Structural Analysis of Ascending Auctions: an Application to Wholesale Used-car Auctions[*]

Kyoo il Kim and Joonsuk Lee[†]

Department of Economics

UCLA

March 31, 2004

Abstract

With the availability of a rich data set, we will develop a new method of conducting structural analysis of ascending auctions under the simplest valuation paradigm, the symmetric IPV model. The data set allows us to adopt a nonparametric approach and make strong interpretation of observed bids, including losing bids, while making a few assumptions about bidding behavior of the model. Identification and estimation are based on the recent work by Song (2003) that we extend by using one more order statistics and developing better procedure to control auction heterogeneity. We then implement nonparametric tests by Athey and Haile (2002).

[*]This version is very preliminary and incomplete.

[†]e-mails: kyookim@ucla.edu and joonsuk@ucla.edu

1 Introduction

Auctions have become one of the most important research topics in both theoretical and empirical microeconomics. Over the last decade, economic analysis of auctions has received enormous attentions not only from game theorists and applied econometricians but also from practitioners in government and related industries due to the important success of spectrum auctions in many countries and the phenomena of eBay. Within empirical analysis of auctions, structural analysis has received growing attentions over the last decade since Paarsch (1992) first conducted a parametric test of common values and private values paradigms.

Structural analysis of auctions assume equilibrium behavior of bidders and then try to estimate the underlying distribution of bidders' valuations directly from observed bidding data. After recovering the underlying data generating process (DGP), researchers can study many interesting policy-relevant questions of auction design; for example, the optimal choice of auction formats, reserve prices, information revelation structures, and so on. Since Paarsch (1992)'s seminal work, there has been a burgeoning empirical literature in structural approaches developing new econometric methodologies to identify, estimate, and test using various data sets from real-world auctions.[1] [2]

Among the dominant formats for auctions, ascending auctions are most often used. They

---

[1]Paarsch (1992) adopted a parametric approach to distinguish between IPV (independent private values) and PCV (pure common values) models. See Hendricks and Paarsch (1995) and Laffont (1997) for surveys of the early empirical works.

[2]We use common values (CV) as a synonym for interdependent values and, to avoid confusion, we distinctively use pure common values (PCV) whenever we need to describe models with the same, unknown valuations for all bidders. Models of bidders' valuations can be classified according to informational structures. Valuations can be either from private values (PV) paradigm or from common values (CV) paradigm. Private values are the cases where a bidder's valuation depends only on her own private signal. Common values refer to all the other general cases. A special case of common values is PCV. Most of the theoretical and empirical analyses in the literature have been done assuming one of the two extreme cases, IPV and PCV, because of their simplicities.

also have interesting features that distinguish itself from other formats.[3] There are many different variants of ascending auctions in the real-world. However, there is only one dominating theoretical and empirical model of ascending auctions in the literature. More than twenty years ago, while presenting a quite general modeling framework for various auctions, Milgrom and Weber (1982) (hereafter MW) modeled ascending auctions as a button auction, an auction with full observability of bidders' actions and, most importantly, with the irrevocable exits assumption, an assumption that a bidder is not allowed to place a bid at any higher price once she drops out at a lower price. This assumption significantly restricts each bidder's strategy space and makes the auction game simple enough to analyze. MW must have put this assumption in their model because ascending auctions are not easy to model due to its dynamic features which allows bidders to update their information (and therefore valuations) continuously and to re-optimize themselves during an auction. After MW, the button auction assumption was widely accepted by almost all the following works on ascending auctions.[4]

However, in almost all the real-world ascending auctions, we do not really observe such irrevocable exits. Only recently, there was a nice attempt by Haile and Tamer (2003) (hereafter HT) to conduct empirical research of ascending auctions without specifying such details as irrevocable exits in the model. This project follows HT in our attempt to conduct structural analysis of ascending auctions.

In the empirical literature of auctions, while there are quite a few works on the first-price sealed-bid auctions, there are not as many works on the ascending auctions. One of the possible reasons for this scarcity is the discrepancies between the theoretical model of

---

[3]They are Ascending (or English), Descending (or Dutch), First-price Sealed-bid and Second-price Sealed-bid (or Vickrey) auctions. To avoid confusion, we use the term ascending auctions for the wide variety of English auctions.

[4]There are few exceptions, e.g. Harstad and Rothkopf (2000) and Izmalkov (2003) in theoretical literature and Haile and Tamer (2003) in empirical literature. Also see Bikhchandani and Riley (1991, 1993) for extensive discussions of modeling ascending auctions.

ascending auctions, especially the button auction, and the way real-world ascending auctions, where the data comes from, are conducted. Another important reason preventing empirical analyses is the difficulty of getting a rich and complete data set that records ascending auctions.

This project contributes to the empirical auctions literature as follows. First, we develop a new method of conducting structural analysis of ascending auctions with the availability of a rich data set. In this first stage of our project, we will assume the simplest valuation paradigm, i.e. the symmetric IPV model. However, we will follow the incomplete modeling technique of HT (2003) removing the button auction assumption. As HT (2003) did in their paper, we will also adopt a nonparametric approach while developing necessary empirical and econometric methodologies. Our main difference from HT (2003) is that we are able to make stronger interpretation of observed bids with a few reasonable assumptions on the bidding behavior because of the availability of the rich data set.

Our second contribution will be the implementation and the evaluation of existing methods, especially Athey and Haile (2002) (hereafter AH), with our rich data set. AH (2002) provided quite general nonparametric identification results and proposed some nonparametric tests for common values for all four standard auction formats, however, they did not provide any exact statistics or empirical results. After we successfully conduct these two analyses, third and the ultimate goal of this research project will be the development of a nonparametric test to distinguish between common values and private values paradigms in ascending auctions.

2   Literature Review

In their nonparametric analysis of ascending auctions with IPV, Haile and Tamer (2003) adopted an incomplete model approach relaxing the button auction assumption and imposing only two assumptions on bidding behavior. The first assumption of HT is that bidders do

not bid more than they are willing to pay. And their second assumption is that bidders do not allow an opponent to win at a price they are willing to beat. In most auctions, these two assumptions seem very reasonable and innocuous. In their analysis, the independence assumption is crucial and they tried to relax this and extend to a model with affiliated private values (APV) in Haile and Tamer (2001), but it seems that they have not made much progress yet in that direction.

With the two assumptions and some known statistical properties of order statistics, HT (2003) nonparametrically estimated upper and lower bounds of the underlying distribution function. The reason they could only estimated bounds and could not make any exact interpretation of losing bids are because they did not impose the button auction assumption and allow quite free forms of ascending auctions. AH (2002) noted this in a footnote saying that "*In oral 'open outcry' auctions we may lack confidence in the interpretation of losing bids below the transaction price even when they are observed.*" Actually, within ascending auctions, there are a few variants that differ from each other slightly in the exact way of conducting auctions. Among those variants, the distinction between one in which bidders call prices and another one in which an auctioneer raises prices has very important theoretical and empirical implications. The former is what AH called as "open outcry" auctions and the latter is what we are going to exploit with our data. The main difference is that the former allows jumps in prices but the latter does not allow those jumps.

Our idea is that assuming IPV and a fixed discrete increments of prices raised by an auctioneer and also assuming similar axiomatic assumptions made by HT, we are able to make strong, exact interpretations of all the losing bids above the reserve price. That means we are able to treat the observed bids as if they come from a button auction without actually imposing the button auction assumption. And then we may use these information from observed bids to estimate the exact underlying distribution.[5] We can show that within the

---

[5] Actually, HT noted that if the true underlying model is the button auction, then their two bounds collapsed to a single distribution, which is also the exact estimate.

private values paradigm, in ascending auctions without irrevocable exits, the last price at which each bidder shows her willingness to win, i.e. each bidder's final exit price, can be directly interpreted as her private signal, and with symmetry, to relevant order statistics because it is weakly dominant strategy for a bidder to place a bid at the highest price she can afford.[6] [7] [8] Basically we may get information about all the order statistics of signals except the highest one.[9] Then, using the properties of the order statistics, we can identify the distribution of valuations. While doing this, we will utilize a recent development by Song (2003) which enables us to identify and estimate the underlying distribution without requiring the information on the exact number of potential bidders in each auction.

Since our approach is nonparametric, it can be also compared to the work by Guerre, Perrigne, and Vuong (2000) (GPV) and Li, Perrigne, and Vuong (2000, 2002) (LPV) for the first-price seal-bid auctions (FPSB). Within the IPV framework, GPV conducted a complete analysis of nonparametric estimation of FPSB auctions. They developed the two-step approach. We may have used their first-step in our estimation, however, the problem is we do not observe the highest order statistic in ascending auctions.

Regarding the test of common values and private values, it is well known in the literature that it is empirically difficult to distinguish between private value and common value models with actual auction data. However, there have been a few recent attempts to develop these tests. Hong and Shum (2003) estimated and tested general private values and common values models with a certain parametric modeling assumption using a quantile esti-

---

[6] In CV model, this is not the case and the analysis is much more complicated because a bidder may try not to press her button unless it is necessary because of strategic consideration to conceal her information. See Riley (1988) and Bikhchandani and Riley (1991, 1993).

[7] We ignore any possible cost associated with each bidding action and assume it is zero or negligible. This assumption seems reasonable for our auction where bidding is just pressing a button.

[8] Bikhchandani, Haile, and Riley (2002) showed there are generally multiple equilibria even with symmetry in ascending auctions so that we have to be careful interpreting observed bidding data. However, they also showed that with PV and weak dominance, we have uniqueness.

[9] See Arnold et al. (1992) and David (1981) for extensive statistical treatments on order statistics.

mation method in ascending auctions. Hendricks, Pinkse, and Porter (2003) developed a test based on data of the winning bids and the ex post values in first-price sealed-bid auctions. Athey and Levin (2001) also used an ex post data to test the existence of common values in first-price sealed-bid auctions. Most recently, Haile, Hong, and Shum (2003) developed a nonparametric test for common value components in first-price sealed-bid auctions using GPV's two-stage estimation method.

3   The Data

The auction data comes from an offline auction house located in Suwon, Korea.[10] It opened in May 2000 and has held wholesale used-car auctions weekly ever since.[11] The auction house mainly plays a role of an intermediary as many other auction houses do. While sellers can be anyone who wants to sell her own car through the auction, only used-car dealers who register as a member of the auction house can participate in the auctions. At the beginning, the number of total members was around 250, and now it has grown to about 350. Actually, this set of members can be viewed as a relatively stable panel, which makes the data from this auction more reliable to conduct meaningful analyses than those from online auctions in general.

Roughly, about a half of the members come to the auction each week. 600-1000 cars are auctioned on a single auction day and 40-50 per cent of those cars are actually sold through the auctions, which implies that a typical dealer who comes to the auction house on an auction day gets 2-4 cars/week on average. Bidders, i.e. dealers, in the auction have resale markets and, therefore, we can view this as if they try to win these used-cars auctioned only

---

[10]In June 2002, they started an online version of their offline auctions. Even though this new format may provide another interesting research agenda, for now we will focus on the offline auctions only. Suwon is located within an hour drive south of Seoul, the capital of Korea.

[11]The entire data set consists of all the auctions from a period between May 2000 and December 2002.

to resell them to the final consumers or to other dealers.

The auction house's objective should be long-term profit maximization. Since there exist repeated relationships between dealers and the house, it might be important for the house to build certain reputations.[12] An original seller's goal is to sell her car at the highest possible price near the time she wants to sell it after considering the trade-off between the price and the possibility of being sold.

The auction itself is an interesting variant of ascending auctions. There is a reserve price set by a seller with a consultation from the auction house. An auction starts from an opening bid below the reserve price. The opening bid is made public on the auction house's website two or three days before an auction day. After an auction starts, the current price of the auction increases by a fixed increment, which is about 30 US dollars for all price ranges, when there is any one bidder who presses a button beneath her desk in the auction room.[13] There is a big screen in front which displays pictures and key descriptions of a car auctioned currently as well as the current price.

In front of the auction room, there are also two important devices for information disclosure. One of them resembles a traffic light with green, yellow, and red lights, and the other is a sign that turns on whenever the current price is above the reserve price, which means that the reserve price is made public once the current price reaches the level.

The traffic lights indicates the number of bids at the current price. The green light indicates three or more bids, yellow means two bids, and red indicates that there is only one bid at the current price while the current price is above the reserve price. When the current price is below the reserve price, they are indicating two or more, one, and zero bids respectively. This traffic light is needed because unlike the usual open ascending auctions

---

[12]There also exists a competition among three similar wholesale used-car auction houses. In this paper, we ignore any effects from the competition and view the auction house as a single monopolist for simplicity.

[13]When the current price is above the reserve price, of course, there should be at least two bidders to continue. This auction might be one of the closest real-world application of MW's button auctions except the auctions they created.

the bidders in this auction do not see who are pressing their buttons and therefore do not know how many bidders are placing their bids at the current price. With the traffic light, bidders only get somewhat incomplete information on the number of current bidders and they never observe the identities of the current bidders.

There is a very short length of a single time period, eighty milliseconds, such that all the bids made in the same period are considered as made at the same price. A bidder can indicate her willingness to buy at the current price by pressing her button at any time she wants, i.e. exit and reentry are freely allowed. The auction ends when three seconds have passed after there is only one bid remains at the current price and no more bids at the time. When an auction ends at the price above the reserve price, the item goes to the bidder who presses last, but when it ends at the price below the reserve price, the item is not sold.

Available data includes the detailed bid-level (button-pressing) log data for every auctions.[14] Auction covariates, very detailed characteristics of cars auctioned, are also available. The covariates available includes each car's make, model, production date, engine-size, mileage, rating[15] , transmission-type, fuel-type, color, options, purpose, body-type etc. We also observe the starting prices and the reserve prices of all auctions. Some bidder-specific covariates such as identities, locations, 'members since' date are also available. And the date of title change is available for each car sold, which may be used as a proxy for the resale date. Last, we only observe the information on 'who' come to the auction house at 'what time' of an auction day for a very rough estimate on the potential set of bidders for an auction.

Here is a descriptive snapshot of a typical auction day, September 4th, 2002, which is randomly picked. Total 567 cars were auctioned on the day, 386 cars of which were passenger cars and the remaining 181 were full-size vans, trucks, buses, etc. Since this auction day

---

[14]Some logs are incomplete because only last fifty scans (the server records one to several "scans" at each price.) in each auction are recorded. Also, we do not have any data on the speed of price increase although it may represent the intensity of competition and therefore will affect bidders' valuations and their strategies.

[15]The auction house inspects each car and gives a 10-0 scaled rating to each car. These ratings are important to determine the value of a used-car.

was the first week of the month, there were relatively small number of cars[16]. 248 cars (43 percent) were sold through the auction and, among those unsold, at least 72 cars sold afterwards through post-auction bargaining, or re-auctioning next week, etc. The log data shows that the first auction of the day started at 13:19 PM and the last auction of the day ended at 16:19 PM. It only took 19 seconds per auction and 43 seconds per transaction. 152 ID cards (132 dealers since some dealers have multiple IDs) were recorded as entered the house. On average each ID placed bids for 7.82 auctions during the day. There were 98 bidders who won at least one car but 40 bidders did not win a single car. On average, each bidder won 1.8 cars and there are three bidders who won more than 10 cars. Among 386 passenger cars, there were at least one bid in 218 auctions. Among those 218 auctions, 170 cars were successfully sold through auctions and 48 were unsold.

The data we are using in this paper are from those auctions of fourteen weeks from September to December 2002. We only considered passenger cars for controlling heterogeneity and there were 5965 passenger cars auctioned in this period. Among those, we only use the data from auctions which have at least four observed bids above the reserve price and there are 717 of those. Here are summary statistics of the sample.

Table 1. Summary Statistics (Sample Size:717)

	Age (years)	Engine Size (ℓ)	Mileage (km)	Rating (0-10)	2nd highest	3rd -	4th -	Reserve price	Opening bid
Mean	6.1219	1.7369	98,225	3.2671	397.09	391.54	384.23	355.97	324.30
S.D.	2.3995	0.3679	48,991	1.0761	313.36	313.03	312.23	503.10	299.47
Med	6.4167	1.4980	94,060	3.5	321	316	308	285	250
Max	11.4167	3.2060	326,398	6	1,933	1,930	1,921	1,900	1,870
Min	0.5	1.4390	215	0	10	7	7	5	0

The market shares of major car makers in the sample is presented in Table 2.

---

[16]The number of cars auctioned are the most in the last auctions of months.

	Hyundai	Daewoo	Kia	Ssangyong	Others
Market Share(%)	45.19	30.26	20.92	2.65	0.98

Table 2. Market Shares in the Sample

4 Model and Estimation

4.1 Model and Identification

This section describes the basic set-up of an IPV model we analyze. Consider a wholesale used-car auction[WUCA] of a single object with the number of risk-neutral potential bidders, $N \geq 2$, drawn from $p_n = Pr(N = n)$. Each potential bidder i has the valuation V^i, which is independently drawn from the absolutely continuos distribution $F(\cdot)$ with support $[\underline{v}, \bar{v}]$. Each bidder knows only his valuation but the distribution $F(\cdot)$ and the distribution p_n are common knowledge. By the design of WUCA, we can treat it as a button auction if we disregard the minimum increment. Actually the minimum increment (about 30 dollars) in WUCA is negligible relative to the average car value (around 3,000 dollars) sold in WUCA, which is about one percent of the average car value.

Hence, in what follows, we simply disregard the existence of the minimum increment in WUCA to make our discussion simple and the bounds estimation implied by the minimum increment is handled in Section 4.6. Therefore, if we observe the number of potential bidders and any i^{th} order statistic of the valuation (identical to i^{th} order statistic of the bids), then we can identify the distribution of valuations from the cumulative density function (CDF) of the ith order statistic as done in many previous literatures. Define the CDF as

$$G^{(i:n)}(x) = H(F(x); i : n) = \frac{n!}{(i-1)!(n-i)!} \int_0^{F(x)} t^{i-1}(1-t)^{n-i} dt \qquad (1)$$

Then, we obtain the distribution of the valuations $F(\cdot)$ from

$$F(x) = H^{-1}(G^{(i:n)}(x); i : n) \qquad (2)$$

11

However, in the auction we consider, we do not know the exact number of potential bidders in a given auction and the number of potential bidders vary over different auctions. Nonetheless we can still identify the distribution of valuations $F(\cdot)$ following the methodology proposed by Song (2003), since we observe several order statistics in a given auction. Song (2003) showed that an arbitrary absolutely continuous distribution $F(\cdot)$ is nonparametrically identified from observations of any pair of order statistics from an iid sample, even when the sample size, n, is unknown and stochastic. The idea is that we can reinterpret the density of the k_1^{th} highest value Y conditional on the k_2^{th} highest value X as the density of the $(k_2 - k_1)^{th}$ order statistic from a sample of $(k_2 - 1)$ following $F(\cdot)$. In other words, the density of Y conditional on X, $p_{(k_2,k_1)}(y|X=x)$ can be written

$$
\begin{aligned}
p_{(k_2,k_1)}(y|x) &= \frac{(k_2-1)!}{(k_2-k_1-1)!(k_1-1)!}\frac{(F(y)-F(x))^{k_2-k_1-1}(1-F(y))^{k_1-1}f(y)}{(1-F(x))^{k_2-1}}I_{\{y\geq x\}} \qquad (3)\\
&= \frac{(k_2-1)!}{(k_2-k_1-1)!(k_1-1)!}((1-F(x))F(y|x))^{k_2-k_1-1}\\
&\quad \times \frac{((1-F(x))(1-F(y|x)))^{k_1-1}f(y|x)(1-F(x))}{(1-F(x))^{k_2-1}} \cdot I_{\{y\geq x\}}\\
&= \frac{(k_2-1)!}{(k_2-1-k_1)!(k_2-1-(k_2-k_1))!}\\
&\quad \times F(y|x)^{k_2-1-k_1}(1-F(y|x))^{k_1-1}f(y|x) \cdot I_{\{y\geq x\}}\\
&= g^{(k_2-k_1:k_2-1)}(y|x),
\end{aligned}
$$

where $f(y|x)(g^{(\cdot)}(y|x))$ denotes the truncated density of $f(\cdot)(g^{(\cdot)}(\cdot))$ truncated at x and $F(y|x)$ denotes the truncated distribution of $F(\cdot)$ truncated at x. This interpretation comes from the probability density function (PDF), $g^{(i;n)}(x)$ of the ith order statistic of the n sample, where

$$
g^{(i:n)}(x) = \frac{n!}{(i-1)!(n-i)!}[F(x)]^{i-1}[1-F(x)]^{n-i}f(x) \qquad (4)
$$

Then, the identification of the distribution of valuations is straightforward by Theorem 1 in Athey and Haile (2002) saying that the parent distribution is identified whenever the distribution of any order statistic (here $k_2 - k_1$) with a known sample size (here $k_2 - 1$) is

12

identified.

4.2 Auction Heterogeneity

In practice, the valuation of objects sold in WUCA (as in other auctions) varies according to several observed characteristics, for example, car types/ makes/ mileages/ year, etc. We want to control the effect of these observables on the individual valuation to obtain the homogeneity of the idiosyncratic factors such as idiosyncratic tastes, cost shocks, or demand shocks). For this purpose, we assume the following nonparametric form of the valuation $V(X_i)$ as

$$\ln V(X_i) = W(l^*(X_i)) + v_i, \tag{5}$$

where $W(\cdot)$ is a known link function and X_i is a vector of observable characteristics of auction and we assume that v_i is independent of X_i. Thus, we do assume the additively(or multiplicatively) separable structure of the value function, which is preserved by equilibrium bidding. In the auction we consider, ignoring the minimal increment, we have

$$\ln B(V_{ij}, X_i) = W(l^*(X_i)) + B(v_{ij}), \tag{6}$$

where V_{ij} is the valuation of a bidder j on an auction i, $B(V_{ij}, X_i)$ is a bidding function of a bidder j with observed heterogeneity of an auction i and $B(v_{ij})$ is a bidding function of homogeneous auctions. Under the IPV assumption, we have $B(V, X) = V(X)$ and $B(v) = v$ as before.

Here we do not make any parametric assumption on either $l(\cdot)$ or the distribution of v, $F_v(\cdot)$. In this case, we can identify both $l(\cdot)$ and $F_v(\cdot)$ up to location (see Athey and Haile (2002)). Thus, we need a normalization. We assume

$$l^*(0) = 0 \tag{7}$$

13

In what follows[17], we also assume $W(\cdot)$ is an identity function. Thus, we have

$$\ln V(X_i) = l^*(X_i) + v_i, \tag{8}$$

4.3 Estimation

4.3.1 Control on the Observed Heterogeneity

Assuming the independence of the idiosyncratic factor, v_i, on the observables, X_i, we can approximate the unknown function $l(X_i)$ in (8) using a sieve estimation such as power series sieve. We first approximate the function space $\mathcal{L}$ containing $l^*(X_i)$ with the following power series sieve space $\mathcal{L}_T$

$$\mathcal{L}_T = \{l(X) | l(X) = R^{k_1(T)}(X)'\pi \ \text{ for all } \pi \text{ satisfying } \|l\|_{\Lambda^{\gamma_1}} \leq c_1\}, \tag{9}$$

where $R^{k_1}(X)$ is a triangular array of some basis polynomials with the length of k_1. The function $\mathcal{L}_T$ is getting dense as $T \rightarrow \infty$ but not that fast, i.e. $k_1 \rightarrow \infty$ as $T \rightarrow \infty$ but $k_1/T \rightarrow 0$. Then, according to Theorem 8, p.90 in Lorentz (1986), there exists a π_{k_1}[18] such that for $R^{k_1}(x)$ on the compact set $\mathcal{X}$ (the support of X)

$$\sup_{x \in \mathcal{X}} |l^*(x) - R^{k_1}(x)'\pi_{k_1}| < c_1 k_1^{-\frac{[\gamma_1]}{d_x}}, \tag{10}$$

where $[s]$ is the largest integer less than s and d_x is the dimension of X. Thus, in what follows, we approximate the pseudo-value v_i in (8) as

$$v_i^{k_1} = \ln V_i - l_{k_1}(X), \tag{11}$$

where $l_{k_1}(x) = R^{k_1}(x)'\pi_{k_1}$.

Specifically, we consider the following polynomial basis considered by Newey, et al (1999). First let $\mu = (\mu_1, \ldots, \mu_{d_x})'$ denote a vector of nonnegative integers with the norm $|\mu| =$

---

[17]The identification and the estimation with a general link function and a general transformation of v_i will be a straightforward extension of what we consider here

[18]I will suppress the argument T in $k_1(T)$, unless otherwise noted from now on

$\sum_{j=1}^{d_x} \mu_j$, and let $x^\mu \equiv \prod_{j=1}^{d_x} (x_j)^{\mu_j}$. For a sequence $\{\mu(k)\}_{k=1}^\infty$ of distinct such vectors, we construct a tensor-product power series sieve as

$$R^{k_1}(x) = (x^{\mu(1)}, \ldots, x^{\mu(k_1)})' \tag{12}$$

Then, replacing each power x^μ by the product of orthonormal univariate polynomials of the same order, we may reduce collinearity.

4.3.2 Distribution of Valuations

Here we use the semi-nonparametric (SNP) estimation procedure developed by Gallant and Nychka (1987) and Coppejans and Gallant (2002). We implement a particular sieve estimation of the unknown density function using a Hermite series. First, we approximate the function space, $\mathcal{H}$, containing the true density function with a sieve space of the Hermite series, $\mathcal{H}_T$. Once we set up the objective function based on a Hermite series approximation of the unknown density function, then the estimation procedure is just a finite dimensional parametric problem. In particular, we use the maximum likelihood methods. What remains is to specify the particular rate in which a sieve space, $\mathcal{H}_T$, gets closer to $\mathcal{H}$ achieving the consistency of the estimator. We specify several regularity conditions for this.

Since we observe at least the second-, third- and fourth-highest bids in each auction of WUCA. We can estimate several different versions of the distributions of valuations ($F(\cdot)$), since any pair of order statistics can identify the parent distribution according to Song (2003) under the number of potential bidders unknown or unobserved. Here, we use two pairs of order statistics (second-, fourth-) and (third-, fourth-) highest bids and obtain two different values of $F(\cdot)$, which provides us an opportunity to test the hypothesis that WUCA is the IPV. This testable implication comes from the fact that under the IPV, value of $F(v)$ implied by the distributions of different order statistics must be identical for all v.[19]

---

[19] For detailed discussion, see Athey and Haile (2003).

Once we show that the IPV assumption holds, then we can combine several order statistics to identify $F(\cdot)$ extending Song (2003) to the case of more than three bids observed. This version of estimator is better than the version that use a pair of order statistics in the sense that we are using more information.

First, consider the estimation of the distribution of valuations using the second- and fourth-highest bids in each auction. Let (Y_i, X_i) denote the second- and fourth-highest pseudo-bids for each auction i and let $c = \min_i x_i$.[20] Then, $F(v)$ for $v < c$ can not be recovered form the data. Hence, we treat $F^*(\cdot) = F(\cdot|c)$ as the model primitive of interest, where $F(\cdot|c)$ denotes the truncated distribution of $F(\cdot)$ from below at c as

$$F^*(v|c) = \frac{F(v) - F(c)}{1 - F(c)} \tag{13}$$

Then, we obtain the density of Y_i conditional on X_i, $p_{(4,2)}(y_i|X_i = x_i)$ from (3) as

$$p_{(4,2)}(y|X = x) = \frac{6(F^*(y) - F^*(x))(1 - F^*(y))f^*(y)}{(1 - F^*(x))^3} \quad \text{for } y \geq x \geq c. \tag{14}$$

To estimate the unknown function $f^*(z)$ (hence, $F^*(z) = \int_c^z f^*(t)dt$), we first approximate $f^*(z)$ with the following specification of $f^K(z)$ up to the order $K(T)$:[21]

$$f^K(z) = \frac{\left(1 + \sum_{j=1}^K a_j \left(\frac{z-\mu}{\sigma}\right)^j\right)^2 \phi(z; \mu, \sigma, c)}{\int_c^\infty \left(1 + \sum_{j=1}^K a_j \left(\frac{t-\mu}{\sigma}\right)^j\right)^2 \phi(t; \mu, \sigma, c)dt}, \tag{15}$$

where $\phi(\cdot; \mu, \sigma, c)$ is the density of $N(\mu, \sigma)$ truncated below at c. Then, we construct the sample likelihood based on $f^K(\cdot)$ instead of the true $f(\cdot)$ using (14):

$$L(f^K) = \frac{1}{T} \sum_{i=1}^T \ln \frac{6(F^K(y_i) - F^K(x_i))(1 - F^K(y_i))f^K(y_i)}{(1 - F^K(x_i))^3}, \tag{16}$$

---

[20]Note that c is a consistent estimator of $\underline{v}$ under no binding reserve price and a consistent estimator of the reserve price under the binding case.

[21]I will suppress the argument T in $K(T)$ unless noted otherwise

where $F^K(z) = \int_c^\infty f^K(t)dt$. Noting that (16) is a parametric estimation problem, we approximate $f^K(\cdot)$ with $\hat{f}(\cdot)$ as the maximum likelihood estimator:

$$\hat{f}(z) = \frac{\left(1 + \sum_{j=1}^K \hat{a}_j \left(\frac{z-\hat{\mu}}{\hat{\sigma}}\right)^j\right)^2 \phi(z; \hat{\mu}, \hat{\sigma}, c)}{\int_c^\infty \left(1 + \sum_{j=1}^K \hat{a}_j \left(\frac{t-\hat{\mu}}{\hat{\sigma}}\right)^j\right)^2 \phi(t; \hat{\mu}, \hat{\sigma}, c)dt}, \tag{17}$$

where

$$(\hat{a}_1, \ldots, \hat{a}_K, \hat{\mu}, \hat{\sigma}) = \arg\max_{a_1,\ldots,a_K,\mu,\sigma>0} L(f^K) \tag{18}$$

Now note that actually a pseudo-bid z is defined as the residual in (8) and is approximated as the residual in (11). Thus, we have another set of parameter (π_{k_1}) to estimate in (18) as

$$\hat{f}(\hat{z}) = \frac{\left(1 + \sum_{j=1}^K \hat{a}_j \left(\frac{\hat{z}-\hat{\mu}}{\hat{\sigma}}\right)^j\right)^2 \phi(\hat{z}; \hat{\mu}, \hat{\sigma}, c)}{\int_c^\infty \left(1 + \sum_{j=1}^K \hat{a}_j \left(\frac{t-\hat{\mu}}{\hat{\sigma}}\right)^j\right)^2 \phi(t; \hat{\mu}, \hat{\sigma}, c)dt}, \tag{19}$$

where

$$(\hat{\pi}, \hat{a}_1, \ldots, \hat{a}_K, \hat{\mu}, \hat{\sigma}) = \arg\max_{\pi,a_1,\ldots,a_K,\mu,\sigma>0} L(f^K). \tag{20}$$

Denote the estimator $\hat{f}(\cdot)$ in (18) as $\hat{f}_1(\cdot)$ to distinguish this with other versions of estimator addressed later.

Note that our estimator requires a rich data set, since we estimate two nonparametric functions at the same time. The approximation precision depends on the choice of smoothing parameters k_1 and K. Here, we pick the optimal length of series (the dimension of the sieve space $\mathcal{H}_T$), K^*, following the Coppejans and Gallant (2002)'s method, which is a cross-validation strategy as used in a Kernel density estimation. Appendix Acontains a detailed discussion of choosing the optimal combination of K^* and k_1^*.

Similarly we can also identify and estimate $f^*(\cdot)$ using the pair of third- and fourth-highest bids from

$$p_{(4,3)}(y|X = x) = \frac{3(1 - F^*(y))^2 f^*(y)}{(1 - F^*(x))^3} \quad \text{for } y \geq x \geq c, \tag{21}$$

which is again obtained from (3). Denote the estimate of $f^*(\cdot)$ based on (21) as $\hat{f}_2(\cdot)$.

4.3.3 Simple Two Step Estimation

Though the estimation procedure considered up to now is a feasible one-step method and may be more efficient, we rather use a two-step estimation method as follows so that we can avoid the computational burden involved in estimating two unknown nonparametric functions $l^*(\cdot)$ and $F_v(\cdot)$ at the same time. First, we approximate the function $l(\cdot) = R^{k_1}(\cdot)'\pi_{k_1}$ estimating the following equation using the OLS

$$\ln V_{ij} = D'_{ij}\gamma + R^{k_1}(X_i)'\pi + \epsilon_{ij}, \tag{22}$$

where D_{ij} is a vector of dummy variables indicating that $j+1$ highest bids. Then, construct the residuals for each order statistics as

$$\hat{v}_{ij} = \ln V_{ij} - R^{k_1}(X_i)'\hat{\pi} \tag{23}$$

Here we are willing to assume the following regularity conditions

Assumption 4.1 $(V_{ij}, X_i), \ldots (V_{Tj}, X_T)$ *are i.i.d. for all j and $\mathrm{Var}(V_j|X)$ is bounded for all j.*

Assumption 4.2 *(i) the smallest and the largest eigenvalue of $\mathrm{E}[R^{k_1}(X)R^{k_1}(X)']$ is bounded away from zero uniformly in k_1 and; (ii) there is a sequence of constants $\zeta_0(k_1)$ satisfying $\sup_{x\in X} \left\|R^{k_1}(x)\right\| \leq \zeta_0(k_1)$ and $k_1 = k_1(T)$ such that $\zeta_0(k_1)^2 k_1/T \to 0$ as $T \to \infty$, where the matrix norm $\|A\| = \sqrt{trace(A'A)}$.*

Under Assumption 4.1 and 4.2, we have the consistency of $\hat{l}(\cdot) = R^{k_1}(\cdot)'\hat{\pi}$ in the mean squared norm or in the sup-norm from Newey (1997), since (10) implies Assumption 3 in Newey (1997) for the polynomial series approximation. Newey (1997) also showed $\zeta_0(k_1) \leq O(k_1)$ for power series sieves. Hence, $k_1 = O(T^\vartheta)$ with $0 < \vartheta < \frac{1}{3}$ satisfies Assumption 4.2.

In the second step, based on the estimated pseudo values $\hat{v}_{ij}$, we estimate $f_v(\cdot)$ based on (18). We pick the optimal length of series k_1^* and K^* using again cross-validation strategies.

For K^*, we follow the Coppejans and Gallant (2002)'s method and for k_1^*, we employ a similar cross-validation method common in a usual kernel regression but we choose k_1^* as the minimizer of the sample average mean squared error. For detailed discussion, again see Appendix A.2.

4.3.4   Testable Restriction: Part 1

As noted in the previous section, we can test the IPV assumption, since several versions of distribution of valuations are identified under availability of three order statistics. In particular, one is from the pairs of the second- and fourth- highest bids and the second one is from the pairs the third- and fourth- order statistics. Therefore, by comparing $\hat{f}_1(\cdot)$ and $\hat{f}_2(\cdot)$, we can test the following hypothesis H_0 against H_0

$$H_0 \quad : \quad \text{WUCA is an IPV auction} \tag{24}$$
$$H_A \quad : \quad \text{WUCA is not an IPV auction,}$$

since under H_0, there should be no significant difference between $\hat{f}_1(\cdot)$ and $\hat{f}_2(\cdot)$.

4.3.5   Test Statistics

Tests based on Means or Higher Moments   We can test (24) based on the means or higher moments implied by f_1 and f_2 as

$$H_0^j \, (IPV) \quad : \quad \mu_1^j = \mu_2^j \tag{25}$$
$$H_A^j \, (NIPV) \quad : \quad \mu_1^j \neq \mu_2^j, \quad j = 1, 2, \ldots, J$$

where $\mu_k^j = \int_c^\infty v^j f_k(v) dv$, $k = 1, 2$, since (24) implies (25) and (24) implies (24) as $J \to \infty$. We can compare several estimates of moments implied by $\hat{f}_1$ and $\hat{f}_2$ and test the significance difference of each pair by constructing a standardized test statistics. One difficulty is to reflect the fact that we used pre-estimated functions in obtaining $\hat{f}_1$ and $\hat{f}_2$ in calculating the asymptotic variance of each moment estimate. [To be completed]

A Sup-Norm Test   Another possible test is based on a Kolmogorov-Smmirnov-type (KS) statistic testing equal distribution as

$$\delta_T = \sup_{v \in [\underline{v}, \bar{v}]} |\hat{F}_1(v) - \hat{F}_2(v)|, \tag{26}$$

where $\hat{F}_k = \int_c^\infty \hat{f}_k(v)dv$, $k = 1, 2$. Again the difficulty lies in the fact that the estimates of $\hat{F}_1$ and $\hat{F}_2$ contain some pre-estimated functions, which may invalidate any bootstrap based inference method without any knowledge on the first-order asymptotics of δ_T. [To be completed]

4.3.6   Combining Several Order Statistics

Once we show the several versions of estimates for the distribution of valuations are statistically not different each other, we may obtain a better estimate by combining these. One way to do this is to consider the joint density function of two or more order statistics conditional on a certain order statistic. Assume that we have the k_1^{th}, k_2^{th}, and k_3^{th}-highest order statistics, which are the $(n - k_1 - 1)^{th}$, $(n - k_1 - 1)^{th}$, $(n - k_1 - 1)^{th}$ order statistics respectively ($1 \leq k_1 < k_2 < k_3 \leq n$). Denote the joint density of these three order statistics as $\tilde{g}^{(k_1, k_2, k_3:n)}(.)$[22]

$$\tilde{g}^{(k_1, k_2, k_3:n)}(x, y, z) = \frac{n!}{(n - k_3)!(k_3 - k_2 - 1)!(k_2 - k_1 - 1)!(k_1 - 1)!} \tag{27}$$
$$\times F(x)^{n-k_3} f(x)[F(y) - F(x)]^{k_3 - k_2 - 1} f(y)[F(z) - F(y)]^{k_2 - k_1 - 1} f(z)[1 - F(z)]^{k_1 - 1},$$

where Z denotes the k_1^{th}-, Y denotes k_2^{th}- and X denotes k_3^{th}-highest order statistics. Using this joint density function with (4), we obtain the conditional joint density of the k_1^{th} and

---

[22]We use this notation $\tilde{g}^{(\cdot)}$ to distinguish it from $g^{(\cdot)}$ so that k_i denotes the k_i^{th} highest order statistics

the k_2^{th}-highest order statistics conditional on the k_3^{th}-highest statistics as

$$
\begin{aligned}
p_{(k_1,k_2|k_3)}(z,y|x) &= \frac{(k_3-1)!}{(k_3-k_2-1)!(k_2-k_1-1)!(k_1-1)!} \\
&\times \frac{[F(y)-F(x)]^{k_3-k_2-1}f(y)[F(z)-F(y)]^{k_2-k_1-1}f(z)[1-F(z)]^{k_1-1}}{[1-F(x)]^{k_3-1}} \\
&\cdot I_{\{y\geq x, z\geq y\}} \\
&= \frac{(k_3-1)!}{(k_3-k_2-1)!(k_2-k_1-1)!(k_1-1)!}[(1-F(x))F(y|x)]^{k_3-k_2-1} \\
&\times f(y|x)(1-F(x))[(F(z|x)-F(y|x))(1-F(x))]^{k_2-k_1-1} \\
&\times \frac{f(z|x)(1-F(x))[(1-F(z|x))(1-F(x))]^{k_1-1}}{[1-F(x)]^{k_3-1}} \cdot I_{\{y\geq x, z\geq y\}} \\
&= \frac{(k_3-1)!}{(k_3-k_2-1)!(k_2-k_1-1)!(k_1-1)!}F(y|x)^{k_3-k_2-1}f(y|x) \\
&\times [F(z|x)-F(y|x)]^{(k_3-k_1)-(k_3-k_2)-1}f(z|x)[1-F(z|x)]^{(k_3-1)-(k_3-k_1)} \\
&= g^{(k_3-k_1,k_3-k_2:k_3-1)}(z,y|x),
\end{aligned}
\tag{28}
$$

where $F(\cdot|x)$ and $f(\cdot|x)(g(\cdot|x))$ are the truncated CDF and PDF truncated at x respectively and $I_{\{\cdot\}}$ is the indicator function whose value equals to one when $\{\cdot\}$ is true. Therefore we can interpret $p_{(k_1,k_2|k_3)}(\cdot)$ as the joint density of $(k_3-k_1)^{th}$ and $(k_3-k_2)^{th}$ order statistic from a sample of which size is (k_3-1) from $F(\cdot)$. When $(k_1,k_2,k_3)=(2,3,4)$, (28) becomes

$$
p_{(2,3|4)}(z,y|x) = \frac{6f(y)f(z)[1-F(z)]}{[1-F(x)]^3}
\tag{29}
$$

Based on (29), we can estimate the distribution of valuations, $f(\cdot)$, following the method proposed in Section 4.3.2. The resulting estimator is more efficient than $\hat{f}_1(\cdot)$ or $\hat{f}_2(\cdot)$ in the sense that it uses more information than the others.

4.4   Estimation Results

4.4.1   Benchmark Monte Carlo

In this section, we perform several Monte Carlo experiments to illustrate the validity of our estimation strategy. First, we generate artificial data of $T = 1000$ auctions as follows. The

number of potential bidders, $\{N_i\}$, are drawn from a Binomial distribution with $(n,p) = (50, 0.1)$ for each auction $(i = 1, \ldots, T)$. N_i potential bidders are assumed to value the object according to:

$$\ln V_{ij} = \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_3 X_{3i} + v_{ij}, \tag{30}$$

where $\alpha_1 = 1$, $\alpha_2 = -1$, $\alpha_3 = 0.5$, $X_{1i} \sim N(0,1)$, $X_{2i} \sim Exp(1)$, $X_{3i} = X_{1i} \cdot X_{2i} + 1$, and $v_{ij} \sim Gamma(9,3)$[23]. $X_{\cdot i}$'s represent the observed auction heterogeneity and v_{ij} is bidder j's private information in auction i, whose distribution is our primary interests here. To consider the case of biding reserve prices, we also generate the reserve prices equation as

$$\ln R_i = \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \eta_i, \tag{31}$$

where $\eta_i \sim Gamma(9,3) - 2$. Note that by construction V_{ij} and R_i are independent conditional on $X_{\cdot i}$'s. Artificial actual bidders bid only when those V_{ij} are greater than R_i. Here we assume our imaginary researcher do not know the presence of potential bidders with valuations below R_i. Thus, in each experiment, she has a data set of $X_{\cdot i}$'s, and the second-, the third-, and the fourth-highest among actual bidder's bids. Auctions with fewer than four actual bidders are dropped. Hence, our research has the sample size less than $T = 1000$ on average around $T = 700$. Our researcher estimates α_1, α_2, α_3 and $f_v(\cdot)$ by varying the smoothness (K) of the SNP estimator, from 0 to 9 without knowing the specification of the distribution of v_{ij} in (30). Figure 1 and Figure 2 illustrate the performance of the SNP estimator considered here. By construction of the data generation, the two versions of estimates for the density function of valuations should be almost identical (one is based on $(2^{nd}, 4^{th})$ order statistics and the other is on $(3^{rd}, 4^{th})$).

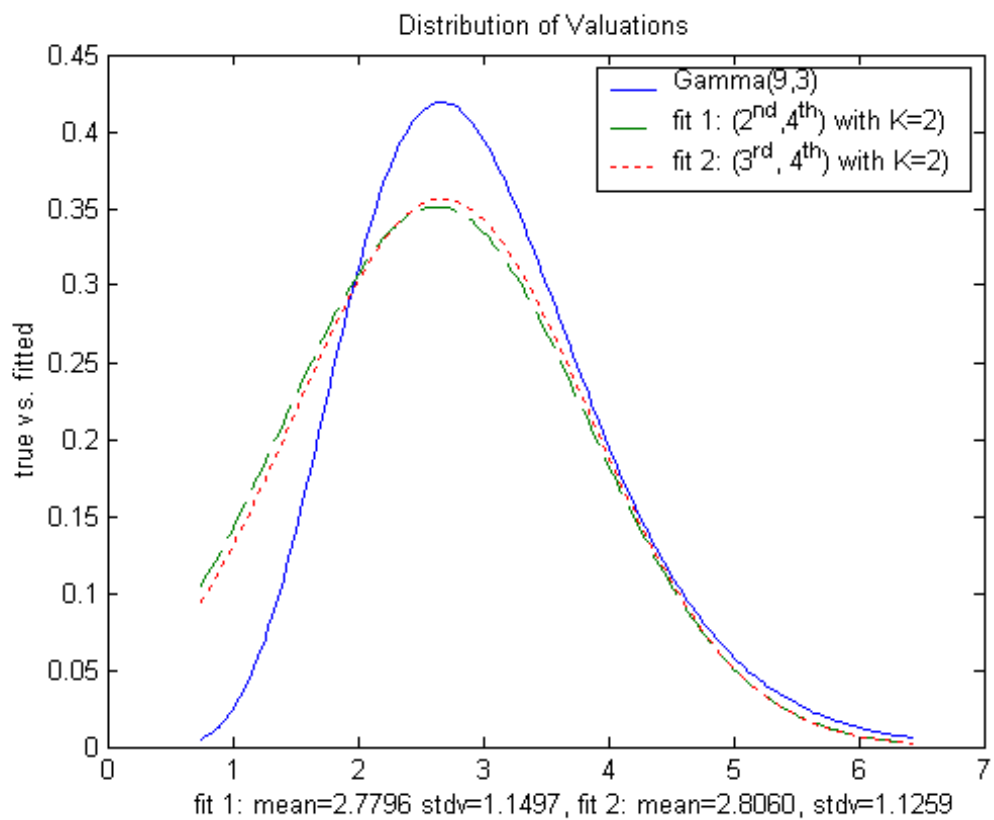[23]Note that for $X \sim Gamma(9,3)$, $E(X) = 3$ and $Var(X) = 1$.

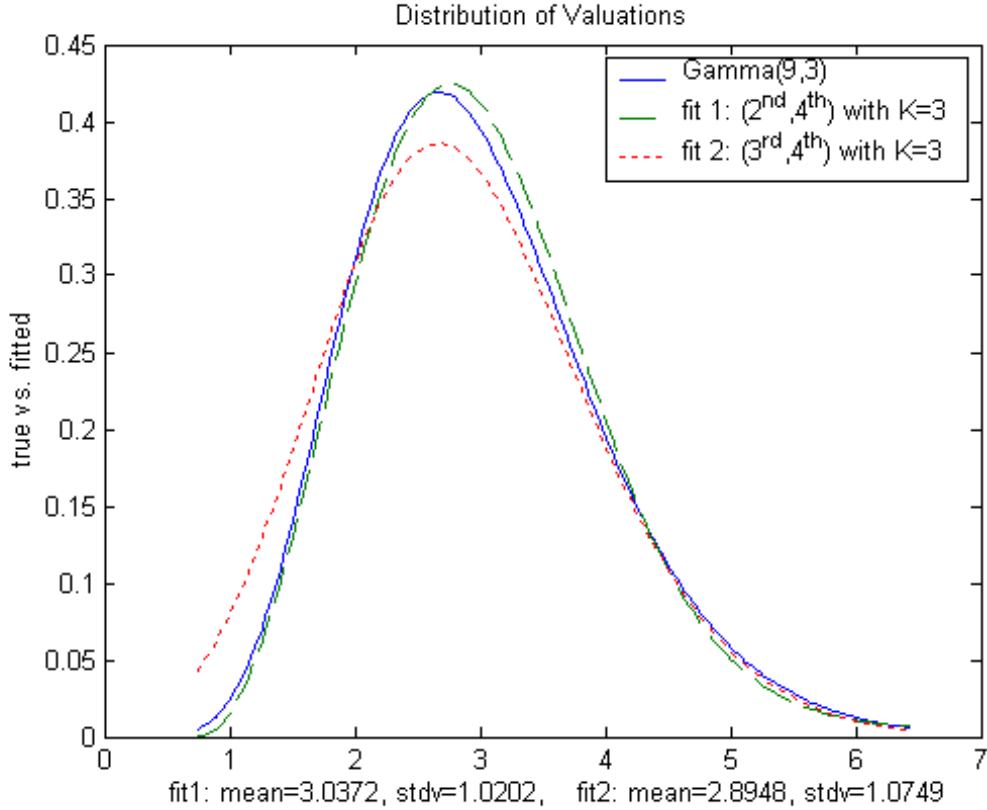Figure 1. SNP Density function estimation with $K = 2$ for the Simulated Data

Figure 2. SNP Density function estimation with $K = 3$ for the Simulated Data

Cross-Validation   [To be completed]

4.4.2   Estimation Result

In the first stage regression obtaining the approximated function of the observed hetero-geneity part, $\hat{l}(x)$. We consider the following covariates: X_1 is the vector of dummy vari-ables indicating the make such as Hyundai, Daewoo, Kia or Others; X_2 is the age; X_3 is the mileage; X_4 is the engine size and X_5 is the inspection score. We estimate $l^*(x)$ separately for each car make and obtain the estimate of pseudo valuations as residuals im-posing the restriction $l^*(0) = 0$ for identification. Thus, we actually use the basis func-

tional form $l_H(X_2, X_3, X_4, X_5)$ for Hyundai and use $c_m + l_m(X_2, X_3, X_4, X_5)$ for others, $m \in \{$Daewoo, Kia, Others$\}$.
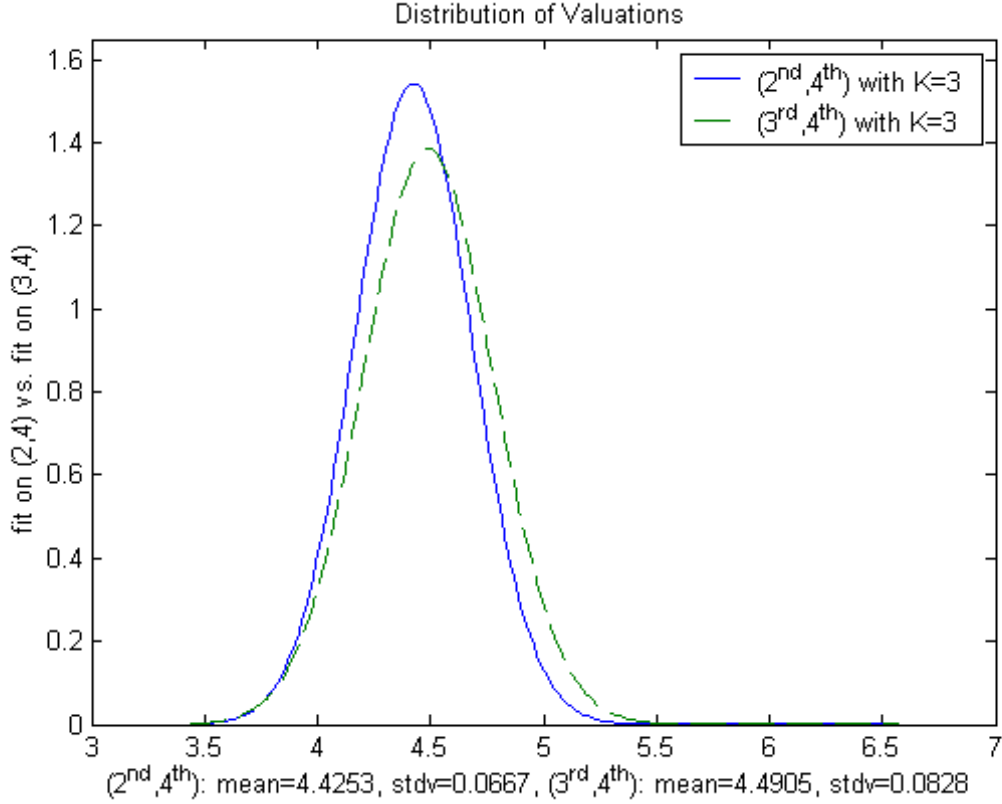

Figure 3. Benchmark SNP Density function estimation with $K = 3$ for the WUCA Data

Based on the estimated pseudo valuations, in the second step, we estimate the distribution of valuations using two pairs of order statistics $(2^{nd}, 4^{th})$ or $(3^{rd}, 4^{th})$. Figure 3 illustrates the estimated density function of valuations using OLS estimation in the first stage as a benchmark. Figure 4 shows the density function estimates based on the series estimation in the first stage. Following the cross-validation strategy explained in Appendix A, we can pick the optimal lengths of series k_1^* for the first stage regression and K^* for the SNP estimator.
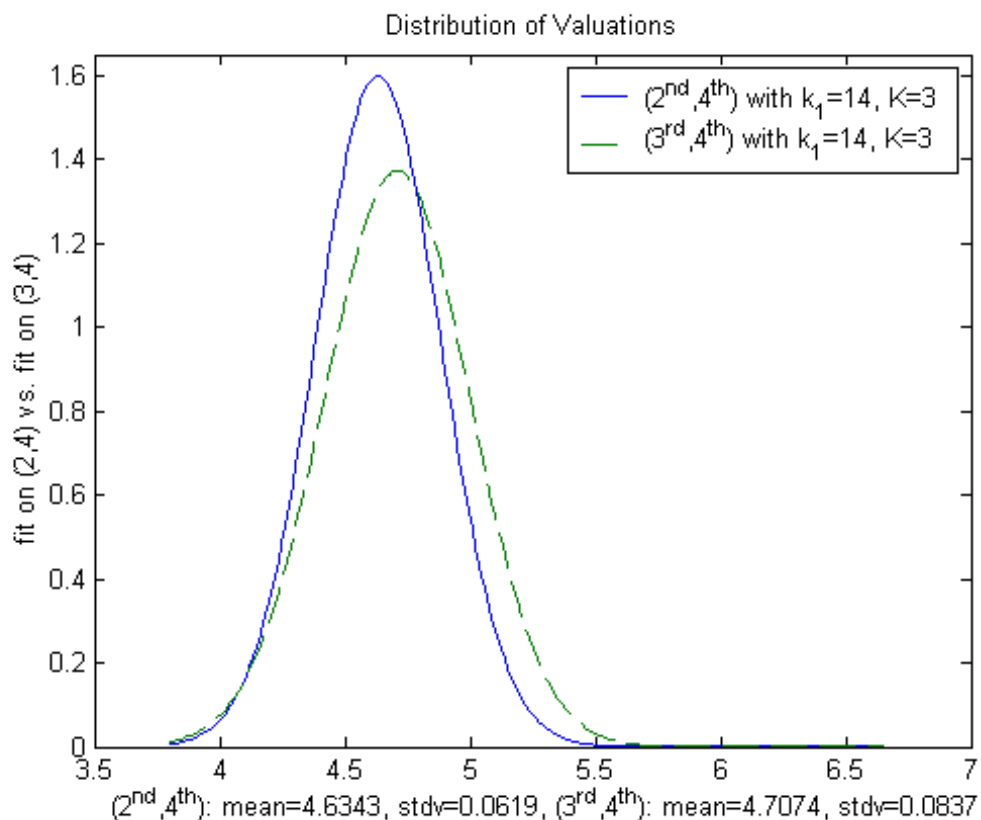
Figure 4. SNP Density function estimation with $K = 3$ and $k_1 = 14$ for the

WUCA Data

Cross-Validation [To be completed]

4.5 Optimal Reserve Price

The key policy issue for the seller is the reserve price. The seller wants to maximize the expected profit by setting a minimum acceptance price so that only bidders have higher valuations than the reserve price attend the auction. The optimum depends on the distribution of valuations, which is our primary interests and derived in previous sections. We are willing to assume the following, which is implied by the standard regularity condition of Myerson

(1981) and assumed in Haile and Tamer (2003).

Assumption 4.3 $(p - vc_0)[1 - F_v(p)]$ *is strictly pseudo-concave in p on $(\underline{v}, \bar{v})$,*

where vc_0 is the cost associated with the auction. The pseudo-optimal reserve price (without the observed heterogeneity) is characterized by

$$p^* = \arg\max_p (p - vc_0)[1 - F_v(p)], \tag{32}$$

which becomes, under Assumption 4.3

$$p^* = vc_0 + \frac{1 - F(p^*)}{f(p^*)} \tag{33}$$

One nice feature of the additively separability assumed in (5) is that the equilibrium bidding is preserved under the observed heterogeneity as $B(V(x)) = l^*(x) + B(v)$, where $B(v)$ is the bidding function under being absence of the observed auction heterogeneity. Thus, the optimal reserve price also has the simple additive form of the observed heterogeneity part and the pseudo-optimal reserve price:

$$p^*(X) = l^*(X) + p^* = l^*(X) + vc_0 + \frac{1 - F_v(p^*)}{f_v(p^*)} \tag{34}$$

Thus, we can estimate $p^*(x)$ using the previous estimates of $\hat{l}(\cdot)$, $\hat{f}_v(\cdot)$ and $\hat{F}_v(\cdot)$ as

$$\hat{p}(x) = \hat{l}(x) + \hat{p}, \tag{35}$$

where $\hat{p}$ solves $p = \hat{vc} + \frac{1 - \hat{F}_v(p)}{\hat{f}_v(p)}$ and $\hat{vc}$ is a consistent estimator of vc_0. It will be very interesting to compare these implied optimal reserve prices from the distribution of valuations and the actual reserve prices recorded in each auction of WUCA, since the actual reserve price data is readily available in our data set. If significant difference emerges between these two and a particular pattern is found in there difference , then it may shed lights on the seller's strategic behavior in WUCA, if any.

Comparison of the Optimal Reserve Price and the Actual Reserve Price [To be addressed later]

4.6 Bounds Estimation

Until now, we have disregarded the minimum increment of around 30 dollars in WUCA. In this section, we discuss how to obtain the bounds of the distribution of valuations incorporating the fact that there exists the minimum increment in WUCA. The bounds considered here is much simpler than those considered in Haile and Tamer (2003), since in WUCA, by construction, any order statistic of valuations other than the first highest one is bounded as

$$b_{(i:n)} \leq v_{(i:n)} \leq b_{(i:n)} + \Delta, \text{ for all } i = 1, \ldots, n-1, \tag{36}$$

where $(i:n)$ denotes the ith order statistic out of the n sample. By the first-order stochastic dominance, noting $G_{b_{(i:n)}+\Delta}(v) = G_{b_{(i:n)}}(v - \Delta)$, (36) implies

$$G_{b_{(i:n)}}(v) \geq G_{v_{(i:n)}}(v) \geq G_{b_{(i:n)}}(v - \Delta), \tag{37}$$

, where $G.(\cdot)$ is the distribution of the order statistics. Then, using the identification method discussed in previous sections, we have

$$F_b(v) \geq F_v(v) \geq F_b(v - \Delta) \cong F_b(v) - f_b(v)\Delta, \tag{38}$$

where $F_b(\cdot)(f_b(\cdot)$ is the distribution (PDF) of valuations based on bids and the last weak equality comes from the first-order taylor series expansion. Therefore, we can estimate the bounds of $F_v(v)$ as

$$\hat{F}_b(v) \geq F_v(v) \geq \hat{F}_b(v) - \hat{f}_b(v)\Delta, \tag{39}$$

where $\hat{f}_b(\cdot)$ the SNP estimator based on the certain observed order statistics of bids, $\hat{F}_b(x) = \int_{\min(b)}^{x} \hat{f}_b(v)dv$ and $\min(b)$ is the minimum among the observed bids considered.

5 Concluding Remarks

[To be added]

Appendix

A   Choosing the optimal smoothing parameters

A.1   Choosing k_1^*

Here we use a sample version of the Mean Squared Error criterion for the cross-validation as

$$SMSE(\hat{l}) = \frac{1}{T}\sum_i^T [\hat{l}(X_i) - l^*(X_i)]^2, \tag{40}$$

where $l(\hat{\cdot}) = R^K(\cdot)'\hat{\pi}$. Instead of using the Leave-one-out method, we will partition the data into P groups, making the size of each group as equal as possible and use the *Leave-one partition-out* method. This is because it will be computationally too expensive to use the Leave-one-out method, since the data size is so large. Namely, we estimate the function $l^*(\cdot)$ from the sample after deleting the pth group with the length of the series equal to k_1 and denote this as $\hat{l}_{p,k_1}(\cdot)$. As a next step, we choose k_1^* such that

$$\arg\min_{k_1} CV(k_1) \tag{41}$$

where $\{p\}$ denotes the set of the data indices belonging to the pth group.

A.2   Choosing K^*

Coppejans and Gallant (2002) employ a cross-validation method based on the ISE (Integrated Squared Error) criteria. The ISE is defined for $\hat{h}(x)$, a density estimate of $h(x)$

$$\begin{aligned} \text{ISE}(\hat{h}) &= \int \hat{h}^2(x)dx - 2\int \hat{h}(x)h(x)dx + \int h(x)^2dx \tag{42} \\ &= M_{(1)} - 2M_{(2)} + M_3. \end{aligned}$$

To approximate the ISE in terms of $p^*(y|x)$, we again use the cross-validation strategy with the data partitioned into P groups. We first approximate $M_{(1)}$ with

$$
\begin{aligned}
\widehat{M}_{(1)}(K) &= \int (\hat{p}^*{}_K(y|x))^2 \\
&= \int \left(\frac{6(\hat{F}_K(y) - \hat{F}_K(x))(1 - \hat{F}_K(y))\hat{f}_K(y)}{(1 - \hat{F}_K(x))^3} \right)^2,
\end{aligned}
\tag{43}
$$

where $\hat{f}_K(\cdot)$ denotes the SNP estimate with the length of the series equal to K and $\hat{F}_K(z) = \int_c^z \hat{f}_K(t)dt$. For $M_{(2)}$, we consider

$$
\begin{aligned}
\widehat{M}_{(2)}(K) &= \frac{1}{T} \sum_{p=1}^P \sum_{t \in \{p\}} \hat{p}^*{}_{p,K}(y_t|x_t) \\
&= \frac{1}{T} \sum_{p=1}^P \sum_{t \in \{p\}} \frac{6(\hat{F}_{p,K}(y_t) - \hat{F}_{p,K}(x_t))(1 - \hat{F}_{p,K}(y_t))\hat{f}_{p,K}(y_t)}{(1 - \hat{F}_{p,K}(x_t))^3},
\end{aligned}
\tag{44}
$$

where $\hat{f}_{p,K}(\cdot)$ denotes the SNP estimate obtained from the sample excluding p^{th} group with the length of the series, K and $\hat{F}_{p,K}(z) = \int_c^z \hat{f}_{p,K}(t)dt$. Noting $M_{(3)}$ is not a function of K, we pick K^* such that

$$
K^* = \arg\min_K CVK(K)
\tag{45}
$$

31

References

[1] Arnold, B., Balakrishnan, N., and Nagaraja, H. (1992), *A First Course in Order Statistics.* (New York: John Wiley & Sons.)

[2] Athey, S. and Haile, P. (2002), "Identification of Standard Auction Models," *Econometrica*, 70(6), 2107-2140.

[3] Athey, S. and Levin, J. (2001), "Information and Competition in U.S. Forest Service Timber Auctions," *Journal of Political Economy*, 109, 375-417.

[4] Bikhchandani, S., Haile, P. and Riley, J. (2002), "Symmetric Separating Equilibria in English Auctions," *Games and Economic Behavior*, 38, 19-27.

[5] Bikhchandani, S. and Riley, J. (1991), "Equilibria in Open Common Value Auctions," *Journal of Economic Theory*, 53, 101-130.

[6] Bikhchandani, S. and Riley, J. (1993), "Equilibria in Open Auctions," mimeo, UCLA.

[7] Coppejans, Mark and A. Ronald Gallant (2002), "Cross-validated SNP density estimates," *Journal of Econometrics*, 110, 27-65.

[8] Gallant, A. Ronald and Douglas Nychka (1987), "Semi-nonparametric Maximum Likelihood Estimation," *Econometrica*, 55, 363-390.

[9] David, H. A. (1981), *Order Statistics.* Second Edition. (New York: John Wiley & Sons.)

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

[11] Haile, P., Hong, H., and Shum, M. (2003), "Nonparametric Tests for Common Values in First-Price Sealed-Bid Auctions," mimeo, Yale University.

[12] Haile, P. and Tamer, E. (2001), "Inference from English Auctions with Asymmetric Affiliated Private Values," mimeo, University of Wisconsin.

[13] Haile, P. and Tamer, E. (2003), "Inference with an Incomplete Model of English Auctions," *Journal of Political Economy*, 111(1), 1-51.

[14] Harstad, R. and Rothkopf, M. (2000), "An "Alternating Recognition" Model of English Auctions," *Management Science*, 46, 1-12.

[15] Hendricks, K. and Paarsch, H. (1995), "A Survey of Recent Empirical Work concerning Auctions," *Canadian Journal of Economics*, 28, 403-426.

[16] Hendricks, K., Pinkse, J., and Porter, R. (2003), "Empirical Implications of Equilibrium Bidding in First-Price, Symmetric, Common Value Auctions," *Review of Economic Studies*, 70(1), 115-146.

[17] Hong, H., and Shum, M. (2003), "Econometric Models of Asymmetric Ascending Auctions," *Journal of Econometrics*, 112, 327-358.

[18] Izmalkov, S. (2003), "English Auctions with Reentry," mimeo, MIT.

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

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

[21] Li, T., Perrigne, I., and Vuong, Q. (2002), "Structural Estimation of the Affiliated Private Value Auction Model," *RAND Journal of Economics*, 33, 171-193.

[22] Lorentz, G. (1986), *Approximation of Functions*, New York: Chelsea Publishing Company.

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

[24] Myerson, Roger B. (1981), "Optimal Auction Design," *Math. Operations Res.*, 6, 58-73.

[25] Newey, Whitney K (1997), "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79, 147-168.

[26] Newey, W.K., J.L. Powell and F. Vella (1999), "Nonparametric Estimation of Triangular Simultaneous Equations Models," *Econometrica*, 67, 565-603.

[27] Paarsch, H. (1992), "Deciding Between the Common and Private Value Paradigms in Empirical Models of Auctions," *Journal of Econometrics*, 51, 191-215.

[28] Riley, J. (1988), "Ex Post Information in Auctions," *Review of Economic Studies*, 55, 409-429.

[29] Song, U. (2003), "Nonparametric Estimation of an eBay Auction Model with an Unknown Number of Bidders," mimeo, University of Wisconsin-Madison.