The Role of Quality in Internet Service Markets*

Elena Krasnokutskaya†
Johns Hopkins University

Kyungchul Song‡
University of British Columbia

Xun Tang§
Rice University

March 20, 2017

Abstract

The paper proposes an empirical framework to study online procurement markets. In these markets projects are often allocated through a mechanism which allows buyers to take into account seller's non-price characteristics as well as his bid. The proposed methodology is designed to recover primitives of the environment in the presence of unobserved seller heterogeneity while accommodating two important features of such markets: buyer-specific choice sets and the high turnover of sellers.

We apply our method to the data from an online market for programming services in order to assess buyers' welfare gains associated with the globalization enabled by the Internet. We focus on the increased variety and competitive effects which arise due to the presence of low cost foreign sellers as the main welfare-improving consequences of globalization. We find that the Internet enables buyers to substantially improve on their outside (local) option, with a large part of the gains arising from access to the international markets.

Keywords: quality, services, procurement, multi-attribute auctions, unobserved agent heterogeneity, Internet

JEL Classification: C14, C18, D22, D44, D82, L15, L86.

*This version: March 20, 2017. We would like to thank Susan Athey, Phil Haile, Ken Hendricks, Ariel Pakes, Dan Ackerberg, Greg Lewis and Marc Rysman for helpful discussions. We are also grateful to seminar participants at the University of Wisconsin-Madison, Rice University, Harvard University, Columbia University, 2012 Winter Meeting of Econometrics Society, 2012 Cowles Foundation Conference for Applied Microeconomics, and 2012 Society for Economic Dynamics Annual Meeting.

†Email: ekrasno1@jhu.edu.

‡Email: kysong@mail.ubc.ca.

§Email: xun.tang@rice.edu.

1 Introduction

Until recently, markets for professional services[1] were local for all but a very few (large) buyers because the cost of searching for non-local providers, assessing their quality, and maintaining communication throughout the process was prohibitively high. The Internet facilitated the entry of intermediaries who were able to substantially mitigate such costs. The objective of this paper is to study the sources and magnitudes of the gains to the buyers, many of whom were previously confined to small local markets, from the ability to access a large globalized pool of diverse sellers online. In addition to the main substantive question, this paper delivers a number of novel insights into the operation of these fast growing yet relatively little studied markets.

The design and the data generated by online markets present a set of challenges that are not yet addressed in the literature. Methodological contribution of this paper lies in developing a tractable framework as well as identification and estimation strategies that allow us to answer the substantive question we study. Our analysis is based on data from a prominent online procurement market for programming services which is representative of many online markets that appeared during the last decade. We begin by summarizing the features of the market which are pertinent to our analysis and which distinguish this setting from other settings studied in the literature.

Transactions in this market are implemented in the form of auctions for individual projects where a buyer may either choose a seller from those who participated in the auction for his project or opt out in favor of offline opportunities.[2] The platform uses a multi-attribute auction format that allows each buyer to deviate from allocation based solely on price (as in standard auctions) and to choose instead a seller with a higher buyer-specific value. Such market design indicates potentially significant seller heterogeneity as well as the buyers' interest in having an option to differentiate among sellers on the basis of characteristics other than price. Under the multi-attribute format the weights for various seller attributes are not announced and may potentially differ from buyer to buyer. This means that the allocation rule is stochastic from sellers' point of view in contrast to standard auctions where the allocation rule is deterministic.[3]

[1] Services generate around 80% of the U.S. gross domestic product, a share that has increased by 20% over the last fifty years, with professional services accounting for half of this growth (according to Herrendorf, Rogerson, and Valentinyi (2009)).

[2] Recent analysis of eBay on-line auctions is represented by Bajari and Hortacsu (2003), Cabral and Hortacsu (2010), Ackerberg, Hirano, and Shahriar (2006), Lewis (2011) and Backus and Lewis (2012), Decarolis, Goldmanis, and Penta (2014), Athey and Nekipelov (2012), Hendricks and Sorensen (2014), Ackerberg, Hirano, and Shahriar (2014). For a survey of the older literature on on-line auctions, see Bajari and Hortacsu (2004).

[3] It is this unstructured nature of the auction format that distinguishes the service market we study from those studied in the previous auction literature, including the recent literature on "non-standard auction formats," which assumes that the decision rule is known to the bidders, e.g., standard auctions with discrimination or preferential treatment, as in Marion (2007), Krasnokutskaya and Seim (2011), and Swinkels (2009), or scoring auctions where the award is based on a rule that aggregates several bid components, as in Athey and Levin (2001), Asker and

This feature affects sellers' pricing in an important way that to the best of our knowledge has not been previously studied in the literature.

In the data, buyers frequently chose sellers that charge prices above the lowest price submitted in the auction. Additionally, a descriptive analysis that projects buyers' choices onto sellers' observable characteristics and prices reveals that buyers prefer sellers that charge higher prices, everything else equal. This suggests that some characteristics observed and valued by buyers are not recorded in the data available to the researcher. This is not surprising as the platform encourages and facilitates extensive buyer-seller communication related to sellers' qualifications and examples of their past work. Specifically, buyers may be specifically interested in certain seller characteristics such as the seller's country of origin which may indicate presence or absence of potential language barriers. However, they are likely to use most of the available seller characteristics to form an opinion about the seller's ability and, thus, to assess the quality of the product that he will deliver if chosen. The descriptive analysis mentioned above documents a positive relationship between the seller's price and the probability of winning. This indicates that the unobserved seller characteristics must be vertical, i.e., positively related to price. Thus, we refer to it as *unobserved quality* in this paper.

We formalize the features of this market in a model where each project attracts a set of sellers who submit bids for the buyer's consideration. The project is awarded to a seller who delivers the highest value over price only if it exceeds the value of the buyer's outside option. The buyer's valuation of a given seller is a function of the seller's characteristics which are evaluated on the basis of the buyer-specific private weights.[4] Our model assumes that the buyers are risk neutral and have full information on the sellers' characteristics. This assumption reflects features of many on-line settings such as the one we study. On-line platforms are often designed to minimize buyers' uncertainty about sellers' characteristics and to protect the participants from the ex post risks. In fact, the online platform we study maintains a database of performance-related measures, provides an arbitration service, and administers payments from an escrow account only after the buyer is satisfied with the delivered service. Our estimation results support this assumption as well.

The structure of the data generated by online markets presents a set of novel challenges despite being characterized by features typically arising in the discrete choice or auction settings. Specifically, while the overall number of the sellers present in this market is very large, the number of the sellers actively bidding for an individual project is relatively small. This means that a very large number of different choice sets are observed in the data and each choice set

Cantillon (2010), Asker and Cantillon (2008), and Bajari and Lewis (2011). While the multi-attribute auctions format is also prevalent in offline industry procurement, it is little studied, with the exception of Greenstein (1993, 1995).

[4]These buyer-specific weights are are not observed by sellers or the researcher.

has only a negligibly small number of buyers choosing from this same set. As a result, "market shares" of individual sellers conditional on the choice set, which are conventionally used in the analysis of discrete choice settings, cannot be precisely estimated.

The analysis is further complicated by the fact that, similar to other service markets, online markets are characterized by high turnover. A large fraction of sellers leave the market after submitting only a small number of bids and winning only one or two auctions. We refer to such sellers as *transitory* sellers as opposed to *permanent* sellers who participate in many auctions.[5] The presence of transitory sellers and their competitive pressure play an important role in online markets and in markets for services in general. For example, in our market every auction attracts several transitory sellers and projects are allocated to transitory sellers with high probability (38%), even in the presence of permanent sellers with comparable prices. Given the mechanism of buyer-seller communication it is entirely plausible that buyers are equally able to collect information about permanent and transitory sellers' quality. Thus, the challenges created by the presence of "unobserved quality" apply to transitory sellers as well.

Our approach builds on the insights from the discrete choice and empirical auction literatures. Specifically, we treat the unobserved qualities of permanent sellers as parameters of the model which may take a (relatively small) finite number of values. This naturally gives rise to a quality group structure where sellers within the same group are characterized by the same level of unobserved quality. In contrast, the transitory sellers' qualities are modeled as random variables which could be correlated with their bids and observed characteristics. Since observable seller characteristics in our setting are discrete, both permanent and transitory sellers can be summarized by their membership in one of a finite number of groups.[6] We derive pairwise inequality restrictions which link sellers' relative performance to the ordering of their qualities. This property, in turn, allows us to recover the quality group structure of the population of permanent sellers through the testing procedure which does not require knowledge of the distributions of buyers' weights or other model primitives. Empirical implementation of this step relies on the classification algorithm developed in Krasnokutskaya, Song, and Tang (2016).

The subsequent estimation step exploits the fact that once permanent sellers' group memberships are identified, we can partially characterize the buyers' choice sets in terms of the group composition of the participating permanent sellers. Our estimation procedure is based on the moments related to the permanent sellers' winning probabilities conditional on the choice sets characterized this way. We use this procedure to recover the quality levels associated with various quality groups as well as the distributions of buyers' weights and outside options. Finally, we rely

[5] In the auction literature transitory sellers are sometimes referred to as "fringe" sellers.

[6] If an observable characteristic is continuous, our approach would require discretizing it. Such discretization is adopted in Chiappori and Salanie (2001) as well as in Ciliberto and Tamer (2009).

on the structure of the sellers' pricing problem in the auction setting to recover the distribution of sellers' costs conditional on all characteristics (observable and unobservable).[7][8]

From our analysis, we find that seller heterogeneity is important in the market for programming services. While buyers are willing to pay a substantial premium to sellers who are from certain countries or have a high level of performance measures, the premiums explained by the variations in observable characteristics are relatively small compared to the 50% of the project value premium that an average buyer is willing to pay for the increase in unobserved quality from the lowest possible to the highest possible level. Our estimates reveal substantial heterogeneity in the unobserved seller quality within the groups of sellers with the similar observable characterisitcs, as well as significant differences in the distributions of the quality across different groups of country affiliation or performance measures.[9]

We use the estimated parameters to evaluate the buyers' welfare gains from the availability of the online market. Since the buyer's outside option includes hiring from the offline local market, the difference in the net values from hiring in this market over that from the outside option provides a lower bound on the gains from market globalization. We estimate this lower bound on the average gain over a buyer's local option to be 73% of the project value. This number reflects the gain in utility from access to a more diverse set of sellers (both in terms of quality and in terms of costs). We further inquire into the source of the gains by examining the effect of having access to international sellers facilitated by the Internet. In this analysis, we limit the diversity of choices available to buyers by replacing foreign sellers with US sellers of similar quality rank. Interestingly, this change impacts the market mainly by affecting the sellers' participation. US sellers present in this market, who are revealed in estimation to be weak competitors (they have higher costs and lower quality levels), participate at substantially lower rates relative to foreign sellers. As a result, under a counterfactual scenario a buyer faces a reduced set of alternatives which, given the higher realized prices, are less attractive to him. In the end, despite the number of *potential* bidders remaining the same, the reduction in the variety of potential bidders leads

[7]We accomplish this by relying on the inversion method first proposed by Guerre, Perrigne, and Vuong (2000) and later applied in various environments by Li, Perrigne, and Vuong (2000), Jofre-Bonet and Pesendorfer (2003), Li, Perrigne, and Vuong (2002), Krasnokutskaya (2011), Athey and Haile (2002) and others.

[8]It is worth noting that this approach enables us to estimate the model's primitives jointly with the mixture components which capture the relationship between the transitory sellers' bids and their qualities. This offers a viable alternative to recovering this relationship from the model within the estimation routine which would be computationally infeasible in our environment, or to estimating some ad hoc parametric mixing distribution (reflecting such dependence) jointly with other elements of the model (which has been shown to perform poorly by Heckman and Singer (1984) in the context of duration models).

[9]The estimation results confirm our surmise that the buyers are informed about the transitory sellers' qualities. For details see section 6. Our results are consistent with findings in Cabral and Hortacsu (2010), who find that performance measures collected by the e-Bay platform were likely to serve an enforcing rather than an informative role and with Lewis (2011) who finds that the e-Bay auto market is able to deliver sufficient information about used product properties to buyers so as to overcome the "lemons problem."

to the 32% decline in gains from the Internet market.

The paper is organized as follows. The descriptive statistics are reported in Section 2; and the basic model is summarized in Section 3. Sections 4 and 5 discuss our empirical methodology. Section 6 reports the results of the empirical analysis whereas Section 7 describes the analysis of counterfactual settings. Section 8 summarizes the findings and outlines directions for further research.

2 Data Description

We have access to the data for an online procurement market for computer programming services. The data cover the first six years of markets operation and include information on more than 600,000 projects that attract participation from close to 50,000 different sellers. For every project, we observe the description of work required, the size of the project as assessed by the platform, the deadline for the completion when it is imposed, and the location of the buyer. We also observe all bids submitted, characteristics of sellers who submitted bids, the identity of the winner, and measures of the winner's subsequent performance.

The projects fall into several broad classes, such as system-based programming, databases, graphics programming and website design. The work is then further divided into finer categories within these classes. For example, one of the recurrent requirements is the specification that a particular programming language should be used. We focus on the projects requesting graphics-related programming which tend to be relatively homogeneous. In addition, the sellers participating in other segments of the market rarely submit bids for this type of projects due to the high degree of specialization required in this area of work. We further restrict our attention to US-based buyers who submit an overwhelming majority of projects in our dataset.

Outsourcing. The market we study is representative of the recent trend often referred to as '1099-economy' where self-employed individuals contract for jobs through on-line platforms. Such markets, especially those associated with business services, are often international in their nature. Specifically, buyers participating in the market we study choose among sellers from a diverse set of countries. Table 1 lists seller countries with the largest presence in this market as well as reports seller presence by region. As the table indicates, the majority of bids are submitted by sellers from North America (16%), Eastern Europe (14%) and South and East Asia (48%). Respectively, a large fraction of the US projects are allocated to sellers from foreign countries: 32% to South and East Asia and 19% to Eastern European sellers. This market, therefore, provides the US buyers with an opportunity to acquire services of foreign sellers. In this analysis we investigate whether foreign sellers differ from those located in the US and, thus, whether buyers gain from being able to access these additional varieties of sellers through online market.

Table 1: Sellers' Composition by Country

Seller Country	Participation (share of bids)	Allocation (share of projects)
Major participants:		
US	0.132	0.205
Canada	0.025	0.044
UK	0.028	0.033
India	0.299	0.210
Pakistan	0.135	0.084
Bangladesh	0.024	0.017
Phillipines	0.024	0.013
Romania	0.096	0.121
Russia	0.018	0.025
Ukraine	0.022	0.029
By region:		
North America	0.157	0.249
South and East Asia	0.482	0.323
Eastern Europe	0.135	0.185
Number of Projects	24,116	
Number of Bids	128,580	

The entries in this table are based on a sample of projects with graphics-related programming posted by US buyers.

Project-Level Statistics. Table 2 provides some project-level statistics for the period covered by our data. Each row of the table summarizes a marginal distribution of the corresponding variable. Table 2 shows that a sizable number of the projects are very small (below $150).[10] On the other hand, some of the projects are quite big (above $875). The projects are fairly short: the deadline for the majority of the projects is between one to three weeks.

The majority of buyers in our data are one-time participants. Less than 2% of buyers post multiple projects. In addition, returning buyers do not post the same type of projects. As a result, a buyer very rarely works with the same seller repeatedly.

Sellers Characteristics. Table 3 describes sellers present in the graphics segment of online market. As we emphasize in the introduction, our market attracts a large number of short-lived sellers. We define a seller's tenure as the length of time that elapses between his first and his last posting. In our market 65% of the seller population (not in the table) has a tenure under three

[10] We have access to a variable reflecting project size as it is assessed by the platform. We have also constructed additional measure of project size using assessments of two independent specialists. This measurement is highly correlated with the measure provided by the platform ($\rho = 0.93$). The statistic in Table 2 is based on the variable provided by the platform.

Table 2: Data Summary Statistics

	Mean	Std.Dev.	25%	50%	90%
Number of projects (per buyer)	1.05	0.02	1	1	1
Project Characteristics					
Size	$525	$221	$150	$500	$875
Duration (days)	12	11	5	10	21
Number of Projects	24,116				

The results in this table are based on a sample of projects with graphics-related programming posted by the US buyers. Duration of project is measured in days. Each row summarizes the inverse cumulative function of the corresponding variable.

Table 3: Sellers' Characteristics

	Mean	Std.Dev.	25%	50%	75%	90%
Tenure (weeks)	23.66	54.42	1	1	11	144
Permanent Sellers						
Number of Projects Won	45.8	69	32	107	154	250
Average Score	9.8	0.061	9.7	9.87	10	10
Disputes	0.021	0.082	0	0	0	1
Delays	0.075	0.096	0	0	0	1
Fraction of Bids Resulting in Winning	0.11	0.09	0.045	0.082	0.143	0.219
Number of Bids Before First Success	11.3	6.7	5	9	17	42
Transitory Sellers						
Number of Projects Won	0.51	1.36	0	1	1	2
Average Score	9.58	0.09	9.5	9.78	10	10
Fraction of Bids Resulting in Winning*	0.12	0.13	0.035	0.076	0.151	0.25
Number of Bids Before First Success*	8.5	6.3	3	7	15	36
Number of Bids Before First Success	4.8	3.4	1	2	3	12
Number of Bids	128,580					
Number of Projects	24,116					

The results in this table are based on a sample of projects with graphics-related programming posted by US buyers. Each row summarizes the inverse cumulative function of the corresponding variable. Tenure is defined as the number of weeks between the last and the first posting of a given sellers. Disputes and delays variables reflect the number of disputes mediated by the platform and the number of missed deadlines reported to the platform respectively. For transitory sellers, 'Fraction of Bids Resulting in Winning*' and 'Number of Bids before First Success*' are computed conditional on winning at least one bid. 'Number of Bids before First Success' reports the overall number of bids a transitory sellers submitted before he wins his first project or exits the market.

weeks and 75% of the seller population has a tenure under eleven weeks (slightly less than three months). On the other hand, 10% of the seller population remains active for more than 144 weeks (or more than two years). The share of sellers with short tenure is larger in the beginning years but settles down, so that the distribution of tenure is almost constant over the last three years in the sample period. In these years, 10% of the sellers stayed in the market for more than three years, whereas 85% of the sellers left the market in less than three months.[11] Substantial seller turnover is an important feature of our market as well as of many other markets for services.

In the subsequent analysis we treat a bidder as permanent if he appears in our data for more than 6 months. Otherwise, we label a bidder as transitory. In our sample, all permanent sellers complete more than twenty projects with the mean equal to 46 projects and the median to 47. The platform records the history of sellers' performance, i.e., the instances of delays and disputes, as well as buyers' feedback about working with a given seller in the form of numerical reputation scores or ratings. The distribution of reputations scores for permanent sellers is quite tight with the mean score equal to 9.8 and the standard deviation of 0.06. Disputes and delays are rare, involving around 2% and 8% of permanent sellers respectively.

Roughly, half of transitory sellers ever registered with the platform leave the market before winning a single project but the other half completes at least one project. Overall, due to their significant presence in the market, transitory sellers win 38% of all projects (see Table 4). Further, transitory sellers' performance on the platform is comparable to that of the permanent sellers. Indeed, the distribution of average reputation scores in the population of transitory sellers who have completed at least one project is very similar to the distribution of average reputation scores in the population of permanent sellers. Additionally, the ratio of the number of bids to the number of projects won is very similar between these two populations. On the basis of these observations, it appears likely that transitory sellers are very similar to permanent sellers in their characteristics.[12]

Finally, very little information about transitory sellers is publicly available. Indeed, public information is released when a seller completes a project, and transitory sellers usually complete one or two projects and leave the market. It is plausible, therefore, that competing sellers are not informed about transitory sellers' qualities. The situation is different for permanent sellers since the market may infer their quality from the long-run rate of their successes.

[11]Some of the sellers who entered the market during the last year of our estimation sample might be permanent sellers but can not be identified as such due to their short tenure. This introduces potential right-censoring into our definition of the set of permanent sellers. This groups of sellers is rather small: much less than 1% of the total number of sellers. Further, the issue of right-censoring does not impact our estimation results due to the design of our empirical methodology.

[12]Further evidence supporting the conclusion above concerns the number of bids a seller submits before he wins his first project. As Table 3 indicates the distribution of this statistic in the population of transitory sellers who have completed at least one project is very similar to analogous distribution in the population of permanent sellers. At the same time the distribution of the number of bids before the first success in the full population of transitory bidders is shifted to the left relative to the distribution for the permanent sellers. This indicates that transitory sellers tend not to wait long enough for their first project. On the basis of these observations, it appears likely that transitory sellers are quite similar to permanent sellers in their characteristics but may have better outside opportunities.

Details of Allocation Process. The platform keeps daily record of the sellers who visit its website (and, specifically, the page with the links to the posted projects). We use this information in conjunction with sellers' history of bids to define the set of potential bidders for the set of graphics-related projects. Specifically, we assume that the set of potential bidders for project l auctioned in week t consists of all the sellers who submit at least one message for projects of the same type of work as project l during this week. This definition ensures that included sellers (a) might reasonably be expected to compete with each other in the auction for a given project; (b) are aware of each other's presence in the market during the auction.

Table 4 summarizes the number of potential as well as actual bidders at the project level. It indicates that the average number of permanent potential bidders is around eight, and the average number of transitory potential bidders is twenty three. The table further shows that only a fraction of potential sellers submits a bid. Specifically, on average only two permanent sellers and three transitory sellers from the set of potential bidders participate in an auction.

In this market, a prospective seller does not observe the set of his competitors. Instead, his price quote is likely to be based on the information about the set of potential rather than actual competitors. Table 4 describes the distribution of bids (normalized by the project size) submitted by participating sellers. As the statistics reported in the table indicates, the distributions of bids submitted by permanent and transitory sellers are very similar.

Determinants of Buyers' Choices. Table 4 further documents that the multi-attribute feature of the allocation mechanism is strongly supported in the data. Indeed, in our sample, 72% of the projects are awarded to a seller who quotes a price above the lowest price submitted in the auction. When such a seller is chosen, his price on average exceeds the lowest price submitted in the auction by 56%. These results indicate that buyers consider seller characteristics other than price when choosing a winner. Thus, a model that takes sellers' heterogeneity into account is required to study this environment. The "Money-Left-on-the-Table" measure, which is computed as the average difference between the second lowest and the lowest bid submitted in the auction divided by the lowest bid, is equal to 0.33 in our sample. In the context of standard auction markets this statistic is often interpreted as indicative of the presence of private information about sellers' costs. The interpretation is less straightforward in the multi-attribute setting. Indeed, the difference between the lowest and the second to the lowest bid may arise due to the difference in private costs but also due to the premium charged by sellers for the characteristics preferred by buyers.

We use a multinomial logit model to further explore how buyers' choices are influenced by sellers' characteristics. Here we assume that the award decision ($Y_{j,l} \in \{0, 1\}$) depends on the buyer's net value from a specific alternative (seller), $Y^*_{j,l}$, which is modeled as a linear function of

Table 4: Allocation Process

	Mean	Std.Dev.
Participation (at the project level):		
Potential Bidders:		
Permanent sellers	8.58	4.01
Transitory sellers	22.71	4.79
Actual Bidders:		
Permanent sellers	1.87	1.23
Transitory sellers	3.32	1.45
Bids (normalized by project size):		
Permanent sellers	1.55	0.21
Transitory sellers	1.63	0.24
Buyers' Choice:		
Awarded to Lowest Bid	0.28	0.20
(Winning Bid-Lowest Bid)/Lowest Bid	0.56	0.29
Money-Left-on-the-Table	0.33	0.26
Share of Projects Won:		
Permanent sellers	0.59	
Transitory sellers	0.38	
Number of Projects	24,116	

The results in this table are based on a sample of projects with graphics-related programming posted by the US buyers. Standard deviations are reported in parenthesis. 'Money-Left-on-the-Table' variable is computed as the difference between the second lowest bid and the lowest bid divided by the lowest bid.

seller characteristics, $X_{p,j,l}$ (the number of ratings, delays, disputes and the average reputation score), seller location dummies, $\mu_{c(j)}$, and a seller's bid, $B_{j,l}$:

$$Y^*_{j,l} = \sum_p X_{p,j,l}\beta_p + \gamma B_{j,l} + \mu_{c(j)} + \epsilon_{j,l}. \qquad (1)$$

The project is awarded to bidder j, $Y_{j,l} = 1$, if and only if $Y^*_{j,l} \geq 0$ and $Y^*_{j,l} \geq Y^*_{i,l}$ for all $i \neq j$ who are present in the auction; $Y_{j,l}$ is equal to zero otherwise.[13] The results of this analysis are reported in Table 5.

We estimate the price coefficient to be positive and statistically significant. This result sug-

[13]This analysis approximates the partially linear specification:

$$Y^*_{j,l} = m(X_{j,l}, \mu_{c(j)}) + \gamma B_{j,l} + \epsilon_{j,l}.$$

Specifically, we control for the effect of the number of scores by considering several intervals for the number of scores. The dependence of the probability of winning on this factor is therefore permitted to be quite non-linear. The impact of an average score is also allowed to vary across different values of the number of scores which permits this dependence to be very nonlinear. All the other variables are dummies. We have considered specification where more terms which interacted these dummies with other variables were included. Such terms, however, were not statistically significant. In the interest of precision we did not include them in the final specification.

Table 5: Multinomial Logit

Variables	Estimates	Std.Errors
Constant	1.069	4.461
Number of Scores (NS):		
$1 \leq$NS≤ 3	1.331 ***	0.171
$3 <$NS≤ 6	1.496 ***	0.368
$6 <$NS≤ 12	1.511 ***	0.408
NS≥ 12	1.513 ***	0.517
Average Score:		
Average Score if NS≤ 3	0.016	0.041
Average Score if $3 <$NS≤ 6	0.048	0.075
Average Score if $6 <$NS≤ 12	0.058	0.094
Average Score if NS≥ 12	0.158	0.119
Disputes	-0.541***	0.064
Delays	0.232	0.164
Price	1.937 ***	0.073
Seller Country Dummies	Yes	
Number of Projects	24,116	
Number of Bids	128,580	

The results in this table are based on a sample of projects with graphics-related programming posted by US buyers. Country dummies for the seller countries with bid share exceeding 0.05% are included. The omitted category is that of the US sellers who have not completed any projects yet. Above '***' indicates significance at 1% level.

gests an omitted variable bias since, in most markets, buyers prefer to pay less, other things equal. This means that some additional characteristic, not recorded in the data, affects buyers' choice in conjunction with the price, location and performance measures. Such an omitted variable should be positively aligned with the price and is, therefore, some vertical characteristic such as quality. Thus a model that rationalizes this pattern should allow for unobserved quality-like sellers' attributes.

As for the other variables, the results indicate that having earned at least a few reputation scores increases the probability of winning in a statistically significant way. After that, however, the impact of the subsequent scores flattens out. Similarly, the impact of an extra point in the average reputation score does not appear to matter in a statistically significant way. Disputes (arbitrations) decrease future probabilities of winning whereas delays have no statistically significant effect.

To summarize, the preliminary analysis of our data indicates that (a) the buyers' utility from hiring a given seller should non-trivially depend on sellers' attributes; (b) it is important to account for the presence of a large number of transitory sellers; (c) the model should allow for the presence of an unobserved quality-like sellers' attribute for permanent as well as for transitory sellers.

3 Model

Let $\bar{N}$ denote the set of sellers who operate in an on-line programming market. Each seller $j \in \bar{N}$ is characterized by a vector of attributes $x_j \in \mathcal{X} \equiv \{\bar{x}_1, \bar{x}_2, ..., \bar{x}_M\}$ and a quality index q_j which takes values from a discrete set $\mathcal{Q}(x_j) = \{\bar{q}_1(x_j), ..., \bar{q}_{K(x_j)}(x_j)\}$. Notice that distribution of quality indexes is x-specific in the sense that both the number of quality levels, $K(x_j)$, and the quality levels themselves may depend on x.

Seller j additionally belongs to one of two types $\rho_j \in \{p, t\}$ where "p" denotes permanent and "t" denotes transitory sellers. A permanent seller's quality is known to all market participants; a transitory seller's quality is his private information which is drawn from a distribution with the discrete support $\mathcal{Q}(x)$ for a seller characterized by a vector of attributes x. Each transitory seller reports his quality draw to the buyer when submitting a bid.

A buyer l seeks to procure services for a single indivisible project using a multi-attribute auction. Throughout the paper we use l to index a buyer or his project. Each project is associated with a set $N_l \subset \bar{N}$ of potential bidders who are available and interested in providing service. Each seller j in N_l decides whether to participate in the auction for project l and what bid $(B_{j,l})$ to submit if he participates.

Buyers' Choice. Let $A_l \subset N_l$ denote the set of sellers who participate in bidding for project l. The buyer either chooses a seller from A_l, or opts for an outside option that gives him a payoff $U_{0,l}$. The buyer is informed about the attributes x_j and the quality index q_j for each seller $j \in A_l$. The buyer's payoff from choosing j as the service provider is

$$U_{j,l} = \alpha_l q_j + x_j \beta_l + \epsilon_{j,l} - B_{j,l},$$

where the random vector $\gamma_l = (\alpha_l, \beta_l, \epsilon_l)$ with $\epsilon_l \equiv \{\epsilon_{j,l} : j \in N_l\}$ denotes the buyer's utility weights (or tastes) for seller characteristics. The vector γ_l and the outside option $U_{0,l}$ summarize the buyer's preferences. The buyer chooses an option that maximizes his payoff. In keeping with the definition of a multi-attribute auction, sellers do not observe the utility weights or the outside option of a specific buyer, and consider these to be random draws from the corresponding distribution.

Sellers' Strategies. We use I_{N_l} to denote the composition of the set N_l with respect to sellers' quality and attributes based on information that is common knowledge among sellers.[14] We will

[14]Specifically, I_{N_l} summarizes the set N_l by the number of permanent sellers in each of the (x, q)-group and

formally define I_{N_l} in Section 4.2.2.

All potential bidders for project l know I_{N_l}. Each seller $j \in N_l$ draws a private entry cost $E_{j,l}$ from a distribution $F_{E|x,q}$ where $x_j = x$, $q_j = q$, and decides whether to participate in bidding or not based on $E_{j,l}$ and I_{N_l}.[15] If the seller decides to participate, he pays $E_{j,l}$, draws a private cost for completing the project $C_{j,l}$ from a distribution $F_{C|x,q}$ and submits a bid $B_{j,l}$ to the buyer. We assume that private costs are independent across sellers. Sellers who do not participate obtain zero payoffs. Participation decisions of sellers are not observed by their competitors. This means that sellers do not acquire any new information about the composition of the set of their competitors between the entry and bidding stage.

Equilibrium. Let $\mathcal{I}$ denote the collection of all possible compositions of subsets of sellers in $\bar{N}$. For seller j, his *pure strategy* consists of two functions: an entry strategy $\tau_j : \mathcal{Q}_j \times \mathbb{R}_+ \times \mathcal{I} \to \{0, 1\}$, and a bidding strategy $\sigma_j : \mathcal{Q}_j \times \mathbb{R}_+ \times \mathcal{I} \to \mathbb{R}_+$, where $\mathcal{Q}_j$ is shorthand for the support of quality for a seller with characteristics x_j, or $\mathcal{Q}(x_j)$. Specifically, the entry strategy maps his realization of quality, entry costs, and the composition of the set of potential bidders for auction l, $I_{N_l} \in \mathcal{I}$, into participation decision $\{0, 1\}$ with 1 denoting the decision to participate; whereas the bidding strategy maps his realization of project costs and the composition of the set of potential bidders into the seller's bid. Let (τ_{-j}, σ_{-j}) denote a strategy profile of the sellers in N_l other than j. For any given composition $I_{N_l} \in \mathcal{I}$ of sellers in N_l, let $A_{j,l}(I_{N_l}; \tau_{-j})$ be the set of sellers *other than* j who decide to participate given their entry costs and strategies τ_{-j}. Note that by construction $A_{j,l}$ is a random set that depends on $E_{-j,l}$, which we suppress in notation for simplicity.

Given a composition $I_{N_l} \in \mathcal{I}$ of potential bidders in auction l, the ex ante payoff for a seller j who decides to participate, draws a project cost c and quotes b when other sellers use strategies (τ_{-j}, σ_{-j}) is:

$$\Pi_j(b, c; I_{N_l}; \tau_{-j}, \sigma_{-j}) = (b - c) P_j(b; I_{N_l}; \tau_{-j}, \sigma_{-j}) \tag{2}$$

where $P_j(b; I_{N_l}; \tau_{-j}, \sigma_{-j})$ denotes the conditional probability that bidder j wins the auction and is equal to

$$\Pr\left(\max\left(U_{0,l}, \max_{i \in A_{j,l}(I_{N_l};\tau_{-j})} \{\alpha_l q_i + x_i \beta_l + \epsilon_{i,l} - \sigma_i(C_{i,l}, I_{N_l})\}\right) \le \alpha_l q_j + x_j \beta_l + \epsilon_{j,l} - b\right). \tag{3}$$

The probability in (3) is with respect to the joint distribution of $(\alpha_l, \beta_l, \epsilon_l, U_{0,l})$, $C_{-j,l}$ and $E_{-j,l}$.

We follow the convention in the literature and focus on type-specific equilibria where sellers of the same type $\theta \equiv (\rho, x, q)$ with $\rho \in \{p, t\}$, use the same strategy. A *type-symmetric pure-strategy Bayesian Nash equilibrium (p.s.BNE)* is a profile of strategies (τ^*, σ^*) such that for any

by the number of transitory sellers in each x-group.

[15] Note that a transitory seller observes his quality realization prior to making entry decision. That is why he also draws his costs realizations from the (x, q)-group specific distribution.

j with type $\theta_j = \theta$, $\tau_j^* = \tau_\theta$ and $\sigma_j^* = \sigma_\theta$, where for each c and e,

$$\sigma_\theta(c, I_{N_l}) = \arg\max_{b \geq 0} (b - c) P_j(b; I_{N_l}; \tau_{-j}^*, \sigma_{-j}^*) \text{ and}$$
$$\tau_\theta(e, I_{N_l}) = 1\{e \leq \mathbf{E}\left[\Pi_j\left(\sigma_\theta(C_{j,l}, I_{N_l}), C_{j,l}; I_{N_l}; \tau_{-j}^*, \sigma_{-j}^*\right)\right]\}.$$

Here $\mathbf{E}\left[\Pi_j\left(\sigma_\theta(C_{j,l}, I_{N_l}), C_{j,l}; I_{N_l}; \tau_{-j}^*, \sigma_{-j}^*\right)\right]$ summarizes the expected profit of seller j conditional on participation where expectation is further taken with respect to the distribution of own project costs, $C_{j,l}$. Notice that the strategy of a transitory seller depends on the realization of his quality Q_j since it is realized before the seller decides on his strategy and is observable to the buyer of the project where he submits his bid. Further, the strategy used by a transitory seller of type (x, q) differs from the strategy used by a permanent seller of the same type. This because the quality of a permanent seller is observable to his competitors whereas the quality of a transitory seller is not.

Discussion. Notice that the multi-attribute environment substantially differs from a scoring auction, another mechanism which allows buyers to take into account seller's attributes other than price at the allocation stage. First, in our case, a seller's quality q_i and characteristics x_i are exogenously given as opposed to being part of the sellers' strategic choices as in scoring auctions. Second, the allocation rule in scoring auctions is explicitly specified before bidding begins whereas buyer preferences for quality and seller attributes are not known to sellers in a multi-attribute auction. Thus in the multi-attribute setting the allocation rule is stochastic from a seller's point of view. The latter property has important implications for seller strategies, which to the best of our knowledge has not been studied in the empirical literature.

4 Identification

In this section we discuss how the primitives can be recovered from the data. Suppose that data contain information on many auctions that proceed as in the model described above. For each auction, the researcher observes the set of potential and actual bidders, submitted prices, and the buyer's choice. For every seller the researcher observes the seller's x-attributes but not his quality. Furthermore, the researcher observes whether each seller is permanent or transitory.

The model primitives to be identified include: (a) the quality of each permanent seller; (b) the distribution of a transitory sellers' qualities given observable characteristics; (c) the distribution of buyers' utility weights $\alpha_l, \beta_l, \epsilon_l, U_{0,l}$; and (d) the distribution of sellers' participation and project costs given sellers' observable and unobservable characteristics.

Recovering the primitives of the model from the available data is quite challenging. To see this let us first consider an environment without transitory sellers. In this setting we only need to focus on recovering the quality levels of permanent sellers. Recall that in a traditional discrete choice setting unobserved heterogeneity associated with different alternatives are identified from the observed probabilities that a given alternative is chosen conditional on the choice set. In our setting choice sets are buyer-specific since sellers' participation varies across auctions. Due to the large number of sellers, conditional choice probabilities cannot be precisely estimated. To get a sense of magnitudes consider that the number of permanent sellers present in the market for a given type of work is around 300 to 500 whereas only 2 or 3 permanent sellers participate in any given auction. This means that the number of possible choice sets is at least $C^3_{300} = \frac{300!}{3!297!} = 8,910,200$ which exceeds the number of projects we have in our dataset. In fact, the *highest* number of projects sharing the same set of participating *permanent* sellers in our data is five. One way to deal with this issue would be to consider probabilities that aggregate over buyers' choice sets, such as:

$$\Pr(j \text{ wins}|j \in A_l) = \textstyle\sum_{a:j\in a} \Pr(j \text{ wins}|A_l = a)\Pr(A_l = a, j \in A_l), \text{ and}$$
$$\Pr(j \text{ wins}|B_l = b, j \in A_l) = \textstyle\sum_{a:j\in a} \Pr(j \text{ wins}|B_l = b, A_l = a)\Pr(A_l = a, j \in A_l),$$

where the sum above is over the choice sets a that contain j. While such aggregation is appealing, it is far from obvious that these moments could be used to identify seller-specific fixed effects and the distribution of buyers' tastes. Specifically, the invertibility argument underlying the standard approach (the most well-known exposition can be found in Berry, Levinsohn, and Pakes (1995)) does not apply to these moments because the probability of observing a given choice set, $\Pr(A_l = a|j \in A_l)$, depends on the unobserved qualities of potential (and, in the consequence, actual) sellers. So further insight is necessary on how to achieve invertibility in this context.[16] It also has to be established that such moments allow us to exploit the exogenous variation present in the data to recover the distribution of buyers tastes. Further, even if this mechanism would work in theory, it is not certain that it would perform well in practice given that the weighting probabilities used in aggregation above, $\Pr(A_l = a|j \in A_l)$, are very small. It might be preferable to consider an entirely different basis for aggregation which would maximize the performance of the estimator given the available data structure.

Let us now return to the realistic setting where transitory sellers differ in their qualities and these qualities are observable to buyers. Note that transitory sellers are an important part of this market since in the data every project attracts several transitory sellers, and has a 38% chance of being allocated to a transitory seller. Buyers often chose transitory sellers over permanent ones even when the prices are comparable.

[16]For example, it is possible that the inversion could be made to work if we use empirical probabilities of observing different choice sets in the expression on the right-hand side.

In contrast to permanent sellers we cannot use transitory sellers' identities as proxies for their quality. Thus, the information which underlies buyer's choice is not observed in the data. Instead, a researcher has to deal with a mixture problem where the probability distribution over the transitory sellers' qualities depends on these sellers' bids and observable attributes. More specifically, suppose the support of a transitory seller h's quality $Q_{h,l}$ is $\{\bar{q}_1, \bar{q}_2\}$ and let x and b be the vectors of observable attributes and bids characterizing the entrants in the auction respectively. Then, the probability that the buyer chooses a permanent seller j while his choice set includes a single transitory seller h, $\Pr(j \text{ wins}|B_l = b, x)$, is a mixture of the following form:

$$\sum\nolimits_{s=1,2} \Pr(j \text{ wins}|Q_{h,l} = \bar{q}_s, B_l = b, x) \Pr(Q_{h,l} = \bar{q}_s|B_l = b, x).$$

The mixing weights $\Pr(Q_{h,l} = \bar{q}_s|B_l = b, x)$ are unknown and correlated with the conditional choice probability through the bid vector b and attributes vector x. In practice, even writing the choice probability in this mixture form is not straightforward, because we do not observe the support of $Q_{h,l}$.

One might attempt to deal with this problem by solving for mixing probabilities from the model within the estimation routine. However, solving one such bidding and participation game is computationally expensive and solutions can be very fragile if parameter values are far from the truth. Further, a large number of possible seller types results in a very large number of possible choice sets for which the problem would have to be solved. These issues combined make such an approach computationally infeasible. Alternatively, one may adopt an ad hoc functional form assumption for the mixing distributions and attempt to recover them jointly with other primitives of the model. It is doubtful that separate identification of these components can be established formally. In practice, such an approach has been shown to perform poorly.[17,18]

To overcome these difficulties we propose a methodology where we first classify permanent sellers into groups of equal quality. Next, we use this grouping to recover other primitives of the model. Such an approach facilitates the analysis in several ways. First, the buyers' choice sets may now be represented in terms of the participating sellers' group memberships rather than their identities. This offers a natural way for partial aggregation of buyers' choice sets and permits exploiting variation in buyers choices across choice sets which could be harnessed to identify the distribution of buyers' tastes. Further, this approach, when combined with the second insight that the support of the transitory sellers' quality distribution can be linked to the support of the permanent sellers' quality distribution, allows us to identify payoffs associated

[17] See Heckman and Singer (1984) for details.

[18] A researcher may also consider an approach proposed by Kasahara and Shimotsu (2009) in the context of a dynamic discrete choice model. However, the model considered by these authors does not readily map into our environment so the applicability of this method, if possible, is far from obvious.

with various bundles of (x, q)-attributes separately from the identification of the frequencies with which such bundles are observed in the population of transitory sellers. In the next section, we illustrate this identification strategy using a simple model.

4.1 Heuristics for Identification

Classification into Quality Groups. Consider sellers i and j with $x_i = x_j$ who participate in two auctions that are *ex ante* identical (i.e., the project characteristics and the set of competitors are the same, and both i and j are in the set of potential bidders) and submit equal bids. Under such circumstances a seller with the higher value of q has a higher chance of winning. This ranking of winning probabilities is preserved after aggregating over possible sets of competitors, as long as the chance of encountering any given set of competitors is the same for both sellers. This condition holds if, for example, the pool from which competitors are drawn does not include either i or j. Specifically, for any pair of sellers i and j such that $x_i = x_j$, define:

$$r_{i,j}(b) \equiv \Pr(i \,\text{wins} \,|\, B_{i,l} = b,\ i \in A_l,\ j \notin A_l,\, i, j \in N_l) \tag{4}$$

where A_l denotes the set of entrants.[19] Then

$$\begin{aligned} &r_{i,j}(b) > r_{j,i}(b) \text{ if and only if } q_i > q_j, \\ &r_{i,j}(b) < r_{j,i}(b) \text{ if and only if } q_i < q_j \text{ and} \\ &r_{i,j}(b) = r_{j,i}(b) \text{ if and only if } q_i = q_j. \end{aligned}$$

For a formal statement of results and conditions, see Proposition 1 in Section 4.2. As long as the conditional winning probabilities defined above are identified from data, we can use them to order sellers i and j with respect to their qualities. By implementing such comparison for every pair of permanent sellers within x-group, the quality ranking of the sellers within this group can be recovered.[20] This identifies the quality group structure.

A Simple Example. Let us see how to identify the rest of the model primitives, given the quality classification of the permanent sellers recovered above. Consider a simple setting with two groups

[19] One could use an alternative, similar index conditional on $i \in A_l$, $j \in A_l$, $B_{i,l} = B_{j,l} = b$. In our data, for many pairs of bidders, there is only a small number of auctions where both i and j participate and submit similar bids. Hence the estimation of such an alternative index is much more problematic than that of $r_{i,j}(b)$ defined above. In this paper we do not pursue such an alternative strategy.

[20] Intuitively, if comparisons for all pairs of permanent sellers are available, we can always split a given x-group into two subgroups where the first subgroup consists of the sellers with the lowest quality among all the sellers in the x-group and the second subgroup consists of the remaining sellers. Then we split this second subgroup similarly into two further subgroups so that the first further subgroup consists of the lowest quality sellers within this second subgroup and the other further subgroup consists of the rest of the sellers. By continuing this process, we can identify the quality group structure.

of sellers defined by observable characteristics $\bar{x}_1$ and $\bar{x}_2$. Each group is further partitioned into two unobservable subgroups based on quality levels $\bar{q}_1(\bar{x}_1), \bar{q}_2(\bar{x}_1)$ and $\bar{q}_1(\bar{x}_2), \bar{q}_2(\bar{x}_2)$ respectively. Some sellers are permanent and others transitory. For simplicity assume the components in buyers' weights are mutually independent, which is relaxed in our formal identification results in Section 4.2 below. The remaining aspects of the model are as described in Section 3.

Suppose there is a large number of sellers in each observable group defined by x. The number of choice sets defined in terms of specific identities of sellers is large. Nevertheless, this number can be drastically reduced if choice sets are defined in terms of quality groups instead of specific identities.[21] Such a definition of choice sets is feasible only after we use the argument above to classify sellers into groups based on unobserved quality levels.

In what follows, we discuss how to identify the payoff structure (specifically the distributions of ϵ_l, α_l, β_l and quality levels associated with each quality group, $\bar{q}_k(x_m)$ where $k = 1, 2$ and $m = 1, 2$). After that, we identify the remaining model primitives (such as F_{U_0} and $F_{Q^t|B_l^t}$).

The Distribution of Payoffs. To identify components in the buyer's payoffs, we exploit how a buyers' decisions vary with the choice set (defined in terms of (x, q)-groups of active permanent bidders). Let us first demonstrate how to identify the distribution of a stochastic component ϵ_l. For this, we focus on auctions which attract at least two permanent sellers with the same observed characteristics $x_i = x_j$ and unobserved quality $q_i = q_j$. We allow for the presence of a transitory seller h. The payoffs from different options are:

$$U_{i,l} = \alpha_l q_i + \beta_l x_i - B_{i,l} + \epsilon_{i,l} \ ; \ U_{j,l} = \alpha_l q_j + \beta_l x_j - B_{j,l} + \epsilon_{j,l};$$
$$U_{h,l} = \alpha_l Q_{h,l} + \beta_l x_1 - B_{h,l} + \epsilon_{h,l} \text{ and } U_{0,l}.$$

The scale of α_l and the quality levels can not be jointly identified; hence we normalize $\mathbf{E}[\alpha_l] = 1$.

When $B_{i,l} = -t_2$ and $B_{j,l} = t_1 - t_2$ for some t_1, t_2, the buyer chooses seller i with probability

$$\begin{aligned}
& \Pr(i \text{ wins in auction } l | B_{i,l} = -t_2, B_{j,l} = t_1 - t_2, B_{h,l} = b_h) \\
= \ & \Pr(\epsilon_{j,l} - \epsilon_{i,l} \leq t_1 \text{ and } Y_{i,l}(x_h) - \epsilon_{i,l} \leq t_2 | B_{i,l} = -t_2, B_{j,l} = t_1 - t_2, \ B_{h,l} = b_h) \\
= \ & \Pr(\epsilon_{j,l} - \epsilon_{i,l} \leq t_1 \text{ and } Y_{i,l}(x_h) - \epsilon_{i,l} \leq t_2 | B_{h,l} = b_h) \equiv F(t_1, t_2 | b_h),
\end{aligned}$$

where we let $Y_{0,l}(x_h) \equiv \max\{\alpha_l Q_{h,l} + \beta_l x_h - B_{h,l} + \epsilon_{h,l}, U_{0,l}\}$ and $Y_{i,l}(x_h) \equiv Y_{0,l}(x_h) - \alpha_l q_i(x_i) - \beta_l x_i$. The last equality follows because under our model assumptions the bids are independent across sellers and independent of buyers' tastes, and because $(\epsilon_{i,l}, \epsilon_{j,l}, Y_{i,l}(x_h), B_{h,l})$ are independent of $(B_{i,l}, B_{j,l})$.

[21] Suppose there are 100 sellers. Then the number of choice sets that consist of three sellers is $C^3_{100} = 100!/97!3! = 100 * 99 * 98/6 = 161,700$. However, if we define choice sets in terms of groups, we can reduce the number of distinct choice sets to $4^3 = 64$.

The winning probability on the left-hand side is directly identifiable from the data. Hence the joint distribution F on the right hand side is identified. Since $\epsilon_{j,l}$, $\epsilon_{i,l}$ and $(Y_{i,l}(x_h), B_{h,l})$ are independent, the conditional distribution of $Y_{i,l}(x_h)$ given $B_{h,l} = b_h$ and the distributions of $\epsilon_{j,l}$ and $\epsilon_{i,l}$ are identified up to a location normalization if the support of $(B_{i,l}, B_{j,l})$ is large enough.[22] This intuition continues to hold when the conditional winning probabilities are aggregated over distinct choice sets that include two permanent sellers from the same (x, q)-group.

The quality levels $\bar{q}_1(\bar{x}_1), \bar{q}_1(\bar{x}_2), \bar{q}_1(\bar{x}_2), \bar{q}_2(\bar{x}_2)$ and the distributions of α_l and β_l are identified similarly. Specifically, we identify the distribution of $\alpha_l(\bar{q}_1(x) - \bar{q}_2(x))$ by applying similar arguments to the subset of auctions with choice sets consisting of permanent sellers i and j from the same observable group x but different quality groups, $\bar{q}_1(x)$ and $\bar{q}_2(x)$, a transitory seller h and the outside option. The mean of the distribution $\alpha_l(\bar{q}_1(x) - \bar{q}_2(x))$ identifies $\bar{q}_1(x) - \bar{q}_2(x)$ under the normalization $\mathbf{E}[\alpha_l] = 1$. Further, we consider the subset of auctions with choice sets consisting of permanent sellers i and j from the lowest quality from the observable groups, $\bar{x}_1$ and $\bar{x}_2$, a transitory seller h from $\bar{x}_2$ and the outside option. Then we can identify the distribution of β_l under an additional normalization that $\bar{q}_1(\bar{x}_1) = \bar{q}_1(\bar{x}_2) = 0$, i.e., the lowest quality in each observable group is normalized to zero. This restriction on the lowest quality levels can be relaxed if β_l is a fixed constant parameter (equal to β) rather than a random variable. In this case it is enough to normalize the lowest quality level for a single observable group, e.g., $\bar{q}_1(\bar{x}_1)$, whereas $\bar{q}_1(\bar{x}_2) - \bar{q}_1(\bar{x}_1) = \bar{q}_1(\bar{x}_2)$ is identified with the rest of the model.

Outside Option and Transitory Sellers' Quality Distribution. We now explain how to identify the distribution of payoffs from the outside option $U_{0,l}$, and the conditional quality distribution $\Pr(Q_{h,l} = \bar{q}_k(x_h)|B_{h,l} = b_h, X_h = x_h)$ for transitory sellers. The latter depends on sellers' bidding strategies in equilibrium.

Consider auctions with one permanent bidder and one transitory bidder, and assume that $U_{0,l}$, β_l and α_l are independent. (Our formal results in the next subsection allow for the correlation between $U_{0,l}$ and α_l.) Recall that $(\bar{q}_1(\bar{x}_1), \bar{q}_1(\bar{x}_2))$ and the distributions of α_l and β_l and that of $Y_{1,l}(x_h) = Y_{0,l}(x_h) - \alpha_l \bar{q}_1(\bar{x}_1) - \beta_l \bar{x}_1$ conditional on $B_{h,l} = b_h$ are identified in the previous step. Hence the conditional distribution of $Y_{0,l}(x_h)$, which is the maximum of the payoff from the outside option and that from the transitory seller, is also identified.

Next, let us now argue that knowledge of the conditional distributions of $Y_{0,l}(x_h)$ helps to identify the distribution of the outside option and the conditional distribution of a transitory seller's quality. With $U_{0,l}$, β_l and α_l independent of each other, the payoff to the outside option $(U_{0,l})$ and the payoff to the transitory seller $(U_{h,l})$ are independent. Hence, for each fixed pair of

[22] This is a consequence of Kotlarski Theorem. See Rao (1992) for details. The formal support requirements are stated in Section 5.2.2 and discussed in the Web Supplement to the paper.

numbers (y_0, b_h),

$$\Pr(Y_{0,l}(x_h) \leq y_0 | B_{h,l} = b_h, X_h = x_h) = \Pr(U_{0,l} \leq y_0) \Pr(U_{h,l}(x_h) \leq y_0 | B_{h,l} = b_h, X_h = x_h). \quad (5)$$

This leads to two equations, one with $x_h = \bar{x}_1$ and the other with $x_h = \bar{x}_2$. After rearranging terms in these equations, we have:

$$\begin{aligned} & g_1(y_0; b_h) \Pr(Q_{h,l} = \bar{q}_1(\bar{x}_2) | B_{h,l} = b_h, X_h = \bar{x}_2) \\ & - g_2(y_0; b_h) \Pr(Q_{h,l} = \bar{q}_1(\bar{x}_1) | B_{h,l} = b_h, X_h = \bar{x}_1) = g_3(y_0; b_h), \end{aligned}$$

where for each b_h, $g_s(y_0; b_h)$, $s = 1, 2, 3$, are known functions of y_0, and $\Pr(Q_{h,l} = \bar{q}_1(\bar{x}_j) | B_{h,l} = b_h, X_h = \bar{x}_j)$, $j = 1, 2$, are unknown probabilities to be recovered.[23] These probabilities are over-identified since we have infinitely many linear equations associated with different values of y_0. Once these probabilities are identified, the distribution of the payoff from the outside option $U_{0,l}$ is identified from (5).

4.2 Formal Results

This section presents our identification results formally.

4.2.1 Quality Classification of Permanent Sellers

The first step is to recover the classification of permanent sellers from the data. Let N_l^t denote the set of transitory potential bidders in auction l.

Assumption 1 *(i) (a) For each $j \in N_l$, $E_{j,l}$ and $C_{j,l}$ are independent. (b) For each $j \in N_l$, $(\alpha_l, \beta_l, \epsilon_l, U_{0,l})$ and $(E_{j,l}, C_{j,l})$ are independent. (c) $C_{j,l}$'s are continuously distributed and i.i.d. across the sellers $j \in N_l$, with its distribution (and potentially its support) depending on (x_j, q_j) but not on $\rho_j \in \{p, t\}$. (d) $(\alpha_l, \beta_l, \epsilon_l, U_{0,l})$ is independent from $(x_i)_{i \in N_l}$. (ii) If we let $\tilde{\alpha}_l \equiv (\alpha_l, U_{0,l})$, then $\tilde{\alpha}_l, \epsilon_l$, and β_l are mutually independent, each having a connected support, and $\epsilon_{j,l}$'s are i.i.d. across the sellers. (iii) The quality of transitory potential bidders in auction l, denoted by $Q_l^t \equiv (Q_{j,l} : j \in N_l^t)$, is independent of $(\alpha_l, \beta_l, \epsilon_l, U_{0,l})$ and $Q_{j,l}$ are independent across $j \in N_l^t$.*

The independence between $(\alpha_l, U_{0,l})$ and β_l is not necessary for Propositions 1, 2 and the first part of Proposition 3 below. Let $\mathcal{B}_i$ denote the support of the price quoted by a seller i in a type-symmetric pure-strategy Bayesian Nash equilibrium (p.s.BNE).

---

[23] Specifically,
$$\begin{aligned} g_1(y_0; b_h) &= \Pr(Y_0(\bar{x}_1) \leq y_0 | b_h, \bar{x}_1)[J_1(b_h, \bar{x}_2) - J_2(b_h, \bar{x}_2)], \\ g_2(y_0; b_h) &= \Pr(Y_0(\bar{x}_2) \leq y_0 | b_h, \bar{x}_2)[J_1(b_h, \bar{x}_1) - J_2(b_h, \bar{x}_1)], \\ g_3(y_0; b_h) &= \Pr(Y_0(\bar{x}_2) \leq y_0 | b_h, \bar{x}_2) J_2(b_h, \bar{x}_1) - \Pr(Y_0(\bar{x}_1) \leq y_0 | b_h, \bar{x}_1) J_2(b_h, \bar{x}_2), \end{aligned}$$
where $J_k(b, x)$ denotes $\Pr(U_h(x) \leq y_0 | B_h = b_h, x, Q_h = \bar{q}_k(x))$ for $k = 1, 2$. The functions J_k, the conditional distribution of Y_0, and therefore $g_s(y_0; b)$ for $s = 1, 2, 3$, are recovered from the previous steps. See Section B2 in the Web Supplement to the paper for details.

Assumption 2 *For any permanent sellers i, j with $x_i = x_j$, $\mathcal{B}_i \cap \mathcal{B}_j$ contains an interval with a non-empty interior.*

For each pair of permanent sellers i and j, let the index $r_{i,j}$ be defined as in (4):

$$r_{i,j}(b) \equiv \Pr(i \text{ wins} \mid B_{i,l} = b,\ i \in A_l,\ j \notin A_l,\ i, j \in N_l)$$

Note that A_l is a random set of entrants in the bidding stage, and $r_{i,j}$ does not condition on the identities of the entrants other than i and j. In practice, we construct this index by pooling all the auctions that satisfy the event conditioned on.

Proposition 1 *Suppose Assumptions 1 and 2 hold. Then for each pair of permanent sellers i, j with $x_i = x_j$ and $b \in \mathcal{B}_i \cap \mathcal{B}_j$,*

$$sign\,(r_{i,j}(b) - r_{j,i}(b)) = sign(q_i - q_j),$$

where $sign(z) \equiv 1\{z > 0\} - 1\{z < 0\}$, for $z \in \mathbb{R}$.

The proposition says that for each pair of permanent sellers i and j, we can determine their quality ordering by looking at the sign of $r_{i,j}(b) - r_{j,i}(b)$. Thus if such a sign is available for each pair of permanent sellers, we can identify the quality group structure among the permanent sellers.

4.2.2 Quality Indices and the Distribution of Buyer Preferences

Next, we identify the quality levels and the distribution of α_l and ϵ_l. Let $\overline{N}^p$ be the total set of permanent sellers in the population, and $\overline{N}^t$ that of transitory sellers in the population. For a generic set $a = a^p \cup a^t$ with $a^p \subset \overline{N}^p$ and $a^t \subset \overline{N}^t$, we denote (x, q)-specific subsets of a by:

$$\begin{aligned} a^p(x, q) &\equiv \{i \in a^p : x_i = x \text{ and } q_i = q\}, \text{ and} \\ a^t(x) &\equiv \{i \in a^t : x_i = x\}. \end{aligned}$$

Hence $a^p(x, q)$ denotes the (x, q)-subgroup of the permanent sellers in a, and $a^t(x)$ the x-subgroup of the transitory sellers in a. For each (x, q), define $\lambda^p(a, x) \equiv (|a^p(x, q)| : q \in \mathcal{Q}(x))$ and $\lambda^t(a, x) \equiv |a^t(x)|$, i.e., the collections of cardinalities of subgroups. Then we define the *composition* of a generic set of sellers a as

$$I_a \equiv (\lambda^p(a, x), \lambda^t(a, x))_{x \in \mathcal{X}}.$$

Note that I_a depends on a only through the sizes of the subgroups contained, not through the identities of sellers.

For any generic set $a \subset \overline{N}$, define $b_a \equiv (b_s : s \in a)$. For a given seller $j \in N_l$ and $a \subset N_l \backslash \{j\}$, define a random variable $Y_{j,l}(b_a, I_a)$ as the maximum of $U_{0,l} - \alpha_l q_j - x_j \beta_l$ and $(x_s - x_j)\beta_l + \alpha_l(Q_{s,l} - q_j) - b_s + \epsilon_s$ for $s \in a$. By construction, the distribution of $Y_{j,l}$ depends on (b_a, I_a) but not on the specific identities of sellers in a. For a pair of permanent sellers $\{i, j\}$ and a subset $a \subseteq N_l \backslash \{i, j\}$, define

$$V_{i,j,l}(b_a, I_a) \equiv (\alpha_l(q_j - q_i) + \Delta x_{j,i}\beta_l + \epsilon_{j,l} - \epsilon_{i,l}, Y_{i,l}(b_a, I_a) - \epsilon_{i,l}),$$

where $\Delta x_{j,i} \equiv x_j - x_i$.

Assumption 3 *For each pair of groups $((x, q), (x', q'))$, and any pair of permanent sellers i, j such that i belongs to (x, q)-group and j belongs to (x', q')-group, there exists some composition I_a where $a \subset N_l \backslash \{i, j\}$ and a related bid vector $b_a \equiv (b_k)_{k \in a}$ such that (i) the support of $(B_{j,l} - B_{i,l}, -B_{i,l})$ contains the conditional support of $V_{i,j,l}(b_a, I_a)$ given (b_a, I_a), and (ii) $V_{i,j,l}(b_a, I_a)$ conditional on (b_a, I_a) has a non-vanishing characteristic function.*

To illustrate the support condition in Assumption 3, consider a pair of permanent sellers i, j such that $(q_i, x_i) = (q_j, x_j)$. Then this support condition holds with the set of other sellers being empty if the joint support of $\epsilon_{j,l} - \epsilon_{i,l}$ and $U_{0,l} - \alpha_l q_i - x_i \beta_l - \epsilon_{i,l}$ is a subset of the support of $(B_{j,l} - B_{i,l}, -B_{i,l})$. This requires the buyer-seller components to have a smaller support relative to that of the bids in equilibrium. In another example, again consider a pair of permanent sellers i, j such that $(q_i, x_i) = (q_j, x_j)$, and I_a consists of a single transitory seller k with $x_k = x_i$. Then this support condition holds if there exists an equilibrium bid $b_{k,l}$ such that the joint support of $\epsilon_{j,l} - \epsilon_{i,l}$ and $\max\{U_{0,l} - \alpha_l q_i - x_i \beta_l, \alpha_l(Q_{k,l} - q_i) - B_{k,l} + \epsilon_{k,l}\} - \epsilon_{i,l}$ is a subset of the support of $(B_{j,l} - B_{i,l}, -B_{i,l})$. In Section B4 of the Web Supplement to the paper, we discuss how our model is capable of generating the price variation in Assumption 3.

Proposition 2 *Suppose Assumptions 1, 2 and 3 hold and $\epsilon_{j,l}$'s have non-vanishing characteristic functions. Then the distribution of $\epsilon_{j,l}$ is identified up to a location normalization (e.g., $\mathbf{E}(\epsilon_{j,l}) = 0$); and the distribution of α_l and the differences in quality levels are jointly identified up to a scale normalization (e.g., $\mathbf{E}(\alpha_l) = 1$).*

The next proposition identifies the distribution of the buyer's outside option $U_{0,l}$ and tastes for observed characteristics β_l under additional conditions stated as Assumptions 4 and 5 in the Web Supplement to the paper.

Proposition 3 *Suppose that Assumptions 1, 2, 3, 4 and 5 hold, and that the lowest quality level for a permanent seller is the same across groups with different observed characteristics x. Then the distribution of the value of outside option conditional on α_l, the distribution of the transitory seller's quality conditional on his bids, and the distribution of β_l are jointly identified.*[24]

[24]Since neither the quality levels nor the buyer's random tastes $(\alpha_l, \beta_l, \epsilon_l, U_{0,l})$ are recorded in the data, a location normalization is required for full identification of the quality levels. Under the condition on lowest quality levels in Proposition 3, we identify the model by normalizing the lowest quality to zero.

To identify the distribution of β_l, we use the variation in bids submitted by permanent sellers. To do so, we need an extended version of the support condition in Assumption 3. We state the new condition as Assumption 5 in Section B2 of the Web Supplement to the paper. In the special case where β_l is a constant vector, identification only requires a weaker location normalization that fixes the lowest quality level for permanent sellers in a single group with a fixed observed characteristic. We discuss this with more details in Section B2 of the Web Supplement to the paper.

The formal results concerning identification of the distributions of sellers' costs are provided in Section B3 of the Web Supplement to the paper.

5 Estimation

We estimate the model in two steps. In the first step, we use a classification algorithm to uncover the (unobserved) group memberships for the permanent sellers, and in the second step, we perform GMM estimation to recover the rest of the model primitives.

Classification Algorithm. Our classification algorithm is based on Proposition 1. More specifically, for two permanent sellers i and j, we define $\hat{\delta}_{ij}(b) \equiv \hat{r}_{i,j}(b) - \hat{r}_{j,i}(b)$, where

$$\hat{r}_{i,j}(b) \equiv \frac{\sum_{l=1}^{L} 1\{i \text{ wins }\} K_h(B_{i,l} - b) 1\{j \notin A_l\} 1\{i, j \in N_l\}}{\sum_{l=1}^{L} K_h(B_{i,l} - b) 1\{j \notin A_l\} 1\{i, j \in N_l\}},$$

where $K_h(v) = K(v/h)/h$ for a univariate kernel function K whereas N_l and A_l are the sets of potential and actual bidders in auction l respectively.[25] Then we construct test statistics: $\hat{\tau}_{ij}^{+} = \int \max\{\hat{\delta}_{ij}(b), 0\} db$, $\hat{\tau}_{ij}^{-} = \int \max\{-\hat{\delta}_{ij}(b), 0\} db$, and $\hat{\tau}_{ij}^{0} = \int |\hat{\delta}_{ij}(b)| db$, where the integral domain is taken to be the sample version of $\mathcal{B}_i \cap \mathcal{B}_j$ that is the intersection of the bid supports of bidders i and j. For example, the test statistic $\hat{\tau}_{ij}^{+}$ is used to check whether $r_{i,j}(b) > r_{j,i}(b)$, i.e., whether $q_i > q_j$ (from Proposition 1.)

Next, we construct a pairwise bootstrap p-value for testing the quality ordering between i and j. For this we generate $\hat{r}_{i,j}^{*}(b)$ and $\hat{\delta}_{ij}^{*}(b)$ using the bootstrap sample of data, and construct the re-centered bootstrap test statistics, $\hat{\tau}_{ij,s}^{*+}$, $\hat{\tau}_{ij,s}^{*-}$ and $\hat{\tau}_{ij,s}^{*0}$ in the same way as we constructed $\hat{\tau}_{ij}^{+}$, $\hat{\tau}_{ij}^{-}$, and $\hat{\tau}_{ij}^{0}$ except that we use re-centered bootstrap quantities $\hat{\delta}_{ij,s}^{*}(b) - \hat{\delta}_{ij}(b)$ in place of $\hat{\delta}_{ij}(b)$.[26] From these we find bootstrap p-values.

We use the algorithm proposed in **?** which is designed to recover the quality group structure so

[25] Test statistics is constructed using a triweight kernel function: $K(u) = 1\{|u| \le 1\}(35/32)(1 - u^2)^3$. The bandwidth selection follows the usual Silverman's rule of thumb. Other parameters use in implementation are reported in the Web Supplement.

[26] Lee, Song, and Whang (2015) established the asymptotic validity of such bootstrap tests of nonparametric inequality restrictions in a more general set-up.

that transitivity of the quality ordering is retained in finite samples. Heuristics for the algorithm is as follows. For each seller i in an x-group, we first divide the remaining sellers into two groups, one with sellers likely to have higher quality than i and the other with sellers likely to have lower quality than i. We obtain this division by comparing the p-values from two pairwise bootstrap tests of the inequality restrictions $r_{i,j} \geq r_{j,i}$ and $r_{i,j} \leq r_{j,i}$. Next, we place seller i in one of the two groups depending on whether seller i is likely to have the same quality as the other sellers in the group. Thus we obtain one group structure for each seller i, and choose one of these structures (specifically, the one that has strongest empirical support in terms of average p-values). This gives the first division of the sellers into two subgroups.

We then sequentially select a subgroup with sellers most likely to have heterogeneous qualities, and divide the group similarly as before. To prevent overfitting (i.e., ending up with too many subgroups), we stop the division process when a goodness-of-fit measure defined in terms of average p-values is dominated by a penalty term. (See the Web Supplement for further details on the implementation of the algorithm.)

GMM Estimation: Moments. Given the estimated quality group structure in the first step, we proceed with GMM estimation of the model primitives.[27] For GMM, the moment conditions are primarily built around the permanent seller's winning probability given the seller's attributes, quality group membership and for a given configuration of the set of active permanent sellers.

To be specific, let B_l be the vector of submitted bids in auction l. For each bidder j in auction l, define

$$m_{j,l} = 1\{j \text{ wins } l\} - \Pr(j \text{ wins } |B_l, I_{A_l}),$$

where $\Pr(j \,\text{wins}\, l\,|B_l, I_{A_l})$ is the conditional winning probability of seller j in auction l having a given composition of active sellers I_{A_l} when the vector of submitted bids is B_l. Then we construct a moment condition as follows:

$$\mathbf{E}\left[\sum_j g_j(B_l, I_{A_l}) m_{j,l}\right] = 0,$$

where $g_j(B_l, I_{A_l})$ is a function of B_l and I_{A_l}, and the summation is over j in the set of active permanent bidders at auction l.

For the implementation, it remains to choose functions $g_j(B_l, I_{A_l})$. The functions are chosen to exploit variations of the sellers' (x, q)-group memberships and variations in the compositions I_{A_l} of the active sellers. Specifically, we consider two types of moments. The moments of the first

[27]The estimation error due to using the estimated quality groups does not affect the asymptotic distribution of the GMM estimator, because the number of quality groups is a discrete parameter and the probability that our estimator is equal to the actual number of quality groups converges to one as the sample size increases.

type are based on the subset of auctions such that the set A_l includes at least two permanent sellers from the same (x, q)-group. The moments of the second type are confined to the auctions where the set A_l contains permanent sellers belonging to two specific groups, (x, q) and (x', q'), for every possible pair of groups.

Formally, for each choice of (x, q)-, (x', q')-, and x_h-groups and each composition I_{A_l}, we form $g_j(B_l, I_{A_l})$'s as functions of $(B_{i,l}, B_{j,l}, B_{h,l})$, where $B_{j,l}$ is the bid of winning permanent seller j from group (x, q), $B_{i,l}$ the bid of another permanent seller i from group (x', q') and $B_{h,l}$ the bid of a transitory seller from group x_h. As for the functions of $(B_{i,l}, B_{j,l}, B_{h,l})$, we consider: constant (equal to 1); $B_{j,l}$ and $B_{j,l}^2$; $B_{j,l} - B_{i,l}$ and $(B_{j,l} - B_{i,l})^2$; $(B_{j,l} - B_{i,l})B_{j,l}$; $B_{j,l}B_{h,l}$ and $B_{j,l}^2 B_{h,l}$ for a transitory seller h; $B_{j,l}x_h$ and $B_{j,l}^2 x_h$. See the Web Supplement to the paper for details of our choice of the functions g_j's.

GMM Estimation: Accounting for Transitory Seller's Qualities. To use the moment conditions summarized above, we need to evaluate the conditional winning probability. However, note that a buyer observes (x, q)-group memberships of all sellers in his choice set, whereas the econometrician does not observe the transitory sellers' quality groups. This means that the conditional winning probability is the winning probability after integrating out the vector of participating transitory sellers' qualities. In other words,

$$\Pr(j \text{ wins } |B_l, I_l) = \sum_{\bar{q}_t} \Pr(j \text{ wins } l|B_l, I_{A_l}, Q_l^t = \bar{q}_t) \Pr(Q_l^t = \bar{q}_t|B_l, I_l), \tag{6}$$

where Q^t is the participating transitory sellers' quality vector in the auction, and $I_l = \{I_{A_l}, I_{N_l}\}$ denotes the compositions of the sets of actual and potential sellers.

The conditional probability $\Pr(j \text{ wins } l|B_l, I_{A_l}, Q_l^t = \bar{q}_t)$ reflects buyers' decisions, and is determined by the distribution of buyers' weights and outside option. To obtain a theoretical expression for this probability we parametrize the distributions of $\epsilon_{j,l}$ and $(\alpha_l, \beta_l, U_{0,l})$ in a standard way. However, it is not immediately obvious how to parametrize $\Pr(Q_l^t = q_t|B_l, I_l)$, because it involves the transitory sellers' behavior. Assume that the auction at hand contains only one transitory actual bidder, say, h. Then we express[28,29] $\Pr(Q_{h,l} = q_h|B_l = b, x_h, I_l)$ as

$$\frac{\Pr(Q_{h,l} = q_h|x_h) \Pr(h \text{ is active}|x_h, Q_{h,l} = q_h, I_{N_l}) f_b(b_h|x_h, q_h, I_{N_l})}{\sum_{q'_h} \Pr(Q_{h,l} = q'_h|x_h) \Pr(h \text{ is active}|x_h, Q_{h,l} = q'_h, I_{N_l}) f_b(b_h|x_h, Q_{h,l} = q'_h, I_{N_l})}, \tag{7}$$

where the subscript h of a variable indicates that it belongs to the transitory seller h. In esti-

[28]The derivation for a general case is available in the Web Supplement to the paper.

[29]Notice that information about x_h is contained in I_l. We condition on it separately here to make the expression more accessible to the reader.

mation we parameterize the bid density function $f_b(b_h|x_h, Q_{h,l} = q_h, I_{N_l})$ and the probability of entry $\Pr(h \text{ is active}|x_h, Q_{h,l} = q_h, I_{N_l})$.

Finally, we impose additional restrictions in estimation that allow us to recover $\Pr(h \text{ is active}|x_h, Q_{h,l} = q_h, I_{N_l})$, $f_b(b_h|x_h, Q_{h,l} = q_h, I_{N_l})$, and $\Pr(Q_{h,l} = q_h|x_h)$ separately.[30] Specifically, we impose that the conditional bid distributions and conditional participation probabilities aggregated to the level observed in the data correspond to their empirical counterparts. We also impose optimality of the transitory sellers' participation decisions. (See the Web Supplement for details.)

The details of the estimation strategy used to recover the distributions of sellers' costs are provided in Section D4 of the Web Supplement to the paper.

6 Empirical Results

This section summarizes the estimation results. We begin by describing the implementation details of the classification procedure and the estimated quality group structure. We then turn to the details and the results of the parametric estimation.

Implementation Details. We assume that the buyer's utility from using a specific seller depends on the seller's country affiliation, average reputation score, and his (residual) quality in addition to the seller's bid. The seller's country affiliation proxies for things such as work culture, convenience of working with a given seller related to the time difference, and the likelihood of language proficiency, whereas the reputation score may reflect public information about the seller's quality. The distribution of residual quality may also plausibly depend on such factors so we allow the unobserved group structure, which captures this distribution, to depend on the sellers' countries, and the long-run averages of the sellers' reputation scores.[31]

We are interested in the long-run differences among sellers. That is why, we focus our analysis on the last four years captured by our dataset (i.e., years three to six of the market operation). This ensures that the majority of permanent sellers (more than 95%) have been with the platform for more than a year by the starting date of our estimation sample.

We divide all the sellers into three cells according to the long-run average reputation score:

[30] Non-parametric identification imposes continuum of restrictions on the shape of $\Pr(Q_h = \bar{q}_h|b_h, x_h, I)$. However, the mapping between such restrictions and the parameterized components listed above is less straightforward than in the case of other primitives. Given how the variation in bids is exploited in other moments, some of the parameters of the three components above may not be identified. We impose additional restrictions to ensure parametric identification. The formal argument which establishes that such restrictions are sufficient for identification will be provided upon request.

[31] We have also verified the robustness of our results by repeating the analysis while including the number of arbitrations and delays as additional observable measures of quality. The results of this analysis are less precise since each cell contains a smaller number of observations but they are very similar to the results we report in the paper.

average reputation score less than 9.7 (cell 1), average reputation score above 9.7 and below 9.9 (cell 2), average reputation score above 9.9 (cell 3). This results approximately in an allocation of 30%, 30%, and 40% of the sellers to the three cells.

We further group sellers into country groups by geographic proximity and similarity of language and economic conditions. We end up with seven country groups: North America (USA and Canada), Latin America, Western Europe, Eastern Europe, Middle East and Africa, South and East Asia, and Australia (grouped with New Zealand). In our data, North America, Eastern Europe and South or East Asia account for the majority of submitted bids.

6.1 Classification Results

The classification index is constructed for pairs of permanent sellers on the basis of projects where they both belong to the set of potential bidders. We follow the steps summarized in Section 5 and described in detail in the Web Supplement. That is, we start by estimating a group structure for different numbers of groups. We then apply a criterion function to select the structure with the number of groups most supported by the data. For this structure we then compute confidence sets. We demonstrate the first two steps for the group of Eastern European sellers with a medium level of average reputation score in a table included in the Web Supplement to the paper.

We have estimated the model for several cells of projects defined in terms of size and duration. The difference in the results across cells is not sufficiently large to warrant a separate discussion in the paper. The results presented below are for the projects owned by US buyers that are of medium size (between \$400 to \$600) and have the specified duration of two to three weeks. We use the results for all cells in our counterfactual analysis.

Table 6 reports the estimated group structures with corresponding confidence sets for cells of North American, Eastern European and East Asian sellers.[32] We estimate multiple quality groups in each cell and the confidence sets associated with each group structure are quite tight. It is difficult to draw any substantive conclusions about the quality distribution on the basis of these results, since the classification into groups is ordinal and does not allow for the comparison of levels across countries or reputation scores. We note here that even the cells that correspond to a very narrow range of reputation scores (such as medium or high reputation scores) are classified into multiple quality groups. Also, allocation of mass between quality groups differs across cells.

We conduct extensive robustness analysis in order to verify robustness of our results to the assumptions of our model and various implementation details. Among other things we explore the potential importance of unobserved auction heterogeneity and the possibility that the sellers'

[32]Table 3 in Section F of the Web Supplement reports the smallest and the average number of observations across pairs of sellers available for classification analysis.

Table 6: Estimated Quality Groups by Seller Covariates

Country Group	Average Score	Total Number of Sellers	$Q=1$ Number of Sellers	$Q=1$ C.S.	$Q=2$ Number of Sellers	$Q=2$ C.S.	$Q=3$ Number of Sellers	$Q=3$ C.S.
North America	low	12	4	(6)	8	(10)		
North America	medium	13	4	(6)	9	(11)		
North America	high	17	12	(13)	5	(6)		
Eastern Europe	low	18	6	(8)	12	(14)		
Eastern Europe	medium	52	33	(37)	12	(14)	7	(9)
Eastern Europe	high	83	6	(7)	65	(69)	12	(15)
East Asia	low	91	62	(68)	18	(22)	11	(13)
East Asia	medium	66	6	(8)	53	(57)	7	(9)
East Asia	high	58	50	(53)	8	(11)		

This table shows the estimated group structure and a consistently selected number of groups for each cell determined by covariate values. Column 3 indicates the total number of the sellers in the cell. Columns 4-6 report the size of the estimated quality group. The number of the sellers in the corresponding confidence set with 90% coverage is reported in parenthesis. Note that the confidence set with the level (1 - α) for a given quality group is defined to be a random set whose probability of containing this quality group is ensured to be asymptotically bounded from below by (1-α).

quality may vary across projects. We find that the results of classification analysis are quite robust and change very little across specifications that we consider. The results also indicate that unobserved auction heterogeneity, if present, plays a limited role in our environment and that the sellers' qualities appear constant across projects. The results of this analysis are summarized in the Web Supplement to the paper.

6.2 The Results of GMM Estimation

In this section we present the results of the GMM estimation. We begin by summarizing our specification and then discuss the estimates of the objects of interest: the distribution of the buyers' utility weights, the quality distributions for a range of covariate values, the sellers' bidding strategies and the recovered cost distributions. The estimates of objects which are auxiliary to our analysis (such as the distribution of bids and participation probabilities) are reported and discussed in the Web Supplement to this paper.

Baseline Specification. We modify the specification of the buyer's utility function for the purpose of estimation. Specifically, we impose that $\beta_l = \beta\alpha_l$, where the multiplier is constant across buyers, as we find that such specification achieves the best fit to the data. Thus, the utility function used in estimation is given by $\alpha_l\tilde{q}_i - B_{i,l} + \epsilon_{i,l}$ where $\tilde{q}_i = q_i + x_i\beta$.

Our baseline specification imposes that the distribution of the sellers' qualities conditional on the vector of observable covariates are the same in the populations of the permanent and transitory sellers. This restriction is more stringent than is necessary for our methodology which requires only that the supports of the quality distributions should coincide. Imposing the equality of distributions enhances robustness and precision of our estimates. Such an assumption also appears to be consistent with the regularities documented in the data. The Web Supplement reports the estimation results for two alternative specifications. We comment on these results later in this section.

We assume that buyer-seller-specific components, $\epsilon_{j,l}$, follow the Extreme Value Type I distribution while weight parameter α_l and the buyer's outside option are assumed to be distributed according to the joint normal distribution.[33] We impose the normalizations implied by our identification argument. That is, we normalize the expected value of $\epsilon_{j,l}$ to be equal to zero, the expected value of α_l to be equal to one, and one of the quality levels (quality level 1 of the low average score group, the South and East Asian country group) to be equal to zero. Further, we assume that transitory and permanent sellers' bid distributions are well approximated by normal distributions.[34,35,36] Similarly, we approximate permanent and transitory bidders' respective probabilities of participation by normal distribution functions.[37]

The majority of transitory sellers complete only one or two projects. As a result their long-run average reputation scores are not observed in the data. We assume that buyers use public information to form beliefs about the probability that a beginning seller belongs to a particular long-run average score group which we recover non-parametrically using the long-run data on permanent bidders.

Quality and Other Attributes as Determinants of Buyer's Choice. Table 7 shows the estimated parameters of the distribution of the buyers' utility weights and estimated quality levels across cells corresponding to different values of sellers' $x-$attributes. In the estimation the prices are normalized by project size; therefore, the estimates for quality levels reflect the buyers'

[33]Strictly speaking, the distribution of α_l should have been chosen to have a non-negative support. However, we estimate the standard error of this distribution to be quite small so that this assumption does not make any practical difference. The same comment applies to our assumption on the distribution of bids below.

[34]See the comment for the distribution of α_l above.

[35]A figure in Section F3 of the Web Supplement to the paper demonstrates that this assumption is a good approximation of the distribution of bids observed in the data.

[36]The means of the bid distribution depend on the seller's quality, average reputation score and country group, as well as on the number of potential permanent competitors by group. We allow the bid distribution of transitory sellers to depend on the number of reputation scores and on both the current and the long-run average scores. This is because the long-run average score is not observed in the data for transitory sellers and the buyer has to base his expectation of the long-run average reputation score on contemporaneously available measures when awarding the project.

[37]We assume that these functions depend on linear indexes of the seller's quality, long-run average score and country group, the numbers of potential competitors by group (as well as the current number of reputation scores, and the current average of reputation scores in the case of transitory sellers).

willingness to pay in terms of the percentage of the project size.

The differences in the estimated quality levels are substantial in magnitude. Specifically, an average buyer is willing to pay an average premium of (0.5 $\times$ the project size) in order to obtain services of a seller with the highest rather than the lowest quality. The quality levels have expected signs and are increasing according to group ranking. We observe that the quality levels are consistent across covariate cells. There appear to be roughly three quality levels present in this market, with the lowest normalized to be around zero, the medium quality level estimated to be somewhere in the range 0.1-0.3, and the highest quality level is between 0.45-0.68. The exact levels differ across country groups, with Eastern Europe characterized by the highest values for each quality level and North America characterized by the lowest "high" quality levels.

Having established that the quality levels are very similar across covariate groups, we can conclude, based on the results from the previous section (Table 6), that there exist important differences in the distribution of quality mass across covariate levels. In particular, North America is missing a middle quality level, whereas the lowest average score cell for Eastern Europe and the highest average score cell for South and East Asia are missing the lowest quality levels. Similarly, the medium score cell for Eastern Europe allocates the most mass to the lowest and medium quality levels, whereas the highest score cell allocates the most mass to the medium and high quality levels. We observe similar regularities in the case of South and East Asia. Hence, the distribution of qualities varies significantly with covariate values.

The country affiliation and the long-run average reputation score appear to have separate effects on the buyer's utility. These effects, however, are rather small relative to the differences in quality levels. For example, using the quality levels reported in Table 7 we compute that a buyer with the average taste for quality ($\alpha = 1$) would be willing to pay almost 9% more of the project size, ($0.507 - 0.413 = 0.094$), to obtain the service of a high-quality North American seller with a high reputation score rather than a high-quality North American seller with a low reputation score. Similarly, a buyer with the average taste for quality would be willing to pay 12% more of the project size, $(0.668 - 0.544) = 0.124$, to hire a medium score, high-quality supplier from Eastern Europe rather than a medium score, high-quality supplier from South or East Asia.

Notice that buyers are quite heterogeneous in their willingness to pay for quality. For example, whereas an average buyer would be willing to pay a 0.507 premium for a high-score, high-quality North American seller, about 5% of the buyers would pay less than 0.20 ($0.507 \times (1 - 1.96 \times \sigma_\alpha) = 0.202$) premium and another 5% would pay more than a 0.81 ($0.507 \times (1 + 1.96 \times \sigma_\alpha) = 0.812$) premium.

The estimated mean of the outside option, μ_{U_0}, measured relative to the quality level of a South or East Asian, low-score, low-quality seller is somewhat lower than the average value

Table 7: Buyers' Tastes and Quality Levels

	Score Group	Quality Group	Estimated Parameters	Standard Errors
Estimated Quality Levels				
North America	Low	1	-0.016 ***	(0.007)
North America	Low	2	0.413 ***	(0.009)
North America	Medium	1	-0.016 ***	(0.008)
North America	Medium	2	0.433 ***	(0.008)
North America	High	1	-0.016 ***	(0.003)
North America	High	2	0.507 ***	(0.004)
Eastern Europe	Low	1	0.263 ***	(0.003)
Eastern Europe	Low	2	0.625 ***	(0.005)
Eastern Europe	Medium	1	-0.103 ***	(0.005)
Eastern Europe	Medium	2	0.255 ***	(0.003)
Eastern Europe	Medium	2	0.648 ***	(0.003)
Eastern Europe	High	1	-0.107 ***	(0.006)
Eastern Europe	High	2	0.263 ***	(0.005)
Eastern Europe	High	3	0.668 ***	(0.004)
South and East Asia	Low	1	normalized to 0	
South and East Asia	Low	2	0.089 ***	(0.008)
South and East Asia	Low	3	0.449 ***	(0.008)
South and East Asia	Medium	1	-0.019 ***	(0.003)
South and East Asia	Medium	2	0.105 ***	(0.007)
South and East Asia	Medium	3	0.544 ***	(0.006)
South and East Asia	High	1	0.105 ***	(0.004)
South and East Asia	High	2	0.556 ***	(0.007)
Other Parameters				
$\log(\sigma_\epsilon)$			-0.915 ***	(0.041)
$\log(\sigma_\alpha)$			-1.118 ***	(0.012)
μ_{U_0}			-1.841 ***	(0.035)
$\log(\sigma_{U_0})$			-0.329 ***	(0.046)
σ_{α,U_0}			0.242 ***	(0.063)

The quality level for South and East Asia, low score, $Q = 1$, is normalized to be zero. The columns in the table show the estimated coefficients and corresponding standard errors for our baseline specification in which the distributions of qualities for transitory and permanent sellers are restricted to be equal. The stars, $***$, indicate that a coefficient is significant at the 95% significance level.

from this inside option.[38] The variance of the outside option is larger than the variance of the stochastic match component (ϵ). In our sample the outside option is positively correlated with

[38]The relevant average price scaled by the project size is 1.34.

price sensitivity, i.e., buyers with the high outside option also tend to be more price sensitive.

Fit to the Data. The estimates reported in Table 7 allow us to predict empirical market shares of different seller groups with a precision of one to two percentage points. In addition, using Efron definition of pseudo R^2, we determine that our model explains auction outcomes at the level of individual bidder for 75% of our observations.[39] This is a big improvement relative to the multinomial logit estimates reported in Section 3 that explain auction outcomes only for 18% of observations.

It is also worth noting that the estimated distribution of transitory sellers' bids and their participation probabilities (reported in the Web Supplement to the paper) indicate a statistically significant dependence of these objects on the transitory sellers' quality levels. Our estimates, therefore, support the assumptions of our model as well as validate our identification strategy.

To summarize, our estimates indicate significant difference in quality levels across sellers. In addition, accounting for unobserved quality substantially improves the fit of the model to the data which indicates that quality plays an important role in this environment.

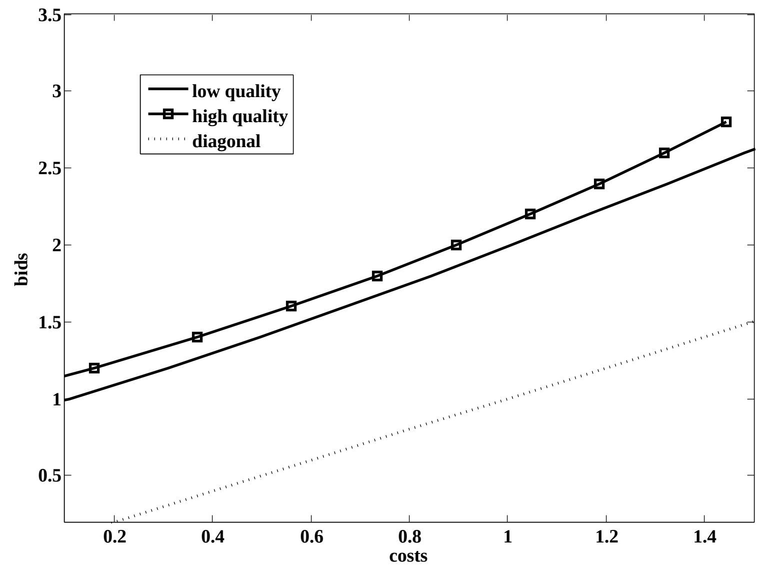
Alternative Specifications. We have estimated two alternative specifications which differ from the baseline specification in their treatment of transitory sellers. The results for these specifications are reported in the Web Supplement.

The first specification aims to demonstrate the ability of our methodology to handle more general specifications. In particular, it allows the distributions of the transitory and permanent sellers' qualities potentially to be different. Under this specification the frequencies of different quality groups in the population of transitory sellers are estimated from the data. As the results in the Web Supplement show, the methodology performs quite well and the parameter estimates obtained in the context of this more general specification are broadly consistent with baseline estimates.

The second specification maintains the 'no unobserved heterogeneity' assumption for transitory sellers. The estimates obtained under this specification are less plausible. This specification also performs poorly in terms of fit to the data. Indeed, both the baseline specification and the first specification (which permits unobserved heterogeneity for transitory sellers) predict quite precisely the probability that a project is allocated to a transitory seller. This probability is equal to 0.38 in the data (see Table 4) whereas the predicted probabilities computed from the baseline estimates and the estimates obtained under specification one are 0.36 and 0.41 respectively. However, the probability computed from the specification without unobserved heterogeneity of transitory sellers (0.23) substantially underpredicts the probability observed in the data. On the

[39] Efron pseudo R^2 is defined as $R^2 = 1 - \sum_{l,j}(y_{l,j} - \pi_{l,j})^2 / \sum_{l,j}(y_{l,j} - \bar{y})^2$ where $y_{l,j}$ is an indicator variable which is equal to one if bidder j wins auction l; $\pi_{l,j}$ is the predicted probability that bidder j wins auction l; and $\bar{y}$ is the predicted probability which ignores bidder heterogeneity.

Figure 1: Bid Functions



The figure shows the equilibrium bidding strategies of North American permanent sellers with medium score levels. Bidding strategies are recovered from the first order conditions of the bidders' optimization program. The convexity at the upper end of the costs' support arises due to the presence of stochastic component in buyers' tastes.

basis of these results we conclude that the assumption that buyers are not informed about the qualities of transitory sellers does not appear to be consistent with the data.

Pricing Strategies and Project Cost Distribution. We recover the sellers' bidding strategies and the distributions of project costs following the approach summarized in Section 5.

Figure 1 demonstrates the properties we observe in the estimated bid functions across countries and score levels. The full set of estimated bid functions is reported in the Web Supplement. We find that the estimated bid functions are increasing in costs, which is consistent with the theoretical predictions for the environment with private values. The low quality group always follows the most aggressive bidding strategy. The mark-ups over the sellers' costs change very slowly with the cost level and, in fact, for some groups increase as costs reach the upper end of the support. This feature arises because the buyer's choice is based in part on a purely stochastic (from the seller's point of view) component, $\epsilon_{j,l}$. At the high cost realizations where the seller's ability to compete on price is low, his probability of winning increasingly depends on the realization of the $\epsilon_{j,l}$ component, which in turn induces him to choose less aggressive bids. This effect essentially reflects the "gambling" behavior of bidders in the presence of uncertainty about the allocation rule used by buyers. In general, stochasticity plays an important role in our environment: sellers are uncertain about buyers' utility weights as well as their actual competition. This accounts for the relatively large mark-ups we document in our environment.

The estimated project cost distributions (means and standard deviations are reported in the Web Supplement) are typically "increasing" in sellers' quality. More specifically, the cost distribution of the high-quality group is always shifted to the right relative to the distribution of the medium-quality group. However, the low-quality group often has costs that are comparable to the costs of the high-quality group. Further, the distribution of costs differ across countries. In general, high quality US sellers tend to have higher costs relative to foreign sellers of comparable rank.

Many researchers have commented on the fact that the distributions of bids in the online setting tend to have high variance relative to other environments. We document similar regularity: the estimated standard deviation of the bid distribution (reported in the Web Supplement) is equal to 0.24. Notice that the estimated project cost distributions have substantially lower standard deviations relative to the standard deviations of the bid distributions. Thus, our model is capable of rationalizing the highly variable pricing environment through reasonably tight cost distributions. The "gambling" property of the bid functions described above explains this effect. Indeed, convexity or increasing mark-up near the end of the support induces a high variance in sellers' prices and also explains the presence of high bids in this environment.

The Distribution of Entry Costs. The estimated means and standard deviations of the distributions of entry costs are reported in the Web Supplement. The results indicate that North American sellers tend to have lower entry costs relative to sellers from other country groups. At the same time the entry costs are very similar for various average reputation score levels and residual quality groups within country. We use these estimates to assess that the participation costs incurred by entrants in this market constitute around 8% to 12% of the mean project costs. This number is slightly higher than that documented in other markets.[40] The relatively large entry costs estimated in this market may reflect large opportunity costs and the fact that active bidding for a project involves substantial interaction with the buyer and possibly the preparation of supplementary materials.

Last, we would like to comment on the limitations of the analysis presented in this section. In this analysis, we take the seller's reputation score as given and ignore the possible dynamic considerations associated with reputation building. To mitigate this concern, we base our estimation of the distribution of the sellers' costs on the optimization problem of a permanent seller. While permanent sellers may still take reputation-related concerns into account, the incentives associated with these concerns are likely to be quite weak. A single score does not make a large impact on the average reputation score once a seller has completed three or more projects. In-

[40]Studies of the US highway procurement market have estimated entry costs to be around $2-5\%$ of the engineer's estimate.

deed, we find through the data analysis that a bad score does not make a statistically significant impact on the probability of winning or on the bid of an established seller.

7 Buyers' Gains from Market Globalization

We use the estimated parameters to evaluate the average gain in value over the outside option collected by buyers in our market using the following measure:

$$\mathbf{E}\left[\max_{i\in A_l\cup\oslash} U_{i,l} - U_{0,l}\right] = \frac{1}{L}\sum_l \mathbf{E}_{\alpha_l,\epsilon_l,U_{0,l},Q_l^t|b_l^t}\left[\max_{i\in A_l\cup\oslash} U_{i,l} - U_{0,l}\right]$$

Here denotes an outside option and $\max_{i\in A_l\cup\oslash} U_{i,l}$ represents the buyer's utility collected from participating in the online market.

Recall that bids are scaled by the size of the project; thus, the welfare gain is measured as a fraction of the project size. We find that the buyers who had access to this market on average are able to improve their welfare relative to the outside option by 73% of the project value. The outside option in our setting represents a traditional procurement process, which implies hiring somebody locally or not hiring anyone at all. In this case our measurement captures the value of the Internet as an alternative marketplace.

This assessment has a number of caveats. First, we are working with a selected set of buyers who perhaps are best able to extract value from the online market. It is possible that the general buyer population still perceives an Internet transaction as high-cost (perhaps in terms of psychic cost) and prefers to use traditional markets. So, perhaps, our finding mostly applies to the "sophisticated" segment of the demand. Second, the offline markets are likely to respond to the emergence of the online market by adjusting prices or product selection. In such a case our measurement would provide a lower bound on the gains to buyers from the Internet since we assess buyers' gains relative to such improved outside options. Finally, the outside option may potentially include using an alternative online platform or re-auctioning the project on our platform but to a different set of sellers. Even if these concerns are valid it would only indicate that our measurement may underestimate the value of this market.

Welfare Gains from International Trade. We expect that welfare gains created by the Internet arise in part because it provides opportunity for international trade. This potentially creates multiple benefits. First, the number of sellers participating in the market might increase and this would intensify competition and potentially result in lower prices. Further, our estimates indicate that international sellers attracted to this market tend to have lower costs conditional on quality and higher quality levels within the quality rank. Thus, through the Internet, buyers may be able to access these low-cost and high-quality options that are not otherwise available to them. In addition, the presence of foreign (more competitive) sellers may put downward pressure on the prices in this online market.

In this section we assess the importance of sellers' heterogeneity as a contributing factor to the gains from globalization (the last two effects) since this issue is central to our paper. Specifically, we compare buyers' welfare under the market conditions in the data to the welfare which obtains when country affiliations of all foreign potential bidders are changed to US country affiliation. In practical terms, when implementing this analysis we keep the long-run average score group, and the quality rank (high, medium or low) of foreign sellers fixed but replace their quality levels and their distributions of private costs (both participation and project costs) with those which are estimated for the US sellers of the same average score group and the same quality rank. Thus, this experiment is implemented in such a way that the number of potential bidders remains unchanged but the set of potential bidders is homogenized in terms of quality levels and the distributions of costs.[41,42] We then solve for the equilibrium participation and bidding strategies for the baseline and counterfactual environments. These strategies are used to simulate average auction outcomes which are reported in columns one and two of Table 8 respectively.[43]

The change in the country affiliation of potential bidders has a substantial negative impact on the market. Specifically, the overall participation under the counterfactual scenario is reduced and the set of active sellers has a higher proportion of low quality sellers relative to the overall participation and the composition of sellers under the baseline market structure.[44] As a result of these participation outcomes, the prices charged by high quality sellers are higher on average under the counterfactual scenario. Thus, buyers in this setting are presented with fewer choices

[41]In this analysis we additionally replace all the transitory sellers with permanent sellers (both in the baseline and counterfactual scenario) to simplify the task of solving the model. We believe that this does not impact our assessment of the variety effect to an important degree since transitory sellers are substantively very similar to the permanent sellers in our setting, although their presence causes a number of methodological difficulties in the empirical analysis.

[42]Since medium quality is not present among the US sellers, we replace the medium-quality level by the low- and high-quality levels while maintaining the relative proportions of the high and low quality sellers constant. We verify the impact of this assumption by re-computing equilibrium outcomes for the intermediate step where the set of the quality levels is reduced to "high" and "low" while the original country affiliation is preserved for all potential bidders. The results of this analysis are reported in the last column of Table 8.

[43]This assessment reflects only short-run benefits since entry/exit of US buyers and sellers in the absence of international trade may result in the change of numbers as well as of composition of participating sellers. We leave the analysis of this latter effect for the future research since it requires conceptually different modeling framework as well as data on sellers opportunities in the offline markets.

[44]This regularity arises because the high quality US sellers have higher costs on average and slightly lower quality levels in comparison to the high quality foreign sellers. Thus, the post-entry profit that high quality sellers are able to obtain in the counterfactual scenario is lower than the post-entry profit achieved by high quality foreign sellers in the baseline scenario. This effect is somewhat mitigated by the higher participation of low quality sellers under counterfactual scenario relative to the baseline scenario (since the low quality US sellers have lower costs relative to the low quality foreign sellers) and by the fact that the US sellers of all quality levels have lower entry costs relative to the foreign sellers.

Table 8: Welfare Gain from International Internet Trade

	All Groups	US sellers Only	Low and High Quality Only
Participation (average number of bidders):			
Total	3.813	2.151	4.104
Low Quality	0.380	0.781	0.501
Medium Quality	1.151	-	-
High Quality	2.281	1.37	3.603
Allocation (share of projects):			
Low Quality	0.067	0.283	0.081
Medium Quality	0.279	-	-
High Quality	0.610	0.569	0.884
Outside Option	0.043	0.148	0.040
Prices:			
Average	1.563	1.626	1.631
Low Quality	1.647	1.450	1.616
Medium Quality	1.412	-	-
High Quality	1.607	1.726	1.651
Welfare Measures:			
Average Buyers Surplus	0.746	0.565	0.771
Average Seller Profit before Entry Costs	0.266	0.180	0.274
Average Profit (Low Quality)	0.067	0.098	0.065
Average Profit (Medium Quality)	0.147	-	-
Average Profit (High Quality)	0.315	0.220	0.350
Fixed Participation:			
Average Price	-	1.643	-
Price (Low Quality)	-	1.513	-
Price (High Quality)	-	1.680	-
Average Buyers Surplus	-	0.744	-

This table reports the results of a counterfactual analysis investigating the welfare gains to US buyers from access to the international market. The first column presents the results for the benchmark setting when all quality groups are present. The second column presents the outcomes from the setting where foreign potential bidders are replaced by US potential bidders while preserving sellers' quality ranks (medium-quality is replaced by high- and low-quality while preserving original shares of these quality levels in population). The last column is included for comparison purposes. It reports the results for the intermediate step where medium-quality potential bidders are replaced by high- and medium-quality potential bidders without changing the seller's country of origin. Average prices are computed as a share-weighted average of submitted bids. Buyer surplus is measured relative to the expected value of an outside option.

and the choices are less attractive (lower quality at higher price) than in the baseline scenario.[45]

[45] Prices charged by low quality sellers are lower than in baseline case but in general the choices available to buyers are less attractive than before: choice sets contain higher fraction of low quality sellers who even with low prices deliver lower surplus in comparison to high quality sellers.

As a result buyers' surplus conditional on purchase declines and the fraction of buyers choosing the outside option increases from 3% to 14.5%. The overall buyer surplus declines by 32% relative to the baseline scenario.

The analysis above highlights the importance of the sellers' participation in this market. In fact, the main channel through which elimination of 'foreign' variety of sellers impacts the market is through the reduction in participation which is driven not by the reduction in the *number* of potential bidders but by the change in the *characteristics* of potential bidders. We illustrate this point by re-computing the equilibrium while holding participation frequency fixed at the levels arising in the intermediate step. Under this restriction the reduction in the buyers' surplus relative to the level under the intermediate step is only 3.63%.[46]

The algorithm we use to solve for bidding and participation strategies combines a numeric method, which relies on the local approximation of seller objective function by means of Taylor expansion (first proposed by Marshall, Meurer, Richard, and Stromquist (1994)), with the projection method which allows us to solve for the whole vector of Taylor coefficients at a given grid point simultaneously.[47] Participation strategies, which are type-specific, are summarized by the equilibrium probabilities of participation. We use a system of equations similar to the one in equation (31) (Part D4 of the Web Supplement) to solve for participation probabilities imposing that the seller entry threshold is given by expected profit conditional on participation which depends on the competitors' participation strategies.[48] The full details are reported in the Web Supplement.

8 Conclusion

This paper makes a two-fold contribution to the literature. First, it develops a tractable framework that enables the analysis of online markets and, specifically, allows researchers to account for unobserved seller heterogeneity characterizing the data generated by these markets. Second,

[46]This result arises because under fixed participation the difference in prices is substantially reduced (the price of high quality is 1.680 comparing to 1.726 under the counterfactual scenario with full adjustment in participation) and the differences in the size and the composition of the buyers' choice sets are eliminated. The reduction in surplus occurs because the overall price level and the prices charged by high quality sellers remain somewhat higher relative to those in the intermediate step since they are based on less advantageous cost distributions and quality levels of the US sellers.

[47]The use of the projection method is necessary because the first order condition in our case contains the competitors' inverse bid functions evaluated at several different points on the support. This is in contrast to the standard case where a single bid level is present. As a result, the existing iterative procedure is not feasible in our setting. On the other hand, in our algorithm, the projection method is applied locally, which is why, it retains the precision and robustness properties that are so attractive in the algorithms based on local approximation.

[48]It is well known that multiple equilibria in participation strategies may arise in the settings such as the one we study. This issue does not impact our estimates since we are not solving the model in estimation. This problem might potentially affect our counterfactual analysis. To address this issue, we re-solve the model using 500 different starting points. We do not find any indication of multiple equilibria.

it exploits the structure of an online service market to provide an assessments of the welfare gains associated with the globalization of trade in services facilitated by the Internet.

We find that the gains to the buyers are quite substantial, at 73% of the project value. The paper emphasizes two channels through which globalization impacts the buyers' welfare: the increase in the variety of available seller types and the competitive effect of the presence of low-cost providers. The analysis of these effects is enabled by our methodology, which allows us to account for the sellers' quality differences that are not observable in the data and to obtain unbiased estimates for the distribution of the buyers' weights, outside options, and the distribution of sellers' costs conditional on sellers' characteristics (observable and unobservable) in the presence of potential endogeneity of sellers' observable attributes and prices.

The methodological part of the paper contains several innovative steps. First, we deviate importantly from the traditional discrete choice approach by structuring our estimation in two steps such that the unobserved group structure of permanent sellers is recovered in the first step and then is subsequently used in the second step to facilitate the identification of the distribution of the buyers' weights and outside options as well as to relieve the computational burden associated with accounting for endogeneity of transitory sellers' observable characteristics and prices. An important insight underlying this procedure is that the unobserved group structure could be recovered separately from the estimation of the buyers' components.

Second, our estimation procedure does not rely on the moments that condition on the buyers' choice set, as is typical in discrete choice estimation. Instead, we exploit moment conditions that aggregate over the choice sets that have certain common properties. This is necessitated by the presence of transitory sellers and buyer-specific choice sets that are prevalent in our setting.

Third, our estimation approach leverages a large amount of data typically available to a researcher in the Internet markets. Specifically, we are able to uncover permanent sellers (unobserved) groupings nonparametrically in the presence of selection into participation (especially by transitory sellers whose selection is difficult to control for in estimation) by conditioning on buyers' and sellers' observable characteristics in estimation.

We obtain a number of important insights into the operation of online procurement markets. To the best of our knowledge, our paper is the first one to inquire into the competitive implications of the market's organization in the form of multi-attribute auctions, which is becoming prevalent in the online (as well as offline) procurement markets. Specifically, we document "gambling"-motivated pricing at high cost realizations that arises due to uncertainty about the buyer's allocation rule. This regularity works well in rationalizing the high variability of prices in our data and is likely to explain similar price variability that has been observed in other online markets.

Some of the features of our setting, such as buyer-specific choice sets and self-selection into participation by agents, have been previously addressed in the context of college choice, online marriage markets, enrollment in residential medical training and other environments with a matching component. Our environment differs from these settings in that in addition to unobserved heterogeneity of the supply-side agents we also have private information on both sides of the market, and that the pricing, which is based on private information, has an important impact on equilibrium outcomes. Further, the large number and anonymity of players on the demand side in our setting facilitates identification of unobserved player heterogeneity despite the endogeneity of the sellers' participation and pricing decisions. In this we have advantage over other studies where researchers have to impose stronger assumptions in order to uncover primitives in the presence of unobserved heterogeneity.

To the best of our knowledge, this paper marks the first effort to estimate a tractable model of the online procurement market. Consequently, we focus on the factors we believe are of the first-order importance - unobserved heterogeneity of the sellers, private information of the sellers about their costs and private information of the buyers about the weights they use and their outside options – while making simplifying assumptions about the issues that are likely to be less important. We expect that the basic insights of our methodology will carry over to richer settings that elaborate on these issues in future research.

References

Ackerberg, D., K. Hirano, and Q. Shahriar (2006): "The Buy-it-Now Option, Risk Aversion and Impatientce in Empirical Model of eBay Bidding," Working paper, University of Michigan.

——— (2014): "Identification of Time and Risk Preference in Buy Price Auctions," Working paper, University of Michigan.

Asker, J., and E. Cantillon (2008): "Properties of Scoring Auctions," *RAND Journal of Economics*, 39.

——— (2010): "Procurement when Both Price and Quality Matter," *RAND Journal of Economics*, 41.

Athey, S., and P. A. Haile (2002): "Identification of Standard Auction Models," *Econometrica*, 70(6), 2107–2140.

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

Athey, S., and D. Nekipelov (2012): "A Structural Model of Sponsored Search Advertizing

Auctions," Working paper, Stanford University.

Backus, M., and G. Lewis (2012): "A Demand System for a Dynamic Auction Market with Directed Search," Working paper, Harvard University.

Bajari, P., and A. Hortacsu (2003): "The Winner's Curse, Reserve Prices, and Endogenous Entry: Empirical Insights from eBay Auctions," *RAND Journal of Economics*, 34(2), 329–355.

——— (2004): "Economic Insights form Internet Auctions: A Survey," *Journal of Economic Literature*, 42(June), 457–486.

Bajari, P., and G. Lewis (2011): "Procurement with Time Incentives," *Quarterly Journal of Economics*, 126.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile prices in market equilibrium," *Econometrica*, 63, 841–890.

Cabral, L., and A. Hortacsu (2010): "Dynamics of Seller Reputation: Theory and Evidence from eBay," *Journal of Industrial Economics*, 58(1), 54–78.

Chiappori, P.-A., and B. Salanie (2001): "Testing for Asymmetric Information in Insurance Markets," *Journal of Political Economy*, 108, 56–78.

Ciliberto, F., and E. Tamer (2009): "Market Structure and Multiple Equilibria in the Airline Industry," *Econometrica*, 77, 1791–1828.

Decarolis, F., M. Goldmanis, and A. Penta (2014): "Common Agency and Coordinated Bids in Sponsored Search Auctions," Working paper, Boston University.

Greenstein, S. (1993): "Did Installed Base Give an Incumbent any (Measurable) Advantage in Federal Computer Procurement," *Rand Journal of Economics*, 24(1), 19–39.

——— (1995): "Sole-Sourcing versus Competitive Bidding: U.S. Government Agencies' Procedural Choices for Mainframe Computer Procurement," *Journal of Industrial Economics*, XLIII(2), 125–140.

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

Heckman, J., and B. Singer (1984): "A Method of Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52, 271–320.

Hendricks, K., and A. Sorensen (2014): "Evaluating the Role of Intermediary in On-line Auctions," Working paper, University of Wisconsin-Madison.

Herrendorf, B., R. Rogerson, and A. Valentinyi (2009): "Two Perspectives on Preferences and Structural Transformation," Working paper 15416, NBER.

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

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

Krasnokutskaya, E. (2011): "Identification and Estimation in Highway Procurement Auctions under Unobserved Auction Heterogeneity," *Review of Economic Studies*, 28, 293–323.

Krasnokutskaya, E., and K. Seim (2011): "Bid Preference Programs and Participation in Highway Procurement," *American Economic Review*, 101.

Krasnokutskaya, E., K. Song, and X. Tang (2016): "Estimating Unobserved Agent Heterogeneity Using Pairwise Comparisons," Working paper, Johns Hopkins University.

Lee, S., K. Song, and Y.-J. Whang (2015): "Testing for a General Class of Functional Inequalities," *arXiv:1311.1595v4 [math.ST]*.

Lewis, G. (2011): "Asymmetric Information, Adverse Selection and Online Disclosure: The Case of eBay Motors," *American Economic Review*, 101(4), 1535–1546.

Li, T., I. Perrigne, and Q. Vuong (2000): "Conditionally independent private information in OCS wildcat auctions.," *Journal of Econometrics*, 98(1), 129–161.

——— (2002): "Structural Estimation of the Affiliated Private Value Auction Model," *RAND Journal of Economics*, 33(2), 171–193.

Marion, J. (2007): "Are bid preferences benign? The effect of small business subsidies in highway procurement auctions," *Journal of Public Economics*, 91(7-8), 1591–1624.

Marshall, R. C., M. J. Meurer, J.-F. Richard, and W. Stromquist (1994): "Numerical Analysis of Asymmetric First Price Auctions," *Games and Economic Behavior*, 7(2), 193–220.

Rao, B. (1992): *Identifiability in Stochastic Models: Characterization of Probability Distributions*. Academic Press, New York.

Swinkels, J. (2009): "First and Second Price Mechanisms in Procurement and Other Asymmetric Auctions," *Working Paper*.