A NON-PARAMETRIC ESTIMATOR FOR RESERVE PRICES IN PROCUREMENT AUCTIONS

MARTIN BICHLER and JAYANT KALAGNANAM
IBM T. J. Watson Research Center
Yorktown Heights, NY 10598, USA
{bichler, jayant}@us.ibm.com

Abstract. Electronic auction markets collect large amounts of auction field data. This enables a structural estimation of the bid distributions and the possibility to derive optimal reserve prices. In this paper we propose a new approach to setting reserve prices. In contrast to traditional auction theory we use the buyer's risk statement for getting a winning bid as a key criterion to set an optimal reserve price. The reserve price for a given probability can then be derived from the distribution function of the observed drop-out bids. In order to get an accurate model of this function, we propose a nonparametric technique based on kernel distribution function estimators and the use of order statistics. We improve our estimatior by additional information, which can be observed about bidders and qualitative differences of goods in past auctions rounds (e.g. different delivery times). This makes the technique applicable to RFQs and multi-attribute auctions, with qualitatively differentiated offers.

Keywords: Reserve prices, auction theory, non-parametric estimation

1. Introduction

During the past few years, electronic reverse auctions have become a very popular economic institution for automating procurement negotiations. The competitive process of these auctions serves to aggregate the scattered information about supplier's costs and to dynamically set a price. The literature on *optimal auction design* tries to find the auction mechanism that provides the greatest revenue/profit for the seller, or the lowest cost for the buyer in a procurement auction, respectively. Such a question is of considerable practical value. In the well-studied *independent private value model* with risk neutral bidders any auction mechanism such as first-price, second-price, English, and Dutch auction, generates the same revenue for the seller. Therefore, the problem of optimal auction design reduces to determining an optimal reserve price [1]. The latter is expressed as a functional of the distribution of private values and its corresponding density function [2, 3].

With the advent of large-scale electronic auction markets, the access to large amounts of transaction data has become considerably easier. This enables a structural estimation of the empirical distributions relevant to setting an optimal reserve price, and has led to an increased interest in empirical applications of optimal auction theory (see section 5). This paper is motivated by our work on a large-scale electronic procurement platform for the retail industry. On this platform retail companies conduct repeated purchases of retail commodities using classic reverse auctions, mostly in an open-cry or English format. These commodities are purchased from an established set of suppliers for the respective commodities. Providing good decision support for setting appropriate reserve prices is an important feature for purchasing managers on this platform.

The results of optimal auction theory, however, have been criticized because they seem to be of theoretical rather than practical significance so far. "The 'optimal' auctions are usually quite complex, and there is no evidence for their use in practice." [4] As argued by McAfee and Vincent [5], a major difficulty in implementing the optimal reserve price is the use of unobservables such as the distribution of supplier's costs and the bid taker's own valuation for the good (i.e. her opportunity cost). First, there are methodological difficulties in estimating the latent supplier costs from the observed bids in the transaction data, which essentially assume knowledge of a bidder's strategy. This is also called the *identification problem*. The accuracy of this estimate, however, has a big impact on the reserve price. Second, also assessing the buyer's own opportunity cost can be non-trivial. Often, there is no alternative market value for the good. Additionally, a procurement auction is not necessarily a one-time event, and it is common procurement practice to repeat unsuccessful auctions with varied reserve prices. Where should the optimal reserve price be set, if there is a possibility to repeat the auction with a different reserve price?

In this paper we propose a new approach to setting reserve prices in such a procurement environment. Our contribution is two-fold: In contrast to traditional auction theory we use the buyer's risk statement for getting a winning bid as a key criterion to set an optimal reserve price. The reserve price for a given probability can then be derived from the distribution function of the observed drop-out bids. In order to get an accurate model of this function, we propose a nonparametric technique based on kernel distribution function estimators and the use of order statistics. We improve our estimatior by additional information, which can be observed about bidders and qualitative differences of goods in past auctions rounds (e.g. different delivery times). This makes the technique applicable to RFQs and multi-attribute auctions, with qualitatively differentiated offers.

The paper is structured as follows. The next section summarizes the relevant results of optimal auction theory. We will focus on two important aspects, namely asymmetry of bidders costs and correlation. We will then describe a univariate and a multivariate estimator for bid prices in section 3. Section 4 will present some first evidence from a Monte Carlo study. We then point to related literature in Section 5 and provide a brief summary and conclusions in section 6.

2. Theory of Optimal Auctions

There is a considerable academic literature on the effects of reserve prices in auctions. The basic theory has been developed by Vickrey [6] and extended by Riley and Samuelson [2], Levin and Smith [7] and Monderer [8]. Many laboratory experiments have tested different predictions of auction theory [9]. Empirical work using field data is summarized in Hendricks and Paarsch [10], or more recently in a so called "field experiment" by Lucking-Reiley [11]. Lucking-Reiley's analysis shows that implementing reserve prices (1) reduces the number of bidders, (2) increases the frequency with which goods go unsold, and (3) increases the revenues received on the goods conditional on their having been sold.

McAfee and McMillan state that the IPV model applies to (government) contract bidding when each bidder knows what his own production cost will be if he wins the contract [12]. Therefore, in the following section we will introduce the basic IPV model and derive the main results with respect to setting reserve prices. We will also discuss relevant factors such as the number of bidders, the asymmetry of bidders' costs, and correlation among the bidders' costs. Although relevant in many instances we will ignore other aspects such as collusion among bidders, their risk attitudes, or royalities for the sake of brevity.

2.1. Optimal Auctions in the Independent Private Values Model

The most thoroughly researched auction model is the symmetric independent private values (IPV) model. In this model applied to reverse auctions:

- A single indivisible object or task is put up for auction to one of several bidders.
- Each bidder i knows her true cost, $c_i \in \Re$, and can revise her signal when that of rival bidders are disclosed. If c_i is lower than the bid $b_i \in \Re$, then the bidder makes a profit of $b_i - c_i$.
- All bidders are symmetric/indistinguishable, i.e., the costs are drawn from a common distribution $G(\cdot)$ with support $[\underline{c}, c]^n$, which is known to all bidders.
- The unknown costs c_i are statistically independent, identically distributed, and continuous random variables.
- The bidders are risk neutral concerning their chance of winning the auction, and so is the seller.

The vector $(c_1, \ldots, c_n)$ is a realization of a random vector whose n-dimensional cumulative distribution function is $G(c)$. Denoting $c_{(i)}$ as the i^{th} smallest order statistic for a sample of size n from the distribution of c, the bidder of the auction will be the player with the lowest cost $c_{(1)}$. That is, $c_{(1)}$ is the first order statistic, and $c_{(2)}$ is the second order statistic [13]. In an English auction the second-last bidder will drop out of the bidding as soon as the price is below her own cost of the item. From the point of view of the winning bidder, her expected rent is the expected difference between $c_{(1)}$ and $c_{(2)}$, which is the difference between the first order statistic and second order statistic, given by $-G(c)/g(c)$, where c is the bidder's production cost, and G and g are the probability distribution function and density function of bidders' costs. Consequently, the *expected buyer's payment* is the winning supplier's cost plus the winning supplier's rent:

$$J(c_{(1)}) = c_{(1)} + \frac{G(c_{(1)})}{g(c_{(1)})} \quad (1)$$

Riley and Samuleson [2] reduce the optimal auction problem to the optimal choice of the reservation price. That is, an optimal auction requires the buyer to set a reserve price, r, above which she will not buy the item and make it public (i.e., a maximum bid). This price is set to mimic the expected bid of the second lowest bidder and is lower than the buyer's valuation, i.e. cost for not getting the good, c_0, namely,

$$r = J^{-1}(c_0) < c_0 \tag{2}$$

This reserve price minimizes the expected cost of the buyer, based on the distribution of costs in the market. For the IPV model, any of the English, Dutch, first-price sealed-bid, and second-price auctions is optimal, provided the reserve price is set optimally as in (2). Remarkably, this optimal reserve price is independent of the number of bidders n. This is a powerful result, as no restrictions have been placed on the types of policies the seller can use. For instance, the seller can have several rounds of bidding, or charge entry fees, or allow only a limited time for the submission of bids. None of these procedures would increase the expected price for the seller.

This optimal level of the reserve price is determined by a trade-off. The disadvantage of setting a reserve price is that it is possible for the remaining bidder to have a valuation that lies between the sellers valuation and the reserve price, $c_0 > c > r$. In this case, the buyer doesn't find a supplier even though the bidder would have been willing to charge a lower price than what the buyer was willing to pay. On the other hand, if the reserve price is below the second-lowest bidder's cost, the bidder charges less than she would have in absence of the reserve price. In summary, the buyer imposes the reserve price in order to capture some of the informational profits that would otherwise have gone to the winner. However, this can have a negative impact on the efficiency of the auction. Bulow and Roberts [14] point out the relationship of optimal auction theory to the theory of price differentiation in monopolies.

2.2. Number of Bidders and Distribution of Costs

The number of bidders and the particular type of cost distribution have a considerable impact on the outcome of an auction. Increasing the number of bidders increases the revenue of the bid taker on average. This is because the second-lowest cost approaches the lowest possible cost. In addition to the number of bidders, the variance of the distribution of costs has an impact. The larger the variance, the larger on average is the difference between the lowest cost and the second lowest cost. This means the average revenue of the buyer and the winning supplier are increased. As can be seen in (1) and (2) the distribution of bidder's costs as well as the buyer's cost for not getting the good, c_0, are the two parameters impacting the reserve price. In the following we will analyze the impact of these state variables for two common types of distributions, the uniform and the normal distribution. For a uniform distribution the reserve price, r, is

$$r = J^{-1}(c_0) = \frac{c_0 + lo}{2} \tag{4}$$

where lo is the lower bound of the uniform distribution. The reserve price calculation for the case of a normal distribution is more complicated. The distribution function of the normal distribution can be computed based on the error function, which can be integrated numerically, or approximated as in [15]. This leads to a J-function as in equation (5). Function J is continuous and monotonically increasing. Therefore, the inverse to function J^{-1} for the optimal reserve price can be found through bisection or other root finding methods.

$$J(r) = r + \frac{\int_{-\infty}^{r} e^{-\frac{1}{2}\left(\frac{x-m}{s}\right)^2} dx}{e^{-\frac{1}{2}\left(\frac{r-m}{s}\right)^2}} = r - e^{-\frac{1}{2}\left(\frac{r-m}{s}\right)^2} \sqrt{\frac{p}{2}}\, s \left(1 - Erf\left(\frac{m-r}{\sqrt{2}s}\right)\right) \qquad (5)$$

Figure 1 shows the shapes of J^{-1} for different levels of c_0 using three different distributions. The solid line illustrates the reserve price for a uniform distribution with a lower bound of 1. The dashed line shows a normal distribution with a μ of 50 and a σ of only 4, whereas the dotted line shows a normal distribution with a μ of 30 and a σ of 20. The intuition is that with only a low variance of cost, it does not make sense to raise the reserve price way beyond the mean of the normal distribution, even with a high opportunity cost c_0.

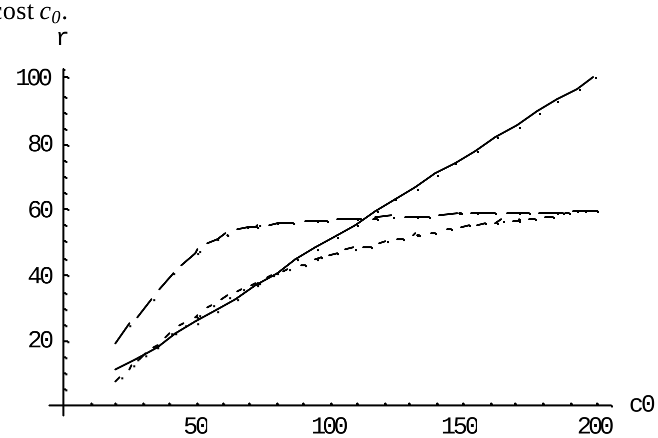

Figure 1: Shapes of J^{-1} functions in reverse auctions for different distributions

The figure illustrates the strong impact the buyer's opportunity cost c_0 on the choice of the optimal reserve price. A bias in c_0 leads to a considerable bias in r, which makes the application of this formula difficult in many real-world cases.

2.3. Asymmetry of Bidders

Although, IPV assumptions are similar to the conditions one can find in typical procurement negotiations, they need to be carefully evaluated before applying the model to auctions in the field. One IPV assumption, which is often violated in practice, is the symmetry of bidders. In many procurement situations bidders fall into recognizably different classes, i.e., bidders are asymmetric. It would therefore not be appropriate to represent all bidders as drawing their valuations from the same probability distribution F. An example might be suppliers from different countries, where there are systematic cost differences between domestic and foreign firms. Asymmetry of bidders leads to a breakdown of the well-known revenue equivalence theorem and therefore impacts the choice of the auction format. In absence of a reserve price, an English auction yields an efficient outcome, whereas a sealed-bid auction may yield an inefficient outcome.

In an optimal asymmetric auction, the seller sets a different reserve price for each type of bidder, computed as in (2). This means, that the optimal auction is *discriminatory* between the types. By setting k different reserve prices, each type has an incentive to bid for the contract. However, there is a possibility that one bidder wins despite another bidder's having a higher valuation. This is true because asymmetry implies that the probability distributions F are different, so that it is possible that the buyer's expected payment $J_k(c^k_{(1)}) > J_{k+1}(c^{k+1}_{(1)})$ even though $c^k_{(1)} < c^{k+1}_{(1)}$, with k being the different types of bidders. Because this optimal policy leaves a positive probability of the item being awarded to someone other than the bidder with the lowest cost, the policy is not Pareto efficient. If the distributions of valuations are identical except for their means, then the class of bidders with the higher average cost receives preferential treatment in the optimal auction. Because it might happen that the higher type bidder is below her reserve price, whereas the lower type bidder is not.

In general, the IPV assumes that the distributions are observable by all of the bidders, which might not be given in many procurement auctions. For example, in many private-sector sealed-bid auctions bidders do not even know who has been invited to bid. In addition, empirical distributions might be complex mixture distributions, which cannot be assumed known by the bidders.

2.4. Correlated Costs

Yet another assumption, which is sometimes violated in procurement negotiations is the independence of bidders' costs. Often the bidders estimates about the cost of executing a contract are somewhat correlated. Two other models have been discussed in the auction literature, namely the *common-value model* and the *affiliated-values model* [16]. The common-value model describes situations where the good has a single objective value to all bidders, and the bidders have different guesses about how much the item is objectively worth. This model is particularly apt to situations where goods have a resale value, such as securities, antiques or the amount of gold in a mine. While most procurement auctions, which can be observed in practice do not exhibit the characteristic of a single objective value of the good in question, it is likely that the valuation for a contract is "somewhat" dependent on the bids of other bidders. For example, if there is a common element of technological uncertainty (e.g., in long-term contracts), then the appropriate assumption is a degree of affiliation among the bidder's bids.

The affiliated-values model accounts for this influence. With n bidders, let $\mathbf{x}=\{x_1, \ldots, x_n\}$ represent the private signals about the item's value observed by bidders, $1 < i < n$; Let $\mathbf{s}=\{s_1, \ldots, s_m\}$ be a vector of variables that measure the quality of the item for sale. The bidder's valuation may depend not only upon his own signal, but also upon the other bidders' private signals and the true quality of the item, $v_i(\mathbf{s}, \mathbf{x})$. If variables are affiliated, then they are positively correlated. In other words, affiliation means that large values for some of the components make the other components more likely to be large than small. Although, affiliation plays a role in the auctioning of contracts with uncertain estimates about the actual costs incurred for the supplier, we will focus on the simpler IPV model in the following, assuming for example the purchase of direct or indirect materials, where the costs are known by the bidders and affiliation can be ignored. For a theory of reserve prices in an auction with affiliated values, see [7].

3. Estimation

As shown in equation (2), the optimal reserve price for an auction is determined by the latent distribution of costs in the auction and the buyer's cost for not getting the good, c_0. The price of the good on a secondary market, or the loss incurred through not getting the good in the subsequent production step, might serve as an estimate for c_0. Nevertheless, this variable is often difficult to set, in particular, since in many settings the procurement manager can initiate a second or third round of auctions, if the first round was unsuccessful. In contrast to traditional auction theory we use the buyer's risk statement for getting a winning bid as a key criterion for setting a reserve price. For example, a buyer wants to find the best reserve price given a probability Pr = 30% of getting a winner. The key technique to deriving such a reserve price is a good fitting estimation for the distribution of the winning bid. In particular, in situations with only a few bidders the accuracy of the estimate can have a big impact.

In the following, we will describe a non-parametric estimator for the probability distribution function of prices in a new auction. The technique is based on well-known kernel density estimators [17] and the theory of order statistics. We will first describe the one-dimensional case and then extend it to a multivariate case, which considers the impact of qualitative differences in the goods and services put up for auction. Due to their popularity in procurement, we will focus on English auctions, where the bidding strategy is simple and the drop out bids equal the true costs of the suppliers. We will in general make IPV assumptions, with the notable exception of symmetry in cases. A basic assumption in this section is that the bid prices in the transaction data do not contain any systematic seasonal or long-term trends.

3.1. Univariate Bid Price Estimators

In a first step, we will estimate the distribution of bids without considering qualitative differences in previous auction rounds. There are two approaches to estimating bid distributions. The *parametric approach* assumes a particular functional form with some unknown parameters for the valuation distributions. The *non-parametric approach* does not assume the valuation distributions to be part of a specified parametric family. Since little is known about the actual shape of the empirical bid distribution, kernel estimation has been chosen as a non-parametric approach to estimating the bid distribution. Besides the missing prior knowledge about the shape of the bid distribution, kernel estimation has a number of additional advantages over parametric estimations with respect to outliers or missing values [17].

A goal of density estimation is to approximate the probability density function (pdf) $g(\cdot)$ of a random variable C, which describes the bidders' cost or drop-out bid in case of an English auction. Assume we have n independent observations $c_1, \ldots, c_n$ from the random variable C. The kernel density estimator $\hat{g}_h(c)$ for the estimation of the density value $g(c)$ at point c is defined as

$$\hat{g}_h(c) = n^{-1} \sum_{i=1}^{n} K_h(c - C_i) \tag{6}$$

which is also called the Rosenblatt-Parzen kernel density estimator [18, 19] of C, where $K_h(u) = h^{-1}K(u/h)$ is the kernel with scale factor h. The shape of the kernel weights is determined by K, whereas the size of the weights is parametrized by h, which is called *bandwidth*. Commonly used kernel functions are the Epanechnikov kernel (7)

$$K(u) = 0.75(1-u^2)I(|u| \le 1) \tag{7}$$

which has a parabolic shape, or the Gaussian function (8) with its bell shape.

$$K(u) = \frac{1}{\sqrt{2p}}\exp\left(-\frac{1}{2}u^2\right) \tag{8}$$

It is easy to see that for estimating the density at point c, the relative frequency of all observations c_i falling in an interval around c is counted. The factor $1/(nh)$ in (9) is needed to ensure that the resulting density estimate has integral $\int \hat{g}_h(z)dz = 1$. For more detailed information on the choice of kernel functions and appropriate bandwidth, we refer to [17].

$$\hat{g}_h(c) = 1/(nh)\sum_{i=1}^{n} 0.75(1-(u/h)^2)I(|u| \le 1) \tag{9}$$

Instead of a kernel density estimator in (9), we propose a kernel distribution function estimator $\hat{G}_h(c)$ (KDF) for $G(c)$. For c 2 <, the KDF of G is given by

$$\hat{G}_h(c) = n^{-1}\sum_{i=1}^{n} \overline{K}_h(c - C_i) \tag{10}$$

where for u 2 <

$$\overline{K}_h(u) = \int_{]-\infty,u]} K_h(v)dv. \tag{11}$$

$\overline{K}_h(u) = \overline{K}(u/h)$, with $\overline{K}(u) = \int_{]-\infty,u]} K(v)dv$ being the probability distribution function. The KDF with an Epanechnikov kernel (9) leads to

$$\hat{G}_h(c) = \frac{1}{n}\sum_{i=1}^{n} 0.75\left(u - \frac{u^3}{3}\right)I(|u| < 1) \tag{12}$$

The probability $\hat{G}_h(c)$ in (12) is a continuously and monotonically increasing function of the bid price c. Using bisection or the Newton-Rhapson method one can find the reserve price to a particular probability of getting a winner. Following IPV assumptions, we suppose that the drop-out bids are n independent variates $C_1, C_2, \ldots, C_n$, each with the cdf $G(c)$, i.e. the bids are independent, identically distributed (iid). Therefore, we are particularly interested in the distribution function of the smallest order statistic $C_{(1)}$, which happens to be the lowest bid. The KDF can then be rewritten as $\mathrm{KDF}_{(1)}$.

$$\hat{G}_{(1)}(c) = 1 - [1 - \hat{G}_h(c)]^n \quad (13)$$

3.2. Multivariate Estimators

Field data from large-scale procurement auction platforms offer the possibility of getting transaction data from repeated auctions on the same good and service with essentially the same supplier pool. Nevertheless, these auctions are not completely homogeneous over time. For example, they might have differing qualitative attributes such as delivery time or bio degradability. In addition, buyers might invite different numbers of bidders to the auction. Multivariate estimators are a possibility to take this additional information into account.

The kernel density estimator can be generalized to the multivariate case in a straightforward way. Suppose we now have observations $\mathbf{c}_1, \ldots, \mathbf{c}_n$ where each of the observations is a d-dimensional vector $\mathbf{c}_i=(c_{i1}, \ldots, c_{id})^T$. The multivariate kernel density estimator at a point $\boldsymbol{c}$ is defined as

$$\hat{g}_h(\mathbf{c}) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_1 \ldots h_d}\mathrm{M}\left(\frac{c_{i1}-c_1}{h_1},\ldots,\frac{c_{id}-c_d}{h_d}\right) \quad (14)$$

with M denoting a multivariate kernel function, i.e. a function working on d-dimensional arguments. Note, that (14) assumes that the bandwidth $\mathbf{h}$ is a vector of bandwidths. A possibility for a multivariate kernel is the radial symmetric Epanechnikov kernel

$$M(\mathbf{u}) \propto (1-\mathbf{u}^T\mathbf{u})I(\mathbf{u}^T\mathbf{u} \le 1) \quad (13)$$

Radial symmetric kernels can be obtained from univariate by defining $M(\mathbf{u}) \propto K(\|\mathbf{u}\|)$, where $\|\mathbf{u}\| = \sqrt{\mathbf{u}^T\mathbf{u}}$ denotes the Euclidean norm of the vector $\mathbf{u}$. The $\propto$ indicates that the appropriate constant has to be multiplied. Radial symmetric kernels use observations from a ball around $\mathbf{c}$ to estimate the density at $\mathbf{c}$. For the bivariate case with one qualitative attribute and price, the function (9) with an Epanechnikov kernel can be rewritten as

$$\hat{g}_h(\mathbf{c}) = n^{-1}\sum_{i=1}^{n}\frac{1}{h_1 h_2}0.75\left(1-\left(\left(\frac{c_{i1}-c_1}{h_1}\right)^2+\left(\frac{c_{i2}-c_2}{h_2}\right)^2\right)^2\right) \quad (14)$$

Similar to (10) and (11), the integral of the kernel function leads to a multivariate kernel distribution function estimator (MKDF). Note that the arguments of the MKDF $\hat{G}_h(\mathbf{c})$, $(c_{ij}\text{-}c_i)$ are constrained between -1 and 1. Figure 1 illustrates the shape of a bivariate MKDF. For a given quality of a newly auctioned good and probability of getting a winner, the reserve price can be derived through slicing the landscape at the appropriate quality level. The resulting function needs to be normalized by dividing through its asymptotic probability value. The estimator for the first-order statistic $\text{MKDF}_{(1)}$ can then be derived as in (13). This procedure can also be used for multi-attribute bid data as can be found in RFQs and multi-attribute auctions.

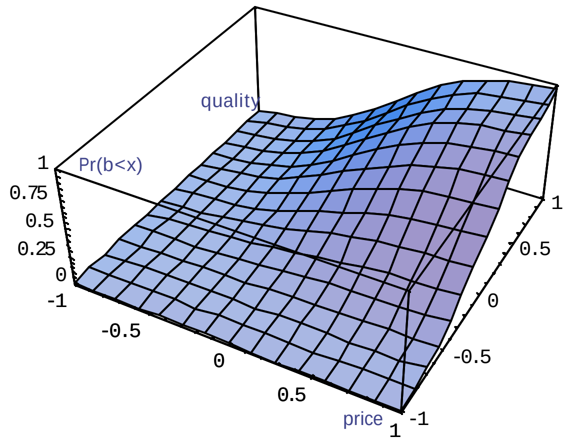

Figure 2: Shape of MKDF

4. Some Monte Carlo Evidence

In this section, we use Monte Carlo methods to compare different estimation techniques. We analyze the results of a naïve approach, which fits a Gaussian cdf to the data based on empirical sample moments, with the univariate and the multivariate kernel estimator described in the previous section. Since data are often scarce, we considered the effect upon the estimators of small to medium sized samples of 20 training auctions.

4.1. Experimental Design and Data Generating Process

In all of our simulation experiments, we assumed that the latent distribution of costs c follows a mixture distribution, similar to an example with asymmetric bidders. Allowing c to have a diffuse distribution also mimics some of the empirical evidence, which we have encountered in field data. We have conducted 20 training auction rounds with 4 bidders each, from which we estimate $\hat{G}_h(\mathbf{c})$. In a consecutive set of 20 test auction rounds, we have set a reserve price with a probability Pr = 30% for getting a winner, using four different estimators:

- no reserve price at all
- the naïve Gauss estimator
- the $KDF_{(1)}$
- the $MKDF_{(1)}$.

After the simulation we analyzed the total cost for the buyer, and the number of successful auctions. In the following sections we will describe the results of two treatments:

1. Treatment A: We assume to have auction data from completely identical auction rounds (exactly the same quality, number of bidders and identity of bidders).
2. Treatment B: We assume differentiated quality in the field auction data (e.g. different delivery times)

4.2. Discussion of Simulation Results

The first simulation with treatment A had 4 bidders with different costs. The bids in each auction round were drawn from normal distributions with a standard deviation of 4 and a mean of 20, 23, 38, and 109, respectively. After 20 training runs reserve prices were set for the two estimators at a probability of 30% for getting a winner.

- Naïve Gauss-based estimator: 28.1
- KDF: 21.6

We have evaluated the results for 20 additional auction rounds and achieved the following results:

No Reserve Price	
Total cost for 20 auction rounds without reserve price	430.20
Naïve Gauss-based estimate for Pr = 30%	0
Auctions with savings	0
Successful auctions	0
Total cost using the estimator	0
Total cost without using the estimator	0
$KDF_{(1)}$ for Pr = 30%	16.34
Auctions with savings	4
Successful auctions	5 (25%)
Total cost using the estimator	80.40
Total cost without using the estimator	99.41

Table 1: Results of Simulation with Treatment A and 4 Bidders

The Naïve Gauss estimator was too low and as a consequence all auctions were unsuccessful. For the $KDF_{(1)}$, only 5 or 25% of all the test auctions had a winner, and 4 of them had savings. Figure 3 shows both estimators as a function of the bid prices. $KDF_{(1)}$ 40 shows the $KDF_{(1)}$ after the 20 training auction rounds plus the 20 test auction rounds. The results illustrate that goodness-of-fit of the estimator is important for setting a reserve price at a certain level. With an increasing number of bidders, the difference between the first-order and the second-order statistic decreases and also the savings will are less significant.

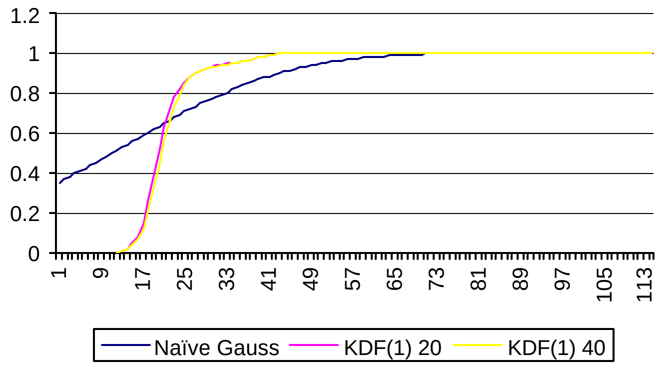

Figure 3: Estimators as a function of bid price

A second simulation used treatment B with differentiated qualities of the goods to be purchased. Throughout the 20 test auction rounds without reserve price, we introduced 3 different qualities L(ow), M(edium), and H(igh). Depending on the quality level demanded by the buyer, we introduced increased costs for quality M (+10) and H (+20) for all bidders. In the 20 subsequent auction rounds, the buyer purchased goods of quality L.

No Reserve Price	
Total cost for 20 auction rounds without reserve price	483.10
Naïve Gauss-based estimate for Pr = 30%	8.40
Auctions with savings	0
Successful auctions	0
Total cost using the estimator	0
Total cost without using the estimator	0
$KDF_{(1)}$ for Pr=30%	23.80
Auctions with savings	10
Successful auctions	19 (95%)
Total cost using the estimator	430.30
Total cost without using the estimator	456.90
$MKDF_{(1)}$ for Pr=30% and quality L	20.60
Auctions with savings	14
Successful auctions	17 (85%)
Total cost using the estimator	348.32
Total cost without using the estimator	404.30

Table 2: Results of Simulation with Treatment B and 4 Bidders

This time, we used three different estimators, the naïve Gauss estimator, the univariate, and the multivariate estimator for a probability of 30% for getting a winner. MKDF takes the different quality levels into account and provides an estimate for a particular quality in question (L), whereas the other estimators ignore these qualitative differences. The simulation illustrates that ignoring qualitative differences is penalized. The Gauss-based estimator is again too optimistic. With $KDF_{(1)}$ 95% of all auctions have a winner, and 10 achieve savings, whereas with the $MKDF_{(1)}$ for quality L 85% of the auctions have a winner and 14 achieve savings. Figure 4 shows all estimators as a function of the bid prices, including the three MKDF for different quality levels.

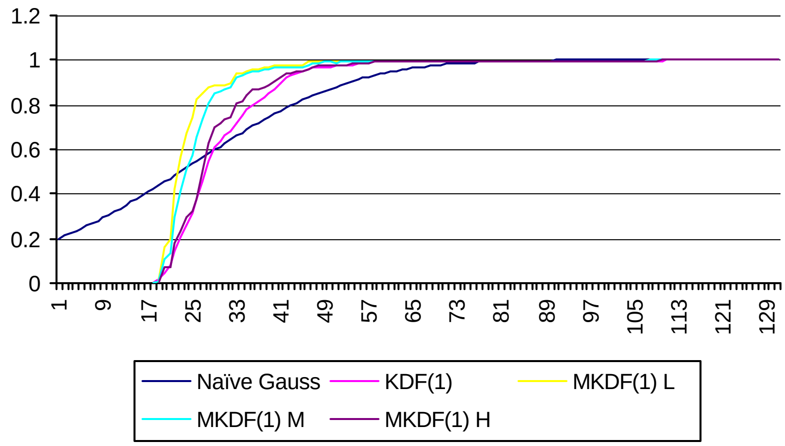

Figure 4: Estimators as a function of bid price

5. Related Literature

Recently, there have been a number of approaches in the econometrics literature focusing on the empirical analysis of auctions. Hendricks and Paarsch (1995) classify empirical work in auctions into two categories, *structural* and *non-structural/reduced-form* approaches. The *non-structural approach* tests necessary conditions of auction theory using *reduced form econometric models*. The reduced form provides a data admissible statistical representation of the economic system, whereas the structural form can be seen as a reformulation of the reduced form in order to impose a particular view suggested by economic theory [20]. An example for non-structural analysis is the detection of bid rigging in Porter and Zona [21]. A key question in the *structural analysis of auctions* is the estimation of the latent distributions that generate bidder valuations in the auction from observed bids. The strategy is to estimate the distribution of bids and then to retrieve the distribution of costs. An issue that all structural estimations have to address is the issue of *identification*, that is, the question of the extent to which the unobservable cost distributions can be recovered from the observed bid distributions. This approach relies upon the hypothesis that observed bids are the equilibrium bids of the auction model under consideration. In the first price sealed bid auction the focus lays on Bayesian-Nash bidding strategies.

Some of these structural estimation procedures are parametric and assume a particular functional form with some unknown parameters for the valuation distributions. Leading examples that analyze the first price sealed bid auction with private values include Donald and Paarsch [22], and Laffont, Ossard, and Vuong [23]. Donald and Paarsch [24] present methods, which consist of finding estimators maximizing the likelihood function for the symmetric IPV, where the valuation distributions are identical across bidders. Much of the literature is concerned with the identification problem under various conditions such as asymmetry [25], or affiliations among the valuations [26]. Some newer approaches also use non-parametric estimators, which do not make assumptions on the shape of the distribution function (see [27], [28] and [29]). Note, that also in non-parametric estimations, some assumptions are made, such as that the distributions are identical across bidders and continuous. In addition, structural estimations make assumptions about the type of equilibrium in an auction. In contrast to existing analysis, our approach relies on the bid taker's risk statement. Therefore, we do not necessarily need to know the latent valuations/costs of suppliers. More relevant is the goodness of fit of our estimator. For this reason we choose a non-parametric technique and take additional information into account, such as qualitative differences in the auctioned goods, the number, and identity of the bidders.

Another relevant stream of literature deals with multi-attribute auctions and RFQs [30]. Multi-attribute reverse auctions allow negotiation over price and qualitative attributes such as color, weight, or delivery time. A thorough analysis of the design of multi-attribute auctions has been provided by Che [31]. He derived a two-dimensional version of the revenue equivalence theorem. Che also designs an optimal scoring rule based on the assumption that the buyer knows the probability distribution of the supplier's cost parameter, and proves that using this scoring function is in fact an optimal mechanism. More recently, Beil and Wein [32] suggested an inverse-optimization based approach that allows the buyer via several changes in the announced scoring rule, to determine an optimal scoring rule. While elegant, these approaches assume a number of prerequisites (e.g., knowledge of the parametric shape of the supplier's cost functions), which are hardly given in practice. An alternative approach to increase the buyer's revenue/utility is to set reserve prices, based on the attribute values a supplier has specified. This is applicable to both, multi-attribute auctions and RFQs, which do not even use public scoring functions.

6. Summary and Conclusions

Game-theoretic auction theory is based on the assumption that the distribution of valuations, or costs respectively, is known among the participants in an auction. In many procurement auctions, this assumption is not given. Knowledge about empirical distribution enables a bid taker to fine-tune reserve prices. We have proposed a multivariate estimator, which is useful in estimating empirical bid distributions in auctions. The estimator allows us to set reserve prices, based on the risk statement provided by a buyer. We plan to incorporate the estimator into a tool, which takes new auction rounds into account and suggests reserve prices based on past auction rounds. Essentially, the user determines a description of the good in question, the invited suppliers and the probability for getting a winner, and the software suggests a reserve price.

There are multiple ways how we intend to improve the estimator and make it more applicable to different environments. In many settings, we will have past auction rounds with reserve prices. It is important in these settings to take into account the fact that certain bidders did not submit bids. Another issue is the consideration of underlying trends in the bid data, in particular, if the data is collected over a longer time period, where trends and seasonal deviations play a role. A general problem in multivariate prediction is also called the “curse of dimensionality”. The basic element of nonparametric smoothing – averaging over neighborhoods – will often be applied to a relatively meager set of points since even large samples are surprisingly sparsely distributed in the higher dimensional Euclidean space. We plan to investigate additive models such as the projection pursuit regression [33], or stochastic gradient boosting [34] for these cases. Setting reserve prices is of course not the only application of such an estimator. It can as well be used for bid pricing, in order to help a bidder in estimating the likelihood of winning, or to compare the attractiveness of different auction markets for buying or selling goods.

7. References

[1] E. Wolfstetter, "Auctions: An Introduction," *Journal of Economic Surveys*, vol. 10, pp. 367-420, 1996.

[2] J. G. Riley and J. G. Samuleson, "Optimal auctions," *American Economic Review*, vol. 71, pp. 381-392, 1981.

[3] R. B. Myerson, "Optimal auction design," *Mathematics of Operations Research*, vol. 6, pp. 58-73, 1981.

[4] M. H. Rothkopf and R. M. Harstad, "Modeling Competitive Bidding: A Critical Essay," *Management Science*, vol. 40, pp. 364-384, 1994.

[5] R. P. McAfee and D. Vincent, "Updating the Reserve Price in Common-Value Auctions," *American Economic Review*, vol. 82, pp. 512-518, 1992.

[6] W. Vickrey, "Counterspeculation, Auctions, and Competitive Sealed Tenders," *Journal of Finance*, pp. 8-37, 1961.

[7] D. Levin and J. L. Smith, "Optimal Reservation Prices in Auctions," *Economic Journal*, vol. 106, pp. 1271-1282, 1996.

[8] D. Monderer and M. Tennenholtz, "Optimal Auctions Revisited," presented at AAAI 98, Madison, Wisconsin, 1998.

[9] J. H. Kagel, "Auctions: A Survey of Experimental Research," in *The Handbook of Experimental Economics*, J. H. Kagel and A. E. Roth, Eds. Princeton: Princeton University Press, 1995, pp. 501-587.

[10] K. Hendricks and H. J. Paarsch, "A Survey of Recent Empirical Work Concerning Auctions," *Canadian Journal of Economics*, vol. 28, pp. 315-338, 1995.

[11] D. Lucking-Reiley, "Auctions on the Internet: What's Being Auctioned, and How?,", vol. 2000, 1999.

[12] R. McAfee and P. J. McMillan, "Auctions and Bidding," *Journal of Economic Literature*, vol. 25, pp. 699-738, 1987.

[13] S. M. Ross, *Probability Models*. San Diego, et al.: Harcourt Academic Press, 2000.

[14] J. I. Bulow and D. J. Roberts, "The Simple Economics of Optimal Auctions," *Journal of Political Economy*, vol. 97, 1989.

[15] M. Abramovitz and A. Stegun, *Handbook of Mathematical Functions*. Dover, 1964.

[16] P. R. Milgrom and R. J. Weber, "A Theory of Auctions and Competitive Bidding," *Econometrica*, vol. 50, pp. 1089-1122, 1982.

[17] W. Haerdle, *Smoothing Techniques with Implementation in S*. New York: Springer Verlag, 1991.
[18] E. Parzen, "On estimation of a probability density function and mode," *Ann. Math Statist*., vol. 33, pp. 1065-1076, 1962.
[19] M. Rosenblatt, "Remarks on some non-parametric estimates of a density function," *Ann. Math.Statist*., vol. 27, pp. 832-837, 1956.
[20] T. Haavelmo, "The Probability Approach in Econometrics," *Econometrica*, 1944.
[21] R. H. Porter and D. J. Zona, "Detection of Bid Rigging in Procurement Auctions," *Journal of Political Economy*, vol. 101, pp. 518-538, 1993.
[22] S. G. Donald and H. J. Paarsch, "Piecwise Pseudo-Maximum Likelihood Estimation in Empirical Models of Auctions," *International Economic Review*, vol. 34, pp. 121-148, 1993.
[23] J.-J. Laffont, H. Ossard, and Q. Vuong, "Econometrics of First-Price Auctions," *Econometrica*, vol. 63, pp. 953-980, 1995.
[24] S. G. Donald and H. J. Paarsch, "Maximum Likelihood Estimation when the Support of the Distribution depends upon Some or All of the Unknown Parameters," Department of Economics, University of Western Ontario, Ontario 1992.
[25] P. Bajari, "Econometrics of the First Price Auction with Asymmetric Bidders," Harvard University, Boston, MA, USA February 6 1998.
[26] T. Li, I. Perrigne, and Q. Vuong, "Structural Estimation of the Affiliated Private Value Auction Model," University of Southern California, San Diego, Working Paper 2000.
[27] E. Guerre, I. Perrigne, and Q. Vuong, "Nonparametric Estimation of First-Price Auctions," Economie et Sociologie Rurales, Toulouse 95-14D, 1995.
[28] B. Elyakime, J.-J. Laffont, P. Loisel, and Q. Vuong, "First-Price Sealed-Bid Auctions with Secret Reserve Prices," *Annales d'Economie et Statistiques*, vol. 34, pp. 115-141, 1994.
[29] P. A. Haile and E. Tamer, "Inference with an Incomplete Model of English Auctions," *Journal of Political Economy*, vol. forthcoming, 2002.
[30] M. Bichler, *The Future of eMarkets: Multi-Dimensional Market Mechanisms*. Cambridge, UK: Cambridge University Press, 2001.
[31] Y.-K. Che, "Design Competition through Multidimensional Auctions," *RAND Journal of Economics*, vol. 24, pp. 668-680, 1993.
[32] D. R. Beil and L. M. Wein, "An Inverse-Optimization-Based Auction Mechanism To Support a Multi-Attribute RFQ Process," MIT, Boston, MA, USA, Research Report December 2001.
[33] J. Friedman and W. Stuetzle, "Projection Pursuit Regression," *Journal of the American Statistical Association*, vol. 76, 1981.
[34] J. Friedman, "Stochastic Gradient Boosting," Stanford University, Stanford, Technical Report 1999.