Identification and Estimation in Highway Procurement Auctions under Unobserved Auction Heterogeneity

Elena Krasnokutskaya*
Department of Economics, Yale University

November 24, 2002

Abstract

The accurate assessment of participants' private information may critically affect policy recommendations in auction markets. The estimation of the private information distribution may be complicated by the presence of unobserved heterogeneity. This problem arises when some of the information available to all bidders at the time of the auction is subsequently not observed by the researcher. This paper develops a semi-parametric method that allows a researcher to uncover the distribution of bidders' private information in a standard First-Price procurement auction when unobserved auction heterogeneity is present. Sufficient identification conditions are derived and a two-stage estimation procedure to recover bidders' private information is developed. The procedure is applied to data from Michigan highway procurement auctions and compared to the estimation procedures traditionally used in the context of highway procurement auctions. The estimation results suggest that ignoring unobserved auction heterogeneity is likely to result in substantially biased estimates and may lead to erroneous policy recommendations.

Keywords : Unobserved heterogeneity, First-Price auctions, empirical analysis of auction data, highway procurement auctions

JEL classification: C1, C7, D8, L0, L5

*I would like to thank my advisors Martin Pesendorfer, Steve Berry and Hanming Fang for their help and encouragement. I also thank Don Andrews, Han Hong, Philip Hail and Katsumi Shimotsu for helpful conversations, and my friends Zoran Crnja, Slava Kargin, David Love, Lingfeng Li and Michael Vaney for their support and help. I am grateful to the Michigan Department of Transportation for providing data for my research and to Mr. Wayne Roe in particular for answering my questions and helping me to understand the industry better.

1 Introduction

The accurate assessment of bidders' private information may crucially affect policy recommendations in auction markets. The evaluation of the magnitude of informational rents (mark-ups over the bidders' costs), the choice of an optimal reserve price and the design of an optimal auction mechanism depend crucially on the distribution of bidders' private information. Several methods have recently been proposed that allow us to infer bidders' private information from the bids they placed. These methods maintain the assumption of no unobserved heterogeneity. If unobserved heterogeneity is present they may produce biased estimates. Unobserved heterogeneity arises when the researcher does not observe some relevant information available to all bidders at the time of the auction.

This paper develops a semi-parametric method that allows the researcher to recover the bidders' cost distribution in a standard First-Price auction under unobserved auction heterogeneity. I derive identification conditions, develop an estimation method that yields uniformly consistent estimates, and apply this method to the data from Michigan highway procurement auctions.[1] The estimation results suggest that the bias introduced by disregarding unobserved auction heterogeneity can be severe.[2]

The paper builds on the literature that uses the equilibrium relationship between costs and bids to recover the distribution of bidders' private information. This literature relies on the assumption that all variation in bidders' costs is attributable to variation in bidders' private information and variation in observable characteristics of the project. In a number of environments this assumption can be violated. One of the possible violations is related to the presence of unobserved auction heterogeneity. Unobserved auction heterogeneity may arise if important information used by bidders is not included in data set due to legal restrictions or other considerations. For example, in highway procurement a detailed description of the project is available to all bidders but usually is not observed by the researcher.

I consider a simple First-Price procurement auction model with unobserved heterogeneity. The model assumes that a bidder's cost equals the product of a common and an individual component. The common component consists of cost attributes which are observed by all bidders but not by the researcher. The individual component consists of additional cost attributes privately observed by each bidder. The cost structure readily translates into equilibrium bids separable into a common and an individual bid component. The common component varies from auction to auction, and

[1] According to the U.S. Census of Construction industries, every year \$30 to \$40 billion is spent on highway and street construction activities. In Michigan, the average amount spent on highway procurement is around \$1.4 billion per year. Given the magnitude of the expenditures, ensuring the integrity and efficiency of these auctions is of considerable social interest.

[2] In particular, at the bid level of \$570,000, the model with unobserved heterogeneity estimates a mark-up over the average bidders costs equal to 6.3%. For the same bid value, the model with independent private values estimates a mark-up of 13%; the model with affiliated private values estimates a mark-up of 19%.

so do all distributions of bidders' cost.

The separability of equilibrium bids forms the basis to my identification result for the distributions of the common and the individual cost components. Identification is possible if the individual cost component is independently distributed across bidders and is independent from the common component. The identification result relies on a theorem by Kotlarski (1967), which provides conditions for identification of distributions of three independent components from the joint distribution of two random variables. Identification of the individual cost component distribution additionally relies on the result by Laffont and Vuong (1996) who establish identification of the model with independent private values and asymmetric bidders.

Having dealt with identification, I explain how the validity of the model's assumptions can be tested from observable data. Specifically, I describe those properties of the data that ensure the existence of a model with unobserved heterogeneity that could have generated the data. I also derive restrictions which independence of cost components imposes on the data.

Next, I describe an estimation procedure that can be used to recover the bidders' cost distribution under the presence of unobserved heterogeneity. The reasoning noted in the context of model identification still applies - the distribution of costs cannot be directly uncovered from the distribution of observable bids. The multiplicative structure of cost, however, allows me to decompose the distribution of bids into the distribution of common cost component (unobserved heterogeneity) and individual bid component. I use the multiple indicators estimator developed by Li and Vuong (1998) to estimate the distributions of the common and individual bid components conditional on bidder characteristics. I then extend the independent values estimator for the model of symmetric bidders, developed by Guerre, Perrigne, and Vuong (2000), to the asymmetric case. This extension allows me to uncover the distribution of the individual cost component from the distribution of the individual bid component. I show that estimators are uniformly consistent.

Using a Monte-Carlo study I show that the estimation procedure behaves well even in small samples. I further demonstrate that models which ignore unobserved heterogeneity tend to underestimate costs and overestimate markups. They also tend to allocate larger mass to the left tail of the cost distribution. These biases may result in erroneous policy recommendations. The over-estimated mark-ups exaggerate inefficiencies that arise in First-Price auctions due to bidder asymmetry, while the described distortions in the shape of the cost distribution function tend to induce the auctioneer to set the reserve price at a lower than optimal level. The importance of both biases increases as the variation in the unobserved heterogeneity component increases.

Finally, I analyze highway procurement auction data obtained from the Michigan Department of Transportation. The sample period is 1997 to 2002. Regression analysis suggests that a large part of the variation in both the mean and the variance of bids arises from an unobserved auction-specific component. That is why I use the methodology developed in the earlier sections of the paper to recover the distribution of bidders costs. I estimate three models: a model with unobserved heterogeneity; an independent private values model, and a model with affiliated private values. I depict the estimated bid functions and the density functions of total costs for all three models. The estimates suggest that variation in the common component explains a large part of the bid variation. Under the null hypothesis of unobserved heterogeneity both the independent and the affiliated private values model tend to under-estimate the cost and over-estimate the mark-up. They also allocate larger mass to the left tail of the cost distribution. Estimation results suggest that the normalization of bids by engineer's estimate, an independent cost proxy for the project, does not fully account for unobserved heterogeneity.

I conclude with a summary and suggestions for further research. A four-part Appendix completes the paper: part A contains proofs, part B depicts results of the Monte Carlo study, part C presents results of the descriptive analysis, and part D describes estimation results.

2 Related Literature

Hendricks and Porter (1988) verify how implications of auction theory conform to bidders' behavior observed in real data.[3] Donald and Paarsch (1993, 1996), and Laffont, Ossard and Vuong (1995) develop parametric methods to recover the distribution of cost from the observed distribution of bids. Elyakime, Laffont, Loisel and Vuong (1994, 1997) propose a nonparametric method to estimate distribution of cost. Guerre, Perrigne and Vuong (2000) study identification of the First-Price auction model with symmetric bidders. They establish that the distribution of bidders' valuations can be identified from bid data if and only if the empirical inverse bid function is increasing. They propose a uniformly consistent estimation procedure. Li, Perrigne and Vuong (2001a,b) extend the result to the affiliated private values and the conditionally independent private values models. Campo, Perrigne and Vuong (2001) prove identification and develop a uniformly consistent estimation procedure for First-Price auctions with asymmetric bidders and affiliated private values. They assume that a vector recording the number of bidders by type summarizes the effect of unobserved heterogeneity on the bid distribution. In all cases, identification relies on the monotonicity of the appropriate inverse bid function. Athey and Haile (2002) establish identification results for auction models with special emphasis on possible data limitations. They address the issue of unobserved heterogeneity in the context of Second-Price and English auctions.

Chakraborty and Deltas (1998) propose a methodology for estimation of the bidders' valuation distribution under unobserved heterogeneity. They assume that the distribution of bidders' valuations belongs to a two-parameter distribution family. This assumption is used to derive small sample estimates for the corresponding parameters of the auction-specific valuation distributions. These estimates are later regressed on the observable auction characteristics to determine the per-

[3] Hendricks and Paarsch (1995) and Laffont (1997) provide surveys of empirical auction studies. Perrigne and Vuong (1995) describe estimation methods developed in the context of auction data.

cent of values variation that could be attributed to unobserved heterogeneity. The methodology is applied to data for packages of real estate loans.

Highway procurement auctions have already been studied in the literature. Porter and Zona (1990) find evidence of collusion in Long Island highway procurement auctions. Bajari and Ye (2002) reject the hypothesis of collusive behavior in the procurement auctions conducted in Minnesota, North Dakota and South Dakota. Jofre-Bonet and Pesendorfer (2002) find evidence of capacity constraints in California highway procurement auctions. Hong and Shum (2002) find some evidence of common values in the bidders' costs in the case of New Jersey highway procurement auctions. They account for unobserved auction heterogeneity by modelling the median of the bid distribution as a normal random variable with a mean that depends on the number of bidders.

3 Model

This section describes the First-Price auction model under unobserved heterogeneity and summarizes properties of the equilibrium bidding strategies.

The seller offers a single project for sale to m bidders. Bidder's i cost is equal to the product of two components: one is common and known to all bidders, the other is individual and private information of the firm i. Both the common and the individual cost components are random variables. In the paper, the capital letters Y and X denote random variables representing the common component and the vector of individual components; the small letters y and x denote realizations of the common component and the vector of individual components. The two random variables $(Y,\ X)$ are distributed on $[\underline{y},\overline{y}]\times[\underline{x},\overline{x}]^m$, $\underline{y}>0$, $\underline{x}>0$, according to the probability distribution function H,

$$\Pr(Y\leq y_0, X\leq x_0)=H(y_0,x_0).$$

Asymmetries between bidders: I assume that there are two types of bidders,[4] m_1 bidders are of type 1, $m_2=(m-m_1)$, bidders are of type 2. Thus the vector of independent components is given by $X=(X_{11},..,X_{1m_1},X_{2(m_1+1)},..,X_{2m})$.

Assumptions $(D_1)-(D_4)$[5] are maintained throughout the paper.

(D_1) Y and X_j's are mutually independent and distributed according to

[4]The model and all the results can easily be extended to the case of m types. I focus on the case of two types for the sake of expositional clarity.

Types are defined from the observable characteristics of bidders.

[5]Assumption (D_2) ensures equilibrium existence. Identification result relies on assumptions (D_1) and (D_3). In particular, assumption (D_3) is used to fix the scale of the distribution of individual cost component. (D_4) summarizes miscellaneous assumptions about auction environment.

$$H(y_0, x_{10}, .., x_{m0}) = H_Y(y_0) \prod_{j=1}^{j=m_1} H_{X_1}(x_{j0}) \prod_{j=m_1+1}^{j=m} H_{X_2}(x_{j0}),$$

where H_Y, H_{X_1} and H_{X_2} are marginal distribution functions of Y, X_{1j} and X_{2j} respectively.

(D_2) The probability density functions of the individual cost components, h_{X_1} and h_{X_2}, are continuously differentiable and bounded away from zero on $[\underline{x}, \overline{x}]$.

(D_3) $EX_{1j} = 1$ for all $j = 1, ..., m_1$.

(D_4) (a) The number of bidders is common knowledge;

(b) There is no binding reservation price;

(c) The cost of preparing the bid is proportional to the realization of y.

The auction environment can be described as a collection of auction games indexed by the different values of the common component. An auction game corresponding to the common component equal to y, $y \in [\underline{y}, \overline{y}]$, is analyzed below.

The cost realization of bidder i is equal to $x_i * y$, where x_i is the realization of the individual cost component. The information set of bidder i is given by $P_{yi} = \{x_i | x_i \in [\underline{x}, \overline{x}]\}$. A bidding strategy of bidder i is a real-valued function defined on $[\underline{x}, \overline{x}]$

$$\beta_{yi} : \; [\underline{x}, \overline{x}] \rightarrow [0, \infty].$$

I use small greek letter β with subscript yi to denote the strategy of bidder i as a function of the individual cost components and small Latin letter b to denote the value of this function at a particular realization x_i.

Expected profit. The profit realization of the bidder i, $\pi_{yi}(b_i, b_{-i}, x_i)$, equals $(b_i - x_i * y)$ if bidder i wins the project and zero if he loses. The symbol b_i denotes the bid submitted by bidder i, and the symbol b_{-i} denotes the vector of bids submitted by bidders other than i. At the time of bidding, bidder i knows y and x_i but not b_{-i}. The bidder who submits the lowest bid wins the project. The interim expected profit of bidder i is given by

$$E[\pi_{yi} | X = x_i, Y = y] = (b_i - x_i * y) * \Pr(b_i \leq b_j, \forall j \neq i | Y = y).$$

A Bayesian Nash Equilibrium is then characterized by a vector of functions $\beta_y = \{\beta_{y1}, ..., \beta_{ym}\}$ such that $b_{yi} = \beta_{yi}(x_i)$ maximizes $E[\pi_i | X = x_i, Y = y]$, when $b_j = \beta_{yj}(x_j)$, $j \neq i$, $j = 1, .., m$; for every $i = 1, .., m$ and for every realization of X_i.

LeBrun (1999) and others establish that, under assumptions $(D_1) - (D_2)$, a vector of equilibrium bidding strategies $\beta_y = \{\beta_{y1}, ..., \beta_{ym}\}$ exists. The strategies are strictly monotone and differentiable. Maskin and Riley (2000) show that under these assumptions there is a unique vector of equilibrium strategies, $\beta_y = \{\beta_{y1}, ..., \beta_{ym}\}$, which satisfy the following boundary condition: for all i $\beta_{yi}(\overline{x}) = \overline{x}$, and there exists $d_{yi} \in [\underline{x}, \overline{x}]$ such that $\beta_{yi}(\underline{x}) = d_{yi}$.

These results accordingly establish equilibrium existence and uniqueness in the game where the common cost component equals y.

Next, I characterize a simple property of the equilibrium bidding strategies.

Proposition 1

If $(\alpha_1(.), ..., \alpha_m(.))$ *is a vector of equilibrium bidding strategies in the game with* $y = 1$*, then the vector of equilibrium bidding strategies in the game with* y*,* $y \in [\underline{y}, \overline{y}]$*, are given by* $\beta_y = \{\beta_{y1}, ..., \beta_{ym}\}$*, such that* $\beta_{yi}(x_i) = y * \alpha_i(x_i)$*,* $i = 1, ..., m$*.*

The proposition shows that bid function is multiplicatively separable into a common and an individual bid component, where individual bid component is given by $\alpha_i(.)$. The proof of this proposition is based on the comparison of the two sets of first-order conditions and follows immediately from the assumption that costs are multiplicatively separable and that the common component is known to all bidders.

Next, I characterize the necessary first-order conditions for the set of equilibrium strategies when $y = 1$. Note, that $\alpha_i(.)$ denotes a strategy of bidder i as a function of individual cost component and a_i the value of this function for a particular realization of X_i. The equilibrium inverse bid function of the individual bid component for a type k bidder is denoted by ϕ_k. Since, the function $\alpha_k(.)$ is strictly monotone and differentiable, the function $\phi_k(.)$ is well-defined and differentiable.

The probability of winning in this game can be expressed as

$$\Pr(a_j \geq a_i,\ j \neq i) = \Pr(x_j \geq \phi_{k(j)}(a_i),\ j \neq i),$$

where $k(j)$ denotes the type of player j.

Further,

$$\Pr(a_j \geq a_i, \forall j \neq i) = [(1 - H_{X_{k(i)}}(\phi_{k(i)}(a_i)))]^{(m_{k(i)}-1)}[(1 - H_{X_{-k(i)}}(\phi_{-k(i)}(a_i)))]^{m_{-k(i)}},$$

where $k(i)$ denotes bidder $i's$ type and "$-k(i)$" denotes the complimentary type.

The objective function of the bidder i can be written as

$$(a_i - x_i) * [(1 - H_{X_{k(i)}}(\phi_{k(i)}(a_i)))]^{(m_{k(i)}-1)}[(1 - H_{X_{-k(i)}}(\phi_{-k(i)}(a_i)))]^{m_{-k(i)}}.$$

The necessary first-order conditions are, then, given by

$$\frac{1}{a-\phi_{k(i)}(a)} = (m_{k(i)}-1)\frac{h_{X_{k(i)}}(\phi_{k(i)}(a))\phi'_{k(i)}(a)}{1-H_{X_{k(i)}}(\phi_{k(i)}(a))} + m_{-k(i)}\frac{h_{X_{-k(i)}}(\phi_{-k(i)}(a))\phi'_{-k(i)}(a)}{1-H_{X_{-k(i)}}(\phi_{-k(i)}(a))}, \qquad (1)$$

where $\phi'_k(.)$ denotes the derivative of $\phi_k(.)$.

Equation (1) characterizes the equilibrium inverse individual bid function when $y = 1$. It describes a trade-off the bidder faces when choosing a bid: an increase in the markup over the cost may lead to a higher ex-post profit if bidder i wins, but it reduces the probability of winning. The bid a is chosen in such a way that the marginal effects of an infinitesimal change in a bid on the winner's profit and the probability of winning sum to zero.

The next section uses properties of the equilibrium bidding functions to show how the primitives of the First-Price auction model can be recovered from the submitted bids in the presence of unobserved auction heterogeneity.

4 Identification and Testable Implications of the Model

The first part of this section formulates an identification problem and provides conditions under which a First-Price auction model with unobserved heterogeneity is identified. The second part describes restrictions which model imposes on the data. The third part discusses possible extensions.

4.1 Identification

I assume that the econometrician has access to the bid data, generated by n independent draws from the joint distribution of (Y, X). The observable data are in the form $\{b_{ij}\}$, where i denotes identity of the bidder, $i = 1, .., m$; and j denotes project, $j = 1, ..., n$. If the data represent equilibrium outcomes of the model with unobserved heterogeneity, then

$$b_{ij} = \beta_{y_j k(i)}(x_{ij}) \qquad (2)$$

(i.e., b_{ij} is a value of bidder i's equilibrium bidding strategy corresponding to y_j evaluated at the point x_{ij}).

As was shown in the previous section, b_{ij} depends on the realizations of the common and individual components as well as on the joint distribution of the individual cost components. This section examines under what conditions on available data there exists a unique triple $\{\{x_{ij}\}, \{y_j\}, H_X\}$ that satisfies (2), i.e. under what conditions the model from a previous section is identified.

Guerre, Perrigne and Voung (2000) obtain an identification result by transforming the first-order conditions of optimal bids to express a bidder's cost as an explicit function of the submitted bid, the bid probability density function, and the bid distribution function. Under unobserved heterogeneity, the necessary first-order condition yields an expression for $x_{ij} \cdot y_j$ as a function of b_{ij} and the conditional bid probability density function and the conditional bid distribution function conditional on $Y = y_j$. The econometrician does not observe the realization of Y, and, consequently, does not know the conditional distribution of bids for $Y = y_j$. Hence, it is not possible to establish identification based on the above first-order conditions. Notice that the econometrician does observe the joint distributions of bids that share the same common cost component.

The idea of my approach is to focus on the joint distributions of bids instead of the marginal bid distributions in order to identify the model with unobserved heterogeneity. I use B_i to denote the random variable that describes the bid of bidder i with distribution function $G_{B_{k(i)}}(.)$ and the associated probability density function $g_{B_{k(i)}}(.)$; b_{ij} denotes the realization of this variable in the auction j. The econometrician observes the joint distribution function of $(B_{i_1}, .., B_{i_l})$ for all subsets $(i_1, ..., i_l)$ of $(1, ..., m)$.

Proposition 1 establishes that

$$b_{ij} = y_j * a_{ij},$$

where a_{ij} is a hypothetical bid that would have been submitted by bidder i if y were equal to one. I use A_i to denote the random variable with realizations equal to a_{ij}. The associated distribution function is denoted by $G_{A_{k(i)}}(.)$ with the probability density function $g_{A_{k(i)}}(.)$. Notice that the econometrician does not observe y_j, and, neither therefore a_{ij}. The distribution of A_i is latent.

My identification result is established in two steps. First, it is shown that the probability density functions of Y, A_i's can be uniquely determined from the joint distribution of two bids that share the same cost component under the following mild regularity condition :

(D_5) The characteristic functions of Y , A_{1i_1} and A_{2i_2}, $i_1 = 1, .., m_1$; $i_2 = m_1 + 1, .., m$, are non-vanishing.

Second, the condition (D_6) is used to establish identification of the probability density functions H_{X_1} and H_{X_2} from the distributions of the individual bid components, G_{A_1} and G_{A_2}.

Let the inverse individual bid function be given by

$$\phi_k(a) = a - \frac{(1 - G_{A_k}(a))(1 - G_{A_{-k}}(a))}{(m_k - 1)g_{A_k}(a)(1 - G_{A_{-k}}(a)) + m_{-k}g_{A_{-k}}(a)(1 - G_{A_k}(a))}, \quad k = 1, 2.$$

Condition (D_6), then states that

(D_6) The inverse individual bid function, $\phi_k(.)$, is a strictly increasing for every k.

The following theorem is the main result of this section. It formulates sufficient identification conditions for the model with unobserved heterogeneity.

Theorem 1

If conditions $(D_1)-(D_6)$ are satisfied, then probability density functions $h_Y(.)$, $h_{X_1}(.)$ and $h_{X_2}(.)$ are uniquely identified from the joint distribution of (B_{i_1}, B_{i_2}), where (i_1, i_2) is any pair such that $i_1 \in \{1, .., m_1\}$; $i_2 \in \{m_1 + 1, .., m\}$.

Theorem 1 states that the distribution functions of cost components, $H_{X_k}(.)$ and $H_Y(.)$ are identified. The proof of this theorem consists of two steps and is given in Part A of the Appendix. In the first step, a statistical result by Kotlarski[6] (1967) is applied to the log-transformed random variables B_{i_1} and B_{i_2} given by

$$\begin{aligned} \log(B_{i_1}) &= \log(Y) + \log(A_{i_1}), \\ \log(B_{i_2}) &= \log(Y) + \log(A_{i_2}). \end{aligned}$$

Kotlarski's result is based on the fact that the characteristic function of the sum of two independent random variables is equal to the product of characteristic functions of these variables. This property allows us to find the characteristic functions of $\log(Y)$, $\log(A_{i_1})$ and $\log(A_{i_2})$ from the joint characteristic function of $(\log(B_{i_1}, \log(B_{i_2}))$. It leads to the following three equations:

$$\begin{aligned} \Phi_{\log(Y)}(t) &= \exp(\int_0^t \frac{\Psi_1(0, u_2)}{\Psi(0, u_2)} du_2), \qquad (3) \\ \Phi_{\log(A_1)}(t) &= \frac{\Psi(t, 0)}{\Phi_{\log Y}(t)}, \\ \Phi_{\log(A_2)}(t) &= \frac{\Psi(0, t)}{\Phi_{\log Y}(t)}, \end{aligned}$$

where $\Psi(.,.)$ and $\Psi_1(.,.)$ denote the joint characteristic function of $(\log(B_{i_1}), \log(B_{i_2}))$ and the partial derivative of this characteristic function with respect to the first component respectively. Since there is a one-to-one correspondence between the set of characteristic functions and the set of probability density functions, the probability density functions of Y, A_{i_1}, A_{i_2} can be uniquely deduced from the characteristic functions of $\log(Y)$, $\log(A_{i_1})$, and $\log(A_{i_2})$ as $\log(.)$ is a strictly increasing function; $\alpha_k(.)$, $k = 1, 2$, are increasing functions of x; $[\underline{x}, \overline{x}] \subset (0, \infty)$, $[\underline{y}, \overline{y}] \subset (0, \infty)$. Notice that the marginal distribution of a single bid per auction may not allow us to identify the distribution functions of Y, A_{i_1}, A_{i_2} because there is no unique decomposition of the sum (or

[6] See Prakasa Rao (1992).

product) into its components.

The second step in the proof establishes that the distribution of the individual cost component is identified with (possibly) asymmetric bidders and independent private values. It is similar to the argument given in Laffont and Vuong (1996).

A related question concerns identification of specific realizations x_{ij} and y_j corresponding to a particular bid b_{ij}. In this case, the answer is negative: x_{ij} and y_j cannot be separately identified. The reason is that for every value of y from the support of the distribution $H_Y(.)$, we can find values $\{x_{ij}\}$, $i = 1, .., m$, such that a vector $(x_{1j}, .., x_{mj}, y)$ together with the distribution functions $H_{X_k}(.)$, $k = 1, 2$, rationalizes the vector of bids $\{b_{ij}\}$, $i = 1, .., m$. More details are provided in Part A of the Appendix after the proof of Theorem 1.

4.2 Testable Implications

The previous section derived conditions under which primitives of a model with unobserved heterogeneity can be uniquely recovered from bids generated within the framework of this model. We now describe properties of the data that arise if the underlying data generating process satisfies a particular assumption of the model with unobserved heterogeneity. These properties allow us to test validity of the model's assumptions.

The first set of conditions describes a set of joint restrictions imposed on the data by all the assumptions of the model with unobserved heterogeneity.

$(W_{\mathbf{1}})$ For every pair (i_l, i_p), $i_l = 1, ..., m_1$; $i_p = m_1 + 1, .., m$, the functions $\Phi_{\log(Y)}(.), \Phi_{\log(A_{i_l})}).), \Phi_{\log(A_{i_p})}(.)$ given by (2) represent characteristic functions of real-valued variables.

(W_2) The characteristic functions $\Phi_{\log(Y)}(.), \Phi_{\log(A_{i_l})}).)$ and $\Phi_{\log(A_{i_p})}).)$ do not depend on the pair of (i_l, i_p), $i_l = 1, ..., m_1$; $i_p = m_1 + 1, .., m$, which was used to derive them.

(W_3) The inverse bid functions

$$\phi_k(a) = a - \frac{(1 - G_{A_k}(a))(1 - G_{A_{-k}}(a))}{(m_k - 1)g_{A_k}(a)(1 - G_{A_{-k}}(a)) + m_{-k}g_{A_{-k}}(a)(1 - G_{A_k}(a))}, \ k = 1, 2,$$

are strictly increasing in a.

Proposition 2

If available data satisfy conditions $(W_1) - (W_3)$, then there exists a model with unobserved heterogeneity that could have generated the data.

The first condition guarantees that two independent random variables Y and A_i exist with the property $B_i = Y * A_i$. The third condition ensures that $A_i's$ are consistent with the equilibrium behavior under independent private values assumption. The first and the second assumptions guarantee that bidders within each of the types are identical.

The next set of conditions allows testing of the independence assumption while relying on the functional form assumption, $c_{ij} = x_{ij} * y_j$. Notice, that Kotlarski's result can be applied to the variables $\log(\frac{B_{i_1}}{B_{i_3}})$ and $\log(\frac{B_{i_2}}{B_{i_3}})$ since $\log(\frac{B_{i_1}}{B_{i_3}}) = \log(A_{i_1}) - \log(A_{i_3})$ and $\log(\frac{B_{i_2}}{B_{i_3}}) = \log(A_{i_2}) - \log(A_{i_3})$. If the individual cost components X_{i_1}, X_{i_2} and X_{i_3} are independently distributed, then so are $\log(A_{i_1})$, $\log(A_{i_2})$ and $\log(A_{i_3})$. The characteristic functions of these variables can be computed using the joint characteristic function of $(\log(\frac{B_{i_1}}{B_{i_3}}), \log(\frac{B_{i_2}}{B_{i_3}}))$, which I denote by $\Theta(.,.)$, according to a formula similar to equation (3).[7] Specifically,

$$\begin{aligned} \Lambda_{\log(A_{i_3})}(t) &= \exp(\int_0^t \frac{\Theta_1(0,u_2)}{\Theta(0,u_2)} du_2), \\ \Lambda_{\log(A_{i_1})}(-t) &= \frac{\Theta(t,0)}{\Lambda_{\log(A_{i_3})}(t)}. \end{aligned} \tag{4}$$

So far, I have only relied on the functional form and the independence of individual cost components assumptions to obtain $\Lambda_{\log(A_{i_1})}(.)$. The assumption of independence of Y and X the implies that $\Lambda_{\log(A_{i_3})}(.)$ and $\Lambda_{\log(A_{i_1})}(.)$ have to coincide with the functions given by (3). These observations provide a basis for conditions (W_4) and (W_5) that play a central role in the next result.

(W_4) For any triple (i_1, i_2, i_3) such that $i_1 = 1, .., m_1$ and $i_3 = m_1 + 1, .., m$, a function $f_1(.)$ given by

$$f_1(t) = \exp(\int_0^t \frac{\Theta_1(0,u_2)}{\Theta(0,u_2)} du_2)$$

is a characteristic function.

(W_5) For any triple (i_1, i_2, i_3) such that $i_1 = 1, .., m_1$ and $i_3 = m_1 + 1, .., m$,

$$\begin{aligned} \Phi_{\log(A_{i_3})}(t) &= \Lambda_{\log(A_{i_3})}(t), \\ \Phi_{\log(A_{i_1})}(t) &= \Lambda_{\log(A_{i_1})}(t) \end{aligned}$$

for every $t \in [-\infty, \infty]$. Here $\Phi_{\log(A_{i_3})}(t)$ and $\Phi_{\log(A_{i_1})}(t)$ denote the characteristic functions of the log of the individual bid components defined earlier in the identification section.

[7] The symbol $\Theta_1(.,.)$ denotes the partial derivative of $\Theta(.,.)$ with respect to the first argument.

Condition (W_6) uses the functional form of the cost function to derive implications of the independence of individual cost components.

(W_6) For any quadruple $(i_1, i_2, i_3, i_4) \subset \{1, ..., m\}$, $\frac{B_{i_1}}{B_{i_2}}$ and $\frac{B_{i_3}}{B_{i_4}}$ are independently distributed.

Proposition 3 describes implications of the independence assumption.

Proposition 3

Let bidder i's cost for the project j be given by $c_{ij} = x_{ij} * y_j$.
If the individual cost components are independent, then (W_4) and (W_6) have to be satisfied.
Further, if Y is independent of X, then W_4 and W_5 hold.

The proof of proposition 3 is given in Part A of the Appendix. Results given by Proposition 2 and 3 allow us to test the validity of model's assumptions in a particular data set.

4.3 Extensions and limitations

My model with unobserved heterogeneity requires the assumption that bidder $i's$ cost of completing the project equals the product of the common and the individual cost components. This functional form emerges when the cost distribution for a particular project is scaled by a project-specific common variable, in which case mean and variance vary with the common component in a coordinated way.

A more general model may allow for the common component to have distinct effects on the mean and variance of the cost distribution function. Such a model can be constructed using a two-dimensional project heterogeneity. Bidder $i's$ cost of the project is, then, equal to

$$c_{ij} = y_{1j} + y_{2j} * x_{ij},$$

where (y_{1j}, y_{2j}) is a realization of a two-dimensional cost component that is common knowledge among all bidders; x_{ij} is a realization of an individual component, which is private information of firm i. This specification has the following interpretation: the average cost of the project j equals y_{1j}, and the individual cost deviations have auction specific scale. It can be shown that the described model is identified under conditions similar to those in Theorem 1. The exact conditions and the proof are given in Part A of the Appendix.

My identification result requires the independence assumption for individual cost components. The model with unobserved heterogeneity includes auction model with independent private values as a particular case which arises when common component is constant across auctions. The nested structure enables a test of the null of unobserved heterogeneity against the alternative of independent private values. Further, we may be interested in testing the null of unobserved heterogeneity against the alternative of affiliated private values. While I do not have a formal proof, my conjecture is that the model with unobserved heterogeneity and the affiliated private values model cannot be distinguished based on the bid data.

5 Estimation

This section describes the estimation method, derives properties of the estimators, and discusses practical issues related to the estimation procedure.

5.1 Estimation Method

The econometrician has data for n auctions. For each auction j, $(m_j, \{b_{ij}\}_{i=1}^{i=m_j}, z_j)$ are observed, where m_j is the number of bidders in the auction j, with m_{j1} bidders of type 1 and m_{j2} bidders of type 2; $\{b_{ij}\}_{i=1}^{i=m_j}$ is a vector of bids submitted in the auction j; z_j is a vector of auction characteristics. The estimation procedure is described for the case of discrete covariates. It can be extended to the case of continuous z_j.

The estimates are obtained conditional on the number of bidders, $m_j = m_0$, $m_{1j} = m_{01}$ and $z_j = z_0$. Let n_0 denote the number of auctions that satisfy these restrictions. The estimation procedure closely follows the identification argument described in the proof of Theorem 1. It consists of two steps. First, the joint characteristic function of two log transformed bids is estimated for every (t_1, t_2) as a sample average of the $\exp(it_1 \cdot B_{lj} + it_2 \cdot B_{pj}))$, where the average is taken across auctions with $m_j = m_0$, $m_{1j} = m_{01}$ and $z_j = z_0$. Then the joint characteristic function is used to infer the characteristic functions of the logs of the common and the individual bid components according to the formulas given by (3). The inversion formula is used to recover the probability density functions for the logs of the common and the individual bid components from the characteristic functions. Finally, the probability density functions of logs are used to recover the probability density functions of the common and the individual bid components.

In the second step, the probability density functions of the individual bid components are used to obtain estimates of the probability density function of the individual cost component. For that, a sample of pseudo-bids is drawn form the probability density function of the relevant individual bid component. This sample is then used to obtain the sample of pseudo-costs with the help of the corresponding inverse bid function. Finally, the sample of pseudo-costs is used to non-parametrically estimate the probability density function of the individual cost component.

To estimate the probability density function of the total cost of the bidder i at a point c, I compute an integral of the function $h_{X_i}(\frac{c}{y}) * h_Y(y)$ with respect to y over the interval $[\underline{y}, \overline{y}]$. To evaluate this integral, I perform Monte-Carlo integration with respect to $h_Y(.)$.[8] The value of an average inverse bid function at a point b is estimated as the mean of the value of the individual bid function at a point $\frac{b}{y}$ with respect to the distribution of y. Again, Monte-Carlo integration methods are used to compute the mean.

The rest of this subsection describes the details of the estimation procedure beginning with Step 1:

Step 1.

1. The log transformation of bid data is performed to obtain $LB_{i_l} = \log(B_{i_l})$ and $LB_{i_p} = \log(B_{i_p})$, where $i_l = 1,..,m_{01}$ and $i_p = m_{01}+1,..,m_0$.

2. The joint characteristic function of an arbitrary pair (LB_{i_l}, LB_{i_p}) is estimated by

$$\widehat{\Psi}(t_1,t_2) = \frac{1}{m_{01}m_{02}} \sum_{1\leq l\leq m_{01}, m_{01}+1\leq p\leq m_0} \frac{1}{n_0}\sum_{j=1}^{n_0} \exp(it_1 \cdot B_{i_l j} + it_2 \cdot B_{i_p j})$$

and the derivative of $\Psi(.,.)$ with respect to the first argument, $\Psi_1(.,.)$, by

$$\widehat{\Psi_1}(t_1,t_2) = \frac{1}{m_{01}m_{02}} \sum_{1\leq l\leq m_{01},\ m_{01}+1\leq p\leq m_0} \frac{1}{n_0}\sum_{j=1}^{n_0} iB_{i_l j}\exp(it_1 \cdot B_{i_l j} + it_2 \cdot B_{l_p j})$$

I average over all possible pairs to enhance efficiency.

3. The characteristic functions of the log of individual bid components LA_k, $k = 1,2$, and the log of the common cost component LY are estimated as

$$\widehat{\Phi}_{LY}(t) = \exp(\int_0^t \frac{\widehat{\Psi}_1(0,u_2)}{\widehat{\Psi}(0,u_2)}du_2),$$

$$\widehat{\Phi}_{LA_1}(t) = \frac{\widehat{\Psi}(t,0)}{\widehat{\Phi}_{\log Y}(t)},$$

$$\widehat{\Phi}_{LA_2}(t) = \frac{\widehat{\Psi}(0,t)}{\widehat{\Phi}_{\log Y}(t)}.$$

4. The inversion formula is used to estimate densities g_{LA_k}, $k = 1,2$, and g_{LY}.

$$\widehat{g}_{LA_k}(u_1) = \frac{1}{2\pi}\int_{-T}^{T} \exp(-itu_1)\widehat{\Phi}_{LA_k}(t)dt,$$

$$\widehat{h}_{LY}(u_2) = \frac{1}{2\pi}\int_{-T}^{T} \exp(-itu_2)\widehat{\Phi}_{LY}(t)dt$$

[8] Judd (2000) provides a detailed explanation of a Monte-Carlo integration method.

for $u_1 \in [\log(\underline{a_k}), \log(\overline{a_k})]$, and $u_2 \in [\log(\underline{y}), \log(\overline{y})]$, where T is a smoothing parameter.

5. The densities of A_k and Y are obtained as

$$\begin{aligned} \widetilde{g}_{A_k}(a) &= \frac{\widehat{g}_{LA_k}(\log(a))}{a}, \\ \widetilde{h}_Y(y) &= \frac{\widehat{h}_{LY}(\log(y))}{y}, \end{aligned}$$

for $a \in [\underline{a_k}, \overline{a_k}]$, and $y \in [\underline{y}, \overline{y}]$.

Step 2

1. The estimate of the individual bid component density, $\widetilde{g}_{A_k}(.)$, $k = 1, 2$, is used to generate a sample of pseudo-bids $\{\widetilde{a}_{kj}\}$, $j = 1, ..., L$.

2. The sample of pseudo bids is used to generate a sample of pseudo-costs as

$$\begin{aligned} \widetilde{x}_{1j} &= \widetilde{a}_{1j} + \frac{(1 - \widetilde{G}_{A_1}(a_{1j})) \cdot (1 - \widetilde{G}_{A_2}(a_{1j}))}{(m_1 - 1) \cdot \widetilde{g}_{A_1}(a_{1j}) \cdot (1 - \widetilde{G}_{A_2}(a_{1j})) + m_2 \cdot \widetilde{g}_{A_2}(a_{1j}) \cdot (1 - \widetilde{G}_{A_1}(a_{1j}))}, \\ \widetilde{x}_{2j} &= \widetilde{a}_{2j} + \frac{(1 - \widetilde{G}_{A_1}(a_{2j})) \cdot (1 - \widetilde{G}_{A_2}(a_{2j}))}{m_1 \cdot \widetilde{g}_{A_1}(a_{2j}) \cdot (1 - \widetilde{G}_{A_2}(a_{2j})) + (m_2 - 1) \cdot \widetilde{g}_{A_2}(a_{2j}) \cdot (1 - \widetilde{G}_{A_1}(a_{2j}))}, \end{aligned}$$

where

$$\widetilde{G}_{A_k}(a) = \int_{\widehat{\underline{a}}_k^1}^{a} \widetilde{g}_{A_k}(z)dz$$

and $\widehat{\underline{a}}_k^1$ is an estimate of the lower bound of the support of $g_{A_k}(.)$ (see part A of the Appendix for the discussion of the support estimation).

3. The density of the individual cost component is non-parametrically estimated from the sample of pseudo-cost

$$\widetilde{h}(x) = \frac{1}{L\delta_{h_k}} \sum_{j=1}^{n} K_h(\frac{x - \widetilde{x}_{kj}}{\delta_k}),$$

where $K_h(.)$ is a kernel function, and δ_{h_k} is the bandwidth.

3a. The estimation procedure described in Step 1 leads to a zero-mean distribution of $\log(A_1)$, which does not necessarily correspond to the random variable X_1 such that $EX_1 = 1$. To arrive at the final estimates of the distributions in question we have to perform an adjustment. Let e denote the mean of the estimated distribution of random variable X_1. Then $\widehat{h}_{X_k}(x) = \frac{\widetilde{h}_{X_k}(ex)}{e}$, $\widehat{h}_Y(y) = e\widetilde{h}_Y(\frac{y}{e})$ are the final estimates of the individual and common cost component probability density functions.

4. I have also constructed an estimate of the cost density function

$$\widehat{h}_{C_k}(c) = \int_{\underline{y}}^{\overline{y}} \widehat{h}_{X_k}(\frac{c}{y})\widehat{h}_Y(y)dy,$$

for $c \in [\underline{x} \cdot \underline{y}, \overline{x} \cdot \overline{y}]$.

5. An average inverse bid function was estimated as

$$\widehat{\vartheta}_k(b) = \int_{\underline{y}}^{\overline{y}} y \cdot \widehat{\phi}_k(\frac{b}{y})\widehat{h}_Y(y)dy,$$

where $\widehat{\phi}_k(.)$ is an estimate of the individual inverse bid function given by

$$\widehat{\phi}_1(a) = a + \frac{(1-\widetilde{G}_{A_1}(a)) \cdot (1-\widetilde{G}_{A_2}(a))}{(m_1-1) \cdot \widetilde{g}_{A_1}(a) \cdot (1-\widetilde{G}_{A_2}(a)) + m_2 \cdot \widetilde{g}_{A_2}(a) \cdot (1-\widetilde{G}_{A_1}(a))},$$

$$\widehat{\phi}_2(a) = a + \frac{(1-\widetilde{G}_{A_1}(a)) \cdot (1-\widetilde{G}_{A_2}(a))}{m_1 \cdot \widetilde{g}_{A_1}(a) \cdot (1-\widetilde{G}_{A_2}(a)) + (m_2-1)\widetilde{g}_{A_2}(a) \cdot (1-\widetilde{G}_{A_1}(a))}.$$

Both integrals were computed using Monte-Carlo integration with respect to $\widehat{h}_Y(.)$.

5.2 Properties of the estimator

This subsection shows that the estimation procedure yields uniformly consistent estimators of the relevant distributions. Conditions are stated under which the first step estimation procedure results in uniformly consistent estimators. It is further shown that if the first step estimators are uniformly consistent then the second step procedure also leads to uniformly consistent estimators.

Li and Vuong (1998) establish the uniform consistency of the first-stage estimator when the densities of interest are either ordinary-smooth or super-smooth with respect to the tail behavior of their characteristic functions. Following Fan (1991),

Definition 1 The distribution of random variable Z is *ordinary-smooth of order* $\varkappa$ if its characteristic function $\phi_z(t)$ satisfies

$$d_0|t|^{-\varkappa} \leq |\phi_z(t)| \leq d_1|t|^{-\varkappa}$$

as t$\rightarrow \infty$ for some positive constants d_0,d_1,$\varkappa$.

The distribution of the random variable Z is *super-smooth of order* $\varkappa$ if $\phi_z(t)$ satisfies

$$d_0|t|^{\varkappa_0}\exp(\frac{-|t|^{\varkappa}}{\lambda}) \leq |\phi_z(t)| \leq d_1|t|^{\varkappa_1}\exp(\frac{-|t|^{\varkappa}}{\lambda})$$

as t$\rightarrow\infty$ for some positive constants d_0,d_1,$\varkappa$, λ and constants $\varkappa_0$ and $\varkappa_1$.

Following Li and Vuong (1998) I make the following assumptions,

(D_8) The characteristic functions $\phi_{LY}(.)$ and $\phi_{LA_k}(.)$ are ordinary smooth with $\varkappa > 1$ or super-smooth.

(Note that the characteristic functions $\phi_{LY}(.)$ and $\phi_{LA_k}(.)$ are necessarily both integrable.)

(D_9) The supports of $h_Y(.)$ and $h_{A_k}(.)$ are bounded intervals of R.

A technique similar to the one proposed by Guerre, Perrigne and Vuong (2000) can be used to establish the uniform consistency of the second-stage estimator. Their proof relies on the fact that uniformly consistent estimators of $G_{A_k}(.)$ and $g_{A_t}(.)$ can be obtained from the data when the number of auctions diverges to infinity. In our case, this fact follows from the uniform consistency of the first-step estimator established by Li and Vuong (1998).[9]

Proposition 4

If conditions (D_1)-(D_9) *are satisfied then* $\widehat{h}_Y(.)$ *and* $\widehat{h}_{X_k}(.)$ *are uniformly consistent estimators of* $h_Y(.)$ *and* $h_{X_k}(.)$, $k = 1, 2$, *respectively.*

The proof of Proposition 4 is presented in Part A of the Appendix.

5.3 Practical issues

This subsection addresses additional estimation issues. In the application I describe estimates of the unobserved heterogeneity model for the case of symmetric bidders. I focus on the symmetric case here as well.

Several important comments must be made about the first step. First, to reduce the error in the characteristic function estimation, I scale bids to fit into the interval $[0, 2\pi]$. Second, as noted by Diggle and Hall (1993), the estimators for $\widehat{h}_{LY}(.)$ and $\widehat{h}_{LA}(.)$, which are obtained by truncated inverse Fourier transformation, may have fluctuating tails.[10] This feature can be alleviated by

[9] See discussion above.

[10] Li,Perigne and Vuong (2001) encountered this problem as well and dealt with it in a similar way.

adding a damping factor to the integrals in $\widehat{h}_{LY}(.)$ and $\widehat{g}_{LA}(.)$. Following Diggle and Hall (1993), I introduce a damping factor defined as

$$d_T(t) = \left\{ \begin{array}{c} 1 - \frac{|t|}{T},\ if\ |t| \leq T \\ 0,\ otherwise \end{array} \right\}.$$

Thus, the estimators are generalized to

$$\begin{aligned} \widehat{g}_{LA}(a) &= \frac{1}{2\pi} \int_{-T}^{T} d_T(t) \exp(-ita) \widehat{\Phi}_{LA}(t) dt, \\ \widehat{h}_{LY}(y) &= \frac{1}{2\pi} \int_{-T}^{T} d_T(t) \exp(-ity) \widehat{\Phi}_{LY}(t) dt. \end{aligned}$$

Third, the smoothing parameter T should be chosen to diverge slowly as $n \to \infty$, so as to insure uniform consistency of the estimators. However, the actual choice of T in finite samples has not yet been addressed in the literature. I follow Li, Perrigne and Vuong (2001) in choosing T through a data-driven criterion. In particular, I use the bid data to obtain estimates of the means and variances for distributions[11] of LY and LA, $\widehat{\mu}_{LY}$, $\widehat{\mu}_{LA} = 0$, $\widehat{\sigma}_{LY}$, $\widehat{\sigma}_{LA}$. These estimates are then used to choose a value of T. Specifically, I try different values of T and obtain estimates of $h_{LY}(.)$ and $h_{LA}(.)$. From each estimated density I compute the means and variances $\widetilde{\mu}_{LY}$, $\widetilde{\mu}_{LA}$, $\widetilde{\sigma}_{LY}$, $\widetilde{\sigma}_{LA}$ respectively. This gives goodness-of-fit criterion $|\widehat{\mu}_{LY} - \widetilde{\mu}_{LY}| + |\widehat{\sigma}_{LY} - \widetilde{\sigma}_{LY}|$ for LY, and similarly for LA. The value of T that I choose minimizes the sum of these errors in percentage of $\widehat{\sigma}_{LY}$ and $\widehat{\sigma}_{LA}$. In the estimation the optimal T equals 50.

The second step in the estimation involves taking random draws from the estimated density. I use a rejection method.[12] In this method random pairs (z_j, a_j) are drawn from the uniform distribution on $[0, r] \times [\underline{a}, \overline{a}]$, where r is the maximum value that $\widehat{h}_A(.)$ attains on the support of the distribution of A. Then, a_j is added to the sample of pseudo-bids if $z_j \leq \widehat{h}_A(a_j)$. The resulting sample of pseudo-bids is distributed according to $\widehat{h}_A(.)$.

The second step of the estimation involves non-parametric estimation of the density and distribution function . In the density estimation a tri-weight kernel is used, because it satisfies conditions of compact support and continuous differentiability on the support including the boundaries.[13] The tri-weight kernel is defined as

$$K(u) = \frac{35}{32}(1 - u^2)^3 1(|u| \leq 1).$$

[11] The estimates for the first two moments of the distributions of LY, LA_1 and LA_2 can be obtained as follows: $\widehat{\mu}_{LY} = \frac{\sum \log(b_i)}{n*m}$, $\widehat{\mu}_{LA} = 0$, $\widehat{\sigma}_{LA} = \frac{\sum (\log(b_{i_1}) - \log(b_{i_2}))^2}{2*n*m}$, $\widehat{\sigma}_{LY} = \frac{\sum (\log(b_i))^2}{2*n*m} - (\widehat{\mu}_{LY})^2 - \widehat{\sigma}_{LA}$.

[12] The rejection method was proposed by Newmann (1951). We need to know the support of the distribution in question to apply this method. A procedure for the supports estimation is described in the Part A of the Appendix.

[13] Conditions given in Li, Perrigne and Vuong (2001) ensure uniform consistency of the second stage estimator.

I follow Guerre, Perrigne and Vuong(2000) in my choice of bandwidth, $\delta_g = d_g(L)^{-\frac{1}{6}}$, where d_g is computed according to a "rule of thumb." Specifically, I use $d_g = 2.978 \times 1.06\widehat{\sigma}_a$, where $\widehat{\sigma}_a$ is the standard deviation of the logarithm of $(1 + bids)$, and 2.978 follows from the use of tri-weight kernel.[14]

Confidence intervals for the estimates are obtained through a bootstrap procedure.

5.4 Alternative Model Specifications

I also estimate two alternative models of competitive bidding. First, I use the method described in Guerre, Perrigne and Vuong (2000) to estimate the model of symmetric and independent values. The estimation procedure involves construction of a sample of pseudo-costs from the observable bids and non-parametric estimation of the cost distribution from the constructed sample. In the estimation, I again use a tri-weight kernel and bandwidth given by the formula $\delta_g = d_g(L)^{-\frac{1}{6}}$, with d_g chosen as explained above.

Second, I estimate the affiliated private values model with symmetric bidders. The estimation procedure is described in Li, Perrigne and Vuong (2001). Again, I use tri-weght kernels to estimate the distribution function $G_{b_1B_1}(u_1, u_2)$ and the probability density function $g_{b_1B_1}(u_1, u_2)$ by

$$\begin{aligned} G_{b_1B_1}(u_1, u_2) &= \frac{1}{nm\delta_g} \sum_{j=1}^{n} \sum_{i=1}^{m} K_G(\frac{u_2 - B_{ij}}{\delta_G}) I(b_{ij} \leq u_1), \\ g_{b_1B_1}(u_1, u_2) &= \frac{1}{nm\delta_g} \sum_{j=1}^{n} \sum_{i=1}^{m} K_G(\frac{u_2 - B_{ij}}{\delta_g}) K_G(\frac{u_1 - b_{ij}}{\delta_g}), \end{aligned}$$

where b_{ij} is the bid submitted by bidder i in the auction j, B_{ij} is minimum of the bids submitted by competitors of bidder i in auction j. Following Li, Perrigne and Vuong (2001) in the choice of bandwidth yields, $\delta_g = d_g(L)^{-\frac{1}{6}}$ and $\delta_G = d_G(L)^{-\frac{1}{5}}$, where $d_g = 2.978 \times 1.06\widehat{\sigma}_b$, $d_G = 2.978 \times 1.06\widehat{\sigma}_b$ and $\widehat{\sigma}_b$ is the standard deviation of the logarithm of $(1 + bids)$.

Guerre, Perrigne and Vuong (2000) and Li, Perrigne and Vuong(2001) establish that the described estimators are uniformly consistent under the null hypothesis of independent private values or affiliated private values, respectively.

6 Monte-Carlo Study

This section describes results of a Monte-Carlo study. The study has two purposes: first, to illustrate the behavior of the proposed estimator in small samples; second, to determine the nature, magnitude and direction of the bias that arises, when a misspecified model is estimated

[14]See Hardle, 1991.

from data generated by the model with unobserved heterogeneity.

Design: The simulated data sets are generated as follows. The individual cost component is distributed according to the power distribution with exponent equal to three and mean fixed at zero. The support of the distribution is $[-0.75, 0.25]$. In the first part of the study the common component is chosen to be distributed according to the power distribution with the exponent equal to three and mean fixed at 10.5. The support of this distribution is $[9.75, 10.75]$. In the second part of the study the common component is chosen to be uniformly distributed with a mean equal to 10.5 and a spread equal to s. To study the effect of an increase in the correlation of bids on the performance of misspecified models, I use several different values of the spread, in particular $s = 0.5, 1, 5, 10$. The cost of bidder i is assumed to equal the sum of the common and the individual cost components, $c_i = y + x_i$. To create a typical data set describing n procurement auctions with k bidders, $k*n$ independent draws from the power distribution ($\alpha = 3$) are combined with n draws from the corresponding uniform distribution such that $\{c_{ij}, c_{ij} = y_j + x_{ij}, i = 1, ., k_j = 1, .., n\}$ is a matrix of simulated costs. The matrix of associated bids is calculated according to the equilibrium bid function:

$$\{b_{ij},\ b_{ij} = y_j + x_{ij} + \frac{(1 - 4 * x_{ij}) - (1 - (x_{ij} + 3/4)^4)}{4 * (1 - (x_{ij} + 3/4)^3},\ i = 1, .., k,\ j = 1, .., n\}.$$

The values of n and k are set to equal 500 and 2 respectively. I replicate the described experiment 500 times and illustrate the resulting distributions of the estimators.

Part B of the Appendix presents the results of the study.

Model with unobserved heterogeneity

The estimation results are presented in Figures 1b-3b. The estimates suggests that in general proposed estimator behaves well in small samples. Figures 1b and 2b show the true densities of the common and the individual cost components as well as 5% and 95% pointwise quantiles of the estimators. The notable feature of the estimator is its bias at the boundary of the support. The boundary effect is similar to the effect described by Guerre, Perrigne and Vuong (2000). Due to the shape of the density function chosen to represent the individual component this effect is more visible at the upper end of the support. The true densities lie in the upper part of the interquantile range. This suggests that the proposed estimator may somewhat underestimate the value of the density function at a point. Figure 3b shows the true individual bid function and the 5% and 95% pointwise quantiles of the estimator distribution. The estimator of the individual bid function demonstrates features similar to those of the density estimators described above. It is noticeably biased near the lower end of the support. The true value of the bid function lies in the lower part of the interquantile range, which indicates that the proposed estimator may somewhat overestimate the value of the individual inverse bid function at a point.

Analysis of misspecified models

Figures 4b and 5b depict the true bid functions, and median or 5% - 95% pointwise quantiles for the bid function estimators based on the assumptions of independent and affiliated private values models for different values of the spread and of the common component distribution. Figures 6b and 7b present the true probability density functions for different spread values and median or 5%-95% pointwise quantiles of the corresponding estimators. The important thing to notice is that the difference between the estimated and true bid functions as well as the probability density functions increases, as the spread of the common component distribution, and thus the correlation between the bidders' costs (and bids) increases. In particular, this effect becomes significant, when the correlation coefficient is around or exceeds 0.3, which corresponds to the case of the spread equaling 1 or more.

The bidders' costs estimated through the IPV or APV procedures are significantly lower than the true ones. The bid functions estimated by the IPV and APV procedures are flatter than the true inverse bid function for most of the support. In particular, the IPV estimator gets increasingly flatter at the lower end of the support. Intuitively, the presence of the known common component leads to a bid distribution with very thin tails. Under the assumption of independent private values, such a distribution reflects extreme bid shading at the lower end of the support, or an extremely flat bid function.

Table 1b presents the estimated average mark-ups over the bidders' costs. To arrive at the average mark-up, the difference between the bid and the estimated cost associated with this bid is constructed, and the resulting mark-up function is integrated with respect to the bid distribution. Thus the constructed average mark-ups put all the estimators into a similar framework and provide an appropriate benchmark for comparison between different models. As may be expected after the earlier findings on the bid functions estimators both, the IPV and APV estimators, lead to higher expected mark-ups. The difference between the true average markup and APV or IPV estimated mark-ups increases, as the variance of the common component increases.

The probability density functions estimated under the independent and affiliated values assumptions allocate more mass to the left tail of the distribution than the true probability density function does. In particular, the probability density function estimated under the assumption of independent private values is substantially more spread out than the true probability density function. This happens because the IPV procedure ignores the presence of the known common component and misinterprets the concentration of bids as high "shading," which in turn leads to a long, thin tail. This tendency persists as the spread of the common component distribution increases.

Table 2b presents estimates of the mean, the variance, and the skewness of the costs distribution across estimation procedures and for different values of the spread. The results suggest that the cost distributions estimated under both the independent and affiliated private values assumptions tend to have lower means and higher skewness as compared to the true cost distribution. The variance of the APV estimated cost distribution is about equal to and the variance of the IPV estimated distribution is higher than the variance of the true distribution.

So far, I have described identification conditions, proposed an estimation procedure, and discussed small sample properties of the estimator using a Monte-Carlo study.

7 Michigan Highway Procurement Auctions

This section describes characteristics of the Michigan highway procurement auctions. Section 7.1 and 7.2 present the data and report descriptive evidence on auction specific variation in the bids distribution. Section 7.3 describes estimation results for the model with unobserved heterogeneity and compares them to the estimates obtained under the assumption of independent and affiliated private values. The estimates suggest that unobserved heterogeneity may account for a large part in bid variation. If unobserved heterogeneity is present, estimators obtained under alternative assumptions may substantially exaggerate bidders' markups and misrepresent the shape of the cost distribution.

7.1 Description of the Data

I have data for highway procurement auctions held by the Michigan Department of Transportation between February 1997 and August 2002. I select highway maintenance projects with bituminous resurfacing as the main task. These projects usually involve additional tasks such as marking, sign installation, landscaping, and so on. The data set consists of a total of 1,260 projects. My information includes the letting date, the completion time, the location, the tasks involved, the identity of all the bidders, their bids, an engineer's estimate, and a list of planholders for all projects in my data set.

Letting process. The Department of Transportation (DoT) advertises projects 4 to 10 weeks prior to the letting date. Advertisement usually includes a short description of the project including the location and completion time. Companies interested in the project can obtain a detailed description from the DoT. Bids have to be submitted 48 hours prior to the letting date. The company that specifies the lowest total bid wins the project. To bid for the project, companies have to be pre-qualified for at least 40% of the tasks involved. They are allowed to subcontract the rest.

Estimated cost. The DoT constructs a cost estimate for every project. This estimate is based on the engineer's assessment of the work required to perform each task and prices derived from the winning bids for the similar projects let in the past. The costs are then adjusted through a price deflator.

Federal law requires that the winning bid should be lower than 110% of the engineer's estimate. If a state decides to accept a bid that is higher than this threshold, it has to justify this action in writing. In this case the engineer's estimate has to be revised and verified for any possible mistake. In my data set, I observe a number of bids higher than 110% of the engineer's estimate. On multiple occasions, the winning bid is higher than this threshold. These facts suggest that bidders consider the probability of an event when this restriction comes into effect to be rather small. The assumption of no reserve price is justified in this environment.

Cost of bidding. Interviews with industry representatives suggest that the costs of preparing a bid are non-trivial and mostly related to the effort of avoiding mistakes in the bid calculation. I assume that the cost of bidding is proportional to the "size" of the project as represented by the common component of the cost. Presumably, the more expensive projects require a larger amount of information to be gathered and processed.

Number of bidders. It is unclear if the auction participants have a good idea about the number of their competitors. The existing literature on highway procurement auctions tends to argue that this is a small market where participants are well informed about each other and can accurately predict the identities of auction participants.[15] I follow this tradition and assume that the number of actual bidders is known to auction participants.

Bidders' heterogeneity. The main sources of heterogeneity between bidders participating in the market are specialization (single vs. multiple tasks), size (number of plants), and location. Location reflects the bidder's cost for the company of moving equipment, materials, and labor to the work site. Specialization may be important because projects in my data set usually involve some auxiliary work such as marking or landscaping. A firm that specializes in paving may have to subcontract these tasks, whereas a firm with multiple specializations may be able to perform all the tasks internally. This may result in cost differences.

Identification of the independent and affiliated private values models, which traditionally have been used to describe highway procurement auctions, requires that auctions with the same distribution of costs should be used in the estimation. The projects in my data set have distinct features including location, completion time, indication whether drainage, curb, sidewalk, marking, signing, or landscaping is included, type of the highway and the need for traffic control, the current state of the road, properties of the site, and the size of each task. Many, but not all, of these features are documented in the detailed description of the project. The characteristics which are not recorded in the data set may affect the cost distribution as well.

The engineer's estimate may seem a natural candidate to select projects with similar costs. However, projects with similar engineer's estimates need not have similar cost distributions. In my data set, the engineer's estimate can deviate from the winning bid by as much as 50% in either

[15] See, for example, P. Bajari and L. Ye (2002)

direction. It differs by almost 200% from the highest bid submitted in the same auction.

Summarizing, an auction model that accounts for unobserved auction heterogeneity may be an appropriate model for Michigan highway procurement auctions.

7.2 Descriptive statistics

Table 1c in part C of the Appendix depicts the distribution of the number of bidders. More than 80% of projects attracted between 2 to 6 bidders with the mean number of bidders equalling 3.7 and standard deviation of 2.6.

Table 2c tabulates the size distribution of projects. About 69% of the projects have an engineer's estimate between \$100,000 and \$1,000,000, five per cent are below \$100,000 and 24% above \$1,000,000.

Table 3c gives summary statistics of some other important variables. Statistics for the winning bid normalized by the engineer's estimate indicate that on average the engineer's estimate is a good measure of project cost since the normalized winning bid is not statistically different from one. Another important variable is "money left on the table" as represented by the difference between the lowest and second-to-lowest bid divided by the engineer's estimate. This variable is usually taken to indicate the extent of uncertainty present in the market. "Money left on the table" is statistically different from zero and is on average equal to 8% of the engineer's estimate which indicates that cost uncertainty may be substantial.

Table 4c reports OLS regression results of the log transformed bid normalized by the engineer's estimate on the project characteristics and a set of project fixed effects. It provides evidence in favor of unobserved heterogeneity. The first set of independent variables include log of the estimate, time to completion, indication if any of the supplementary work such as curb, sidewalk, drainage, marking, signing or landscaping is required. The estimates show that these variables explain about 60% of the variation in the dependent variable. Results of the second regression suggest that auction-specific dummies explained an additional 20%. Even after controlling for the engineer's estimate and main characteristics of the project, there is substantial residual variation in the bids across auctions. Table 5c presents results of random auction effects regression. It provides a variance decomposition of the residuals into auction specific and idiosyncratic components. The results of decomposition suggest that, in total, about 46% of the residual variation in the log bids may be auction specific.

Thus,the regression analysis provides strong evidence for the importance of unobserved heterogeneity in Michigan highway procurement auctions.

7.3 Estimation results

This section reports the estimation results for the multiplicative component model under the assumption of bidder symmetry.[16] Two specifications are estimated. In the first, the bid is a function of the cost, i.e. $b_i = b(c_i)$. In the second, the bid function is scaled by the engineer's estimate, $b_i = est * b(c_i/est)$.

Figures 1d-7d depict results for the first specification. Bids and costs are measured in tens of thousands of dollars. Figure 1d depicts the estimated density of the common cost component. The common component ranges between \$100,000 and \$700,000, with a mean of \$304,000 and standard deviation of \$194,255.

Figure 2d shows the probability densities of the individual bid and individual cost component.[17] The individual cost component has a higher variance and thus a flatter density than the individual bid component. The upper end of the support is common, but the lower end of the individual cost component extends further to the left than the lower end of the individual bid component. The mean of the individual component is fixed at one and standard error equals 0.33. The estimation results imply that the individual costs for the same project can differ by more than 100%.

Variance decomposition

Recall that bidder $i's$ cost for project j is given by $c_{ij} = y_j * x_{ij}$. The delta-method implies that asymptotically

$$Var(c) = (EY)^2 Var(X) + (EX)^2 Var(Y).$$

If $(EY)^2 Var(X)$ and $(EX)^2 Var(Y)$ are taken to represent parts of the cost variation generated by the variation in the individual and common cost components respectively, then it can be calculated that the individual cost component accounts for almost 20% of variation in the cost, whereas the common cost component accounts for the residual 80%.[18]

Mark-ups over the bidders costs

Figure 3d depicts the individual component of the inverse bid function. The estimated inverse

[16] The estimation results in Part D of the Appendix correspond to the case of four bidders. The estimation was also performed for the auctions with three and five bidders. Results look qualitatively similar for the higher number of bidders.

[17] The mean of individual cost component was fixed at one.

[18] Note that this decomposition does not depend on our choice of mean normalization. Suppose that X_0 and Y_0 are true random variables representing the individual and common cost components respectively. Due to normalization we are working with $X = \frac{1}{k}X_0$ and $Y = kY_0$, for some $k > 0$. Then

$$(EY)^2 Var(X) = k^2 (EY_0)^2 \frac{1}{k^2} Var(X_0),$$

i.e

$$(EY)^2 Var(X) = (EY_0)^2 Var(X_0).$$

Similarly,

$$(EX)^2 Var(Y) = (EX_0)^2 Var(Y_0).$$

bid function is used to compute mark-ups over the bidders costs. The normalized markup, $\frac{b-c}{c} = \frac{a-x}{x}$, $x = \phi(a)$, ranges from 0.15% to 30% and on average is equal to 11%. Mark-up for the winning bid is on average equal to 19%.

Comparison to the alternative auction models

Figures 4d and 5d compare the expected bid function to the bid function recovered under the APV and IPV models respectively.[19] Both the IPV and APV procedures estimate total costs that are substantially lower than the cost estimated under the unobserved heterogeneity assumptions. For example, at the bid level of $570,000, the model with unobserved heterogeneity estimates the average bidder's cost at $536,300, with a mark-up of $33,800 or 6.3%. For the same bid value, the model with independent private values estimates the bidder's cost at $504,300 with a mark-up of $65,800 or 13%; the model with affiliated private values estimates the bidder's cost at $479,000, with a mark-up of $91,100 or 19%. The bid function under heterogeneity is much steeper than the IPV and APV bid functions. In both cases, confidence intervals for the IPV and APV estimates intersect the confidence interval constructed under the null of unobserved heterogeneity only for a very small part near the upper end of the support. These results suggest that the APV and IPV models may lead to significant overestimation of markups and thus erroneous policy conclusions if the data are generated by the model with unobserved heterogeneity.

Figures 6d and 7d compare the total cost densities across the three models.[20] The IPV and APV densities are shifted to the left relative to the unobserved heterogeneity. The IPV density is flatter and has a longer left-side tail than the unobserved heterogeneity density. The densities differ mostly at the lower bound of the support. In both cases, confidence intervals for the IPV and APV estimates intersect the confidence interval constructed under the null of unobserved heterogeneity only for a very small part near the upper end of the support.

Normalization by engineer's estimate

The second specification normalizes bids by the engineer's estimate. Figure 8d depicts the individual bid component density, the individual costs component density and the common cost component density. As expected, the densities of the individual components are the same as in the first specification. The density of the common component is shifted to the left relative to the density of the individual component. The mean of the individual cost component is still fixed at one; the mean of the common component equals 0.88. In the second specification, the individual component becomes more important then the common component on average. However, the common component has a non-trivial effect even after normalization by engineer's estimate: it accounts for more than 35% of costs variation. Normalizing by engineer's estimate does not fully

[19] To compute the value of the expected inverse bid function at a point b, I first derived total costs for every value of the common component that could have resulted in a bid b and then computed an expectation of total costs with respect to the distribution of the common component.

[20] The total costs density is computed as a convolution of the densities of the common and the individual cost component.

account for the unobserved heterogeneity.

Figures 9d and 10d compare the expected bid function to the inverse bid function recovered under the APV and IPV models respectively. The estimates for the second specification suggest smaller differences in markups estimated from the three models. Confidence intervals for the IPV and APV estimates intersect the confidence interval constructed under the null of unobserved heterogeneity for a part of the support. However, the slopes of the bid functions remain very different (unobserved heterogeneity estimates are flatter), which leads to important differences in the shape of the total cost distribution. In particular, both the IPV and APV total cost densities, Figures 11d and 12d, are shifted to the left as compared to the UH estimate. They also allocate less mass to right tail of the cost distribution than the unobserved heterogeneity density. The last observation suggests that normalizing by the engineer's estimate may reduce bias in the estimates for the average mark-ups. It cannot, however, reduce bias in the total cost density estimate.

8 Summary and Conclusions

This paper addresses the issue of uncovering private information from observable bids in the presence of unobserved heterogeneity. A simple model is considered in which bidders' costs equal the product of an individual and a common component. New identification conditions are given, and it is shown how the result can be extended to more general settings. An associated estimation procedure is developed. Further, data for Michigan highway procurement auctions are analyzed. Regression analysis provides strong evidence of the importance of unobserved auction heterogeneity. That is why, the methodology developed in the earlier sections of the paper is used to estimate bid function and the distribtution of bidders' costs. The estimation results suggest that unobserved heterogeneity may account for a significant proportion of bid variation. Furthermore, it is illustrated that auction models that ignore the presence of unobserved heterogeneity may lead to substantially biased estimates of the markups and cost distribution function.

The estimation results suggest that, if the true environment is well described by the model with unobserved auction heterogeneity, then the magnitude of private information present in these auctions may be much smaller than is suggested by the estimates based on the alternative informational assumptions.

This paper focuses on the auction environment when there is no binding reserve price. The results can easily be extended to allow for a reserve price that is proportional to the common component. If the reserve price is not proportional to the common cost component, the common cost component and individual cost components are correlated conditional on participation in the auction. For this reason, the framework of this paper cannot be directly applied in such environments and further research of this issue is required.

Bidders may not know the actual number of their competitors. My model can easily be extended to environments in which the number of bidders is random and independent of the common cost component. Environments where the number of bidders is correlated with the common cost component will be studied in the future.

This paper studies unobserved auction heterogeneity. An unexplored issue concerns bidders' unobserved heterogeneity. While unobserved auction heterogeneity introduces the same change in the cost distribution for all bidders, unobserved bidder heterogeneity implies changes in the bidders' cost distribution which are bidder specific. The environments with unobserved bidder heterogeneity are more complex and require further research.

The methodology, developed in the context of the First-Price auction, can also be extended to other environments such as multi-unit auctions.

9 Bibliography

[1] Athey, S. and P. Haile (2002), "Identification of Standard Auction Models," *Econometrica*, 70(b), p. 2107-2140.

[2] Athey, S. and P. Haile (2000), "Identification of Standard Auction Models," MIT working paper 00-18 and SSRI working paper 2013.

[3] Bajari, P. and L. Ye (2001), "Competition versus Collusion in Procurement Auctions: Identification and Testing," Working Paper.

[4] Blum, J.R., J.Kiefer and M. Rosenblatt (1961), "Distribution Free Tests of Independence Based on the Sample Distribution Function," *Annals of Mathematical Statistics* 32, p. 485-498.

[5] Chakraborty,I., and G.Deltas (1998): "A Two-stage Approach to Structural Econometric Analysis of First-Price Auctions," Working Paper, University of Illinois.

[6] Campo S., Perrigne I., and Q. Vuong (2001), "Asymmetric Bidders in First-Price Auctions with Affiliated Private Values," forthcoming in *Journal of Applied Econometrics.*

[7] Diggle, P. J., and P. Hall (1993), "A Fourier Approach to Nonparametric Deconvolution of a Density Estimate," *J.Royal Statist. Soc.* Ser. B 55, p. 523 - 531.

[8] Donald, S., and H. Paarsch (1996), "Identification, Estimation and Testing in Parametric Empirical Models of Auctions within the Independent Private Values Paradigm," *Econometric Theory*, 12, p. 517-567.

[9] Elyakime, B., J. J. Laffont, P. Loisel and Q. Vuong (1994), "First-Price Sealed Bid Auctions with Secret Reservation Prices," *Annales d'Economie et de Statistique*, 34, p. 115-141.

[10] Elyakime, B., J. J. Laffont, P. Loisel and Q. Vuong (1997), "Auctioning and Bargaining: an Econometric Study of Timber Auctions with Secret Reservation Prices," Journal of Business and Economic Statistics, 15, p. 209-220.

[11] Fan, J. Q. (1991), "On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems," *Ann. Statist.* 19, p. 1257-1272.

[12] Friedman, M. (1991), "How to sell Government Securities," Wall Street Journal, August 28, p. A8.

[13] Guerre, E., I. Perrigne, and Q. Vuong (2000), "Optimal Nonparametric Estimation of First-Price Auctions," *Econometrica*, 68, p.525-574.

[14] Hardle, W. (1991), "Smoothing Technique with Implementation," S.Springer, New York.

[15] Hendricks, K. and H. J. Paarsch (1995), "A survey of recent empirical work concerning auctions," *Canadian Journal of Economics*, 28, p. 403-426.

[16] Hendricks, K. and R. Porter (1988), "An Empirical Study of an Auction with Asymmetric Information," *American Economic Review*, 78, p. 865-883.

[17] Hong H., and M. Shum (2002), "“Increasing Competition and the Winner's Curse: Evidence from Procurement," *Review of Economic Studies*, 69, p. 871-898.

[18] Horovitz J. L., and M. Marakatou (1996), "Semiparametric Estimation of Regression Models for Panel Data," *Review of Economic Studies*, 63, p.145-168.

[19] Jofre-Bonet M. and M. Pesendorfer (2001), "Estimation of a Dynamic Auction Game," NBER Working Paper #8626.

[20] Kotlarski, Ignacy (1966), "On Some Characterizations of Probability Distributions in Hilbert Spaces," *Annal di Matematica Pura et Applicate*, 74, p. 129-134.

[21] Laffont, J. J. (1997), "Game Theory and Empirical Economics: The Case of Auction Data," *European Economic Review*, 41, p. 1-35.

[22] Laffont, J. J., and Q. Vuong (1996), "Structural Analysis of Auction Data," *American Economic Review, Papers and Proceedings*, 86, p.414-420.

[23] LeBrun, B. (1999), "First-Price Auctions in the Asymmetric N Bidder Case," *International Economic Review*, 40, p. 125-142.

[24] Li, T., I. Perrigne, and Q. Vuong (2000), "Structural Estimation of the Affiliated Private Values Model," *Rand Journal of Economics*, forthcoming .

[25] Li, T., I. Perrigne, and Q. Vuong (2000), "Conditionally Independent Private Information in OCS Wildcat Auction," *Journal of Econometrics*, 98, p. 129-161.

[26] Li, T., and Q. Vuong (1998), "Nonparametric Estimation of Measurement Error Model Using Multiple Indicators," *Journal of Multivariate Analysis*, 65, p. 135-169.

[27] Lukacs, E. (1970), "Characteristic Function," Griffin, London.

[28] Maskin, E. and J. Riley (2000a), "Asymmetric Auctions," *Review of Economic Studies*, 67, p.413-438.

[29] Maskin, E. and J. Riley (2000b), "Existence of Equilibrium in Sealed High Bid Auctions," *Review of Economic Studies*, 67, 439-454.

[30] McAffee R. P. and J. MacMillan (1996), "Auctions and Bidding," *Journal of Economic Literature*, 25 p. 669-738.

[31] Milgrom, P. R. and R. J. Weber (1982), "A Theory of Auctions and Competitive Bidding," *Econometrica*, 50(5), p. 1089-1122.

[32] Milgrom, P. R. (1996), "Procuring Universal Services: Putting Auctions to Work," Lecture at the Royal Swedish Academy of Sciences, December 9, 1996.

[33] Paarsch, H.J. (1992), "Empirical Models of Auctions and an Application to British Columbian Timber Sales," Research Report 9212, University of Western Ontario.

[34] Perrigne, I. and Q. Vuong (1999): "Structural Econometrics of First-Price Auction: a Survey of Methods," *Canadian Journal of Agricultural Economics*, 47, p. 203-223.

[35] Porter, R. and D. Zona (1993), "Detection of Bid Rigging in Procurement Auctions," *Journal of Political Economy*, 101, p. 518-538.

[36] Rao, B. L. S. P. (1992): "Identifiability in Stochastic Models: Characterization of Probability Distributions," Academic Press, New York.

[37] Stone, C. J. (1992), "Optimal Rates of Convergence for Non-parametric Regressions," *Annals of Statistics*, 10, p. 1040-1053.

10 Appendix

Part A

Proof of Proposition 1:

LeBrun (1999) and Maskin and Riley (2000), establish that $(\alpha_{1y}(.), \alpha_{2y}(.))$ constitutes a unique vector of equilibrium strategies conditional on $Y = y$ if and only if functions $\alpha_{ky}(.)$ satisfy a system of differential equations for every a

$$\frac{1}{a - y * x} = \frac{\frac{(m_k-1)}{y} h_{X_k}(\frac{\alpha_{ky}^{-1}(a)}{y})}{(1 - H_{X_k}(\frac{\alpha_{ky}^{-1}(a)}{y}))\alpha'_{ky}(\alpha_{ky}^{-1}(a))} + \frac{\frac{m_{-k}}{y} h_{X_k}(\frac{\alpha_{ky}^{-1}(a)}{y})}{(1 - H_{X_k}(\frac{\alpha_{ky}^{-1}(a)}{y}))\alpha'_{ky}(\alpha_{ky}^{-1}(a))} \tag{5}$$

with boundary conditions given by

(A_1) $\alpha_{ky}(y * \overline{x}) = \overline{x}$ and (2) there exists $d_{ky} \in [y * \underline{x}, y * \overline{x}]$ such that $\alpha_{ky}(y * x) = d_{ky}$.

The vector of equilibrium strategies conditional on $y = 1$ satisfies the system of differential equations

$$\frac{1}{a - x} = \frac{(m_k - 1)h_{X_k}(\alpha_{k1}^{-1}(a))}{(1 - H_{X_k}(\alpha_{k1}^{-1}(a)))\alpha'_{k1}(\alpha_{k1}^{-1}(a))} + \frac{m_{-k}h_{X_k}(\alpha_{k1}^{-1}(a))}{(1 - H_{X_k}(\alpha_{k1}^{-1}(a)))\alpha'_{k1}(\alpha_{k1}^{-1}(a))}. \tag{6}$$

Let us consider a set of functions $(\gamma_{1y}, \gamma_{2y})$, $\gamma_{ky} : [y * \underline{x}, y * \overline{x}] \to (0, \infty)$ such that

$$\begin{aligned} \gamma_{ky}(z) &= y * \alpha_k(\frac{z}{y}), \\ \gamma_{ky}(y * \overline{x}) &= y * \overline{x}, \\ \gamma_{ky}(y * \underline{x}) &= y * d_{y1}. \end{aligned}$$

Then

$$\alpha_k^{-1}(\frac{a}{y}) = \frac{\gamma_{ky}^{-1}(a)}{y}$$

and

$$(\alpha_k^{-1}(z))' = (\frac{\gamma_{ky}^{-1}(y * z)}{y})' = (\gamma_{ky}^{-1}(y * z))'.$$

If we substitute functions γ_{1y}, γ_{2y} in the system of equations (5), then (5) can be transformed to

$$\frac{1}{\frac{a}{y} - x} = \frac{(m_k - 1)h_{X_k}(\alpha_{k1}^{-1}(\frac{a}{y}))}{(1 - H_{X_k}(\alpha_{k1}^{-1}(\frac{a}{y})))\alpha'_{k1}(\alpha_{k1}^{-1}(\frac{a}{y}))} + \frac{m_{-k}h_{X_k}(\alpha_{k1}^{-1}(\frac{a}{y}))}{(1 - H_{X_k}(\alpha_{k1}^{-1}(\frac{a}{y})))\alpha'_{k1}(\alpha_{k1}^{-1}(\frac{a}{y}))}. \tag{7}$$

Replacing $\frac{a}{y}$ with z in the system of equations (7), we are back to the system of equations (7), which we know is satisfied by $\alpha_{k1}(z)$, which implies that γ_{1y}, γ_{2y} satisfy the system of equations (5). They also satisfy corresponding boundary conditions by definition if d_{ky} is set equal to $y * d_{y1}$. Since the solution to the system (5) that satisfies boundary conditions (A_1) is unique and constitutes the set of equilibrium functions, γ_{1y}, γ_{2y} coincide with α_{1y}, α_{2y}. Thus

$$\alpha_{ky}(z) = y * \alpha_{k1}(\frac{z}{y}).$$

When $Y = y$, $z = y * x$, where $x \in [\underline{x}, \overline{x}]$, then

$$\alpha_{ky}(z) = y * \alpha_{k1}(x).$$

Proof of Theorem 1:

As has been established in Proposition 1, $B_{kij} = y_j * a_{kij}$, where a_{kij} is an individual bid component. Two bids per auction produce two relationships $B_1 = Y * A_1$ and $B_2 = Y * A_2$. Since Y and A_k's take only positive values, these relationships can be rewritten as

$$\begin{aligned} \log(B_1) &= \log(Y) + \log(A_1), \\ \log(B_2) &= \log(Y) + \log(A_2). \end{aligned}$$

The theorem by Kotlarski (1967)[21] applies directly in this environment and ensures that distributions of $\log(Y)$, $\log(A_1)$, and $\log(A_2)$ are identified up to a constant. To fix the constant we assume that $E(\log(A_1)) = 0$. Then, since the distribution functions of $\log(Y)$, $\log(A_1)$, and $\log(A_2)$ are uniquely identified, and, since $\log(.)$ is a strictly monotone function, then the distribution functions of Y, A_1 and A_2 are uniquely identified as well.

Since the individual bid components represent bids that would have been submitted in the auction game without unobserved heterogeneity and with asymmetric bidders, then the identification of the distribution of the individual cost component from the distribution of the individual bid component follows according to the results established by Laffont and Vuong (1996).

Remark

The realizations of the common component and individual cost component corresponding to bid b_{ij} are not uniquely identified. In particular, let us denote by $b_j = \{b_{ij}\}$ the vector of bids submitted in the auction j and by $x_j = \{x_{ij}\}$ a vector of individual cost component draws in the auction j. We will show now that for a generic b_j , (i.e. $b_j = \{b_{ij},\ \underline{y} * \alpha_i(\underline{x}) < b_{ij} < \overline{y} * \alpha_i(\overline{x})\}$) there exist multiple pairs (y_j, x_{ij}) such that $b_{ij} = \beta_{y_j}(x_{ij})$.

Consider $\{[\underline{y_i^0}, \overline{y_i^0}],\ \underline{y_i^0} = \max(\frac{b_{ij}}{\alpha_i(\overline{x})}, \underline{y}),\ \overline{y_i^0} = \min(\frac{b_{ij}}{\alpha_i(\underline{x})}, \overline{y})\}$.

[21] See Rao (1992).

If data are generated by the model with unobserved heterogeneity, then $b_{ij} = y_j * \alpha_{y_j}(x_{ij})$ and

$$\max(\frac{y_j * \alpha_{y_j}(x_{ij})}{\alpha_i(\overline{x})\}}, \underline{y}) \leq y_j \leq \min(\frac{y_j * \alpha_{y_j}(x_{ij})}{\alpha_i(\underline{x})}, \overline{y}),$$

with the equality on either side occurring only if $x_{ij} = \underline{x}$ or $x_{ij} = \overline{x}$. Since the event where $x_{ij} = \underline{x}$ or $x_{ij} = \overline{x}$ for some i has a probability of zero, generally

$$\max(\frac{y_j * \alpha_{y_j}(x_{ij})}{\alpha_i(\overline{x})\}}, \underline{y}) < y_j < \min(\frac{y_j * \alpha_{y_j}(x_{ij})}{\alpha_i(\underline{x})}, \overline{y}).$$

Thus,

$$\underline{y^0} = \max(y_i^0) < y_j < \min(y_i^0) = \overline{y^0}.$$

For any $y \in [\underline{y^0}, \overline{y^0}]$, let us define $a_{iy} = \frac{b_{ij}}{y}$, $i = 1,..,m$. Notice that $\alpha_i(x) \leq x_{iy} \leq \alpha_i(x)$ by construction. This means that the inverse bid function from (D_6) could be used to find x_{iy} such that $a_{iy} = \alpha_i(x_{iy})$. Thus, I have shown that there are multiple pairs $(y, \{x_{iy}\})$ that rationalize b_j.

Two-dimensional model with unobserved heterogeneity

The cost of bidder i is equal to $c_i = y_1 + y_2 * x_i$, where $Y = (Y_1, Y_2)$ is a random vector representing the common cost component and $X = (X_{11}, .., X_{1m_1}, X_{2,m_1+1}, .., X_{2m})$. Random variables (X, Y) are distributed on $[\underline{y_1}, \overline{y_1}] \times [\underline{y_2}, \overline{y_2}] \times [\underline{x}, \overline{x}]^m$ according to the probability distribution function $H(., .., .)$ with the associated probability density function $h(., .., .)$.

Assumptions $(F_1) - (F_6)$ are analogous to $(D_1) - (D_6)$ of the one-dimensional case.

(F_1) The components of Y and X are independent:

$$H(y_{10}, y_{20}, x_{10}, .., x_{m0}) = H_{Y_1}(y_{10}) H_{Y_2}(y_{20}) \prod_{j=1}^{j=m_1} H_{X_1}(x_{j0}) \prod_{j=m_1+1}^{j=m} H_{X_2}(x_{j0}),$$

where H_{Y_1}, H_{Y_2}, H_{X_1} and H_{X_2} are marginal distribution functions of Y_1, Y_2, X_{1j} and X_{2j} respectively.

(F_2) The probability density functions of the individual cost component, h_{X_1} and h_{X_2}, are continuously differentiable and bounded away from zero on $[\underline{x}, \overline{x}]$.

(F_3) $EX_{1j} = 1$ for all $j = 1, ..., m_1$.

(F_4) (a) The number of bidders is common knowledge;

(b) There is no binding reservation price;

(c) The cost of preparing the bid is proportional to the realization of y.

(F_5) The characteristic functions of $\log(Y_2)$, $\log(A_{1i_l} - A_{ki_p})$, $\log(A_{2i_q} - A_{ki_p})$, and the joint characteristic function of $(Y_2 X_{1i_l}, Y_2 X_{2i_q})$ are non-vanishing everywhere;

(F_6) The inverse bid functions given by

$$\phi_k(a) = a - \frac{(1 - G_{A_k}(a))(1 - G_{A_{-k}}(a))}{(m_k - 1)g_{A_k}(a)G_{A_{-k}}(a) + m_{-k}g_{A_{-k}}(a)G_{A_k}(a)},\ k = 1, 2,$$

are strictly increasing functions of a. Here $G_{A_k}(.)$ is the probability distribution function of the type k individual bid component with associated probability density function.

The condition (F_2) insures the existence and uniqueness of the equilibrium in the auction game corresponding to the realization (y_1, y_2) of (Y_1, Y_2). The conditions (F_1),(F_3),(F_4), and (F_5) provide a basis for the identification of the probability density functions of Y_1,Y_2, A_1 and A_2. The condition (F_6) guarantees that the probability density functions of X_1 and X_2 are uniquely identified.

Theorem 1a

If conditions (F_1)–(F_6) are satisfied, then the probability density functions h_{Y_1},h_{Y_2},h_{X_1}, h_{X_2} are uniquely identified from the joint distribution of four arbitrary bids $(B_{i_1}, B_{i_2}, B_{i_3}, B_{i_4})$.

Sketch of the proof

(a) Applying Kotlarski's argument to the log-transformed random variables $(B_{i_1} - B_{i_2})$ and $(B_{i_3} - B_{i_4})$ allows identification of the probability density function h_{Y_2}.

(b) The joint characteristic function of $((B_{i_1} - B_{i_3}), (B_{i_2} - B_{i_3}))$ in conjunction with the characteristic function of Y_2 (identified in (a)) allows identification of the joint characteristic function of $((A_{i_1} - A_{i_3}), (A_{i_2} - A_{i_3}))$, which according to the Kotlarski argument in turn implies that the probability density functions of A_{i_1} and A_{i_2} are identified.

(c) The probability density functions $g_{A_{i_1}}$, h_{Y_2} uniquely determine the probability distribution and thus characteristic function of $Y_2 * A_{i_1}$, which allows uniquely identification of the probability distribution of Y_1 from the characteristic function of B_{i_1}.

(d) The argument developed in Laffont and Vuong (1996) can be applied to establish identification of the probability density functions from the probability distribution of A_{i_1} and A_{i_2}.

Thus I have established that h_{Y_1}, h_{Y_2}, h_{X_1}, h_{X_2} are identified from the joint distribution of four arbitrary bids.

Similar to the one-dimensional case, the exact realizations of y_{1j}, y_{2j} and $\{x_{ij}\}$ are not uniquely identified.

Proof of proposition 2:

According to (W_2), $\Phi_{\log(Y)}(.)$, $\Phi_{\log(A_{1i_l})}(.)$ and $\Phi_{\log(A_{2i_p})}(.)$ are the same for all pairs (i_l, i_p) such that $i_p = 1, .., m_1$ and $i_l = m_1 + 1, .., m_2$. This implies that i indices can be dropped, so that we can focus on just three functions $\Phi_{\log(Y)}(.)$, $\Phi_{\log(A_1)}(.)$, and $\Phi_{\log(A_2)}(.)$.

If (W_1) is satisfied, then there exist independent random variables Y, A_1 and A_2 such that the characteristic functions of $\log(Y)$, $\log(A_1)$, and $\log(A_2)$ are given by $\Phi_{\log(Y)}(.)$, $\Phi_{\log(A_1)}(.)$, $\Phi_{\log(A_2)}(.)$ respectively. Kotlarski (1967) shows that

$$\Phi_{\log(Y)}(t_1 + t_2)\Phi_{\log(A_1)}(t_1)\Phi_{\log(A_2)}(t_2) = \Psi(t_1, t_2).$$

This equality implies that $(\log(Y) + \log(A_1), \log(Y) + \log(A_2))$ are distributed the same as $(\log(B_1), \log(B_2))$.

Let us consider $X_k = \phi_k(A_k)$. Then Y, X_1 and X_2 define the model with unobserved heterogeneity that rationalizes the data.

Proof of proposition 3

(1) The proof follows from the property of independent variables: if the random variables Z_1 and Z_2 are independent then so are $f(Z_1)$ and $f(Z_2)$, for any function $f(.)$.

(2) If X_i's are independent then so are $log(X_i)$. The structure of the bidder's cost, $c_{ij} = y_j * x_{ij}$, implies that $\log(\frac{B_{1i_1j}}{B_{1i_1j}}) = \log(A_{1i_1j}) - \log(A_{2i_2j})$ and $\log(\frac{B_{1i_1j}}{B_{2i_2j}}) = \log(A_{1i_3j}) - \log(A_{2i_2j})$. Then by Kotlarski (1967) theorem the characteristic function of $log(A_1)$ is given by

$$f_1(t) = \int_0^t \frac{\Theta_1(0, u_2)}{\Theta(0, u_2)} du_2.$$

Therefore if X_i's are independent and $c_{ij} = y_j * x_{ij}$ then $f_1(t)$ is a characteristic function.

(3) If Y and X_i's are independent, the cost structure is given by $c_{ij} = y_j * x_{ij}$, then Kotlarski (1967) theorem applied to $(\log(B_{i_1j}), \log(B_{i_2j}))$ implies that the characteristic function of $\log(A_{1i_1j})$ are given by the function $\Phi_{\log(A_1)}(t)$. Kotlarski (1967) theorem applied to $\log(\frac{B_{1i_1j}}{B_{1i_1j}})$ and $\log(\frac{B_{1i_1j}}{B_{2i_2j}})$ implies that the characteristic function of $\log(A_{1i_1j})$ is given by $f_1(t)$. Thus the following equality has to hold

$$\Phi_{\log(A_1)}(t) = f_1(t).$$

Estimation of the support bounds

An important part of the first step is the estimation of the supports of distributions of the individual bid and the common cost components. Notice, that the distributions of the components are identified up to the location only. So, I start with an arbitrary choice of supports, then estimate the shift in supports that accompanies the first stage estimation. Finally, after the second stage I adjust supports, so that estimated distributions satisfy assumption (D_3).

Initially, I ignore the assumption (D_3), $EX_{i1} = 1$. Instead, I assume that there are no restrictions on the means of the distributions. To fix the supports of the distributions in question, I assume that support of LY is symmetric around zero. I denote the support of the log of the common component by $[-y^0, y^0]$ and the support of the log of the individual bid component of the type 1 bidder by $[\underline{a_1}^0, \overline{a_1}^0]$. Then the support of the log of bids for type 1 is given by $[\underline{a}_1^0 - y^0, \overline{a}_1^0 + y^0]$, and the support of the differences in logs of bids is given by $[\underline{a}_1^0 - \overline{a}_1^0, \overline{a}_1^0 - \underline{a}_1^0]$. Since the bounds of these supports can be estimated as $[min(\log(b_{1lj})), max(\log(b_{1lj}))]$ and $[min(\log(b_{1lj}) - \log(b_{1pj})), max(\log(b_{1lj}) - \log(b_{1pj}))]$, we arrive at the system of equations

$$\begin{aligned} min(\log(b_{1lj})) &= \widehat{\underline{a}}_1^0 - \widehat{y}^0, \\ max(\log(b_{1lj})) &= \widehat{\overline{a}}_1^0 + \widehat{y}^0, \\ max(\log(b_{1lj}) - \log(b_{1pj})) &= \widehat{\overline{a}}_1^0 - \widehat{\underline{a}}_1^0, \end{aligned}$$

which can be solved to get

$$\begin{aligned} \widehat{y}^0 &= \frac{max(\log(b_{1lj})) - min(\log(b_{1lj})) - max(\log(b_{1lj}) - \log(b_{1pj}))}{2}, \\ \widehat{\underline{a}}_1^0 &= \frac{min(\log(b_{1lj})) + max(\log(b_{1lj})) - max(\log(b_{1lj}) - \log(b_{1pj}))}{2}, \\ \widehat{\overline{a}}_1^0 &= \frac{min(\log(b_{1lj})) + max(\log(b_{1lj})) + max(\log(b_{1lj}) - \log(b_{1pj}))}{2}. \end{aligned}$$

Formulas for the estimation of the characteristic function of the common cost component in (2) and (4) have been derived under the assumption that $E(LA_1) = 0$. Hence, the mean of the common component equals the bids' mean. Thus, the probability density functions $\widehat{h}_Y(.)$ and $\widehat{h}_{LA_1}(.)$ are shifted so as to achieve a zero mean for the distribution of the log of individual bid component LA_1. If the symmetrization of the common component support initially assigned a mean of e_1 to the individual bid component of type 1, then Step 1 is going to produce density $\widehat{h}_{LY}(.)$ with the support $[-\widehat{y}^0 + e_1, \widehat{y}^0 + e_1]$ and $\widehat{g}_{LA_1}(.)$ with the support $[\widehat{\underline{a}}_1^0 - e_1, \widehat{\underline{a}}_1^0 - e_1]$, where the shift factor e_1 is given by the solution to the equation

$$\int_{\widehat{\underline{a}}_1^0 - e_1}^{\widehat{\underline{a}}_1^0 - e_1} a\widehat{g}_{LA_1}(a)da = 0.$$

I use this equation to estimate e_1 through a line search method.

The procedure described above produces intermediate estimates for the supports of Y and A_k

$$[\widehat{\underline{y}}_1, \widehat{\overline{y}}_1] = [\exp(-\widehat{y}^0 + \widehat{e}_1), \exp(\widehat{y}^0 + \widehat{e}_1)],$$
$$[\widehat{\underline{a}}_1^1, \widehat{\overline{a}}_1^1] = [\exp(\widehat{\underline{a}}_1^0 - \widehat{e}_1), \exp(\widehat{\overline{a}}_1^0 - \widehat{e}_1)],$$
$$[\widehat{\underline{a}}_2^1, \widehat{\overline{a}}_2^1] = [\exp(min(\log(b_{2lj})) + \widehat{y}^0 - \widehat{e}_1), \exp(max(\log(b_{2lj})) - \widehat{y}^0 + \widehat{e}_1)].$$

Sketch of the proof of proposition 4 :

(1) Uniform consistency of the estimators for the probability density functions of the common cost component, $\widehat{h}_Y$, and individual bid components, $\widetilde{g}_{A_k}$, follows from the Theorem 3.1 - 3.4 in Li and Vuong (1998).

(2) Uniform consistency of the estimators for the individual inverse bid function and the probability density function of individual component follows the logic of Proposition 3 and Theorem 3 of Guerre, Perrigne, Vuong (2000). Notice that I do not derive an optimal rate of convergence.

Recall that Guerre, Perrigne, Vuong (2000) rely on four sets of assumptions. Assumption 1(i) states that the observations on number of bidders and exogeneous covariates are independent draws from corresponding distributions. I do not need this assumption because I assume that all exogenous covariates are discreet and perform estimation conditional on the realization of covariates values and the number of bidders. Notice that d, which in Guerre, Perrigne, Vuong (2000) is equal to the number of continuous exogenous variables, is equal to zero in my case. I do maintain the Assumption 1(ii) which postulates that individual cost components of biders, $\{x_{ij}\}$, $i \in \{1, .., m\}$ and $j \in \{1, ..., n\}$, are independent draws from corresponding distributions H_{X_1} and H_{X_2}.

Assumptions 2, 3 and 4 are modified in the following way:

Assumption 2:

(i) The supports of H_{X_k} are given by $S(H_k) = [\underline{x}, \overline{x}]$, $x \geq 0$, $x \leq x$, for $k \in \{1, 2\}$;

(ii) For every $x \in S(H_k)$, $h_k(x) \geq c_h > 0$;

(iii) $H_k(.)$ admit up to $R+1$ continuous bounded derivatives, with $R \geq 1$.

Notice, that Assumption 2 basically repeats assumption (D_2) which assumes $R = 1$. The proof is given for a more general case of $R \geq 1$.

Assumption 3:

(i) The kernel $K_h(.)$ is symmetric with bounded hypercube support and twice continuous bounded derivatives;

(ii) $\int K_h(x)dx = 1$

(iii) $K_h(.)$ is of order $R+1$. Thus moments of order strictly smaller than given order vanish.

Assumption 4:

(i) The bandwidth δ_{h_k} is of the form

$$\delta_{h_k} = \lambda_{h_k}(\frac{\log(L)}{L})^{R/(2R+1)},$$

where λ_{h_k} is a strictly positive constant and L the number of pseudo bids draws in the second stage estimation.

We also additionally assume that $h_Y(.)$ is a continuous function.

The proof consists of several steps.

(a) First, we establish that the distribution function and the probability density functions of the individual bid component inherit properties of the distribution function and the probability density functions of the individual cost component. Namely,

Lemma 1

Given Assumption 2, the distribution functions $G_{A_k}(.)$ satisfy:

(i) its supports $S(G_{A_k})$ are given by $[\underline{a}_k, \overline{a}_k]$ with $\overline{a}_k = \overline{x}$;

(ii) for $a \in S(G_{A_k})$, $g_{A_k}(a) \geq c_g > 0$;

(iii) G_{A_k} admit up to $R+1$ continuous bounded derivatives on $S(G_{A_k})$;

(iv) g_{A_k} admit up to $R+1$ continuous bounded derivatives on every closed subset of the interior of $S(G_{A_k})$.

(v) g_{B_k} admit up to $R+1$ continuous bounded derivatives on every closed subset of the interior of $S(G_{A_k})$ and for every $b \in [\underline{y}\underline{a}_k, \overline{y}\overline{a}_k]$, $g_{B_k}(b) \geq c_{g_B} > 0$.

The proof of Lemma 1, (i)-(iv) closely follows the proof of Proposition 1 from Guerre, Perrigne and Vuong (2000). Part (v) follows from

$$g_{B_k}(b) = \int_{\underline{y}}^{\overline{y}} g_{A_k}(\frac{b}{y})h_Y(y)dy,$$

properties (i)-(iv) of g_{A_k} and continuity of $h_Y(.)$.

(b) Next, we derive the rate of convergence for the support bounds, $\underline{a}_k$ and $\overline{a}_k$. Recall that bounds of supports have been derived in several steps. First, supports of the distributions of LB_{1i} and $(LB_{1i_1} - LB_{1i_2})$ have been estimated as

$$[min(\log(b_{1lj})), max(\log(b_{1lj}))]$$
$$[min(\log(b_{1lj}) - \log(b_{1pj})), max(\log(b_{1lj}) - \log(b_{1pj}))].$$

These are maximum likelihood estimators for the support bounds of corresponding densities. (They are well defined due to (v) of Lemma1.) We know that they converge to the true value of the support bounds at the rate of n. The preliminary estimates for the bounds of LA_k supports, $\widehat{\underline{a}}_k^0$ and $\widehat{\overline{a}}_k^0$, are obtained as linear functions of the supports bounds for LB_{1i} and $(LB_{1i_1} - LB_{1i_2})$. Therefore they also converge to the true support bounds at the rate of n. Next stage obtains intermediate estimates of the support bounds, $\widehat{\underline{a}}_k^1$ and $\widehat{\overline{a}}_k^1$. They are obtained from $\widehat{\underline{a}}_k^0$ and $\widehat{\overline{a}}_k^0$ through a shift by an adjustment factor e_1. An extremum estimator for e_1 is obtained by minimizing

$$\widehat{Q}_n = \left(\int_{\widehat{\underline{a}}_1^0 - \widehat{e}_1}^{\widehat{\overline{a}}_1^0 - \widehat{e}_1} a \widehat{g}_{LA_1}(a) da \right)^2,$$

or

$$\widehat{Q}_n = \left(\int_{\widehat{\underline{a}}_1^0}^{\widehat{\overline{a}}_1^0} a \widehat{g}_{LA_1}(a + \widehat{e}_1) da - \widehat{e}_1 \right)^2.$$

The usual results for extremum estimators apply. Notice that $\widehat{Q}_n \to \left(\int_{\underline{a}_1^0}^{\overline{a}_1^0} a g_{LA_1}(a + e_1) da - e_1 \right)^2$ at the same rate as $\widehat{g}_{LA_1}$ converges to g_{LA_1} (see Li and Vuong (1998) for an appropriate rate of convergence). Let us denote this rate by d_n. It can be shown that all standard conditions for the convergence of extremum estimators hold and $\widehat{e}_1$ converges to e_1 at the rate d_n. Thus intermediate estimators of the supports bounds of LA_k, $\widehat{\underline{a}}_k^1$ and $\widehat{\overline{a}}_k^1$, converge to the corresponding true values at the rate d_n. The bounds of supports for A_k are estimated as $\widehat{\underline{a}}_k = exp(\widehat{\underline{a}}_k^1)$ and $\widehat{\overline{a}}_k = exp(\widehat{\overline{a}}_k^1)$ respectively. The smoothness of the exponential function ensures consistency of these estimators. The argument similar to delta method can be used to show that the rate of convergence remains equal to d_n.

(c) The rate of convergence for $\widehat{g}_{A_k}(.)$ is established in Li and Vuong (1998). Recall that here we denote it d_n. Now we derive a rate of convergence for $\widehat{G}_{A_k}$. The estimator for G_{A_k} is defined as

$$\widehat{G}_{A_k}(a) = \int_{\widehat{\underline{a}}_k}^{a} \widehat{g}_{A_k}(a) da.$$

To establish consistency we consider

$$\left| \widehat{G}_{A_k}(a) - G_{A_k}(a) \right| \leq \left| \int_{\widehat{\underline{a}}_k}^{\underline{a}_k} \widehat{g}_{A_k}(a) da \right| + \left| \int_{\underline{a}_k}^{a} (\widehat{g}_{A_k}(a) - g_{A_k}(a)) da \right|.$$

Since g_{A_k} is a continuous function with bounded support, (D_9), then g_{A_k} is a bounded function. For large enough n, $\widehat{g}_{A_k}$ is also bounded a.s. due to uniform convergence of $\widehat{g}_{A_k}$ to g_{A_k}. Then, part (b) implies that the first summand converges to zero at the rate d_n. The second summand also converges to zero at the rate d_n since support of g_{A_k} is bounded. Therefore, $\widehat{G}_{A_k}$ converges to G_{A_k} at the rate d_n.

(d) Next, we prove uniform consistency of the estimator for the individual cost component. The proof closely follows the argument of the Proposition 3 of Guerre, Perrigne, Vuong (2000). Recall that the individual cost components corresponding to the individual bid components a_k are estimated as

$$\begin{aligned}
\widetilde{x}_1 &= \widetilde{a}_1 + \frac{(1-\widetilde{G}_{A_1}(a_1))\cdot(1-\widetilde{G}_{A_2}(a_1))}{(m_1-1)\cdot\widetilde{g}_{A_1}(a_1)\cdot(1-\widetilde{G}_{A_2}(a_1))+m_2\cdot\widetilde{g}_{A_2}(a_1)\cdot(1-\widetilde{G}_{A_1}(a_1))},\\
\widetilde{x}_{2j} &= \widetilde{a}_{2j} + \frac{(1-\widetilde{G}_{A_1}(a_2))\cdot(1-\widetilde{G}_{A_2}(a_2))}{m_1\cdot\widetilde{g}_{A_1}(a_2)\cdot(1-\widetilde{G}_{A_2}(a_2))+(m_2-1)\cdot\widetilde{g}_{A_2}(a_2)\cdot(1-\widetilde{G}_{A_1}(a_2))}.
\end{aligned}$$

Similar to Guerre, Perrigne and Vuong (2000) I restrict my attention to the subset of the support

$$V(G_{A_k}) = \{a \in [\underline{a}_k, \overline{a}_k] \text{ such that } (a + S(2\delta_k)) \in S(H_k).$$

Notice that for every $a_{1j} \in V(G_{A_k})$ corresponding x_{1j} is finite. For every $a \in V(G_{A_k})$, $\widehat{g}_{A_k}(a) \geq c_g$ and $(1-\widehat{G}_{A_k}(a)) \geq c_G$ for some c_g and c_G, since $\widehat{g}_{A_k}$ and $\widehat{G}_{A_k}$ uniformly converge to g_{A_k} and G_{A_k} respectively and (ii) of Lemma 1.

Below I sketch the argument that establishes uniform convergence of $\widetilde{x}_{1j}$ to x_{1j}.

Let us denote

$$\begin{aligned}
\xi_1(a_1) &= \frac{(1-G_{A_1}(a_1))\cdot(1-G_{A_2}(a_1))}{(m_1-1)\cdot g_{A_1}(a_1)\cdot(1-G_{A_2}(a_1))+m_2\cdot g_{A_2}(a_1)\cdot(1-G_{A_1}(a_1))},\\
\widetilde{\xi}_1(a_1) &= \frac{(1-\widetilde{G}_{A_1}(a_1))\cdot(1-\widetilde{G}_{A_2}(a_1))}{(m_1-1)\cdot\widetilde{g}_{A_1}(a_1)\cdot(1-\widetilde{G}_{A_2}(a_1))+m_2\cdot\widetilde{g}_{A_2}(a_1)\cdot(1-\widetilde{G}_{A_1}(a_1))},
\end{aligned}$$

$$\begin{aligned}
\zeta_1(a_1) &= (m_1-1)\cdot g_{A_1}(a_1)\cdot(1-G_{A_2}(a_1))+m_2\cdot g_{A_2}(a_1)\cdot(1-G_{A_1}(a_1)),\\
\widetilde{\zeta}_1(a_1) &= (m_1-1)\cdot\widetilde{g}_{A_1}(a_1)\cdot(1-\widetilde{G}_{A_2}(a_1))+m_2\cdot\widetilde{g}_{A_2}(a_1)\cdot(1-\widetilde{G}_{A_1}(a_1)),
\end{aligned}$$

$$\begin{aligned}
\varepsilon_1(a_1) &= (1-G_{A_1}(a_1))\cdot(1-G_{A_2}(a_1)),\\
\widetilde{\varepsilon}_1(a_1) &= (1-\widetilde{G}_{A_1}(a_1))\cdot(1-\widetilde{G}_{A_2}(a_1)).
\end{aligned}$$

Then

$$|\widetilde{x}_{1j} - x_{1j}| = \left|\widetilde{\xi}_1(a_1) - \xi_1(a_1)\right|,$$

which in turn can be bounded by

$$\left|\widetilde{\xi}_1(a_1) - \xi_1(a_1)\right| \leq \frac{1}{\widetilde{C_1}C_1}\left|\widetilde{\varepsilon}_1(a_1)\zeta_1(a_1) - \varepsilon_1(a_1)\widetilde{\zeta}_1(a_1)\right|$$

or

$$\left|\widetilde{\xi}_1(a_1) - \xi_1(a_1)\right| \leq \frac{1}{\widetilde{C_1}C_1}(\left|\widetilde{\varepsilon}_1(a_1) - \varepsilon_1(a_1)\right| \cdot \left|\zeta_1(a_1)\right| + \left|\widetilde{\zeta}_1(a_1) - \zeta_1(a_1)\right| \cdot \left|\varepsilon_1(a_1)\right|),$$

or

$$\left|\widetilde{\xi}_1(a_1) - \xi_1(a_1)\right| \leq \frac{1}{\widetilde{C_1}}\left|\widetilde{\varepsilon}_1(a_1) - \varepsilon_1(a_1)\right| + \frac{\widetilde{c}_G c_G}{\widetilde{C_1}C_1}\left|\widetilde{\zeta}_1(a_1) - \zeta_1(a_1)\right|,$$

where $C_1 = (m_1 + m_2 - 1)c_g c_G$ and $\widetilde{C}_1 = (m_1 + m_2 - 1)\widetilde{c}_g \widetilde{c}_G$.

The analog of delta method allows us to conclude that

$$\begin{aligned}\left|\widetilde{\varepsilon}_1(a_1) - \varepsilon_1(a_1)\right| &= O_p(d_n), \text{ a.s.}\\ \left|\widetilde{\zeta}_1(a_1) - \zeta_1(a_1)\right| &= O_p(d_n), \text{ a.s.}\end{aligned}$$

Then

$$\left|\widetilde{\xi}_1(a_1) - \xi_1(a_1)\right| = O_p(d_n), \text{ a.s.}$$

To conclude the proof we note that δ_{h_k} converges to zero as n diverges to infinity (we choose L so that it diverges to infinity together with n) and thus the statement above holds everywhere on the interior of the support.

(e) Finally, we establish uniform convergence of the probability density function of the individual cost component. Here again we consider closed subsets of the support interior. Recall that

$$\widetilde{h}_{X_k}(x) = \frac{1}{L\delta_{h_k}}\sum_{j=1}^{n} K_h(\frac{x - \widetilde{x}_{kj}}{\delta_{h_k}}).$$

here L is the size of pseudo-sample. We will make it a function of n, i.e. $L = L(n)$. Let us denote

$$\widetilde{\widetilde{h}}_{X_k}(x) = \frac{1}{L\delta_{h_k}}\sum_{j=1}^{n} K_h(\frac{x - x_{kj}}{\delta_{h_k}}).$$

Then

$$\left|\widetilde{h}_{X_k}(x) - h_{X_k}(x)\right| \leq \left|\widetilde{h}_{X_k}(x) - \widetilde{\widetilde{h}}_{X_k}(x)\right| + \left|\widetilde{\widetilde{h}}_{X_k}(x) - h_{X_k}(x)\right|.$$

The rate of convergence for the second term depends solely on L and is equal to $(\frac{\log(L)}{L})^{R/(2R+1)}$

(see Stone 1982). Next, we focus on the first term:

$$\left|\widetilde{h}_{X_k}(x) - \widetilde{\widetilde{h}}_{X_k}(x)\right| = \left|\frac{1}{L\delta_{h_k}}\sum_{j=1}^{n}(K_h(\frac{x-\widetilde{x}_{kj}}{\delta_{h_k}}) - K_h(\frac{x-x_{kj}}{\delta_{h_k}})\right|.$$

A second-order Taylor expansion gives

$$\left|\widetilde{h}_{X_k}(x) - \widetilde{\widetilde{h}}_{X_k}(x)\right| \leq \left|\frac{1}{L\delta_{h_k}}\sum_{j=1}^{n}\frac{1}{\delta_{h_k}}\frac{dK_h}{dx}(\frac{x-x_{kj}}{\delta_{h_k}})\cdot(x-\widetilde{x}_{kj}) + \frac{1}{L\delta_{h_k}}\sum_{j=1}^{n}\frac{1}{\delta_{h_k}^2}\cdot\frac{d^2K_h}{dx^2}(\frac{x-x_{kj}}{\delta_k^2})\cdot(x-\widetilde{x}_{kj})^2\right|$$

or

$$\left|\widetilde{h}_{X_k}(x) - \widetilde{\widetilde{h}}_{X_k}(x)\right| \leq \left|\frac{1}{L\delta_{h_k}}\sum_{j=1}^{n}\frac{dK_h}{dx}(\frac{x-x_{kj}}{\delta_{h_k}})\right|\cdot\left|\frac{x-\widetilde{x}_{kj}}{\delta_{h_k}}\right| + \left|\frac{1}{L\delta_{h_k}}\sum_{j=1}^{n}(\frac{1}{\delta_{h_k}})\frac{d^2K_h}{dx^2}(\frac{x-x_{kj}}{\delta_{h_k}})\right|\cdot\left|\frac{x-\widetilde{x}_{kj}}{\delta_{h_k}}\right|^2.$$

The terms

$$\frac{1}{L\delta_{h_k}}\sum_{j=1}^{n}\frac{dK_h}{dx}(\frac{x-x_{kj}}{\delta_{h_k}}),$$
$$\frac{1}{L\delta_{h_k}}\sum_{j=1}^{n}(\frac{1}{\delta_k})\frac{d^2K_h}{dx^2}(\frac{x-x_{kj}}{\delta_{h_k}})$$

can be considered as kernel estimators. It can be shown that they converge to

$$h_{X_k}(x)\int\frac{dK_h}{dx}(x-z)dzdx,$$
$$h_{X_k}(x)\int\frac{d^2K_h}{dx^2}(x-z)dzdx$$

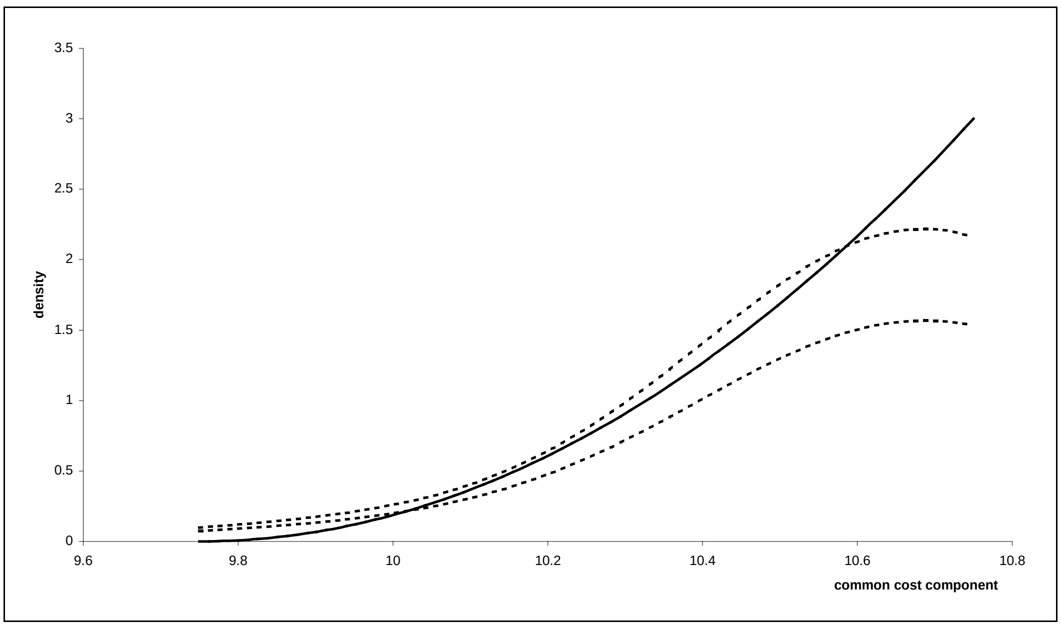
respectfully, which ensures that corresponding terms are bounded. Recall that the rate of convergence for $(x - x_{kj})$ is given by d_n. If δ_{h_k} is of the order r_{h_k} then the rate of convergence for the first term is given by δ_k/r_{h_k}. This is also the rate of convergence for $\widetilde{h}_{X_k}(x) - \widetilde{\widetilde{h}}_{X_k}(x)$, since the second term has a smaller order of magnitude. The bandwidth is chosen as a function of the number of random draws L, which is in turn is a function of the number of auctions in the data set n. the number of draws L can always be chosen so that $\delta_{h_k}/r_{h_k} \to 0$. The rate of convergence for $(\widetilde{h}_{X_k}(x) - h_{X_k}(x))$ is then given by $max\{\delta_{h_k}/r_{h_k}, (\frac{\log(L)}{L})^{R/(2R+1)}\}$. This concludes the proof of the Proposition 4.

Part B: Results of a Monte Carlo Study

Simulated data:

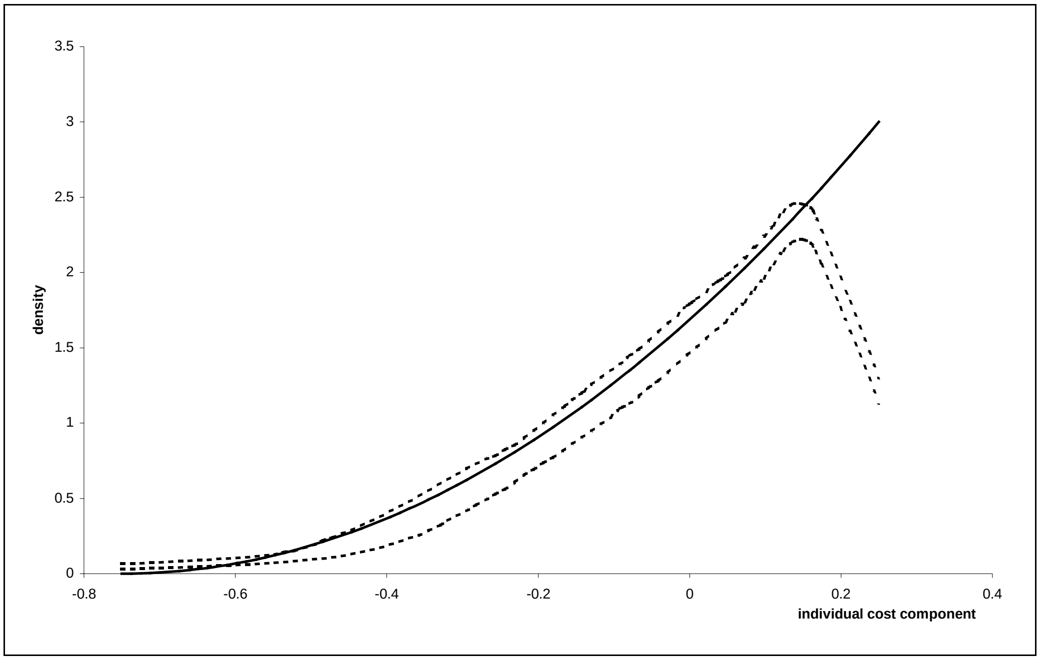
The individual cost component, X, is distributed on [-0.75, 0.25] according to the distribution function $H_X(x)=(x+0.75)^3$; the common component is distributed on [9.25, 10.75] according to the distribution function $H_Y(x)=(x-9.25)^3$; the total cost is equal to $c_{ij}=x_{ij}+y_j$, where y_j is a realization of common component, whereas x_{ij} is a realization of individual component in the draw j; the simulated bids are calculated using formula for equilibrium bid function
$\{b_{ij}=y_j + x_{ij}+ ((1+4*x_{ij})-(1-(x_{ij}+0.75)^4))/(4*(1-(x_{ij}+0.75)^3)), i=1,..,m\ \ j=1,..,n\}$;
n=500, m=2, the experiment was repeated 500 times.

Figure 1b: Probability density function of the common cost component
(model with unobserved heterogeneity)



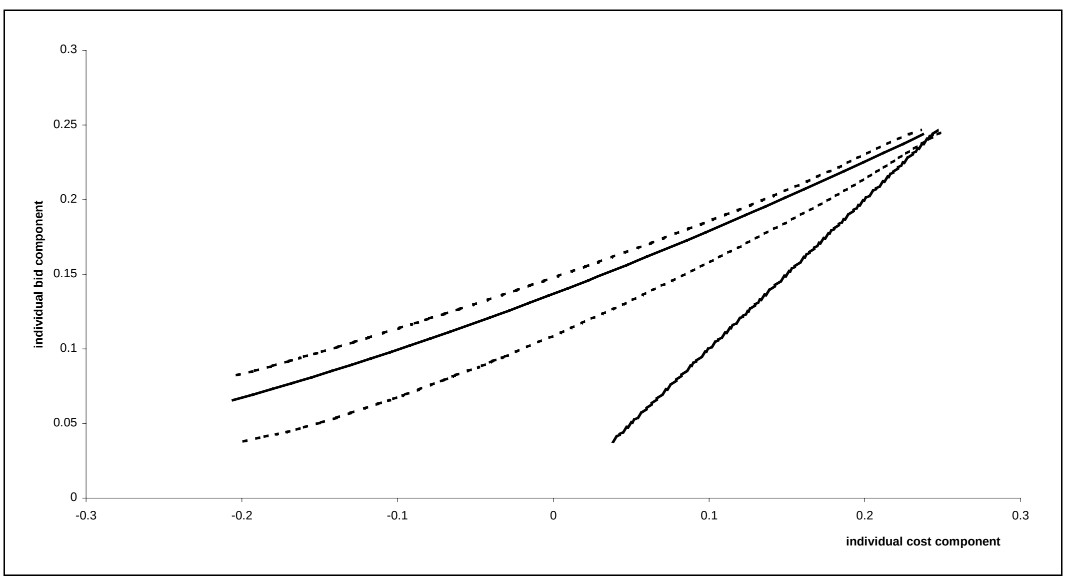
Solid line - true probability density function
Dotted lines - 5% and 95% pointwise quantiles of the estimator distribution

Figure 2b: Probability density function of the individual cost component
(model with unobserved heterogeneity)



Solid line - true probability density function
Dotted lines – 5 % and 95 % pointwise quantiles of the estimator distribution

Figure 3b: Individual bid function
(model with unobserved heterogeneity)

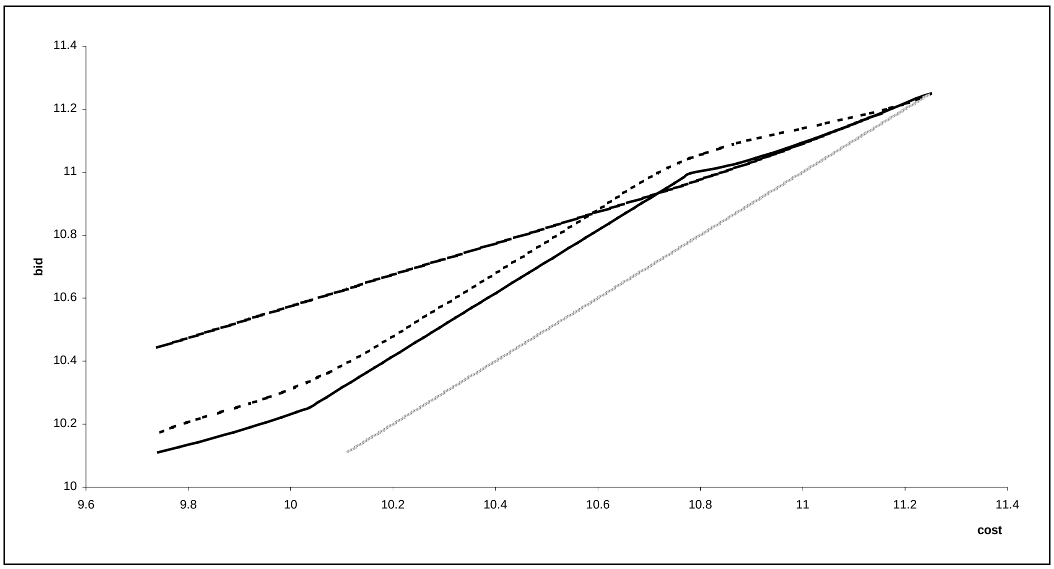


Solid line - true individual bid function; dashed line – 45º line
Dotted lines – 5 % and 95 % pointwise quantiles of the estimator distribution

Simulated data:

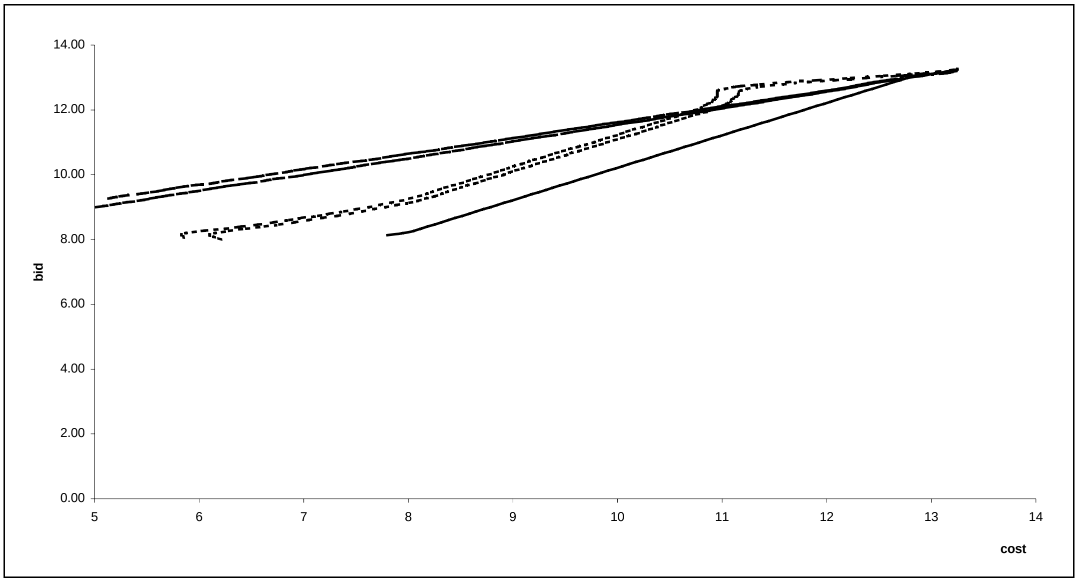
The individual cost component, X, is distributed on [-0.75, 0.25] according to the distribution function $H_X(x)=(x+0.75)^3$; the common cost component is uniformly distributed on [10.5-s/2, 10.5+s/2], s=0.5, 1, 5, 10; the total cost equals $c_{ij}=x_{ij}+y_j$, where y_j is a realization of common component, whereas x_{ij} is a realization of the individual component in the draw j; the simulated bids are calculated using formula for equilibrium bid function
$\{b_{ij}=y_j + x_{ij}+ ((1+4*x_{ij})-(1-(x_{ij}+0.75)^4))/(4*(1-(x_{ij}+0.75)^3))$, i=1,..,m j=1,..,n};
n=500, m=2, the experiment was repeated 500 times.

Figure 4b: Bid functions, s = 0.5



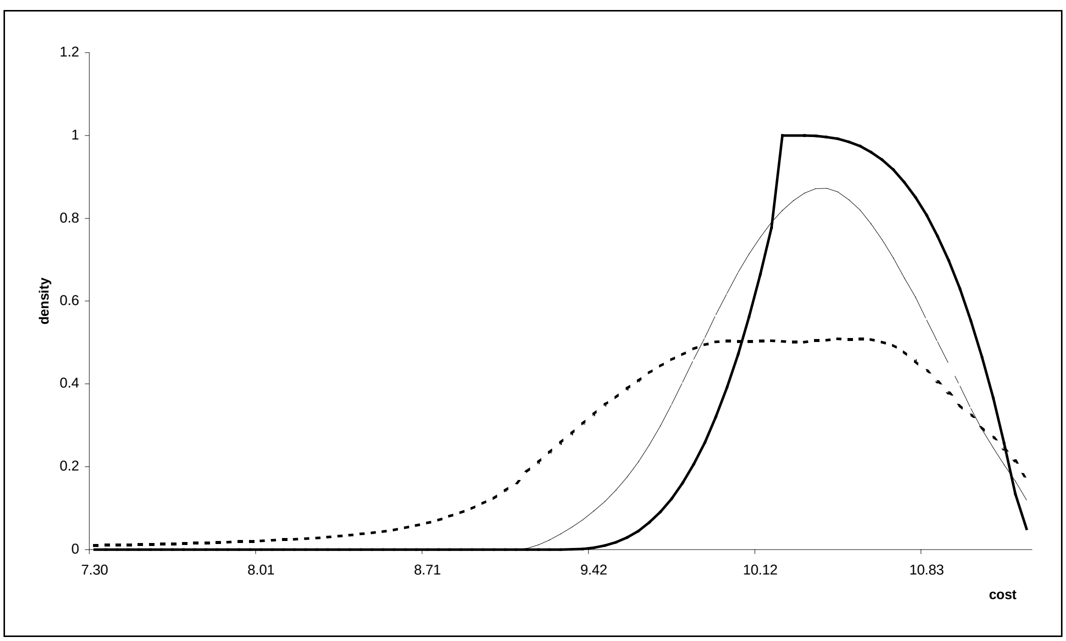
Solid line – true average bid function; fine dotted line – 45º line
Dotted line – bid function estimated under APV assumption (median)
Dashed line – bid function estimated under IPV assumption (median)

Figure 5b: Bid function, s = 5, (5% and 95% pointwise quantiles of estimators)



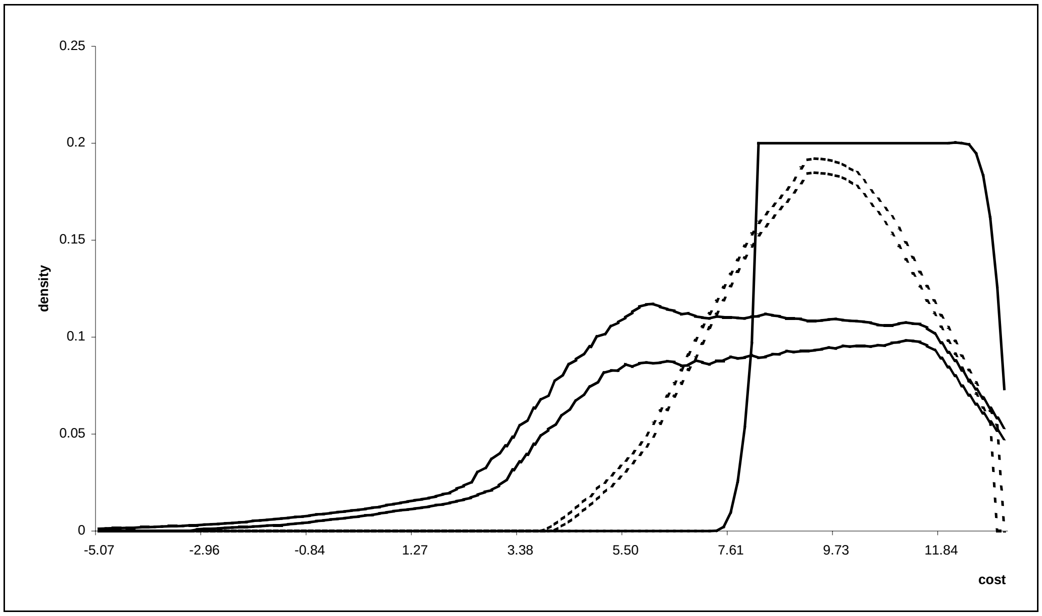
Solid line – true average bid function
Dotted line – bid function estimated under APV assumption
Dashed line – bid function estimated under IPV assumption

Figure 6b: Probability density function of total cost, s = 1



Solid line – true probability density function of total cost
Dotted line – probability density function estimated under IPV assumption
Dashed line – probability density function estimated under APV assumption

Figure 7b: Probability density function of total cost, s= 5,
(5% and 95% pointwise quantiles of estimators)



Solid line – true probability density function of total cost
Dotted line – probability density function estimated under APV assumption
Dashed line – probability density function estimated under IPV assumption

Table 1b: Average mark-ups over the bidders costs
(means of the simulated distributions of the estimators)

	true	UH estimator	IPV estimator	APV estimator
s=0.5	0.0136	0.0202	0.0615	0.0263
s=5	0.0150	0.0208	0.2782	0.1306
s=10	0.0151	0.0222	0.2817	0.2786

	true	UH estimator		IPV estimator		APV estimator	
		5%	95%	5%	95%	5%	95%
s=0.5	0.0136	0.0128	0.0224	0.06	0.0631	0.0259	0.0266
s=4	0.0150	0.0141	0.0215	0.2553	0.3063	0.1279	0.1329
s=10	0.0151	0.0150	0.0242	0.2566	0.3028	0.2730	0.2828

Table 2b: Moments of the estimated cost distribution[1]

	True			APV estimator			IPV estimator		
	mean	variance	skewness	mean	variance	skewness	mean	variance	skewness
s=0.5	10.5	0.058	-0.4475	10.38	0.125	-0.041	10.03	0.439	-5.085
s=5	10.5	2.861	-0.0012	9.48	2.641	-0.065	7.97	5.150	-0.689
s=10	10.5	8.337	-0.0003	8.37	10.16	-0.069	5.34	19.96	-0.614

	APV					
	mean		variance		skewness	
	5%	95%	5%	95%	5%	95%
s=0.5	9.91	11.23	0.112	0.139	-0.05	-0.03
s=5	9.05	9.92	2.489	2.764	-0.06	-0.07
s=10	7.93	8.81	8.87	11.03	-0.06	-0.075

	IPV					
	mean		variance		skewness	
	5%	95%	5%	95%	5%	95%
s=0.5	9.83	10.21	0.402	0.481	-4.85	-5.23
s=5	7.45	8.24	4.889	5.274	-0.599	-0.797
s=10	5.01	5.92	17.87	21.03	-0.506	-0.725

[1] The first table presents medians of relevant distributions.

Part C: Descriptive Analysis of the Data

Table 1c: Frequency distribution of the number of bidders

number of bidders	number of auctions	frequency
≤1	55	4.34
2	302	23.97
3	276	21.90
4	218	17.30
5	168	13.33
6	109	8.65
≥7	132	10.47

Table 2c: Probability distribution of the engineer's estimate

Percentile	Centile
1	50,020
5	100,264
25	263,102
50	536,417
75	1,291,964
95	4,597,571
99	10,220,000

Table 3c: Summary statistics of the data

$B_{(k)}$ is the k'th lowest bid submitted in the auction
Estimate refers to the engineer's estimate for a given project

variable	mean	standard error
$b_{(1)}$ / estimate	0.93	0.14
($b_{(1)}$ - estimate)/estimate	0.07	0.14
$b_{(2)}$-$b_{(1)}$	89,758	236,089.20
($b_{(2)}$-$b_{(1)}$)/estimate	0.081	0.033

Table 4c: Parameter estimates of OLS regression

The dependent variable is log of the bid normalized by engineer's estimate.
Log(estimate) refers to the log of engineer's estimate; Tcompletion is the time to complete the project; sign, marking, landscaping, drainage, curb are dummy variables, indicating whether a particular task is included in the project description; Nbidders1 – Nbidders6 are dummy variables corresponding to the number of auction participants. Two models are estimated: the set of independent variables of the first model includes the variables described above, in the second model a set of auction specific dummies is additionally included.

variable	**coefficient**	**standard error**
Log(estimate)	-0.018	0.004
Tcompletion	0.005	0.002
sign	0.084	0.149
marking	0.091	0.059
landscaping	0.035	0.019
drainage	0.105	0.049
curb	0.023	0.012
Nbidders1	-0.051	0.041
Nbidders2	-0.391	0.015
Nbidders3	0.022	0.012
Nbidders4	0.005	0.013
Nbidders5	-0.018	0.013
Nbidders6	0.007	0.013
Constant	0.146	0.050
R^2 (no dummies)	0.59	
R^2 (with dummies)	0.786	

Table 5c: Parameters of the linear model with auction-specific random effects

The estimated model is given by $\log(b_{ij}/est_j) = X_{ij}\beta + u_i + e_{ij}$, where index refers to a bidder, index j refers to an auction, X's are independent variables from the regression in Table 4, u_i denotes the auction-specific random effect. GLS is used in the estimation.

variables	**coefficient**	**standard error**
Log(estimate)	-0.017	0.006
Tcompletion	0.004	0.004
sign	0.067	0.159
marking	0.076	0.052
landscaping	0.036	0.020
drainage	0.895	0.040
curb	0.017	0.012
Nbidders1	-0.051	0.041
Nbidders2	-0.481	0.022
Nbidders3	0.011	0.020
Nbidders4	0.049	0.021
Nbidders5	-0.029	0.022
Nbidders6	0.007	0.022
Constant	0.164	0.080
R^2 (within)	0.008	
R^2 (between)	0.622	
R^2(overall)	0.501	
sigma u_j	0.116	
sigma e_{ij}	0.121	
ρ[2]	0.462	

[2] ρ is the fraction of variance attributed to the auction specific component.

Part D: Estimation Results[3]

Figure 1d: Probability density function of the common component

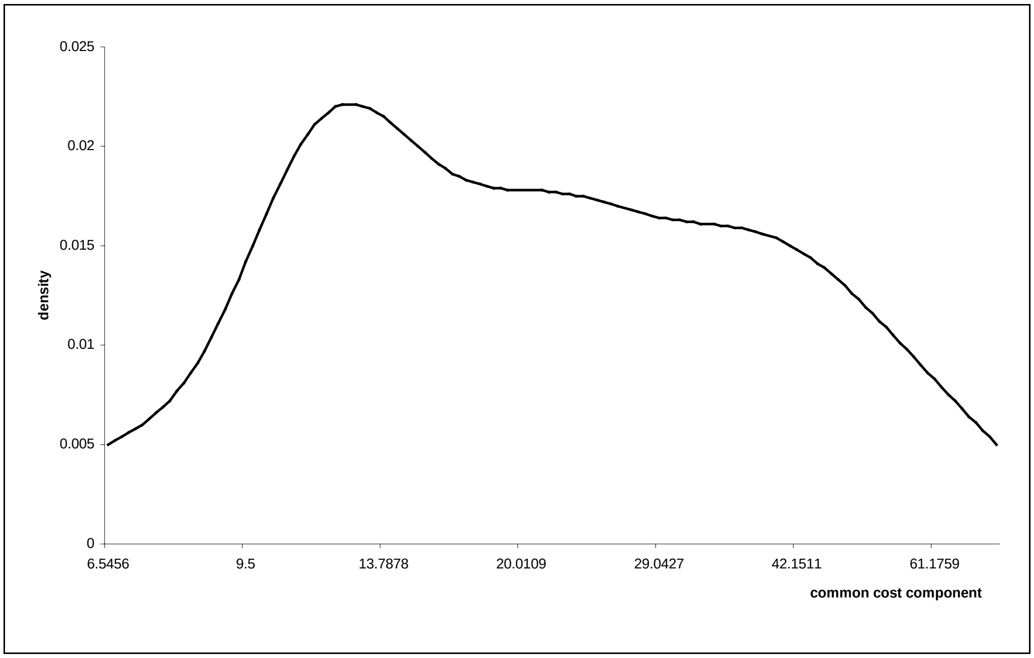
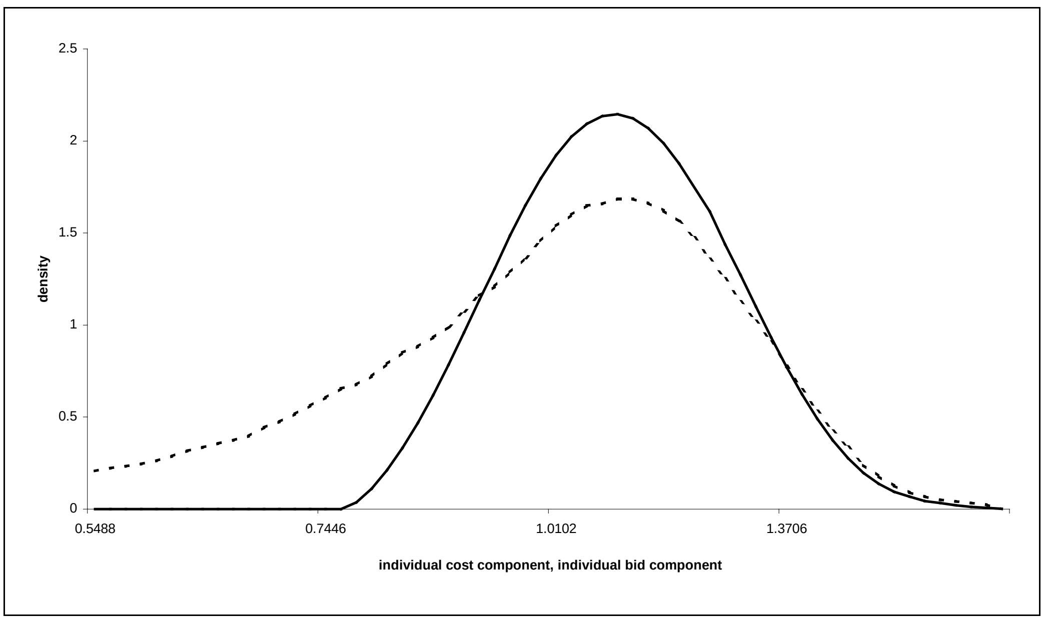


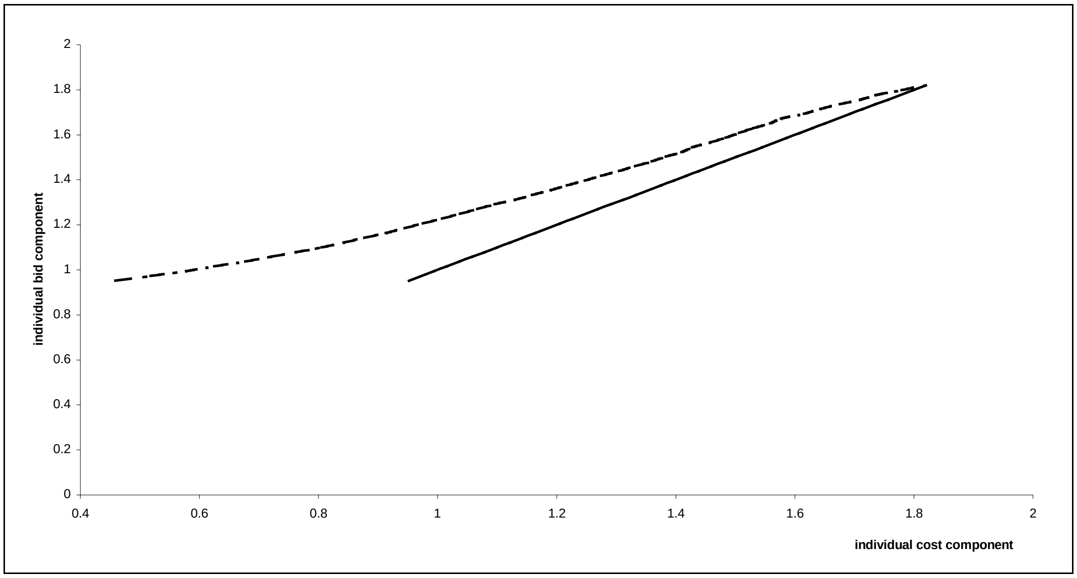
Figure 2d: Probability density functions of individual components



Solid line - probability density function of the individual bid component
Dotted line – probability density function of the individual cost component

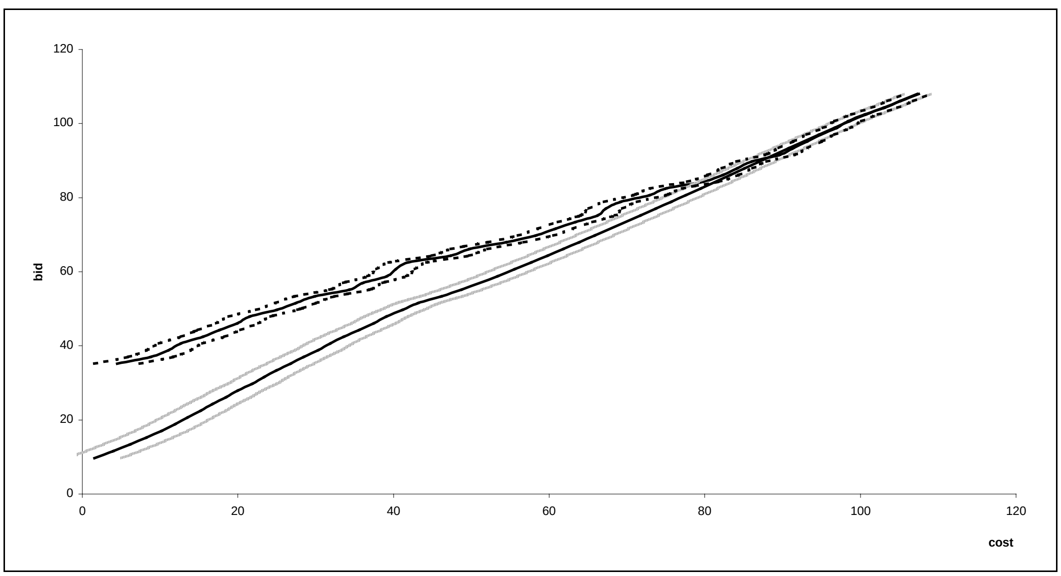
[3] The estimates are given for the set of auctions with the number of bidders equal to four.

Figure 3d: Individual bid function



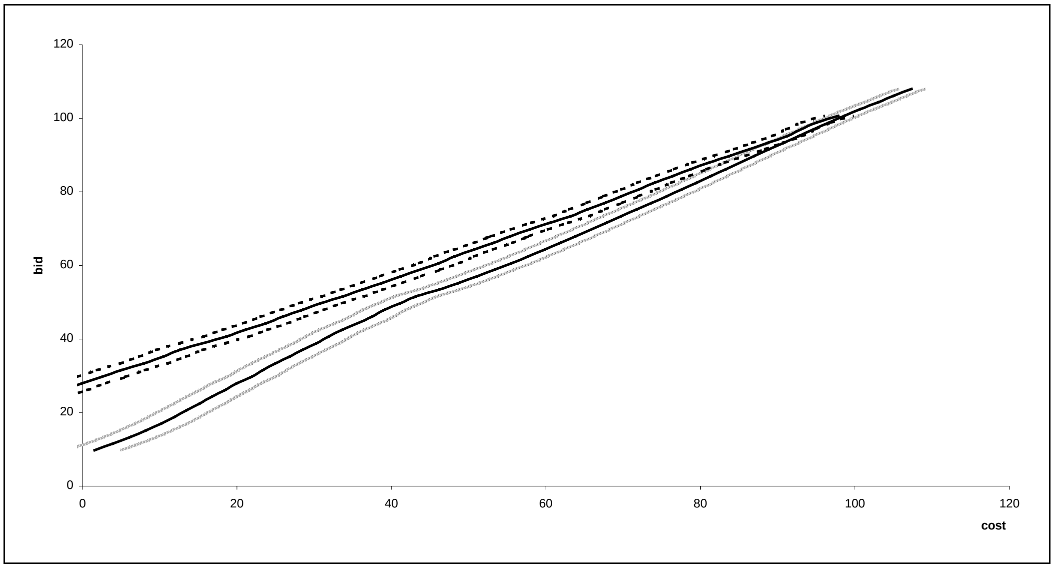
Dashed line – the estimated individual bid function
Solid line – 45º line

Figure 4d: Bid function



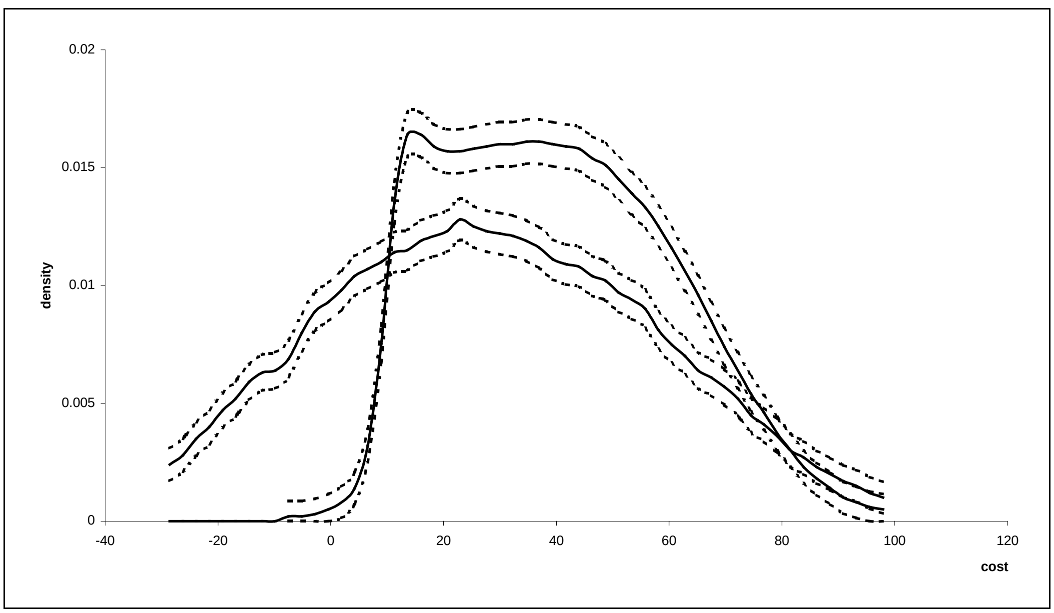
Solid line (first from below) – average bid function estimated under the assumption of unobserved heterogeneity; solid line (second from below) – bid function estimated under the APV assumption; dotted lines – 95% confidence intervals

Figure 5d: Bid function



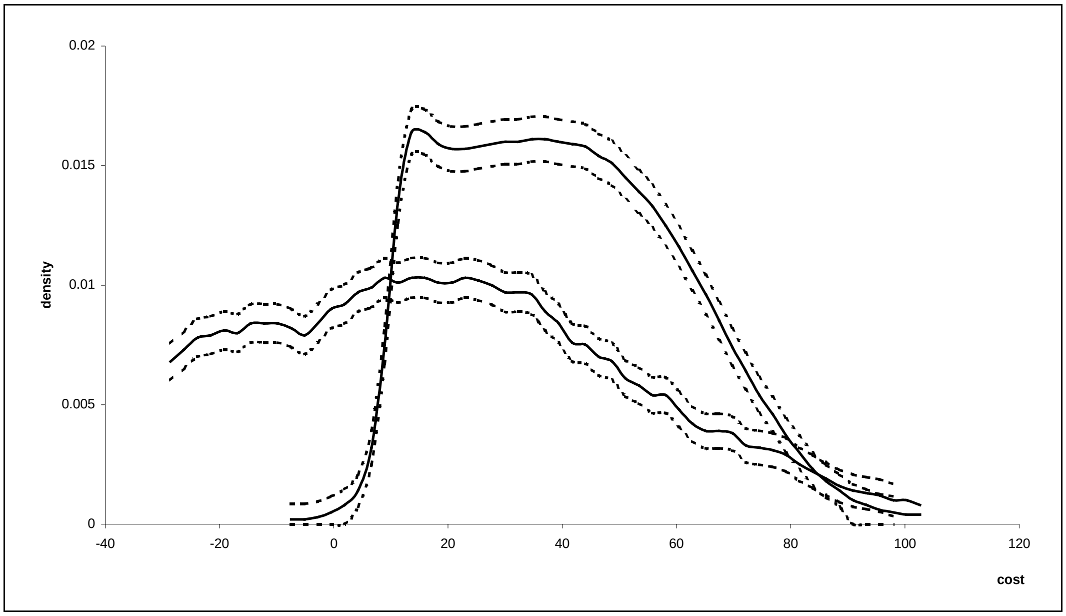
Solid line (first from below) – average bid function estimated under the assumption of unobserved heterogeneity; solid line (second from below) – bid function estimated under the IPV assumption; dotted lines – 95% confidence intervals

Figure 6d: Probability density functions of total cost



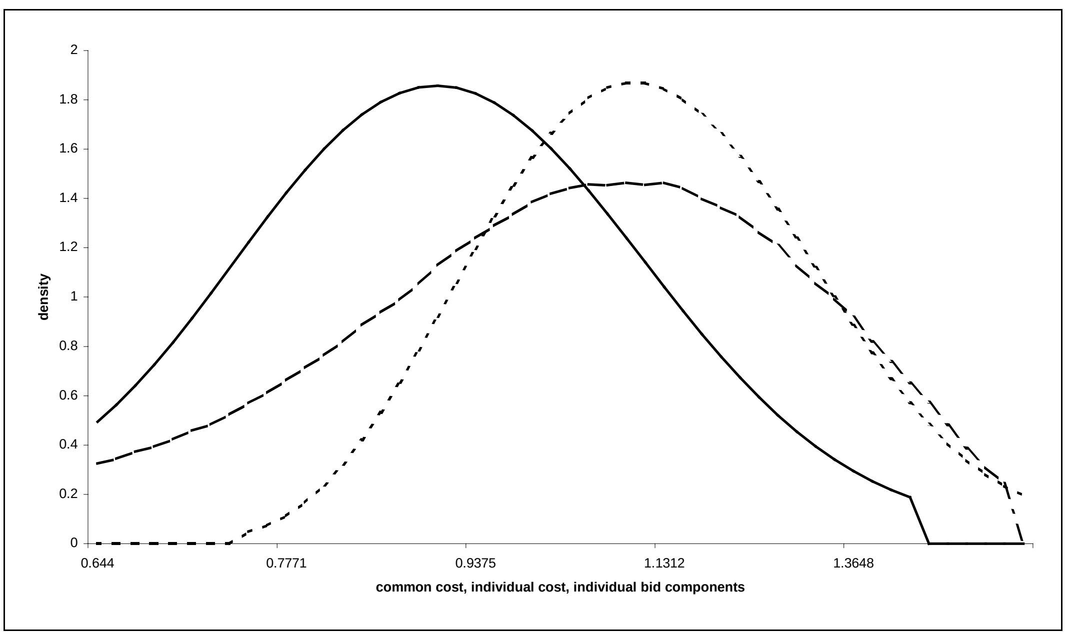
Solid line (first from above) – probability density function estimated under the UH assumption; solid line (second from above) – probability density function estimated under the APV assumption; dotted lines – 95% confidence intervals

Figure 7d: Probability density functions of total cost



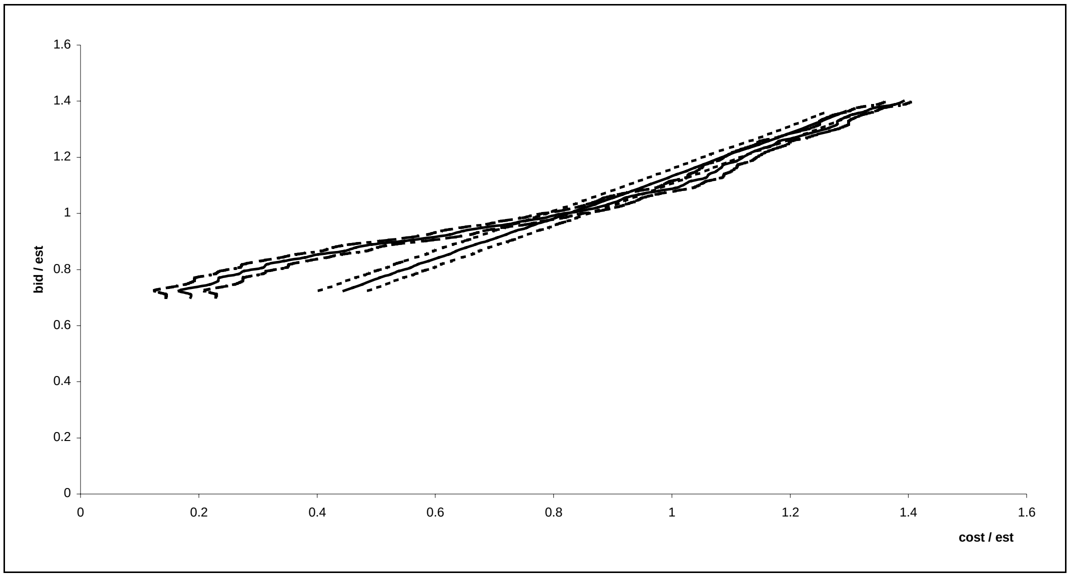
Solid line (first from the above) – probability density function estimated under the assumption of unobserved heterogeneity; solid line (second from above) – probability density function estimated under the IPV assumption; dotted lines – 95% confidence intervals

Figure 8d: Probability density functions of cost components



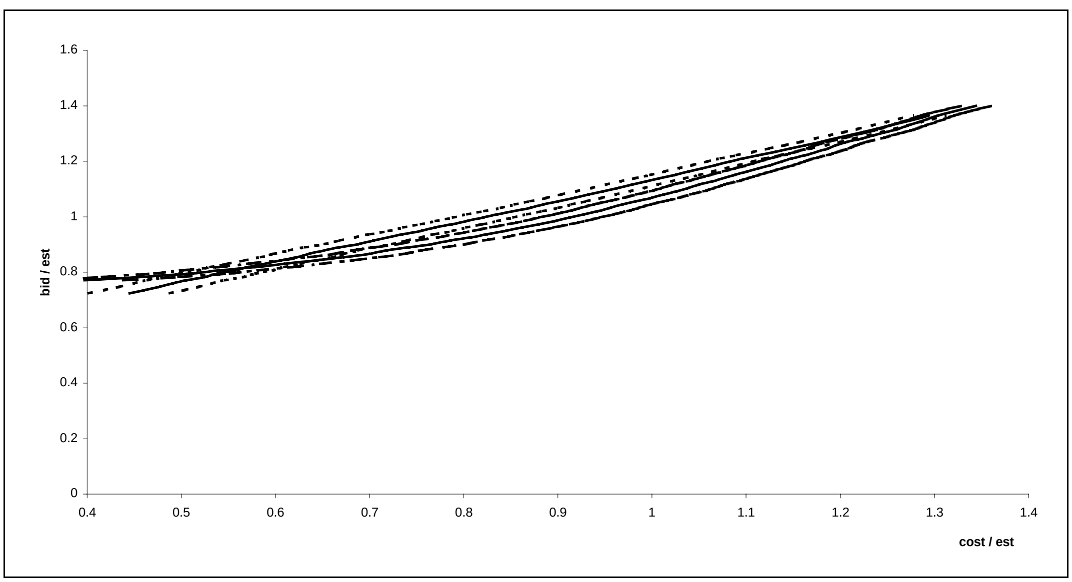
Solid line – probability density function of the common cost component
Dotted line - probability density function of the individual bid component
Dashed line - probability density function of the individual cost component

Figure 9d: Bid functions



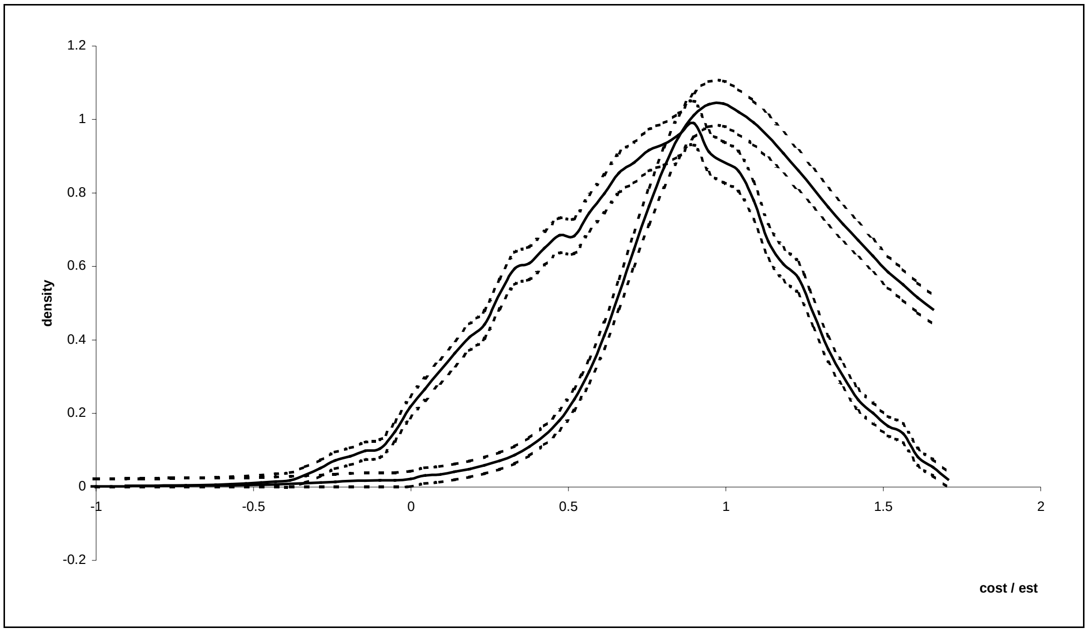
Solid line (steeper) – average bid function estimated under UH assumption; solid line (flatter) – bid function estimated under APV assumption; dotted lines – 95% confidence intervals

Figure 10d: Bid function



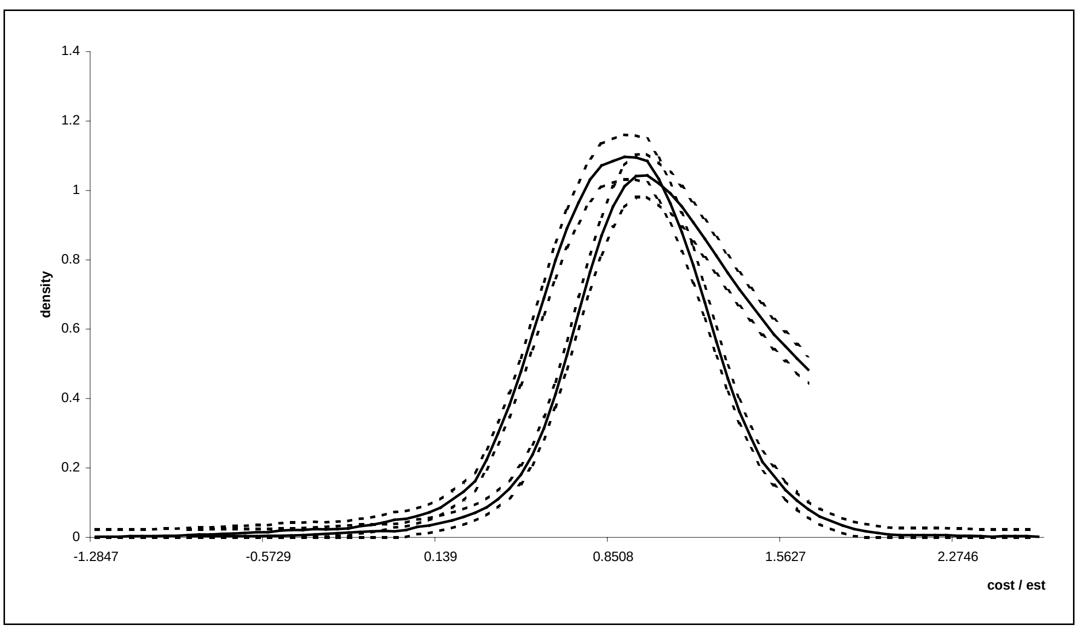
Solid line (steeper) – average bid function estimated under UH assumption; solid line (flatter) – bid function estimated under IPV assumption; dotted lines – 95% confidence intervals

Figure 11d: Probability density functions of total cost



Solid line (first from the left) – probability density function of total cost estimated under APV assumption; solid line (second from the left) – probability density function of total cost estimated under UH assumption; dotted lines – 95% confidence intervals

Figure 12d: Probability density functions of total cost



Solid line (first from the left) – probability density function of total cost estimated under IPV assumption; solid line (second from the left) – probability density function of total cost estimated under UH assumption; dotted lines – 95% confidence intervals