Arcidiacono

and Miller (2011), which allows for persistent unobservables in dynamic settings. The common theme in both approaches is the use of finite types to account for unobserved heterogeneity. However, a key difference is that in the first step we use a smoothed non-parametric EM-like algorithm proposed by Benaglia et al. (2009a), which employs a kernel smoothing procedure to accommodate continuous state spaces instead of the binning procedure advocated by Arcidiacono and Miller (2011). Further, in the second-step, we use the F.O.C condition to do a straightforward inversion of cost instead of using a maximum likelihood procedure.

7 Applying the Model and Estimation to Freelance Setting

We now adapt the framework described in §5 and §6 to suit our setting. Note that we have four seller attributes – two continuous (number of ratings and mean rating) and two discrete (seller geographic region and indicator of past buyer-seller interaction). The outline of our estimation strategy is as follows:

- First, we specify nonparametric models of the two continuous seller attributes, as functions of observed buyer and auction attributes A_i. See Web Appendix §A.4.1.

- Second, we specify a multinomial logit model of seller's geographic region as a function of observed buyer and auction attributes A_i, and seller's ratings (number and mean). See Web Appendix §A.4.2.

- Third, we specify a logit model of buyer-seller past interaction, as a function of observed buyer and auction attributes A_i, and seller's ratings (number and mean) and geographic region. See Web Appendix §A.4.3.

- Fourth, we distribute all the bids in the data based on observed buyer and seller characteristics, and derive nonparametric mixture distributions of bids. See §7.1 below.

- Fifth, we specify a flexible nested logit model of buyer decisions. See §7.2 below.

- Sixth, we utilize all the estimates from the previous steps to obtain nonparametric distributions of seller costs for the K unobserved auction types. See §7.3 below.

In principle, the second, third, and fifth steps can also be nonparametric. However, this is difficult in practice due to the size of the state space. Parametrization of some aspects of structural auction models to address state space issues has precedents in the literature (Bajari and Ye, 2003; Athey et al., 2011).

7.1 Nonparametric Estimation of Mixtures of Bid Prices

Figure 4 depicts the kernel density estimate of all the bids in the data. Note that it is lumpy and does not resemble any parametric distribution. Therefore, we employ a fully nonparametric estimation method. To estimate the nonparametric mixtures of bid prices, we first distribute all the bids in the data into bins based on auction and seller attributes as follows:

- Three groups based on number of bids: (a) No. of bids ≤ 13, (b) $13 <$ No. of bids ≤ 30, and (c) No. of bids > 30.

- Three groups based on sum of seller ratings: (a) Sum seller ratings $= 0$, (b) $0 <$ Sum seller ratings ≤ 90, and (c) Sum seller ratings > 90.

- Two groups based on seller geographic region: (a) Seller region $= 1$, and (b) Seller region $\neq 1$.

Note that our binning is coarser than our state space. As discussed earlier, this is necessary in large state space settings with finite samples. We experimented with a large number of binning strategies before finalizing this one. The chosen binning strategy is the one which accomplishes the following three goals to the best

possible extent – (1) Bin cut-offs should be such that the distributions of bid prices across bins should be as different as possible. (2) Bin cut-offs should be such that each bin has sufficient data for nonparametric estimation of K unobserved types. (3) Bin cut-offs should be such that each bin has approximately the same amount of data. Unbalanced binning will give rise to some bins with a large number of datapoints and others with very few datapoints. In such cases, estimation in smaller bins is likely to be biased.

Our binning strategy gives us a total of $3 \times 3 \times 2 = 18$ bins. Further, in all our analysis, we consider three unobserved types, $K = 3$. So we estimate a total of $18 \times 3 = 54$ bid distributions. Since we have around 44,000 bids, this gives us approximately 1000 datapoints in each distribution, if the unobserved types are equally distributed (which they are not). With an unbalanced distribution of unobserved types, some distributions will have fewer points. This is the main reason why further binning is not feasible for this dataset; doing so would introduce significant finite sample bias. Nevertheless, we performed many robustness checks with different binning strategies, and found that the results are, in general, robust to modifications in the bin cut-offs and the number of bins used.

For each of the 18 bins, we employ the nonparametric EM-like algorithm described in detail in Web Appendix §A.3, and retrieve the 54 component distributions and their population probabilities.

7.2 Nested Logit Model of Buyers' Equilibrium Allocation Rule

We model buyers' equilibrium allocation rule using a flexible nested Logit model. That is, in addition to the observed state variables, we introduce ϵ_{ji}, which is an unobserved seller-auction specific state variable that captures the unobserved preference of the buyer in auction i for seller j. ϵ_{ji} is a $(q_i+1) \times 1$ mean-zero vector with support R^{q_i+1} and is assumed to be independent of observed seller attributes X_{ji}. ϵ_{ji} is buyer i's private information and not observable to sellers. Hence, sellers cannot condition their bids on the realizations of ϵ_{ji}s. We further assume that the errors, ϵ_{ji}s, are drawn from a Generalized Extreme Value (GEV) distribution. All the bid options are in one nest, and the cancel option is in a separate singleton nest. Let $\sigma \in [0,1]$ be the correlation of errors in the nest with the bid options, where $\sigma = 0$ implies perfect correlation and $\sigma = 1$ indicates no correlation. Errors across nests are not correlated.

The probability that buyer i will choose to cancel, and the probability that i will choose the bid from seller j, given state variables $X_{ji}^n = \{A_i, X_{ji}, b_{ji}, X_{-ji}, b_{-ji}\}$ are given by:

$$\mathcal{P}(cancel_i | X_{ji}^n, \theta_w, \theta_v, \sigma) = \frac{1}{1 + \exp[\mathcal{W}(O_i, \theta_w) + \sigma \mathcal{I}(X_{ji}, b_{ji}, X_{-ji}, b_{-ji}, \theta_v, \sigma)]} \tag{19}$$

$$\mathcal{P}(\text{bid}_{ji} | X_{ji}^n, \theta_w, \theta_v, \sigma) = \frac{\exp[\mathcal{W}(O_i, \theta_w) + \sigma \mathcal{I}(X_i, b_i, X_{-ji}, b_{-ji}, \theta_v, \sigma)]}{1 + \exp[W(A_i, \theta_w) + \sigma \mathcal{I}(X_{ji}, b_{ji}, X_{-ji}, b_{-ji}, \theta_v, \sigma)]} \frac{\exp \frac{\mathcal{V}(X_{ji}, b_{ji}, \theta_v)}{\sigma}}{\sum_{k=1}^{q_i} \exp\left[\frac{\mathcal{V}(X_{ki}, b_{ki}, \theta_v)}{\sigma}\right]} \tag{20}$$

where $\mathcal{W}(O_i, \theta_w)$ is a function of observable buyer/auction level variables and parameter vector θ_w, $\mathcal{V}(X_{ji}, b_{ji}, \theta_v)$ is a function of seller j's attributes and bid price and parameter vector θ_v, and $\mathcal{I}(X_{ji}, b_{ji}, X_{-ji}, b_{-ji}, \theta_v, \sigma) = \ln\left[\sum_{k=1}^{q_i} \exp\left(\frac{\mathcal{V}(X_{ki}, b_{ki}, \theta_v)}{\sigma}\right)\right]$ is the inclusive value of the bid nest. This model is estimated using a maximum

likelihood procedure. The parameters $\{\theta_w, \theta_v, \sigma\}$ are obtained from the following maximization:

$$\underset{\theta_w, \theta_v, \sigma}{\mathrm{argmax}} \sum_{i=1}^{n} \left[\ln\left[\mathcal{P}(cancel_i|X_{ji}^n, \theta_w, \theta_v, \sigma)\right]^{I(d_i)=cancel} + \sum_{j=1}^{q_i} \ln\left[\mathcal{P}(\mathrm{bid}_{ji}|X_{ji}^n, \theta_w, \theta_v, \sigma)\right]^{I(d_i)=j} \right] \quad (21)$$

where d_i is buyer i's observed decision.

7.3 Estimation of Seller Costs

Finally, using the first step estimates, we estimate seller costs. Since we consider three unobserved types and 18 bins of observed states when binning the bid prices, we estimate 54 kernel density functions of costs, as described in Web Appendix A.3.2.

In general, in finite samples, some regions in the bid range could be empty either due to data deficiencies or because they are off-equilibrium. We handled these gaps by smoothing both $P(\cdot)$ and $G(\cdot)$. We now clarify the need and the impact of this smoothing. Smoothing over off-equilibrium paths is both necessary and common in two-step estimators (Bajari et al., 2007). In our model, we don't need $G(\cdot)$s to be smooth, since they only represent sellers' equilibrium beliefs on the distributions of other bids in the market. We use kernel smoothing for bids simply because from a conceptual standpoint, we believe that cost distributions should have full support. However, we do need mild smoothness requirements for the probability of winning $P(\cdot)$ in order to calculate the derivative of $P(\cdot)$ at optimal observed bids (Assumption 2). To do that, we need estimates of agents' beliefs on winning probabilities in off-equilibrium paths arbitrarily close to the equilibrium path. For example, if we only see \$10 and \$20 bids in data, we not only need estimates of $P(\cdot)$ at 10 and 20, but also at points very close to 10 and 20 to obtain the derivative of $P(\cdot)$ at these values. Note that this smoothness requirement is weaker than that needed in two-step estimators for discrete choice models, where researchers need to smooth over all off-equilibrium paths.

Empirically, large gaps in bid data can affect the estimates of the derivative of $P(\cdot)$ by making it heavily dependent on the smoothness operator used. To ensure that this is not an issue here, we examined the data further. First, we tested the data and found that there 372 unique prices between 85 and 500, *i.e.*, most integer prices are covered and the data is far from sparse. Second, we found that the average distance between two adjacent unique bids is only 1.398, *i.e.*, observed bids are mostly close by and hence the effect of any smoothing operator is likely to be insignificant. Third, in terms of estimation, we tried to make $P(\cdot)$ as flexible as possible.[11]

8 Results

8.1 First Step Estimates: Seller Attributes

First, we discuss estimates of the nonparametric joint distributions of number and average rating of bidders. There are no parametric results in this context except the bandwidths (μ_t) for the 16 categories. Since these bandwidths are not very informative in and of themselves, we do not present them here. However, we note that the kernel density estimates are very good at approximating the joint distributions of these two attributes.

Second, we discuss the results from the Multinomial logit model of seller region; see Table 14 in the Web

---

[11]Of course, data sparseness is a non-issue if we take a parametric stance on buyers' decision since the derivative of $P(\cdot)$ would then be available from the parametric model.

Appendix for estimates. In this model, we include all the buyer and auction specific attributes, seller mean rating, number of seller ratings, and their interactions as explanatory variables. There are several interesting points of note. First, we find that, everything else being constant, buyers in regions 1 and 2 are more likely to attract sellers from their own region. This effect likely stems from lower communication costs and similarities in Intellectual Property restrictions within a region. Next, we find that sellers bidding in popular auctions (with a large number of bids) are more likely to belong to the Indian sub-continent (region 1) and less likely to belong to developed countries (region 2). Buyers' past reputation on the website – number of uncanceled auctions hosted in the past, number of previous ratings, mean rating, *etc.* – also has a significant impact on the geographic location of the sellers they attract. For example, buyers who have hosted many auctions in the past and have canceled few auctions are more likely to attract sellers from developed countries and the Indian sub-continent (regions 1 and 2); and buyers with high average ratings are more likely to attract sellers from the Indian sub-continent and less likely to attract those from developed countries. Similarly, seller's reputation and interactions of seller and buyer reputation variables are also correlated with seller's geographic location. For example, sellers with the lowest mean ratings are likely to belong to the Indian sub-continent (region 1), followed by those from developed countries (region 2), followed by those from Eastern Europe (region 3).

Third, we discuss the findings from the logit model for the indicator of past buyer-seller interaction. (See Table 12 in the Web Appendix for estimates.) For this model, we include all the buyer and auction specific attributes, seller mean rating, number of seller ratings, the seller's geographic region, and their interactions as explanatory variables. The large negative constant term implies that, on average, only a small percentage of sellers are likely to have interacted with the buyer. Auctions that receive many bids and have detailed project attachments are less likely to attract sellers with whom the buyer has interacted in the past. Further, buyers who have a good amount of history on the website – many past auctions, high success ratio, few canceled auctions – are more likely to draw sellers with whom they have interacted in the past. Similarly, sellers with a good reputation on the site – high number of ratings and average rating – are more likely to have interacted with the buyer. Seller's geographic location also has a significant impact on the indicator for past interactions. For example, sellers from the Indian sub-continent and Eastern Europe (regions 1 and 3) are less likely, and those from developed countries (region 2) are more likely, to bid on auctions by buyers with whom they have interacted.

Finally, note that, for all the models discussed above, it is possible to include more elaborate combinations of buyer and bidder specific state variables. However, we found no improvement in fit or explanatory power with the increased complexity.

8.2 First Step Estimates: Nonparametric Mixture Distributions of Bid Prices

Because we consider three unobserved types of auctions $\{v_1, v_2, v_3\}$, we recover three nonparametric bid distributions for each type. Technically, the bids are also binned by their observed state variables, but to give a general idea of the extent to which unobserved type v_i affects bid prices, we pool all bids based on v_is (without separating them based on observables), and present three kernel densities, one for each type, in Figures 5, 6, and 7. The complete distribution of all the bids is shown in Figure 4 to give us a sense of how the overall distribution looks like before the split.

The three unobserved auction types are referred to as Low, Medium, and High. They are distributed as follows: Low type – 17.07%, Medium type – 46.47%, and High type – 36.45%. That is, nearly 17% of the

25

auctions consist of low difficulty or low cost jobs, nearly a third of them consist of very high difficulty or high cost jobs, whereas the majority of the jobs ($\approx 46\%$) are medium cost.

One common pattern in all the three KDEs is that there are clear modalities at multiples of 50. However, the similarities end there. Notice that the Low type KDE is first order stochastically dominated (FOSD) by the Medium type KDE, which in turn is FOSD by the High type KDE (Figure 8). For the Low type KDE, bids are dispersed over a large range of $100 to $500. The median of this distribution is $300, and the 75th percentile is 400. Less than 20% of the bids on this distribution are close to the MaxBid $500. In contrast, for the Medium KDE, the weight is more skewed towards the right, near $500, though there is significant amount of weight near $400 and $450, and some weight near $300 and $350. Here, the 25th percentile is $300, while the median is much higher at $450. Further, close to 40% of the bids in this type are $500 or more unlike the Low type KDE, where less than 20% of the bids were $500 or more. Finally, in the high KDE, almost all the weight is near the MaxBid $500. These are auctions for which all bidders uniformly bid either $500, or very close to it.

8.3 First Step Estimates: Buyers' Equilibrium Allocation Rule

The parameter estimates from the Nested Logit model need not be interpreted as primitives of buyers' utilities. Treating them as CCPs is sufficient to infer seller costs. However, if we want to model buyer entry or run full equilibrium counterfactuals, we need to treat them as structural parameters (see §8.6 and §10).

The estimates from the nested logit model are shown in Model N1 in Table 6. We find that buyers' probability of picking a bid is decreasing in price. The interaction effect of price and number of deadline days is positive – with longer deadlines, buyers tend to be less price sensitive. Next, we find that buyers tend to pick sellers with a good reputation on the website. There is a large negative coefficient for sellers with no past ratings. Also, buyers tend to pick sellers with more ratings compared to those with fewer past ratings, *i.e.*, the coefficient of the ln of number of ratings is positive. However, the marginal impact of each additional rating is decreasing, possibly because the new information in each new rating decreases as the number of ratings increase. This effect is consistent with the 'imperfect monitoring' problem in online reputation systems (Holmstrom, 1999; Cripps et al., 2004), because the number of reviews would be irrelevant if reviews were perfectly informative. The interaction of the ln of number of ratings and the centered mean rating is also positive. This implies that the impact of changes in a seller's rating (over the population mean) is proportional to the ln of the number of ratings she has received. That is, a seller with a mean rating of 9.5 and 20 ratings is less likely to be picked compared to a seller with a mean rating of 9.5 and 25 ratings, even after controlling for the main effects of the number of ratings. Overall, these results suggest that buyers view ratings on the website as an informative, but noisy measure of seller quality.

Buyers exhibit a strong tendency to pick sellers with whom they have worked in the past. In principle, the directionality of these effects could go either way, depending on whether the previous interactions went well or not. However, in the data, we find that, usually sellers avoid buyers who gave them low ratings previously. So the indicator for previous interaction almost always indicates good previous interactions. We also find that buyers exhibit a preference for sellers from eastern Europe, followed by those from developed countries, and buyers prefer not to pick sellers from the Indian sub-continent.

Buyers are more likely to choose a bid (as opposed to canceling the auction) if they post auctions with long deadlines and project attachments. This is possibly because serious buyers who intend to procure from the website are more likely to specify reasonably long deadlines and take the time to spell out project details. Buyers

who have few uncanceled and many successful past auctions are less likely to cancel. Further, buyers from region 1 are more likely, and those from region 2 are less likely to cancel. This effect likely stems from local labor costs, which are high in region 2 and low in region 1. Finally, the nesting parameter is estimated to be 0.371, which suggests that buyers' unobservable preferences for bids have a component that is correlated across bid options.

8.4   Second Step Results: Seller Cost and Market Power

Finally, we present the results on seller cost distributions. As with the bid distributions, these costs distributions can be further partitioned based on observables. However, we first present the cost distributions for the three unobserved types, and then deconstruct the sources of cost variation in §8.5.

CDFs of seller costs for Low, Medium, and High type auctions are shown in Figure 9 and their summary statistics in Table 8. The three cost distributions are significantly different from each other. As in the case of bid distributions, the cost distribution of Low type auctions is FOSD by that of Medium type auctions, which in turn is FOSD by the cost distribution of High type auctions. For Low type auctions, the costs are, on average, quite low and there is significant variation in costs across sellers. The median cost is \$195.55 while the 75th percentile is \$323.13. For Medium type auctions, the costs are generally higher; the median of the distribution is \$322.49. Finally, for High type auctions, the costs are even higher, with a median of \$365.

Next, we present the estimated distributions of margins for the three unobserved types in Figure 10. Margin distributions give us a measure of sellers' marketpower and the extent of competitiveness in the marketplace. We find that seller margins are, on average, 15%. That is, on average, if the seller bids \$100 and wins, her cost is \$70, she pays a commission of \$15 to the site, and gets to keep \$15. In terms of dollar amounts, almost all the margins lie between \$35 and \$100. Specifically, the median margins for sellers in Low, Medium, and High type auctions are \$61.81, \$72.16, and \$74.53, respectively. Our findings thus suggest that this marketplace is quite competitive and that sellers don't wield much marketpower.

Importantly, the estimated margin distributions for the three types are significantly different from each other. Recall that if the unobserved auction type is multiplicatively/additively separable from costs, then the distribution of seller margins would just be scaled up across auction types. However, that is not the case here. Thus, our findings affirm the invalidity of the multiplicative and additive separability assumptions here, and highlight the importance of allowing for non-multiplicatively separable common shocks to seller costs in auctions.

8.5   Explaining Differences in Sellers' Costs

We now examine how auction, buyer, and seller characteristics affect seller costs. Recall that costs are functions of auction and seller characteristics as well as that of seller's private shock, *i.e.*, $c_{ji} = c_{ji}(A_i, v_i, X_{ji}, \tilde{c}_{ji})$. In small state spaces, the impact of auction and seller attributes can be understood by simply examining cost distributions at each combination of these state variables. However in large state spaces, we need to impose some structure to further explore this issue. Hence, following Haile et al. (2003), we expand cost as:

$$c_{ji} = \mathcal{T}(A_i, X_{ji}) + \sum_{k=2}^{3} \omega_k I(v_i = v^k) + \tilde{c}_{ji} \tag{22}$$

where $\mathcal{T}(\cdot)$ is a function of observed auction and seller characteristics, $I(v_i = v^k)$ is the indicator that auction i is of unobserved auction type k (the effect of the Low type, $k = 1$, is set to zero), ω_ks are the coefficients for

the indicators, and $\tilde{c}_{ji}$ are i.i.d draws of sellers' private costs. $\mathcal{T}(A_i, X_{ji})$ can be estimated nonparametrically in small state spaces or through flexible parametric functions in larger state spaces. We can thus retrieve $\mathcal{F}(\tilde{c}_{ji})$, which is the distribution of sellers' costs after accounting for observable auction and bidder characteristics and unobserved auction type v_i. Then, the expected cost of seller j in auction i is:

$$E[c_{ji}] = \mathcal{T}(A_i, X_{ji}) + \sum_{k=2}^{3} \lambda_{ik} \omega_k + \tilde{c}_{ji} \tag{23}$$

where λ_{ik}s are posterior probabilities of auction i belonging to type k. Based on Equation (23), we regress the expected cost on A_i, X_{ji}, and posterior-weighted unobserved types. The results are shown in Table 7.

First, the unobserved type of the auction has a significant impact on sellers' costs. Compared to a low type auction, a medium type auction costs a seller \$109 more, and a high type auction costs her \$190 more. This reiterates the importance of accounting for unobserved auction heterogeneity. Second, we find that buyer-specific variables have a significant impact on seller costs. Sellers find it cheaper to work with buyers from Eastern Europe, buyers who give detailed project instructions (through project attachments), and those who have posted many successful auctions on the site. They also prefer experienced buyers and those who have a good reputation on the site. Recall that sellers also face an information asymmetry problem in this context. If a transaction runs into difficulties, then the seller may have to initiate a lengthy arbitration process through the site or forfeit her earnings. In general, experienced buyers who have successfully conducted many auctions in the past are less likely to make unreasonable demands and more likely to be clear about their requirements. Thus transacting with them is likely to be less costly compared to a new buyer who may be difficult to deal with both due to his ignorance of the landscape as well his lack of incentives to behave well. Third, sellers find it significantly more costly to transact with buyers with whom they have not interacted before. This mirrors our previous finding on buyers' preference for sellers with whom they have interacted.

Fourth, we find that the seller's own geographic location and past experience on the site have a significant impact on her costs. An interesting finding in this context is that new sellers have significantly lower costs compared to experienced and reputed sellers. At first glance, this may seem surprising because experienced and higher reputation sellers are more likely to have lower programming costs. However, this becomes understandable once we realize that there are other aspects to costs. New sellers need to build a reputation, so they have higher marginal benefits of winning a job and getting good ratings. On the other hand, the marginal benefit of winning another job and another rating is low for sellers who have done many jobs already. Sellers with few past ratings also have lower opportunity costs compared to experienced sellers, because the latter are more likely to win other auctions that they may bid on in the near future.[12]

8.6 Endogenous Buyer Entry

We now present a model of buyer entry. Note that modeling buyer or seller entry is not essential for obtaining unbiased estimates of seller costs as long as the maximum bid is non-binding and there are no acquisition costs (a natural assumption in Internet settings, since sellers don't need to spend significant resources to understand

---

[12]We do not include the number of bidders in this regression because we have an independent private values setting – the number of other sellers in the market has no impact on a seller's private cost of doing the job.

their own valuation); see Athey and Haile (2007). However, estimating a buyer entry model can give us insight into how different auction specific variables affect buyers' decision to post auctions, which in turn would allow us to run more realistic counterfactuals. Nevertheless, in order to do so, we need to take a stance on buyers' choices. Specifically, we need to treat the estimates from the nested logit model of buyers' decision (Table 6) as structural parameters that define the utility of profit-maximizing buyers.

8.6.1 Model of Buyer Entry

Before entering the auction, the buyer has uncertainty on both the number of bids he will get and their attributes. We have already modeled and retrieved the equilibrium distribution of bid attributes, $\mathcal{G}(X_{ji}, b_{ji} | O_i, q_i, v_i)$. So we now model buyers' expectations on the number of bids they expect to receive after entry.

Bid Arrival Model: A Poisson model is appealing for two reasons here. First, it models count data from independent events, and it is therefore appropriate for our sealed bid setting, where bids arrive independently. Second, a key feature of data generated by Poisson arrivals is the equality of mean and variance, an empirical regularity shared by our data (see Table 3).

The expected conditional probability of observing q_i bids as a function of auction specific observables is:

$$\mathcal{H}(q_i | O_i, r_i, \theta_p) = \sum_{k=1}^{K} \pi_k \frac{e^{\eta_{ik}} \cdot (\eta_{ik})^{q_i}}{q_i!} \tag{24}$$

where $\eta_{ik} = \exp(\{O_i, I(v_i = v^k), r_i\}' \cdot \theta_p)$, and θ_p is a parameter vector to be estimated. In the above equation, the unobserved type v_i is integrated out since it is not known to us.

For estimation, we include this bid arrival model within the EM-like loop, and recover both nonparametric estimates of bids and the parameters of the Poisson model as functions of the unobservable v_i, along with population distribution of types. These first step estimates are then used to recover the cost distributions.

Buyers' Decision to Post the Auction: Using the estimates from above, we now specify a buyer entry model. A buyer i chooses to post an auction/enter the market if his expected utility from doing so is greater than that from not entering. If we normalize the utility of not posting an auction to zero (similar to that from canceling), and assume that buyers' costs of making the actual post is negligible, we can write out i's entry decision as:

$$\mathcal{EU}_i + \epsilon_{i0}^{enter} > \epsilon_{i0}^{no-enter} \tag{25}$$

where the first term is the expected value of entering the auction (and making optimal decisions henceforth) and right hand side is the utility from not entering. The two error terms, $\epsilon_{i0}^{enter}, \epsilon_{i0}^{no-enter}$ are assumed to be i.i.d extreme value. The expected utility from entry can be expanded as:

$$\mathcal{EU}_i = \int_{q_i} \int_{(X_{ji}, b_{ji})|_{j=1}^{q_i}} \left[\mathcal{U}_i \cdot \mathcal{G}\left((X_{ji}, b_{ji})|_{j=1}^{q_i} | O_i, r_i, q_i, v_i \right) \cdot \mathcal{H}(q_i | O_i, r_i, v_i, \theta_p) \right] d\left(X_{ji}, b_{ji})|_{j=1}^{q_i} | O_i, r_i, q_i, v_i \right) dq_i \tag{26}$$

where $\mathcal{U}_i = 1 + \exp\left[\mathcal{W}(O_i, \theta_w) + \sigma \mathcal{I}((X_{ji}, b_{ji})|_{j=1}^{q_i}, \theta_v, \sigma) \right]$ is the expected utility from the auction for a given

29

draw of bids and $\mathcal{I}((X_{ji}, b_{ji})|_{j=1}^{q_i}, \theta_v, \sigma) = \ln\left[\sum_{j=1}^{q_i} \exp\left(\frac{\mathcal{V}(X_{ji}, b_{ji}, \theta_v)}{\sigma}\right)\right]$ is the inclusive value of the bid nest for

a given draw of number of bids and their attributes. Since the buyer doesn't know how many bids or what kinds of bids he will get upon entering the auction, we need to integrate $\mathcal{U}_i$ over his beliefs on the number of bids he expects as well as their attributes. Thus, the inside integral in Equation (26) is the integral over bid attributes for all the bids for a given draw of number of bids (or q_i) and the outside integral is the summation over the distribution of the number of bids, q_i.[13] Thus, in our estimation, we obtain numerical estimates of $\mathcal{EU}_i$ for each buyer simulating from the estimated equilibrium distribution of bids derived in §8.2 and the Poisson bid arrival model. Once we have $\mathcal{EU}_i$, we can obtain the entry probability of buyer i as:

$$P(enter|O_i, r_i, v_i) = \frac{e^{\mathcal{EU}_i}}{1 + e^{\mathcal{EU}_i}} \tag{27}$$

8.6.2 Results of Poisson Model and Entry Probabilities

We find that buyers with longer tenure on the site, and those who have posted many successful auctions and canceled few auctions, are likely to get more bids. (Please see Table 13 in the Web Appendix for the parameter estimates of the Poisson model.) We also find that buyers from the Indian sub-continent and Eastern Europe attract more bids, followed by those from developed countries. Finally, High and Medium type auctions attract fewer bids than Low type auctions, possibly because the supply of sellers who can perform low type jobs is larger.

In Figure 13, we present box plots of entry probabilities of buyers for the three auction types. There are two points of note here. First, entry probability is decreasing with unobserved auction difficulty/type. The average entry probability for a Low type auction is 0.635, whereas it is 0.532 for a High type auction. This difference ($\approx 19\%$) reiterates the importance of accounting for auction specific unobservables. The discrepancy in entry probabilities stems from the difference in the number of bids received and the equilibrium distribution of bid prices across auction types – Low type auctions get many more bids and significantly cheaper ones, increasing the expected value of entry for buyers. Second, after accounting for the entry model, the a priori (before entry) population distribution of the Low, Medium, and High type jobs/auctions is found to be: 16.78%, 45.92%, and 37.29%, respectively. While these post-hoc findings have no impact on the estimates of seller costs, they can have significant implications for counterfactuals; see §10 for details.

9 Validation and Robustness Checks

9.1 Validation

We now compare the results from our model with two others models – one with no unobserved heterogeneity (one type) and another with two unobserved types. The bias due to ignoring unobserved heterogeneity is measured using two metrics: 'Absolute Horizontal Distance' ($\mathcal{D}_A$) and 'Relative Horizontal Distance' ($\mathcal{D}_R$).

$$\mathcal{D}_A(\mathcal{F}_a, \mathcal{F}_b) = \int_0^1 |F_a^{-1}(s) - F_b^{-1}(s)| \mathrm{d}s; \qquad \mathcal{D}_R(\mathcal{F}_a, \mathcal{F}_b) = \int_0^1 \frac{|F_a^{-1}(s) - F_b^{-1}(s)|}{F_a^{-1}(s)} \mathrm{d}s \tag{28}$$

---

[13]Details of this simulation are available from the author for interested readers.

where $\mathcal{F}_a$ and $\mathcal{F}_b$ are the CDFs of the two distributions.

The model with no unobserved heterogeneity gives us a single cost distribution, shown in Figure 11. Let $\mathcal{F}_{3,L}, \mathcal{F}_{3,M}, \mathcal{F}_{3,H}$ be the CDFs of cost distributions for the three types when we allow for three unobserved types, and $\mathcal{F}_{1,L} \equiv \mathcal{F}_{1,M} \equiv \mathcal{F}_{1,H}$ be the CDFs of cost distributions for the three types when we allow for no unobserved type. Then the distance metrics comparing the respective distributions for each type are:

- $\mathcal{D}_A(\mathcal{F}_{3,L}, \mathcal{F}_{1,L}) = 56.67; \quad \mathcal{D}_R(\mathcal{F}_{3,L}, \mathcal{F}_{1,L}) = 90.83\%.$
- $\mathcal{D}_A(\mathcal{F}_{3,M}, \mathcal{F}_{1,M}) = 23.34; \quad \mathcal{D}_R(\mathcal{F}_{3,M}, \mathcal{F}_{1,M}) = 7.94\%$
- $\mathcal{D}_A(\mathcal{F}_{3,H}, \mathcal{F}_{1,H}) = 1.07; \quad \mathcal{D}_R(\mathcal{F}_{3,H}, \mathcal{F}_{1,H}) = 17.79\%$

Note that the performance of the model without unobserved heterogeneity is quite inferior compared to our main model with three unobserved types (see Figure 9 and Table 8). The 25th, 50th, and 75th percentiles of this distribution are $\{262.53, 363.49, 365.59\}$. Recall that the Low type distribution is characterized by $25^{th}, 50^{th}, 75^{th}$ percentiles of $\{107.70, 195.55, 323.13\}$. Hence, using the estimates from a model with no unobserved types will lead to significant overprediction of seller costs for Low type auctions. This is also affirmed by the high values of distance metrics. This model also overestimates the costs of sellers in Medium type auctions, where the 25th, 50th, and 75th percentiles are $\{236.13, 322.49, 365.28\}$, though the extent of overestimation is lesser. The distance metrics are lower for this type, though even here the relative distance metric is nearly 8%. Finally, not allowing for unobserved heterogeneity also leads to significant underprediction in the costs for more than 40% of the sellers in High type auctions. The distance metrics are lower than that for Low type auctions, but much higher than that for Medium type auctions.

Now consider a model with two unobserved types – Low type and High type. This gives us two cost distributions ($\mathcal{F}_{2,L}$ and $\mathcal{F}_{2,H}$) shown in Figure 12, and the inferred population probabilities of the two types are: Low type – 62.42% and High type – 37.57%. Further, we have:

- $\mathcal{D}_A(\mathcal{F}_{3,L}, \mathcal{F}_{2,L}) = 56.67; \quad \mathcal{D}_R(\mathcal{F}_{3,L}, \mathcal{F}_{2,L}) = 50.73\%.$
- $\mathcal{D}_A(\mathcal{F}_{3,M}, \mathcal{F}_{2,L}) = 23.34; \quad \mathcal{D}_R(\mathcal{F}_{3,M}, \mathcal{F}_{2,L}) = 12.73\%$
- $\mathcal{D}_A(\mathcal{F}_{3,H}, \mathcal{F}_{2,H}) = 1.07; \quad \mathcal{D}_R(\mathcal{F}_{3,H}, \mathcal{F}_{2,H}) = 0.37\%$

Note that the population probability of the High type auctions in this model is very similar to that in the model with three unobserved types (36.45%). The cost distributions for the High types in the two models are also very similar; see the distance metrics above and Figures 9 and 12. This implies that the model with only two unobserved types is able to almost perfectly recover the cost distribution and population distribution of High type auctions; however it puts both the Low and Medium type auctions into one single group (which we call Low type here). Thus the population probability of Low type auctions in a model with two unobserved types (62.42%) is approximately equal to the sum of population probabilities of Low and Medium type auctions from the model with three unobserved types (17.07% + 46.47% = 63.54%). Hence, the model with only two unobserved types overpredicts the costs of Low type auctions and underpredicts the costs of Medium type auctions. Because the population weights of Low type auctions is almost one third of the Medium type auctions, the overprediction of costs for Low types is much more severe than the underprediction for Medium type auctions.

A natural question that arises at this juncture is whether there are more than three unobserved types in the data, especially since the KDEs of the Low and Medium type auctions are multi-modal (Figures 5 and 6). Further exploration of the data however revealed that these modes lie at multiples of 50 and mainly stem from

31

within auction heterogeneity. Nevertheless, we did experiment with more than three types. Recall that in order to recover four types, we need seven or more bids per auction (see §5.5). Since more than 50% of the auctions receive at least seven bids, identification is theoretically feasible, if there are indeed four unobserved types in the marketplace. But we could not recover a fourth type. It is well-known that trying to recover more types than the number of true types leads to convergence issues in finite mixture algorithms because of identification problems, *i.e.*, it is not possible to identify four types if there exist only three types in the data. We thus conclude that three unobserved types is sufficient for this setting.

9.2 Robustness Checks

9.2.1 Allowing Buyer Choices to Depend on Unobservable Auction Heterogeneity

First, we examine the impact of relaxing Assumption 5, so that buyers' decisions are allowed to be functions of the unobserved auction specific variable, v_i. In this case, we include a parametric model of buyer behavior within the EM-like loop, and recover both nonparametric estimates of bids and a parametric model of buyer decisions as functions of the unobservable v_i, along with population distribution of types.[14] These first step estimates are then used to recover the cost distributions, as before. Estimating this expanded version of the algorithm takes significantly longer because the buyers' decision model has to re-estimated with posterior weights at every step of the EM-like loop.

The results from this extension are very similar to those from our base model. First, we present the expanded version of the Nested logit model of buyers' decision (Model N2) in Table 6. Note that except for an increased tendency of medium type bidders to cancel, the other estimates and the Log-likelihood are not very different. In fact, the estimated population distribution of types is almost the same as that in the base case: Low – 14.96%, Medium – 48.1%, and High – 36.94%. Cost estimates here are also quite similar to the base case. Overall, our findings suggest that the identification of the unobserved types comes from the variation in bids, and not from the variation in buyer decisions. This could be due to three reasons. First, it is possible that decisions are not fundamentally different across buyer types, after controlling for bid prices and other observables. Second, it is possible that we are unable to retrieve such patterns even if they exist because buyer decisions are not nonparametrically identified as functions of v_i. Third, the amount of information in bids may simply overwhelm any minor information from buyer decisions. Recall that an auction gets eleven bids on average; compared to the information in those bids, one decision (which is also a function of a large number of state variables) cannot be expected to contribute much to the identification of unobserved types.

Finally, while there exists a small discrepancy in the estimates from this model and that from the base case, we would not claim that this is more correct/appropriate, especially given the non-identification issue. The difference could simply stem from parametric bias introduced into the EM-like algorithm. So in our counterfactuals, we stick with the original model.

9.2.2 Testing the Validity of Assumption 6

Recall that our EM-like algorithm is essentially a fixed point algorithm. In an identified model, there is a unique fixed point. Hence, to test the validity of Assumption 6, we only need to examine whether the fixed point from an EM-like algorithm with Assumption 6 is the same as that from an algorithm that doesn't make

---

[14]The details of this expanded version of the EM-like algorithm are available from the author.

this assumption. The two fixed points can be equivalent if and only if $\mathcal{G}^X(X_{ji}|A_i) = \mathcal{G}^X(X_{ji}|A_i, v_i = k) \, \forall k$. Since we have estimated a fixed point of the system using Assumption 6, we already have an estimate of $\hat{\mathcal{G}}^X(X_{ji}|A_i, v_i)$. So if we can now show that:

$$\hat{\mathcal{G}}^X(X_{ji}|A_i, v_i = Low) = \hat{\mathcal{G}}^X(X_{ji}|A_i, v_i = Medium) = \hat{\mathcal{G}}^X(X_{ji}|A_i, v_i = High) \tag{29}$$

using the current posterior estimates, then we have essentially shown that the fixed point from the two algorithms are the same.

Using our estimated posteriors, we test the distributional equality of seller attributes for the three unobserved types. For each type k, we generate a nonparametric distribution $\hat{\mathcal{G}}^X(X_{ji}|A_i, v_i = k)$ by using X_{ji}s as datapoints and the estimated posterior probabilities $\hat{\lambda}_{ik}$s as weights for the datapoint, and then compare them across ks. We find that the distributions are statistically indistinguishable from each other. We present the findings for the reputation variables, ln(Sum of seller ratings), in Figure 18 in the Web Appendix. The distributions for all three types – Low, Medium, and High – are very similar; Kolmogorov-Smirnov tests confirm the statistical equality. The findings for seller's geographic region and indicator for past interaction are also similar. Together these tests confirm that Assumption 6 is valid for our setting.

More generally, we advocate the use of this assumption in future research since it has tremendous speed advantages and modest data requirements compared to the full algorithm. Further, since it is easily testable, the researcher can always go back to the full version of the algorithm if the validity tests fail.

9.2.3   Uncertainty on Number of Bids Received by the Auction

When placing her bid, a seller knows the number of bids that the auction has received so far, but does not know the exact number of bids it will receive when the auction ends. In the main model, we assumed that sellers have perfect information on the total number of bids received by the auction. We now relax that assumption and allow sellers to have uncertainty on the total bids the auction will eventually receive. Specifically, we model the tth seller's expectations on the number of bids she expects the auction to receive using a truncated version of the Poisson model:

$$\mathcal{H}(q_i|(q_i > t), O_i, v_i, r_i, \theta_p) = \frac{e^{\eta_i} \cdot (\eta_i)^{q_i}}{q_i!} \cdot \frac{1}{\Pr(q_i > t)} \tag{30}$$

Using the estimates from the Poisson model in §8.6, we now draw from the truncated Poisson specified above to generate seller t's beliefs on the number of bids received by auction i and use it to generate the expected probabilities of winning as follows:

$$\mathcal{S}(X_{ji}, b_{ji}|O_i, q_i, r_i, v_i) = \int\limits_{(q_i|q_i \geq t)} \int\limits_{(X_{ji}, b_{ji})|_{j=1}^{q_i}} \left[\begin{array}{l} \mathcal{P}(X_{ji}, b_{ji}, X_{-ji}, b_{-ji}|O_i, q_i, r_i, v_i) \cdot \mathcal{G}(X_{-ji}, b_{-ji}|O_i, q_i, r_i, v_i) \cdot \\ \mathcal{H}[q_i|(q_i > t), O_i, r_i, v_i, \theta_p] \cdot d(X_{-ji}, b_{-ji}|O_i, q_i, r_i, v_i) \cdot d(q_i|q_i > t) \end{array} \right]$$

The cost estimates from this more complex model (Model C2) and comparisons to the original (Model C1) are shown in Table 9. Overall, we find the cost estimates from the two models to be quite similar.

10 Counterfactual Simulations

We now use our estimates to answer policy questions that are of importance to managers of freelance sites. We present counterfactual simulations that examine the impact of changes to auction and site design on equilibrium outcomes. Each counterfactual experiment requires us to numerically re-solve for the equilibrium following a policy change. We briefly explain the solution process below.

However, before doing so, we make note of a couple of points. First, as usual, all our counterfactuals should be interpreted cautiously with the necessary caveats. For instance, buyers may change their preferences under counterfactual scenarios. Both buyers and sellers may be solving a dynamic across-auction optimization problem rather than a static within-auction problem; see §11.2 for details. They may also be substituting across different freelance sites. All these factors can influence outcomes in the real world. Second, because of these factors, we do not make broader social welfare comparisons or predictions. Rather, we prefer to stick to managerially relevant questions from the site's perspective.

For any regime change, there are two possible methods to solve for counterfactuals. First, treat the estimates of buyers' decisions as simply CCPs, ignore buyer entry, and re-solve sellers' decisions given CCPs. While this approach does not require us to take a parametric stance on buyer decisions, the downside is that it ignores the fact that buyers may choose not to enter the market under different counterfactual scenarios. The second option is to treat the estimates of buyers' decision models as structural parameters, and re-solve buyers' entry and choice decisions simultaneously with seller decisions to obtain the new equilibrium. With reliable estimates of buyers' entry and choice models (§8.6 and §7.2), this method is likely to give more realistic estimates of counterfactual outcomes because it solves for a full, rather than a partial equilibrium. In two-sided markets such as ours, full counterfactuals are valuable; so we take the latter approach.

To obtain the counterfactuals, we employ a backward solution strategy, *i.e.*, we first derive the equilibrium bidding strategies of sellers given buyers' entry and choice behavior. Next, we go back and solve for optimal entry from buyers' perspective. Then we combine these two steps to compute the equilibrium outcome for the entire system. Please see §A.5 in the Web Appendix for step-by-step details.

10.1 Who Are More Important: Buyers or Sellers?

Given our two-sided market setting, the most important question from a manager's perspective is: who are more important to the freelance site – sellers or buyers? While there may be positive externalities to growing both sides of the market, it may be profitable to focus on one of them due to resource constraints.

To answer this question, we conduct two sets of counterfactuals. In the first set, we increase the number of sellers in each auction by a specific amount, while keeping their distribution constant. We then recompute the new equilibrium outcomes (buyer entry, bidding strategy, and buyer choice) and derive the new auction clearance rates and revenue for the site. In the second set, we increase the number of buyers and proportionally redistribute the sellers in the system among all the auctions (so each auction gets fewer bids).[15] The results from the two experiments are shown in Figures 14 and 15. In both experiments, the effect of increased supply is positive. This is expected – everything else being constant, attracting either more sellers or more buyers has

---

[15]It is possible to do this without decreasing the number of bids in each auction, in which the results are stronger. However, it is more reasonable to model some drop-off in the number of bids per auction when increasing buyer supply.

a positive impact on the auction clearance rate and site revenue. However, the source of this increased revenue and its magnitude are quite different across the two cases.

Increasingly seller supply has the following effect – from sellers' perspective, it increases competition, which induces them to bid lower in equilibrium. This has a positive impact on buyer entry because more buyers enter the market in anticipation of better bid prices. Further, lower prices also mean that a larger fraction of posted auctions are successful since buyers are more likely to pick a bid rather than cancel. So overall, the site mediates more transactions (Figure 14). However, the increase in revenue is lower than that implied by the increase in clearance rate. This is because successful auctions are now clearing at lower prices due to increased competition. For instance, a 40% increase in number of sellers per auction leads to a 22% increase in the number of successful auctions, but only a 5.5% increase in revenue. Thus the lower clearance prices due to the competition effect largely overwhelm the higher clearance rates due to the market expansion.

Now consider the effect of increasing buyer supply. In this case, there are fewer bids per auction, and sellers increase their prices in response to lower competition. From buyers' perspective, this is doubly harmful because not only do they have fewer bids to choose from, but these bids are also more expensive. So they are less likely to enter the auction. Thus, the increase in buyer/auction supply doesn't translate to a proportional increase in successful auctions. However, the auctions that do clear, now do so at a higher price, leading to a significant increase in site revenues. In fact, in this case, the softer competition dominates the market contraction effect. For example, a 40% increase in auction/buyer supply only leads to a 17% increase in successful auctions, but a 25% increase in revenues; see Figure 15.

The recurring theme in these results is that the competition effect is stronger than the market expansion effect in this marketplace. Hence, for the same percentage increase, buyers always provide higher revenues than sellers. An important managerial implication of these findings is that the site should focus on attracting more buyers since they are more valuable. Note that the site can grow its buyer-base using many mechanisms – premium services for buyers, targeted advertising to attract buyers in developed countries, better monetary incentives, *etc*. We don't recommend a specific mechanism that the site should use since we do not have information on the costs of implementing these mechanisms, and hence cannot evaluate their relative merits.

10.2   Which Auction Type is More Important: Low, Medium, or High?

In the second counterfactual, we continue to study site revenues, but we focus on auction types instead of site participants. Recall that we retrieved three types of auctions – Low, Medium, and High, with significant price differentials across auction types; *e.g.*, the average price difference between High and Low type auctions is $150. Since the site's profits come from commission revenues, an important question that a manager faces is: should the site encourage high value auctions (*e.g.*, through lower commissions or other subsidies) and discourage low and medium value auctions.

To answer this question, we run three counterfactuals. In the first, we replace all High type auctions with Low and Medium type auctions using a proportional redistribution. Then we simulate the outcomes and recalculate site revenues. Similarly, in the second and third experiments, we replace Medium and High type auctions, respectively, and recalculate site revenues. The results from the three experiments are shown as stacked bar graphs in Figure 16, with the first bar denoting the revenues in the base case (normalized to 1). The stacking denotes the relative contribution of a specific auction type to the total revenue, and the size of the total

bar depicts the relative increase/decrease in total revenues compared to the base case.

Interestingly, we find that all three auction types are almost equally valuable. In the base case, the Low, Medium, and High type auctions contribute at almost the same proportion as their population distribution. When we replace High type auctions with Medium and Low type auctions, surprisingly the total revenue and the relative contribution of Low and Medium type auctions is almost the same. Note that while High type auctions invite higher bids and bring in larger commissions, they also clear at a lower rates due to higher prices. Thus, their overall contribution is not much higher than Low/Medium type auctions, which clear at lower prices, but do so with higher frequency. The same pattern repeats in the second and third experiments, for similar reasons. So the site should not promote one specific type of auction at the expense of others – all are equally valuable.

10.3 Value of Sellers from Different Geographic Regions

We now examine whether sellers from a specific geographic region are more valuable than those from other regions. If so, the site can incentivize these sellers over others. While it may be difficult and/or illegal to discriminate by region of residence, the site can adopt indirect mechanisms. For example, sellers in developing countries prefer trading using Paypal accounts while those from developed countries typically use credit cards or bank accounts. If the site varies commission rates by payment mechanisms, then it can accomplish its goals without direct discrimination. The issue of seller value by region is of relevance to policymakers too, since online freelancing is a growing industry that is increasingly contributing offshore outsourcing of jobs.

We conduct four experiments in this context. In each experiment we replace sellers from a specific region by proportionally substituting sellers from the other three regions. Both buyers and sellers are allowed to respond to the changes in the system by modifying their own strategic behavior. For example, sellers are allowed to change their bids in response to the change in competition, and buyers are allowed to modify their entry and choice decisions. We also preserve the correlations between other seller attributes within the seller geographic regions to ensure that the results are as realistic as possible. We present the results in Figure 17. The first two bars show the number of successful auctions and revenue in the base case, which are normalized to one. The next set of bars depicts these metrics (relative to the base case) for a scenario where sellers from Indian sub-continent (region 1) are proportionally replaced by sellers from other regions. Similar bar graphs for Regions 2, 3, and 4 are shown next.

Buyers' high and low preferences for sellers from developed countries (region 2) and the Indian sub-continent (region 1), respectively, has already been established (see Table 6). However, it does not necessarily follow that sellers from region 2 (region 1) are more (less) valuable to the site since sellers compensate for this in their bids. Consider the effects at play in the first experiment. First, knowing buyers' preferences, sellers from region 2, 3, and 4 are likely to submit higher bids than those from region 1. This is the primary effect of changing the distribution of seller types. Of course, since this is an equilibrium setting, sellers will increase their prices a bit more than that implied by the first-order change because they realize that their competitors are also increasing their prices. This second-order reaction has the effect of softening competition even more. From the buyers' perspective, this is not unalloyed good news – while they have access to more sellers from preferred regions, these sellers are also charging higher prices. If the increase in bid prices is not offset by the increase in utility from seller regions, buyers are not only less likely to enter, but also more likely to cancel posted auctions. If the fraction of auctions clearing falls, then this loss has to be offset by the increase in commission revenues from higher prices for the experiment to be profitable. Thus, the overall directionality and magnitude of the

effect of changing seller regions is not clear a priori.

The main finding from our numerical experiments is that sellers from the Indian sub-continent are the least valuable to the site, and if the site can replace them with sellers from other regions, it is likely to be better off. Similarly, we find that sellers from developed countries are the most valuable to the site and losing them can negatively impact site revenues. The role of sellers from regions 3 and 4 is negligible. Note that we are not advocating that the site dissuade sellers from the Indian sub-continent; as shown in §10.1, any increase in seller supply is good for the site. However, when choosing between expending resources to attract two sellers with similar attributes, the site should choose those from region 2 over region 1.

11   Discussion of Limitations and Directions for Future Research

We now discuss some limitations of our approach, their potential solutions, and avenues for future research.

11.1   Seller level Unobservables

In our setting, the site does not allow private communication between buyers and sellers before the auction ends. Also, most sellers are small time players with very little or non-existent brand equity. So we believe seller-specific unobservables are not a first order concern in this setting. Nevertheless, buyers may be able to observe aspects of seller ability unobservable to the researcher, *e.g.*, written comments accompanying seller ratings. Sellers may condition their prices, and buyers their choices, on such unobservables. Not accounting for such unobserved heterogeneity can, in principle, bias our estimates of CCPs. This problem is similar to the price endogeneity problem in the demand estimation literature (Nevo, 2000). In our data, we don't observe sufficient repeat observations per seller to nonparametrically identify seller level unobservables. Nonetheless, we discuss a few approaches to address this issue, if data is available.

First, if we assume that all sellers know the population distribution of unobserved ability in the market, then it may be possible to extend our estimator as follows: assume that unobserved seller abilities are drawn from a set of finite types that affect seller costs, and then follow the general estimation method proposed, with the EM-like algorithm looping over both seller and buyer types. While this method preserves the nonparametric nature of our estimator, it is not feasible if there are a large number of sellers in each auction, or if seller churn is high, or if there are not sufficient repeat observations on sellers (as in our setting).

Second, if we are willing to impose some structure on buyers' unobserved preferences, we can adopt the well-known BLP estimation (Berry et al., 1995) procedure to address price endogeneity in the CCPs. However, BLP requires accurate estimates of market shares obtained from a large number of repeat interactions between the same set of sellers. This is generally not available in auction contexts (especially online ones) because there are hardly any settings where the same sellers repeatedly bid against each other.

Alternately, one could take the novel approach proposed in a recent working paper by Krasnokutskaya et al. (2013), which does not require market-share data; instead it tries to address this issue by deriving the relative ranking between sellers using data from a large number of repeat observations of sellers. However, a significant downside of this approach, is that its identification and estimation firmly rests on the assumption of no unobserved auction heterogeneity, an untenable assumption in many auction settings. Recall that over 30% of the variation in bids in our data is explained by unobserved auction fixed effects. Further, it requires the distribution of the unobserved quality of transitory sellers to be the same as that of long term sellers. This

is not true in most online settings (including ours) because of the well-known cheap-pseudonym phenomenon – low quality sellers will repeatedly milk their reputation and reappear under new pseudonyms; so the quality distribution of transitory sellers will be decidedly lower than that of long term players (Kreps and Wilson, 1982; Friedman and Resnick, 2001; Dellarocas, 2003). Finally, note that none of these methods can account for seller-specific unobservables that vary across auctions; *e.g.*, the match of a web-development expert will be higher in auctions for web-development jobs compared to Java coding jobs. Nevertheless, we hope that some aspects of the methods discussed above can be incorporated into our estimation to simultaneously address both auction and seller specific unobservables, in the future.

11.2 Dynamics

Finally, there could be across auction dynamics that affect buyer and seller behavior. First consider buyer-side dynamics; *e.g.*, a buyer may be more likely to cancel the auction or be extremely selective in his decision if he has the chance to post the auction again in the future. However, buyers' dynamic considerations will not bias estimates of seller costs because our estimator does not require us to take a stance on buyer's decision process, *i.e.*, we only use CCPs of buyers' decisions in our estimation. In fact, this is one of the advantages of being agnostic to the buyer's decision process and simply working with CCPs. However, if we want to run counterfactuals where we modify the buyer's decision, then we need to take a parametric stance on buyer's decision (as we did in §10), in which case the counterfactual results may suffer from dynamic bias. Thus, the decision to model entry or conduct counterfactuals where buyer behavior is modified and the faith that a researcher places on such results should depend on the application context and goals of the research.

Next, consider sellers' side dynamics, *e.g.*, if buyers tend to re-post canceled auctions in the future, or if there are multiple auctions being conducted at the same time and sellers are resource constrained, they may shade their bids higher, which can bias our estimates of seller costs (Zeithammer, 2006, 2007).

In our application setting, we found that less than 5% of buyers post an auction in the same category within a month of canceling one. Moreover, closer inspection of these auctions revealed that most of them were for new projects, and not re-posts of previous auctions. Hence, dynamics due to buyer-side considerations is not an issue here. Further, the site is extremely competitive with an average of $11+$ bids per auction and a high cancel rate. Hence, sellers don't have to worry about winning too many auctions or trading-off between auctions. In fact, of the 810 winning sellers in our dataset, 683 are unique. So we don't expect across sellers' side dynamics to be a significant source of bias either.

Nevertheless, extending beauty contests to allow for dynamics is an important next step. We are optimistic that some of the methods proposed in Jofre-Bonet and Pesendorfer (2003) and Backus and Lewis (2012) for first and second price auctions can be imported to solve dynamic beauty contests.

12 Conclusion

Beauty contests are procurement mechanisms where the auctioneer does not specify an allocation rule; instead he picks a winner based on price as well as other considerations such as seller reputation. Unlike traditional price-only auctions, beauty contests have no closed-form bidding strategies and suffer from non-multiplicatively separable unobserved auction heterogeneity, which makes their estimation challenging.

In this paper, we present an empirical framework to model and estimate seller costs in beauty contest

auctions. We formulate beauty contests as two stage games of strategic interaction with incomplete information, and present a two step method to estimate them. Our proposed method of formulating and estimating beauty contest auctions offers many advantages. First, it can be used to estimate auctions that don't have pre-specified allocation rules. Second, it is computationally simple and does not require us to solve for equilibrium bidding strategies, which is a challenging task in this complex setting. Third, it does not impose any optimality assumptions on the buyer or third-party site that conducts the auction. Fourth, it does not require any parametric assumptions on seller types, seller attributes, or bid distributions.

Importantly, our method provides a clean solution to the problem of auction specific unobserved heterogeneity that does not require multiplicative separability assumptions, and it can be applied to a wide range of auction settings. We show that it is possible to accommodate auction specific unobservables through finite unobserved types in our two step estimator. We present a constructive proof of identification and propose a nonparametric EM-like algorithm to recover the nonparametric estimates of underlying bid distributions as well as the population distribution of unobserved types in the first step itself. These first stage estimates are then used to derive the nonparametric distribution of sellers' private costs.

We apply our method to data on beauty contest auctions from a leading freelance site. We derive the inferred dollar values of sellers' costs in this marketplace and show that cost differences across freelancers can be explained by heterogeneity in geographic location, past experience on the site, previous interactions with the buyer, and by unobserved auction difficulty. We show that there are three unobserved types of auctions, with the following distribution: Low type – 17.07%, Medium type – 46.47%, and High type – 36.45%. The inferred cost distribution for Low type auctions is FOSD by that of Medium type auctions that have moderate costs, which in turn is FOSD by the distribution of costs for High type auctions. The sellers in this marketplace don't enjoy high margins/marketpower, with average percentage margins around 15%, *i.e.*, the marketplace is quite competitive. We also find that not accounting for unobserved heterogeneity can significantly bias the estimates of seller costs, *e.g.*, the Relative Horizontal Distance between the estimated distributions for Low types with and without unobserved heterogeneity is more than 90%. Based on our estimates, we run counterfactuals to provide recommendations to the freelance site, *e.g.*, placing a greater focus on growing its buyer side.

In sum, our paper makes three key contributions to literature. First, from a methodological perspective, we provide an empirical framework to model and estimate beauty contest auctions. Our framework is fairly general and can be adapted to suit a large class of auction problems that lack observable buyer allocation rules and closed form solutions to seller strategies. An important contribution is the method's ability to handle non-multiplicatively separable auction specific unobservables. Second, from a substantive perspective, we derive the cost distributions of sellers in a prominent freelance marketplace and show that freelancing is a very competitive industry with low seller margins. We also show that not accounting for unobservables in this market can bias estimates of seller costs. Third, from a normative perspective, our work offers guidelines to managers of freelance sites regarding optimizing the incentive mechanisms on their websites.

Figures and Tables
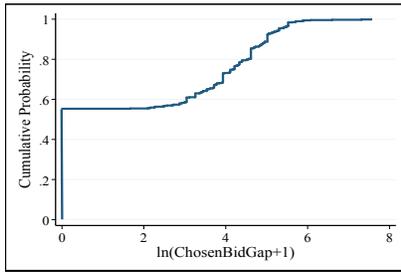
Figure 1: ln(ChosenBidGap + 1) for auctions with two or more bids.
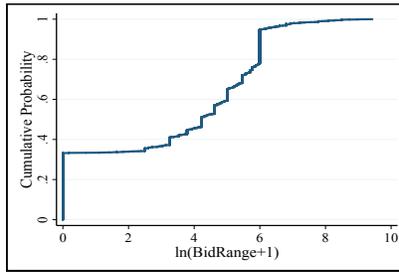
Figure 2: ln(BidRange + 1) for auctions with two or more bids.
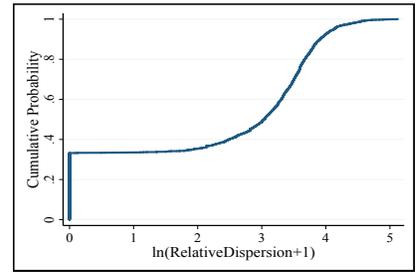
Figure 3: ln(RelativeDispersion + 1) for auctions with two or more bids.
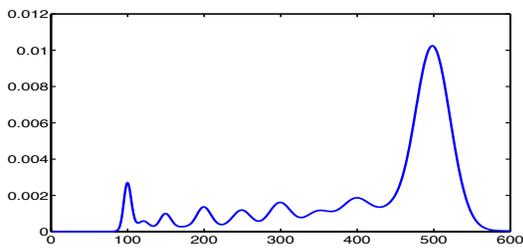
Figure 4: Kernel density estimate of all the bids in the data.
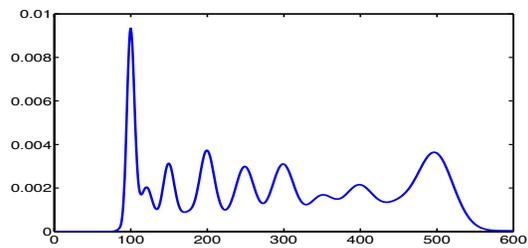
Figure 5: Kernel density estimate of bids from Low type auctions.
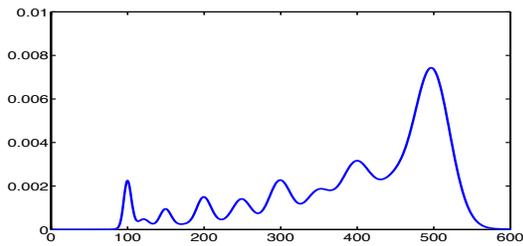
Figure 6: Kernel density estimate of bids from Medium type auctions.
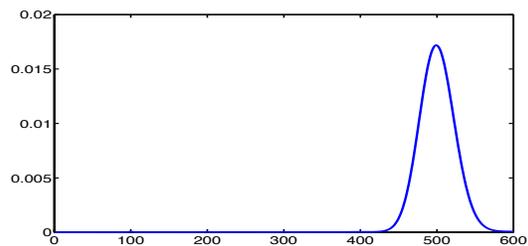
Figure 7: Kernel density estimate of bids from High type auctions.
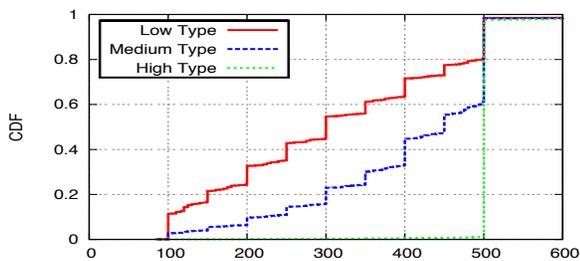
Figure 8: Cumulative density functions of bid distributions for Low, Medium, and High types.
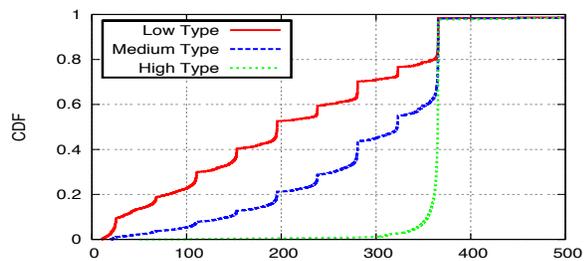
Figure 9: Cumulative density functions of estimated costs for Low, Medium, and High types.
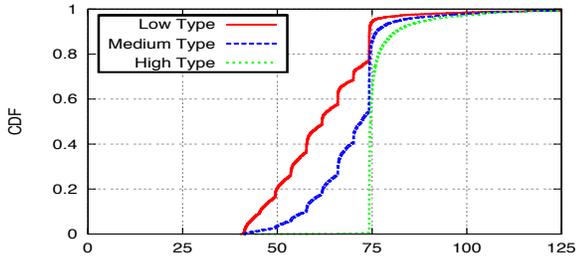
Figure 10: Cumulative density function of estimated margins for Low, Medium, and High types.
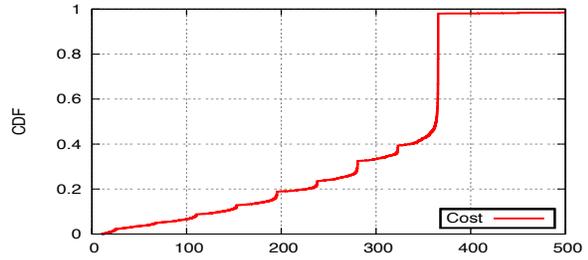
Figure 11: Cumulative density function of estimated cost with one type.
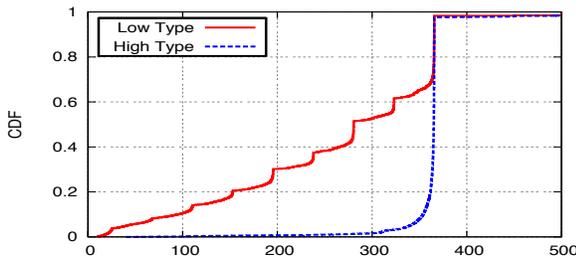
Figure 12: Cumulative density function of estimated costs for two unobserved types.
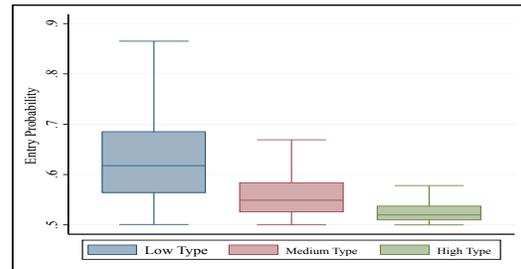
Figure 13: Distribution of buyer entry probability by unobserved auction type.
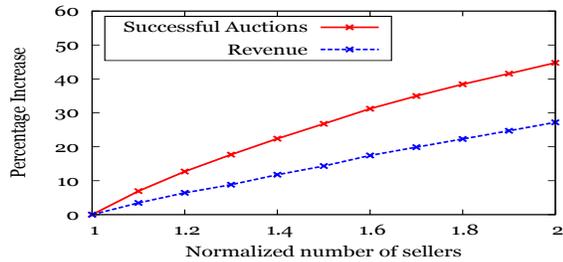
Figure 14: Impact of increasing seller supply.
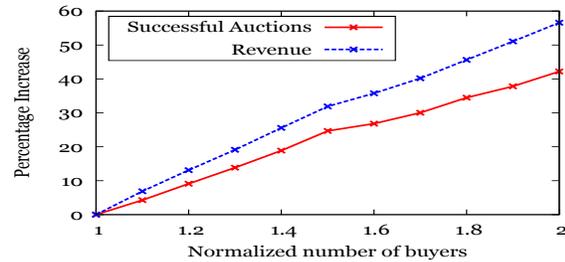
Figure 15: Impact of increasing buyer supply.
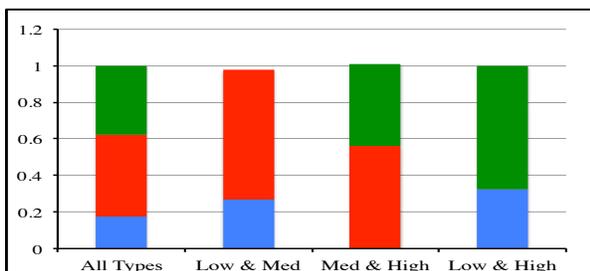
Figure 16: Impact of excluding a specific auction types.
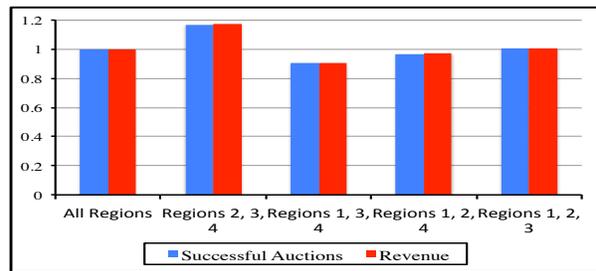
Figure 17: Impact of excluding sellers from a region.

Region Code	Countries
1	Indian sub-continent – India, Pakistan, *etc.*
2	Developed countries – USA, Western Europe, *etc.*
3	Eastern Europe – Romania, Russia, *etc.*
4	Everything else – Philippines, China, *etc.*

Table 1: Code for buyer and seller geographic regions.

Buyer Region (%)	Seller Region (%)			
	Region 1	Region 2	Region 3	Region 4
1 : 5.67	66.83	10.14	10.03	13.00
2 : 82.68	53.47	18.85	14.30	13.39
3 : 2.25	58.35	8.93	16.78	15.95
4 : 9.40	58.80	12.57	14.47	14.16

Table 2: Region distributions of auction and bids in the data . E.g., the leftmost column shows the percentage of auctions initiated by buyers from different regions.

Auction and buyer attributes	Mean	Std. Dev	25th	50th	75th	(Min, Max)	Sample size
	All auctions						
Number of bids received	11.06	12.27	3	7	14	(0, 137)	4002
Deadline (days)	19.21	32.05	0	13	30	(1, 1132)	4002
Number of buyer ratings	9.96	42.66	0	0	5	(0, 723)	4002
Avg. ratings	4.62	4.89	0	0	10	(0, 10)	4002
Avg. ratings (if rated)	9.73	0.83	9.89	10	10	(1, 10)	1900
No. of past uncanceled auctions	7.02	18.09	0	0	5	(0, 262)	4002
No. of past canceled auctions	8.51	29.18	0	1	6	(0, 597)	4002
Indicator for attachment with auction	Freq. $0 = 3300\,(82.46\%)$, Freq. $1 = 702\,(17.54\%)$						4002
	Auctions where buyer chose a bid						
Number of bids received	12.52	11.14	6	10	16	(1, 87)	810
Deadline (days)	26.21	51.72	5	14	30	(1, 1132)	810
Number of buyer ratings	13.28	57.54	0	1	8	(0, 603)	810
Avg. ratings	5.50	4.91	0	9.6	10	(0, 10)	810
Avg. ratings (if rated)	9.85	0.46	9.96	10	10	(6.88, 10)	452
No. of past uncanceled auctions	9.62	16.81	0	2	11	(0, 119)	810
No. of past canceled auctions	4.49	11.33	0	1	5	(0, 211)	810
Indicator for attachment with auction	Freq. $0 = 581\,(71.73\%)$, Freq. $1 = 229\,(28.27\%)$						810

Table 3: Summary statistics of auction and buyer attributes.

Seller and bid attributes	Mean	Std. Dev	25th	50th	75th	(Min, Max)	Sample size
	All bids received						
Bid price	434.83	230.87	350	500	500	(85, 5000)	44274
Bid price (if ≤ 500)	416.98	122.85	350	500	500	(85, 5000)	43393
Number of seller ratings	18.64	49.41	0	2	16	(0, 1343)	44274
Avg. ratings	5.71	4.52	0	8.54	9.79	(0, 10)	44274
Avg. ratings (if rated)	8.99	1.62	8.79	9.6	10	(1, 10)	28101
Indicator for past interaction with buyer	Freq. $0 = 44045\,(99.48\%)$, Freq. $1 = 229\,(0.52\%)$						44274
	Accepted bids						
Bid price	306.84	151.79	180	300	500	(85, 500)	810
Bid price (if ≤ 500)	306.84	151.79	180	300	500	(85, 500)	810
Number of seller ratings	46.35	94.46	6	19	52	(0, 1343)	810
Avg. ratings	8.85	2.77	9.44	9.79	10	(0, 10)	810
Avg. ratings (if rated)	9.7	0.41	9.56	9.83	10	(6.5, 10)	739
Indicator for past interaction with buyer	Freq. $0 = 732\,(90.37\%)$, Freq. $1 = 78\,(9.63\%)$						810

Table 4: Summary statistics of seller and bid attributes.

Auction, buyer & seller specific	Model M1		Model M2	
explanatory variables	Coefficient	Std. error	Coefficient	Std. error
Buyer region = 1	3.273×10^{-2}	1.516×10^{-2}	1.723	1.296
Buyer region = 2	-1.564×10^{-2}	1.889×10^{-2}	-1.487	9.013×10^{-1}
Buyer region = 3	3.878×10^{-2}	1.337×10^{-2}	-9.763×10^{-1}	1.236
Number of bids	-2.863×10^{-3}	2.634×10^{-4}	-9.115×10^{-2}	5.128×10^{-2}
Square of number of bids	1.750×10^{-5}	2.780×10^{-6}	2.939×10^{-3}	1.398×10^{-3}
Ind. auction attachment = 1	-9.390×10^{-3}	6.002×10^{-3}	-1.796	9.941×10^{-1}
ln(Deadline_days+1)	1.146×10^{-2}	1.531×10^{-3}	5.610×10^{-1}	2.400×10^{-1}
Buyer's success ratio	-1.533×10^{-2}	1.389×10^{-2}	-1.036	8.449×10^{-1}
ln(No. of auctions posted by buyer)	3.385×10^{-2}	5.537×10^{-3}	-7.113×10^{-1}	9.542×10^{-1}
ln(No. of auctions uncanceled by buyer)	-1.118×10^{-2}	8.376×10^{-3}	-4.135×10^{-1}	6.285×10^{-1}
ln(Buyer tenure in days + 1)	5.125×10^{-3}	1.445×10^{-3}	3.975×10^{-1}	2.241×10^{-1}
Indicator no. of buyer ratings = 0	8.455×10^{-2}	6.775×10^{-2}	-1.565×10^{2}	5.265×10^{1}
ln(No. of buyer ratings + 1)	-2.726×10^{-3}	5.916×10^{-3}	6.023×10^{-1}	5.675×10^{-1}
Buyer mean rating (centered)	4.172×10^{-3}	6.862×10^{-3}	-1.609×10^{1}	5.446
ln(No. of buyer ratings + 1) × Buyer mean rating (centered)	-1.070×10^{-2}	4.905×10^{-3}	8.596	2.977
Indicator no. of seller ratings = 0	-3.127×10^{-2}	1.542×10^{-2}	1.076×10^{-2}	1.417×10^{-2}
ln(No. of seller ratings + 1)	3.022×10^{-2}	2.136×10^{-3}	3.445×10^{-2}	1.901×10^{-3}
Seller mean rating (centered)	-8.723×10^{-3}	1.755×10^{-3}	-1.493×10^{-3}	1.611×10^{-3}
Buyer mean rating (centered) × Seller mean rating (centered)	-1.749×10^{-4}	1.033×10^{-4}	-1.886×10^{-4}	8.940×10^{-5}
Seller region = 1	9.950×10^{-2}	1.733×10^{-2}	6.180×10^{-2}	1.393×10^{-2}
Seller region = 2	4.206×10^{-2}	2.455×10^{-2}	5.030×10^{-2}	1.979×10^{-2}
Seller region = 3	3.560×10^{-3}	2.222×10^{-2}	1.517×10^{-2}	1.833×10^{-2}
Seller region = 1 & Buyer Region = 2	-2.786×10^{-3}	1.894×10^{-2}	-7.567×10^{-3}	1.541×10^{-2}
Seller region = 2 & Buyer Region = 2	-1.066×10^{-2}	2.629×10^{-2}	-2.625×10^{-2}	2.129×10^{-2}
Seller region = 3 & Buyer Region = 2	-3.996×10^{-2}	2.443×10^{-2}	-4.150×10^{-2}	2.005×10^{-2}
Seller region = Buyer region $\neq$ 2	3.778×10^{-3}	1.480×10^{-2}	-8.242×10^{-3}	1.249×10^{-2}
Indicator for no interactions b/n buyer & seller	1.570×10^{-2}	3.999×10^{-2}	1.434×10^{-1}	3.107×10^{-2}
Constant	5.807	4.524×10^{-2}	5.181	6.860×10^{-1}
Auction Dummies	No		Yes	
Adjusted R-squared	0.0278		0.3548	
No. of Observations	44274		44274	

Table 5: Bid price regressions. Dependent variable: $\ln(bid_{ji})$. Robust standard errors shown.

Explanatory variables (X_{ji}^n)	Coefficients varying within the bids nest(θ_v)			
	Model N1		Model N2	
	Coefficient	Std. error	Coefficient	Std. error
price	-6.007×10^{-3}	4.608×10^{-4}	-6.246×10^{-3}	4.791×10^{-4}
price $\times$ Buyer's success ratio	2.532×10^{-4}	4.196×10^{-4}	2.633×10^{-4}	4.363×10^{-4}
price $\times$ ln(Deadline_days+1)	2.529×10^{-4}	1.303×10^{-4}	2.629×10^{-4}	1.354×10^{-4}
Indicator for zero seller ratings	-3.560×10^{-1}	8.132×10^{-2}	-3.702×10^{-1}	8.456×10^{-2}
ln(Number of seller ratings + 1)	1.637×10^{-1}	2.129×10^{-2}	1.702×10^{-1}	2.213×10^{-2}
ln(Number of seller ratings + 1) $\times$ Seller mean rating (centered)	2.065×10^{-1}	1.962×10^{-2}	2.147×10^{-1}	2.040×10^{-2}
Seller region = 1	-1.891×10^{-1}	6.627×10^{-2}	-1.966×10^{-1}	6.891×10^{-2}
Seller region = 2	2.556×10^{-1}	1.179×10^{-1}	2.657×10^{-1}	1.226×10^{-1}
Seller region = 3	3.093×10^{-1}	6.662×10^{-2}	3.216×10^{-1}	6.927×10^{-2}
Seller region = Buyer region	2.548×10^{-1}	1.136×10^{-1}	2.649×10^{-1}	1.181×10^{-1}
Indicator for no buyer-seller past interaction	-1.242	1.510×10^{-1}	-1.291	1.570×10^{-1}
	Coefficients common for bid nest (θ_w)			
ln(Deadline_days+1)	1.526×10^{-1}	5.316×10^{-2}	1.552×10^{-1}	3.400×10^{-2}
ln(No. of buyer ratings + 1)	-4.726×10^{-1}	1.101×10^{-1}	-4.624×10^{-1}	1.091×10^{-1}
ln(No. of auctions uncanceled by buyer)	1.392	1.322×10^{-1}	1.368	1.161×10^{-1}
ln(No. of auctions canceled by buyer)	-9.540×10^{-1}	8.495×10^{-2}	-9.496×10^{-1}	7.669×10^{-2}
Indicator that auction has attachment	1.048	1.124×10^{-1}	1.058	1.126×10^{-1}
Buyer region = 1	-1.930	5.135×10^{-1}	-1.927	5.131×10^{-1}
Buyer region = 2	4.286×10^{-1}	1.853×10^{-1}	4.157×10^{-1}	1.855×10^{-1}
Buyer region = 3	-2.733×10^{-1}	3.863×10^{-1}	-2.802×10^{-1}	3.846×10^{-1}
Indicator high type			1.567×10^{-1}	1.554×10^{-1}
Indicator medium type			-4.067×10^{-1}	1.275×10^{-1}
Constant	-5.327×10^{-1}	2.788×10^{-1}	-3.065×10^{-1}	2.250×10^{-1}
Nest correlation	3.710×10^{-1}	2.343×10^{-2}	3.858×10^{-1}	2.200×10^{-2}
No. of auctions, Log-likelihood	$4002, -2323.155$		$4002, -2312.002$	

Table 6: Nested Logit Estimates of Buyers' Equilibrium Allocation Rule.

Explanatory variables	Coefficient	Std. error
Buyer region = 1	−1.377	3.633
Buyer region = 2	−2.131	2.749
Buyer region = 3	−10.97	3.447
Ind. auction attachment = 1	−7.940	1.725
ln(Deadline_days+1)	1.561	0.576
Buyer's success ratio	−7.678	3.217
ln(No. of auctions uncanceled by buyer)	−5.469	2.121
ln(Buyer tenure in days + 1)	2.338	0.462
Indicator no. of buyer ratings = 0	42.27	16.08
ln(No. of buyer ratings + 1)	5.247	2.258
Buyer mean rating (centered)	3.009	1.621
ln(No. of buyer ratings + 1) × Buyer mean rating (centered)	−2.533	1.097
Indicator no. of seller ratings = 0	−8.252	4.253
ln(No. of seller ratings + 1)	5.671	0.859
Seller mean rating (centered)	−1.474	0.474
Seller region = 1	20.47	2.104
Seller region = 2	10.41	3.836
Seller region = 3	−8.664	2.725
Seller region = Buyer region	−3.482	3.142
Indicator for no interactions b/n buyer & seller	228.1	33.09
Indicator high type auction	190.3	2.521
Indicator medium type auction	109.0	2.560
Constant	−80.67	33.92
Adjusted R-squared		0.1511
No. of Observations		44274

Table 7: Explaining Variations in Seller. Dependent variable: $E[cost_{ji}]$. Robust standard errors shown.

Number of types	Type	Average	25th perc.	50th perc.	75th perc.	Pop. prob. (%)
	Low type	206.53	107.70	195.55	323.13	17.07
Three types	Medium type	286.55	236.13	322.49	365.28	46.47
	High type	362.06	362.93	365.14	365.54	36.45
Two types	Low type	263.20	193.32	280.63	365.03	62.42
	High type	361.56	362.87	365.15	365.55	37.57
One type		301.04	262.53	363.49	365.59	100

Table 8: Summary statistics of cost distributions, with three, two, and one unobserved types.

Percentile	Low Type		Medium Type		High Type	
	Model C1	Model C2	Model C1	Model C2	Model C1	Model C2
10%	30.77	32.46	144.33	143.73	354.74	363.70
20%	80.33	82.77	195.46	195.45	361.69	364.92
30%	112.46	118.69	255.40	256.54	363.80	365.35
40%	153.12	153.58	280.48	280.53	364.71	365.51
50%	195.55	195.45	322.49	322.76	365.14	365.58
60%	246.61	254.95	353.36	360.50	365.37	365.61
70%	280.65	285.76	364.71	364.88	365.50	365.63
80%	361.46	362.38	365.49	365.53	365.57	365.64
90%	365.62	365.55	365.62	365.63	365.62	365.65

Table 9: Cost distributions without and with uncertainty in number of bids.

References

V. Aguirregabiria and P. Mira. Sequential Estimation of Dynamic Discrete Games. *Econometrica*, 75(1):1–53, 2007.

G. M. Allenby and P. E. Rossi. Marketing Models of Consumer Heterogeneity. *Journal of Econometrics*, 89(1–2):57–78, 1998.

E. S. Allman, C. Matias, and J. A. Rhodes. Identifiability of Parameters in Latent Structure Models with Many Observed Variables. *The Annals of Statistics*, 37(6A):3099–3132, 2009.

A. Ansari, R. Montoya, and O. Netzer. Dynamic Learning in Behavioral Games: A Hidden Markov Mixture of Experts Approach. *Quantitative Marketing and Economics*, 10(4):475–503, 2012.

P. Arcidiacono and R. A. Miller. Conditional Choice Probability Estimation of Dynamic Discrete Choice Models With Unobserved Heterogeneity. *Econometrica*, 79(6), 2011.

J. Asker and E. Cantillon. Properties of Scoring Auctions. *The RAND Journal of Economics*, 39(1):69–85, 2008.

J. Asker and E. Cantillon. Procurement when Price and Quality Matter. *The RAND Journal of Economics*, 41(1):1–34, 2010.

S. Athey. Single Crossing Properties and the Existence of Pure Strategy Equilibria in Games of Incomplete Information. *Econometrica*, 69(4):861–889, 2001.

S. Athey and P. Haile. Nonparametric Approaches to Auctions. *Handbook of Econometrics*, 6:3847–3965, 2007.

S. Athey and J. Levin. Information and Competition in US Forest Service Timber Auctions. *Journal of Political Economy*, 109(2):375–417, 2001.

S. Athey, J. Levin, and E. Seira. Comparing open and Sealed Bid Auctions: Evidence from Timber Auctions. *The Quarterly Journal of Economics*, 126(1):207–257, 2011.

M. Backus and G. Lewis. A Demand System for a Dynamic Auction Market with Directed Search. *Working Paper*, 2012.

P. Bajari and A. Hortacsu. Economic Insights from Internet Auctions. *Journal of Economic Literature*, 42(2):457–486, 2004.

P. Bajari and L. Ye. Deciding Between Competition and Collusion. *The Review of Economics and Statistics*, 85(4): 971–989, 2003.

P. Bajari, C. Benkard, and J. Levin. Estimating Dynamic Models of Imperfect Competition. *Econometrica*, 75(5): 1331–1370, 2007.

P. Bajari, H. Hong, J. Krainer, and D. Nekipelov. Estimating Static Models of Strategic Interactions. *Journal of Business & Economic Statistics*, 28(4):469–482, 2010.

T. Benaglia, D. Chauveau, and D. Hunter. An EM-like Algorithm for Semi and Nonparametric Estimation in Multivariate Mixtures. *Journal of Computational and Graphical Statistics*, 18(2):505–526, 2009a.

T. Benaglia, D. Chauveau, D. R. Hunter, and D. S. Young. mixtools: An R Package for Analyzing Mixture Models. *Journal of Statistical Software*, 32(6), 2009b.

S. Berry, J. Levinsohn, and A. Pakes. Automobile Prices in Market Equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890, 1995.

F. Branco. The Design of Multidimensional Auctions. *The RAND Journal of Economics*, pages 63–81, 1997.

O. Bruttel. Contracting-out the Public Employment Service and the Consequence for Hard-to-place Jobseekers: Experiences from Australia, the Netherlands and the UK. Technical report, 2004.

M. Cabizza and G. D. Fraja. Quality Considerations in Auctions for Television Franchises. *Information Economics and Policy*, 10(1):9–22, 1998.

S. Campo, I. Perrigne, and Q. Vuong. Asymmetry in First-price Auctions with Affiliated Private Values. *Journal of Applied Econometrics*, 18(2):179–207, 2003.

Y. Che. Design Competition through Multidimensional Auctions. *Rand Journal of Economics*, 24:668–680, 1993.

P. K. Chintagunta, D. C. Jain, and N. J. Vilcassim. Investigating Heterogeneity in Brand Preferences in Logit Models for Panel Data. *Journal of Marketing Research*, pages 417–428, 1991.

M. W. Cripps, G. J. Mailath, and L. Samuelson. Imperfect Monitoring and Impermanent Reputations. *Econometrica*, 72(2):407–432, 03 2004.

S. Datta and K. Sudhir. The Agglomeration-Differentiation Trade-off in Spatial Location Choice. Working Paper, 2011.

C. Dellarocas. The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms. *Management Science*, 49(10):1407–1424, October 2003.

A. P. Dempster, M. N. Laird, and D. B. Rubin. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 39:1–22, 1977.

M. Draganska, M. Mazzeo, and K. Seim. Beyond Plain Vanilla: Modeling Joint Product Assortment and Pricing

Decisions. *Quantitative Marketing and Economics*, 7:105–146, 2009.

R. P. W. Duin. On the Choice of Smoothing Parameters for Parzen Estimators of Probability Density Functions. *IEEE Trans. Comput.*, 25:1175–1179, November 1976. ISSN 0018-9340.

P. Ellickson and S. Misra. Enriching Interactions: Incorporating Outcome Data into Static Discrete Games. *Quantitative Marketing and Economics*, 10:1–26, 2012.

P. Ellickson, S. Misra, and H. Nair. Repositioning Dynamics and Pricing Strategy. *Journal of Marketing Research*, 2012.

P. B. Ellickson and S. Misra. Supermarket Pricing Strategies. *Marketing Science*, 27(5):811–828, 2008.

P. B. Ellickson and S. Misra. Structural Workshop Paper – Estimating Discrete Games. *Marketing Science*, 30(6): 997–1010, 2011.

P. B. Ellickson, S. Houghton, and C. Timmins. Estimating Network Economies in Retail Chains: A Revealed Preference Approach. *The RAND Journal of Economics*, 44(2):169–193, 2013. ISSN 1756-2171.

R. T. Elmore, T. P. Hettmansperger, and H. Thomas. Estimating Component Cumulative Distribution Functions in Finite Mixture Models. *Communications in Statistics – Theory and Methods*, 33(9):2075–2086, 2004.

E. J. Friedman and P. Resnick. The Social Cost of Cheap Pseudonyms. *Journal of Economics & Management Strategy*, 10(2):173–199, 06 2001.

Google. Google Inc. 2013 Financial Tables, September 2013. URL `http://investor.google.com/financial/tables.html`.

A. Gray and A. Moore. Rapid Evaluation of Multiple Density Models. In *Artificial Intelligence and Statistics*, 2003.

E. Guerre, I. Perrigne, and Q. Vuong. Optimal Nonparametric Estimation of First-price Auctions. *Econometrica*, 68 (3):525–574, 2000.

E. Guerre, I. Perrigne, and Q. Vuong. Nonparametric Identification of Risk Aversion in First-Price Auctions Under Exclusion Restrictions. *Econometrica*, 77(4):1193–1227, 2009.

P. Haile, H. Hong, and M. Shum. Nonparametric Tests for Common Values at First-price Sealed-bid Auctions. Technical report, National Bureau of Economic Research, 2003.

P. Hall and X. Zhou. Nonparametric Estimation of Component Distributions in a Multivariate Mixture. *Annals of Statistics*, pages 201–224, 2003.

W. Hartmann. Demand Estimation with Social Interactions and the Implications for Targeted Marketing. *Marketing Science*, 29(4):585–601, 2010.

E. Haruvy and S. D. Jap. Differentiated Bidders and Bidding Behavior in Procurement Auctions. *Journal of Marketing Research*, 50(2):241–258, 2013.

T. Hettmansperger and H. Thomas. Almost Nonparametric Inference for Repeated Measures in Mixture Models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4):811–825, 2000.

B. Holmstrom. Managerial Incentive Problems: A Dynamic Perspective. *Review of Economic Studies*, 66(1):169–82, January 1999.

H. Hong and M. Shum. Increasing competition and the winner's curse: Evidence from procurement. *The Review of Economic Studies*, 69(4):871–898, 2002.

J. V. Hotz and R. A. Miller. Conditional Choice Probabilities and the Estimation of Dynamic Models. *Review of Economic Studies*, 60(3):497–529, July 1993.

A. Ihler. Kernel Density Estimation Toolbox for MATLAB, 2003. URL `http://www.ics.uci.edu/~ihler/code/kde.html`.

M. C. W. Janssen. Auctions and Beauty Contests: A Policy Perspective. Working Paper, Erasmus University of Rotterdam, 2002.

S. D. Jap and E. Haruvy. Interorganizational relationships and bidding behavior in industrial online reverse auctions. *Journal of Marketing Research*, 45(5):550–561, 2008.

M. Jofre-Bonet and M. Pesendorfer. Estimation of a Dynamic Auction Game. *Econometrica*, 71(5):1443–1489, 2003.

W. A. Kamakura and G. J. Russell. A Probabilistic Choice Model for Market Segmentation and Elasticity Structure. *Journal of Marketing Research*, pages 379–390, 1989.

H. Kasahara and K. Shimotsu. Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices. *Econometrica*, 77(1):135–175, 01 2009.

P. Klemperer. Spectrum on the Block. Technical report, October 5 2000.

P. Klemperer. What Really Matters in Auction Design. *Journal of Economic Perspectives*, 16(1):169–189, September 2002.

C. Kozierok. Online Freelancing Guide, 2011. URL `http://www.freelancingguide.net/index.htm`.

E. Krasnokutskaya. Identification and Estimation of Auction Models with Unobserved Heterogeneity. *The Review of*

Economic Studies, 78(1):293–327, 2011.

E. Krasnokutskaya, K. Song, and X. Tang. The Role of Quality in Service Markets Organized as Multi-Attribute Auctions. Working Paper, 2013.

D. M. Kreps and R. Wilson. Reputation and Imperfect Information. *Journal of Economic Theory*, 27(2):253–279, August 1982.

M. Lacity and M. Rottman. *Offshore Outsourcing of IT Work: Client and Supplier Perspectives*. Technology, Work, and Globalization. Palgrave Macmillan, 2008.

M. Levine, D. Hunter, and D. Chauveau. Maximum Smoothed Likelihood for Multivariate Mixtures. *Biometrika*, 98 (2):403–416, 2011.

T. Li, I. Perrigne, and Q. Vuong. Structural Estimation of the Affiliated Private Value Auction Model. *RAND Journal of Economics*, pages 171–193, 2002.

T. Li, I. Perrigne, and Q. Vuong. Semiparametric Estimation of the Optimal Reserve price in First-price Auctions. *Journal of Business & Economic Statistics*, 21(1):53–64, 2003.

J. C. Liechty, M. Wedel, and R. Pieters. The Representation of Local and Global Scanpaths in Eye Movements Through Bayesian Hidden Markov Models. *Psychometrika*, 68(4):519–541, 2003.

N. G. Mankiw and P. Swagel. The Politics and Economics of Offshore Outsourcing. Working Paper 12398, National Bureau of Economic Research, July 2006.

D. McAdams. Isotone Equilibrium in Games of Incomplete Information. *Econometrica*, 71(4):1191–1214, 2003.

G. McLachlan and D. Peel. *Finite Mixture Models*. Wiley-Interscience, 2004.

S. Misra and H. Nair. A Structural Model of Sales-force Compensation Dynamics: Estimation and field implementation. *Quantitative Marketing and Economics*, 9(3):211–257, September 2011.

S. Moon, W. A. Kamakura, and J. Ledolter. Estimating Promotion Response When Competitive Promotions are Unobservable. *Journal of Marketing Research*, pages 503–515, 2007.

M. Morgan. The Freelance Marketplace Annual Review, 2010. WhichLance.com, 2011.

S. Narayanan. Bayesian Estimation of Discrete Games of Complete Information. *Quantitative Marketing and Economics*, 11:39–81, 2013.

O. Netzer, J. M. Lattin, and V. Srinivasan. A hidden markov model of customer relationship dynamics. *Marketing Science*, 27(2):185–204, 2008.

A. Nevo. A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand. *Journal of Economics & Management Strategy*, 9(4):513–548, 2000.

Y. A. Orhun. Spatial Differentiation in the Supermarket Industry: The Role of Common Information. *Quantitative Marketing and Economics*, 11:3–37, 2013.

A. Pakes, J. Porter, K. Ho, and J. Ishii. Moment inequalities and their application. Working Paper, 2011.

M. Pesendorfer and P. Schmidt-Dengler. Asymptotic Least Squares Estimators for Dynamic Games. *The Review of Economic Studies*, 75(3):901–928, 2008.

P. Resnick and R. Zeckhauser. Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System. In M. R. Baye, editor, *The Economics of the Internet and E-Commerce*, volume 11 of *Advances in Applied Microeconomics*. Elsevier Science, 2002.

SEC. Securities and Exchange Commission Filing by Facebook Inc., July 2012. URL http://pdf.secdatabase.com/700/0001193125-12-316895.pdf.

S. Shriver. Network Effects in Alternative Fuel Adoption: Empirical Analysis of the Market for Ethanol. Working paper, 2011.

B. W. Silverman. *Density Estimation for Statistics and Sata Analysis*. Monographs on Statistics and Applied Probability. Chapman and Hall, 1986. ISBN 9780412246203.

V. Singh and T. Zhu. Pricing and Market Concentration in Oligopoly Markets. *Marketing Science*, 27(6):1020–1035, 2008.

S. Yao and C. Mela. Online Auction Demand. *Marketing Science*, 27(5):861–885, 2008.

H. Yoganarasimhan. The Value of Reputation in an Online Freelance Marketplace. *Marketing Science*, 27(5):861–885, 2013.

R. Zeithammer. Forward–Looking Bidding in Online Auctions. *Journal of Marketing Research*, 43(3):462–476, 2006.

R. Zeithammer. Optimal Selling in Dynamic Auctions: Adaptation Versus Commitment. *Marketing Science*, 26(6): 859–867, 2007.

R. Zeithammer and A. Adams. The Sealed-Bid Abstraction in Online Auctions. *Marketing Science*, 29(6):964–987, 2010.

A Web Appendix for "Estimation of Beauty Contest Auctions"

A.1 Nonparametric Identification of Component Mixtures of Bid Distributions and Mixing Probabilities with Unobserved Auction Heterogeneity

We assume that bids arrive independently in this sealed bid setting. Let Q be the number of bids that an auction receives and K the total number of unobserved buyer/auction specific state variables. We present the proof for the case where price is the only bid attribute b_{ji}. Expanding to include more bid-specific variables is straightforward.

Let $\zeta \in \mathbb{R}$ be a cut-off in the positive Real line, and define:

$$\mu_{ji}(A_i,\zeta) = \begin{cases} 1, & \text{if } b_{ji} \leq \zeta \\ 0, & \text{otherwise} \end{cases} \tag{31}$$

Then, $\mu_i(A_i,\zeta) = \sum_{j=1}^{Q} \mu_{ji}(A_i,\zeta)$ is the number of times the ith auction, with observed auction attributes A_i, received bids which are less than or equal to ζ. Based on this expansion, at each ζ, $\mu_i(A_i,\zeta), \mu_2(A_i,\zeta), ..., \mu_n(A_i,\zeta)$ are i.i.d. Then, we can write out the mixture model as:

$$\mathcal{L}(m,Q) = \sum_{k=1}^{\mathcal{K}} \pi_k \mathcal{B}(m;Q,\mathcal{T}_k(\zeta)) \tag{32}$$

where $\mathcal{T}_k(\zeta) = \int_0^{\zeta} \mathcal{G}_k(b|A_i,v^k) \mathrm{d}b$ is the CDF that a bid-draw b will be less than or equal to ζ, if the unobserved auction type is k, and $\mathcal{B}(m;Q,\mathcal{T}_k(\zeta)) = \binom{Q}{m}[\mathcal{T}_k(\zeta)]^m [1-\mathcal{T}_k(\zeta)]^{Q-m}$ is the binomial probability function.

At any given ζ, the number of such equations is Q. The number of parameters to be identified consist of $K-1$ mixing probabilities and K CDFs, $\mathcal{T}_k(\zeta)$s. Thus, at each point in the non-negative real line, ζ, this system of equations is identified if:

$$Q \geq 2K-1 \tag{33}$$

Thus, the CDFs of the component distributions and the mixing probabilities can be retrieved at each point in the non-negative real line if the number of bids received is greater than or equal to $2K-1$. Note that this is a strong identification condition since it requires that the mixing probabilities be identified at each ζ, which is, of course, not necessary since identifying π_ks just once is sufficient for identification.

A.2 Estimation Steps without Auction Specific Unobservables

A.2.1 First Step Estimation

In the first step, we estimate the conditional choice probabilities (CCP), their derivatives, and the equilibrium distributions of bids and seller attributes. We describe each in detail now:

- Conditional choice probability – $\mathcal{P}(X_{ji},b_{ji},X_{-ji},b_{-ji}|A_i)$
  This is the probability that seller j will win auction i given state variables $\{A_i, X_{ji}, b_{ji}, X_{-ji}, b_{-ji}\}$. Given the assumption of no common knowledge unobservables, these probabilities are directly available from data following Hotz and Miller (1993). They are identified and can be estimated without making functional form assumptions.

- Derivative of the CCPs – $\left(\frac{\partial \mathcal{P}(X_{ji},b_{ji},X_{-ji},b_{-ji}|A_i)}{\partial b_{ji}} \right)^{-1}$
  This is readily available from $\mathcal{P}(\cdot)$ by computing the numerical derivative.

- Joint distributions of equilibrium bids and seller attributes – $\mathcal{G}(X_{ji},b_{ji}|A_i)$
  These distributions represent sellers' expectations on the competitors' bids and attributes in equilibrium, and with sufficient data, can be obtained from the data directly.

All of the above distributions can estimated using sieve estimators or kernel densities when the researcher has large samples and a relatively small state space. However, with large state spaces, finite samples may not be amenable to purely nonparametric methods. In such cases, the researcher may employ semi-parametric or parametric methods, with the goal of maximizing the fit or the predictive ability of these first stage models.

A.2.2 Second Step Estimation

Given the first step results, we obtain the numerical estimates of $\widehat{\mathcal{S}}(\cdot)$ and $\left(\frac{\partial \widehat{\mathcal{S}}(\cdot)}{\partial b_{ji}}\right)^{-1}$ for each seller j, in each auction i in the data, as follows:

- Step 1: Make $(q_i - 1)$ draws of equilibrium seller attributes and bids from $\widehat{\mathcal{G}}(X_{ji}, b_{ji}|A_i)$. Denote these draws as: $\tilde{X}_{-ji} = \{\tilde{X}_{1i}, \ldots, \tilde{X}_{(j-1)i}, \tilde{X}_{(j+1)i}, \ldots, \tilde{X}_{q_i i}\}$ and $\tilde{b}_{-ji} = \{\tilde{b}_{1i}, \ldots, \tilde{b}_{(j-1)i}, \tilde{b}_{(j+1)i}, \ldots, \tilde{b}_{q_i i}\}$. Together with j's own attributes and bid, this constitutes one simulation of auction i.
- Step 2: Using the $q_i - 1$ simulated draws from Step 1 and j's own attributes and bid, obtain the probability of being chosen and its derivative, $\tilde{\mathcal{P}}\left(X_{ji}, b_{ji}, \tilde{X}_{-ji}, \tilde{b}_{-ji}|A_i\right)$ and $\frac{\partial \tilde{\mathcal{P}}\left(X_{ji}, b_{ji}, \tilde{X}_{-ji}, \tilde{b}_{-ji}|A_i\right)}{\partial b_{ji}}$.
- Step 3: Repeat Steps 1 and 2 L times and take the averages to obtain:

$$\widehat{\mathcal{S}}(X_{ji}, b_{ji}|A_i) = \frac{\sum_{l=1}^{L} \tilde{\mathcal{P}}\left(X_{ji}, b_{ji}, \tilde{X}_{-ji}^l, \tilde{b}_{-ji}^l|A_i\right)}{L} \tag{34}$$

$$\frac{\partial \widehat{\mathcal{S}}(X_{ji}, b_{ji}|A_i)}{\partial b_{ji}} = \frac{\sum_{l=1}^{L} \frac{\partial \tilde{\mathcal{P}}\left(X_{ji}, b_{ji}, \tilde{X}_{-ji}^l, \tilde{b}_{-ji}^l|A_i\right)}{\partial b_{ji}}}{L} \tag{35}$$

where $\{\tilde{X}_{-ji}^l, \tilde{b}_{-ji}^l\}$ is the lth set drawn (from Step 2). While one set of draws is sufficient for consistency, we set $L = 1000$ in our estimation to improve the efficiency of the results.

Finally, using Equation (12), we infer the costs of each seller j in each auction i, which are then used to nonparametrically estimate the distribution of costs for different levels of observed auction characteristics.

A.3 Estimation with Assumption 6

A.3.1 Modified EM-like Algorithm in the First Step

Consider the following nonparametric EM-like algorithm to estimate the bid prices. At this stage, $\mathcal{G}^X(X_{ji}|A_i)$ is assumed to have been already estimated. Since it is not dependent on the unobserved type, its estimation is straightforward. Now, let $AX_{ji} = \{A_i, X_{ji}\}$ take on $\bar{H}$ possible levels, $AX_{ji} \in \{AX^1, AX^2, \ldots, AX^{\bar{H}}\}$. Then all the bids in the data can be partitioned into $\bar{H}$ groups based on observed auction and seller attributes AX_{ji}. Note that because seller attributes (X_{ji}s) can vary across bids within the same auction, bids from the same auction can belong to different partitions. This is one of the key ways in which this algorithm differs from the standard nonparametric mixture algorithms. Then, the three-step iterative algorithm is as follows:

KDE-Step – Let $\mathcal{G}_{h,k}^t(b|AX^h, v^k)$ denote the probability density function of observing bid price b at observed state AX^h and unobserved type v^k in iteration t. We now define:

$$\mathcal{G}_{h,k}^t(b|AX^h, v^k) = \frac{1}{\mu_h^t\left[\sum_{m=1}^{n_h} \lambda_{mhk}^{t-1}\right]} \sum_{m=1}^{n_h} \lambda_{mhk}^{t-1} \mathcal{K}\left(\frac{b - b_m}{\mu_h^t}\right) \tag{36}$$

where μ_h^t is the bandwidth for group h in iteration t, $\mathcal{K}(\cdot)$ is a univariate kernel such that $\int_R \mathcal{K}(b) d(b) = 1$.

E-Step – In the E-step, we update the posterior probabilities λ_{ik}^ts, for each auction, for this iteration, as follows:

$$\lambda_{ik}^t = \frac{\pi_k^{t-1} \prod_{j=1}^{q_i} \mathcal{G}_{h,k}^t\left(b_{ji}|AX_{ji} = AX^h, v^k\right)}{\sum_{k=1}^{K} \pi_k^{t-1}\left[\prod_{j=1}^{q_i} \mathcal{G}_{h,k}^t\left(b_{ji}|AX_{ji} = AX^h, v^k\right)\right]} \forall k \tag{37}$$

where π_k^{t-1}s is the population probability of unobserved type k from the previous iteration. An important point of note is that even when updating the posterior probabilities for a single auction, we might have to use many different bid distributions because the group that a bid belongs to depends on both A_i and X_{ji}.

M-Step – The Maximization step is the same as before.

We iteratively perform the above steps till convergence, at which point, we have consistent estimates of the population probabilities of unobserved types, posterior probability of an auction belonging to a given unobserved type, and $\bar{H} \times K$ probability density functions of bid prices $\mathcal{G}_{h,k}(b|AX^h, v^k)$.

A.3.2 Modified Second Step Estimation

The second step estimation is similar to that outlined in §6.2, with the following changes to Step i:

- We first make $(q_i - 1)$ draws of equilibrium seller attributes from $\widehat{\mathcal{G}}^X(X_{ji}|A_i)$. Denote these draws as: $\tilde{X}_{-ji} = \{\tilde{X}_{1i}, ..., \tilde{X}_{(j-1)i}, \tilde{X}_{(j+1)i}, ..., \tilde{X}_{q_i i}\}$.
- Then, for each draw of X_{ji}, based on $AX_{ji} = \{A_i, X_{ji}\}$, make a draw of bid price from $\widehat{\mathcal{G}}_{h,k}(b_{ji}|AX_{ji} = AX^h, v^k)$. Denote these as $\tilde{b}_{-ji} = \{\tilde{b}_{1i}, ..., \tilde{b}_{(j-1)i}, \tilde{b}_{(j+1)i}, ..., \tilde{b}_{q_i i}\}$. Together with j's own attributes and bid, this constitutes one simulation of auction i for $v_i = v^k$.

Continue with the estimation as before.

A.4 Details of the First Step Estimation for the Freelancing Context

A.4.1 Non-Parametric Joint Distributions of Number of Ratings and Mean Rating

We model the joint distributions of these two attributes using bivariate kernel density functions. First, we classify all the bids according to the number of bids received in the corresponding auction (Table 10) and then sub-classify the bids based on the buyer's reputation (Table 11). These classifications allow us to capture the differences in sellers' expectations about the attributes of her competitors. For example, buyers with a large number of high ratings (sub-class 4) receive better bids (*e.g.*, low bid price, high bidder reputation), on average, compared to those with no ratings (sub-class 1). Ideally, of course, we would prefer a more fine-grained classification. However, further classification is not feasible given the size of our data.

Classes	Number of bids
Class 1	$0 <$ Number of bids ≤ 5
Class 2	$5 <$ Number of bids ≤ 10
Class 3	$10 <$ Number of bids ≤ 20
Class 4	$20 <$ Number of bids

Sub-classes	Number of ratings	Avg. rating
Sub-class 1	Number of ratings $= 0$	–
Sub-class 2	$0 <$ Number of ratings	Avg. rating ≤ 9.3
Sub-class 3	$0 <$ Number of ratings ≤ 10	Avg. rating > 9.3
Sub-class 4	Number of ratings > 10	Avg. rating > 9.3

Table 10: The four classes of auctions.   Table 11: The four sub-classes of buyers.

Let $\{1, ..., T\}$ be the set of categories, where $T = 16$ since we have four classes for number of bids, each with four sub-classes. $t = 1$ denotes class 1 and sub-class 1, $t = 2$ denotes class 1 and sub-class 2, and so on. The observed bids in each category are indexed by $m \in \{1, 2, .., M_t\}$, where M_t is the total number of bids in category t. The mth bid in any category is denoted by vector Y_m, where the two elements of Y_m are the number of ratings and mean rating of the mth bid. We model the probability density function at a point Y in the two dimensional space, in category t, using the multivariate kernel density estimator:

$$\psi_t(Y, \mu_t) = \frac{1}{M_t \mu_t^2 r(l_t, Y)^2} \sum_{m=1}^{M_t} \mathcal{K}\left\{ \frac{1}{\mu_t r(l_t, Y)} (Y - Y_m) \right\} \tag{38}$$

where μ_t is the optimal bandwidth window for category t, $\mathcal{K}(\cdot)$ is the two dimensional kernel function satisfying the property $\int_{R^2} \mathcal{K}(Y) d(Y) = 1$. $r(l_t, Y)$ is a scaling parameter that represents the Euclidean distance from Y to the l_t^{th} nearest point in the data.[16]

The choice of the bandwidth is crucial to the quality of the kernel estimator. We estimate the optimal bandwidths, $\mu_t \,\forall\, t$, using likelihood cross-validation (Duin, 1976; Silverman, 1986). Let $\hat{\psi}_t(\mu, Y)$ and $\hat{\psi}_{t,-m}(\mu, Y)$ be the PDF estimate of point Y from the tth category using bandwidth μ and datasets $\{Y_1, ..., Y_{M_t}\}$ and $\{Y_1, ..., Y_{m-1}, Y_{m+1}, ..., Y_{M_t}\}$, respectively. Then the cross-validation score of μ for category t is given by:

$$CV_t(\mu) = M_t^{-1} \sum_{m=1}^{M_t} \ln[\hat{\psi}_{t,-m}(\mu, Y)] \tag{39}$$

---

[16]In the absence of $r(l_t, Y)$, the same bandwidth is used for all parts of the distribution. This is problematic in finite samples because it is difficult to pick one optimal bandwidth for the entire range; low bandwidths lead to spurious noise in the tails of the distribution, while high bandwidths cause over-smoothing in the main parts of the distribution (Silverman, 1986). Scaling the bandwidth locally using $r(l_t, Y)$ provides a simple but effective solution to this problem. Further, as is common in the literature, we set $l_t = \sqrt{M_t}$.

The likelihood cross-validation choice of the optimal bandwidth is the value that maximizes $CV_t(\mu)$. Intuitively, the cross-validation score $CV_t(\mu)$ is the log-likelihood of observing the dataset. $\hat{\psi}_{t,-m}(\mu,Y)$ is the probability of drawing the data-point A (assuming it is not part of the dataset). So $M_t^{-1}\Pi_{m=1}^{M_t}\hat{\psi}_{t,-m}(\mu,Y)$ is the total probability of observing the dataset.

While the maximization is conceptually simple, it is computationally intensive. To evaluate the likelihood at a given bandwidth, we need to evaluate the density at each data point at that bandwidth and then sum over the density contributions of all data points. Moreover, at each data point, we need to find the l_t^{th} nearest point in the Euclidean space to calculate its density contribution. This becomes prohibitively expensive as the size of the dataset and the number of dimensions increase. Moreover, this has to be done multiple times to reach the optimal μ_t. To address these computational issues, we follow the recent method proposed by Gray and Moore (2003), which is based on *k-d trees* and has been shown to be much faster than previous methods. We use the MATLAB-based KDE toolbox to perform the estimation (Ihler, 2003).

A.4.2 Multinomial Logit Model of Bidder's Geographic Region

Sellers can belong to one of four discrete geographic regions (Table 1). Conditional on buyer and auction specific state variables, and a given draw of bidder mean rating and number of ratings, we model the distribution of seller's geographic region using a Multinomial Logit model. Let $\mathcal{Z}_g(region_{ji}|X_{ji}^g,\theta_g)$ be the conditional probability of seller j's region, where X_{ji}^g is the set of variables that influences the draw of the bidders' geographic region and $\theta_g=\{\theta_{g1},\theta_{g2},\theta_{g3}\}$ are the parameter vectors associated with regions 1, 2, and 3, respectively. Then, the probability that bidder j in auction i belongs to geographic region k is:

$$
\begin{aligned}
\mathcal{Z}_g(region_{ji}=\gamma|X_{ji}^g,\theta_g) &= \frac{\exp(X_{ji}^g\theta_{gk})}{1+\exp(X_{ji}^g\theta_{g1})+\exp(X_{ji}^g\theta_{g2})+\exp(X_{ji}^g\theta_{g3})}\forall\gamma\in\{1,2,3\}\\
\mathcal{Z}_g(region_{ji}=4|X_{ji}^g,\theta_g) &= \frac{1}{1+\exp(X_{ji}^g\theta_{g1})+\exp(X_{ji}^g\theta_{g2})+\exp(X_{ji}^g\theta_{g3})}
\end{aligned}
\tag{40}
$$

$\theta_{g4}=0$ because $\gamma=4$ is the base region. The set of parameters to be estimated in this context is $\theta_g=\{\theta_{g1},\theta_{g2},\theta_{g3}\}$. The log-likelihood of drawing the sellers' geographic regions observed in the data is:

$$
L_g(\theta_g)=\sum_{i=1}^{n}\sum_{j=1}^{q_i}\sum_{\gamma=1}^{4}\ln[\mathcal{Z}_g(region_{ji}=\gamma|X_{ji}^g,\theta_g)^{I(region_{ji}=\gamma)}]
\tag{41}
$$

Maximizing the above log-likelihood gives us a consistent estimate of θ_g.

A.4.3 Logit Model of Buyer-Bidder Past Interaction Indicator

The indicator for whether a buyer and seller have interacted in the past is modeled using a binary Logit model. Let $\mathcal{Z}_l(int_{ji}|X_{ji}^l,\theta_l)$ be the conditional probability of seller j's having positive past interactions with buyer i, given state variables X_{ji}^l and parameter vector θ_l. Then, the probability of drawing int_{ji} is:

$$
\mathcal{Z}_l(int_{ji}=1|X_{ji}^l,\theta_l)=\frac{e^{X_{ji}\theta_l}}{1+e^{X_{ji}\theta_l}}; \qquad \mathcal{Z}_l(int_{ji}=0|X_{ji}^l,\theta_l)=\frac{1}{1+e^{X_{ji}\theta_l}}
\tag{42}
$$

The parameters of interest in this context are θ_l, which can be estimated using maximum likelihood. The log-likelihood of observing the indicators of past buyer-seller interactions observed in the data is given by:

$$
L_l(\theta_l)=\sum_{i=1}^{n}\sum_{j=1}^{q_i}\sum_{\gamma=0}^{1}\ln[\mathcal{Z}_l(int_{ji}=\gamma|X_{ji}^l,\theta_l)^{I(int_{ji}=\gamma)}]
\tag{43}
$$

A.5 Step-by-Step Procedure for Counterfactual Simulations

Step 1: Solving for sellers' bidding strategy

First, we solve for sellers' bidding strategy given a set of auction attributes and number of bids, $\{A_i,v_i\}$. Since unobserved auction heterogeneity v_i is known during counterfactuals, we essentially have the problem of obtaining optimal bids

given a set of state variables. Recall that the FOC of seller's optimization is:

$$
\begin{aligned}
b_{ji} &= \beta(A_i, X_{ji}, c_{ji}, \mathcal{G}(\cdot), \mathcal{P}(\cdot)) = \beta(O_i, q_i, r_i, v_i, X_{ji}, c_{ji}, \mathcal{G}(\cdot), \mathcal{P}(\cdot)) \\
&= \frac{c_{ji}}{1 - r_i} - \left(\frac{\partial \mathcal{S}_{ji}(X_{ji}, b_{ji} | O_i, r_i, q_i, v_i)}{\partial b_{ji}} \right)^{-1} \mathcal{S}_{ji}(X_{ji}, b_{ji} | O_i, r_i, q_i, v_i)
\end{aligned}
\tag{44}
$$

Throughout this section, we use the expanded nomenclature $A_i = \{O_i, q_i, r_i\}$ to be explicit about the primitives that are modified in counterfactuals. To derive a seller's equilibrium bid under a counterfactual scenario, we need information on both the equilibrium distribution of bids, $\mathcal{G}(\cdot)$, and the buyers' decision, $\mathcal{P}(\cdot)$. With a structural interpretation of buyers' decisions, we can treat the estimates from the Nested logit model from §7.2 as primitives of buyer utilities. Since primitives of utility are unlikely to change under counterfactual scenarios, we can continue using them. However, equilibrium distribution of bids will change, and we need to estimate the new $\mathcal{G}^{new}(\cdot)$.

To obtain the new equilibrium distribution of bids and seller attributes, $\mathcal{G}^{new}(X_{-ji}, b_{-ji} | O_i, q_i, r_i, v_i)$, we start by assuming a distribution, $\mathcal{G}^1(X_{-ji}, b_{-ji} | O_i, q_i, r_i, v_i)$, in the first iteration. Then, for each seller j, in each auction i, we solve for the optimal bid b_{ji}^1 in iteration 1 based on the current estimate of seller attributes and bids, $\mathcal{G}^1(X_{-ji}, b_{-ji} | O_i, q_i, r_i, v_i)$, using Equation (44). Note that solving for the optimal bid is not straightforward because the FOC is an implicit function of b_{ji}. Thus, for each seller j in each auction i, we not only need to use a root-finding algorithm such as Newton-Raphson[17] to obtain the new b_{ji}^1, but we also need to numerically simulate the expected probabilities of winning and its derivatives at each step of the root finding algorithm. Next, with the new estimates of b_{ji}^1s and X_{ji}s, we update our estimate of the distribution of seller attributes and bids to $\mathcal{G}^2(X_{-ji}, b_{-ji} | O_i, q_i, r_i, v_i)$. These in turn are used to generate the new estimates of bids, b_{ji}^2s. This process continues till the joint distributions of equilibrium bids and seller attributes converge.

This process is computationally intensive, because in order to reach the overall fixed-point of the system, we need to calculate the fixed-point of each agent using a root-finding algorithm at each iteration, which in turn requires numerical simulations at each of its iterations.[18] Since we need to derive the equilibrium bids for all auction-seller combinations observed in the data, this can take some time.

Step 2: Solve for buyers' entry decisions

Once we have estimates of optimal bids for each auction-seller combination, we solve for buyers' entry decisions. For each combination of auction attributes, $\{O_i, r_i, v_i\}$, simulate entry probability as follows:

- Step (a): Simulate a draw of number of bids, q_i^{new}, from the estimated Poisson bid arrival model. Next, simulate the seller-bid attributes for the q_i^{new} bids using the bid distribution $\mathcal{G}^{new}(X_{-ji}, b_{-ji} | O_i, q_i^{new}, r_i, v_i)$ estimated in Step 1. Then using the Nested Logit estimates, derive the Inclusive value from entry, $\mathcal{U}_i^{new}$, as specified in §8.6.1. This constitutes one realization of the auction.
- Step (b): Perform Step (a) a large number of times and average to derive the new expected Inclusive value $\mathcal{EU}_i^{new}$ for the given combination of auction attributes.
- Step (c): Derive the new entry probability $P^{new}(enter | O_i, r_i, v_i)$ using Equation (27).

Using these steps, derive the entry probability for each auction type.

Step 3: Combine buyer and seller decisions to obtain new system-level equilibrium

Start with the pre-entry population distribution of auction types $\{O_i, r_i, v_i\}$. Draw a large number of auctions from this distribution. For each draw, simulate: (a) buyer's entry decision using estimates from Step 2, (b) number of bid arrivals using estimates of Poisson bid arrival model (Table 13), (c) seller-bid attributes using the bid distributions estimated in Step 1, and (d) buyer's choice decision using estimates from the Nested logit model (Table 6). Keep track of the outcomes to calculate the new clearance rates and site revenues.

A.6 Additional Tables and Figures

---

[17] Since there exists a unique best-response for a given $G(\cdot)$ and $\mathcal{P}(\cdot)$, any root-finding algorithm will reach the unique optimal bid.

[18] It is possible to ease the computational burden by avoiding root-finding algorithms in the initial iterations. For example, the researcher may simply substitute the previous estimate of the bid in the right hand side of the FOC and obtain the new estimate of the bid. This requires only one set of numerical simulations of probabilities of winning and its derivative per iteration, as opposed to repeated simulations at each step of the root-finding algorithm. We found that employing this method for the first few steps and then switching to the full solution speeds up the convergence considerably without compromising convergence.

Explanatory variables (X_{ji}^l)	Coefficient	Std. error
Buyer region = 1	-4.585×10^{-1}	6.472×10^{-1}
Buyer region = 2	-3.480×10^{-1}	2.889×10^{-1}
Buyer region = 3	-6.335×10^{-1}	7.785×10^{-1}
Number of bids	-8.369×10^{-2}	1.072×10^{-2}
Square of number of bids	6.262×10^{-4}	1.255×10^{-4}
Indicator for auction attachment	-6.495×10^{-1}	2.279×10^{-1}
Buyer's success ratio	3.485	4.861×10^{-1}
ln(Number of past auctions canceled by buyer)	-4.027×10^{-1}	1.983×10^{-1}
ln(Number of past auctions posted by buyer)	1.132	2.105×10^{-1}
ln(Buyer tenure in days + 1)	-2.543×10^{-1}	6.099×10^{-2}
Indicator for zero buyer ratings	-8.531	4.050
ln(No. of buyer ratings + 1)	-1.634×10^{-1}	1.449×10^{-1}
Buyer mean rating (centered)	-8.833×10^{-1}	4.082×10^{-1}
ln(No. of buyer ratings + 1) × Buyer mean rating (centered)	9.916×10^{-1}	3.045×10^{-1}
ln(No. of seller ratings + 1)	5.664×10^{-1}	5.742×10^{-2}
Seller mean rating (centered)	5.581×10^{-1}	1.365×10^{-1}
Square of seller mean rating (centered)	4.161×10^{-2}	1.782×10^{-2}
Seller mean rating (centered) × Buyer mean rating (centered)	3.320×10^{-2}	7.916×10^{-3}
Seller region = 1	-7.604×10^{-1}	2.301×10^{-1}
Seller region = 2	5.399×10^{-1}	2.129×10^{-1}
Seller region = 3	-5.493×10^{-1}	2.649×10^{-1}
Constant	-7.506	6.015×10^{-1}
No. of observations = 44274; Log likelihood = -868.526		

Table 12: Estimates of Logit model of indicator for buyer-seller past interactions.
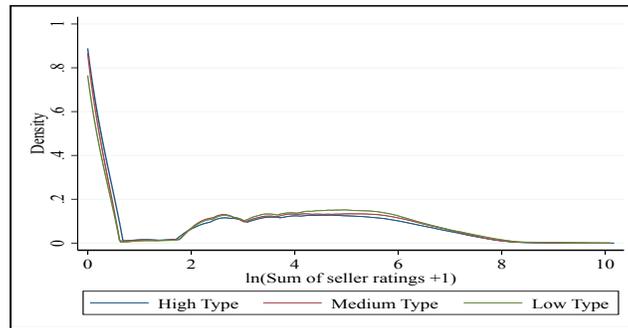
Figure 18: CDF of ln(Sum of seller ratings + 1) for the three auction types – Low, Medium, and High.

Explanatory variables (η_i)	Coefficient	Std. error
ln(Sum of buyer ratings + 1)	-3.418×10^{-2}	2.084×10^{-2}
Buyer region = 1	3.179×10^{-1}	9.178×10^{-2}
Buyer region = 2	1.300×10^{-1}	6.112×10^{-2}
Buyer region = 3	3.458×10^{-1}	1.123×10^{-1}
ln(Buyer tenure in days + 1)	1.769×10^{-2}	9.403×10^{-3}
ln(Number of past auctions canceled by buyer)	-5.849×10^{-2}	2.437×10^{-2}
ln(Number of auctions uncanceled by buyer)	9.663×10^{-2}	3.545×10^{-2}
Indicator high type	-2.099×10^{-1}	7.118×10^{-2}
Indicator medium type	-2.612×10^{-1}	7.890×10^{-2}
Constant	2.430	8.783×10^{-2}
No. of observations = 4002; Log likelihood = -27557.068		

Table 13: Estimates of Poisson bid arrival model, θ_p.

Explanatory variables (X_{ji}^{g})	Region 1		Region 2		Region 3	
	Coefficient	Std. error	Coefficient	Std. error	Coefficient	Std. error
Buyer region = 1	1.488×10^{-1}	7.576×10^{-2}	-1.281×10^{-1}	1.020×10^{-1}	-2.631×10^{-1}	1.010×10^{-1}
Buyer region = 2	-4.009×10^{-2}	5.262×10^{-2}	4.392×10^{-1}	6.805×10^{-2}	3.647×10^{-2}	6.639×10^{-2}
Buyer region = 3	-1.523×10^{-1}	9.585×10^{-2}	4.888×10^{-1}	1.372×10^{-1}	1.370×10^{-2}	1.194×10^{-1}
Number of bids	1.186×10^{-2}	1.880×10^{-3}	-8.098×10^{-3}	2.188×10^{-3}	-1.124×10^{-3}	2.294×10^{-3}
Square of number of bids	-9.520×10^{-5}	2.050×10^{-5}	7.450×10^{-5}	2.330×10^{-5}	2.950×10^{-5}	2.430×10^{-5}
Indicator for auction attachment	-9.026×10^{-2}	3.786×10^{-2}	-2.077×10^{-1}	4.597×10^{-2}	-6.432×10^{-2}	4.760×10^{-2}
ln(Deadline_days+1)	-1.181×10^{-2}	1.014×10^{-2}	2.320×10^{-3}	1.213×10^{-2}	2.815×10^{-2}	1.282×10^{-2}
Buyer's success ratio	-3.046×10^{-1}	1.118×10^{-1}	-4.237×10^{-1}	1.314×10^{-1}	-3.408×10^{-1}	1.420×10^{-1}
ln(Total number of past auctions of buyer)	1.544×10^{-1}	7.487×10^{-2}	2.841×10^{-1}	8.655×10^{-2}	7.499×10^{-2}	9.384×10^{-2}
ln(Number of past auctions canceled by buyer)	-1.037×10^{-1}	7.206×10^{-2}	-4.069×10^{-1}	8.300×10^{-2}	-1.377×10^{-1}	9.023×10^{-2}
ln(Sum of buyer ratings + 1)	5.931×10^{-1}	2.026×10^{-1}	1.058×10^{-1}	2.420×10^{-1}	3.968×10^{-1}	2.560×10^{-1}
ln(No. of buyer ratings + 1)	-7.187×10^{-1}	2.329×10^{-1}	-6.453×10^{-2}	2.784×10^{-1}	-4.384×10^{-1}	2.948×10^{-1}
Buyer mean rating (centered)	2.024×10^{-1}	1.169×10^{-1}	-2.197×10^{-1}	1.426×10^{-1}	-3.431×10^{-2}	1.451×10^{-1}
Square of buyer mean rating (centered)	3.197×10^{-2}	1.291×10^{-2}	-2.095×10^{-2}	1.557×10^{-2}	2.381×10^{-3}	1.600×10^{-2}
ln(No. of buyer ratings + 1) × Buyer mean rating (centered)	-7.772×10^{-2}	4.582×10^{-2}	7.884×10^{-2}	5.736×10^{-2}	-3.167×10^{-2}	5.693×10^{-2}
Indicator for zero seller ratings	-2.439×10^{-1}	2.197×10^{-1}	-9.696×10^{-2}	2.934×10^{-1}	-6.915×10^{-1}	3.087×10^{-1}
ln(Sum of seller ratings + 1)	-7.459×10^{-1}	1.818×10^{-1}	7.098×10^{-1}	2.165×10^{-1}	3.550×10^{-1}	2.193×10^{-1}
ln(No. of seller ratings + 1)	8.575×10^{-1}	2.036×10^{-1}	-9.223×10^{-1}	2.438×10^{-1}	-2.772×10^{-1}	2.449×10^{-1}
Seller mean rating (centered)	-1.675×10^{-1}	2.578×10^{-2}	1.331×10^{-1}	3.234×10^{-2}	4.111×10^{-1}	3.503×10^{-2}
Square of seller mean rating (centered)	-2.828×10^{-2}	5.560×10^{-3}	3.099×10^{-2}	6.982×10^{-3}	6.383×10^{-2}	7.349×10^{-3}
Buyer mean rating × Seller mean rating	-1.736×10^{-3}	6.589×10^{-4}	-9.303×10^{-4}	7.757×10^{-4}	3.151×10^{-4}	8.450×10^{-4}
Constant	1.433	4.844×10^{-1}	-1.090	5.767×10^{-1}	-1.572	6.016×10^{-1}
No. of observations = 44272;　Log likelihood = -50928.8						

Table 14: Estimates of Multinomial Logit model of seller regions. Region 4 is the base outcome.