It's All in the Interval
An imperfect measurements approach to estimate bidders' primitives in auctions

David Sundström
Department of Economics
Umeå School of Business and Economics
Umeå University

Abstract

A novel method to measure bidders' costs (valuations) in descending (ascending) auctions is introduced. Based on two bounded rationality constraints bidders' costs (valuations) are given an imperfect measurements interpretation that is robust to behavioral deviations from the traditional rationality assumptions. Theory provides no guidance as to the shape of the cost (valuation) distributions while empirical evidence suggests them to be positively skew. Consequently, a flexible distribution is employed in an imperfect measurements framework. An illustration of the proposed method on Swedish public procurement data is provided along with a comparison to a traditional Bayesian Nash Equilibrium approach.

Keywords: latent variable, log-generalized gamma distribution, maximum likelihood, prediction, public procurement

JEL classification: C24, C51, D22, D44

1 Introduction

This paper deals with the methodological issue of measuring bidders' costs (valuations) in descending (ascending) auctions. The proposed method relies on very few (two) economic assumptions in the form of bounded rationality constraints and this constitutes the primary novelty of the current paper. It is robust to deviations from the strongly rational behavior commonly assumed in the literature, e.g., it allows for disequilibrium bidding, optimization errors and agents that lack well-defined strategies. The methodological philosophy gathers more inspiration from imperfect measurements in statistics and less influence by game theory than is common in cost (valuation) estimation using auction setups. Hereafter I will consider descending auctions and the estimation of costs as opposed to ascending auctions and the estimation of valuations. This makes the proposed method no less general as the two types of setups is the inverse of the other.

Measuring costs is useful in the following sense. In studies of auctions (e.g., Paarsch, 1992; Laffont et al., 1995) one sometimes observes empirical specifications such as

$$\boldsymbol{c} = m\left(\boldsymbol{X}\right) + \boldsymbol{\varepsilon} \tag{1}$$

where $\boldsymbol{c}$ is a vector of costs, $\boldsymbol{X}$ contains regressors, and $\boldsymbol{\varepsilon}$ is unobserved random heterogeneity. Often $m\left(\boldsymbol{X}\right) = \boldsymbol{X}\boldsymbol{\beta}$ where $\boldsymbol{\beta}$ is a vector of parameters. Using a specification like (1), rather than a specification of the type $\boldsymbol{b} = f\left(\boldsymbol{X}\right) + \boldsymbol{\nu}$, where $\boldsymbol{b}$ is a vector of bids, would most likely provide better empirical relationships because in contrast to the expression for $\boldsymbol{b}$, (1) is structural as it captures the underlying relationship between the firm's cost and relevant conditioning information (Rezende, 2008); a firm's cost reveals something about its fundamental structure while the size of its bid may be contaminated by strategic behavior. A complication is that, as compared to the bid, the cost is unobserved.

According to mainstream economic theory, it is the underlying structure of the firm, essentially the production function that determines the firm's cost. Consider a single firm's bid of size b and its underlying cost, c, to carry out the contract for which it requests b. Intuitively $b > c$; otherwise it would not profit from the transaction. It is plausible that the difference $b - c$, often called markup, changes with respect to the institutional setting. If the structure of the market in question is that of perfect competition then economic theory implies that the bid perfectly represents the firm's cost. In reality however, as compared to theory, not many markets are perfectly competitive. This is one of the reasons why effort has been made to estimate firms' costs, starting with Rosse (1970) in its modern form. Much of the research since then is based on game-theoretical models and many results rely entirely on game-theoretical arguments grounded on strong rationality and behavioral assumptions. While this strand of research has provided many useful insights, the current paper adopts another approach, mainly due to robustness arguments against violations of the traditional behavioral rationality assumptions.

The general idea is to see the observed bids as noisy information on firms' costs that, despite their noisiness, carry useful information. Using appropriate methods to handle the noise, estimates of the latent cost variable may be obtained; a measure of the firm's underlying structure that is sought. This measure can then be utilized to estimate structural parameters and markups of firms. It is a common exercise within the Empirical Industrial Organization literature to estimate the distributions of such costs.[1] In the majority of these studies the costs are obtained using guidance from game theory. Typically, by assuming Bayesian Nash Equilibrium (BNE) behavior one can obtain an expression for the optimal bid. The BNE approach, however, is laden with assumptions. Most work in the Bayesian Nash-paradigm is based on Vickrey's (1961) auction theoretical contribution in conjunction with Harsanyi's writings on strategic behavior under imperfect competition (Harsanyi, 1967, 1968a,b). This strand of research generated many structural econometric works on auctions in the New Empirical Industrial Organization literature (see footnote 1) where Paarsch (1992) is considered to be the first such application and Guerre et al. (2000) perhaps the most influential. Porter (1995), Laffont (1997) and the textbook of Paarsch and Hong (2006) provide surveys of the research area.

The aim of the present paper is to estimate the same phenomenon, the cost, but from a different angle relying on a small and transparent set of intuitively appealing assumptions. The endeavor here is to provide a serious approach to estimate costs given weak economic structure. I want to break free from the Bayesian Nash straitjacket but still take the particular structure of auctions seriously. To do this I invoke statistical methods of imperfect measurements. The proposed method is potentially robust to the restrictive rationality assumptions of the BNE approach. So, in a sense this paper is a novel methodological contribution increasing the breadth of ways to approach and think about the problems in estimating costs in the framework of auctions.

The outline of the paper is as follows. The next section describes the general idea of the proposed method. Section 3 deals with the interpretation of heterogeneity. Section 4 provides an empirical illustration including a comparison between the method presented in this paper and a mainstream BNE approach. The final section concludes by discussing the content of the paper and possible extensions of this study.

2 The General Idea

The assumptions I make are bounded rationality constraints partly inspired by Haile and Tamer (2003)[2] and entirely intuitively appealing. They are

[1]Two interesting texts on the method of Empirical Industrial Organization is the progress report by Einav and Levin (2010) and the critique by Angrist and Pischke (2010).

[2]In addition to Haile and Tamer (2003), the spirit of Varian's (1982; 1983; 1984; 1985) suite of papers on a nonparametric approach to consumer and producer behavior looms over the general

A1 The firm will not bid under its cost.

A2 Competition is an efficient sorting mechanism.

The first assumption is rather weak and straightforward; it states that $b > c$. The second assumption says that no firm will allow a rival to make profit if the firm's cost structure allows it to beat its rival's bid. In symbols this is stated as $c_1 < b_1 < c_2 < b_2$ where the subindex specifies firm. It implies that a low-cost firm will cast a lower bid than a high-cost firm will do, i.e. that bids are monotone in costs. That is, the auction allocates the contract efficiently in the economic sense of the word.

Putting the two assumptions to work, I can bound the firms' costs into intervals. To describe the procedure, index the firms by $i = 1, 2, ..., N$. As an example, suppose that we observe bids of firms $i = (1, 2, 3)$ in a given auction and that their magnitudes can be described by

$$b_1 < b_2 < b_3 \tag{2}$$

That is, firm 1 is the lowest bidder and firm 3 the highest. Then, considering assumptions A1 and A2 above, the cost of firm 1 is in the interval between zero and its bid, that is $c_1 \in [0, b_1]$. Also, $c_2 \in [b_1, b_2]$ and $c_3 \in [b_2, b_3]$. That is, the cost of the firm is bounded from above by its bid and bounded from below by the bid of the competitor that just underbids the firm. Consider firm 3. It is unlikely that $c_3 \geq b_3$ as then the firm would not profit from that transaction; this event is ruled out by assumption A1. If $c_3 < b_2$ then it would profit from casting a lower bid than b_3 and this event is ruled out by assumption A2.

Here, as in the BNE literature, I assume that the firm knows its cost. That is, the cost of a firm to carry out the project in a given auction, c, is deterministic from the viewpoint of the firm. On the other hand, c is stochastic to the econometrician and its generating process is generally seen as

$$b = cu \tag{3}$$

where $u > 1$. Hence, u can be seen as a markup-factor which is close to one if the markup is low and larger if the markup is higher. Taking logarithms (but skipping the ln-operator to avoid cluttering notation) and rearranging we obtain

$$c = b - u \tag{4}$$

One reason why I take the logarithm in (4) will become clear in the next section. Note that the ln-transformation is monotonic, so the discussion below (2) still holds. Also, note that $u > 0$ holds in (4) after the ln-transformation.

Now, let us interpret the cost of a firm in an auction as an imperfectly observed variable. Consider us wanting to measure firm 2's cost of fulfilling the

idea of this paper.

contract in a given auction. In the light of assumptions A1 and A2 as well as the examples above, we are able to to conclude that, e.g., $c_2 \in [b_1, b_2]$. In other words, c_2 is an interval-censored random variable. Using techniques taking the random cost variable to be interval censored, point estimates of the firms' costs can be obtained along with estimates of the cumulative distribution functions (cdf) and the probability density functions (pdf) of the costs. That is, features that enable structural analysis in the sense discussed in the introduction can be recovered. Naturally, this implies that one will be able to draw structural conclusions as well.

Just as in this approach, the more traditional BNE literature provides an expression for the cost as composed of two parts, the bid minus the markup. The BNE type of models assume that knowing the bid and the markup will enable the calculation of the cost as a point on the real line. The strong rationality assumptions used in the BNE framework in conjunction with calculating the cost as a point on the real line are potentially sensitive and prone to give erroneous estimators if real-world behavior deviates from the model. This is why I use this more conservative bounding approach in the spirit of Manski (1999). The BNE approach will be described in more detail in Section 4.1, where I compare it to the new method that I present in this paper.

So, the discussion above implies that the event $(b_1 < c_2 < b_2)$ is considered when collecting information on firm 2's cost. The probability of the event is $P(b_1 < c_2 < b_2)$. Using (4) I write this as

$$P(b_1 < c_2 < b_2) = P(c_1 + u_1 < c_2 < c_2 + u_2) \tag{5}$$

When imposing a stochastic structure on this economic framework, the model generates u_1 and u_2 as independent random variables. We can see this by subtracting c_2 from the inequality in (5) as follows

$$P(u_2 > 0 \cap c_2 - c_1 > u_1) = P(u_2 > 0)\, P(c_2 - c_1 > u_1)$$

which can be written in terms of the distribution functions of the random variables as

$$F_{u_1}\left(c_2 - c_1, \sigma^2_{u_1}\right)\left[1 - F_{u_2}\left(0, \sigma^2_{u_2}\right)\right]$$

where $\sigma^2_{u_i}$ is the variance of the distribution. So, the model assumes that markups are independent. That is, it allows for boundedly rational behavior, such as optimization errors that are independent across the bidders. It does not allow for correlatedness of, e.g., optimization error between the agents. The independence-relation over firms' markups is in line with earlier work in this research area, see the surveys mentioned in the introduction of this paper.

This framework and bid data enable the researcher to obtain an estimate of the cost of each firm in a sample, and estimate a density over each auction, a density of the costs of all firms in all auctions, costs in different years, and so forth. It also allows one to study a firm's markup, that is, the general difference between its bid and its cost and how markups differ with respect to the number of bidders and other institutional changes.

3 Heterogeneity and Estimation

Turning from the theoretical to the real world forces one to handle heterogeneity that is unobserved. How this heterogeneity manifests itself is key when the choice of a proper way to handle it is to be made. If all auctions were homogenous, a firm's cost of delivering a service or good would not differ between the auctions. In practice, of course, the auctions are not homogenous, but heterogenous. An example is that it is cheaper to provide cleaning services for one square meter of office space than to clean one square meter of a hospital (due to, e.g., different sanitary requirements). Therefore the firm's cost will naturally differ in fulfilling these contracts. Accordingly, a method to homogenize the cost intervals with respect to contract heterogeneity would be useful. I choose a parametric estimation framework for the estimation procedure to calculate c in (4) from data. The major argument for assuming a parametric setting is that in most applications, researchers face auctioned objects that are not identical as the office/hospital case exemplifies. If the researcher observes some conditioning information in the form of covariates, say $\boldsymbol{X}$, she can homogenize the costs with respect to the information contained in $\boldsymbol{X}$, i.e. make the auctioned objects more similar. The exercise of doing so is straightforward in parametric setups. As opposed to parametric approaches, nonparametric approaches suffer from the curse of dimensionality, i.e. that the amount of data points needed increases exponentially in the number of covariates. If the dimensionality of $\boldsymbol{X}$ is large and/or if $\boldsymbol{X}$ contains continuous variables this problem becomes severe; it hinders the researcher to use the conditioning information in practice. On the other hand, parametric estimation frameworks suffer from the inverse problem, i.e. the observations in data are seen as indirect observations on the parameters themselves. This makes the choice of the particular parametric model crucial as the correct model has to be assumed at the outset (see, e.g., the excellent discussion in Mittelhammer et al. (2000)). Hence, the more flexible parametric specification, the better.

Empirical evidence (O'Hagan et al., 2003) suggests that costs do not seem to be symmetrically distributed. In addition, costs assume values on the positive part of the real line, ruling out distributions supported on $(-\infty, \infty)$. In many studies of auctions, the log-normal distribution is chosen due to flexibility arguments. In this paper I make an even more flexible parametric assumption: I assume the costs to be log-generalized gamma distributed. This distributional assumption includes a skewness parameter. It contains the log-normal distribution as a special case and thereby the estimation of costs is made more flexible than in the studies utilizing the log-normal distribution. So, I want to make use of conditioning information that is common to find in auction datasets, but still keep some flexibility. Additionally, as the log-normal distribution is a special case of the log-generalized gamma distribution, I can test the assumption of log-normally distributed costs often made in other parametric studies.

Now, I describe the estimation procedure in more detail. As already discussed, the idea is to estimate in a regression setup and then use the predicted costs from which observed heterogeneity has been partialed out. Then, one can use the model to predict the expected value conditional on it being in the observed interval, i.e. obtain

$$E\left[c_2 \mid b_1 < c_2 < b_2\right] \tag{6}$$

The observed heterogeneity accounts for some of the variance in the dependent variable, the cost. This heterogeneity is a matrix of p variates, $\boldsymbol{X} = \left(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p\right)$ of which a row, such as $\boldsymbol{x}^i = \left(x_{i1}, x_{i2}, \ldots, x_{ip}\right)$, $i = 1, ..., N$, is observed heterogeneity associated with cost observation c_i. That is, $\left(c_i, \boldsymbol{x}_i\right)$ is generated by each bid and can be linked as in the following logarithmic cost function:

$$c_{ik} = \boldsymbol{x}_{ik}'\boldsymbol{\beta} + \sigma\eta_{ik},\ i = 1, ..., N;\ k = 1, ..., K \tag{7}$$

where σ and $\boldsymbol{\beta} = \left(\beta_1, \beta_2, \ldots, \beta_p\right)'$ are unknown parameters, k indicates the auction, and η_{ik} is distributed as

$$f\left(\eta; \lambda\right) = \begin{cases} \frac{|\lambda|\left(\lambda^{-2}\right)^{\lambda^{-2}}}{\Gamma\left(\lambda^{-2}\right)} \exp\left[\lambda^{-1}\eta - \lambda^{-2}\exp\left(\lambda\eta\right)\right], & \lambda \neq 0 \\ \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\eta^2}{2}\right), & \lambda = 0 \end{cases} \tag{8}$$

Here λ is an unknown parameter that determines the skewness of the density and (8) is the pdf of the log-generalized gamma distribution. This distribution was discussed by Prentice (1974) who discussed it in the light of the generalized gamma distribution proposed by Stacy (1962). The density is positively skewed whenever $\lambda > 0$, negatively skewed when $\lambda < 0$, and symmetric in case $\lambda = 0$. This is advantageous in the light of the discussion on flexibility above. Further, (8) includes some well-known densities as special cases: if η is on log-scale it, e.g., includes the normal distribution when $\lambda = 0$ and the log-gamma distribution when $\lambda > 0$ and $\sigma = 1$ (Farewell and Prentice, 1977).

In this setting c_{ik} is observed up to an interval. To be consistent with earlier notation I call the lower interval bound b_1 and the upper interval bound b_2. The cost is denoted by c_2. When a particular auction k is considered they are denoted as b_{1k}, b_{2k} and c_{2k}, respectively. The log-likelihood function is

$$\ln L(\boldsymbol{\theta}) = \sum_{i \in \overline{w}} \sum_{k=1}^{K} \ln L_{\overline{w}k}\left(\eta_{b_1}, \eta_{b_2}; \boldsymbol{\theta}\right) + \sum_{i \in w} \sum_{k=1}^{K} \ln L_{wk}\left(\eta_{b_2}; \boldsymbol{\theta}\right) \tag{9}$$

where $\boldsymbol{\theta} = \left(\lambda, \sigma, \boldsymbol{\beta}'\right)'$, w is the set of all winning bids, $\overline{w}$ is the set of all other bids,[3] K is the total number of auctions, and

[3] Note that the set of winning bids, w, consists of the typical observation $c_1 \in [0, b_1]$ (in levels, in logs $\ln(c_1) \in [-\infty, \ln(b_1)]$), i.e. left-censored random cost variables with 0 as lower bound as the gamma distribution is supported on $[0, \infty]$ (when c is in levels but its logarithm is supported on $[-\infty, \infty]$), while $\overline{w}$, the complement of the set w contains all other observations, such as $c_2 \in [b_1, b_2]$, i.e. interval-censored random cost variables.

$$\ln L_{\overline{w}k}\left(\eta_{b_1},\eta_{b_2};\boldsymbol{\theta}\right)=\begin{cases}\ln\left[I\left(\lambda^{-2},\psi_1\right)-I\left(\lambda^{-2},\psi_2\right)\right], & \lambda>0\\ \ln\left[I\left(\lambda^{-2},\psi_2\right)-I\left(\lambda^{-2},\psi_1\right)\right], & \lambda<0\\ \ln\left[\Phi\left(\eta_{b_{2k}}\right)-\Phi\left(\eta_{b_{1k}}\right)\right], & \lambda=0\end{cases}\tag{10}$$

where $\psi_j=\lambda^{-2}\exp\left[\lambda\left(\eta_{b_{jk}}\right)\right]$, $\eta_{b_{jk}}=\left(b_{jk}-\boldsymbol{x}'_{jk}\boldsymbol{\beta}\right)/\sigma$ and $j=1,2$. Also, $\Phi\left(\cdot\right)$ is the cdf of the standardized normal distribution. Further,

$$\ln L_{wk}\left(\eta_{b_2};\boldsymbol{\theta}\right)=\begin{cases}\ln\left[1-I\left(\lambda^{-2},\psi_2\right)\right], & \lambda>0\\ \ln\left[I\left(\lambda^{-2},\psi_2\right)\right], & \lambda<0\\ \ln\left[\Phi\left(\eta_{b_{2k}}\right)\right], & \lambda=0\end{cases}\tag{11}$$

In (10) and (11), $I\left(\cdot\right)$ is the incomplete gamma integral, defined as

$$I\left(k,b_2\right)=\frac{1}{\Gamma\left(k\right)}\int_0^{b_2}x^{(k-1)}e^{-x}dx$$

The log-likelihood function in (9) is maximized with respect to $\boldsymbol{\theta}$ to obtain the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}=\arg\max_{\boldsymbol{\theta}}\ln L(\boldsymbol{\theta})$. Using $\widehat{\boldsymbol{\theta}}$ and observed covariates, one can calculate the expected value of the costs given that they are contained in the proper interval as in (6). In general, this expected value $E\left[c_2\mid b_1<c_2<b_2\right]$ is defined as

$$E\left[c_2\mid b_1<c_2<b_2\right]=\frac{1}{F\left(b_2\right)-F\left(b_1\right)}\int_{b_1}^{b_2}tf(t)dt\tag{12}$$

where $F\left(\cdot\right)$ is a cdf and $f(\cdot)$ the corresponding pdf. The integral in (12) can be calculated using standard numerical methods. In the case when $\lambda=0$ in $f(\cdot)$, the natural logarithm of the costs is normally distributed and the raw cost variable is log-normally distributed. In these two cases closed form expressions for (12) are available and displayed in the Appendix.

When the costs have been estimated according to the procedure above, they can be used in regression setups to empirically study the relationship between bids and costs, i.e. expression (4). An illustration of this is provided in the section below.

4 Empirical Illustration and Comparison to a Mainstream Approach

The method outlined above will be applied to data on Swedish public procurements and compared to a mainstream BNE approach. The data consist of 3399 bids from two datasets on fixed price cleaning service contracts auctioned between 1992-1998 and 2009-2010. I use the data to study how the price of buying and cost of selling these services have evolved over the years when the Swedish

Table 1: Descriptive statistics (sample size $N = 3399$)

Variable	Mean	s.d.	Min	Max
1. Bid (to clean one square meter over a year)	149.3	94.16	13.03	1676.4
2. Local level (YES=1, NO=0)	0.731	0.443	0.000	1.000
3. Office (YES=1, NO=0)	0.340	0.474	0.000	1.000
4. Wage (in $SEK \times 1000^{-1}$)	18.94	4.698	13.43	24.43
5. Length of contract (in years)	2.116	0.777	0.170	4.250
6. Option to prolong contract (YES=1, NO=1)	0.934	0.555	0.000	1.000
7. Number of bids in auction	8.335	4.771	1.000	37.00

Public Procurement Act was young as compared to more recent times. This information will not only give an opportunity to study whether the effective transaction price, i.e. the winning bid has changed, but also whether the cost of providing the services has changed. This will implicitly assess whether the market power has changed over the years allowing for interesting policy implications.

Descriptive statistics are found in Table 1. A short description of the variables follows. Bid is the size of the bid standardized to be interpretable as the bid a firm requests to clean one square meter during a period of one year adjusted for inflation. Local level is a dummy variable that indicates whether the procurement was conducted by a municipality authority or not. The third variable in the table, office, indicates whether the lot to be cleaned was an office or not. This variable describes the cost shift between offices and other types of lots to be cleaned. It is not as costly to clean an office as compared to, e.g., a hospital as a hospital requires more rigorous cleaning procedures. Wage is the inflation-adjusted[4] mean wage. The length of the contracts is measured in years. A longer contract is advantageous to the firm as it does not have to participate in new procurements as often. Option to prolong contract is whether there is a clause in the contract that makes it possible to prolong the contract for an additional period of time without a new bidding process. If there is such an option an incentive is provided to behave more orderly. The final variable shows how many bids that were cast in the auction. Admittedly, some of the variables in Table 1 are non-standard in cost-function setups. Nevertheless, it is better to use some relevant conditioning information than none, to increase the precision of the estimator, and even so, the variables have economic interpretation as argued above.

The parameters of the interval log-generalized gamma regression model in (7) are estimated, i.e., a cost function is estimated and variables two to six in Table 1 are used as conditioning information. The likelihood is maximized using

[4]Reference year for the inflation adjustment is 2000 and this variable is collected from Statistics Sweden. Wage statistics for the cleaning industry are unavailable for parts of the time period under consideration. Instead I use information on the average wage for the cohorts that did and did not graduate high school but was never enrolled in college or at a university.

Table 2: Parameter estimates from model in expression (7)

Variable	$\hat{\beta}$	s.e.
Intercept	4.647	0.069
Local level (YES=1, NO=0)	0.172	0.034
Office (YES=1, NO=0)	-0.314	0.028
Wage (in $SEK \times 1000^{-1}$)	0.010	0.003
Length of contract	-0.015	0.012
Option to prolong contract (YES=1, NO=0)	-0.089	0.024
$\hat{\sigma}$	0.508	0.007
$\hat{\lambda}$	0.028	0.022
N	3399	
R^2_{pseudo}	0.02	
Log-likelihood	10599.98	

the BFGS algorithm. Starting values are obtained by regressing the mid-points of the intervals on the regressors using ordinary least squares and the starting value of λ was set to zero.[5] The parameter estimates are presented in Table 2. The intercept is interpreted as the firms' fixed cost of carrying out a project. A positive (negative) $\hat{\beta}$ reveals a positive (negative) effect of the corresponding variable on the firms' costs. The parameter estimates in the mean function are all signed as expected, e.g., higher wages lead to larger costs and it is less costly to clean offices as compared to other objects. The $\hat{\lambda}$ is not different from zero in a statistical significance sense. As logged bids are utilized in (9) this implies that I cannot reject that the logged costs are normally distributed. Hence, I cannot reject that the costs measured in levels are log-normally distributed. The pseudo-R^2 in Table 2 is defined in McFadden (1974) and calculated as follows: Let $\ln L(\boldsymbol{\theta})$ denote the log-likelihood function being maximized, $\ln L_0(\boldsymbol{\theta})$ and $\ln L_{fit}(\boldsymbol{\theta})$ its value in the intercept-only model and the fitted model, respectively. Then

$$R^2_{pseudo} = 1 - \frac{\ln L_{fit}(\boldsymbol{\theta})}{\ln L_0(\boldsymbol{\theta})}$$

Moreover, a likelihood ratio-test rejects, on the five percent significance level the null hypothesis that the parameters in the fitted model are jointly equal to zero.

The idea is to use the predicted values of the estimated model as the cost estimates, given that these predicted costs are contained in the proper intervals, as in (12).[6] The result from this exercise is visualized in Figure 1 where we observe

[5] In addition I tried 1 as starting value for λ but $\hat{\lambda}$ differed only on the third decimal when doing so.

[6] As $\hat{\lambda} = 0.028$ I used the normal distribution when we estimated the conditional mean in (12). The reason is that the pdf of the log-generalized gamma distribution contains the number $\Gamma\left(\lambda^{-2}\right)$ which is extremely large when $\lambda = 0.028$, in fact it is so large that it cannot be handled by most computer programs. As an example, the software R that is used for estimations in this paper, can

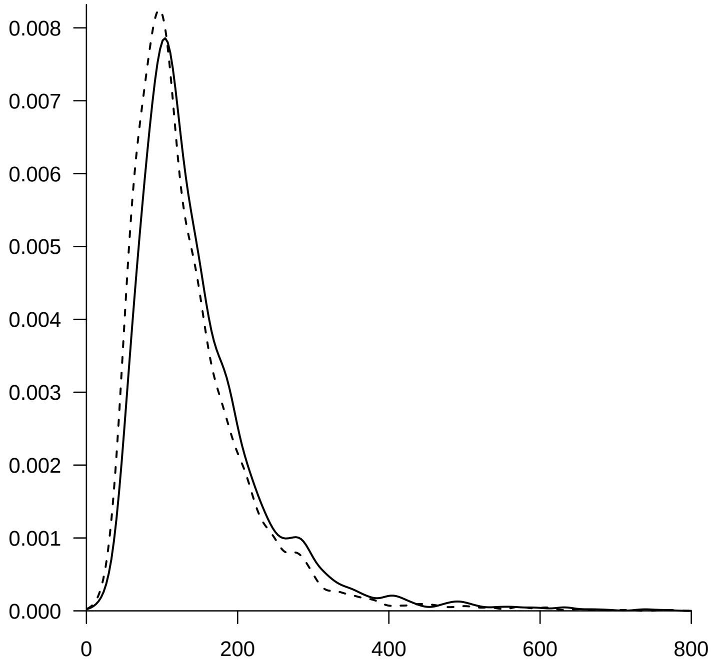

Figure 1: Bid (solid) and cost (hatched) distributions. Note: bids and costs in levels as in Table 3.

the intuitive result that the bids dominate the costs stochastically. Summary statistics of bids and predicted costs across years are given in Table 3.

Now, I use the predicted costs to study expression (4) empirically, i.e. I estimate the parameters in

$$b_{ik} = \alpha \hat{c}_{ik} + u_{ik} + z_{ik}' \gamma,\ i = 1, ..., N;\ k = 1, ..., K \tag{13}$$

where $\hat{c}_{ik}$ is the predicted cost of firm i in auction k, α is the effect of the predicted cost on the bid, z_{ik} is a vector of strategic variables and γ their effect on the bid. The z_{ik} may contain information that is assumed to influence the size of the firm's bid via its strategic behavior but not through its cost *per se*, such as a measure of competition in the auction. The $u_{ik} = \mu + \xi_{ik}$ is the markup and $u_{ik} > 0$ (in logs and $u_{ik} > 1$ in levels) according to Section 2.

Table 4 displays ordinary least squares estimates of the parameters in specification (13). Note that the estimates are generated by a model where both bids

at the very least handle a value of λ approximately equal to 0.080, that is $\Gamma\left(0.080^{-2}\right)$ which is approximately equal to 1.692×10^{274}.

Table 3: Bids and estimated costs across years

Year	Median bid	Median cost	Mean bid	Mean cost
1992	92.749	64.595	116.324	90.957
1993	156.089	136.011	208.204	184.453
1994	141.612	126.292	153.960	139.724
1995	140.728	129.305	159.494	145.610
1996	140.476	124.509	159.020	145.546
1997	185.514	151.777	210.203	178.939
1998	117.542	109.215	147.277	126.907
2009/1	109.627	101.948	115.494	105.638
2009/2	119.783	107.694	129.069	116.993
2010/1	108.091	99.805	114.371	103.338
2010/2	112.733	101.931	157.709	134.225

Note: 2009/1 means January-June 2009 and 2009/2 July-December 2009. Interpretable as the bid/cost to clean one square meter during one year.

and costs are measured in natural logarithms (as in expression (4) in Section 2). We see that the parameter estimates of α vary from 0.898 to 1.058 over the years. A parameter estimate of 0.898 indicates that a one percent increase in cost for a certain contract will lead to an approximately 0.9 percent increase in the firm's bid in that year. That is, here the firms' markups decrease when their costs increase.

Figure 2 shows a picture of the estimated markups, $\hat{u}_{ik}$, using the full sample. The $\hat{u}_{ik}$ in Figure 2 are associated to bids and costs measured in levels, and we see that all $\hat{u}_{ik}$ are larger than one, as the simple model (3) requires. The observation that all $\hat{u}_{ik} > 1$ illustrates that the parametric approximations correspond well to the model (3). Descriptive statistics relating to Figure 2 are found in Table 5 below in the IAI column.

Also, the $\hat{\gamma}$ is the magnitude of the effect that one additional bid in an auction has on the bid.

All in all, I conclude that there is no obvious trend in the markup over the years in the data at hand.

4.1 Comparison to a Mainstream Bayesian Nash Equilibrium Approach

Now, I provide a comparison of my proposed method to the common procedure proposed by Guerre et al. (2000). Their proposition has its foundations on a BNE assumption which restricts behavior to error-free equilibrium rational behavior as was discussed in the introduction. Given the BNE assumption it is very flexible as it is nonparametric in an econometric sense. But this is also a drawback as observed heterogeneity cannot be used in other than very limited ways, due to the curse of dimensionality. Another drawback of their approach

Table 4: Parameter estimates from model in expression (13). Note: bids and costs in natural logarithms

Year	Intercept	s.e.	$\hat{\alpha}$	s.e.	$\hat{\gamma}$	s.e.
1992	-0.112	0.343	1.058	0.054	0.018	0.079
1993	0.557	0.150	0.898	0.035	0.020	0.015
1994	0.416	0.061	0.946	0.014	-0.007	0.003
1995	0.445	0.038	0.946	0.008	-0.010	0.001
1996	0.369	0.055	0.961	0.011	-0.010	0.002
1997	0.945	0.203	0.940	0.034	-0.079	0.016
1998	0.675	0.200	0.940	0.035	-0.059	0.020
2009/1	0.549	0.083	0.916	0.019	-0.008	0.002
2009/2	0.463	0.087	0.932	0.018	-0.004	0.001
2010/1	0.580	0.065	0.911	0.014	-0.008	0.001
2010/2	0.600	0.083	0.931	0.016	-0.018	0.003

Note: 2009/1 means January-June 2009 and 2009/2 July-December 2009.

as compared to mine is that their estimation procedure requires the researcher to supply information on the number of competitors. In many cases the question on how many competitiors the bidders observe is dubious. That is, it is not clear to the econometrician how many firms the bidders see as a real threat: if not all firms participate, it is unclear whether it is the potential or the actual number of bidders that firms' see as competition as I discuss elsewhere (Sundström, 2016). Nevertheless, according to a BNE analysis where agents play monotone strategies, the optimal bid of firm i is

$$b_i = c_i + \frac{[1 - G(b_i)]}{g(b_i)} \frac{1}{N-1} \tag{14}$$

where the second term on the right hand side is the markup and $G(\cdot)$ and $g(\cdot)$ denote the cdf and pdf of the bids, respectively. Guerre et. al (2000) suggests substituting the nonparametric estimators $\hat{G}(\cdot)$ and $\hat{g}(\cdot)$ for $G(\cdot)$ and $g(\cdot)$ in (14). As $\hat{G}(\cdot)$ and $\hat{g}(\cdot)$ are nonparametric, a choice of bandwidth must be supplied to the estimation procedure. The N is the degree of competition where some measurement of competition must be used, e.g., either the potential or actual number of competitors depending on assumptions and available measurements. Here I choose N to be the number of received bids, i.e. the actual number of bids. To calculate $\hat{g}(\cdot)$ I use the Triweight kernel, $(35/32)\left(1 - z^2\right)^3 \mathbb{1}(|z| < 1)$, whose bandwidth is chosen as $h = 1.06\hat{\sigma}_b (M)^{-1/5}$ where $\hat{\sigma}_b$ is the sample standard deviation of the bids and M the sample size. These choices of kernel and bandwidth are the same as in the Monte Carlo experiments of Guerre et. al (2000). Here, $\hat{h} = 19.63$. Some observations drop out due to the trimming procedure to make the estimator $\hat{g}(\cdot)$ unbiased in the tails as described by Guerre et. al (2000). In total, a number of 7 observations drop out, effectively reducing the sample size from 3399 to 3392 observations. Table 5 provides descriptives of the esti-

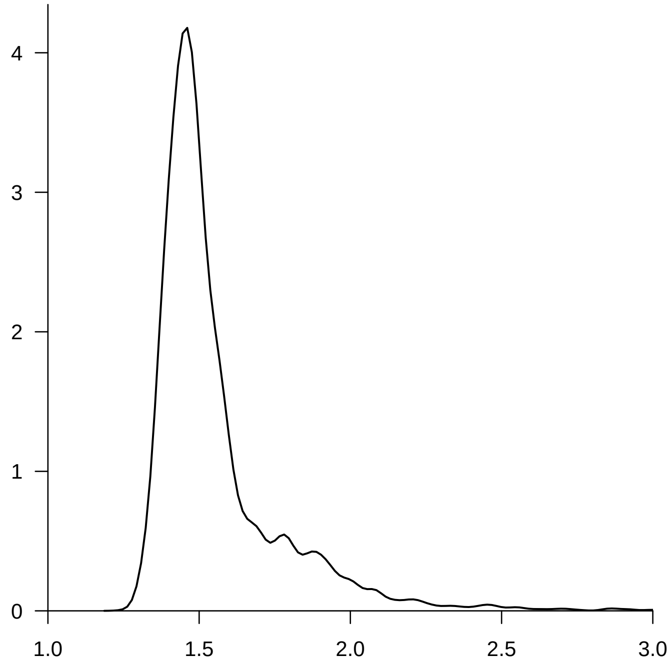

Figure 2: Density of the estimated multiplicative factors $\hat{u}_{ik}$ for the full sample corresponding to u in expression (3). The horizontal axis shows the size of u, i.e. the multiplicative factor on the cost. The horizontal axis is truncated to make the exposition clearer, but there are only 20 $\hat{u}_{ik} > 3$, and $\max \hat{u}_{ik} = 8.9$.

mated costs, $\hat{c}_{ik}$ and estimated markup factors, $\hat{u}_{ik}$, for the full sample using the method described in this paper (IAI) and the method proposed by Guerre et al. (2000) (GPV), respectively. In the table we see that they differ quite substantially. We also see that the GPV approach generates very small multiplicative markups. Further, we observe that the GPV method yields a mean of the estimated costs equal to 149.2 while the mean of the bids is 149.9. On the other hand, the mean estimated cost is 133.8 according to the approach proposed in this paper.

I also test whether the predicted costs differ in a distributional sense. Here I perform two sample Kolmogorov-Smirnov tests of equality of the distributions of $\hat{c}_{ik}$ as well as the $\hat{u}_{ik}$ estimated in the IAI and GPV frameworks, respectively. The null hypothesis is that the distributions are equal. The Kolmogorov-Smirnov test statistic considers the maximum absolute difference of the empirical cdf of the IAI and GPV cost estimates, respectively, as $D = \max_{\hat{c}} \mid \hat{F}_{IAI}(\hat{c}) - \hat{F}_{GPV}(\hat{c}) \mid$ and analogously for $\hat{u}$. As can be seen in Table 5 the null hypothesis is rejected

Table 5: Comparison between estimated costs and markup factors using the method studied in this paper (IAI) and Guerre et. al (GPV)

	$\hat{c}_{ik}$		$\hat{u}_{ik}$		
Statistic	IAI	GPV	IAI	GPV	Bids
Mean	133.815	149.241	1.572	1.011	149.280
Median	112.475	124.861	1.488	1.010	124.823
s.d.	79.503	94.164	0.322	0.008	94.164
Min	11.841	10.134	1.252	1.009	13.031
Max	769.436	1676.427	8.917	1.425	1676.427
Kolmogorov-Smirnov test	$\hat{c}_{ik}$		$\hat{u}_{ik}$		
$\hat{D}$	0.093		0.999		
p-value	2.3×10^{-13}		$< 2.2 \times 10^{-16}$		
Sample size	3399	3392	3399	3392	3399

Note: The descriptives of the bids are given here for convenience. They were given earlier in Table 1. The sample size for the GPV statistics is smaller because of the trimming of the extreme observations in the GPV procedure.

in both the cost- and the markup distributions.[7] Further, considering Table 5 we see that the mean multiplicative markup factor is 1.57 calculated using the IAI approach and 1.011 using the GPV approach.

To summarize the main differences between the two approaches we see that in the approach suggested by Guerre et al. (2000), one has to make two choices: kernel[8] and bandwidth. As it is nonparametric there is a problem with respect to the curse of dimensionality. On the other hand, misspecifications due to parametric assumptions is not an issue. In the approach suggested in this paper I can use conditioning information in a straightforward way as this is a parametric approach and it does not suffer from the dimensionality curse. That being said, imposing a parametric structure is always restrictive as compared to a nonparametric alternative. However, the log-generalized gamma parametric setup used here enjoys more flexibility than what is common in most parametric auction setups.

5 Concluding Remarks

In this paper, I have discussed a novel approach to extract information about bidders' costs (and valuations) in auctions. It is novel in the sense that it im-

[7] I also conducted Kolmogorov-Smirnov tests on the equality of distributions between the actual bids and the estimated cost distributions under the IAI and GPV frameworks, respectively. I reject that the bids and costs estimated in the IAI framework share the same distribution (p-value 1.9×10^{-13}) while I cannot reject the bids and costs estimated in the GPV framework are equally distributed (p-value 1).

[8] I also tried the Epanechnikov kernel and, as expected, the results did not change much.

poses only minimal economic structure in the form of two bounded rationality constraints. This makes the proposal robust to disequilibrium behavior, optimization errors and play in strategies that are not well-defined. The estimated costs allow the applied researcher to perform structural analyzes in auction settings, i.e. to study how different market characteristics influence unobserved phenomena such as the bidders' costs and markups. Conditional on the intuitively appealing assumptions, a latent variable interpretation of bidders' costs in the form of an interval censored random variable was given. Also, costs may be asymmetrically distributed. Therefore, a log-generalized gamma maximum likelihood estimator for interval censored data was proposed to make the procedure robust to and enabling tests for deviations from symmetry.

I detached from the traditional approach to empirically assess auctions when I separated from the traditional research domain of the Bayesian Nash-paradigm. Bayesian Nash analysis has contributed substantially to the understanding of auctions, but may not be very robust to behavioral departures from the strong rationality assumptions on the agents that it presumes. The assumptions made in this paper, that (1) no one bids above their cost and that (2) competition is an efficient sorting mechanism, impose less structure on bidders' behavior and are more robust to, e.g., disequilibrium play and optimization errors made by the economic agents.

Conditional on the two assumptions, bidders' costs were given an interval censored random variable interpretation. Using this interpretation, the rest of the estimation philosophy was parametric, albeit a robust one thanks to the flexible properties of the log-generalized gamma distribution. The parametric maximum likelihood approach is motivated as there are observed heterogeneities that should be conditioned on. Some recent estimation methods of auctions relies on nonparametric techniques. Such nonparametric techniques suffer from the curse of dimensionality as the need for data points increases exponentially in the number of covariates that is utilized. This is unfortunate as most data sets on auctions consist of a quite small number of observations, and a fair amount of conditioning variables.

The illustration on Swedish public procurement data showed that there is no trend in the evolution of firms' markup over years since the Public Procurement Act was implemented. The distribution of costs in this particular dataset was skewed to the right, but I could not reject that its natural logarithm was symmetrically distributed. This can be seen in Figure 1 and also be deduced from the estimate $\hat{\lambda} = 0.028$ along with its standard error (0.022) in Table 2 and implies that it cannot be rejected that the raw cost variable is log-normally distributed. In this case the standard procedure of interval regression as proposed by Stewart (1983) (that is based on a normality assumption) could have been utilized with logged bids. However, this may not be true for other data and the log-generalized gamma approach is well motivated by robustness arguments. Also, the test of log-normality was enabled by this approach.

Further, the method introduced in the current paper is easily implemented as it basically requires (1) the maximization of a likelihood function, (2) prediction using the estimated model obtained in (1) and (3) using estimated costs in other inquiries. Sometimes, obtaining the costs is the objective and then (3) is superfluous. The main advantage is, however, that the economic model forming the foundation of the method is easily understood by practitioners, for whom this method is designed.

The comparison of the method discussed in this paper to a traditional BNE approach as suggested by Guerre et al. (2000) showed that the mean markup factors were 1.572 and 1.01, respectively, which reveals quite a wide difference in results. This discrepancy opens up for interesting discussions about models and data generating processes in general. Say that the assumptions A1 and A2 of Section 2 generally holds, and further that (14) actually accurately describes the behavior of at least a subset of the economic agents. Can we determine that a certain subset of the economic agents that adheres to the decision rule (14), i.e. agents whose behavior can be described by rational equilbrium play? If so, can we increase the precision in cost estimation by using a combination of the model outlined in Section 2 and the specification in (14)? These are questions that are interesting on a both philosophical and practical level that deserve attention in future studies.

Acknowledgments

Many thanks to Kurt Brännäs for guidance and comments. I am grateful to Suzanne Bijkerk, Bengt Kriström, Sofia Tano and Mattias Vesterberg for helpful comments and suggestions. I also thank seminar participants at the Department of Economics, USBE, Umeå University, at the Research Institute of Industrial Economics (Stockholm, November 2015) and conference sessions at the Spring Meeting of Young Economists (Ghent, May 2015) and the Annual Conference of the International Association for Applied Econometrics (Thessaloniki, June 2015) for valuable comments. Sofia Lundberg and Elon Strömbäck are thanked for sharing their data for the empirical illustration. All errors are mine.

References

Angrist, J. and Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, 24:3–30.

Einav, L. and Levin, J. D. (2010). Empirical industrial organization: A progress report. *Journal of Economic Perspectives*, 24:145–162.

Farewell, V. T. and Prentice, R. L. (1977). A study of distributional shape in life testing. *Technometrics*, 19:69–75.

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

Haile, P. A. and Tamer, E. (2003). Inference with an incomplete model of english auctions. *Journal of Political Economy*, 111:1–51.

Harsanyi, J. C. (1967). Games with incomplete information played by bayesian players, i–iii: Part i. *Management Science*, 14:1804–1817.

Harsanyi, J. C. (1968a). Games with incomplete information played by bayesian players, i–iii: Part ii. *Management Science*, 14:320–334.

Harsanyi, J. C. (1968b). Games with incomplete information played by bayesian players, i–iii: Part iii. *Management Science*, 14:486–502.

Laffont, J.-J. (1997). Game theory and empirical economics: The case of auction data 1. *European Economic Review*, 41:1–35.

Laffont, J.-J., Ossard, H., and Vuong, Q. (1995). Econometrics of first-price auctions. *Econometrica*, 63:953–980.

Manski, C. F. (1999). *Identification problems in the social sciences*. Harvard University Press.

McFadden, D. (1974). *Conditional logit analysis of qualitative choice behavior*, volume 1 of *Frontiers in Econometrics*, pages 105–142. Academic Press.

Mittelhammer, R. C., Judge, G. G., and Miller, D. J. (2000). *Econometric Foundations*, volume 1. Cambridge University Press.

O'Hagan, A., Stevens, J. W., et al. (2003). Assessing and comparing costs: how robust are the bootstrap and methods based on asymptotic normality? *Health economics*, 12:33–49.

Paarsch, H. J. (1992). Deciding between the common and private value paradigms in empirical models of auctions. *Journal of Econometrics*, 51:191–215.

Paarsch, H. J. and Hong, H. (2006). An introduction to the structural econometrics of auction data. *MIT Press Books*, 1.

Porter, R. (1995). The role of public information in u.s. offshore oil and gas lease auctions. *Econometrica*, 63:1–28.

Prentice, R. L. (1974). A log gamma model and its maximum likelihood estimation. *Biometrika*, 61:539–544.

Rezende, L. (2008). Econometrics of auctions by least squares. *Journal of Applied Econometrics*, 23:925–948.

Rosse, J. N. (1970). Estimating cost function parameters without using cost data: Illustrated methodology. *Econometrica*, 38:256–275.

Stacy, E. W. (1962). A generalization of the gamma distribution. *The Annals of Mathematical Statistics*, 33:1187–1192.

Stewart, M. B. (1983). On least squares estimation when the dependent variable is grouped. *The Review of Economic Studies*, 50:737–753.

Sundström, D. (2016). The competition effect in a public procurement model: An error-in-variables approach. *Umeå Economic Studies*, 920.

Varian, H. R. (1982). The nonparametric approach to demand analysis. *Econometrica*, 50:945–973.

Varian, H. R. (1983). Non-parametric tests of consumer behaviour. *The Review of Economic Studies*, 50:99–110.

Varian, H. R. (1984). The nonparametric approach to production analysis. *Econometrica*, 52:579–597.

Varian, H. R. (1985). Non-parametric analysis of optimizing behavior with measurement error. *Journal of Econometrics*, 30:445–458.

Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed tenders. *The Journal of Finance*, 16:8–37.

Appendix: The Expected Value in Two Truncated Distributions

In the case of the normal distribution, that is when $\lambda = 0$ in (8), the expected value in (12) can be shown to be

$$E\left[c_2 \mid b_1 < c_2 < b_2\right] = c_2 - \sigma \left[\frac{\phi\left(b_2\right) - \phi\left(b_1\right)}{\Phi\left(b_2\right) - \Phi\left(b_1\right)}\right]$$

where $\Phi\left(\cdot\right)$ and $\phi\left(\cdot\right)$ is the cdf and the pdf of a standardized normal random variable, respectively. When we consider the log-normal distribution (12) is

$$E\left[c_2 \mid b_1 < c_2 < b_2\right] = \exp\left(c_2 + \frac{\sigma^2}{2}\right)\left[\frac{\Phi\left(b_2 - \sigma\right) - \Phi\left(b_1 - \sigma\right)}{\Phi\left(b_2\right) - \Phi\left(b_1\right)}\right] \tag{A.1}$$

I use (A.1) in the empirical illustration of Section 4.