The Role of Quality in On-Line Service Markets

Elena Krasnokutskaya†
Johns Hopkins University

Kyungchul Song‡
University of British Columbia

Xun Tang§
University of Pennsylvania

August 12, 2014

Abstract

We use data from an on-line market for programming services to assess buyers’ welfare gains associated with the globalization of this service market enabled by the Internet. Our study exploits the design of this market where projects are allocated through multi-attribute auctions, a mechanism that takes into account a seller’s non-price characteristics as well as his bid. We focus on the increased variety and competitive effects associated with the presence of low cost sellers as the main welfare-improving consequences of globalization. The paper proposes an empirical methodology to recover primitives of the model in the presence of unobserved seller heterogeneity. The methodology is designed to accommodate two important features of such markets: buyer-specific choice sets and the high turnover of sellers. We find that the Internet enables buyers to substantially improve on their outside (local) option, with the larger 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: July 31, 2014. 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: xuntang@sas.upenn.edu.
1 Introduction

Until recently, markets for professional services\(^1\) were necessarily 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. Our objective in this paper is to understand 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 on-line. In addition to answering the main substantive question, our analysis delivers a number of novel insights into the operation of these fast growing but relatively little studied markets.

Our analysis is based on data from a prominent on-line procurement market for programming services. Transactions in this market are implemented in the form of multi-attribute auctions\(^2\) that allow buyers to deviate from allocation based solely on price (as in standard auctions) in favor of service providers who are associated with higher buyer-specific value. Indeed, in the data, buyers frequently chose sellers that charge prices above the lowest price submitted in the auction. Moreover, 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 researchers. This is not surprising as the platform encourages and facilitates extensive buyer-seller communication related to seller’s qualifications and examples of his past work. The methodological contribution of our paper is to propose an identification and estimation strategy that accommodates the data structures commonly available from on-line platforms.

We begin by developing a tractable framework that can be used to answer our question of interest. We formalize the features of this market in a model where each project attracts a set of sellers who submit bids for a buyer’s consideration. The project is awarded to a seller who delivers the highest value over price if it exceeds the buyer’s outside option; otherwise, the project is not awarded. The buyer’s valuation for a given seller is a function of the seller’s characteristics evaluated according to a vector of buyer-specific weights (which are the buyer’s private information and thus unobserved by sellers and a researcher).\(^3\) We use data on buyers’ choices to recover the distribution of buyers’ weights and outside options. Since the buyer’s outside option includes hiring from the off-line local market, the difference in the net value from hiring in this market and the outside option provides a lower bound on the gains from

\(^1\)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)).


\(^3\)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), Krasokutskaia 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 Cantillon (2010), Asker and Cantillon (2008), and Bajari and Lewis (2011). While the multi-attribute auctions format is also prevalent in off-line industry procurement, it is little studied, with the exception of Greenstein (1993, 1995).
market globalization. Further, we recover the distribution of sellers’ costs conditional on their characteristics. This enables us to assess through counterfactual analysis the relative importance of different channels contributing to buyers’ welfare gains from on-line markets as well as to assess the importance of using a multi-attribute rather than a standard auction in organizing such a market.\footnote{The results of the latter analysis are presented in the supplemental on-line appendix.}

This analysis assumes full information on the buyers part about the sellers characteristics as well as the risk neutrality of buyers. We believe such an approach is justified in the on-line setting such as the one we study, since the on-line platform is specifically designed to minimize buyers’ uncertainty about sellers’ characteristics and to protect participants from the ex post risks. In particular, it maintains a database of performance-related measures, provides an arbitration service, and administers payments from an escrow account only after a buyer is satisfied with the delivered service. Because of this, informational concerns do not appear to be of first-order importance in this market: it is very likely that buyers are able to obtain sufficient information and their risk aversion is not likely to impact their decisions to a large degree. Our estimation results reported below will provide additional evidence supporting this assessment.

Further, buyers may be specifically interested in certain seller characteristics, e.g. the seller’s country of origin may indicate or assuage concerns about 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 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.”

To answer the questions motivating the paper, we need to develop a methodology that would allow us to recover unobserved sellers’ qualities, as well as other primitives of the model, allowing for the possibility that sellers’ prices, costs and possibly other characteristics may depend on the unobserved quality component. The discrete choice literature deals with the issue of recovering the unobserved product quality as well as the distribution of the buyers’ tastes for the product’s characteristics. However, the methodologies developed in the context of this literature cannot be readily extended to on-line markets because they rely on the data structure whereby researchers observe the decisions of a large number of consumers operating in the same market and, therefore, choosing from the same set of products. In contrast, the data structure observed in on-line markets is typical of an auction environment. It is characterized by buyer-specific choice sets (or an auction-specific set of participants), sellers’ self selection into participation (and thus into the buyer’s choice set), as well as by the presence of a large number of sellers who are observed to enter the market for only a very short time. These sellers typically participate in a small number of auctions and leave the market after winning only one or two projects. We refer to such sellers as “transitory”\footnote{In the auction literature such sellers are sometimes referred to as “fringe” sellers.} as opposed to “permanent” sellers who are observed to participate in many auctions and may be assumed to remain in the market indefinitely. These settings are also characterized by the private variation in sellers’ costs both within and across projects.

Let us elaborate on the methodological difficulties associated with the data structures typically observed in on-line markets. First, let us consider a case when buyers’ information about transitory sellers’ qualities is limited to the variables recorded in the data. It would seem that the methodologies developed to estimate discrete choice models\footnote{See McFadden (1989), Nevo (2001), Berry, Levinsohn, and Pakes (1995), and Berry, Levinsohn, and Pakes (2004).} could be used to recover the
distribution of weights (random coefficients) and permanent sellers’ fixed effects. The main “building block” in these methodologies, however, is the probability of choosing a particular alternative conditional on the choice set. It could be consistently estimated from the data typically used in discrete choice analysis. In an on-line setting an individual seller’s market share conditional on the choice set cannot be precisely estimated, since because of the buyer-specific choice sets and a very large number of sellers, the number of buyers choosing from exactly the same set of alternatives is always very small. In such a setting the estimation of the distribution of random coefficients and sellers’ fixed effects would necessarily rely on the probabilities of choice, which aggregate over many choice sets. Whether such aggregate moments are sufficient for identification of these objects remains an open question in the literature.

Further, the assumption of transitory sellers’ homogeneity conditional on observables is not appealing in the context of the market we study. The platform managing the market allows buyers to inquire into the qualities of competing sellers, either permanent or transitory. Thus, unobserved seller heterogeneity and the associated endogeneity concerns are likely to arise in the case of transitory sellers as much as in the case of permanent sellers. Moreover, the presence of and the competitive pressure from transitory sellers appear to be important in markets for services. For example, in our market every auction attracts several transitory sellers and projects are allocated to transitory sellers with high probability (60%), even in the presence of permanent bidders with comparable prices.

As fixed effects or BLP approaches cannot be used to deal with the unobserved heterogeneity of transitory sellers, since they participate in a very small number of auctions and because of the non-linearity of the model, an alternative approach would be to integrate it out in estimation while taking into account the dependence of transitory sellers’ bids on their qualities. The most tight link between the model and the data would be implemented if we solve for this relationship from the model and then use it to integrate transitory sellers out in estimation. Such an approach is computationally prohibitive if sellers’ heterogeneity is defined at the individual level, as one needs to solve a very large number of auctions with asymmetric bidders where the degree of bidder asymmetry depends on parameter values. Alternatively, we could follow a mixture approach and estimate some ad hoc parametric mixing distribution (of transitory sellers’ unobserved heterogeneity) jointly with other elements of the model. This last estimation approach has been shown to perform poorly.⁷

We develop a methodology to recover model components that allows for unobserved seller heterogeneity and works well with the data structures observed in on-line markets. We treat the unobserved qualities of permanent sellers as parameters of the model (fixed effects) and the qualities of transitory sellers as draws from the common distribution (random effects). Further, we summarize the distributions of unobserved qualities (both for permanent and transitory sellers) through a group structure where sellers within the same group are characterized by the same level of unobserved quality. Such a group structure arises naturally when the distribution of qualities is discrete. If it is continuous, our approach would require discretizing sellers’ characteristics.⁸ Our estimation methodology consists of two steps. In the first step we use the classification algorithm proposed in Krasnokutskaya, Song, and Tang (2014) to recover the group structure underlying the unobserved quality of permanent sellers. This step relies on pairwise comparisons of the long-run performance of permanent sellers.⁹ Then, in the second step we estimate the

---

⁸Such discretization is adopted in Chiappori and Salanie (2001) as well as in Ciliberto and Tamer (2009).
⁹In this step we build on the recent econometric literature on the estimation of discrete-valued parameters.
remaining primitives of the model through the Generalized Method of Moments procedure after imposing in the estimation the recovered group structure as well as the equality of the supports of the quality distributions of permanent and transitory sellers.

The first step enables us to avoid complications associated with aggregation over the choice sets raised above because the comparisons are pair-wise and because this step is implemented separately from the estimation of buyers’ weights and outside options. This allows us to organize the set of participants into a quality group structure that replaces the seller’s identity with his group affiliation and thus substantially reduces the dimensionality of the set of “products” considered by various buyers. In practice, however, this may still (and does) leave us with the number of possible choice sets that is large relative to the number of available observations. That is why, in contrast to the discrete choice methodology, we do not rely on the moments that fully condition on the buyers’ choice set either in the form of identities or as summarized by the estimated group structure. Instead, we propose an alternative identification strategy that exploits the residual variation in sellers’ costs remaining after conditioning on sellers’ characteristics. We use this variation in the context of a permanent seller’s probability of winning an auction where the set of participants possesses. We then rely on the structure of the sellers’ pricing problem within the context of an auction setting to recover the distribution of sellers’ costs conditional on all characteristics (observable and unobservable).  

Our method also provides a tractable way of dealing with the unobserved heterogeneity of transitory sellers. As we have already explained, the relationship between transitory sellers’ unobserved heterogeneity and their bids has to be integrated out in estimation. The most tight link between the model and the data would be implemented if we solve for this relationship from the model and then use it to integrate transitory sellers out in estimation. Our method makes such an approach feasible by reducing the dimensionality of unobserved heterogeneity and thus decreasing the number of asymmetric groups that could be present in any given auction as well as the number of different auctions that have to be solved during the estimation. On the other hand, our identification strategy suggests a way to recover the relationship between transitory sellers’ unobserved heterogeneity and their bids directly from the data. This approach exploits the variation in the composition of the sets of permanent auction participants as well as their prices in the second step of estimation. This substantially reduces the computational burden and ensures a more robust performance of the estimation routine at the expense of exploiting data more heavily. Our ability to control for the permanent participant’s group membership in the second step of estimation also takes care of the selection into the participation both for permanent and transitory sellers.

Our results indicate that seller heterogeneity is important in the market for programming services. Buyers are willing to pay a substantial premium to sellers from certain countries, and with higher levels of performance measures. However, the premiums associated with the variation in observable characteristics are relatively small compared to 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 an important variation in the unobserved seller quality conditional on observed characteristics, as well as significant differences in the distributions of this variable across different countries and levels of performance measures. In fact, the variation in this characteristic as well as in buyers’ preference for this characteristic

\footnote{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.}
accounts for more variation in the data than the variation of all other covariates combined.

We use the estimated parameters to evaluate the welfare gains to the buyers from the availability of this on-line market. We estimate the 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 generated by 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 access to international sellers facilitated by the Internet. We find that 65% of buyers’ gains from on-line trade are explained by access to international sellers who are capable of delivering higher quality at lower cost.

Further, the estimation results confirm our surmise that buyers are informed about the quality of transitory sellers\(^\text{11}\) since we recover a statistically significant relationship between transitory sellers’ bids and their unobserved characteristic. This relationship is not directly observed in the data and thus could be estimated from buyers’ choices only if buyers, in fact, are informed about transitory sellers’ qualities. This feature also plays an important role in rationalizing allocative decisions. In fact, the model that assumes that transitory sellers are identical conditional on their observed characteristics substantially under-predicts the probability that a project will be allocated to a transitory seller (37% instead of 60% in the data), whereas our model approximates this probability quite well (64%). Our model fits the data well in general: it is capable of explaining 75% of buyers’ choices in contrast to 25% of choices rationalized by the model without unobserved heterogeneity and 52% of choices rationalized by the model where permanent sellers are heterogeneous in an unobserved way but transitory sellers are homogeneous conditional on observables.

Our estimation results provide a number of interesting insights into the operation of an on-line market for programming services and potentially other on-line markets for services. For example, we find that uncertainty about the buyers’ allocation rule, which is inherent in multi-attribute auctions, induces gambling behavior by sellers with high cost realizations. This feature of the environment is capable of rationalizing the high-variance price distribution often observed in on-line markets by means of relatively tight cost distributions. We also find that while the distributions of costs generally appear to be stochastically increasing in unobserved quality (as well as in observed performance measures), a subset of sellers with the lowest quality levels appear to have costs comparable to the costs of high quality sellers. The latter speaks to the heterogeneity of the participants who are attracted to and are able to survive in on-line markets.

The paper is organized as follows. Section 2 describes our market; the basic model is described in Section 3. Section 4 discusses our empirical methodology. Section 5 describes the data and the results of the empirical analysis. Section 6 summarizes the findings and outlines directions for further research.

2 Market Description

We study a market mediated by an online platform that serves as a match-maker between the demand and supply for services of computer programming. This company provides an environment that allows buyers (the demand side) to post job announcements. It also maintains the registry of potential sellers (the supply side). The registry provides limited information on

\(^{11}\)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.”
verifiable “outside” credentials as well as information about the on-site performance of the seller. The latter includes 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. In the case of a dispute, the company provides professional arbitration services that ensure that a seller is paid if only if the completed job satisfies industry standards.

This intermediary company allocates jobs through multi-attribute auctions. Under the rules of such an auction, a buyer is allowed to take into account seller characteristics other than price. As a result, the selected seller is not necessarily the one who submits the lowest quote. An important feature of this mechanism is that the award rule is not announced and thus remains unknown to other market participants.

Suppliers can communicate with buyers before posting price quotes. Such an exchange of messages is very common. On average, each seller submits three messages per auction in our data. A seller may attach an example of his work or a sketch of the proposed code. The number and the content of these communications are not observed by other sellers. Hence, while the buyer has an opportunity to form an opinion about each sellers’ quality, competing sellers have much less knowledge of their competitors’ characteristics. However, competitors might be able infer a seller’s quality from his long-run rate of winning in a way similar to that proposed in this paper.

A seller who decides to submit a bid has no information about other sellers who are also submitting bids. His choice of price quote is thus based on his beliefs about potential competition.

Finally, most of the buyers in our dataset appear in this market only once. A large number of sellers stay in the market for a long time whereas a considerable fraction participates only in a few auctions and leaves the market. As mentioned in introduction we refer to the first type of sellers as permanent and to the second type of sellers as transitory.

3 The Basic Model

In the interest of transparency, we first present the main ideas underlying our empirical methodology in the context of a simplified model that leaves out several empirically relevant features. In particular, it ignores observable project and seller heterogeneity and assumes that sellers select projects to bid in completely at random, ignoring strategic considerations. We explain how our model and methodology are adjusted to incorporate these features in section 4.4.

3.1 Theoretical Framework

The set of sellers who operate in our service market, \( S \), consists of permanent and transitory sellers (these subsets are denoted \( S^p \subset S \) and \( S^t \subset S \) respectively). Each seller is characterized by a vertical scalar characteristic, \( q \), which admits a finite number of values, \( q^1 < q^2 < \cdots < q^K \), and induces partition of the sets \( S^p \) and \( S^t \) into groups \( S^{r,k} \), \( S^r = \bigcup_{k \leq K} S^{r,k} \), \( r \in \{p,t\} \), such that a group \( S^{r,k} \) represents a subset of sellers of type \( r \) (either permanent or transitory) who are characterized by the level \( q^k \). The frequency distributions of this characteristic is given by, \( \{q_k, \pi_{r,k}\}_{k=1}^K \) where \( \pi_{r,k} = \frac{|S^{r,k}|}{|S^r|} \) denotes the fraction of sellers of type \( r \) who belong to group \( k \). We refer to characteristic \( q \) as quality in the exposition below and assume that it does not change over time or over projects.\(^{12}\) In subsequent analysis we assume that this characteristic is

\(^{12}\)In this analysis we interpret quality as seller’s ability. We do not specifically address the possibility that seller’s quality for a given project may depend on seller’s effort. However, our methodology is not affected by
unobserved to a researcher.

We use \( A_l \) to denote the set of sellers who submit a bid for project \( l \) and refer to such sellers as active bidders.\(^{13}\) For simplicity, we first assume that a decision to become an active bidder, \( D_{l,t} \), is non-strategic, i.e., a seller \( i \in S^{r,k} \) becomes active for project \( l \) at random, with exogenous probability \( \lambda_{r,k} \), such that \( \Pr(D_{l,t} = 1) = \lambda_{r,k} \) for \( i \in S^{r,k} \). These probabilities are common knowledge among all market participants. We further assume that an active bidder does not observe who else is active in the same project.

Upon becoming active, each seller privately observes his cost \( C_{i,l} \in \mathbb{R}_+ \) for completing the project, quotes a bid/price \( B_{i,l} \) and reveals his quality \( q_i \) to the buyer. The costs of seller \( i \) from quality group \( S^{r,k} \) are distributed according to \( F^k_c(\cdot) \). The distributions of costs are common knowledge among all sellers.

The demand side of the market consists of one-time buyers who observe all the relevant sellers’ characteristics including (and restricted in this section to) \( q \) and are endowed with an outside option that delivers utility \( U_{0,l} \). They procure services using multi-attribute auctions. In particular, a buyer \( l \) associates a (private) value, \( \Delta_{l,i} \), with an active seller \( i \in A_l \) and awards his project to an active seller with the highest level of \( \Delta_{l,i} - b_{i,l} \) if this level exceeds \( U_{0,l} \) and otherwise leaves the project unassigned. The buyer’s value is a weighted average of seller’s attributes with buyer-specific weights \( \alpha_l \) and intercept \( \epsilon_{i,l} \) (the residual value assigned by buyer to a match with specific seller), i.e.,

\[
\Delta_{l,i} = \epsilon_{i,l} + \alpha_l q_i.
\]

We let \( \epsilon_l = \{\epsilon_{1,l}, \ldots, \epsilon_{|A_l|, l}\} \) and refer to \((\alpha_l, \epsilon_l)\) as the vector of buyers’ weights. In keeping with the definition of a multi-attribute auction, sellers do not observe the weights or outside option of a specific buyer, and consider it to be a random draw from some joint distribution of weights and outside options characterizing population of buyers.\(^{14}\)

In line with the existing empirical auction literature, we assume that the observed outcomes reflect a type-symmetric pure strategy Bayesian Nash equilibrium (psBNE).\(^{15}\) In such an equilibrium, participants who are \textit{ex ante} identical (i.e. those who belong to the same \( S^{r,k} \)) adopt the same strategy. Thus, the bidding strategy for seller \( i \) who belongs to the group \( S^{r,k} \) is denoted \( \sigma^{r,k} : [c_k, \bar{c}_k] \to \mathbb{R}_+ \), and entrant \( i \)’s expected profit from bidding \( b \), i.e. submitting \( \sigma^{r,k}(c) = b \) when the cost draw is equal to \( c \) is given by

\[
\Pi^{r,k}(b, c; \sigma^{-i}) \equiv (b - c) \Pr(i \text{ wins } | b, i \in S^{r,k}; \sigma^{-i}),
\]

this distinction. In the model we outline below the optimal effort level would be a function of sellers’ ability (or project’s and seller’s observable characteristics as well as sellers’ quality). Because of this property the methodology proposed in this paper could be used to recover the quality levels produced by sellers which reflects sellers’ optimal effort.

\(^{13}\)The set of active bidders is also partitioned into groups, \( A^p_l = A^p_l \cup A^q_l \) with \( A^p_l \cup_{k \leq K} A^q_r \).

\(^{14}\)Unobservability of \( \epsilon_{l,i} \) rules out common (or even correlated) between buyer and seller understanding of the specific seller’s suitability for a given project. This may appear to be restrictive in view of the extensive interaction between the buyer and the seller. However, from the theoretical point of view this component controls the size of (unsystematic part of) surplus generated by the specific buyer-seller match. The simulated solution of the game where \( \epsilon_{l,i} \) is known indicates that knowledge of \( \epsilon_{l,i} \) allows bidder to extract larger part of surplus. Thus, it is not in the interest of the buyer to share any information that would reveal \( \epsilon_{l,i} \) to seller \( i \). More specifically, the interview should be conducted in such a way as to elicit maximum information about seller’s suitability without revealing the match value to the seller.

\(^{15}\)We abstract away from sellers’ dynamic incentives associated with reputation building. This assumption does not impact the methodology for the estimation of buyers’ weights or sellers’ qualities. We provide detailed discussion of this issue in the empirical section of this paper when we recover the distribution of sellers’ costs.
where \( \sigma^{-i} \) denotes a profile of other sellers’ strategies that they would use should they become active in a given project, and \( \Pr(i \text{ wins } | b, i \in S^{r,k}; \sigma^{-i}) \) the probability of seller \( i \) winning the auction by bidding \( b \) and the other active sellers’ bids are consistent with the strategies \( \sigma^{-i} \). Notice, that expression in \( \Pr(i \text{ wins } | b, i \in S^{r,k}; \sigma^{-i}) \) implicitly includes integrating over possible sets of active competitors since information about active competitors is not available to seller \( i \) at the time when he decides on his bid.

Thus, a psBNE is a profile of strategies \( \{\sigma^{r,k}\}_{r \in \{p,t\}, k \in \{1,...,K\}} \) such that

\[
\sigma^{r,k}(c) = \arg \max_b \Pi^{r,k}(b, c; \sigma^{-i}) \text{ for all } c.
\]

We prove the existence of such a psBNE in the supplemental on-line appendix.\(^ {16} \)

We make several simplifying assumptions in this analysis such as that seller’s quality is constant across projects, that buyers have full knowledge of sellers’ qualities, and that sellers’ do not observe buyers’ weights. Our empirical results demonstrate that they work quite well in the setting we study. We also implement a number of robustness checks to verify the sensitivity of our analysis to these assumptions.

### 3.2 Statistical Details

In section 4.4 we introduce observable seller characteristics and explain how they impact buyer’s valuation of a seller. For now, however, our focus remains on \( q \) which in our model captures seller’s vertical characteristic unobserved to the researcher.

Recall that the main distinction between the transitory and permanent sellers is that the number of observations available for each transitory seller is small and finite in a sense that it does not grow as the number of auctions recorded in the data increases. In contrast, the number of observations for each permanent seller grows with the number of auctions. That is why, for the purpose of estimation we treat the quality of a permanent seller as a constant and the quality of a transitory seller as a draw from some distribution. To reflect this we use lowercase letter \( q_i \) to denote the quality of the permanent seller \( i \) and capital letter \( Q_j \) to denote the random variable representing the quality of a transitory seller \( j \). In accordance with our model unconditional distribution of \( Q_j \) in population is summarized by \( \{\Pr(Q_j = q_k) = \pi^{t,k}\}_{k=1,...,K} \).

In our empirical analysis we recover the group affiliation of each permanent seller, that is the map from his identity into his quality group, \( \kappa : S^p \to \{1,...,K\} \) such that \( q_i = q^{\kappa(i)} \) for every seller \( i \in S^p \), parameters \( \{q^k\}_{k=1,...,K} \) characterizing quality levels of these groups, and \( \{\Pr(Q_j = q^k) = \pi^{t,k}\}_{k=1,...,K} \).

Notice that the only restriction that we impose on the distributions of qualities for the sets of permanent and transitory sellers is that they have the same support. That is we assume that selection into long-run participation based on quality if it exists is not too severe as too exclude participation of some quality levels. We believe that this assumption is reasonable in our as well as in many other contexts. Indeed, in many markets factors other then quality, e.g. schedule flexibility and outside option influence sellers’ participation decisions.

The bidding strategies of both types of sellers depend on their qualities. We are able to control for the dependence between the bids and qualities of permanent sellers once the group affiliations of permanent sellers are recovered. In simple terms we use a group-specific dummy variable to control for the quality of an individual permanent seller. On the other hand the quality of a

\(^{16}\text{We extend this model to allow for strategic entry in section 4.4 and prove the existence of a psBNE for such an extended model in supplemental on-line appendix.}\)
transitory seller is a random variable realizations of which could not be recovered from the data. This introduces endogeneity problem for the prices of transitory sellers. We control for this endogeneity by integrating out the relationship between the quality of the transitory seller and his price using restrictions imposed by our model.

4 Estimation Methodology

The primitives of our model are non-parametrically identified as we argue in the section below. However, the complexity of our model makes full nonparametric estimation impractical. Instead, we propose a two-step estimation procedure where we recover the unobserved group structure of the set of permanent sellers through non-parametric classification procedure and then make parametric assumptions about the distribution of buyers’ weights, \((\alpha, U_0)\) and \(\epsilon\), and proceed to estimate the parameters of these distributions using the Generalized Method of Moments (GMM). In this section, we summarize classification procedure, outline identification mechanism, and explain how GMM is implemented in a computationally feasible manner.

4.1 Recovering Quality Group Structure

In the first step, we make use of a non-parametric classification procedure proposed in Krasnokutskaya, Song, and Tang (2014) to recover the mapping \(\kappa(.)\) and thus the quality group structure of the set of permanent sellers, \(\{S_{p,k}\}_{k=1,\ldots,K}\). This procedure is based on the pairwise testing of inequality restrictions which relies on the proposition below. It exploits differences in probability of winning across sellers with different levels of unobserved characteristic.

Intuitively, if \(i\) and \(j\) participate in two separate but ex-ante identical auctions (in terms of the realized set of competitors) and submit the same price then the seller with the higher value of \(q\) has the higher chance of winning. Note that the winner is not deterministic in the presence of uncertainty about buyers’ weights. The ranking of winning probabilities is preserved when aggregated over different sets of competitors as long as the probability of encountering a 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 \(j\).

To formulate the result more formally we need the the following two assumptions:

(A1) Sellers’ private costs \(C_{i,l}\) and the events of being active are independent across all \(i \in S\) and across \(l\). For each seller \(i\) with \(q_i = q^k\), his cost in each auction is an independent draw from continuous distribution \(F^k\) with a density positive over support \([c^k, \bar{c}^k]\).\(^{17}\) The events of being active are independent across the projects and the sellers.

(A2) The three random vectors \((\alpha_l, U_{0,l})\), \(\epsilon_l\) and \(C_l\) are mutually independent; match components \(\epsilon_{i,l}\) are i.i.d. across \(i\)’s; and \(\epsilon_{i,l}\) and \(\alpha_l, U_{0,l}\) are continuously distributed with a density positive over \([\varepsilon, \overline{\varepsilon}]\) and over \([0, \overline{\alpha}] \times [u_0, \overline{u}_0]\) respectively.\(^ {18}\)

\(^{17}\)This assumption does not allow for a persistent unobserved seller-specific cost component in addition to quality. This excludes, for example, differences in opportunity costs associated with sellers’ location (e.g. urban vs. rural) if it is not observed in the data. It might be possible to separately account for this type of unobserved seller heterogeneity since our current strategy identifies unobserved quality from buyers’ choices whereas unobserved cost persistence might be identified from the additional correlation in prices (unaccounted for by quality) over time. We leave this extension to future research.

\(^{18}\)Notice that we require that \(\epsilon_l\) is orthogonal to \((\alpha_l, U_{0,l})\), whereas \(\alpha_l\) and \(U_{0,l}\) are allowed to be dependent.
For the remainder of this section we drop subscript \( l \) (index for auctions/buyers) to simplify notation. Let \( B_i \) denote the support of prices submitted by a seller \( i \) in a psBNE. For any \( b \in B_i \cap B_j \), define a pair-specific index:

\[
 r_{i,j}(b) \equiv \Pr(i \text{ wins} \mid B_i = b, \ i \in A, \ j \not\in A).
\]  

This index reflects the probability that seller \( i \) wins an auction when submitting a bid \( b \) and when the set of his direct competitors does not include \( j \). The proposition below establishes pairwise ranking of bidders \( i \) and \( j \) on the basis of indices \( r_{i,j}(b) \) and \( r_{j,i}(b) \).

**Proposition 1** Under (A1)-(A2),

\[
\text{sign}(r_{i,j}(b) - r_{j,i}(b)) = \text{sign}(q_i - q_j)
\]

for any pair of permanent sellers \( i, j \) and all \( b \) in the interior of \( B_i \cap B_j \).

Since sellers’ ordering with respect to \( q \) is transitive in our model this result applied to a sufficiently large dataset generated by our model allows arranging all sellers in the order of (weakly) increasing quality. In other words, we are able to identify the quality group structure and group affiliations of permanent sellers.

The main issue that we need to overcome in order to translate this identification strategy into a viable estimation method is that, while ordering with respect to \( q \) is transitive in our model, the estimation based on pairwise comparisons may result in estimates that violate transitivity in small samples, even when there are only two quality groups in the population.

The nonparametric classification procedure we use proposes a method to estimate the whole group structure at once in a way that satisfies transitivity. Below, we provide a heuristic summary of how this is achieved in a simple case of two groups (corresponding to high and low values of \( q \)).

The idea is to divide the set of sellers into two groups such that sellers within each group are “closer” to each other than to sellers from the other group according to some metric which is based on index \( r_{i,j} \). More specifically, for each seller \( i \), we first divide the other sellers into two groups, one with sellers more likely to have higher quality than \( i \) and the other with sellers more likely to have lower quality than \( i \). This division is implemented by comparing \( p \)-values from a pairwise bootstrap test of the inequality restrictions \( r_{i,j}(b) \geq r_{j,i}(b) \) for all \( b \). Next, we check whether seller \( i \) is more likely to belong to the first group or to the second group. Thus for each seller \( i \), we estimate a separate group structure. We choose one of these structures to be our estimate of the underlying quality group structure so that the chosen structure has most empirical support (in terms of average \( p \)-values). The formal exposition of the classification method for a more general case of multiple quality groups can be found in Krasnokutskaya, Song, and Tang (2014).

---

19 The index with restriction \( \{i, j \in A\} \) is not monotone in bidders’ quality. In fact, under such restriction the ranking of \( r_{i,j}(b) \) and \( r_{j,i}(b) \) depends on the distributions of buyers’ weights.

20 Proposition 1 also holds if we relax (A2) to allow dependence between \( \alpha \) and \( \epsilon \) and only require \( \epsilon_i \) to be independent conditional on \( \alpha \).

21 Here \( \text{sign}(x) \equiv 1\{x > 0\} - 1\{x < 0\} \) for all \( x \in \mathbb{R} \).
The true number of the quality groups is usually unknown. Thus in estimation we use a consistent group number selection procedure which utilizes the following regularity.\(^{22}\) When the sample size is large, misspecifying the number of quality groups to be smaller than the true number of groups results in the weak empirical support of seller homogeneity for some of the estimated groups. On the other hand, when the number of quality groups is misspecified to be larger than or equal to the true number of groups, the group estimation does not show any sign of misspecification bias.

Note that once the quality group affiliations of sellers are known their identities are no longer important. For example, to condition on the realization of the set of active bidders no longer means to fix the list of bidders identities. Rather, it means to require that the set of active bidders should include specific numbers of sellers by type and quality group. Note that the quality group affiliation is not publicly observed for transitory sellers and thus only information on the overall number of transitory active bidders should be specified. In what follows we will use notation \(I_A\) to reflect such information about a given set of sellers \(A\).

4.2 Identification

We now discuss how to identify the remaining elements of the model, which include (a) the quality levels of unobserved groups, (b) the joint distribution of buyers’ weights for quality, \(\alpha\); match components, \(\epsilon_i\), and the outside option, \(U_0\), as well as (c) the conditional distribution of a transitory bidder’s quality given his bid and information about the set of his active competitors, \(\Pr(Q_j = q^k | B_j = b, I_A)\). The last object permits recovering the probabilities \(\Pr(Q_j = q^k)\), a primitive of the model, since the distribution of \((B_j = b, I_A)\) is observed in the data once the group structure of the set of permanent sellers is identified.

The main idea is that, the independence between buyers’ weights, outside option and sellers’ private costs ensures that the variation in permanent sellers’ bids conditional on seller’s unobserved quality group is exogenous and could be used to recover (a)-(c) from the variation in permanent bidders’ winning probabilities across different sets of active competitors. Here the sets of active competitors are characterized in terms of their group structure as explained at the end of the last section. We provide a summary of the identification mechanism below. The details of the formal argument could be found in the supplemental on-line appendix to the paper.

More specifically, the recovery of (a), (b) and (c) consists of three parts. First, in auctions where a permanent bidder wins in the presence of other permanent bidders from the same quality group, changes in his winning probability in response to the variation in prices submitted by permanent sellers identifies the distribution of match components \(\epsilon_i\) up to a location normalization. Next, with knowledge of the distribution of \(\epsilon_i\), similar changes in the winning probability for a permanent bidder when competing with other permanent bidders from different quality groups identify the quality levels up to a location normalization and the distribution of the weights for quality up to a scale normalization. Finally, the changes in the relation between the winning probability and the prices of permanent bidders that are driven by the variation in the prices of

\(^{22}\)The fact that the number of groups is recovered from data ensures that the assumption of quality discreteness is not overly restrictive. Indeed, any continuous distribution can be approximated by a sequence of discrete distributions with finite supports. Therefore, one can obtain as good approximation of the continuous distribution of qualities by a discrete random variable as information in the data would allow if the support of the discrete random variable is not restricted. Note that modeling unobserved heterogeneity using a discrete distribution is common in empirical studies. For examples, see Heckman and Singer (1984), Keane and Wolpin (1997), and Crawford and Shum (2005) to name just a few.
transitory bidders identify the distribution of transitory bidders’ qualities conditional on their bids and the distribution of outside options conditional on buyer’s weight for quality.

The first two steps essentially generalize identification mechanism which is well-known in the demand estimation literature: the distribution of buyers’ tastes (random coefficients) is identified from the variation in one of the covariates (permanent sellers’ prices) and in choice sets. The last step, however, is unique to our environment. We briefly illustrate it below.

In this step we aim to recover the distribution of $U_0$ given $\alpha$ and that of $Q_k$ given $b_k$. It is convenient to illustrate our argument using a simple case where the set of participants consist of two permanent bidders $i, j$ and a single transitory bidder $k$. If $i$ and $j$ belong to different quality groups then $I_A = \{|A^p_{i(k)}| = 1, |A^p_{i(j)}| = 1, |A^l| = 1\}$. Under the maintained assumption of independence between $(U_0, \alpha, \epsilon)$ and $C_i$’s (and hence $B_i$’s), the probability that $i$ wins conditional on the realized vector of bids $(b_i, b_j, b_k)$ and $I_A$ is

$$
\Pr (\alpha \Delta q_{j,i} + \Delta \epsilon_{j,i} \leq B_i - B_j \text{ and } Y_{i,k} - \epsilon_i \leq -B_i \lvert B_i = b_i, B_j = b_j, B_k = b_k, I_A),
$$

where random variable $Y_{i,k}$ is defined as the maximum of $U_0 - \alpha q_i$ and $\alpha(Q_k - q_i) - b_k + \epsilon_k$,

$$
\Delta q_{j,i} = q_j - q_i, \quad \Delta \epsilon_{j,i} = \epsilon_j - \epsilon_i.
$$

Sufficient independent variation in $b_i$ and $b_j$ ensures that the joint distribution of $\alpha \Delta q_{j,i} + \Delta \epsilon_{j,i}$ and $Y_{i,k} - \epsilon_i$ can be identified from the data. Once the distributions of $\epsilon_i$ and $\alpha$ as well as $q^k$ are recovered as in steps summarized above the marginal distribution of $Y_{i,k} + \alpha q_i$ could be obtained through a change of variables and by integrating one of the marginals out.

The cumulative distribution function of $Y_{i,k} + \alpha q_i$ can be further represented as follows:

$$
\psi_0(y, \alpha, b_k, I_A) \equiv \Pr (Y_{i,k} + \alpha q_i \leq y \lvert b_k, \alpha, I_A)
$$

$$
= \Pr (U_0 \leq y \lvert \alpha) \sum_m \tilde{\pi}_m(b_k, I_A)F_{\epsilon_k}(b_k + y - \alpha q^m),
$$

where $\tilde{\pi}_m(b_k, I_A) \equiv \Pr (Q_k = q^m | B_k = b_k, I_A)$. Note that the equality holds due to independence in (A2) and due to the Law of Total Probability.

We can construct a similar equation for the same $\alpha$, $y$ but $b'_k \neq b_k$. Taking the ratio of these two equations associated with $b_k$ and $b'_k$, and re-arranging terms, we get

$$
\psi(y, \alpha, b'_k, I_A) \sum_m \tilde{\pi}_m(b_k, I_A)F_{\epsilon_k}(b_k + y - \alpha q^m) = \psi(y, \alpha, b_k, I_A) \sum_m \tilde{\pi}_m(b'_k, I_A)F_{\epsilon_k}(b'_k + y - \alpha q^m),
$$

which is a system of linear equations in $2K$ unknown weights $\{\tilde{\pi}_m(b_k, I_A)\}_{m=1,...,K}$ and $\{\tilde{\pi}_m(b'_k, I_A)\}_{m=1,...,K}$ since $\psi(y, \alpha, b'_k, I_A)$, $\psi(y, \alpha, b_k, I_A)$ as well as $F_{\epsilon_k}(b_k + y - \alpha q^m)$ and $F_{\epsilon_k}(b'_k + y - \alpha q^m)$ are identified as explained above.

Evaluating (6) at the same pairs of $(b_k, b'_k)$ but different pairs of $(\alpha, y)$ gives us a linear system of equations in the $2M$ unknown weights. The $2M$ weights are then identified, provided the matrix of coefficients in the linear system formed as above has full rank at $2M$. The conditional distribution $F_{U_0|\alpha}$ is then identified from (4).

Thus, the changes in the relation between the winning probability and the price of permanent winner that are driven by the variation in prices of transitory sellers identify the conditional distribution of transitory seller’s quality conditional on his bid and the distribution of outside options conditional on buyer’s weight for quality.

---

\[23\] Also included into the linear system are two natural constraints on the vectors of weights: $\sum_m \tilde{\pi}_m(b_k, I_A) = 1$ and $\sum_m \tilde{\pi}_m(b'_k, I_A) = 1$. 
option conditional on buyer’s weight for quality.

As for the distribution of bidders’ private costs, it is identified via arguments similar to those in Guerre, Perrigne, and Vuong (2000) once the quality levels and the distribution of buyers’ weights are identified as above.

4.3 Generalized Method of Moments Estimation

The second step of our estimation procedure recovers the remaining structural primitives using Generalized Method of Moments (GMM) and utilizing insights from the previous section. In this step we rely on the group structure recovered in the first step.\textsuperscript{24} The GMM estimation is based on the moment conditions that stem from permanent seller’s winning probability conditional on the seller’s quality group and on some features of the set of active competitors.\textsuperscript{25}

To see our approach more clearly, let us recall that the buyers in our model make choices with full information on both permanent and transitory sellers’ qualities. The probability of buyer choosing $i$ (i.e., the probability of seller $i$ winning the auction) on the basis of his information can be written as

$$
\Pr(i \text{ wins} | i \in S^{p,k}, Q_{A'} = \bar{q}, B, I_A),
$$

where $B$ denotes the vector of submitted bids and $I_A$ represents information on the number of participating permanent sellers by quality group and the number of participating transitory sellers. Recall that a researcher does not observe transitory sellers’ qualities (i.e., $Q_{A'}$). Thus, in order to form the moment conditions corresponding to information available to the researcher, we integrate out transitory sellers’ qualities as follows\textsuperscript{26}

$$
\Pr(i \text{ wins} | i \in S^{p,k}, B, I_A) = \sum_{\bar{q}} \Pr(i \text{ wins} | i \in S^{p,k}, Q_{A'} = \bar{q}, B, I_A) \Pr(Q_{A'} = \bar{q} | B, I_A). \tag{6}
$$

The winning probability on the left hand side is estimable from data. The conditional winning probability $\Pr(i \text{ wins} | i \in S^{p,k}, Q_{A'} = \bar{q}, B, I_A)$ on the right hand side reflects buyers’ decision, and is determined by the distribution of buyers’ weights and outside option. To compute this probability from the model we utilize parametric assumptions for the distributions of $\epsilon_{i,l}$ and $(\alpha, U_0)$.\textsuperscript{27}

The remaining components, the conditional probabilities for the realizations of transitory sellers’ quality vector, $\Pr(Q_{A'} = \bar{q} | B, I_A)$, are equilibrium objects which reflect transitory sellers’ bidding and participation behavior. Proposition 2 obtains a convenient representation of these probabilities.

\textsuperscript{24}The estimation error due to using the estimated quality groups does not affect the asymptotic variance matrix of the GMM estimator because it has a convergence rate that is arbitrarily fast as the number of the auctions increases to infinity due to the finite number of quality groups.

\textsuperscript{25}In the actual empirical work we also condition on the structure of the set of potential bidders which we allow to vary across auctions. This is because our estimation is implemented for a more general version of our model which allows for strategic participation. The details of this approach are explained in Section 5.1 and in the Appendix.

\textsuperscript{26}Notice that $\Pr(Q_{A'} = \bar{q} | i \in S^{p,k}, B, I_A) = \Pr(Q_{A'} = \bar{q} | B, I_A)$ since $I_A$ summarizes all the necessary information about the set of competitors.

\textsuperscript{27}Notice that $\Pr(Q_{A'} = \bar{q} | B, I_A) = \Pr(Q_{A'} = \bar{q} | B', I_A)$. The equality holds because permanent sellers do not observe qualities of transitory sellers – we expand more on this further in the section. Also, from the previous section $\tilde{\pi}_\bar{q}(B', I_A) = \Pr(Q_{A'} = \bar{q} | B', I_A)$. 
Proposition 2 Under (A1′), (A2′), and (A3),

\[ \Pr(Q_{A^t} = \bar{q} \mid B, I_{A,N}) = \frac{g_{\bar{q}}(B^t, I_{A,N})}{\sum_q g_q(B^t, I_{A,N})}. \] (7)

where

\[ g_{\bar{q}}(B^t, I_{A,N}) \equiv \prod_{j \in A^t} \pi_{t,k(j)} \Pr(j \text{ is active} \mid Q_j^f = \bar{q}_j, I_N) f(B_j^t \mid Q_j^f = \bar{q}_j, I_N), \]

\[ B^t = (B_{1^t}, ..., B_{|A^t|}) \] and \( I_{A,N} \) summarizes information about the set of active and potential sellers for a given project.

Note that Proposition 2 above is stated for the general case with strategic (endogenous) participation. In this formulation it relies on assumptions (A1′) and (A2′) that generalize assumptions (A1) and (A2) to such a case and on assumption (A3). The details of the general model and the proof of Proposition 2 are presented in the Appendix. We heuristically explain the derivation of Proposition 2 at the end of this section.

The probability \( \Pr(Q_{A^t} = \bar{q} \mid B, I_{A,N}) \) is now expressed in terms of the distribution of the sellers’ equilibrium bidding and participation strategies as well as the primitive distribution of the transitory sellers’ qualities. It is feasible to compute this object from the model under our methodology which defines sellers’ asymmetry at the level of the group rather than seller identity. However, it is still fraught with complications since solving an asymmetric auction with multiple sellers’ groups numerically is a challenging task and solution is quite often fragile and sensitive to the problem’s parameters. Our identification mechanism suggests an alternative approach.

As we have argued in the previous section the distribution of transitory seller’s quality given his bid is non-parametrically identified from the data jointly with the distributions of weights and the support of quality distribution. We, thus, choose to estimate these objects simultaneously while using the former to integrate transitory sellers’ qualities in estimation.

One approach consistent with this strategy would be to use sieve estimation that would allow recovering the distribution of qualities conditional on bid non-parametrically while restricting the distribution of buyers’ weights to belong to one of the parametric families. On-line appendix to this paper explains how such a procedure could be implemented. However, for the sake of tractability and computational time this paper pursues a fully parametric approach. In estimation, we parameterize the sellers’ bidding and participation strategies and then combine them as in Proposition 2 to obtain \( \Pr(Q_{A^t} = \bar{q} \mid B, I_A) \). This provides us with better intuition for parametrization as well as allows imposing additional restrictions in estimation which strengthen the power of our identification mechanism.

Our GMM estimation is based on two sets of moment conditions. The moments in the first set are based on the expressions from (6) and Proposition 2 and are related to the probability that a permanent seller from the quality group \( S^{p,k} \) wins the auction given certain properties of the set of active bidders such as, for example, that this set include permanent sellers from the same group or permanent sellers from a specific pair of groups. This set of moments additionally includes the expected values of the price differences and squared price differences between the

\[ \text{This argument could easily be extended to parametric setting. More specifically, representation in (6) can be re-written as a system of equations that linearly include } g_{\bar{q}}(\cdot \mid I_{A,N}) \text{ for all possible } \bar{q} \text{ consistent with } A^t. \] The number of equations is virtually unlimited since a separate equation can be written for each possible configuration of the set of permanent sellers as well as for each possible vector of permanent sellers’ prices. This system of equations, thus can be solved for a full set of \( g_{\bar{q}}(\cdot \mid I_{A,N}) \).
winner and some other permanent bidder, the product of such price difference and the price of an active transitory bidder.

In the second set of moments, we additionally impose the rationality of entry decisions in combination with the requirement that the conditional bid distributions and conditional participation probabilities aggregated to the level observed in the data correspond to their empirical counterparts. These restrictions are needed to separate \( \Pr(j \text{ is active} | Q_j^t = q_j, I_N) \) from \( f(B_j^t | Q_j^t = q_j, I_N) \) and \( \pi_{t,k(j)} \). A detailed account of moment conditions used for estimation is found in the Appendix.

Let us conclude this section with heuristics behind Proposition 2. By the Bayes rule and the Law of Total Probability, we first write

\[
\Pr(Q_{A^t} = q | B, I_A) = \frac{f(B | Q_{A^t} = q, I_A) \Pr(Q_{A^t} = q | I_A)}{f(B | I_A)} = \frac{\Pr(Q_{A^t} = q | I_A) \prod_{j \in A^t} f(B_j^t | Q_{A^t}, j = q_j)}{\sum_q \Pr(Q_{A^t} = q | I_A) \prod_{j \in A^t} f(B_j^t | Q_{A^t}, j = q_j)}
\]

where \( f(B | Q^t = q, I_A) \) and \( f(B | I_A) \) are joint conditional density functions of \( B \) given \( Q_{A^t} = q, I_A \) and given \( I_A; B_j^t \) and \( Q_{A^t}, j \) are the bid and the quality of transitory bidder \( j \). The last equality uses independence of bidders’ strategies conditional on \( I_A \).\(^{29}\) The event \( \{Q_{A^t} = q\} \) in the above arises when the bidders in \( A^t \) are active and \( Q_{A^t} \) is realized to be \( q \). For expositional clarity, let us consider a simple case when the set of active participants contains no other permanent active bidders except the winner, i.e., \( I_A \) is given by the number of active transitory bidders such that \( I_A = |A^t| \), and continue to assume non-strategic entry. Then for each integer \( m \),

\[
\Pr(Q_{A^t} = q | |A^t| = m) = \sum_{a : |a| = m} \Pr(Q_{a,j} = q_j \text{ and } D_j = 1, \forall j \in a | |A^t| = m) = |\{a : |a| = m\}| \prod_{j \in a} \Pr(D_j = 1 | Q_{a,j} = q_j) \Pr(Q_{a,j}),
\]

where \( q_j \) denotes the \( j \)-th component of \( q \) and \( D_j \in \{0, 1\} \) takes 1 if and only if player \( j \) is active in the auction. The second equality uses the fact that \( D_j \)'s are exogenous and independent across the players.\(^{30}\)

\(^{29}\)Notice that the distribution of bids of permanent bidders cancels out from the numerator and denominator because this distribution does not depend on the particular realization of the vector qualities of transitory bidders since these qualities remain unknown to permanent bidders. In addition, the conditional distribution of bids for transitory bidder \( j \) depends only on his own quality since he does not observe the quality of his transitory competitors.

\(^{30}\)More specifically,

\[
\sum_{a : |a| = m} \Pr(Q_{a,j} = q_j \text{ and } D_j = 1, \forall j \in a | |A^t| = m) = \sum_{a : |a| = m} \prod_{j \in a} \Pr(D_j = 1 | Q_{a,j} = q_j) \Pr(Q_{a,j})
\]

\[
= |\{a : |a| = m\}| \prod_{j \in a} \Pr(D_j = 1 | Q_{a,j} = q_j) \Pr(Q_{a,j}).
\]

The last equality holds because the terms \( \Pr(D_j = 1 | Q_{a,j} = q_j) \Pr(Q_{a,j}) \) under the product and summation do not depend on a specific realization of \( A^t \) and \( a_0 \) is one of the possible such realizations.
As a final step, recall that $Pr(Q_{a,j} = \bar{q}_j)$ represents the proportion of transitory bidders with quality level $\bar{q}_j$. This proportion is denoted earlier by $\pi_{t,k(j)}$. Furthermore, under non-strategic entry, we can write $Pr(D_j = 1|Q_{A^t,j} = \bar{q}_j) = \lambda^t_{k(j)}$. Thus,

$$Pr(Q_{A^t} = \bar{q}|B, I_A) = \frac{\prod_{j \in A^t} f(B^t_j|Q_{A^t,j} = \bar{q}_j)\lambda^t_{k(j)}\pi_{t,k(j)}}{\sum_q \prod_{j \in A^t} f(B^t_j|Q_{A^t,j} = q_j)\lambda^t_{k(j)}\pi_{t,k(j)}},$$

which is the expression in Proposition 2 in the case of non-strategic entry. In the case of strategic entry, more work is required to associate $Pr(Q_{A^t} = \bar{q}|B, I_A)$ with the bidders’ participation strategies. Its full derivation is provided in the Appendix.

4.4 Extensions

In this section we explain how our simple model can be enriched to allow for strategic auction participation and observable project and seller heterogeneity. All the identification and estimation results can be easily extended to such more general setting.

4.4.1 Endogenous Entry

In this section we extend our model to allow for strategic (endogenous) entry decisions by sellers. It is important to account for strategic participation in our setting due to the endogeneity associated with transitory sellers’ qualities. More specifically, under strategic participation the distribution of transitory sellers’ qualities in the auction differs from the distribution of qualities in the population or within the set of potential bidders if sellers become potential bidders due to the circumstances unrelated to their quality. Thus, the distribution of qualities used to integrate out bid-quality relationship in estimation has to account for participation decision.

We model the entry decisions in the following way. Let $N_l$ denote the set of potential bidders for a given auction $l$. As in previous subsections, we abstract away from auction- and seller-level heterogeneity observed in the data. We explain how such heterogeneity can be introduced into the model and our methodology in the next section.

A set of potential bidders is partitioned into a set of potential permanent bidders, $N^p_l$, and potential transitory bidders, $N^t_l$. Recall, that the qualities of permanent sellers are known to all market participants and considered unknown parameters from a researcher’s point of view. The quality of a transitory seller is only known to himself and to the buyer (if this seller decides to enter the auction by submitting a bid). For the researcher and all other market participants, the qualities of transitory potential bidders are summarized by a random vector $Q_{N^t_l} = \{Q_j : j \in N^t\}$ such that $Pr(Q_j = q_j^t) = \pi_{t,k}$ for $j \in N^t$.

During an auction for project $l$ each potential bidder $i \in N^p_l \cup N^t_l$ observes some private signal, or entry costs, $E_{i,l}$, drawn from distribution $F_E$ and is aware of $N_l$. More specifically, seller $i$’s information set consists of $E_{i,l}$ and $I_{N,l}$, where the later contains information on the numbers of potential permanent bidders by quality group, and the total number of potential transitory bidders. Given this information set, potential bidder $i$ decides whether to participate in the auction or not. His entry strategy $\sigma^E_i$ is a mapping from the supports of $E_{i,l}$ and $I_{N,l}$ into $\{0, 1\}$. We denote the entry decision (outcome) by $D_{i,l}$ ($D_{i,l} = 1$ if enters and $D_{i,l} = 0$ otherwise).

Upon entry, an active bidder observes a private cost $C_{i,l}$ for completing the project. As in the basic model, each active bidder $i$ does not observe participation decisions of other potential
bidders, and is thus unaware of the composition of the set of active bidders. He then submit a price \( B_{i,l} \) based on his information set.

The potential bidders' strategies and psBNE in this environment can be defined in a usual way. We focus on type-symmetric equilibria in which any pair of participants \( i, j \) who are \textit{ex ante} identical (i.e. either \( "i, j \in N^p_l" \) and \( q_i = q_j \) or \( "i, j \in N^t_l" \)) adopt the same strategies. Supplemental on-line appendix provides further details and the proof of the equilibrium existence. It also argues that identification strategies described above remain applicable. One modification concerns the first step of the estimation procedure: pairwise comparison of sellers \( i \) and \( j \) requires that they should be compared on the basis of auctions where they both belong to the set of potential bidders.

### 4.4.2 Project and Seller Heterogeneity

We now discuss how to extend the main methodology in Sections 3 and 4 to accommodate observable project and seller heterogeneity. In our application, projects differ in several observable dimensions such as the date of posting, the nature of the work, and other specification details. Buyers weights may potentially vary with project characteristics. Thus, we perform the analysis conditional on observable project characteristics.

The sellers differ by their country of origin as well as their recorded performance measures such as reputation scores, delays or instances of conflict. These performance measures may reflect market’s information about seller’s quality or may be indicative of other service dimensions that are valued by buyers. In any case, all observable seller characteristics (including country) may plausibly be correlated with seller’s quality. Therefore, in contrast to a standard differentiated product environment the characteristics of competing sellers cannot be used as instruments in our setting. Our approach is to recover permanent sellers grouping conditional on observable characteristics in the first step of the analysis. Once the grouping is recovered the group-specific dummies control for the endogeneity of permanent sellers’ observable characteristics and bids directly. The grouping also allows us to recover the conditional distribution of \( Q^t_i \) conditional on \( B^t_i \) as explained in identification section above.

Formally, each seller \( i \) is characterized by a vector of observable characteristics \( X_i \), and a scalar measure of quality \( q_i \) which is not observed in the data. The support of the distribution of qualities among sellers with \( X_i = x \) is given by \( \{q^k(x) : 1 \leq k \leq K_x\} \) where \( K_x \) is the cardinality of the support given \( x \). The proportion of various quality levels among permanent and transitory sellers with \( X_i = x \) is \( \{\pi_{r,k}(x) : r \in \{p, t\}, 1 \leq k \leq K_x\} \). Finally, we assume that buyers’ value for \((x, q)\)-seller in an auction indexed by \( l \) is:

\[
\Delta_{i,l} = \alpha_l \tilde{q}^{k(i)}(x_i) + x_i \beta_l + \epsilon_{i,l},
\]

where \( \tilde{q} \) refers to the residual quality after the mean quality associated with observed characteristics, \( x_i \beta_l \), is netted out; \( \beta_l \) reflects buyer’s weights for observable seller characteristics. Note, that in the case when \( \beta_l \) are constant across buyers the expression for \( \Delta_{i,l} \) can be re-written as

\[
\Delta_{i,l} = \alpha_l q^{k(i)}(x_i) + \epsilon_{i,l}.
\]

Previous arguments hold once conditioning on the vector of non-quality characteristics of potential bidders, provided the required assumptions are satisfied conditional on this vector. Thus, our classification algorithm is implemented within the subpopulation of sellers characterized by \( X_i = x \).
The argument for identification of the distribution of $\beta$ is quite standard and is presented in the supplemental on-line appendix.

5 Empirical Results

5.1 Data Description

We have access to the data from the starting date of our on-line programming market and for the subsequent 6 years of this company’s operation. The data include information on close to 600,000 projects that involve participation from around 50,000 different sellers. For every project, we observe the description of work required, the approximate size of the project, the time requirements, and the location of the buyer. We also observe all bids submitted, the identity of the winner, and measures of the winner’s subsequent performance.

The projects fall into several broad classes, such as platform 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.

Table 1 provides some descriptive statistics for projects in our data set. Each row of the table summarizes a marginal distribution of the correspondent variable. The table shows that a sizable number of the projects are very small (below $100). On the other hand, some of the projects are quite big (above $1000). The projects are fairly short: the deadline for the majority of the projects is between one to three weeks. The median number of sellers submitting bids for a project is six while the median number of permanent bidders is three. However, about 10% of projects receive more than 18 bids (5 from permanent bidders). The projects with a large number of bids tend to be small.

The table also summarizes the characteristics of permanent sellers. It shows that a median permanent seller has completed 100 projects, while 10% of sellers completed 250 or more projects. The distribution of the average reputation scores appears to be quite tight. A median permanent seller has an average score of 9.87, while less than 25% have an average score below 9.7 or above 9.95. Similarly, a median permanent seller was never involved in an arbitration or had a delay. However, less than 10% of permanent sellers were involved in at least one arbitration or had at least one delay.

5.2 Some Data Regularities

The majority of buyers in our data are one-time participants. Less than 2% of buyers return with multiple projects. In addition, repeat buyers do not return with the same type of project. As a result, they very rarely work with the same seller repeatedly.

The multi-attribute feature of the auction is strongly supported in the data. Indeed, in our sample, 58% of the projects are awarded to a seller who quotes a price above the lowest price submitted in the auction. Table 2 documents the share of such projects as well as an average mark-up over the smallest bid for some project types.

\[ \text{\footnotesize The following anecdotal insight may help to put the size into the right perspective. One of the authors used this market to procure programming services: the project that costs $200 in this on-line market was quoted at $800 in the off-line programming market in Philadelphia.} \]
Table 1: Data Summary Statistics

<table>
<thead>
<tr>
<th></th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>90%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Project Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>$150</td>
<td>$250</td>
<td>$500</td>
<td>$1000</td>
</tr>
<tr>
<td>Duration</td>
<td>5</td>
<td>10</td>
<td>14</td>
<td>21</td>
</tr>
<tr>
<td>Number of Bidders</td>
<td>4</td>
<td>6</td>
<td>11</td>
<td>18</td>
</tr>
<tr>
<td>Number of Permanent Bidders</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td><strong>Permanent Sellers’ Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experience</td>
<td>75</td>
<td>100</td>
<td>150</td>
<td>250</td>
</tr>
<tr>
<td>Average Score</td>
<td>9.7</td>
<td>9.87</td>
<td>9.95</td>
<td>10</td>
</tr>
<tr>
<td>Arbitrations</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Delays</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>Number of Projects</strong></td>
<td>32,679</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The results in this table are based on a sample of projects with graphics-related programming. Duration of project is measured in days. Each row summarizes the inverse cumulative function of the corresponding variable. Experience is defined as the number of completed projects.

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.

We explore the buyers’ choices using a logit model with random coefficients (without alternative-specific fixed effects). In this analysis, we set the dependent variable, \( Y_{li} \), to be a project award dummy that is equal to one if the seller \( i \) won the project \( l \) and zero otherwise. The award depends on the buyer’s net value from a specific alternative (seller), which is modeled as a linear function of seller characteristics, \( X_{ki} \) (the number of ratings (experience), delays, arbitrations and the average reputation score), seller location dummies, \( \mu_c(i) \), and a seller’s bid, \( B_{li} \):

\[
Y_{li} = X_i \alpha_l + \gamma_l B_{li} + \mu_{c(i)} + \epsilon_{li} \tag{11}
\]

We estimate the mean of the price coefficient to be positive and statistically significant.\(^{32}\) This result suggests an omitted variable bias since, in most markets, buyers dislike paying higher prices, 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 describes this setting should allow for an unobserved quality-like sellers’ attributes.

As we emphasized in the introduction our market attracts a large number of short-lived sellers. This feature of the data is summarized in Table 3. Let us define a seller’s tenure as the

\(^{32}\)We estimate the mean of the coefficient on price to be equal to 2.82 and the standard deviation to be equal to 1.76.
Table 2: Projects Awarded at a Price That Exceeds Lowest Price in Auction

<table>
<thead>
<tr>
<th>Type of Work</th>
<th>Project’s Share</th>
<th>Price Mark-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database</td>
<td>64%</td>
<td>41.2%</td>
</tr>
<tr>
<td>Platforms</td>
<td>52%</td>
<td>37.9%</td>
</tr>
<tr>
<td>Graphics</td>
<td>71%</td>
<td>38.4%</td>
</tr>
<tr>
<td>Web-related</td>
<td>52%</td>
<td>41.2%</td>
</tr>
</tbody>
</table>

Note: The results in this table are based on a full sample that includes 600,000 projects. The “Project’s Share” column reports the fraction of projects that have been awarded to bidders with price quotes that exceed the lowest price quote for the respective project. “Price Mark-up” summarizes the average normalized difference between the winning price and the lowest price quote across projects that are awarded at a price that exceeds the lowest price quote. The differences in prices are normalized by the lowest price quote.

length of time that elapses between the date when he submits his last bid and the date when he submits his first bid. 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 these years, 30% of the sellers stayed in the market for more than a year, whereas 65% of the sellers left the market in less than three months. Substantial seller turnover is an important feature of our market as well as many other markets for services.

In contrast to other on-line markets, the sellers’ performance does not appear to be related to their propensity to stay in the market. To see this, we define permanent sellers as the sellers with a tenure longer than one year and transitory sellers as the sellers who left after less than one year. Table 3 documents no significant differences between permanent and transitory sellers in the number of bids submitted before the first success (conditional on achieving at least one success), as well as in the distribution of reputation scores received by these bidders for their first or last projects. We obtain similar results when transitory sellers are redefined to be those who left after six or three months.

An interesting regularity emerges concerning the number of bids before the first success. When we compute the distribution of this variable for all transitory bidders (including those who did not win any projects), the time to the first success for transitory sellers appears to be substantially shorter than that for permanent sellers. This suggests that many transitory bidders do not wait for success and that sorting into permanent or transitory groups is likely driven by sellers’ outside opportunities rather than performance differences among the sellers. Therefore, the quality distributions of permanent and transitory sellers are likely to be quite similar. Thus, the assumption that the supports of quality distributions are the same in permanent and transitory seller populations appears reasonable in this market.

In general, transitory sellers appear to be quite successful: their rate of winning is comparable to that of permanent sellers, and they often beat permanent sellers at comparable prices. Given that extensive communication between buyers and sellers is present, this suggests that buyers may be able to assess the quality of transitory sellers as accurately as they assess the quality of permanent sellers.

On the other hand, 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
Table 3: Analysis of Permanent vs. Transitory Sellers

<table>
<thead>
<tr>
<th>Tenure Distribution</th>
<th>≤ 1m</th>
<th>≤ 3m</th>
<th>≤ 12m</th>
<th>≤ 24m</th>
</tr>
</thead>
<tbody>
<tr>
<td>overall</td>
<td>65%</td>
<td>75%</td>
<td>80%</td>
<td>90%</td>
</tr>
<tr>
<td>annual (last 3 years)</td>
<td>45%</td>
<td>65%</td>
<td>70%</td>
<td>75%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of Bids Before First Success</th>
<th>≤ 10%</th>
<th>≤ 25%</th>
<th>≤ 50%</th>
<th>≤ 75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>tenure ≥ 12m</td>
<td>5</td>
<td>9</td>
<td>17</td>
<td>42</td>
</tr>
<tr>
<td>tenure ≤ 12m (success ≥ 1)</td>
<td>3</td>
<td>7</td>
<td>15</td>
<td>36</td>
</tr>
<tr>
<td>tenure ≤ 12m (all)</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>First Reputation Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>tenure ≥ 12m</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>tenure ≤ 12m</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Last Reputation Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>tenure ≥ 12m</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>tenure ≤ 12m</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Note: The results in this table are based on a full sample that includes 600,000 projects. In these calculations the “Tenure” variable reflects the total length of time a seller is observed to be active in the market, i.e., “Tenure” equals the length of time between the date of the last bid recorded in the data and the date of the first bid. Panel 1 records the cumulative distribution function of the “Tenure” variable, panel 2 records the inverse of the cumulative distribution function of the variable “Number of Bids Before First Success,” and panels 3 and 4 record the probability distributions of the variables “First Reputation Score” and “Last Reputation Score.” This table indicates that permanent and transitory sellers appear to be very similar in their performance.

permanent sellers. The market may infer their quality from the long-run rate of their successes through reasoning similar to that we use in this paper.

To summarize, the preliminary analysis of our data indicates that (a) the buyers’ score should non-trivially depend on sellers’ attributes; (b) the model should allow for the presence of an unobserved quality-like sellers’ attributes; (c) it is important to account for the presence of a large number of transitory sellers; (d) buyers most likely observe the qualities of participating transitory sellers; and (e) the distribution of the seller’s quality does not depend on the seller’s tenure.

We now turn to the discussion of the estimation results. We first summarize estimates from our classification procedure, then we discuss the parametric estimates of the buyers’ weights and the support of sellers’ quality distribution as well as the bidding and participation strategies of transitory sellers.

5.3 Empirical Results: Classification

In this section we summarize the results of the group structure estimation. The classification algorithm is applied to the set of permanent participants specializing in graphics-related pro-
gramming. Projects of this type involve programming computer games, computer-generated animation, and media-related programming. Our choice of project type was mostly motivated by sample size considerations. However, this is also a highly specialized segment of the market. The related work is very sophisticated and is done exclusively by hard-core professionals. This, therefore, is an environment where the seller’s quality is likely to matter. On the other hand, this environment perhaps would be characterized by lower variation in provider qualities as opposed to the less skill-intensive types of projects.

For the sake of tractability we present the results for the projects owned by US buyers that are of medium to medium-large size (between $400 to $700) and have the specified duration of two to three weeks. We have estimated the model for several cells of projects defined in terms of size and duration. The differences in the results of estimation across projects cells are not sufficiently large as to warrant a separate discussion in the paper. However, we use all results in our counterfactual analysis.

We assume that the buyer’s net value depends on the seller’s price, as well as on the seller’s attributes such as quality, seller’s performance indicators, and country affiliation. We remain agnostic about the exact role of performance indicators. It may be that they summarize publicly observable information about the seller’s quality or they may play some other role, e.g., indicate the seller’s reliability. The seller’s country affiliation may also enter buyers’ values since it can proxy for things such as convenience of working with a given seller related to time difference, the likelihood of language proficiency, and work culture. The unobserved group structure, which we allow to be country- and performance-dependent, captures the seller’s residual quality. It is plausible that the distribution of residual quality may vary across countries. At the same time performance measures may be endogenous and thus correlated with the residual quality.

In estimation, we discard the first year of the seller’s tenure and use only observations that correspond to the later years of his career with an on-line market. We divide all the sellers into three cells according to the average reputation score: (cell 1) average reputation score less than 9.7, (cell 2) average reputation score above 9.7 and below 9.9, (cell 3) average reputation score above 9.9. This results approximately in an allocation of 30%, 30%, and 40% across cells. All permanent sellers have a high number of ratings; therefore, we assume that the exact number of ratings is not important.33 We also 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.

The classification index is constructed for a pair of sellers on the basis of projects where they both belong to the set of potential bidders. In this analysis we assume that the set of potential sellers for project \( l \) consists of all sellers who are active in the market (i.e., submitted bids or sent messages to buyers) during the week when project \( l \) is posted and who are qualified for the type of work indicated for project \( l \) (i.e., they bid for similar projects in the past). We define the set of potential bidders in this way because such a definition fulfills two requirements: it includes sellers who (a) might reasonably be expected to compete with each other in a given auction; (b) these sellers are aware of the presence of other sellers from this set in the market.

---

33We have also verified the robustness of our results by repeating the analysis while including the number of arbitrations and delays as additional measures of reliability. 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.
during the auction.

We follow the steps described in Section 4.1. That is, we start by estimating a group structure for a range of the number 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 steps 1 and 2 for the group of Eastern European sellers with a medium level of average reputation score in a table included in the supplemental on-line appendix. The pair-wise nature of our index does not have strong implications for our sample. We are able to compute an index for each pair of sellers within each of our cells.

To verify the robustness of our results we have additionally estimated a quality group structure under the assumptions that the set of potential bidders consists of (a) sellers who contacted the buyer but did not necessarily submit a bid for project $l$ – the most stringent definition; or (b) sellers who were qualified for project $l$ and were active in the market during the two- or three-week window (the week of the auction and a week or two before the auction respectively). The results we report in the paper are very similar to those we find under the definitions (a) or (b). Specifically, the difference between quality group structures estimated under different specifications is under 3% of the total number of permanent sellers. The confidence intervals are somewhat larger under the definition in (a). This is not surprising since the classification procedure is based on a much lower number of auctions in this case.

Recall that we assume that sellers do not observe buyers’ weights in a given auction and that the distribution of sellers’ costs is the same for all projects that share the same observable characteristics. The results summarized in the previous paragraph indicate that unobserved project-level heterogeneity, even if present, is unlikely to be large. Indeed, note that the definitions we consider impose a varying degree of restrictions on the sets of projects that are used to compare a pair of sellers. These sets of projects are likely to differ in the mix of the realized values of unobserved project heterogeneity if it is present. If unobserved project heterogeneity affects this environment to an important degree, we would expect that recovered group structures would depend on the definition of the set of potential bidders. The fact that they do not is perhaps not surprising given that we observe many important parameters of the project and we have access to the third-party assessment of the project size.

Next, we explore the sensitivity of the estimation procedure to the bid levels used for classification. Specifically, we split the interval of bids we used in the original classification procedure into two sub-intervals and then re-do the classification for each sub-interval. We find that the estimated group structures are basically the same, with the difference between the three estimates being under 5% of the number of permanent sellers. Recall that one of the simplifying assumptions we make is an assumptions that seller’s quality is constant for all projects that share the same observable characteristics. These findings alleviate our concerns about the potential variation in sellers’ unobserved quality across projects. Indeed, if such project-specific variation in seller’s quality were important, we would expect that the seller would be more likely to be classified as high quality in the auctions where he submits high bids. However, we do not find any substantial evidence of such regularity.

Table 4 reports the estimated group structures with corresponding confidence sets for cells of North American, Eastern European and East Asian sellers. 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 ordinary 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) allow for
a non-trivial number of quality groups. Also, allocation of mass between quality groups differs across cells. We defer the more interesting substantive inference to the section on the results of the parametric estimation.

<table>
<thead>
<tr>
<th>Country Group</th>
<th>Average Score</th>
<th>Total Number of Suppliers</th>
<th>$Q = L$</th>
<th>$Q = M$</th>
<th>$Q = H$</th>
</tr>
</thead>
<tbody>
<tr>
<td>North America low</td>
<td>12</td>
<td>4</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>North America medium</td>
<td>13</td>
<td>4</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>North America high</td>
<td>17</td>
<td>12</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eastern Europe low</td>
<td>18</td>
<td>6</td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eastern Europe medium</td>
<td>52</td>
<td>33</td>
<td>12</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Eastern Europe high</td>
<td>83</td>
<td>6</td>
<td>65</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>East Asia low</td>
<td>91</td>
<td>62</td>
<td>18</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>East Asia medium</td>
<td>66</td>
<td>6</td>
<td>53</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>East Asia high</td>
<td>58</td>
<td>50</td>
<td>8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 suppliers in the cell. Columns 4-6 report the size of the estimated quality group. The size of the corresponding confidence set with 90% coverage is reported in parenthesis. Note that the confidence set with the level $(1 - \alpha)$ 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 $(1-\alpha)$.

5.4 Empirical Results: Parametric Estimation

In this section we present the results of the parametric analysis. We begin by summarizing our specification and then discuss the estimates of the objects of interest: the parameters of buyers’ weights distribution, quality distributions for a range of covariate values, and sellers’ bidding strategies and recovered cost distributions.

5.4.1 Parametric Specifications

We modify the specification of the buyer’s net value function for the purpose of estimation. More specifically, we divide the expression for the net value function by the quality coefficient $\alpha$. This
obtains a specification which similar to the one often used in the estimation of differentiated product models: \[ \tilde{q}_i(x) + x_i\tilde{\beta}_l - \tilde{\alpha}_l b_{li} + \tilde{\epsilon}_{li}. \]

Recall that \( \tilde{q} \) refers to the residual quality after the mean quality associated with observed characteristics, \( x_i\beta_l \), is netted out.

We have estimated a number of specifications, and in each case, the standard deviations on \( \beta \) coefficients were not statistically significant. That is why the results we report in the paper are for the specification where \( \beta \) coefficients are constant across buyers. As we mentioned, in this case the score function is given by

\[ q_i(x) - \tilde{\alpha}_l b_{li} + \tilde{\epsilon}_{li}. \]

We assume that match components, \( \tilde{\epsilon} \), follow the Extreme Value Type I distribution with standard error \( \sigma_\epsilon \), while weight parameter \( \tilde{\alpha} \), and buyer’s outside option are assumed to be distributed according to the normal distribution \( N(\mu_{a,U_0}, \Sigma_{a,U_0}) \). We impose the normalization assumptions implied by our identification argument. That is, we normalize the expected value of \( \tilde{\epsilon} \) to be equal to zero, the expected value of \( \tilde{\alpha} \) 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. We, therefore, aim to estimate the vector of parameters \( \theta = \{\sigma_\epsilon, \sigma_\alpha, \{q^k(x)\}_{k=1,...,K_x}\}_x \) where \( \{q^k(x)\}_{k=1,...,K_x} \) is the support of quality distributions that correspond to the covariate values \( x \).

As we stated in the previous section, we assume that buyers’ value from a specific seller depends on the seller’s quality, price, country group affiliation and the long-run average reputation score. Since the majority of transitory sellers complete only one or two projects 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 with a given number and sum of scores belongs to a particular long-run average score group. We recover these beliefs non-parametrically using beginning of career and long-run data on permanent bidders.

We assume that transitory and permanent sellers’ bid distributions are well approximated by normal distributions \( N(\mu_{B^t}, \sigma_{B^t}^2) \) and \( N(\mu_{B^p}, \sigma_{B^p}^2) \), respectively. The means of the bid distribution depend on the seller’s quality, country, and average reputation score group, and 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. Similarly, we approximate permanent and transitory bidders’ respective probabilities of participation by normal distribution functions that depend on linear indices 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

\[ 34 We could be worried about such re-parameterization in the case when zero belongs to the support of \( \alpha \). However, this would only mean that infinity belongs to the supports of \( \tilde{\alpha} = \frac{1}{\alpha}, \tilde{\beta} = \frac{2}{\alpha}, \text{ and } \tilde{\epsilon} = \frac{\xi}{\alpha} \), the case that can be easily accommodated.

\[ 35 \text{Strictly speaking, the distribution of } \alpha \text{ should have been chosen to have a non-negative support. However, we estimate the standard error of this distribution to be quite small so that this assumption does not make any practical difference. The same comment applies to our assumption on the distribution of bids below.} \]

\[ 36 \text{See the comment for the distribution of } \alpha \text{ above.} \]

\[ 37 \text{This is because the long-run average score is not observed in the data for transitory sellers. Therefore, the buyer has to base his expectation of the long-run average reputation score on contemporaneously available measures when awarding the project. This, in turn, implies that transitory bidders would incorporate their current average scores into their bids.} \]
of reputation scores.

Transitory sellers are an important feature of our setting that potentially introduces methodological challenges. That is why we estimate several specifications that differ in their treatment of transitory sellers. We discuss the results and compare the fit of these specifications in the next section.

5.4.2 Performance of Estimation Procedure and Model Fit

In this section we present and contrast estimation results for three specifications that differ in their treatment of transitory sellers. Specifications one and two allow for buyers to be informed about the qualities of transitory sellers, whereas specification three assumes that buyers treat transitory sellers as homogeneous conditionally on observable characteristics. Specification one is the most general of the three, since it allows the distributions of 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. In contrast, specification two restricts the frequencies of quality groups in populations of permanent and transitory sellers to be the same.

We first take a look at specifications one and two. Tables 5 and 6 report the parameters estimated in the second step of our estimation procedure. Table 5 reports the frequencies for the population of transitory sellers estimated in the second step of the first specification and compares them to the frequency distribution of quality groups in the population of permanent sellers estimated in the first step. The results in table 5 suggest that the two frequency distributions are very similar, with the transitory sellers’ distribution allocating a slightly larger probability mass to the higher quality cells.

Table 6 shows the estimated parameters of the distribution of buyers’ weights. The results for specifications one and two are reported in columns one and two, respectively. Both specifications report the estimated coefficients that have the expected signs. The estimated variance of $\epsilon$ is quite small, which indicates that the seller’s attributes indeed rationalize buyers’ choice to important degree. We also calculate that both specifications correctly predict around 75 – 78% of buyers’ choices and, thus, significantly improve on the model without unobserved seller heterogeneity, which can only rationalize 25% of the data. Specification one tends to have somewhat larger standard errors in comparison to specification two so that some quality levels appear to be not statistically different from zero under specification one, while they are estimated to be statistically significantly different from zero under specification two. However, the results for these specifications are broadly consistent. The specification in column one is substantially more challenging to estimate within the context of our model. Its performance could possibly be strengthened if the model also described the mechanism by which a seller becomes permanent or transitory. We, however, leave investigation of this issue for a separate project. We focus on specification two from now on since our estimates show sufficient support for this specification.

We perform a further robustness check of our approach with specification three. This specification restricts transitory sellers to be homogeneous (from the buyers point of view) conditional on observable characteristics. The estimated coefficients for this specification are reported in column three of table 6. They differ from those in columns one and two in several important dimensions. First, the estimated variance of $\epsilon$ is much higher under this specification. In addition, the estimated quality levels are less dispersed, with high quality levels being substantially lower.

In some cases, we estimate quality levels that are not statistically distinct for different quality
Table 5: Estimated Quality Distributions of Transitory Sellers

<table>
<thead>
<tr>
<th>Country Group</th>
<th>Average Score</th>
<th>Permanent Sellers</th>
<th>Transitory Sellers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Q = L</td>
<td>Q = M</td>
<td>Q = H</td>
</tr>
<tr>
<td>North America low</td>
<td>0.33</td>
<td>0.67</td>
<td>0.37</td>
</tr>
<tr>
<td>North America medium</td>
<td>0.31</td>
<td>0.69</td>
<td>0.26</td>
</tr>
<tr>
<td>North America high</td>
<td>0.71</td>
<td>0.29</td>
<td>0.65***</td>
</tr>
<tr>
<td>Eastern Europe low</td>
<td>0.33</td>
<td>0.67</td>
<td>0.35***</td>
</tr>
<tr>
<td>Eastern Europe medium</td>
<td>0.63</td>
<td>0.23</td>
<td>0.13</td>
</tr>
<tr>
<td>Eastern Europe high</td>
<td>0.07</td>
<td>0.78</td>
<td>0.14</td>
</tr>
<tr>
<td>East Asia low</td>
<td>0.68</td>
<td>0.20</td>
<td>0.12</td>
</tr>
<tr>
<td>East Asia medium</td>
<td>0.09</td>
<td>0.80</td>
<td>0.11</td>
</tr>
<tr>
<td>East Asia high</td>
<td>0.86</td>
<td>0.14</td>
<td>0.78***</td>
</tr>
</tbody>
</table>

This table compares the estimated distribution of transitory sellers’ qualities (far right panel) to the distribution of permanent sellers’ qualities as implied by the group structure recovered through the classification procedure (see table 4). In this table \((***)\) indicates that the estimated parameter is statistically significant at the 95% significance level.

These differences reflect an attempt by specification three to rationalize buyers’ choices that allocate projects to transitory sellers when permanent sellers with comparable prices are available. Despite this, specification three lags behind specifications one and two in predicting the probability that a project will be allocated to a transitory seller: the predicted probability for specification three is 0.37, whereas specifications one and two get very

\[38\] This moment is not specifically targeted in estimation and thus could be interpreted as a measure of fit.
Table 6: Buyers’ Tastes and Quality levels

<table>
<thead>
<tr>
<th>Variable</th>
<th>(1a)</th>
<th>(1b)</th>
<th>(2a)</th>
<th>(2b)</th>
<th>(3a)</th>
<th>(3b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>log(σε)</td>
<td>-0.732**</td>
<td>(0.223)</td>
<td>-0.615**</td>
<td>(0.041)</td>
<td>-0.228*</td>
<td>(0.192)</td>
</tr>
<tr>
<td>log(σα)</td>
<td>-1.028**</td>
<td>(0.211)</td>
<td>-0.898**</td>
<td>(0.035)</td>
<td>-1.865**</td>
<td>(0.033)</td>
</tr>
<tr>
<td>µU0</td>
<td>-2.213**</td>
<td>(0.332)</td>
<td>-1.840**</td>
<td>(0.035)</td>
<td>-1.113**</td>
<td>(0.009)</td>
</tr>
<tr>
<td>log(σU0)</td>
<td>-0.246*</td>
<td>(0.131)</td>
<td>-0.329**</td>
<td>(0.046)</td>
<td>0.136*</td>
<td>(0.072)</td>
</tr>
<tr>
<td>σα,U0</td>
<td>0.149</td>
<td>(0.092)</td>
<td>0.242**</td>
<td>(0.063)</td>
<td>-0.052</td>
<td>(0.037)</td>
</tr>
<tr>
<td>North America, low score,</td>
<td>1</td>
<td>-0.062</td>
<td>(0.081)</td>
<td>-0.016**</td>
<td>(0.007)</td>
<td>-0.021**</td>
</tr>
<tr>
<td>North America, medium score,</td>
<td>2</td>
<td>0.399**</td>
<td>(0.065)</td>
<td>0.413**</td>
<td>(0.009)</td>
<td>0.298**</td>
</tr>
<tr>
<td>North America, high score,</td>
<td>1</td>
<td>0.001</td>
<td>(0.023)</td>
<td>-0.016**</td>
<td>(0.008)</td>
<td>-0.013</td>
</tr>
<tr>
<td>North America, medium score,</td>
<td>2</td>
<td>0.412**</td>
<td>(0.057)</td>
<td>0.433**</td>
<td>(0.008)</td>
<td>0.235**</td>
</tr>
<tr>
<td>North America, high score,</td>
<td>1</td>
<td>0.003</td>
<td>(0.032)</td>
<td>-0.016**</td>
<td>(0.003)</td>
<td>-0.019**</td>
</tr>
<tr>
<td>North America, medium score,</td>
<td>2</td>
<td>0.488**</td>
<td>(0.071)</td>
<td>0.507**</td>
<td>(0.004)</td>
<td>0.261**</td>
</tr>
<tr>
<td>Eastern Europe, low score,</td>
<td>1</td>
<td>0.112</td>
<td>(0.102)</td>
<td>0.263**</td>
<td>(0.003)</td>
<td>0.116**</td>
</tr>
<tr>
<td>Eastern Europe, medium score,</td>
<td>2</td>
<td>0.703**</td>
<td>(0.012)</td>
<td>0.625**</td>
<td>(0.005)</td>
<td>0.322**</td>
</tr>
<tr>
<td>Eastern Europe, medium score,</td>
<td>3</td>
<td>0.781**</td>
<td>(0.035)</td>
<td>0.672**</td>
<td>(0.009)</td>
<td>0.345**</td>
</tr>
<tr>
<td>Eastern Europe, medium score,</td>
<td>1</td>
<td>0.111</td>
<td>(0.104)</td>
<td>-0.103**</td>
<td>(0.005)</td>
<td>-0.031**</td>
</tr>
<tr>
<td>Eastern Europe, high score,</td>
<td>1</td>
<td>0.001</td>
<td>(0.031)</td>
<td>-0.107**</td>
<td>(0.006)</td>
<td>-0.032**</td>
</tr>
<tr>
<td>Eastern Europe, medium score,</td>
<td>2</td>
<td>0.289**</td>
<td>(0.047)</td>
<td>0.263**</td>
<td>(0.005)</td>
<td>0.129**</td>
</tr>
<tr>
<td>Eastern Europe, high score,</td>
<td>3</td>
<td>0.789**</td>
<td>(0.069)</td>
<td>0.668**</td>
<td>(0.004)</td>
<td>0.351**</td>
</tr>
<tr>
<td>South and East Asia, low score,</td>
<td>1</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>South and East Asia, low score,</td>
<td>2</td>
<td>0.108**</td>
<td>(0.043)</td>
<td>0.089**</td>
<td>(0.008)</td>
<td>0.154**</td>
</tr>
<tr>
<td>South and East Asia, low score,</td>
<td>3</td>
<td>0.512**</td>
<td>(0.034)</td>
<td>0.449**</td>
<td>(0.008)</td>
<td>0.178**</td>
</tr>
<tr>
<td>South and East Asia, medium score,</td>
<td>1</td>
<td>0.001</td>
<td>(0.101)</td>
<td>-0.019**</td>
<td>(0.003)</td>
<td>-0.001</td>
</tr>
<tr>
<td>South and East Asia, medium score,</td>
<td>2</td>
<td>0.201**</td>
<td>(0.059)</td>
<td>0.105**</td>
<td>(0.007)</td>
<td>0.064**</td>
</tr>
<tr>
<td>South and East Asia, medium score,</td>
<td>3</td>
<td>0.535**</td>
<td>(0.042)</td>
<td>0.544**</td>
<td>(0.006)</td>
<td>0.297**</td>
</tr>
<tr>
<td>South and East Asia, high score,</td>
<td>1</td>
<td>0.067</td>
<td>(0.041)</td>
<td>0.105**</td>
<td>(0.004)</td>
<td>0.068**</td>
</tr>
<tr>
<td>South and East Asia, high score,</td>
<td>2</td>
<td>0.586**</td>
<td>(0.012)</td>
<td>0.556**</td>
<td>(0.007)</td>
<td>0.301**</td>
</tr>
<tr>
<td>Pr(transitory seller wins)</td>
<td>0.68</td>
<td></td>
<td>0.64</td>
<td></td>
<td>0.37</td>
<td></td>
</tr>
</tbody>
</table>

The results are based on the data set consisting of 11,300 projects. The quality level for South and East Asia, low score, Q = 1, is normalized to be equal to zero. The columns in the table show the estimated coefficients and corresponding standard errors for several specifications: the numerical part of the column label indicates the specification, whereas the letter (a) denotes the column that contains the estimated coefficients and the letter (b) indicates the column with standard errors. Specification (1) corresponds to the case when the distribution of transitory sellers’ qualities is estimated, whereas specification (2) corresponds to the baseline case when the distributions of qualities for transitory and permanent sellers are restricted to be equal. Specification (3) corresponds to the robustness check where we assume that the buyer is not informed about transitory sellers’ qualities and thus treats them as homogeneous conditional on observable characteristics. The stars, **, indicate that a coefficient is significant at the 95% significance level.
close to the probability in the data (0.6) with predicted probabilities 0.68 and 0.64 respectively. On the basis of these results we conclude that the assumption of the buyer not being informed about the qualities of transitory sellers does not appear to be consistent with the data.

Finally, it is worthwhile noting that the estimated distribution of transitory sellers’ bids and their participation probabilities (estimated under specification two and reported in the supplemental on-line appendix) indicate a statistically significant dependence of these objects on transitory sellers’ quality levels. Recall that the transitory sellers’ bid distribution is estimated jointly with other parameters from the observed buyers’ choices via the set of moments that exploit the structure of our model and proposed identification strategy. In particular, there is no direct link in the data between the transitory sellers’ bids and their quality levels. Our estimates, therefore, support assumptions of our model as well as validate our identification strategy.

5.4.3 Quality and Other Attributes as Determinants of Buyer’s Choice

The last panel of table 6 reports the estimated quality levels across covariate cells. In the estimation the prices are normalized by project size; therefore, these estimates reflect the percentage of the project size that a buyer would be willing to pay for the corresponding quality level.

The estimated levels have the expected sign and are increasing according to group ranking. The differences across quality levels are substantial in magnitude. Since, in addition, the model with quality explains the data substantially better than the model without quality, we can conclude that quality plays an important role in our environment.

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 50% premium for a high-score, high-quality North American seller about 5% of buyers would pay more than an 5% premium and another 10% would pay less than a 15% premium.

Next, we observe that the quality levels are consistent across covariate cells. There appears 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, 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 Middle 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 middle 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. This finding underscores the importance of using our methodology, which allows for such dependence, as opposed to a pure mixture methodology that would have to impose the restriction that the distribution of unobserved heterogeneity is orthogonal to other variables that may enter buyers’ net value function.

Country and the long-run average reputation score appear to have independent effects on the buyer’s utility. These effects, however, are rather small relative to the differences in quality levels. For example, an average buyer 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, an average buyer 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.

Tables 2, and 3 in the supplemental on-line appendix report the estimated coefficients for bid distributions and participation probabilities. We estimate that the number of reputation scores and an average reputation score matter for transitory bidders in a statistically significant way. The results in the supplemental on-line appendix show how these variables impact transitory sellers’ prices (bids). For example, having no reputation scores bears a negative premium of close to 8% relative to the price charged by a seller with more than six scores. On the other hand, having a positive but small number of scores erodes this negative premium to 4% or 3%. The average reputation score does not appear to be important when the number of scores is really small. However, the difference between 9 points and 10 is rewarded with a 5% premium if the number of scores is moderate. This is comparable to the 7% premium documented above for the case of a long-run average reputation score that corresponds to the large number of scores.

5.4.4 Buyers’ Gains from Market Globalization

Our estimation procedure allows us to recover the joint distribution of buyers’ price sensitivity and outside option. The estimated mean of the distribution of outside option, \(\mu_{\text{0}}\), measured relative to the quality level of a South or East Asian, low-score, low-quality seller and shown in table 6 is somewhat lower than the average value from an inside option. The variance of the distribution of the outside option is larger than the variance of the stochastic match component \(\epsilon\). In our sample the outside option is positively correlated with price sensitivity, i.e., buyers with the high outside option also tend to be more price sensitive.

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:

\[
\frac{1}{L} \sum L \mathbb{E}_{\alpha,\epsilon,\mu_{0},Q_{t}|B_{t}}[\max_{\substack{i \in \mathcal{A}_{l} \cup \emptyset}} [U_{i,l} - U_{0,l}| i \text{ wins}]
\]

Notice that in this analysis the expectation is computed using conditional distributions of \(\alpha, \epsilon, \mu_{0}\) that are consistent with the buyer’s choice observed in the data (here \(i \in \mathcal{A}_{l} \cup \emptyset\) and \(\emptyset\) denotes choosing an outside option).\(^{39}\) 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. This is comparable to 50% the premium that buyers are willing to pay on average in order to procure high- rather than low-quality services.

We further document the welfare gain over the outside option for different levels of buyer price sensitivity. The results of this analysis are summarized in table 7. The results indicate that the gains are very similar across different levels of price sensitivity, while the buyers with low price sensitivity gain the most. This happens mostly because the buyers with low price sensitivity also tend to have lower values of the outside option, whereas the buyers with high price sensitivity tend to have a higher outside option. However, when compared in absolute terms the value to the price-sensitive buyer from the inside option is slightly higher (by 0.052 or

\(^{39}\)We compute this expectation using simulation methods.
5% of the project value) than the value to the buyer with low price sensitivity. Thus, it appears that price-sensitive buyers are getting a better deal in this market.

Table 7: Welfare Gain from Internet Markets

<table>
<thead>
<tr>
<th>Price Sensitivity quantile level</th>
<th>Welfare Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau_\alpha = 0.3 ) ( \alpha = 0.55 )</td>
<td>89%</td>
</tr>
<tr>
<td>( \tau_\alpha = 0.5 ) ( \alpha = 1 )</td>
<td>73%</td>
</tr>
<tr>
<td>( \tau_\alpha = 0.7 ) ( \alpha = 1.45 )</td>
<td>65%</td>
</tr>
</tbody>
</table>

An 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 on-line 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 side of the market. Second, the off-line markets are likely to respond to the emergence of the on-line 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. 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.

We expect that an important part of the welfare gains arises because the Internet provides US buyers with access to international markets. We evaluate this effect by re-computing the equilibrium outcomes and buyers’ welfare under the condition that US buyers may only procure services from US sellers and compare these results to the magnitudes documented above. In order to do this, we recover the distributions of sellers’ project costs conditional on seller attributes as well as the distribution of entry costs using the estimates obtained in earlier sections. We explain our methodology for doing this and comment on the recovered objects in the next section. We return to our analysis of the gains from the Internet after that.

5.4.5 Pricing Strategies, Cost Distributions and Quality Heterogeneity

We recover the distributions of the seller’s costs conditional on the seller’s attributes by combining the bid distributions of permanent sellers with the corresponding inverse bid functions:

\[
F_C(c|(q,x)) = F_B(\xi^{-1}(c|(q,x))|(q,x)).
\]

Here the inverse bid function, \( \xi(b|(q,x)) \), is derived from the first order condition of the corresponding permanent seller’s optimization problem:

\[
\xi(b_i|(q,x)_i) = b_i - \frac{P(i \text{ wins } | b_i; \sigma^E_i, \sigma^B_i)}{\frac{\partial}{\partial b} P(i \text{ wins } | b_i; \sigma^E_i, \sigma^B_i)|_{b=b_i}}.
\]
Next, we assess the magnitude of the costs of entering the auction using the model of strategic participation summarized in section 4.4. Under this model, the observed probability of participation satisfies the following equation

\[ F_E(E[\pi^p(x, k)]) = \Pr(i \in A^p(x, k)), \]

where \( F_E(.) \) is the distribution of the entry costs and \( \pi(x, k) \) is an ex-ante expected profit. We estimate the mean and standard deviation of the entry costs distribution by fitting the truncated normal distribution (truncated at 0) to the set of points implied by the ex-ante expected profit and the probability of participation values for various covariate cells and quality groups.

Due to space limitations, the figures depicting the estimated permanent sellers’ bidding functions and the probability density functions are included in the supplemental on-line appendix. Here, we briefly comment on their properties. 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 mark-ups over sellers’ costs change very slowly with 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, \( \epsilon \). As the seller’s costs increase and therefore his ability to compete on price decreases, his probability of winning increasingly depends on the realization of the \( \epsilon \) component, which in turn makes his bidding strategy less aggressive. 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’ 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 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. This indicates substantial cost heterogeneity unrelated to the quality that characterizes the participants in this market.

Further, the estimated project cost distributions have substantially lower variances relative to the variance 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 really high bids in this environment.

The estimated value for the mean and standard deviation of the entry costs are 0.082 and 0.077, respectively. That is, entry costs roughly correspond to 7% of the project cost on average. This number is slightly higher than that documented in other markets.\(^{40}\) The relatively large entry costs estimated in this market may reflect 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 sellers’ costs on the optimization problem of a permanent seller. While permanent sellers may still take reputation-related concerns into account, the incentives

\[^{40}\text{Studies of the US highway procurement market have estimated entry costs to be around 2 – 5\% of the engineer’s estimate.}\]
associated with these concerns are likely to be quite weak. A single score does not make a large impact on average reputation score once a seller has completed ten or more projects. Indeed, in the data a bad score does not make a statistically significant impact on the probability of winning or on the bid of an established seller.

5.4.6 Welfare Gains from International Trade

In this analysis we evaluate how US buyers’ welfare would change if they are shut out from the international market enabled by the Internet. Specifically, we focus on the reduction in variety and on the competitive effects associated with the exclusion of low-cost international sellers. More specifically, we take into account that (a) the distributions of costs conditional on quality differ across countries (US sellers often have higher costs relative to sellers from other countries); (b) US sellers generally have fewer quality levels than sellers from other countries (medium quality is missing).

In order to account for the cost and the variety effects we proceed in two steps. First, we re-compute equilibrium outcomes when the set of quality levels is reduced to “high” and “low” for all countries and reputation score groups. In this step, we replace the medium-quality level by the low- and high-quality levels while maintaining the relative frequencies of high and low quality sellers constant. In the second step, we eliminate international participation, i.e., we replace foreign sellers of high and low quality with US sellers of high and low quality, respectively. The last step effectively means using the quality levels and the cost distributions of US sellers in place of the quality levels and the cost distributions of foreign sellers. An algorithm to solve for bidding and pricing strategies builds on the numerical framework developed in Marshall, Meurer, Richard, and Stromquist (1994). The full details can be found in Krasnokutskaya, Terwisch, and Tiererova (2014).

The welfare analysis is summarized in table 8. We find that welfare gains to US buyers over the outside option would be reduced almost in half to 42% of the project value after the first step. The competitive pressure on both high- and low-quality sellers is reduced substantially as medium-quality sellers, who tend to have the lowest costs on average, are eliminated from the market. As a result of the price increase and the reduction in variety, many buyers choose the outside option. In the second step, prices go up still further, but this effect is small relative to the one observed in the first step, since the costs of low- and high-quality US providers differ only slightly from those of foreign sellers. This price increase induces further re-allocation of buyers toward the outside option. In the end, the gain from access to the Internet is largely reduced, as US buyers are shut out of the international market, but at 35% it remains a quite substantial improvement over the outside option. To conclude, we find that a large part of the gain from Internet trade arises from access to foreign low-cost-per-unit-of-quality sellers.

Two qualifications to this analysis are in order. First, in this analysis we again abstract from the reputation building and maintenance and hold the observed reputation scores fixed. In order to minimize the impact of this assumption, we solve for the equilibrium under the current regime and under the regime with restricted access and compare the computed equilibrium outcomes instead of comparing the outcomes recorded in the data to the those computed from the model. Second, we ignore the possible adjustment in the market participation of US sellers that might result if US buyers were restricted from hiring international sellers. In other words, we hold the number of potential sellers fixed. We thus focus just on the variety and competitive effects as we stated earlier.
Table 8: Welfare Gain from International Internet Trade

<table>
<thead>
<tr>
<th></th>
<th>Data (all quality levels)</th>
<th>High and Low Quality Only</th>
<th>US sellers Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Price</td>
<td>1.45</td>
<td>1.52</td>
<td>1.54</td>
</tr>
<tr>
<td>Average Price ($Q = H$)</td>
<td>1.71</td>
<td>1.68</td>
<td>1.70</td>
</tr>
<tr>
<td>Average Price ($Q = M$)</td>
<td>1.40</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Average Price ($Q = L$)</td>
<td>1.28</td>
<td>1.33</td>
<td>1.35</td>
</tr>
<tr>
<td>Not allocated (%)</td>
<td>10%</td>
<td>33%</td>
<td>37%</td>
</tr>
<tr>
<td>Welfare Gain</td>
<td>78%</td>
<td>42%</td>
<td>35%</td>
</tr>
</tbody>
</table>

This table reports the results of a counterfactual analysis investigating the welfare gains to US buyers from access to the international market. The second column reports the market outcomes in the environment where medium-quality foreign potential sellers are replaced by high- and medium-quality potential sellers whereas the last column presents the outcomes from a setting where foreign potential sellers are further replaced by US potential sellers of corresponding quality levels. Average prices are computed as a share-weighted average of submitted bids.

6 Conclusion

This paper makes a two-fold contribution to the literature. First, it exploits the structure of an on-line service market to provide one of the few available assessments of the welfare gains associated with the globalization of service markets facilitated by the Internet. Second, it develops a tractable framework that enables analysis of such markets.

We find that gains to buyers are quite substantial, at 73% of project value. The paper emphasizes two channels through which globalization impacts buyers’ welfare: the increase in the variety of available quality levels 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 sellers’ quality differences that are not observable in the data and to obtain unbiased estimates for the distribution of 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. Using these estimates, we establish that an important source of welfare improvement for US buyers is derived from the ability to access foreign low-cost providers of good quality on-line.

We obtain a number of important insights into the operation of on-line 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 very prevalent in the on-line (as well as off-line) 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 on-line markets.

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 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 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 buyer-specific choice sets that are prevalent in our setting.

Third, our estimation approach uses “large data” typically available to a researcher in the Internet markets. Some of the features of our setting, such as buyer-specific choice sets and self-selection into participation by sellers, have been previously addressed in the context of college choice, on-line marriage markets, enrollment in residential medical training and other environments with a matching component. The feature that sets us apart from these settings is the fact that we have private information on both sides of the market and that the pricing in our market is based on private information. It is well known that the identification of matching models could be quite challenging in the presence of unobserved heterogeneity. We are able to circumvent some of these challenges by exploiting the large number of individual observations available to us. Specifically, we facilitate uncovering sellers’ unobserved characteristics in the presence of selection into participation by conditioning on buyers’ and sellers’ observable characteristics in estimation. This procedure delivers unbiased estimates, since the selection into participation with a given buyer depends only on the buyer’s observable characteristics (as well as all sellers’ characteristics).

To the best of our knowledge, this paper marks the first effort to estimate a tractable model of the on-line procurement market. Consequently, we focus on the factors we believe are of the first-order importance - unobserved heterogeneity of sellers, private information of sellers about their costs and private information of buyers about the weights they use and their outside options – while making simplifying assumptions about 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.

In summary, we believe that our results shed light on the operation of on-line markets for services and on the gains accrued to the buyers participating in these markets. The methodology developed in this paper can be applied in other settings characterized by unobserved agent heterogeneity. Among other things it could be used to further study various aspects of (on-line) service markets: from optimal pricing and optimal procurement to the analysis of reputation-building in this environment.

References


Appendix

A. Proof of Proposition 1

Fix a set of sellers $S$. Let the set of entrants $A$ be partitioned into those who are preferred to the outside option.(denoted by $\bar{A}$) and those who are not (denoted by $\tilde{A}$). For any pair of permanent sellers $i, j$, let $A_{i,j}$ denote the support of such a partition for entrants excluding $i, j$. That is, $A_{i,j} \equiv \{(a,a') : a \cap a' = \emptyset \text{ and } a \cup a' \subseteq S\backslash\{i,j\}\}$. For any $(a,a') \in A_{i,j}$, define:

$$P_i(b;a,a') \equiv P(\text{i wins } | B_i = b, \bar{A} \backslash \{i\} = a, \tilde{A} = a')$$

for any $b \in B_i$.

**Lemma A1:** Suppose (A1) and (A2) hold. Consider any $i, j$ with $B_i \cap B_j \neq \emptyset$. (a) For any $b \in B_i \cap B_j$ and $(a,a') \in A_{i,j}$,

$$q_i \begin{cases} > & q_j \Rightarrow P_i(b;a,a') \begin{cases} > & \text{if } P_i(b;a,a') \begin{cases} > & \text{if } P_j(b;a,a'). \end{cases} \end{cases} \end{cases} (12)$$

(b) If $a^* \subseteq S\backslash\{i,j\}$ is such that either “$q_k = q_i$ or $q_k = q_j$” for all $k \in a^*$, then

$$\text{sign}(q_i - q_j) = \text{sign}(P_i(b;b^*,a') - P_j(b;b^*,a'))$$

for all $b \in B_i \cap B_j$ and any $a'$ with $(a^*,a') \in A_{i,j}$.

**Proof of Lemma A1.** Part (a). Recall entry decisions are i.i.d. binary variables with success probability $\lambda_{r,k}$ for any $i \in S^r$. In equilibrium, sellers’ bidding strategies are functions of private costs alone and are orthogonal to $(\alpha, \epsilon, U_0)$. Given any pair of disjoint sets $a, a'$ such that $(a \cup a') \subseteq N\backslash\{i,j\}$, let $\mathcal{E}(a,a')$ be a shorthand for the event “max_{s\in a^*} U_s < U_0 \leq \min_{k \in a} U_k”.

Then:

$$P_i(b;a,a') \equiv \Pr \left\{ U_i \geq \max_{k \in a} U_k \text{ and } U_i \geq U_0 \mid B_i = b, \bar{A} \backslash \{i\} = a, \tilde{A} = a' \right\}$$

$$= \int \Pr \left( \Delta_{\epsilon,i} - B_k \leq \alpha \Delta q_{i,k} - b \forall k \in a; \text{ and } U_0 - \epsilon_i \leq \alpha q_i - b \bigg| \alpha, \mathcal{E}(a,a') \right) dF(\alpha|\mathcal{E}(a,a')), (13)$$

where the equality follows from the Law of Total Probability and the facts that entry decisions are independent from realizations of $\alpha, \epsilon, C, U_0$; and that sellers’ private costs are independent across each other as well as from $\alpha, \epsilon, U_0$. By similar arguments, $P_j(b;a,a')$ takes a form that is almost identical to $P_i$ in (13), except with all indices $i$ therein replaced by $j$. By (A1) and (A2), the distribution of $(\Delta_{\epsilon_{i,j}}, B_k)_{k \in a}$ is identical to that of $(\Delta_{\epsilon_{k,j}}, B_k)_{k \in a}$ once conditioning on $\alpha$ and $\mathcal{E}(a,a')$. It then follows that (12) holds for all $b \in B_i \cap B_j$ and any $(a,a') \in A_{i,j}$.

Part (b). It is sufficient to show that weak inequalities in (12) hold strictly for all $b \in B_i \cap B_j$. Part (b).
and any \( a^* \) that satisfies the conditions in part (b). By definition of \( a^* \),

\[
P_i(b; a^*, a') - P_j(b; a^*, a') = \int \left( \Pr \left( \Delta \epsilon_{k,i} - B_k + b \leq \alpha \Delta q_{i,k} \forall k \in a^* \text{ and } U_0 - \epsilon_i + b \leq \alpha q_i \right) \left\{ \alpha, \mathcal{E}(a^*, a') \right\} \right) dF(\alpha|\mathcal{E}(a^*, a'))
\]

for all \( b \in \mathcal{B}_i \cap \mathcal{B}_j \) and \( a' \) with \( (a^*, a') \in \mathcal{A}_{i,j} \). Under (A1) and (A2), \( (B_i)_i \in a^* \) are independent from \( (\epsilon_i)_i \in a^* \) and \( \alpha \) for any given set \( a^* \). Under (A1) and (A2), \( (\Delta \epsilon_{k,i})_{k \in a^*} \) is continuously distributed with positive density conditional on \( \alpha \). Thus support of \( (\Delta \epsilon_{k,i})_{k \in a^*} \) is \( [\mathbb{E} - \varepsilon, \mathbb{E} + \varepsilon]^{\#(a^*)} \), which contains the zero vector in its interior. Likewise for \( (\Delta \epsilon_{k,j})_{k \in a^*} \). For any \( b \) in the interior of \( \mathcal{B}_i \cap \mathcal{B}_j \) there is positive probability that \( (B_k - b)_{k \in a^*} \) is close enough to 0 and \( \alpha \) is small enough so that \( P_i(b; a^*, a') > (=, <) P_j(b; a^*, a') \) for all \( a' \) with \( (a^*, a') \in \mathcal{A}_{i,j} \) whenever \( \Delta q_{i,j} > (\text{and } =, < \text{respectively}) 0 \). \( Q.E.D. \)

**Proof of Proposition 1.** By definition and an application of the Law of Total Probability, we can write \( r_{i,j}(b) \) as:

\[
\sum_{(a,a') \in \mathcal{A}_{i,j}} \Pr(i \text{ wins } | B_i = b, A^1_i = i \cup a, A^0_i = a') \Pr(A^1_i = i \cup a, A^0_i = a' | B_i = b, i \in A, j \notin A).
\]

It follows from Lemma A1 that \( \Pr(i \text{ wins } | B_i = b, A^1_i = i \cup a, A^0_i = a') \geq (\text{or } =, \leq) \Pr(j \text{ wins } | B_j = b, A^1_j = j \cup a, A^0_j = a') \) whenever \( \Delta q_{i,j} > 0 \) (or \( =, < 0 \) respectively) for all \( (a, a') \in \mathcal{A}_{i,j} \). Weak inequalities hold strictly for any \( (a, a') \in \mathcal{A}_{i,j} \) such that “either \( q_k = q_i \) or \( q_k = q_j \)” for all \( k \in a \). Such a pair \( (a, a') \) exists in \( \mathcal{A}_{i,j} \) and occurs with positive probability even after conditioning on \( B_i = b \) and \( i \in A, j \notin A \). This is because entry decisions are exogenous and independent from private costs, as each potential bidder \( i \) entering with probability \( \lambda_{r,k} \) if \( i \in S^{r,k}, r \in \{t, p\} \). The same argument applies as we switch the role of \( i \) and \( j \) in the above sentence. Furthermore, under (A1) and (A2), \( \Pr(A^1_i = i \cup a, A^0_i = a' | B_i = b, i \in A, j \notin A) \) is identical to \( \Pr(A^1_i = j \cup a, A^0_i = a' | B_j = b, j \in A, i \notin A) \) for all \( (a, a') \in \mathcal{A}_{i,j} \). It then follows that \( \text{sign}(r_{i,j}(b) - r_{j,i}(b)) = \text{sign}(q_i - q_j) \). \( Q.E.D. \)

**B. Proof of Proposition 2**

We begin by introducing some notation. For each \( x \) in the common support \( \mathcal{X} \) of \( x_i \), let \( \mathcal{Q}_x \equiv \{q_{1,x}, \ldots, q_{K_{x,x}}\} \) be the set of possible quality levels for a seller \( i \in N \) with \( x_i = x \). With each \( (x,q) \in \mathcal{X} \times \mathcal{Q}_x \) are associated sets of sellers indices, \( A^p_{(x,q),l} \equiv \{i \in A^p_l : (x_i, q_i) = (x,q)\} \), \( N^p_{(x,q),l} \equiv \{i \in N^p_l : (x_i, q_i) = (x,q)\} \), \( A^t_{x,l} \equiv \{i \in A^t_l : x_i = x\} \) and \( N^t_{x,l} \equiv \{i \in N^t_l : x_i = x\} \). It is convenient for exposition to arrange observations in a certain order. More specifically, the observations for permanent and transitory sellers are allocated into separate vectors. We enumerate observations for actual entrants first then for non-entrants, and group the observations for permanent sellers according to \( (x,q) \)-characteristics, and those for transitory sellers according to \( x \)-characteristics. Thus we write \( B_{j,l}^p \) to denote the \( j \)-th transitory seller’s bid at auction \( l \), \( B_{j,l}^p \) the \( j \)-th permanent seller’s bid at auction \( l \), \( Q_{j,l} \) the \( j \)-th transitory seller’s quality at auction \( l \), and \( W_{j,l} \in \{1, 0\} \) taking the value of one if and only if the \( j \)-th permanent seller wins at the
l-th auction. Similarly, we define \( x_{j,l}^p, x_{l,t}^p, \) and \( q_{j,l}^p \). After the rearrangement, the competitive structure of auction \( l \) is summarized by

\[
I_l \equiv \bigcup_{x \in \mathcal{X}, q \in \mathcal{Q}_x} \{ |A_{(x,q),l}^p|, |N_{(x,q),l}^p|, |A_{x,t}^l|, |N_{x,t}^l| \},
\]

where \( |A| \) for any set \( A \) denotes its cardinality. For each auction \( l \), we define \( B_l = [B_l^p, B_l^{p'})' \), where \( B_l^p \) and \( B_l^{p'} \) are random vectors with their \( j \)-th entries given by \( B_{j,l}^p \) and \( B_{j,l}^{p'} \) respectively. We also define \( Q_{N,l}^t \) and \( Q_{A,l}^t \) to be both random vectors of entries \( Q_{j,l}^t \) with \( j = 1, \ldots, |N_l^t| \) and with \( j = 1, \ldots, |A_l^t| \) respectively. We denote the set of values for \( Q_{N,l}^t \) by \( \{ q_{N,1}, \ldots, q_{N,K_{N,l}} \} \) with \( q_{N,k} = (q_{N,1,k} \cdots q_{N,K_{N,l},k}) \). Similarly, the set of values for \( Q_{A,l}^t \) is denoted by \( \{ q_{A,1,k}, \ldots, q_{A,K_{A,l},k} \} \) with \( q_{A,k} = (q_{A,1,k} \cdots q_{A,K_{A,l},k}) \). These sets change across auctions because the dimensions of \( Q_{N,l}^t \) and \( Q_{A,l}^t \) change.

In accordance with the parametric estimation approach, we assume that \( \epsilon_i \) and \( (\alpha, \beta) \) are distributed according to \( F(\epsilon_i \theta_1) \) and \( F(\alpha, \beta; \theta_2) \), distributions known up to a set of parameters \( (\theta_1, \theta_2) \), so that the vector of parameters to be estimated is given by \( \theta = (\theta_1, \theta_2, (\mathcal{Q}_x : x \in \mathcal{X}) \) along with the parameters involved in the parameterization of \( f(B_{l,t}^p | Q_{A,i,l}^t = \mathbf{q}_{A,i,k,l}, I_{l,t}) \) and \( \text{Pr}(i \in A_{x,t}^l | Q_{A,i,l}^t = \mathbf{q}_{A,i,k,l}, I_{l,t}) \).

We begin by deriving a representation of permanent seller’s winning probability conditional on the vector of bids and auction competitive structure as observed by the econometrician. Unlike the econometrician, a buyer observes all the relevant characteristics for all actual competitors. Let \( e_{(x,q),k,l}^p (B_l, I_l; \theta, j) \equiv \text{Pr}(W_{j,l}^p = 1 | B_l, Q_{A,l}^t = \mathbf{q}_{A,k,l}, I_l) \) be the probability that seller \( j \) (with \( (x,q) \)-characteristics) wins conditional on a full competitive structure of the auction, including information on transitory actual bidders’ vector of qualities. More specifically,

\[
e_{(x,q),k,l}^p (B_l, I_l; \theta, j) = \int \text{Pr} \left( \alpha q + \beta x - B_{j,l} \geq \alpha q_i + \beta x_i - B_{i,l} \forall i \neq j \mid \alpha, \beta \right) dF_{\alpha,\beta}(\alpha, \beta)
\]

Also let \( p_{k,l} = \text{Pr}(Q_{A,l}^t = \mathbf{q}_{A,k,l} | B_l, I_l) \) be the probability that the set of transitory actual bidders’ qualities is \( \mathbf{q}_{k,l} \) conditional on the vector of bids \( B_l \), and on information about the auction’s competitive structure as summarized in \( I_l \). Hence using this notation, we can rewrite

\[
\text{Pr}(W_{j,l}^p = 1 | B_l, I_l) = \sum_{k=1}^{K_{A,l}} e_{(x,q),k,l}^p (B_l, I_l; \theta, j) \text{Pr}(Q_{A,l}^t = \mathbf{q}_{A,k,l} | B_l, I_l).
\]

Let

\[
\begin{align*}
I_{l,1} & \equiv \{ |N_{(x,q),l}^p|, |N_{x,t}^l| : (x, q) \in \mathcal{X} \times \mathcal{Q}_x \}, \\
I_{l,2}^p & \equiv \{ |A_{(x,q),l}^p| : x \in \mathcal{X}, q \in \mathcal{Q}_x \} \text{ and } I_{l,2}^t \equiv \{ |A_{x,t}^l| : x \in \mathcal{X} \},
\end{align*}
\]

so that \( I_l = I_{l,1} \cup I_{l,2}^p \cup I_{l,2}^t \). We also let

\[
\omega_{k,l} = \prod_{i \in A_i^l} \text{Pr}(Q_{A,i,l}^t = \mathbf{q}_{A,i,k,l} | \mathbf{x}_{i,l}),
\]

where \( \mathbf{q}_{A,i,k,l} \) represents the vector of transitory actual bidders’ qualities in auction \( l \).
and

\[ g_k(\mathbf{B}_i^t, \mathbf{I}_{t,1}, \mathbf{I}_{t,2}^2) \equiv \omega_{k,l} \prod_{x \in \mathcal{X}} \prod_{i \in \mathcal{A}_{k,l}} f(B_{x,l}^t \mid \mathbf{Q}_{A,i,l}^t = \mathbf{q}_{A,i,k}, \mathbf{I}_{t,1}) \frac{\Pr(i \in A_{x,l}^t \mid \mathbf{Q}_{A,i,l}^t = \mathbf{q}_{A,i,k}, \mathbf{I}_{t,1})}{\mathbf{q}_{A,i,k}, \mathbf{I}_{t,1}}. \]

We prove Proposition 2 for the case with endogenous participation formulated in Section 5. We begin by fully stating conditions (A1) and (A2) in the context of the extended model.

(A1') The private signal \( E_{i,l} \) and the private cost \( C_{i,l} \) are independent from each other, and the random vectors \((E_{i,l}, C_{i,l})\) are independent across all \( i \in N_l \) and across auctions.\(^{41}\) For each \( i \) with \( q_i = q^k \), these costs are independent draws from the continuous distributions \( F_k^E \) and \( F_k^C \) with a density positive over supports \([\underline{e}_k, \bar{e}_k]\) and \([\underline{c}_k, \bar{c}_k]\) respectively.

(A2') The three random vectors \((\alpha_i, U_{0,l}, \epsilon_{i,l})\), \( \epsilon_{i,l} \) and \((E_{i,l}, C_{i,l})\) are mutually independent; match components \( \epsilon_{i,l} \) are i.i.d. across \( i \)'s; \( \epsilon_{i,l} \) and \((\alpha_l, V_{0,l})\) are continuously distributed with a density positive over \([\underline{\xi}, \bar{\xi}]\) and over \([0, \bar{\alpha}] \times [\underline{u}_l, \bar{u}_l]\) respectively.

**Proposition 2.** Suppose (A1') and (A2') hold. For each \( x \in \mathcal{X}, q \in \mathbb{Q}_x \), and for the \( j \)-th permanent seller with \((x, q)\)-characteristic who participated in auction \( l \),\(^{42}\)

\[ \Pr\{\mathbf{Q}_{A,j,l}^t = \mathbf{q}_{A,i,k}, \mathbf{I}_l \} = \frac{g_k(\mathbf{B}_i^t, \mathbf{I}_{t,1}, \mathbf{I}_{t,2}^l)}{\sum_{d=1}^{K_{A,l}} g_d(\mathbf{B}_i^t, \mathbf{I}_{t,1}, \mathbf{I}_{t,2}^l)}. \]

The quantities \( g_k(\mathbf{B}_i^t, \mathbf{I}_{t,1}, \mathbf{I}_{t,2}^l) \) involve \( f(\cdot \mid \mathbf{Q}_{A,i,l}^t = \mathbf{q}_{A,i,k}, \mathbf{I}_{t,1}) \), i.e., the density of a transitory seller’s bids conditional on this seller’s characteristics, and \( \Pr(i \in A_{x,l}^t \mid \mathbf{Q}_{A,i,l}^t = \mathbf{q}_{A,i,k}, \mathbf{I}_{t,1}) \), i.e., the probability of transitory seller \( i \)'s participation in the auction conditional on his characteristics. As mentioned earlier, we estimate these equilibrium objects jointly with the parameters of buyer’s taste distribution and quality levels. In doing so, we do not need to recover these objects separately. Since in our setting the distribution of signals is the same for permanent and transitory bidders, we can use permanent bidder’s optimization problem, bid distribution, and participation frequency to recover the distributions of signals. This requires knowing only \( g_k(\mathbf{B}_i^t, \mathbf{I}_{t,1}, \mathbf{I}_{t,2}^l) \) and not separately \( f(B_{x,l}^t \mid \mathbf{Q}_{A,j,l}^t = \mathbf{q}_{A,i,k}, \mathbf{I}_{t,1}) \) and \( \Pr(i \in A_{x,l}^t \mid \mathbf{Q}_{A,j,l}^t = \mathbf{q}_{A,i,k}, \mathbf{I}_{t,1}) \).

**Proof of Proposition 2:** For the purpose of the derivations below it is convenient to introduce mapping \( \pi(\cdot \mid N, A) : \{1, \ldots, |N|\} \rightarrow N \). This mapping plays the following role. Sometimes we need to consider a scenario where a subset of potential bidders different from the one realized in the data would choose to participate in the auction. In considering such a case we would re-arrange the observations in such a way that the observations for this hypothetical set of actual bidders are listed first and the observations for the remaining potential bidders would be listed after them. The mapping \( \pi(j \mid N, A) \) reflects the original (data set) position of the observation that would be listed in position \( j \) under this re-arrangement. In our analysis the order in which observations are listed within the set of entering or non-entering bidders is not important. Therefore, when re-arranging observations we do not consider all possible permutations (orderings) of

\(^{41}\)Our identification results could be extended to the case when \( E_{i,l} \) and \( C_{i,l} \) are correlated for each seller \( i \) in an auction, but the random vectors \((E_{i,l}, C_{i,l})\) are independent across \( i \in N_l \) and across auctions.

\(^{42}\)In fact, \( E[W_{x,q,l} \mid \mathbf{B}_i, \mathbf{I}_l] = E[W_{x,q,l} \mid \mathbf{B}_i, \mathbf{I}_l] \) for all \( j \) such that \( j \in A_{x,q,l} \) by symmetry. This formulation facilitates its sample analogue when we replace the sample version of the moment conditions for estimation.
the hypothetical set of actual bidders. Instead, we re-allocate them to the front of the vector without changing the order in which they were listed originally.

We use $\bar{N}_x, \bar{A}_x, \bar{N}_t, \bar{A}_t$ to denote the realizations of respective random sets as they are recorded in the data. Notice that $\pi(j; \bar{N}_x^p, \bar{A}_x^p) = j$ and $\pi(j; \bar{N}_t^p, \bar{A}_t^p) = j$. For simplicity, we write $\pi^i(j) = \pi(j; A^i, N_t^p)$ and $\pi^l(j) = \pi(j; A^p, N_t^l)$ whenever it is clear which $A$ and $N$ sets are used.

Notice that we consider the probability of a two-part event: (1) that a given vector of qualities characterizes a subset of potential bidders, (2) potential bidders characterized by these qualities enter.

First, as for $p_{k,l}$, note that by the Bayes rule, we can write

$$ p_{k,l} = \Pr\{Q_{A,l} = \bar{q}_{A,k} | B_l, I_l\} = \frac{f(B_l | Q_{A,l} = \bar{q}_{A,k}, I_l) \Pr\{Q_{A,l} = \bar{q}_{A,k} | I_l\}}{f(B_l | I_l)} $$

$$ = \frac{f(B_l | Q_{A,l} = \bar{q}_{A,k}) \Pr\{Q_{A,l} = \bar{q}_{A,k} | I_l\}}{f(B_l | I_l)} $$

$$ = \frac{f(B_l | Q_{A,l} = \bar{q}_{A,k}, I_{l,1}) \Pr\{Q_{A,l} = \bar{q}_{A,k} | I_l\}}{f(B_l | I_l)}. $$

The first equality holds because the bids of permanent sellers are independent of bids of the transitory sellers and do not depend on the qualities of the transitory sellers,

$$ f(B_l^p | I_l) = f(B_l^p | Q_{A,l} = \bar{q}_{A,k}, I_l). $$

We denote terms in this expression by

$$ (A) = f(B_l^p | Q_{A,l} = \bar{q}_{A,k}, I_{l,1}), \; (B) = f(B_l^p | I_{l,1}), \; (C) = P\{Q_{A,l} = \bar{q}_{A,k} | I_l\}. $$

Next, we work with these terms one by one.

Notice that $B_l^i$ are independent conditional on $Q_{A,l} = \bar{q}_{A,k}$, and $I_{l,1}$. Therefore

$$ (A) = \prod_{x \in X} \prod_{j \in A_{x,l}} f(B_l^i | Q_{A,l} = \bar{q}_{A,k}, I_{l,1}) = \prod_{x \in X} \prod_{j \in A_{x,l}} f(B_l^i | Q_{A,l} = \bar{q}_{A,k}, I_{l,1}). $$

The last equality holds because the transitory seller knows his quality but not the quality of his transitory competitors.

Applying the rule of total probability we obtain

$$ (B) = \sum_{d=1}^{K_A} f(B_l^p | Q_{A,l} = \bar{q}_{A,d}, I_l) \Pr\{Q_{A,l} = \bar{q}_{A,d} | I_l\} \quad (18) $$

$$ = \sum_{d=1}^{K_A} f(B_l^p | Q_{A,l} = \bar{q}_{A,d}, I_{l,1}) \Pr\{Q_{A,l} = \bar{q}_{A,d} | I_l\}. $$

We will return to this expression after we tackle term $(C)$.

Our goal here is to relate an event in $(C)$ to transitory bidders’ participation (entry) decisions, and to express $(C)$ in terms of the participation probabilities of the transitory bidders. First, we consider

$$ (C) = \Pr\{Q_{A,l} = \bar{q}_{A,k} | I_{l,1}, I_{l,2} = \bar{I}_{l,2}\}, $$

where $\bar{I}_{l,2} = (m_{x,q}^p, m_x^l : x \in X$ and $q \in Q_{x,l})$. Then observe that this conditional probability is
equal to

$$\Pr(Q_{A,t}) = \frac{\bar{q}_{A,k} | I_{t,1}, | A^p_{(x,q),t} | = m^p_{(x,q)}, | A^t_{x,t} | = m^t_x \text{ for all } x \text{ and } q}{= \Pr(|A^p_{(x,q),t} | = m^p_{(x,q)}, | A^t_{x,t} | = m^t_x \text{ for all } x \text{ and } q | Q_{A,t} = \bar{q}_{A,k}, I_{t,1}) \Pr(Q_{A,t} = \bar{q}_{A,k} | I_{t,1})} \Pr(|A^p_{(x,q),t} | = m^p_{(x,q)}, | A^t_{x,t} | = m^t_x \text{ for all } x \text{ and } q | I_{t,1})$$

$$= \frac{\Pr(|A^t_{x,t} | = m^t_x \text{ for all } x, Q_{A,t} = \bar{q}_{A,k} | I_{t,1})}{\sum_{d=1}^{K_A} \Pr(|A^t_{x,t} | = m^t_x \text{ for all } x, Q^t_{A,t} = \bar{q}_{A,k,d} | I_{t,1})}.$$  

(19)

The second equality holds because the events $|A^p_{(x,q),t} | = m^p_{(x,q)}$, for all $(x, q)$ and $| A^t_{x,t} | = m^t_x$, for all $x$ are independent conditional on $Q^t_{A,t} = \bar{q}_{A,k}$, $I_{t,1}$, and the event $|A^p_{(x,q),t} | = m^p_{(x,q)}$, for all $(x, q)$ is independent of $Q^t_{A,t} = \bar{q}_{A,k}$ conditional on $I_{t,1}$. We next work on the expression

$$\Pr(|A^t_{x,t} | = m^t_x \text{ for all } x, Q^t_{A,t} = \bar{q}_{A,k} | Q^t_{N,t} = \bar{q}_{N,t}, I_{t,1})$$

in the numerator of equation (19). We then return to equations (19) and (15) to conclude our derivation. We let $Q^t_{N,t} = (Q^t_{N,j,t})_{j \in N_t}$ and $Q^t_{I,t}$ be the set of values that $Q^t_{N,t}$ takes. Then

$$\Pr(|A^t_{x,t} | = m^t_x \text{ for all } x, \text{ and } Q^t_{A,t} = \bar{q}_{A,k} | I_{t,1}) = \sum_{\bar{q} \in Q^t_{N,t}} \Pr(|A^t_{x,t} | = m^t_x \text{ for all } x, \text{ and } Q^t_{A,t} = \bar{q}_{A,k} | Q^t_{N,t} = \bar{q} | I_{t,1}) \Pr(Q^t_{N,t} = \bar{q} | I_{t,1}).$$  

(20)

Further notice that

$$\Pr(Q^t_{N,t} = \bar{q} | I_{t,1}) = \Pr(Q^t_{N,t} = \bar{q} | \bar{q}^t_{I,t}) = \prod_{j \in N_t} \Pr(Q^t_{N,j,t} = \bar{q}_j | \bar{q}^t_{I,t}).$$

The probability $\Pr(Q^t_{N,j,t} = \bar{q}_j | \bar{q}^t_{I,t})$ is primitive in our environment, which characterizes the distribution of sellers’ qualities within $x-$cell. We now show how the expression for

$$\Pr \{|A^t_{x,t} | = m^t_x, \text{ and } Q^t_{A,t} = \bar{q}_{A,k} | Q^t_{N,t} = \bar{q}, I_{t,1}\}$$

can be modified and then return to equation (20). Recall that $\pi^t_i(j)$ links elements from some set $\Omega_x \subset \{1, ..., |N^t_t|\}$ to a vector $\{1, ..., |A^t_t|\}$. Then for a given $\bar{q}_{A,k}$ and $\bar{q}$ we obtain

$$\sum_{\Omega_x \subset N^t_t, \pi^t_i(j) \in \Omega_x} \prod_{j \in A^t_t} \Pr \left\{ \begin{array}{l} j \in A^t_t, \text{ and } Q^t_{A,j,t} = \bar{q}_{A,j,k} \text{ for all } \pi^t_i(j) \in \Omega_x \text{ and } j \in N^t_t, A^t_t, \text{ and } Q^t_{N,s,t} \bar{q}_{N,s,t} \text{ for all } s \in N^t_t, \bar{x}^t_i, I_{t,1} \end{array} \right\}$$

$$= \prod_{\Omega_x \subset N^t_t, \pi^t_i(j) \in \Omega_x} \Pr \left\{ \begin{array}{l} j \in A^t_t, \text{ and } Q^t_{A,j,t} = \bar{q}_{A,j,k} \text{ for all } \pi^t_i(j) \in \Omega_x \text{ and } j \in N^t_t, A^t_t, \text{ and } Q^t_{N,s,t} \bar{q}_{N,s,t} \text{ for all } s \in N^t_t, \bar{x}^t_i, I_{t,1} \end{array} \right\} \times \prod_{\pi^t_i(j) \in N^t_t, \pi^t_i|_{\Omega_x} \cap \Omega_x} \Pr \left\{ \begin{array}{l} j \in A^t_t, \text{ and } Q^t_{A,j,t} = \bar{q}_{A,j,k} \text{ for all } \pi^t_i(j) \in \Omega_x \text{ and } j \in N^t_t, A^t_t, \text{ and } Q^t_{N,s,t} \bar{q}_{N,s,t} \text{ for all } s \in N^t_t, \bar{x}^t_i, I_{t,1} \end{array} \right\},$$

where the sum over all sets $\Omega_x$ that are consistent with the restrictions imposed on the set of
entrants, i.e., $\Omega_x \subset \bar{N}_{x,l}$ such that $|\Omega_x| = m^t_x$, $\tilde{q}_{\pi_l(j)} = \tilde{q}_{A,j,k}$ for all $j$ such that $\pi_l^t(j) \in \Omega_x$. Next,

$$
= \sum_{\Omega_x \subset \bar{N}_{x,l}} \prod_{\pi_l^t(j) \in \Omega_x} \Pr \left\{ j \in A_{x,l}^t, Q_{A,j,l}^t = \tilde{q}_{A,j,k} | Q_{N,l,l}^t = \tilde{q}_{\pi_l^t(s)} \text{ for all } s \in N^t_l, I_{l,1} \right\}
\times \prod_{\pi_l^t(i) \in \bar{N}_{x,l} - \Omega_x} \Pr \left\{ i \in N^t_{x,l} - A_{x,l}^t | Q_{N,l,l}^t = \tilde{q}_{\pi_l^t(s)} \text{ for all } s \in N^t_l, I_{l,1} \right\}
= \sum_{\Omega_x \subset \bar{N}_{x,l}} \prod_{\pi_l^t(j) \in \Omega_x} \Pr \left\{ j \in A_{x,l}^t | Q_{A,j,l}^t = \tilde{q}_{A,j,k}, I_{l,1} \right\}
\times \prod_{\pi_l^t(i) \in \bar{N}_{x,l} - \Omega_x} \Pr \left\{ i \in N^t_{x,l} - A_{x,l}^t | Q_{N,l,l}^t = \tilde{q}_{\pi_l^t(i)}, I_{l,1} \right\}
$$

Notice that for every $\Omega_x$ the set of qualities within $\Omega_x$ and $\bar{N}_{x,l} - \Omega_x$ is the same. Therefore, the expression above can be written

$$
\sum_{\Omega_x \subset \bar{N}_{x,l}} \prod_{\pi_l^t(j) \in \Omega_x} \Pr \left\{ j \in A_{x,l}^t, Q_{A,j,l}^t = \tilde{q}_{A,j,k}, I_{l,1} \right\}
\times \prod_{\pi_l^t(i) \in \bar{N}_{x,l} - \Omega_x} \Pr \left\{ i \in N^t_{x,l} - A_{x,l}^t | Q_{N,l,l}^t = \tilde{q}_{\pi_l^t(i)}, I_{l,1} \right\}
\times \prod_{\pi_l^t(i) \in \bar{N}_{x,l} - \Omega_x} \Pr \left\{ i \in N^t_{x,l} - A_{x,l}^t | Q_{N,l,l}^t = \tilde{q}_{\pi_l^t(i)}, I_{l,1} \right\}
$$

Here, $|\Omega^x|$ denotes the cardinality of set $\Omega^x = \{ \Omega_x : \Omega_x \subset \bar{N}_{x,l}, \text{ such that } |\Omega_x| = m^t_x \text{ and } \tilde{q}_{\pi_l^t(j)} = \tilde{q}_{A,j,k}, \text{ for all } \pi_l^t(j) \in \Omega_x \}$, with $\Omega^0_x$ representing one specific member of $\Omega^x$. For example, we can set $\Omega^0_x = A_{x,l}^t$.

Returning with expression (21) to equation (20) obtains

$$
\sum_{\tilde{q} \in \tilde{Q}_l^N} \Pr(|A_{x,l}^t| = m^t_x \text{ for all } x, \text{ and } Q_{A,l}^t = \tilde{q}_{A,k}, Q_{N,l}^t = \tilde{q}, I_{l,1}) \Pr(Q_{N,l}^t = \tilde{q} | I_{l,1})
= \sum_{\tilde{q} \in \tilde{Q}_l^N} \prod_{x \in X^t} \prod_{\pi_l^t(j) \in \Omega_x^0} \Pr \left\{ j \in A_{x,l}^t, Q_{A,j,l}^t = \tilde{q}_{A,j,k}, I_{l,1} \right\}
\times \prod_{\pi_l^t(i) \in \bar{N}_{x,l} - \Omega_x^0} \Pr \left\{ i \in N^t_{x,l} - A_{x,l}^t | Q_{N,l,l}^t = \tilde{q}_{\pi_l^t(i)}, I_{l,1} \right\} \Pr(Q_{N,l}^t = \tilde{q} | I_{l,1})
= \sum_{\tilde{q} \in \tilde{Q}_l^N} \prod_{x \in X^t} \prod_{j \in A_{x,l}^t} \Pr \left\{ j \in A_{x,l}^t, Q_{A,j,l}^t = \tilde{q}_{A,j,k}, I_{l,1} \right\}
\times \prod_{\pi_l^t(i) \in \bar{N}_{x,l} - A_{x,l}^t} \Pr \left\{ i \in N^t_{x,l} - A_{x,l}^t | Q_{N,l,l}^t = \tilde{q}_{\pi_l^t(i)}, I_{l,1} \right\} \prod_{j \in N^t_l} \Pr(Q_{N,l}^t = \tilde{q} | I_{l,1})
$$

Note that in the last summation over $\tilde{q} \in \tilde{Q}_l^N$, part of the vectors in $\tilde{q}$ such that $(\tilde{q}_j)_{j \in A_{x,l}^t}$, and hence the summation is essentially over vectors in $\tilde{Q}_l^{N-A_{x,l}^t}$ which is the set of values for
\[ Q_{N-A}^t \equiv (Q_{N-A_{j,l}}^t)_{j \in N_i-A^t_i}. \] Thus we write the last sum as

\[
\sum_{A \in Q^{N-A}_{x,A}} \prod_{j \in A} \Pr \{ j \in A_j^t | Q_{A,j,l}^t = \tilde{q}_{A,j,k}, I_{l,1} \} \prod_{j \in A_i^t} \Pr(Q_{N,j,l}^t = \tilde{q}_j | \tilde{x}_j^t) \prod_{i \in N_{x,i}^t-A_{i,l}^t} \Pr \{ i \in N_i^t-A_{i,l}^t | Q_{i,l}^t = \tilde{q}_i, I_{l,1} \} \prod_{j \in A_i^t} \Pr(Q_{N,j,l}^t = \tilde{q}_j | \tilde{x}_j^t) \]

\[ = \prod_{x \in X} \prod_{i \in N_{x,i}^t-A_{i,l}^t} \prod_{A^t_j} \{ \Pr \{ i \in N_i^t - A_{i,l}^t | Q_{i,l}^t = \tilde{q}_i, I_{l,1} \} \Pr(Q_{N,i,l}^t = \tilde{q}_i | \tilde{x}_l^t) \}. \]

The expression in (22) is derived for an arbitrary \( \tilde{q}_{A,k} \). Therefore, we substitute it into both the numerator and the denominator of (19). Notice that the dimensionalities of actual and potential bidders’ x-sets are the same in the numerator and the denominator and that is why both expressions contain the common factor:

\[
\sum_{A \in Q^{N-A}_{x,A}} \prod_{x \in X} \prod_{i \in N_{x,i}^t-A_{i,l}^t} \{ \Pr \{ i \in N_i^t - A_{i,l}^t | Q_{i,l}^t = \tilde{q}_i, I_{l,1} \} \Pr(Q_{N,i,l}^t = \tilde{q}_i | \tilde{x}_l^t) \}
\]

Therefore, after canceling out this factor, the expression in (19) transforms into

\[
\Pr(Q_{A,j,l}^t = \tilde{q}_{A,j,k}, I_{l,1}, I_{l,2} = \tilde{I}_{l,2}) = \frac{\prod_{x \in X} \prod_{j \in A_{j,l}^t} \{ \Pr \{ j \in A_j^t | Q_{A,j,l}^t = \tilde{q}_{A,j,k}, I_{l,1} \} \Pr(Q_{A,j,l}^t = \tilde{q}_{A,j,k} | \tilde{x}_j^t) \}}{\sum_{d=1}^{K_i} \prod_{x \in X} \prod_{j \in A_{j,l}^t} \{ \Pr \{ j \in A_j^t | Q_{A,j,l}^t = \tilde{q}_{A,j,k}, I_{l,1} \} \Pr(Q_{A,j,l}^t = \tilde{q}_{A,j,k} | \tilde{x}_j^t) \}}.
\]

Having obtained an expression for (C), we now return to (B).

Denote \( \omega_{A,k,l} = \prod_{j \in A_i^t} \Pr(Q_{A,j,l}^t = \tilde{q}_{A,j,k} | \tilde{x}_j^t) \) and write

\[
(B) = \sum_{k=1}^{K_A} f(B_j | Q_{A,j,l}^t = \tilde{q}_{A,j,k}, I_{l,1}) \Pr(Q_{A,j,l}^t = \tilde{q}_{A,j,k} | I_l)
\]

\[ = \sum_{k=1}^{K_A} \omega_{A,k,l} \prod_{x \in X} \prod_{j \in A_{j,l}^t} f(B_j | Q_{A,j,l}^t = \tilde{q}_{A,j,k}, I_{l,1}) \Pr \{ j \in A_j^t | Q_{A,j,l}^t = \tilde{q}_{A,j,k}, I_{l,1} \}
\]

Finally, combining (A), (B), and (C) obtains

\[
\frac{\omega_{A,k,l} \prod_{x \in X} \prod_{j \in A_{j,l}^t} f(B_j | Q_{A,j,l}^t = \tilde{q}_{A,j,k}, I_{l,1}) \Pr \{ j \in A_j^t | Q_{A,j,l}^t = \tilde{q}_{A,j,k}, I_{l,1} \}}{\sum_{d=1}^{K_i} \omega_{A,d,l} \prod_{x \in X} \prod_{j \in A_{j,l}^t} f(B_j | Q_{A,j,l}^t = \tilde{q}_{A,j,d}, I_{l,1}) \Pr \{ j \in A_j^t | Q_{A,j,l}^t = \tilde{q}_{A,j,d}, I_{l,1} \}}.
\]

**C. Moments Used in Estimation**

The estimation is based on two sets of moment conditions. The first set of moments relates the probability that a permanent seller wins when the sets of permanent actual and potential sellers satisfy certain restrictions. The second set links transitory and permanent sellers’ empirical distribution of bids and participation frequencies to their theoretical counterparts.
The first set of moments consists of three subsets:

(1a) Moments that are based on the permanent seller’s probability of winning in an auction where two or more active permanent bidders belong to the same group. In these moment conditions, we compute expectations of the following functions: a constant (equal to one), the difference between the winning bid and a bid submitted by a permanent bidder from the same group, and the squared difference between the winning bid and a bid submitted by a bidder from the same group.

(1b) Moments that are based on the permanent seller’s probability of winning in an auction where he competes with one or more active permanent bidders belonging to a different group. In these moment conditions, we compute expectations of the following functions: a constant (equal to one), the difference between the winning bid and a bid submitted by a seller from a different group, the squared difference between the winning bid and a bid submitted by a seller from a different group, respectively. We include moments for all possible pairs of different groups.

(1c) Moments that are based on the permanent seller’s probability of winning in an auction where he competes with one or more active permanent bidders belonging to a different group, and at least one transitory active bidder belonging to a specific country group. In these moment conditions, we compute expectations of the following functions: the product of transitory bid and the differences between the winning bid and the bid of a permanent seller from a different group, the product of transitory bidder’s characteristics other than price (e.g., the number of available scores, and a current average of scores) and the differences between the winning bid and the bid of a permanent seller from a different group. We include moments for all possible pairs of different permanent sellers’ groups and all possible country groups of transitory sellers.

We supplemented moment conditions in the first set by the following moment conditions:

(1d) The probability that project is not allocated.

(1e) The first and the second moments of prices submitted by active permanent bidders of a specific group when project is not allocated. We include such moment for each group of active permanent sellers.

The identification strategy which motivates moments in the first set is shown to identify the composite function $g_k(.)$ from Proposition 2. The remaining challenge, therefore, lies in separately identifying the distribution of transitory sellers’ bids, transitory sellers’ probability of participation and unconditional probability of observing a given group in the population of transitory sellers. These objects are multiplicatively combined in the expression for $g_k(.)$. The identification of these objects is aided by the moments in the second set.

The second set of moments matches the following empirical moments to their theoretical counterparts: the mean and the variance of the transitory bid distributions, as well as the frequencies with which potential transitory sellers submit a bid aggregated to the level observed in the data. We include such moments for every permanent seller group, or correspondingly every combination of transitory seller’s country group, the number of his ratings and the current average of his reputation scores. We additionally imposed the expected profit conditions that summarize the optimal participation decision of transitory bidders for each group of transitory
sellers. These conditions impose the restriction that in equilibrium only potential bidders with entry costs below the ax-ante expected profit value should participate.\textsuperscript{43}

\textsuperscript{43}The identifying assumption here is that the the distribution of the costs of entry is the same for all groups.