This article was downloaded by: [128.95.104.109] On: 21 July 2016, At: 00:33
Publisher: Institute for Operations Research and the Management Sciences (INFORMS)
INFORMS is located in Maryland, USA

Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Estimation of Beauty Contest Auctions

Hema Yoganarasimhan

To cite this article:
Hema Yoganarasimhan (2016) Estimation of Beauty Contest Auctions. Marketing Science 35(1):27-54. http://dx.doi.org/10.1287/mksc.2015.0929

Full terms and conditions of use: http://pubsonline.informs.org/page/terms-and-conditions

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2016, INFORMS

Please scroll down for article—it is on subsequent pages

INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

MARKETING SCIENCE
Vol. 35, No. 1, January–February 2016, pp. 27–54
ISSN 0732-2399 (print) | ISSN 1526-548X (online)

http://dx.doi.org/10.1287/mksc.2015.0929
© 2016 INFORMS

Downloaded from informs.org by [128.95.104.109] on 21 July 2016, at 00:33 . For personal use only, all rights reserved.

Estimation of Beauty Contest Auctions

Hema Yoganarasimhan
University of Washington, Seattle, Washington 98195, hemay@uw.edu

Beauty contests are auction mechanisms used to buy or sell differentiated products where the auctioneer does not specify a decision rule to pick the winning bidder. Beauty contests are widely used in procuring welfare-to-work projects, freelance services, selling online ads, real estate transactions, and hiring, dating/marriage decisions. Unlike price-only auctions, beauty contests have no closed-form bidding strategies and suffer from nonmultiplicatively 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 estimator. 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 Expectation-Maximization (EM)-like algorithm, and that these 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 in this marketplace are around 15% of the bid, and that not accounting for unobserved heterogeneity can significantly bias cost estimates in this setting. Based on our estimates, we run counterfactual simulations and provide guidelines to managers of freelance firms.

Data, as supplemental material, are available at http://dx.doi.org/10.1287/mksc.2015.0929.

Keywords: structural models; auctions; freelance markets; seller marketpower; unobserved heterogeneity; online markets

History: Received: October 31, 2013; accepted: March 9, 2015; Preyas Desai served as the editor-in-chief and Sanjog Misra served as associate editor for this article. Published online in *Articles in Advance* September 9, 2015.

1. Introduction

A beauty contest auction is a procurement mechanism where the auctioneer does not specify an allocation rule or a decision rule to pick the winning bidder. 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 thus 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 in which a seller has a set of attributes and then chooses an optimal bid given her own attributes and beliefs on the buyer's allocation rule. Figure 1 presents a couple of examples of beauty contest auctions.

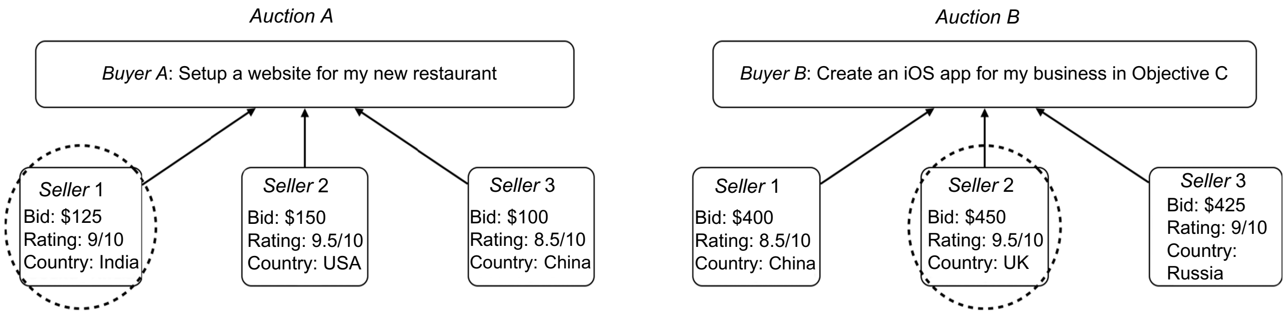
The term "beauty contest" comes from beauty pageants, where the winner is chosen by a committee that does not announce any scoring or selection rule a priori. While beauty is an important attribute, contestants realize that other attributes matter too, e.g., compassion and general knowledge. Auction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

Beauty contests are used in private and public sector procurement. Governments use them to procure welfare-to-work projects (Bruttel 2004), to 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 find freelance programmers (Yoganarasimhan 2013). The popular Google Adwords and Facebook ads auctions, which generate over $4 billion in quarterly revenues

[1] Procurers often prefer beauty contests over scoring auctions for 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. Conversely, because beauty contests lack transparency, they can suffer from agency problems if the procurer is a middleman, 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.

RIGHTSLINK

Figure 1 Two Examples of Beauty Contest Auctions



Note. The circled bid is the one chosen by the buyer.

(Google 2013, SEC 2012), can also be interpreted as beauty contests as they are multiattribute auctions where the allocation rule of the auctioneer is not observable. Apart from these obvious examples, any setting in which an agent invites bids from a set of discrete alternatives and makes an optimal choice without using prespecified allocation rules can be interpreted as a beauty contest, e.g., real estate bidding, hiring or employment transactions, and marriage/dating decisions.

There are three central questions of interest to researchers and managers in this area. These questions are: (1) What is the sellers' equilibrium bidding strategy, and the buyers' equilibrium allocation rule? (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 market and to answering important questions such as: How much market power 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? In this paper, we present a structural framework to estimate beauty contest auctions and answer the above questions.

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 about buyer behavior. Hence, in addition to modeling the seller side, we need to infer buyer behavior from the data. Second, we do not have a closed form solution for a seller's bidding strategy. The multidimensionality of bids and the unobservability of the buyer's 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 be 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, in Figure 1, auction B is a more difficult job on unobservable dimensions (to the researcher) compared to auction A because it requires mobile app development skills. Hence auction B invites higher bids than auction A. A model without unobserved heterogeneity will mistakenly attribute the high bids to high underlying seller costs, when in fact, high bids should be attributed to the unobserved difficulty of the job.

In this paper, we formulate beauty contest auctions as two-stage games of strategic interaction with incomplete information. We present a two-step estimation method that can address these 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).

We show that the joint distributions of seller attributes and bid prices with finite unobserved auction types are nonparametrically identified in sealed bid auctions with independent bidders. We then propose a two-step estimator similar in spirit to Arcidiacono and Miller (2011) to accommodate common knowledge unobservables for continuous state space problems. We use a kernel-smoothed nonparametric Expectation-Maximization (EM)-like algorithm (similar to that in Benaglia et al. 2009a) to recover nonparametric estimates of underlying bid distributions as functions of observable attributes and unobservable auction types, as well as the population distribution of unobserved types in the first step. At this stage, we also estimate the buyers' allocation rule. In the second step, we obtain a numerical estimate of

Downloaded from informs.org by [128.95.104.109] on 21 July 2016, at 00:33 . For personal use only, all rights reserved.

Downloaded from informs.org by [128.95.104.109] on 21 July 2016, at 00:33 . For personal use only, all rights reserved.

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 the 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 observed and unobserved state variables.

Our method requires two key assumptions. First, conditional on bids, a buyer's equilibrium decision is independent of auction-specific unobservables. For example, in procurement auctions, the unobserved difficulty of the job may act as a common shock on all sellers' costs, i.e., all sellers may quote higher prices for a difficult job. These prices obviously affect buyers' decisions. However, the shock itself is assumed to be irrelevant to the buyer's decision. If this assumption is violated, our estimator is still applicable, but requires parametric assumptions on buyers' decisions (an extension we present in §6.6.2). Second, auction-specific unobservables may affect bid prices conditional on entry, but not the types of bidders entering the auction. This assumption is satisfied if the maximum bid is nonbinding and there are no bidding costs.

Our method has many advantages. First, it can be used to estimate auctions that do not have prespecified 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 require any parametric assumptions on seller types, seller attributes or bid distributions. Fourth, it does not require us to take a structural stance on the buyer's decision process. In auctions with sufficiently small state spaces, the entire estimation procedure can be nonparametric. Of particular importance, our method provides a new solution to the problem of unobserved auction 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, i.e., 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 many settings including ours (see §5 for details). 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.

We present an application of our model and estimation framework in the context of online freelancing. Freelance marketplaces are websites that match buyers of electronically deliverable services with sellers or freelancers. Services procured through these websites fall under the categories of Web development, programming, writing, translation, design, and multimedia (Kozierok 2011). These websites typically use beauty contest auctions to match buyers and sellers, that is, the lowest priced bidder is not the default winner; rather the buyer chooses the winner based on her discretion. In doing so she may trade off sellers' reputations, bid prices, and other attributes. In the last few years, online freelancing has grown tremendously and generates over \$360 million in revenue (Morgan 2011). Because online freelancing is playing an increasingly important role in the global labor market, uncovering the underlying distribution of employee costs is of paramount importance to many players. First, information on seller costs can help us retrieve seller margins or market power, and understand the extent of competitiveness in this marketplace. Second, from a managerial perspective, knowing the distributions of seller costs can help managers optimize the procurement mechanism on their sites.

In our application, we estimate the distribution of seller costs in a prominent freelance marketplace and present the following findings. First, we derive the inferred dollar values of sellers' costs for auctions with a buyer specified nonbinding MaxBid 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 seller costs; 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 distribution of costs for High type auctions. The dollar amount of the margins in this marketplace is not very high, with average percentage margins around 15% of the bid, i.e., the marketplace is quite competitive, and consequently sellers do not enjoy much market power.

Second, we find that not accounting for unobserved heterogeneity can significantly bias the estimates of seller costs. Without unobserved heterogeneity, there is a 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.

RIGHTSLINK

Downloaded from informs.org by [128.95.104.109] on 21 July 2016, at 00:33 . For personal use only, all rights reserved.

Third, 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 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 a 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. Because there is a significant difference between the bids placed in High and Low type auctions (over $150), managers face the following important question: 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 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. We find that sellers from the Indian subcontinent 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 the 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. Of particular importance, our method can handle nonmultiplicatively separable auction-specific unobservables. Second, from a substantive perspective, we derive the sellers' cost distributions in a prominent freelance marketplace and show that freelancing is a 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 by helping them to evaluate the relative value of the two sides of the market and the value of players from different geographies.

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 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; 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. Guerre et al. (2000) developed this approach in the context of first price auctions with Independent Private Values (IPV) symmetric bidders. It uses the relationship between the equilibrium distributions of observed bids, a seller's private cost, and her bid, to back out her private cost. The original method has been augmented in many directions; for example, Li et al. (2002) extend it to the 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. The main difference between the previous papers and our paper is that we do not have closed form expressions for sellers' equilibrium strategies. Here, the expected probability of winning is not only a function of the observed distribution of bids but also of the unobserved allocation rule used by buyers. Thus, our proposed estimation and identification framework expands the range of auction settings that are amenable to nonparametric estimation.

Third, our paper relates to the literature on the estimation of strategic games in marketing, e.g., supermarket entry models (Singh and Zhu 2008, Orhun 2013), demand estimation with social interactions and

RIGHTSLINK

Downloaded from informs.org by [128.95.104.109] on 21 July 2016, at 00:33 . For personal use only, all rights reserved.

joint consumption (Hartmann 2010, Narayanan 2013), product introductions (Draganska et al. 2009), network effects (Shriver 2015), and pricing strategies (Ellickson and Misra 2008, Ellickson et al. 2012). See Ellickson and Misra (2011) for an excellent survey of this literature. The main difference between these papers and ours is that they study discrete decisions, whereas in our case the decisions are continuous.

Fourth, our paper relates to the literature on online auctions. Bajari and Hortacsu (2004) provide a detailed discussion of recent advances and the remaining challenges in this area. In the marketing context, Yao and Mela (2008) model seller and buyer behavior and estimate the impact of varying commission rates and the value of sellers to the marketplace. Of course, these papers pertain to price-only auctions, not beauty contests.

Finally, our paper relates to a small, but growing literature on procurement auctions without prespecified 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, her focus is building a partial equilibrium model of the buyer's optimization problem taking the sellers' side as given. Here, we aim to uncover the distribution of private costs by modeling sellers' strategic bidding behavior and buyers' choice decisions in a static setting. Thus, we forgo dynamics to solve a full-equilibrium model. Second, the technical challenges that the two papers address are very different. In the former, the main issue is controlling for dynamic selection within auctions, whereas our challenge is controlling for unobserved auction heterogeneity. Third, because we have a full-equilibrium model, we can answer a broader set of questions that Yoganarasimhan (2013) cannot.

3. Empirical Framework

We first present the basic model and estimation framework, and then expand it to include unobserved auction heterogeneity in §5.

3.1. Set-Up

We use a nomenclature consistent with procurement auctions, i.e., the auctioneer is the buyer and bidders are sellers. All of the players are risk-neutral. We use script letters to denote functions, capital letters to denote sets of variables, and small letters to denote a single variable. Table 1 summarizes all of the nomenclature and variables used in the model set-up.

There are n buyers, indexed by i, who conduct one sealed bid beauty contest auction each, referred to as auction i, with some abuse of notation.

Table 1 Nomenclature

	Variable	Description
Auction-specific observables (A_i)	O_i	Observable auction and buyer attributes of auction i
	q_i	Number of bids received by auction i
	r_i	Commission rate charged by the third-party conducting the auction
Auction-specific unobservable	v_i	Latent type of auction (drawn from k finite types) Population probability of type $k = \pi_k$
Seller attributes	X_{ji}	Seller-specific state variables for bidder j in auction i
	c_{ji}	Seller j's cost doing job i
Seller decision	b_{ji}	Seller j's bid in auction i
Equilibrium distributions	$\mathcal{G}(\cdot)$	Equilibrium distribution of bids and seller attributes
	$\mathcal{P}(\cdot)$	Buyer i's probability of choosing seller j given competing bids $\{X_{-ji}, b_{-ji}\}$ or CCP
	$\mathcal{S}(\cdot)$	Bidder j's expected probability of winning auction i

Note. CCP, Conditional choice probability.

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. It includes buyer-specific variables that are constant for the duration of the auction, properties of the auctioned object, and a nonbinding maximum bid (MaxBid) or the maximum price that the buyer is willing to pay.

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; the commission is 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, where q_i is known to or observed by all sellers.[2] Hereafter, we refer to $\{O_i, q_i, r_i\}$ as A_i, the complete set of observed auction-specific variables. A_i constitutes all of the attributes of the auction that are invariant across sellers and observed by the researcher.

Seller-specific Variables. Sellers are indexed by j, and for auction i, $j \in \{1, 2, \ldots, q_i\}$. Let $\{c_{ji}, X_{ji}\}$ be the state variables that seller j is endowed with at the beginning of auction i. Specifically, X_{ji} is the set of observable seller attributes that are relevant to the auction, i.e., which affect buyer i's probability of choosing

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

RIGHTSLINK

Downloaded from informs.org by [128.95.104.109] on 21 July 2016, at 00:33 . For personal use only, all rights reserved.

seller j. It is assumed to be fixed for the duration of auction i, that is, seller j cannot optimize on it.[3] 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. Further, c_{ji} is seller j's private cost of completing the project specified in auction i. It can be expressed as

$$c_{ji} \equiv c_{ji}(O_i, X_{ji}, \tilde{c}_{ji}). \tag{1}$$

Intuitively, a seller's cost of providing a service can vary with her own attributes, with the auction's attributes, and a private shock $\tilde{c}_{ji}$ (Athey and Haile 2007). For example, experienced programmers may have lower programming costs and experienced buyers may be easier (less costly) to work for. The cost shock $\tilde{c}_{ji}$ is assumed to be drawn from a distribution $\mathcal{F}(\tilde{c}_{ji})$ that is independent of $\{O_i, X_{ji}\}$. For notational simplicity, we henceforth denote $c_{ji}(A_i, X_{ji}, \tilde{c}_{ji})$ 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 posts the auction and the auction-specific variables, $A_i = \{O_i, r_i, q_i\}$, become visible to the sellers. At stage 2, each seller $j \in \{1, \ldots, q_i\}$ submits 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 auction attributes, A_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 A_i, and the joint distribution of seller costs and types, c_{ji} and X_{ji}, given A_i. They do not observe the costs, bids or attributes of the other sellers.

3.2. Sellers' Optimization

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

$$u(A_i, X_{ji}, c_{ji}, b_{ji}) = [b_{ji}(1-r_i) - c_{ji}] \cdot \mathcal{S}_{ji}(X_{ji}, b_{ji} \mid A_i), \tag{2}$$

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

$$\begin{aligned}\mathcal{S}_{ji}(X_{ji}, b_{ji} \mid A_i) &= E[\mathcal{P}_{ji}(X_{ji}, b_{ji}, X_{-ji}, b_{-ji} \mid A_i)] \\ &= \int \mathcal{P}_{ji}(X_{ji}, b_{ji}, X_{-ji}, b_{-ji} \mid A_i) \\ &\quad \cdot \mathcal{G}(X_{-ji}, b_{-ji} \mid A_i)\, d(X_{-ji}, b_{-ji} \mid A_i),\end{aligned} \tag{3}$$

where $\mathcal{P}_{ji}(X_{ji}, b_{ji}, X_{-ji}, b_{-ji} \mid A_i)$ is buyer i's equilibrium probability of choosing seller j given A_i, seller j's own attributes and price $\{X_{ji}, b_{ji}\}$, and the attributes and prices of all other bidders in the auction, $\{X_{-ji}, b_{-ji}\}$.[4] $\mathcal{P}_{ji}(\cdot)$ 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). $\mathcal{G}(X_{-ji}, b_{-ji} \mid A_i)$ 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 nonbinding, we have an unconstrained maximization problem,[5] and the seller's optimization problem can be expressed as

$$\max_{b_{ji}} [b_{ji}(1-r_i) - c_{ji}] \cdot \mathcal{S}_{ji}(X_{ji}, b_{ji} \mid A_i). \tag{4}$$

The First Order Condition (FOC) of this problem is

$$\begin{aligned}&[b_{ji}(1-r_i) - c_{ji}] \cdot \frac{\partial \mathcal{S}_{ji}(X_{ji}, b_{ji} \mid A_i)}{\partial b_{ji}} \\ &\quad + (1-r_i) \cdot \mathcal{S}_{ji}(X_{ji}, b_{ji} \mid A_i) = 0.\end{aligned} \tag{5}$$

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

$$\begin{aligned}c_{ji} &= \xi(A_i, X_{ji}, b_{ji}, \mathcal{G}(\cdot), \mathcal{P}(\cdot)) \\ &= (1-r_i)\left[b_{ji} + \frac{\mathcal{S}_{ji}(X_{ji}, b_{ji} \mid A_i)}{\partial \mathcal{S}_{ji}(X_{ji}, b_{ji} \mid A_i)/\partial b_{ji}}\right].\end{aligned} \tag{6}$$

3.3. Assumptions

We make the following assumptions. 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 of their implications on identification, and the trade-offs involved in relaxing them. See §§3.4 and 6.6.2 for details.

[3] The set of observable seller attributes, X_{ji}, is assumed to be verifiable by the buyer and cannot be falsified by the seller.

[4] The buyer's equilibrium CCP, $\mathcal{P}_{ji}(\cdot)$, 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.

[5] In settings with nonbinding maximum bids, the estimated cost distributions are uncensored (Athey and Haile 2007).

RIGHTSLINK

Downloaded from informs.org by [128.95.104.109] on 21 July 2016, at 00:33 . For personal use only, all rights reserved.

Assumption 1 (Conditional Independence). *In sealed-bid auctions, conditional independence follows naturally, i.e., sellers are drawn independently from a joint distribution* $\mathcal{F}(\tilde{c}_{ji}, X_{ji} | A_i)$.

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. The main advantage of this assumption is that it implies that sellers' bids and attributes $\{X_{ji}, b_{ji}\}$ can be drawn independently from $\mathcal{G}(X_{ji}, b_{ji} | A_i, v_i)$.

Assumption 2 (Continuity and Monotonicity). *We assume that,* $\mathcal{P}_{ji}(X_{ji}, b_{ji}, X_{-ji}, b_{-ji} | A_i)$, *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}.

According to this assumption, 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 (SOC)). *The SOC of the seller's maximization problem is assumed to be satisfied.*

$$[b_{ji}(1-r_i)-c_{ji}]\frac{\partial^2 \mathcal{S}_{ji}(X_{ji}, b_{ji}|A_i)}{\partial b_{ji}^2} + 2(1-r_i)\frac{\partial \mathcal{S}_{ji}(X_{ji}, b_{ji}|A_i)}{\partial b_{ji}} < 0. \tag{7}$$

Note that $\partial \mathcal{S}_{ji}(X_{ji}, b_{ji}|A_i)/\partial b_{ji} < 0$, $1-r_i > 0$, and under individual rationality constraints, $[b_{ji}(1-r_i) - c_{ji}] > 0$. Hence, for condition (7) to be satisfied, we require that $\partial^2 \mathcal{S}_{ji}(X_{ji}, b_{ji}|A_i)/\partial b_{ji}^2$ is not too positive, i.e., we assume that the function $\mathcal{S}_{ji}(X_{ji}, b_{ji}|A_i)$ is not too convex.

Assumption 4 (Symmetry and Private Information). *All auctions and buyers are assumed to be symmetric after accounting for* A_i, *and all sellers are assumed to be symmetric after accounting for* $\{X_{ji}, b_{ji}\}$. *This implies that* $\mathcal{P}_{ji}(\cdot) = \mathcal{P}(\cdot)\ \forall j, i$.

Given a draw of $\{A_i, X_{-ji}, b_{-ji}\}$, seller j's probability of winning auction i is the same as that of another seller k if her attributes and price are the same as k's. Private information 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 which 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. Finally, it assumes that there exists no auction-specific state variable that is observable to the buyer and affects his decision, and is observable to the sellers and affects their costs, but is unobservable to the researcher. In §5, we present some relaxations of this assumption.

Assumption 5 (Zero Bidding Costs). *Bidders are assumed to have zero costs associated with learning their costs of doing the job and in preparing the bid.*

This ensures that sellers do not systematically enter or avoid auctions based on some seller-specific variables. For example, if it were costly to place a bid, those with very high costs or those with unfavorable attributes (e.g., low reputation) may choose not to bid at all because their expected returns from bidding may be lower than the cost of bidding.

This assumption ensures that the estimated cost distributions are uncensored. It is reasonable in Internet settings, where sellers' costs of learning about auctions and submitting bids is negligible, as opposed to timber or art auctions, wherein sellers have to spend a significant amount of time learning about their own valuation of the products and bid preparation.

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

3.4. Identification

With the assumptions listed above, it is easy to see that $\mathcal{P}(\cdot)$ and $\mathcal{G}(\cdot)$ are nonparametrically identified in the data. 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. Combined with the FOC, this 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 heterogeneity in auctions and sellers and the strategic behavior of sellers, the variation in bids is assumed to stem purely from the variations in the private costs of sellers.

RIGHTSLINK

Downloaded from informs.org by [128.95.104.109] on 21 July 2016, at 00:33 . For personal use only, all rights reserved.

Note that all of the unobservables in this model are the private information of agents. Private information is nonstrategic in that it does not enter the opponent's optimization problem; whereas common knowledge information is strategic in that other players can condition their own optimization on it. Allowing for common knowledge unobservables can jeopardize the identification of $\mathcal{P}(\cdot)$ and $\mathcal{G}(\cdot)$; see Einav (2003) for a detailed discussion of this issue. However, it is possible to relax some aspects of this assumption. In §5, we expand the model to allow for auction-specific common knowledge unobservables that affect sellers' costs. We discuss its identification in §5.2.

4. Estimation

Recall that the lack of a prespecified buyer allocation rule and the multidimensionality of bid attributes implies that this problem does not have a closed form bidding strategy. However, our specification of beauty contest auctions as incomplete information games allows us to estimate cost distribution without actually solving for the optimal bidding strategy.

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

$$\begin{aligned} c_{ji} &= \xi(A_i, X_{ji}, b_{ji}, \mathcal{G}(\cdot), \mathcal{P}(\cdot)) \\ &= (1-r_i)\left[b_{ji} + \frac{\hat{\mathcal{S}}(X_{ji}, b_{ji} \mid A_i)}{\partial \hat{\mathcal{S}}(X_{ji}, b_{ji} \mid A_i)/\partial b_{ji}}\right], \end{aligned} \tag{8}$$

where $\hat{\mathcal{S}}(X_{ji}, b_{ji} \mid A_i)$ and $(\partial \hat{\mathcal{S}}(X_{ji}, b_{ji} \mid A_i)/\partial b_{ji})^{-1}$ are the numerical estimates of bidder j's expected probability of winning auction i and its derivative w.r.t. to j's bid. We can write out $\hat{\mathcal{S}}(X_{ji}, b_{ji} \mid A_i)$ as

$$\begin{aligned} \hat{\mathcal{S}}(X_{ji}, b_{ji} \mid A_i) = \int & \hat{\mathcal{P}}(X_{ji}, b_{ji}, X_{-ji}, b_{-ji} \mid A_i) \\ & \cdot \hat{\mathcal{G}}(X_{-ji}, b_{-ji} \mid A_i)\, d(X_{-ji}, b_{-ji} \mid A_i). \end{aligned} \tag{9}$$

Hence, to obtain numerical estimates of $\hat{\mathcal{S}}(\cdot)$ and $(\partial \hat{\mathcal{S}}(\cdot)/\partial b_{ji})^{-1}$, we need to estimate $\mathcal{P}(X_{ji}, b_{ji}, X_{-ji}, b_{-ji} \mid A_i)$ and $\mathcal{G}(X_{-ji}, b_{-ji} \mid A_i)$, which represent sellers' beliefs on the probability of being chosen given a set of state variables and the equilibrium distribution of bids. Under the assumption of rational expectations and unique equilibrium in data, these distributions are available from data, which we estimate in the first step. In the second step, we use first step estimates to simulate the expected probability of winning, and its derivative.

Estimation of beauty contest auctions is thus different from the estimation of the traditional first price auctions (Guerre et al. 2000). 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. By contrast, in beauty contests, the researcher needs to estimate the joint distribution of bids and seller attributes, as well as estimate the buyers' CCP or allocation rule. Moreover, expected probabilities of winning and their derivatives are not directly available from the first step estimates; rather, they have to be obtained using simulations. We discuss the estimation steps in detail below.

4.1. First Step Estimation

In the first step, we estimate the buyers' decision rule or CCP, their derivatives, and the equilibrium distributions of bids and seller attributes. We describe each in detail now:

- CCP—$\mathcal{P}(X_{ji}, b_{ji}, X_{-ji}, b_{-ji} \mid A_i)$.

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—$(\partial \mathcal{P}(X_{ji},\ b_{ji},\ X_{-ji},\ b_{-ji} \mid A_i)/\partial b_{ji})^{-1}$.

This is readily available from $\mathcal{P}(\cdot)$ by computing the numerical derivative. Note that $\mathcal{P}(\cdot)$ and $(\partial \mathcal{P}(\cdot)/\partial b_{ji})^{-1}$ simply capture predicted outcomes and rate of change of the predicted outcomes as functions of observed state variables. Hence, we do not need to model or take any stance on the buyers' optimization problem to consistently estimate them.

- Joint distributions of equilibrium bids and seller attributes—$\mathcal{G}(X_{ji}, b_{ji} \mid A_i)$.

These distributions represent sellers' expectations on the competitors' bids and attributes in equilibrium. With sufficient data, they can be obtained directly from the data.

All of the above distributions can be 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 use semi-parametric or parametric methods to maximize the fit or the predictive ability of these first stage models.

4.2. Second Step Estimation

Given the first step results, we obtain the numerical estimates of $\hat{\mathcal{S}}(\cdot)$ and $(\partial \hat{\mathcal{S}}(\cdot)/\partial b_{ji})^{-1}$ 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 $\hat{\mathcal{G}}(X_{ji}, b_{ji} \mid A_i)$. Denote these

Downloaded from informs.org by [128.95.104.109] on 21 July 2016, at 00:33 . For personal use only, all rights reserved.

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, $\tilde{\mathcal{P}}(X_{ji},b_{ji},\tilde{X}_{-ji},\tilde{b}_{-ji}|A_i)$ and $\partial\tilde{\mathcal{P}}(X_{ji},b_{ji},\tilde{X}_{-ji},\tilde{b}_{-ji}|A_i)/\partial b_{ji}$.
- *Step* 3. Repeat Steps 1 and 2 L times and take the averages to obtain

$$\hat{\mathcal{P}}(X_{ji},b_{ji}|A_i)=\frac{\sum_{l=1}^{L}\tilde{\mathcal{P}}(X_{ji},b_{ji},\tilde{X}^l_{-ji},\tilde{b}^l_{-ji}|A_i)}{L}, \quad (10)$$

$$\frac{\partial\hat{\mathcal{P}}(X_{ji},b_{ji}|A_i)}{\partial b_{ji}}=\frac{1}{L}\sum_{l=1}^{L}\frac{\partial\tilde{\mathcal{P}}(X_{ji},b_{ji},\tilde{X}^l_{-ji},\tilde{b}^l_{-ji}|A_i)}{\partial b_{ji}}, \quad (11)$$

where $\{\tilde{X}^l_{-ji},\tilde{b}^l_{-ji}\}$ is the lth set drawn (from Step 2). While one set of draws is sufficient for consistency, we set $L=1{,}000$ in our estimation to improve the efficiency of the results.

Finally, using Equation (9), we infer the costs of each seller j in each auction i, which are then used to nonparametrically estimate the cost distribution for different levels of observed auction characteristics.

5. Unobserved Auction Heterogeneity

So far, we have assumed that all unobserved variables are private information. However, in many settings, factors that affect seller 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 thus 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.

Auction-specific unobservables can arise from a number of factors, such as the difficulty of the job, the choosiness of the buyer, and the urgency of the job. All of these affect the sellers' bidding strategies, but are unobservable to the researcher. Consider the two similar auctions A and B in Figure 1 that differ in difficulty levels. Auction A is a simple web-development job that requires no special skills, whereas auction B requires mobile app development skills and programming in objective C. Bidders in A will take into account the easiness of the job and bid low (and anticipate low bids from other bidders), whereas those in B will bid high (and anticipate high bids from others). Thus the observed distribution of bids for both auctions would be very different. If we do not account for the unobserved difficulty of the job in our estimation, we will incorrectly attribute the high bids in auction B to high costs of sellers and low bids in auction A to low seller costs. Allowing for auction-specific unobservables avoids this incorrect inference.

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, i.e., the unobservable cost of the seller and the unobservable auction heterogeneity. A seller who bids high may do so because she has high programming costs or because the job is inherently difficult even for skilled programmers. Thus, separating these two instances is essential for identification. In the context of first price auctions, 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 seller costs multiplicatively.[6] She then uses a deconvolution method to recover both of these unobservables. However, Krasnokutskaya's (2011) 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 the seller's margins are multiplicatively separable in v_i, and thereby allows her to use 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, 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 potential solutions from the literature can be imported to solve this problem. First, if we are willing to assume that unobserved heterogeneity affects some, but not all, outcomes, 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

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

RIGHTSLINK

Downloaded from informs.org by [128.95.104.109] on 21 July 2016, at 00:33 . For personal use only, all rights reserved.

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 it are not valid in our setting.[7] 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 homogeneous, then the first stage estimation can be at the market level. Misra and Nair (2011) and Ellickson and Misra (2012) use this approach in the games context. While this method is ideal if the researcher has access to data sets with long panels of buyers and if the only source of unobserved heterogeneity is from the buyer (rather than the auction), such data sets are seldom available in auction settings.

To address these drawbacks, 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 independent and identically distributed (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).

5.1. Model with Unobserved Auction Heterogeneity

We now modify the model in §3 to include an auction-specific unobservable v_i, which is independent of the observables A_i, and which is drawn from a set of finite types $v_i \in \{v^1, v^2, \ldots, v^K\}$. Let π_k be the population probability that an auction is an unobserved type k (see Table 1). Now the cost retrieved from the FOC of the sellers' profits is

$$c_{ji} = (1-r_i)\cdot\left[b_{ji} + \frac{\mathcal{S}(X_{ji}, b_{ji} \mid A_i, v_i)}{\partial \mathcal{S}(X_{ji}, b_{ji} \mid A_i, v_i)/\partial b_{ji}}\right], \qquad (12)$$

where $\mathcal{S}(X_{ji}, b_{ji} \mid A_i, v_i) = \int \mathcal{P}(X_{ji}, b_{ji}, X_{-ji}, b_{-ji} \mid A_i, v_i)\cdot \mathcal{G}(X_{-ji}, b_{-ji} \mid A_i, v_i)\, d(X_{-ji}, b_{-ji} \mid A_i, v_i)$. The main difference between the FOC of this model and the one without unobserved heterogeneity is that $\mathcal{S}(\cdot)$ is now a function of v_i. Recall that cost c_{ji} can vary with seller and auction attributes. While previously, this included only observed seller and auction attributes, now it also includes the unobservable v_i. Hence, c_{ji} in Equation (12) can be expanded as $c_{ji}(A_i, v_i, X_{ji}, \tilde{c}_{ji})$. However, as before, to keep the notation simple, we denote it as c_{ji}.

5.2. Identification with Unobserved Auction Heterogeneity

We now present a discussion of identification with unobserved auction heterogeneity. Now $\mathcal{G}(X_{-ji}, b_{-ji} \mid A_i, v_i)$ and $\mathcal{P}(X_{ji}, b_{ji}, X_{-ji}, b_{-ji} \mid A_i, v_i)$ are not directly available from the data because they are functions of the unobservable state variable v_i. Moreover, the mixture probabilities π_ks must be identified. Since v_i is drawn from a set of finite types, this devolves to identifying a nonparametric mixture model.

We start by considering the identification of component distributions $\mathcal{G}(X_{-ji}, b_{-ji} \mid A_i, v_i)$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. Our identification proof, presented in Web Appendix A (available as supplemental material at http://dx.doi.org/10.1287/mksc.2015.0929), follows along the same lines as theirs. 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.[8]

We now explain this identification constraint using a very simple example. Consider a setting with two unobserved auction types with population probabilities π and $1-\pi$. Suppose that each auction gets two

[7] An example of a case where this would work is: 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 affected. However, in most realistic settings, if auction-specific unobservables influence sellers' entry, then they are also likely to affect bid prices.

[8] 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; 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.

RIGHTSLINK

Downloaded from informs.org by [128.95.104.109] on 21 July 2016, at 00:33 . For personal use only, all rights reserved.

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, $\{\pi, a,$ and $b\}$, to identify. In terms of data, we have three types of 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. These 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 $\{HHH, LLL, LLH,$ and $HHL\}$); so 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 treat 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. 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 nonparametrically identified because we only observe one decision by the buyer in each auction. Without repeat buyer decision observations within the same auction, there is no way 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 nonidentification result is as follows: When 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 to the buyer-auction-specific unobservable v_i. See Kasahara and Shimotsu (2009) for a detailed nonidentification proof. Therefore, we make the following assumption.

Assumption 6 *Conditional on bids, 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 6 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 a buyer only pays the actual bid, this is reasonable.[9]

However, 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 parameterize the buyers' decision rule; then it is possible to simultaneously estimate nonparametric mixtures of bid distributions and a parametric mixture model of buyers' decisions in the first step of estimation. We present this extension in §6.6.2. Alternately, if the researcher is willing to treat v_i as simply buyer-(and not auction-)specific unobservable, and has long buyer panels, then she may estimate the decision rule at the buyer level.

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 conditional cost distribution, $\mathcal{F}(c_{ji} | X_{ji}, A_i, v_i)$, is naturally identified; there is a one-to-one relationship between bids and costs.

5.3. Estimation with Unobserved Auction Heterogeneity

The estimation of $\mathcal{P}(\cdot)$ and its derivative proceeds as before because it is a function of v_i. Hence, the main challenge lies in estimating bid distributions, $\mathcal{G}(X_{-ji}, b_{-ji} | A_i, v_i)$, which now depend on v_i and hence are not directly available from the data. We observe bid distributions $\mathcal{G}(X_{-ji}, b_{-ji} | A_i)$. For each realization of A_i, we need to recover K distributions, $\mathcal{G}(X_{-ji}, b_{-ji} | A_i, v^k)$s, and the population distribution of the K unobserved types. Moreover, this entire exercise must be nonparametric because parameterization of bid distributions can bias the estimates of seller costs. To address these challenges, we use a nonparametric EM-like algorithm to estimate the components $\mathcal{G}(X_{-ji}, b_{-ji} | A_i, v^k)$ and π_ks.

As before, in the first step, we estimate the buyer's decision rule or the CCP, its derivative, the equilibrium distributions of bids and seller attributes, and the population distribution of unobserved types. The estimation of the buyer's decision and its derivative are the same as that described in §4.1 since it is not conditioned on v_i. Below, we discuss the estimation of $\mathcal{G}(X_{ji}, b_{ji} | A_i, v_i)$ and π_ks.

5.3.1. 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

[9] This assumption can be restrictive if the buyer's outside option on canceling a job is to do it himself; the unobserved difficulty of the job will influence his decision. In such cases, our estimator is not directly applicable.

RIGHTSLINK

Downloaded from informs.org by [128.95.104.109] on 21 July 2016, at 00:33 . For personal use only, all rights reserved.

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 nondecreasing 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).

Instead of the EM-like algorithm, we can also use a Majorization-Minimization (MM) algorithm that guarantees convergence. See Levine et al. (2011) for details. However, there have been no documented situations where MM-algorithms converge and EM-like do not. Hence, we stick to an EM-like algorithm.

Basic Setup. Consider a setting wherein seller attributes, X_{ji}, are a set of continuous state variables. Recall that the unobserved type v_i comes from a finite set, $\{v^1, \ldots, v^K\}$. 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, $A_i \in \{A^1, A^2, \ldots, A^H\}$. Then all of 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 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. We now present an iterative algorithm where each iteration consists of three steps, i.e., the Kernel Density Estimation (KDE) step, the Expectation (E) step, and the Maximization (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.

Kernel Density Estimation (KDE)-Step. Let the dimensionality of $\{X_{ji}, b_{ji}\}$ be D, and let Z denote a point in this D-dimensional space. Furthermore, the bids in each group are indexed by $m \in \{1, 2, \ldots, n_h\}$, where n_h is the total number of bids in group h where $h \in \{1, \ldots, H\}$. Then $\mathscr{G}_{h,k}^t(Z \mid A^h, v^k)$ denotes the multivariate KD 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. Because 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 $\mathscr{G}_{h,k}^t(Z \mid A^h, v^k)$ as

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

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

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

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

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

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

$$\pi_k^t = \frac{\sum_{i=1}^n \lambda_{ik}^t}{n}, \quad \forall k. \quad (15)$$

We iteratively perform the three steps until the population probabilities (π_ks) converge, at which point, we have consistent estimates of the population probabilities of unobserved types, the 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, $\mathscr{G}_{h,k}(X_{ji}, b_{ji} \mid A^h, v^k)$.

5.3.2. 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 $\hat{\mathscr{S}}(\cdot)$ and $(\partial \hat{\mathscr{S}}(\cdot)/\partial b_{ji})^{-1}$ for each seller j, in each auction i, for each unobserved state v^k. Below, we describe the steps in detail.

1. For each type v^k belonging to $\{v^1, \ldots, v^K\}$, 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 $\hat{\mathscr{G}}_{h,v^k}(X_{ji}, b_{ji} \mid A_i = A^h, v^k)$. 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 for $v_i = v^k$.

Downloaded from informs.org by [128.95.104.109] on 21 July 2016, at 00:33 . For personal use only, all rights reserved.

(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, $\tilde{\mathcal{P}}(X_{ji}, b_{ji}, \tilde{X}_{-ji}, \tilde{b}_{-ji} \mid A_i)$ and $\partial \tilde{\mathcal{P}}(X_{ji}, b_{ji}, \tilde{X}_{-ji}, \tilde{b}_{-ji} \mid A_i)/\partial b_{ji}$.

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

$$\hat{\mathcal{S}}(X_{ji}, b_{ji} \mid A_i, v^k) = \frac{\sum_{l=1}^{L} \tilde{\mathcal{P}}(X_{ji}, b_{ji}, \tilde{X}^l_{-ji}, \tilde{b}^l_{-ji} \mid A_i)}{L}, \quad (16)$$

$$\frac{\partial \hat{\mathcal{S}}(X_{ji}, b_{ji} \mid A_i, v^k)}{\partial b_{ji}} = \frac{1}{L} \sum_{l=1}^{L} \frac{\partial \tilde{\mathcal{P}}(X_{ji}, b_{ji}, \tilde{X}^l_{-ji}, \tilde{b}^l_{-ji} \mid A_i)}{\partial b_{ji}}, \quad (17)$$

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

(iv) Using the above estimates of $\hat{\mathcal{S}}(X_{ji}, b_{ji} \mid A_i, v^k)$ and $\partial \hat{\mathcal{S}}(X_{ji}, b_{ji} \mid A_i, v^k)/\partial b_{ji}$ and Equation (12), obtain costs c_{ji}.

(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 data points and the posterior probabilities λ_{ik}s as the weights.

5.4. 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 (i.e., all of the X_{ji}s are continuous). Second, it has very high data requirements because it relies on multidimensional 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 multimodalities 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 multimodalities, it is possible to ease the data burden by using large bandwidths. However, large bandwidths will smooth out multimodalities, making the recovery of the underlying KDEs difficult. To address this issue, we suggest the following assumption (which we also implement in our estimation).

Assumption 7. *The joint distribution of seller attributes and bid prices is multiplicatively separable in seller attributes and bid prices as follows*:

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

According to Assumption 7, 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, on 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.

With some small modifications, the estimation algorithm described earlier can be made to accommodate Assumption 7. It would involve a more sequential first step estimation, as follows: Within the first-step, first estimate $\mathcal{G}^X(X_{ji} \mid A_i)$, and then use an EM-like algorithm to estimate $\mathcal{G}^b(b_{ji} \mid X_{ji}, A_i, v_i)$. See Web Appendix B for step-by-step instructions on the modified estimation.

There are two additional points of note about Assumption 7. First, it is an assumption of convenience; it need not be made if the researcher is addressing only a small number of seller attributes and has access to large data sets. Second, and more important, it is testable. In §6.6.2, we present empirical tests to confirm the validity of this assumption.

5.5. Discussion

We now discuss how our estimation method relates to the broader literature on accommodation of common knowledge or persistent unobservables within two-step methods.

First, in the context of auctions, compared to Krasnokutskaya's (2011) 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.

Second, our method 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 nonparametric EM-like algorithm proposed by Benaglia et al. (2009a). They use a kernel smoothing procedure to accommodate continuous state spaces instead of the binning procedure advocated by Arcidiacono and Miller (2011). Furthermore, in the second step, we use the FOC condition to invert costs instead of using a maximum likelihood procedure.

Downloaded from informs.org by [128.95.104.109] on 21 July 2016, at 00:33 . For personal use only, all rights reserved.

Finally, note that our method does not come with guaranteed convergence or asymptotics for rates of convergence because we use an EM-like algorithm in the first step that does not have proven convergence. However, if the first-step algorithm does converge and the model used has been shown to be nonparametrically identified, then we can be assured that the algorithm has converged to the correct parameter values because there is a unique fixed point in an identified model.

6. Application: Online Freelancing

We now present an application of our model and estimation framework in the context of online freelancing. Freelance marketplaces are websites that match buyers of electronically deliverable services with sellers or freelancers. The most popular freelance marketplaces are Elance, Guru, vWorker, ODesk, and Freelancer, and the most popular categories of jobs are Web development, programming, writing, translation, design, and multimedia (Kozierok 2011). These websites typically use beauty contest auctions to match buyers and sellers.

While agents can buy and sell services online without going through a freelance marketplace, freelance sites help mitigate the risks associated with trading online by offering escrow systems, arbitration services, and reputation mechanisms. Specifically, feedback-based reputation mechanisms address information asymmetry problems by collecting information on market participants over long periods of time and making it available to future players. While there is no theoretical consensus on the robustness of reputation systems (Holmstrom 1999, Cripps et al. 2004), Yoganarasimhan (2013) shows that the reputation systems in freelance marketplaces are indeed effective.

Spurred by technological innovations (fast Internet, the growing number of electronically doable and deliverable jobs) and their ability to connect cheap labor from developing countries with small business owners in developed economies, freelance markets have grown tremendously in the last few years (Morgan 2011). Uncovering the underlying distribution of employee costs is paramount to many players. First, from researchers' perspective, information on seller costs can help us retrieve seller margins and understand the extent of competitiveness in this marketplace. Second, from a managerial perspective, knowing the distributions of seller costs can help managers optimize the procurement mechanism on their sites. For example, what is the relative value of attracting more sellers to that of attracting more buyers to the marketplace? Thus, we now estimate the distributions of seller costs in a prominent freelance site, and use our estimates to address these questions.

6.1. Setting and Data

Our data comes from one of the leading online freelance firms in the 2006–2010 time frame. Site membership is free and there are no fees for either posting auctions or for bidding. The site uses a sealed bid auction format. Most auctions are technology-oriented and the majority (over 80%) fall under the information technology (IT) services category. Our data is composed of 4,002 auctions posted from January 1, 2006 to December 31, 2010 that have a nonbinding maximum bid of \$500. The site uses 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 (with optional text comments). A rating of 1 stands for very bad and 10 for excellent, and the site enjoys a high feedback rate with more than 90% of buyers rating the sellers they worked with.

The site charges a 15% commission on the transaction amount, 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 data 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. The site posts the auction on its public forum, which can be browsed by all of its members. Sellers can also obtain up-to-date information on new auctions by subscribing to newsletters from the site. The auction posting contains information provided by the buyer (e.g., project description, auction start date) as well as information on the buyer (e.g., past ratings and geographic location).
- *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. When submitting the bid, sellers can see the number of bids 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 an extension in §6.6.2, we allow sellers to have uncertainty on the total bids the auction will receive and condition their beliefs on the number of bids it has received so far.

RIGHTSLINK

Downloaded from informs.org by [128.95.104.109] on 21 July 2016, at 00:33 . For personal use only, all rights reserved.

Table 2 Code for Buyer and Seller Geographic Regions

Region code	Countries
1	Indian subcontinent—India, Pakistan, etc.
2	Developed countries—USA, Western Europe, etc.
3	Eastern Europe—Romania, Russia, etc.
4	Everything else—Philippines, China, etc.

- *Stage* 4. The buyer makes his decision by 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 2.
  - 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, the quality of his outside options, or both.
  - Number of past ratings and the sum of all past ratings.

Table 4 Region Distributions of Auction and Bids in the Data

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

Note. The leftmost column shows the percentage of auctions initiated by buyers from different regions.

  - 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, i.e., number of days since the buyer signed up.
- The following seller attributes for all bids received:
  - Bid price
  - Seller's geographic region (see Table 2).
  - 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; the rest are canceled. This is because the site does not charge any fees for posting/canceling auctions. So buyers 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. There is considerable heterogeneity across auctions in the number of bids received. An average auction receives about 11 bids, with the median being seven. Auctions in which the buyer picks a bid tend to receive slightly higher numbers of

Table 3 Summary Statistics of Auction and Buyer Attributes

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)	4,002
Deadline (days)	19.21	32.05	0	13	30	(1, 1132)	4,002
Number of buyer ratings	9.96	42.66	0	0	5	(0, 723)	4,002
Avg. ratings	4.62	4.89	0	0	10	(0, 10)	4,002
Avg. ratings (if rated)	9.73	0.83	9.89	10	10	(1, 10)	1,900
No. of past uncanceled auctions	7.02	18.09	0	0	5	(0, 262)	4,002
No. of past canceled auctions	8.51	29.18	0	1	6	(0, 597)	4,002
Indicator for attachment with auction	Freq. 0 = 3,300 (82.46%), Freq. 1 = 702 (17.54%)						4,002
				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

Downloaded from informs.org by [128.95.104.109] on 21 July 2016, at 00:33 . For personal use only, all rights reserved.

Table 5 Summary Statistics of Seller and Bid 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, 5,000)	44,274
Bid price (if $\leq$ 500)	416.98	122.85	350	500	500	(85, 5,000)	43,393
Number of seller ratings	18.64	49.41	0	2	16	(0, 1,343)	44,274
Avg. ratings	5.71	4.52	0	8.54	9.79	(0, 10)	44,274
Avg. ratings (if rated)	8.99	1.62	8.79	9.60	10	(1, 10)	28,101
Indicator for past interaction with buyer	Freq. 0 = 44,045 (99.48%), Freq. 1 = 229 (0.52%)						44,274
				Accepted bids			
Bid price	306.84	151.79	180	300	500	(85, 500)	810
Bid price (if $\leq$ 500)	306.84	151.79	180	300	500	(85, 500)	810
Number of seller ratings	46.35	94.46	6	19	52	(0, 1,343)	810
Avg. ratings	8.85	2.77	9.44	9.79	10	(0, 10)	810
Avg. ratings (if rated)	9.70	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

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 sizable portion of buyers (52.5%) have no past ratings; 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 a 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. 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 reputations are more likely to pick a seller and not cancel the auction. Finally, as shown in Table 4, the majority (82.68%) of buyers are in English-speaking countries, i.e., Region 2.

Table 5 shows the summary statistics of seller and bid attributes for two sets of bids, i.e., all of the bids received and the 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 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 fraction of sellers actually quote prices higher than $500, with $5,000 being the maximum quote observed. Unlike buyers, the majority of the bidders (54.91%) are in the Indian subcontinent (Table 4). The other three regions are about equally represented (≈13%–17%). The distribution of the seller's geographic region also varies with the buyer's region.

There are systematic differences between the distributions of winning bidders and all of the bidders. On average, they quote lower prices (Table 5), have significantly better reputations, are more likely to be in developed countries, and have a higher likelihood (9.63%) of past interaction with the buyer.

6.2. Preliminary Analysis

We now present structure-free measures to quantify the beauty contest aspect of the auctions and the extent of auction-specific unobserved heterogeneity in the marketplace.

- Chosen Bid Gap

$$ChosenBidGap_i = \frac{\text{Chosen bid in auction } i - \text{Minimum bid in auction } i}{\text{Minimum bid in auction } i} \times 100. \tag{19}$$

This is the percentage difference between the chosen bid and the lowest bid. It is a measure of the beauty contest aspect of the auctions: If buyers place significant value on other seller attributes (apart from price), then this metric is high. In price-only auctions, the Chosen Bid Gap is always zero.

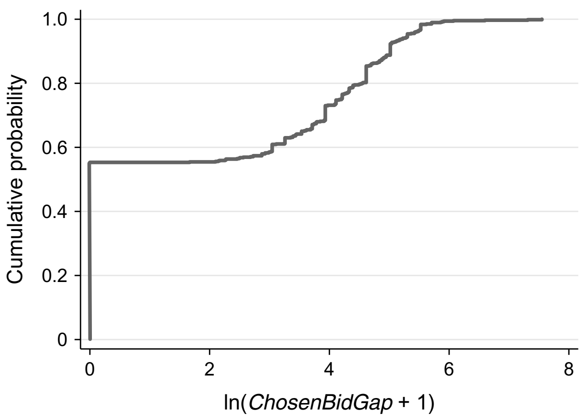
- Relative Dispersion

$$RelativeDispersion_i = \frac{\text{Standard deviation of bids in auction } i}{\text{Mean of bids in auction } i} \times 100. \tag{20}$$

This is 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.

Downloaded from informs.org by [128.95.104.109] on 21 July 2016, at 00:33 . For personal use only, all rights reserved.

Figure 2 ln(*ChosenBidGap* + 1) **for Auctions with Two or More Bids**



First, we present the CDF of ln(*ChosenBidGap* + 1) in Figure 2 for auctions in our data set that have received two or more bids (3,511 auctions). In nearly 40% of the auctions, the Chosen Bid Gap is more than 50%, i.e., 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 CDF of ln(*Relative Dispersion*) in Figure 3. The median Relative Dispersion within auctions is 30%, whereas the Relative Dispersion for all of the bids in the data set is 55%. Because 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, 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 6. 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 subcontinent and Eastern Europe (Regions 1 and 3). Furthermore, buyers with higher past reputations (high mean ratings and number of ratings) receive lower prices. This

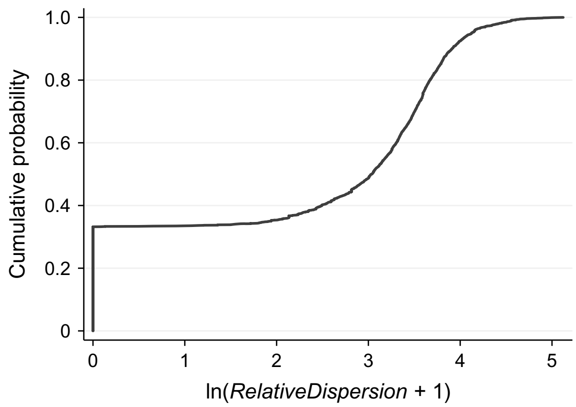
Table 6 **Bid Price Regressions**

Auction-, buyer-, and seller-specific explanatory variables	Model M1		Model M2	
	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 *and Buyer region* = 2	-2.786×10^{-3}	1.894×10^{-2}	-7.567×10^{-3}	1.541×10^{-2}
Seller region = 2 *and Buyer region* = 2	-1.066×10^{-2}	2.629×10^{-2}	-2.625×10^{-2}	2.129×10^{-2}
Seller region = 3 *and Buyer region* = 2	-3.996×10^{-2}	2.443×10^{-2}	-4.150×10^{-2}	2.005×10^{-2}
Seller region = *Buyer region* ≠ 2	3.778×10^{-3}	1.480×10^{-2}	-8.242×10^{-3}	1.249×10^{-2}
Indicator for no interactions between buyer and 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	44,274		44,274	

Notes. Dependent variable: ln(bid_{ji}). Robust standard errors shown.

Downloaded from informs.org by [128.95.104.109] on 21 July 2016, at 00:33 . For personal use only, all rights reserved.

Figure 3 ln(*RelativeDispersion* + 1) **for Auctions with Two or More Bids**



is also true for buyers who have posted many auctions on the site and those with a high success ratio. Sellers from developed countries charge the highest prices, followed by those in the Indian subcontinent. 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 difficulty of the job (or unobserved auction heterogeneity), and (3) the strategic behavior of sellers, which is highly nonlinear in observables and unobservables (Equation (12)), and hence not captured in this simple regression model. Because 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. Because of space constraints, we do not present the auction dummies in the table. Note, however, 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. This suggests that unobserved auction heterogeneity has a significant impact on bids.

To summarize, our preliminary findings suggest that there are significant variations in bid prices, and that such variations are driven, at least in part, by auction-specific unobservables. Armed with these insights, we now specify and estimate a structural model of seller behavior in freelance auctions.

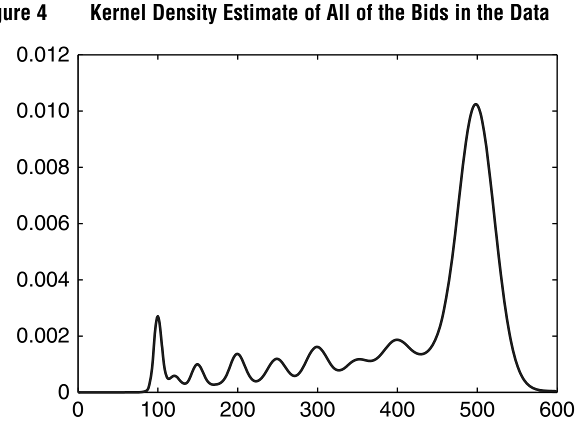
6.3. Applying the Model and Estimation to Freelance Setting

We now adapt our empirical framework to suit our setting. Note that we have four seller attributes, i.e., two continuous (number of ratings and mean rating) and two discrete (seller geographic region and indicator of past buyer-seller interaction). In the first step, we need to model each of these seller attributes, the buyers' decision model, and use the EM-like algorithm to separate the bid prices. In the second step, we use these first step estimates to retrieve the costs.

To avoid repetition and adhere to space limitations, we outline our estimation strategy here and refer interested readers to the Web Appendix for details.

1. First, we specify nonparametric models of the two continuous seller attributes, i.e., number of ratings and mean rating, as functions of observed buyer and auction attributes A_i. See Web Appendix §C.1 for details.
2. 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 rating variables. See Web Appendix §C.2.
3. Third, we specify a logit model of buyer-seller past interaction, as a function of observed buyer and auction attributes A_i, seller's ratings (number and mean), and geographic region. See Web Appendix §C.3.
4. Figure 4 depicts the KDE of all of the bids in the data. Note that it is lumpy and does not resemble any parametric distribution. Therefore, we use the fully nonparametric EM-like separation algorithm. We distribute all of the bids in the data based on observed buyer and seller characteristics, and derive nonparametric mixture distributions of bids. See Web Appendix §C.4.
5. Fifth, we specify a flexible nested logit model of buyer decisions. Purely nonparametric models of the CCP of buyers' decision, $\mathcal{P}(\cdot)$, are not feasible in our large state space setting with finite data. Hence, we parameterize $\mathcal{P}(\cdot)$ using a Generalized Extreme Value (GEV) distribution such that all bid options are in one nest; the cancel option is in a separate singleton nest. See Web Appendix §C.5 for details.

Figure 4 **Kernel Density Estimate of All of the Bids in the Data**



RIGHTSLINK

Downloaded from informs.org by [128.95.104.109] on 21 July 2016, at 00:33 . For personal use only, all rights reserved.

6. Sixth, we use all of the estimates from the previous steps to obtain nonparametric distributions of seller costs for the $K=3$ unobserved auction types. See Web Appendix §D for details.

While we parameterize the estimation of two seller attributes and the buyers' decision because of our large state space, the distributions of bid prices and seller costs are estimated using fully nonparametric procedures.

6.4. Results

6.4.1. First Step Results.

Estimates of 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. Because these bandwidths are not very informative, in and of themselves, we do not present them here. However, we note that the KDEs are very good at approximating the joint distributions of these two attributes.

Second, consider the results from the multinomial logit model of seller region; see Table A3 in the Web Appendix for the parameter estimates. In this model, we include all of the buyer- and auction-specific attributes, seller mean rating, number of seller ratings, and their interactions as explanatory variables. There are a few interesting points of note. First, everything else being constant, buyers in the Indian subcontinent and developed countries 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 be in the Indian subcontinent and less likely to be in developed countries. 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 subcontinent; buyers with high average ratings are more likely to attract sellers from the Indian subcontinent and less likely to attract those from developed countries. Furthermore, sellers with the lowest mean ratings are likely to be in the Indian subcontinent, followed by those from developed countries, and Eastern Europe.

Next, consider the findings from the logit model for the indicator of past buyer-seller interaction. (See Table A4 in the Web Appendix for estimates.) 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. Furthermore, buyers who have a good record on the website, e.g., 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, i.e., high number of ratings and average rating, are more likely to have interacted with the buyer. Finally, sellers from the Indian subcontinent and Eastern Europe are less likely, and those from developed countries are more likely, to bid on auctions by buyers with whom they have interacted.

Estimates of Nonparametric Mixture Distributions of Bid Prices. We consider three unobserved types of auctions $\{v_1, v_2, v_3\}$ and pool the observed state variables into 18 bins. So we recover a total of 54 bid distributions. 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 KDs, one for each type, in Figures 5–7. The complete distribution of all

Figure 5 Kernel Density Estimate of Bids from Low Type Auctions

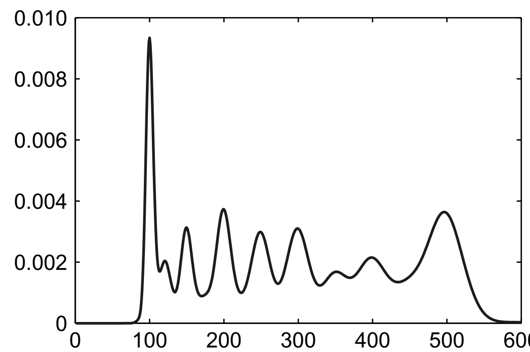


Figure 6 Kernel Density Estimate of Bids from Medium Type Auctions

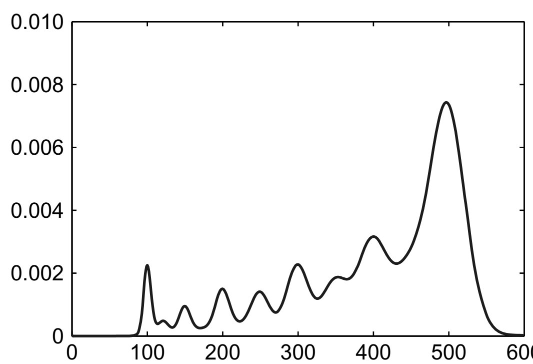
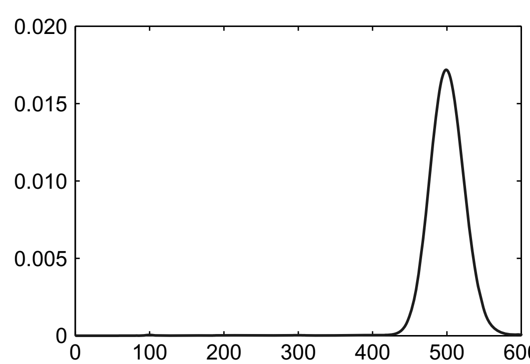


Figure 7 Kernel Density Estimate of Bids from High Type Auctions



RIGHTSLINK

Figure 8 (Color online) Cumulative Density Functions of Bid Distributions for Low, Medium, and High Types

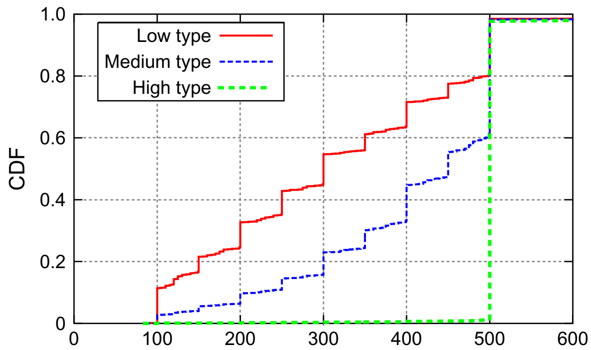
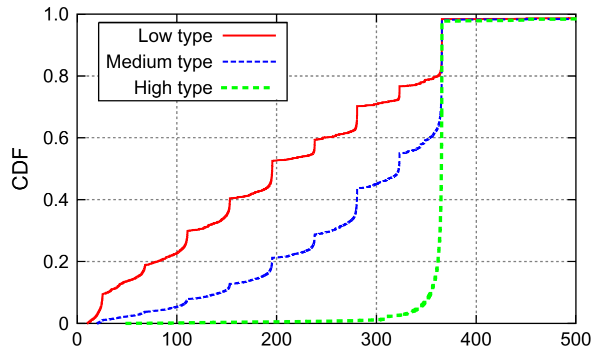


Figure 9 (Color online) Cumulative Density Functions of Estimated Costs for Low, Medium, and High Types



of the bids is shown in Figure 4 to give us a sense of how the overall distribution looks 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 auctions consist of low difficulty or low cost jobs, nearly a third (36.45%) of them consist of very high difficulty or high cost jobs, whereas the majority of the jobs (≈46%) are medium cost.

One common pattern in all three KDEs is that there are clear modalities at multiples of 50. However, the similarities end there. Note that the Low type KDE is 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 wide range of \$100 to \$500. The median of this distribution is \$300, and the 75th percentile is \$400. Less than 20% of the bids in this distribution are close to the MaxBid \$500. By contrast, for the Medium KDE, the weight is more skewed towards the right, near \$500, though there is significant 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. Furthermore, 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 type KDE, almost all of the weight is near the MaxBid \$500. These are auctions in which all bidders uniformly bid either \$500 or very close to it.

Estimates of 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 must treat them as structural parameters (see the entry model presented in Web Appendix §E and the counterfactual discussion in §6.7).[10]

We now briefly discuss the main results here (interested readers should refer to Model N1 in Table A5 in the Web Appendix for details). First, 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. However, the marginal impact of each additional rating is decreasing. Furthermore, buyers have a strong tendency to pick sellers with whom they have worked in the past. Buyers also exhibit a preference for sellers from Eastern Europe, followed by those from developed countries, and buyers prefer not to pick sellers from the Indian subcontinent. Buyers are more likely to choose a bid (as opposed to canceling the auction) if they post auctions with long deadlines and project attachments. Buyers who have few uncanceled and many successful past auctions are less likely to cancel. 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.

6.4.2. Second Step Results: Seller Cost and Market Power. Finally, we present the results on seller cost distributions. As with the bid distributions, these cost 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 §6.5.

CDFs of seller costs for Low, Medium, and High type auctions are shown in Figure 9 and their summary statistics in Table 7. The three cost distributions are significantly different. 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

[10] In fact, given that we do not specify a buyer optimization problem, a structural interpretation of these estimates is likely to be flawed. However, the inferred CCPs themselves are consistent because they capture buyers' equilibrium behavior, on which sellers base their decision. Hence, they are sufficient to infer seller costs.

Downloaded from informs.org by [128.95.104.109] on 21 July 2016, at 00:33 . For personal use only, all rights reserved.

Downloaded from informs.org by [128.95.104.109] on 21 July 2016, at 00:33 . For personal use only, all rights reserved.

Table 7 Summary Statistics of Cost Distributions, with Three, Two, and One Unobserved Types

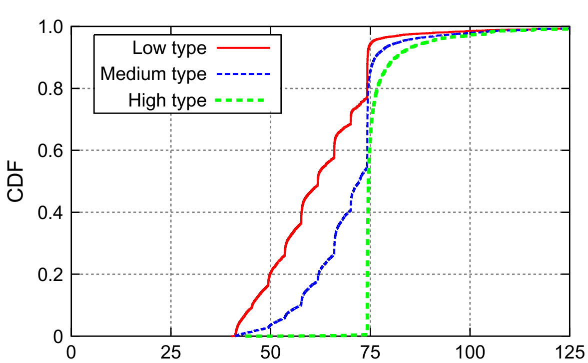
Number of types	Type	Average	25th perc.	50th perc.	75th perc.	Pop. prob. (%)
Three types	Low type	206.53	107.70	195.55	323.13	17.07
	Medium type	286.55	236.13	322.49	365.28	46.47
	High type	362.06	362.93	365.14	365.54	36.45
Two types	Low type	263.20	193.32	280.63	365.03	62.42
	High type	361.56	362.87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

is FOSD by the cost distribution of High type auctions. For Low type auctions, the costs are, on average, quite low and there is a 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, with the median cost being $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' market power and the extent of competitiveness in the marketplace. We find that seller margins are, on average, 15% of the bid. 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 keeps $15. In terms of dollar amounts, almost all of 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 suggest that this marketplace is quite competitive and that sellers do not wield much market power.

Of particular importance, the estimated margin distributions for the three types are significantly different. 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,

Figure 10 (Color online) Cumulative Density Function of Estimated Margins for Low, Medium, and High Types



and highlight the importance of allowing for nonmultiplicatively separable common shocks to seller costs in auctions.

6.5. 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 must impose some structure to further explore this issue. Hence, following Haile et al. (2003), we expand cost as

$$c_{ji} = \mathcal{T}(A_i, X_{ji}) + \sum_{k=2}^{3} \omega_k I(v_i = v^k) + \tilde{c}_{ji}, \qquad (21)$$

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. Then, the expected cost of seller j in auction i is

$$E[c_{ji}] = \mathcal{T}(A_i, X_{ji}) + \sum_{k=2}^{3} \lambda_{ik}\omega_k + \tilde{c}_{ji}, \qquad (22)$$

where λ_{ik}s are posterior probabilities of auction i belonging to type k. Based on Equation (22), we regress the expected cost on A_i, X_{ji}, and posterior-weighted unobserved types. The results are shown in Table 8.

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

RIGHTSLINK

Downloaded from informs.org by [128.95.104.109] on 21 July 2016, at 00:33 . For personal use only, all rights reserved.

Table 8 Explaining Variations in Seller

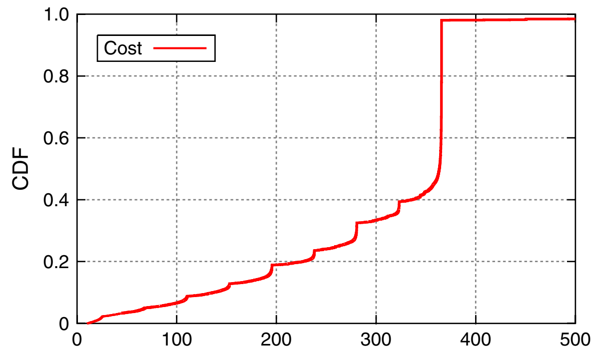
Explanatory variables	Coefficient	Std. error
Buyer region = 1	−1.377	3.633
Buyer region = 2	−2.131	2.749
Buyer region = 3	−10.97	3.447
Ind. auction attachment = 1	−7.940	1.725
ln(*Deadline_days* + 1)	1.561	0.576
Buyer's success ratio	−7.678	3.217
ln(*No. of auctions uncanceled by buyer*)	−5.469	2.121
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 between buyer and 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	44,274	

Notes. Dependent variable: $E[cost_{ji}]$. Robust standard errors shown.

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, 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 are 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 work with both due to ignorance of the landscape as well as 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

Figure 11 (Color online) Cumulative Density Function of Estimated Cost with One Type



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. The latter are more likely to win other auctions that they may bid on in the near future.[11]

6.6. Validation and Robustness Checks

6.6.1. Validation. We now compare the results from our model with two other models, i.e., 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" ($\mathcal{D}_R$).

$$\mathcal{D}_A(\mathcal{F}_a, \mathcal{F}_b) = \int_0^1 |F_a^{-1}(s) - F_b^{-1}(s)|\,ds;$$
$$\mathcal{D}_R(\mathcal{F}_a, \mathcal{F}_b) = \int_0^1 \frac{|F_a^{-1}(s) - F_b^{-1}(s)|}{F_a^{-1}(s)}\,ds, \quad (23)$$

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. Its performance is quite inferior compared to our main model with three unobserved types (see Tables 7 and 9). It significantly overpredicts seller costs for Low and Medium type auctions and underpredicts the costs for High type auctions (for more than 40% of the sellers).

Now consider a model with two unobserved types. This gives us two cost distributions, shown in Figure 12. The inferred population probabilities of the two types are 62.42% and 37.57%. Note that the population probability of the High type auctions in this

[11] 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.

RIGHTSLINK

Downloaded from informs.org by [128.95.104.109] on 21 July 2016, at 00:33 . For personal use only, all rights reserved.

Table 9 **Absolute and Relative Distance Metrics Comparing the Cost Distributions of Models One and Two Types with the Model with Three Types**

Comparing one type with three types		Comparing two types with three types	
Absolute distance	Relative distance (%)	Absolute distance	Relative distance (%)
$\mathcal{D}_A(\mathcal{F}_{3,L}, \mathcal{F}_{1,L}) = 56.67$	$\mathcal{D}_R(\mathcal{F}_{3,L}, \mathcal{F}_{1,L}) = 90.83$	$\mathcal{D}_A(\mathcal{F}_{3,L}, \mathcal{F}_{2,L}) = 56.67$	$\mathcal{D}_R(\mathcal{F}_{3,L}, \mathcal{F}_{2,L}) = 50.73$
$\mathcal{D}_A(\mathcal{F}_{3,M}, \mathcal{F}_{1,M}) = 23.34$	$\mathcal{D}_R(\mathcal{F}_{3,M}, \mathcal{F}_{1,M}) = 7.94$	$\mathcal{D}_A(\mathcal{F}_{3,M}, \mathcal{F}_{2,L}) = 23.34$	$\mathcal{D}_R(\mathcal{F}_{3,M}, \mathcal{F}_{2,L}) = 12.73$
$\mathcal{D}_A(\mathcal{F}_{3,H}, \mathcal{F}_{1,H}) = 1.07$	$\mathcal{D}_R(\mathcal{F}_{3,H}, \mathcal{F}_{1,H}) = 17.79$	$\mathcal{D}_A(\mathcal{F}_{3,H}, \mathcal{F}_{2,H}) = 1.07$	$\mathcal{D}_R(\mathcal{F}_{3,H}, \mathcal{F}_{2,H}) = 0.37$

Notes. $\mathcal{F}_{3,L}, \mathcal{F}_{3,M}, \mathcal{F}_{3,H}$ are the CDFs of cost distributions for the three types when we allow for three unobserved types. Similarly, $\mathcal{F}_{2,L}, \mathcal{F}_{2,M}, \mathcal{F}_{2,H}$ and $\mathcal{F}_{1,L}, \mathcal{F}_{1,M}, \mathcal{F}_{1,H}$ are the CDFs of cost distributions when we allow for two and one types, respectively.

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 in Table 9 and Figures 9 and 12. Thus, this model can recover the cost and population distribution of High type auctions. However, because it pools the Low and Medium type auctions into one group, it overpredicts the costs of Low type auctions and underpredicts the costs of Medium type auctions.

A natural question that arises here is whether there are more than three unobserved types in the data. So we experimented with more than three types. Recall that to recover four types, we need seven or more bids per auction (see §3.4). Because 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. Yet we could not recover a fourth type. We thus conclude that three unobserved types are sufficient for this setting.

6.6.2. Robustness Checks. We now present three empirical tests to confirm the robustness of our results.

First, we examine the impact of relaxing Assumption 6, 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. The results from this extension are very similar to those from our base model (see Model N2 in Table A5 in the Web Appendix). This suggests that identification of the unobserved types comes from the variation in bids, and not from the variation in buyer decisions.

Second, we test the validity of Assumption 7. 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 7, we only need to examine whether the fixed point from an EM-like algorithm with Assumption 7 is the same as that from an algorithm that does not make this assumption. The two fixed points can be equivalent if and only if $\mathcal{G}^X(X_{ji} \mid A_i) = \mathcal{G}^X(X_{ji} \mid A_i, v_i = k)\ \forall k$. Because we have estimated a fixed point of the system using Assumption 7, we already have an estimate of $\hat{\mathcal{G}}^X(X_{ji} \mid A_i, v_i)$. So if we can now show that

$$\hat{\mathcal{G}}^X(X_{ji} \mid A_i, v_i = Low) = \hat{\mathcal{G}}^X(X_{ji} \mid A_i, v_i = Medium) = \hat{\mathcal{G}}^X(X_{ji} \mid A_i, v_i = High) \qquad (24)$$

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

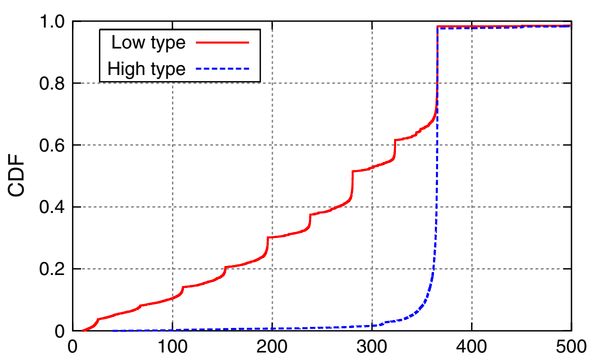
Thus using our posterior estimates, we test the distributional equality of seller attributes for the three unobserved types, and find them statistically indistinguishable; see the findings for the reputation variables, ln(*Sum of seller ratings*), in Figure A4 in the Web Appendix. The findings for other seller-specific variables are similar. These findings confirm that Assumption 7 is valid for our setting.

Finally, we allow bidders to have uncertainty on number of bids received in the auction. We model the t-th bidder's expectations on the number of bids she expects the auction to receive using a truncated Poisson model

$$\mathcal{H}(q_i \mid (q_i > t), O_i, v_i, r_i, \theta_p) = \frac{e^{\eta_i} \cdot (\eta_i)^{q_i}}{q_i!} \cdot \frac{1}{\Pr(q_i > t)}. \qquad (25)$$

Using the estimates from the Poisson model of the number of bids received by an auction (see Web

Figure 12 **(Color online) Cumulative Density Function of Estimated Costs for Two Unobserved Types**



RIGHTSLINK

Downloaded from informs.org by [128.95.104.109] on 21 July 2016, at 00:33 . For personal use only, all rights reserved.

Appendix §E), 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. We find that the cost estimates from this more complex model are similar to that from the original (see Table A7 in the Web Appendix).

6.7. Counterfactual Simulations

We now use our estimates to answer policy questions that are of importance to managers of freelance sites. 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 buyers' decision model estimates as structural parameters, estimate a model of buyer entry, 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, 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 this end, we specify and estimate a structural model of buyer entry in Web Appendix E.[12] A buyer entry model requires us to treat the estimates from the buyer's choice model on entry as structural parameters that define the utility of profit-maximizing buyers. We model buyers' beliefs on the number of bids they expect to receive using a Poisson model, and the actual entry decision using a binary logit model. Both of these models are allowed to be functions of the auction-specific unobservable and estimated within the EM-like loop. See Web Appendix E for details of these models and results.

Each counterfactual requires us to numerically re-solve for the equilibrium following a policy change. To obtain the counterfactuals, we use 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. See §F in the Web Appendix for step-by-step details.

Note that our counterfactuals should be interpreted cautiously with the necessary caveats. First, we assume that sellers face zero bidding costs, which may not always be true, in which case estimates of the Poisson bid arrival model in the entry model may not hold under counterfactual scenarios. Second, buyers and sellers may be solving a dynamic across-auction optimization problem rather than a static within-auction problem. Third, they may also be substituting across different freelance sites. All these factors can influence outcomes in the real world.

6.7.1. Who Is More Important: Buyers or Sellers? Given our two-sided market, the most important question from a manager's perspective is: Who is more important to the 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, 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 of the auctions (so each auction gets fewer bids).[13] The results from the two experiments are shown in Figures 13 and 14. In both experiments, the effect of increased supply is positive. This is expected, everything else being constant: Attracting more sellers or more buyers has a positive impact on the auction clearance rate and site revenue. However, the source of this increased revenue and its magnitude is quite different across the two cases.

Increasingly seller supply has the following effect: From the 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. Furthermore, 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 13). 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.

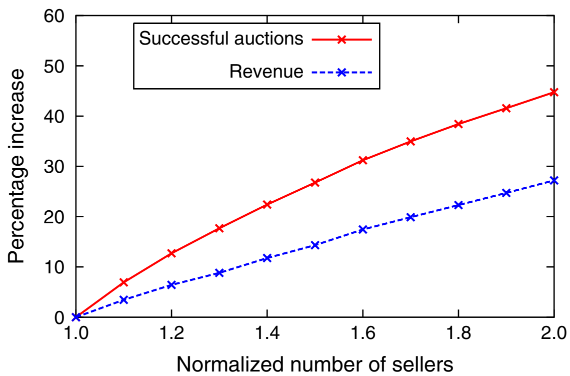
[12] Note that modeling buyer or seller entry is not essential to obtain unbiased seller cost estimates as long as the maximum bid is nonbinding and there are no acquisition costs (a natural assumption in Internet settings, since sellers do not need to spend significant resources to understand their own valuation); see Athey and Haile (2007).

[13] One can do this without decreasing the number of bids in each auction, in which case 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.

RIGHTSLINK

Downloaded from informs.org by [128.95.104.109] on 21 July 2016, at 00:33 . For personal use only, all rights reserved.

Figure 13 (Color online) Impact of Increasing Seller Supply

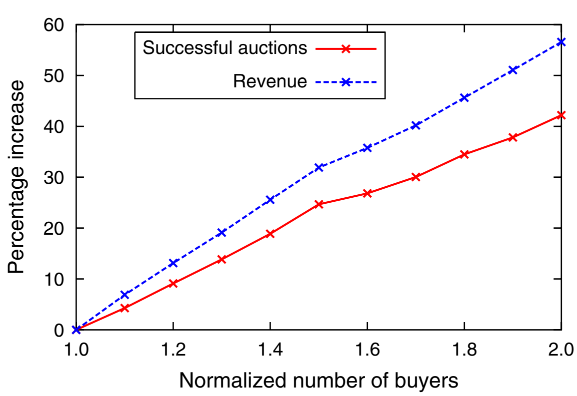


For instance, a 40% increase in the number of sellers per auction leads to a 22% increase in the number of successful auctions, but only a 5.5% increase in revenue. Thus the lower clearance prices due to the competition effect largely overwhelm the higher clearance rates due to the market expansion.

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 the 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 does not 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 14.

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, e.g., premium services for buyers, targeted advertising to attract buyers in developed countries, better monetary incentives, etc. We do not 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.

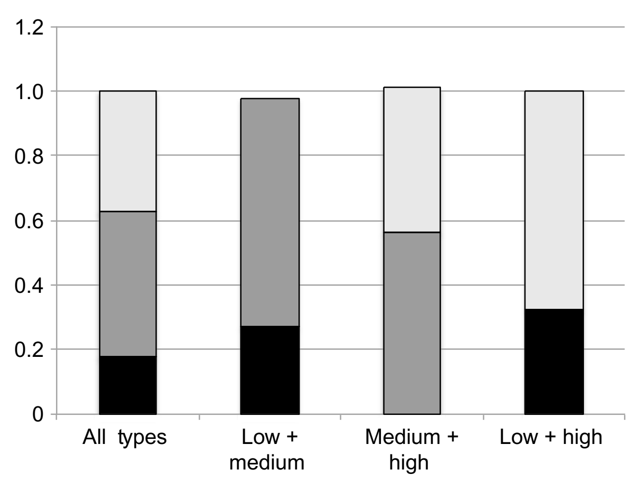
Figure 14 (Color online) Impact of Increasing Buyer Supply



6.7.2. Which Auction Type Is More Important: Low, Medium or High? In the second counterfactual, we continue to study site revenues, but focus on auction types instead of site participants. Recall that we retrieved three types of auctions, i.e., 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. Because 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 15, 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. The size of the total bar depicts the relative increase/decrease in total revenues compared to the base case.

Figure 15 Impact of Excluding Specific Auction Types



RIGHTSLINK

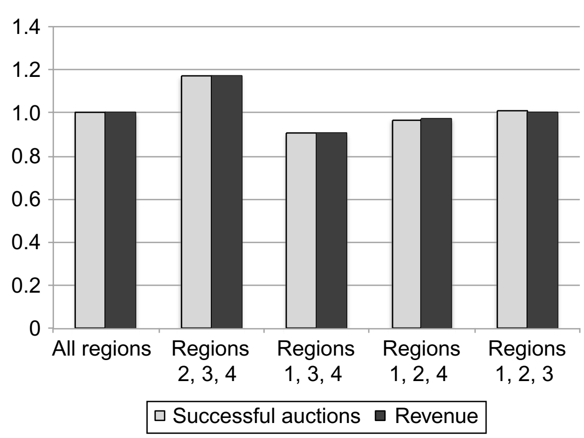
Downloaded from informs.org by [128.95.104.109] on 21 July 2016, at 00:33 . For personal use only, all rights reserved.

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. Surprisingly, when we replace High type auctions with Medium and Low types, 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 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.

6.7.3. 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. This issue is of relevance to policymakers too, since online freelancing is increasingly contributing to offshore outsourcing of jobs.

We conduct four experiments in this context. In each experiment we drop sellers from a specific region and proportionally replace them with sellers from the other three regions. 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 16. 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 wherein sellers from the Indian subcontinent (region 1) are proportionally replaced by sellers from other regions. Similar bar graphs for regions 2, 3, and 4 are shown next.

Figure 16 Impact of Excluding Sellers from a Region



Buyers' high and low preferences for sellers from developed countries (region 2) and the Indian subcontinent (region 1), respectively, have already been established (see Table A5 in the Web Appendix). 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 regions 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 softens 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 are also more likely to cancel posted auctions. If the fraction of auctions clearing falls, then, for the experiment to be profitable, this loss must be offset by the increase in commission revenues from higher prices. Thus, the overall directionality and magnitude of the effect of changing seller regions is not clear a priori.

The main finding from our experiments is that sellers from the Indian subcontinent are the least valuable to the site. 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.

7. Conclusion

Beauty contests are procurement mechanisms whereby 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 nonmultiplicatively separable unobserved auction heterogeneity. This makes their estimation challenging.

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

RIGHTSLINK

Downloaded from informs.org by [128.95.104.109] on 21 July 2016, at 00:33 . For personal use only, all rights reserved.

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 do not have prespecified 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.

Of particular importance, our method provides a clean solution to the problem of auction-specific unobserved heterogeneity. Our method does not require multiplicative separability assumptions, and 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. In the first step, we present 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. These first stage estimates are then used to derive the nonparametric distribution of sellers' private costs.

Our method is applicable to a large range of marketing problems where a decision-maker chooses from a set of interested parties without prespecifying a decision rule, especially when market-specific unobservables play a significant role in the strategic behavior of participants, e.g., online dating settings, digital advertising, real-estate bidding, hiring, and contracting scenarios.

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 and derive their population distributions. We find that the sellers in this marketplace do not enjoy high margins, with average percentage margins around 15% of the bids, i.e., the marketplace is quite competitive. We also find that not accounting for unobserved heterogeneity can significantly bias the seller cost estimates. Based on our estimates, we run counterfactuals to provide recommendations to identify the most valuable players in the marketplace. We find that even though it is a two-sided marketplace, the site should focus on growing its buyers' side (over the sellers' side) and attracting more local sellers from developed countries.

Our approach can be easily extended to allow sellers to optimize not only price but also other attributes (e.g., price and delivery time) by specifying and solving for a system of first order conditions. Future researchers may want to extend our approach in useful directions such as seller-specific unobservables visible to buyers and modeling across auction dynamics.

Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/mksc.2015.0929.

Acknowledgments

The author is grateful to an anonymous firm for providing the data. The author also thanks Bryan Bollinger, Preyas Desai, Harikesh Nair, Greg Lewis, Sanjog Misra, Sungho Park, and Scott K. Shriver for detailed comments that have improved the paper. The author also thanks the marketing seminar attendees at Columbia University, University of Alberta, University of Colorado at Boulder, University of Toronto, and Washington University at St. Louis, Information Systems and Operations Management seminar attendees at the University of Washington, and participants of the University of Texas at Dallas FORMS 2013, Marketing Dynamics Conference 2013, INFORMS Marketing Science Conference 2013, Quantitative Marketing and Economics Conference 2013, NASMEI 2013, and SMART Workshop 2014 for feedback. Please address all correspondence to hemay@uw.edu.

References

Aguirregabiria V, Mira P (2007) Sequential estimation of dynamic discrete games. *Econometrica* 75(1):1–53.
Allman ES, Matias C, Rhodes JA (2009) Identifiability of parameters in latent structure models with many observed variables. *Ann. Statist.* 37(6A):3099–3132.
Arcidiacono P, Miller RA (2011) Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica* 79(6):1823–1867.
Asker J, Cantillon E (2008) Properties of scoring auctions. *RAND J. Econom.* 39(1):69–85.
Asker J, Cantillon E (2010) Procurement when price and quality matter. *RAND J. Econom.* 41(1):1–34.
Athey S (2001) Single crossing properties and the existence of pure strategy equilibria in games of incomplete information. *Econometrica* 69(4):861–889.
Athey S, Haile P (2007) Nonparametric approaches to auctions. *Handbook Econometrics* 6:3847–3965.
Athey S, Levin J (2001) Information and competition in US forest service timber auctions. *J. Political Econom.* 109(2):375–417.
Bajari P, Hortacsu A (2004) Economic insights from Internet auctions. *J. Econom. Literature* 42(2):457–486.
Bajari P, Benkard C, Levin J (2007) Estimating dynamic models of imperfect competition. *Econometrica* 75(5):1331–1370.
Bajari P, Hong H, Krainer J, Nekipelov D (2010) Estimating static models of strategic interactions. *J. Bus. Econom. Statist.* 28(4):469–482.
Benaglia T, Chauveau D, Hunter D (2009a) An EM-like algorithm for semi and nonparametric estimation in multivariate mixtures. *J. Computational Graphical Statist.* 18(2):505–526.

RIGHTSLINK Copyright Clearance Center

Downloaded from informs.org by [128.95.104.109] on 21 July 2016, at 00:33 . For personal use only, all rights reserved.

Benaglia T, Chauveau D, Hunter DR, Young DS (2009b) Mixtools: An *r* package for analyzing mixture models. *J. Statist. Software* 32(6):1–29.

Branco F (1997) The design of multidimensional auctions. *RAND J. Econom.* 28(1):63–81.

Bruttel O (2004) Contracting-out the public employment service and the consequence for hard-to-place jobseekers: Experiences from Australia, the Netherlands and the UK. *Second Annual ESPAnet Conf.* (University of Oxford, Oxford, UK).

Cabizza M, Fraja GD (1998) Quality considerations in auctions for television franchises. *Inform. Econom. Policy* 10(1):9–22.

Campo S, Perrigne I, Vuong Q (2003) Asymmetry in first-price auctions with affiliated private values. *J. Appl. Econometrics* 18(2):179–207.

Che Y (1993) Design competition through multidimensional auctions. *RAND J. Econom.* 24(4):668–680.

Cripps MW, Mailath GJ, Samuelson L (2004) Imperfect monitoring and impermanent reputations. *Econometrica* 72(2):407–432.

Dempster AP, Laird MN, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc.: Ser. B (Statist. Methodology)* 39(1):1–22.

Draganska M, Mazzeo M, Seim K (2009) Beyond plain vanilla: Modeling joint product assortment and pricing decisions. *Quant. Marketing Econom.* 7(2):105–146.

Einav L (2003) A note on the analogies between empirical models of auctions and of differentiated product models. Working paper, Stanford University, Stanford, CA.

Ellickson P, Misra S (2012) Enriching interactions: Incorporating outcome data into static discrete games. *Quant. Marketing Econom.* 10(1):1–26.

Ellickson P, Misra S, Nair H (2012) Repositioning dynamics and pricing strategy. *J. Marketing Res.* 49(6):750–772.

Ellickson PB, Misra S (2008) Supermarket pricing strategies. *Marketing Sci.* 27(5):811–828.

Ellickson PB, Misra S (2011) Structural workshop paper—Estimating discrete games. *Marketing Sci.* 30(6):997–1010.

Elmore RT, Hettmansperger TP, Thomas H (2004) Estimating component cumulative distribution functions in finite mixture models. *Comm. Statist.—Theory Methods* 33(9):2075–2086.

Google, Inc. (2013) 2013 financial tables. (September). Accessed March 9, 2015, http://investor.google.com/financial/2013/tables.html.

Guerre E, Perrigne I, Vuong Q (2000) Optimal nonparametric estimation of first-price auctions. *Econometrica* 68(3):525–574.

Guerre E, Perrigne I, Vuong Q (2009) Nonparametric identification of risk aversion in first-price auctions under exclusion restrictions. *Econometrica* 77(4):1193–1227.

Haile P, Hong H, Shum M (2003) Nonparametric tests for common values at first-price sealed-bid auctions. Technical report, National Bureau of Economic Research, Cambridge, MA.

Hall P, Zhou X (2003) Nonparametric estimation of component distributions in a multivariate mixture. *Ann. Statist.* 31(1):201–224.

Hartmann W (2010) Demand estimation with social interactions and the implications for targeted marketing. *Marketing Sci.* 29(4):585–601.

Haruvy E, Jap SD (2013) Differentiated bidders and bidding behavior in procurement auctions. *J. Marketing Res.* 50(2):241–258.

Hettmansperger T, Thomas H (2000) Almost nonparametric inference for repeated measures in mixture models. *J. Roy. Statist. Soc.: Ser. B (Statist. Methodology)* 62(4):811–825.

Holmstrom B (1999) Managerial incentive problems: A dynamic perspective. *Rev. Econom. Stud.* 66(1):169–182.

Hong H, Shum M (2002) Increasing competition and the winner's curse: Evidence from procurement. *Rev. Econom. Stud.* 69(4):871–898.

Hotz JV, Miller RA (1993) Conditional choice probabilities and the estimation of dynamic models. *Rev. Econom. Stud.* 60(3):497–529.

Janssen MCW (2002) Auctions and beauty contests: A policy perspective. Working paper, Erasmus University of Rotterdam, Rotterdam, Netherlands.

Jap SD, Haruvy E (2008) Interorganizational relationships and bidding behavior in industrial online reverse auctions. *J. Marketing Res.* 45(5):550–561.

Kasahara H, Shimotsu K (2009) Nonparametric identification of finite mixture models of dynamic discrete choices. *Econometrica* 77(1):135–175.

Klemperer P (2000) Spectrum on the block. *Wall Street J.*, http://www.wsj.com/articles/SB970691861257795781.

Klemperer P (2002) What really matters in auction design. *J. Econom. Perspectives* 16(1):169–189.

Kozierok C (2011) Online freelancing guide. Accessed December 1, 2012, http://www.freelancingguide.net/index.htm.

Krasnokutskaya E (2011) Identification and estimation of auction models with unobserved heterogeneity. *Rev. Econom. Stud.* 78(1):293–327.

Levine M, Hunter D, Chauveau D (2011) Maximum smoothed likelihood for multivariate mixtures. *Biometrika* 98(2):403–416.

Li T, Perrigne I, Vuong Q (2002) Structural estimation of the affiliated private value auction model. *RAND J. Econom.* 33(2):171–193.

Li T, Perrigne I, Vuong Q (2003) Semiparametric estimation of the optimal reserve price in first-price auctions. *J. Bus. Econom. Statist.* 21(1):53–64.

McAdams D (2003) Isotone equilibrium in games of incomplete information. *Econometrica* 71(4):1191–1214.

Misra S, Nair H (2011) A structural model of sales-force compensation dynamics: Estimation and field implementation. *Quant. Marketing Econom.* 9(3):211–257.

Morgan M (2011) The freelance marketplace annual review, 2010. Accessed September 1, 2011, http://WhichLance.com.

Narayanan S (2013) Bayesian estimation of discrete games of complete information. *Quant. Marketing Econom.* 11(1):39–81.

Orhun YA (2013) Spatial differentiation in the supermarket industry: The role of common information. *Quant. Marketing Econom.* 11(1):3–37.

Pesendorfer M, Schmidt-Dengler P (2008) Asymptotic least squares estimators for dynamic games. *Rev. Econom. Stud.* 75(3):901–928.

Securities and Exchange Commission (SEC) (2012) Securities and exchange commission filing by Facebook Inc. Accessed March 9, 2015, http://pdf.secdatabase.com/700/0001193125-12-316895.pdf.

Shriver S (2015) Network effects in alternative fuel adoption: Empirical analysis of the market for ethanol. *Marketing Sci.* 34(1):78–97.

Singh V, Zhu T (2008) Pricing and market concentration in oligopoly markets. *Marketing Sci.* 27(6):1020–1035.

Yao S, Mela C (2008) Online auction demand. *Marketing Sci.* 27(5):861–885.

Yoganarasimhan H (2013) The value of reputation in an online freelance marketplace. *Marketing Sci.* 27(5):861–885.

RIGHTSLINK