Econometrics of Auctions by Least Squares

Leonardo Rezende*

PUC-Rio and University of Illinois

September 28, 2007

Abstract

This paper investigates using ordinary least squares (OLS) on auction data. We find that for parameterizations of the valuation distribution that are common in empirical practice, an adaptation of OLS provides unbiased estimators of structural parameters. Under symmetric independent private values, adapted OLS is a specialization of the method of moments strategy of Laffont, Ossard and Vuong (1995). In contrast to their estimator, here simulation is not required, leading to a computationally simpler procedure. The paper also discusses using estimation results for inference on the shape of the valuation distribution, and applicability outside the symmetric independent private values framework.

JEL Codes: D44, C14.

Keywords: Auctions, Ordinary Least Squares, Semiparametric Methods.

*Address: Departamento de Economia, PUC-Rio, Rua Marquês de São Vicente, 225 SL 210F, Rio de Janeiro, Brazil, 22453-900. email: lrezende@econ.puc-rio.br. Tel.: 55-21-3527-1078, Fax: 55-12-3527-1084. Parts of this research are based on chapters 4 and 5 of my Ph.D. dissertation at Stanford University. I gratefully acknowledge the support of a John M. Olin Dissertation Fellowship and a Melvin and Joan Lane Stanford Graduate Fellowship. Thanks to Pat Bajari, Lanier Benkard, Tim Bresnahan, George Deltas, Phil Haile, Ali Hortaçsu, Thierry Magnac, Paul Milgrom, Harry Paarsch and two anonymous referees for valuable suggestions. All errors are mine.

1 Introduction

The field of econometrics of auctions has been successful in providing methods for the investigation of auction data that are well grounded in economic theory and allow for inference on the structure of an auction environment. Today a researcher has a number of alternative structural methods, specially within the independent private values paradigm; an excellent reference to this literature is the Paarsch *et al.* (2006) book. To name a few alternatives, it is possible to use maximum likelihood (Donald and Paarsch, 1996), nonparametric methods (Guerre *et al.*, 2000), simulated nonlinear least squares (Laffont *et al.*, 1995) and bounds estimation of incomplete models (Haile and Tamer, 2003).

And yet, it is still common to find empirical studies of auction markets that do not use these techniques and instead run regressions of the following form:

$$p = X\beta + \epsilon, \tag{1}$$

where p is the transaction price (or its log). For example, several studies have recently investigated the importance attributed by consumers to the reputation of sellers in eBay using specifications of this sort (Houser and Wooders, 2000; Lucking-Reiley *et al.*, 2000; McDonald and Slawson, 2002; Melnik and Alm, 2002). These studies identify interesting empirical relationships, but they are not structural, in the sense of identifying parameters from the valuation distribution. This is a drawback, since these estimates cannot be used for counterfactual analysis of changes in the institutional environment.

What makes specification 1 not structural is not the right-hand side of regression — covariates may indeed impact consumer preferences in a linear fashion — but rather the left-hand side. A superior starting point for an empirical analysis would be a specification such as

$$V_i = X\beta + \epsilon_i, \tag{2}$$

where V_i is consumer i's valuation, or willingness to pay, for the product being auctioned. Unlike price, valuation is a demand concept, that reflects the consumer's preferences in isolation of supply or market institutions effects. A bidder will never elect to pay his or her own valuation for the product — doing so would guarantee that the bidder would not gain anything from participating in an auction. Because of that, we know prices and valuations are not the same. Therefore, whenever the objective is to measure consumers' preferences, a regression like equation 1 would suffer from misspecification bias.

The main point of this paper is that applying the broad method of moments approach from Laffont *et al.* (1995) to a standard specification such as equation 2 leads to the possibility of structural estimation by an adaptation of ordinary least squares.

Laffont *et al.* (1995) exploit the fundamental property of symmetric independent private values auctions, expected revenue equivalence, as a moment condition that can be used to estimate parameters that appear in the valuation distribution in a general way. To obtain a mapping between parameters and moments for estimation purposes they use computer simulation. They show that due to simulation error standard nonlinear least squares on simulated moments is biased; they propose instead a simulated nonlinear least squares estimator (SNLLS) that minimizes an objective function that corrects for the simulation bias.

In this paper, we argue that once one specializes the analysis to parameterizations that are commonly made in empirical work, the method simplifies dramatically. In particular, simulation of the sort used in the SNLLS is no longer necessary, which reduces the computational cost of the algorithm and the interpretation of its statistical properties.

It will also be shown that, while revenue equivalence is helpful, it is not crucial for the method. The key property is that auction games are "linear", in the sense that a change of location or scale of *all* valuations within an auction leads to a corresponding change in the expected price in equilibrium. Using this property, it is possible to generalize the method to some environments where revenue equivalence does not hold.

Of course, the idea (or the practice) of estimation with auction data through OLS is not new. What we will do is propose an adjustment, an additional artificial regressor, that makes the standard OLS estimators unbiased for the parameters of the valuation distribution. In exercise 3.2.b of Paarsch *et al.* (2006), the same adjustment is proposed in the context of a second-price auction.[1] Here we argue that the adjustment is applicable to a wider class of auctions and, when properly generalized, to contexts where revenue equivalence does not hold.

We hope the discussion presented here of the opportunities and limits of OLS for the structural estimation of auction models can help guide researchers interested in estimating factors that affect bidder valuations in a way that is simple to implement and easy to interpret, as well as theoretically sound; conversely, it may also help inform researchers of the potential pitfalls of simple OLS on auction data in many contexts.

We do not see the adjusted OLS method proposed here as a true substitute for more complex estimation procedures found in the literature; the computational simplicity comes with a cost. First, while unbiased the estimators are not as efficient as maximum likelihood. Second, the method depends on a parameterization where only the location and scale of valuations vary across auctions (and the variation is the same for every bidder). This is too restrictive in some

[1] I thank an anonymous referee for pointing me this reference.

circumstances; for example, in a study of an asymmetric auction environment the focus may be in whether a given covariate affects bidder valuations differently in the same auction. Such effect cannot be accounted for in the framework discussed here. Even then, we hope the methods discussed here can find a place in the toolbox of applied economists, as they may help guide preliminary work on auction data.

The paper is organized as follows: Section 2 describes the setting. Section 3 introduces the method. The possibility of exploring the information obtained from the method for inference about the shape of the valuation distribution is discussed in section 4. Section 5 discusses whether the method can be generalized for contexts where revenue equivalence does not hold. An illustration with a practical application of the method to a sample of iPod mini eBay auctions is done in section 6. Section 7 concludes.

2 The Setting

We are interested in estimating the parameters that determine the location and scale of distributions of bidder valuations in a sample of L auctions. Let V_{il} be the valuation of the i-th bidder in the l-th auction. Let μ_l be the mean and σ_l the standard deviation of valuations in auction l. As the subscript suggests, they may vary across auctions. We will assume that variation across auctions only affects the location and scale of valuations, not other aspects of the valuation distribution:

Assumption 1 $V_{il} = \mu_l + \sigma_l \epsilon_{il}$, *where the* ϵ_{il} *are i.i.d. with distribution* F.

The standardized valuation, $\epsilon_{il} = (V_{il} - \mu_l)/\sigma_l$, has a common distribution F that does not vary across auctions or bidders. For the moment we also impose independence, both across auctions and bidders.

Independence across different auctions is an assumption made here for convenience. Relaxing it in what follows would have the same effect of having non-spherical disturbances in a linear regression model: it would affect efficiency, but not unbiasedness of the estimators.[2]

Independence across bidder valuations within an auction is a requirement of the benchmark model in auction theory, the symmetric independent private values auction model. Section 5 investigates the possibility of extending the method to settings where valuations are not independent.

We consider the problem of a researcher interested in evaluating the effect of exogenous covariates on μ_l and σ_l. Let X_l be the vector of covariates that affect the expected valuation μ_l and Z_l the vector of covariates that affect σ_l. It is possible to have regressors that appear in both X_l and Z_l. We assume that $[X_l, Z_l]$ are either deterministic or otherwise publicly known by all bidders before auction l starts. Linearity is imposed:

Assumption 2 $\mu_l = X_l\beta$ *and* $\sigma_l = Z_l\alpha$.

This is imposed solely in the interest of simplicity. All that follows would still hold if these relationships were nonlinear, substituting nonlinear least squares for OLS. Linearity in the specification of σ_l is unusual, since the standard practice is to specify a form that restricts it to be non-negative. This can be done here as well, and again would lead to nonlinear least squares.[3]

We intend to obtain estimates of β and α from a sample of auctions of which we know the covariates $[X_l, Z_l]$, the number of bidders n_l and the winning price

[2] In any case, spherical disturbances in equation 2 is not enough to guarantee efficiency of OLS, since the regression will be heteroskedastic even under this assumption.

[3] It may be worth pointing out that the role played by the specification of σ_l in what follows is not the usual one. When the functional form of σ_l is used to correct for heteroskedasticity in a GLS procedure, allowing for negative values is computationally problematic, since the objective function of the estimation procedure may become ill-behaved. This is not a problem here, since σ_l will appear in the least squares computation in a different way. Hypothetical negative values for σ_l would not lead to any computational problems, although they are hardly justifiable on economic grounds.

p_l. We require

Assumption 3 *The number of bidders, n_l, is exogenous and common knowledge. Bidders are risk-neutral, and maximize their profits at each auction in isolation.*

Here exogeneity is meant both in the game-theoretic sense — n_l is taken as given, and is not determined by each bidder decision-making process — and in the econometric sense — ϵ_{il} and n_l are independent. The number of bidders is also assumed to be publicly known before bidding.

We consider a standard auction, in the following sense:

Assumption 4 *The auction rules are such that the good is always assigned to the bidder with the highest value, and the lowest valuation bidder expects to pay nothing.*

This is a condition satisfied by the English auction, the sealed bid first-price auction, the second-price auction, and also by the all-pay auction.[4]

The method uses the central result in auction theory, the expected revenue equivalence theorem (Vickrey, 1961; Myerson, 1981):[5]

Theorem 1 (Expected Revenue Equivalence) *Under assumptions 1, 3 and 4, the expected payment for the good in auction l is $E[V_{(2:n_l)l}]$.*

Here the expectation is taken with respect to all information that is publicly available at the time of the auction; therefore the expected revenue equivalence theorem establishes that $E[p_l|X_l, Z_l, n_l] = E[V_{(2:n_l)}|X_l, Z_l, n_l]$.

[4] In the case of an all-pay auction, p_l should be interpreted as the sum of the prices payed by all bidders, rather than the amount payed by the winner.

[5] In what follows the notation $x_{(k:n)}$ represents the kth-highest order statistic from an i.i.d. sample of n observations of x.

3 Estimation

In this section, we propose two alternative estimation procedures. The first one requires information about F, that is, knowledge of the valuation distribution up to location and scale. The second method is applicable when F is unknown.

3.1 Estimation when F is known

Suppose a dataset of auctions is available with information about the final selling price p_l, the number of bidders n_l and covariates $[X_l, Z_l]$. Then under the assumptions made we can write

$$\begin{aligned} E[p_l|X_l, Z_l, n_l] &= E[V_{(2:n_l)}|X_l, Z_l, n_l] \\ &= E[\mu_l + \sigma_l \epsilon_{(2:n_l)}|X_l, Z_l, n_l] \\ &= \mu_l + \sigma_l E[\epsilon_{(2:n_l)}], \end{aligned}$$

where the first equality is due to the expected revenue equivalence theorem, the third by linearity of the conditional expectation, and the second is an useful property of order statistics: When an increasing affine transformation is applied to all variables in a sample, the order statistics of the new sample are the same affine transformation of the original order statistics. Paarsch *et al.* (2006) use this property to propose an OLS estimator for μ_l and σ_l in the context of a second-price auction, since then the winning price is the second-order statistic itself.[6]

Defining

$$a(n) \equiv E[\epsilon_{(2:n)}] = n(n-1) \int tF(t)^{n-1}(1 - F(t))dF(t),$$

[6] This property has also been discussed and utilized by Thiel (1988) in the context of a common values auction. Thiel's objective was to obtain for estimation purposes a linear relationship between two different objects, a bidder's own signal and the bid. This procedure was corrected by Levin and Smith (1991) and applied by Paarsch (1992).

we obtain

$$E[p_l|X_l, Z_l, n_l] \quad = \quad X_l\beta + a(n_l)Z_l\alpha. \tag{3}$$

Note that the conditional expectation of the winning bid is *linear* in β and α. This means that OLS is an unbiased, consistent estimation method for these coefficients. This observation gives rise to a straightforward procedure to estimate β and α when F is known:

Method 1 Using the standardized value distribution F, compute $a(n_l)$ for all values of n_l in the sample. Construct the set of regressors $[X_l, a(n_l)Z_l]$, and run OLS of the observed winning bids on these regressors.

It is important to notice that method 1 is not the same as introducing n_l as an additional regressor, as it is sometimes done in the literature, but rather introducing $a(n_l)$, a nonlinear function of n_l. In fact, ignoring the nonlinearity is *never* correct:

Proposition 1 *There is no non-degenerate distribution F such that $a(n)$ is affine in n.*

Because the proof of this proposition depends on tools that will be developed in section 4, it will be deferred to appendix B.

The first step method 1 requires the computation of $a(n) = E[\epsilon_{(2:n)}]$. Values of $a(n)$ are shown for some standardized distributions from selected location-scale families in table 1 below. By standardizing we mean working with parameter values that lead to a distribution with mean 0 and variance 1; thus in table 1 values labeled "uniform" are from the $U[-\sqrt{3}, \sqrt{3}]$ distribution; "logistic" are from the distribution $F(t) = e^{\pi t/\sqrt{3}} / \left(1 + e^{\pi t/\sqrt{3}}\right)$ (that is, the logistic distribution with scale parameter $\sqrt{3}/\pi$); "Laplace" is for the Laplace or double exponential distribution with location 0 and scale $1/\sqrt{2}$, with density $f(t) = e^{-|t|\sqrt{2}}/\sqrt{2}$; and "Gumbel" is the Gumbel or Extreme Value Type I distribution with location parameter $-\sqrt{6}\gamma/\pi$ and scale $\sqrt{6}/\pi$, with $F(t) = e^{-e^{-\pi\, t/\sqrt{6}-\gamma}}$, where $\gamma = 0.577216\ldots$ is the Euler-Mascheroni constant.

Of these five distributions, in four $a(n)$ can be written in closed form for all $n \geq 2$:

- For the uniform distribution, $a(n) = \sqrt{3}\,\frac{n-3}{n+1}$.
- For the logistic distribution, $a(n) = \frac{\sqrt{3}}{\pi}\left(\sum_{k=1}^{n-2}\frac{1}{k} - 1\right)$.
- For the Laplace Distribution, $a(n) = n\,\omega_{1:n-1} - (n-1)\,\omega_{1:n}$, where

  $$\omega_{1:n} = \frac{n}{2\sqrt{2}}\sum_{k=0}^{n-1}\frac{\prod_{t=1}^{k}(t-n)}{2^k k!(k+1)^2} - \frac{1}{n2^n\sqrt{2}},$$

  for all $n \geq 1$ (including $\omega_{1:1} = 0$).
- For the Gumbel or extreme value type I distribution,

  $$a(n) = \frac{\sqrt{6}}{\pi}[n\log(n-1) - (n-1)\log(n)].$$

For the normal distribution, closed forms for $a(n)$ exist only for $n \leq 5$: $a(2) = -1/\sqrt{\pi}$, $a(3) = 0$ (Jones, 1948), $a(4) = \frac{3}{2\sqrt{\pi}}\left(1 - \frac{6}{\pi}\arcsin(1/3)\right)$ and $a(5) = \frac{5}{2\sqrt{\pi}}\left(1 - \frac{6}{\pi}\arcsin(1/3)\right)$ (Godwin, 1949). For higher values of n numerical integration can be used. The results reported in table 1 have been computed using *Mathematica*'s `NIntegrate` algorithm. Tabulations with more accuracy and for higher values of n can be found in Harter (1961) (5 decimal places, n up to 100), Yamauti *et al.* (1972) (10 places, n up to 50) and Parrish (1992) (25 places, n up to 50).

The method can be applied with an arbitrary distribution F with mean 0 and variance 1. In this case numerical integration to evaluate $a(n)$ will be

n	Uniform	Normal	Logistic	Laplace	Gumbel
2	-0.57735	-0.56419	-0.55133	-0.53033	-0.54044
3	0	0	0	0	-0.09184
4	0.34641	0.29701	0.27566	0.24307	0.18367
5	0.57735	0.49502	0.45944	0.40511	0.38495
6	0.74231	0.64176	0.59727	0.53033	0.54410
7	0.86603	0.75737	0.70754	0.63419	0.67588
8	0.96225	0.85222	0.79943	0.72373	0.78842
9	1.03923	0.93230	0.87819	0.80278	0.88665
10	1.10221	1.00136	0.94710	0.87369	0.97383
11	1.15470	1.06192	1.00836	0.93807	1.05219
12	1.19911	1.11573	1.06350	0.99703	1.12336
13	1.23718	1.16408	1.11362	1.05144	1.18857
14	1.27017	1.20790	1.15956	1.10196	1.24872
15	1.29904	1.24794	1.20197	1.14910	1.30456
16	1.32451	1.28474	1.24135	1.19330	1.35665
17	1.34715	1.31878	1.27811	1.23489	1.40548
18	1.36741	1.35041	1.31256	1.27418	1.45142
19	1.38564	1.37994	1.34500	1.31139	1.49480
20	1.40214	1.40760	1.37563	1.34675	1.53590

Table 1: Values of $a(n)$ for selected distributions.

required. Unbiasedness of the least squares regression requires this integration to be accurate, as computational error would lead to error-in-variables bias.

Laffont *et al.* (1995) deal with this problem by modifying the objective function of their estimator to compensate for the computational error bias. In our case the issue is less problematic since 1) closed form solutions for $a(n)$ exist in some cases; 2) even when they do not exist, the computation of $a(n)$ must be done only once, and therefore it is possible to require a high level of accuracy. (In the context of Laffont *et al.* (1995), the computation must be redone for every evaluation of the objective function, and therefore accuracy is much more costly).

Assuming the hypothesis made about F is the correct one and the calculation of $a(n)$ is made without errors, the estimators from method 1 have the standard properties of least squares estimators with non-spherical (and non-normal) disturbances; they are unbiased and $\sqrt{L}$-consistent, but they are not efficient.

Even with independence across auctions, there is heteroskedasticity across auctions with different numbers of bidders and efficiency can be gained with generalized least squares. For some distributional families, closed form solutions for the variance of $\epsilon_{(2:n)}$ are available; for example, for the standardized Gumbel distribution, $\mathrm{Var}(\epsilon_{(2:n)}) = 1 - \frac{6}{\pi^2} n(n-1)[\log(n) - \log(n-1)]^2$. If σ_l is constant across auctions, the variance of the residuals in method 1 regression is a multiple of $\mathrm{Var}(\epsilon_{(2:n)})$, and thus weighted least squares can be done with little additional computational cost.

Also, once F is imposed the model is fully specified, and thus maximum likelihood (ML) is available. Maximum likelihood estimators for the symmetric independent private values framework have been studied by Donald and Paarsch (1996). For first-price auctions, the support of the bid distribution depend on parameters; as Donald and Paarsch (2002) put, "in such situations the maximum-likelihood estimator is often difficult to calculate and usually has a nonstandard limiting distribution that depends on nuisance parameters" (p. 305). Donald and Paarsch (1993) and Donald and Paarsch (2002) propose alternative estimation procedures that are computationally easier and have simpler asymptotic properties.

Similarly, the method proposed here trades off computational simplicity for efficiency. Appendix A provides evidence from Monte Carlo experiments that compare the OLS estimators with maximum likelihood in two simple settings where the ML estimator is computationally less involved. In both cases, there is a loss of efficiency from using OLS, as expected. OLS estimates for the mean μ tend to have acceptably low variances, in fact smaller than the variance of the estimator a researcher would obtain observing a sample of L valuations directly. The same is not true for the estimator of σ, which is unbiased, but imprecise. This is confirmed with field data in section 6.

3.2 Estimation when F is not known

The method proposed in the previous section requires introducing an appropriately chosen nonlinear function of the number of bidders as an additional regressor, and the appropriate function depends on the shape of the valuation distribution. This is a drawback since there may not be a priori information about F. In that case, it is possible to substitute dummy variables to flexibly estimate the unknown function $a(n)$ in the conditional expectation of p_l.

For every number of bidders k observed in an auction in the sample, construct a dummy variable d_{kl} for the event $n_l = k$. We consider two situations in turn: The homoskedastic case where $\sigma_l = \sigma$, and then the general case where σ_l is a function of variable regressors as well.

When $\sigma_l = \sigma$ (that is, Z_l contains only a constant), writing $m_l = X_l\beta = \beta_0 + x_l\beta_1$, we can write the conditional expectation of p_l in terms of these variables as follows:

$$
\begin{aligned}
E[p_l|X_l, \{d_{kl}\}] &= x_l\beta_1 + \sum_k d_{kl}[\beta_0 + \sigma a(k)] \\
&= x_l\beta_1 + \sum_k d_{kl}\delta_k,
\end{aligned}
$$

where the summation over k is over the empirical support of n_l and $\delta_k = \beta_0 + \sigma a(n)$. Again we obtain a linear relationship that can be estimated using OLS. The OLS estimates allows us to recover all β coefficients except the intercept. Both σ and the intercept β_0 will be absorbed in the dummy coefficients, and thus cannot be identified.[7]

When Z_l contains other variables, the conditional expectation will involve interactions of all these variables (including the constant) with the d_{kl}. Adding those in a linear regression is enough to obtain unbiased estimates of the coefficients for regressors that appear on X but not on Z.

To obtain estimates for the other coefficients without imposing a distributional choice for the valuations, nonlinear least squares are required. Let Y be the set of regressors common to both X and Z, and let x and z be the sets of regressors that appear on X but not on Z and vice-versa. Write $\beta = (\beta_1, \beta_2)$ so that $X\beta = Y\beta_1 + x\beta_2$, and likewise for α. Then

$$
\begin{aligned}
E[p_l|X_l, Z_l, \{d_{kl}\}] &= x_l\beta_2 + \sum_k d_{kl}Y_l[\beta_1 + a(k)\alpha_1] + \sum_k d_{kl}z_l[a(k)\alpha_2]; \quad (4) \\
&= x_l\beta_2 + \sum_k d_{kl}Y_l\delta_{k1} + \sum_k d_{kl}z_l\delta_{k2}.
\end{aligned}
$$

The model imposes nonlinear restrictions on $N \times (A + B)$ of the "reduced

[7] If the researcher is willing to impose the condition that F is symmetric, then β_0 can be identified in the following way: for any symmetric distribution F (with finite expectation), $a(3) = 0$. Thus, the coefficient of d_{3l} is $\delta_3 = \beta_0 + \sigma a(3) = \beta_0$.

form" coefficients obtained in this regression, where N is the number of different observed values for n_l, A is the number of regressors that appear both in X and Z, and B is the number of regressors that appear only in Z. Since there are $2A+B+N$ parameters to be estimated (namely α_1, α_2, β_1 and $a(n)$), if $N \geq 4$ the model will be identified.[8]

We thus propose the following method:

Method 2 For every number of bidders k observed in an auction in the sample, construct dummy variables d_{kl} for the event $n_l = k$.

- To obtain estimates of β parameters for the set of regressors x_l in X_l that do not appear in Z_l, run OLS of p_l on x_l and all interactions of Z_l and the dummies d_{kl};
- To obtain estimates of the other parameters, run nonlinear least squares of p_l on the function in the right hand side of equation 4, treating $a(k)$ as parameters to be estimated.

An interesting feature of method 2 is that it may be possible to obtain estimates of moments of order statistics $\hat{a}(n)$ of the underlying distribution F. This information can be used for inference about F. Section 4 discusses this possibility and section 6 provides a practical illustration.

4 Identifying Distributions from Least Squares Coefficients

As mentioned in the previous section, with method 2 it may be possible to obtain estimates for the $a(n)$'s, the first moments of the second order statistics of the F distribution. This section discusses ways to explore this information. It

[8] A notable exception is the previous case where Z contains only a constant, since then $B = 0$ and $A = 1$, and we have N equations and $N + 2$ unknowns.

establishes that knowledge of these moments (for all n) is enough to fully identify the valuation distribution: in principle, it is possible to obtain a nonparametric estimator of the valuation distribution from the estimated coefficients of a least squares regression!

It also provides a straightforward test for specific distributional assumptions. The methodology is illustrated in the next section with an application to a dataset of auctions of iPod minis from eBay.

4.1 Full Identification of F from $\{a(n)\}$

A probability distribution can be fully identified from knowledge of its $a(n)$'s. This has been shown in several versions in the Statistics literature (Hoeffding, 1953; Chan, 1967; Pollak, 1973). Here a constructive proof is provided, that directly shows how to compute F from $\{a(n)\}_{n=2}^{\infty}$.

Theorem 2 *Suppose that F has a finite expectation. Then there is a one-to-one mapping between F and the sequence $\{a(n)\}_{n=2}^{\infty}$.*

The overall strategy of the proof follows Pollak (1973): the construction of F from $a(n)$ will be made in two steps, that we state as lemmas.

Lemma 1 (Recurrence relation) *Let $\omega_{(k:n)} = E[\epsilon_{(k:n)}]$, for all $n = 2, 3, \ldots$ and all $k = 1, 2, 3, \ldots, n-1$. Then the following recurrence relation holds:*

$$\omega_{(k:n-1)} = \frac{n-k}{n}\,\omega_{(k:n)} + \frac{k}{n}\,\omega_{(k+1:n)}.$$

Proof: See appendix B. $\square$

Corollary 1 *Knowledge of $a(n) = E[\epsilon_{(2:n)}]$ for every $n = 2, 3, \ldots$ is sufficient to know all $\omega_{(k:n)} = E[\epsilon_{(k:n)}]$, for all $n = 2, 3, \ldots$ and all $k = 2, 3, \ldots, n$.*

Proof: See appendix B. □

The recurrence relation is valid for any random variable (with finite expectation). If the mean ($=\omega_{(1:1)}$) of it is known, then one can further compute $\omega_{(1:n)}$ using the recurrence formula. However, this will not be needed in what follows.[9]

Another remarkable fact is that the same recurrence relation is valid for the expectation of any measurable function g of the order statistics: $E[g(\epsilon_{(k+1:n)})] = \frac{n}{k}E[g(\epsilon_{(k:n-1)})] - \frac{n-k}{k}E[g(\epsilon_{(k:n)})]$. So the same lemma applies to other moments, such as the variance of the order statistics.[10]

Lemma 2 *Let $\{k_n\}$ be a sequence of integers such that $1 - k_n/n \to \alpha \in [0,1]$. Then $\epsilon_{(k_n,n)} \to F^{-1}(\alpha)$ in probability.*

Proof: See appendix B. □

Proof of Theorem 2: To obtain the quantile $F^{-1}(\alpha)$ for any $\alpha \in [0,1]$, select a sequence $\{k_n\}$ such that $k_n \geq 2$ and $k_n/n \to 1-\alpha$. Use the recurrence formula to compute $\omega_{(k_n:n)}$ from the $a(n)$'s, and then take the limit. The converse is immediate. □

In principle Theorem 2 provides a way to obtain a non-parametric estimate of the distribution of F by ordinary least squares. From method 2 we obtain estimators $\hat{a}(n)$ that are consistent for $a(n)$. Since the recurrence formula is linear, the corresponding $\hat{\omega}_{(k:n)}$ are also consistent (for a fixed n, as the number of auctions goes to infinity). Finally, for large n, from lemma 2 the expectation of these estimators converges to the quantiles of the original distribution.

[9] The recurrence relation can also be used to facilitate computation of $a(n)$; for example, see the closed form solution for the Laplace distribution in section 3.1.

[10] Another application of the recurrence relation in Economics can be found in Athey and Haile (2002). Using the relation with $g(x) = \mathbb{1}\{x < t\}$ one obtains a recurrence relation between distributions of order statistics. Athey and Haile (2002) use this relation to investigate ways of testing for the winner's curse.

However, the linear combinations that arise from the recurrence formula have large coefficients of opposing signs. Therefore the variance of $\hat{\omega}_{(k:n)}$ can be very large even if the $\hat{a}(n)$ are estimated precisely. So this estimation strategy is likely to perform very poorly even in a very large dataset.

To see that, using the recurrence formula the following closed form expression for $\omega_{(k:n)}$ can be obtained:

$$\omega_{(k:n)} = \frac{n!}{(n-k)!(k-1)!} \sum_{j=0}^{k-2} (-1)^{k-j-2} \frac{(k-2)!}{j!(k-2-j)!} \frac{a(n-j)}{(n-j)(n-j-1)}.$$

$\omega_{(k:n)}$ is a linear combination of the $k-1$ previous $a(n)$'s, with coefficients that alternate signs. As k and n grow, not only the number of terms in the sum grow, but also do the coefficients. This means that the estimator for $\omega_{(k:n)}$ is likely to have a variance too large to be of practical use.

Consider for example the case where $\mu_l = 0$ and $\sigma_l = 1$. In this case the estimators for $\hat{a}(n)$ are means of different auction subsamples, and are therefore independent. In this case, the variance of $\hat{\omega}_{(k:n)}$ is simply $\mathrm{Var}(\hat{\omega}_{(k:n)}) = \sum_{j=0}^{k-2} c_{jkn}^2 \mathrm{Var}(\hat{a}(n-j))$, with $c_{jkn} = \frac{n!}{(n-k)!(k-1)!} \frac{(k-2)!}{j!(k-2-j)!(n-j)(n-j-1)}$.

Suppose we are interested in estimating the median of the valuation distribution, and for that we use $\hat{\omega}_{(n/2:n)}$ for some large choice of n. Take the first coefficient c_{0kn}. For $k = n/2$ we find

$$\begin{aligned} c_{0kn} &= \frac{n!}{(n-k)!(k-1)!n(n-1)} = \frac{(n-2)!}{(n-k)!(k-1)!} \\ &\simeq \frac{n!}{((n/2)!)^2} \underset{n\to\infty}{\longrightarrow} \infty. \end{aligned}$$

So the variance of $\hat{\omega}_{(n/2:n)}$ would be arbitrarily large for large values of n even if we had a sample that allowed us to estimate $a(n)$ well. For this reason, one should not expect to obtain accurate estimates of quantiles of F using this method.

However, theorem 2 indicates that there is a significant amount of information in $\{a(n)\}$; in particular, about aspects of the shape of the upper tail that are of interest to study revenue and surplus issues. The next section discusses a practical way to explore this information.

4.2 Hypothesis Testing about Distributions

The information contained on the $\hat{a}(n)$ estimators can be used to test for the hypothesis of a specific distribution F.

Under the null hypothesis that a given F is the standardized valuation distribution, method 1 would be appropriate. Under the alternative hypothesis, method 2 would be. But the regression by method 1 is a restricted version of the regression in method 2, where the $a(n)$ terms are required to follow the shape specified by F. So one can simply test the hypothesis by running both methods and applying an F-test on the R^2 difference.[11] This procedure is illustrated in section 6.

5 Other settings

The presentation of the method was done under the assumption of the benchmark model in auction theory, the independent private values framework with symmetric bidders and exogenous participation. In many applications, some or many of the underlying hypotheses in this model do not hold. This section

[11] If F is known up to a family of distributions indexed by a set of parameters θ, a similar idea can be used in a two step procedure to estimate θ.

Assume that ϵ_{il} is drawn from a distribution $F \in \{F_\theta\}_\theta$, where $\{F_\theta\}_\theta$ is a family parameterized by θ. We are interested in finding θ such that $F = F_\theta$. Then, for every n observed in the sample, we can use the moment condition

$$a(n) = \frac{n!}{(n-2)!} \int x F_\theta^{(n-2)} (1 - F_\theta) dF_\theta$$

for each n to run a GMM method using as the left hand side the estimates that come from method 2. Further research is needed to investigate if this approach is feasible and yields good estimators in small sample.

briefly discusses to what extent the method works when applied in more general contexts.

Given that the presentation relied on the expected revenue equivalence theorem, it may come as a surprise that the method is to some extent applicable even in contexts where the theorem does not hold. The key to the method is a general property of auction games: when an affine transformation is applied to all valuations, then the new equilibrium bids will be an affine transformation of the original equilibrium bids.

More formally, consider the class of auctions where a bidder action can be represented by a number b_i (i's "bid"). Let $b = (b_1, \ldots, b_n)$. Given an auction rule, let $W_i(b)$ be i's probability of winning the item given that bidders in the auction played b, and let $P_i(b)$ be the expected payment given b, conditional upon winning. Note that these are functions of the realized bids of all participants, and thus are determined solely by the auctions rules and not by the data generating process of the valuations. For example, in a first-price auction, $W_i(b) = 1$ if b_i is the sole highest bid and 0 if b_i is not the highest bid[12] and $P_i(b) = b_i$.

Let v_i be i's valuation, not necessarily independent or symmetric. Let $v = (v_1, \ldots, v_n)$, and let μ and $\sigma > 0$ be constants. Then we have the following proposition, that generalizes results found in Bajari and Hortaçsu (2003), Deltas and Chakraborty (2001) and Krasnokutskaya (2002):

Proposition 2 *Suppose bidders are risk neutral and the auction rule is such that (i) $W_i(\mu + \sigma b) = W_i(b)$, (ii) $P_i(\mu + \sigma b) = \mu + \sigma P_i(b)$, for all i and (iii) losing bidders pay nothing. Then, if b^* is a Nash equilibrium of an auction when bidders have valuations v, then $\mu + \sigma b^*$ is a Nash equilibrium of the auction when valuations are $\mu + \sigma v$. As a consequence, if the expected selling price in*

[12]If b_i is tied for highest bid $W_i(b)$ is the probability of i winning as determined by the auction tie-breaking rule.

the former auction is p^*, then the expected price in the latter auction will be $\mu + \sigma p^*$.

Proof: For b^* to be a Nash equilibrium under v, it must be that it prescribes i to bid

$$b_i^* \in \underset{b_i}{\operatorname{argmax}}\, \pi(b_i, b_{-i}^*) = \underset{b_i}{\operatorname{argmax}}\, E[(v_i - P_i(b_i, b_{-i}^*))W_i(b_i, b_{-i}^*)],$$

where this expectation is taken to be conditional on all information available to i (including the contingency of winning).

In the game with valuations $\mu+\sigma v$, suppose the other bidders play $\mu+\sigma b_{-i}^*$. By following $\mu + \sigma b_i^*$ bidder i obtains

$$\begin{aligned}
\tilde{\pi}(\mu + \sigma b^*) &= (E[(\mu + \sigma v_i - P_i(\mu + \sigma b^*))W_i(\mu + \sigma b^*)] \\
&= E[(\mu + \sigma v_i - (\mu + \sigma P_i(b^*)))W_i(b^*)] \\
&= \sigma E[(v_i - P_i(b^*))W_i(b^*)] \\
&= \sigma \pi(b^*).
\end{aligned}$$

Thus, if b_i^* is a best response to b_{-i}^* in the former game, $\mu+\sigma b_i^*$ is a best response to $\mu + \sigma b_{-i}^*$ in the latter game.

As for the expected selling price, this is a consequence of this result combined with the affinity of P_i. $\square$

The three conditions for the proposition are satisfied by standard auction rules. The condition on W_i is satisfied by any rule that assigns a winner based on a ordering of the bids.[13] The condition on P_i is also very easily met, as typically the payment for the winner is either the bidder's own bid (as in a first

[13] It is important to point out that this assumption implicitly imposes a restriction on reserve prices. For W to have the desired property, either the reserve price r must be trivial (in the sense that the probability of v_i and $\mu + \sigma v_i$ be below r is zero) or it must change along with v (so that in the second auction the reserve price is $\mu + \sigma r$).

price auction) or somebody else's bid (as in the Vickrey or English auctions). The condition on losing bidders paying nothing rules out all-pay auctions, but is otherwise usually satisfied.

Consider now a setting with three sets of exogenous variables: (X_l, Z_l, R_l). As before, let X_l and Z_l be regressors that affect the location and scale of bidder valuation. Let R_l be variables that capture all changes in conditions across auctions: for example, changes in auction rules and variation in participation levels (both total number of bidders and the presence of specific bidders if their valuations are asymmetric). Let $V_l = (V_{1l}, \ldots, V_{n_l l})$ be the vector of valuations for bidders in auction l. We impose the following generalization of assumption 1:

Assumption 5 $V_l = \mu_l + \sigma_l \epsilon_l$, *where* ϵ_l, *the vector of standardized valuations, conditional on* R_l *has the same distribution across auctions (and are independent across auctions).*

In other words, after a suitable standardization, the joint distribution of valuations is the same across auctions with the same characteristics R_l. Using proposition 2, we obtain

$$E[p_l|X_l, Z_l, R_l] = X_l\beta + Z_l\alpha\tilde{a}(R_l), \tag{5}$$

where $\tilde{a}(R_l) = E[\tilde{p}_l|R_l]$, where $\tilde{p}_l$ is the equilibrium selling price in an auction with rules R_l and standardized valuations (that, is, if $\mu_l = 0$ and $\sigma_l = 1$).

Conceptually, this is a direct generalization of equation 3; in practice, however, what is lost is the easy statistical interpretation of the additional regressor $a(n_l)$ that is possible when the expected revenue equivalence theorem holds. To compute $\tilde{a}(R_l)$, one needs to fully specify the auction game being played, then solve for the equilibrium bids. Thus, generalizing method 1 becomes computationally burdensome.

It is possible to generalize method 2. What is needed is to add dummy variables for all interactions of variables in R_l rather than just the number of bidders n_l. Here are two examples:

As a first example, consider an affiliated valuations model, and suppose the joint distribution of valuations across auctions varies only in location and scale. Because valuations are not independent, the expected price in the auction depends not only on the number of bidders but also on the specific auction rule utilized; for example, if it is an English auction the expected price will be higher than in a first-price auction. In this case, the successful application of method 2 will require a set of dummies that interact the number of bidders with the particular auction rule used, if it changes across the sample.

Another example would be a situation where valuations are independent, but asymmetric: there are two types of bidders with valuation from different distributions. Now "location" and "scale" coefficients are interpreted as follows: When evaluated at the true parameters, $\mu_l = X_l\beta$ and $\sigma_l = Z_l\alpha$ are such that for all bidders i of the first type, $(V_{il} - \mu_l)/\sigma_l$ is i.i.d. F_1 and for bidders of the second type, $(V_{il} - \mu_l)/\sigma_l$ is i.i.d. F_2, $F_1 \neq F_2$ (Note that this definition is only meaningful with variable regressors: the intercepts of μ_l and σ_l are not identified). In this case one would need all interactions of numbers of bidders for each separate type, plus dummies for changes in the auction rules, since in this case the expected revenue equivalence theorem is invalid as well. Appendix A shows a Monte Carlo experiment involving asymmetric auctions with bidders coming from three different distributions.

In general, two important requirements should be met for this approach to be valid. First, coefficients that can be directly estimated have to impact *all* bidders in the same way. In an asymmetric auction setting, for example, it is not possible to have different coefficients in the location specification for different bidder types.[14] This makes the method unappealing if the focus of the research is investigating the nature of bidder asymmetry.

Second, R_l should be exogenous, and must contain all changes in rules and settings that modify the game being played. If equilibrium behavior is affected by any variable that is observed by bidders but is absent from R_l, then estimates from the method will be biased from spurious correlation between X_l or Z_l and this variable. An important practical setting where this problem may occur (even in the symmetric independent private values case) is when the number of bidders is not observed or is not exogenous.

In conclusion, it is possible to apply method 2 in some circumstances outside the symmetric independent private values environment, but one should keep in mind that the approach is just a way to obtains estimates for location and scale coefficients controlling for "auction effects", not a way to investigate properties of auctions in wider settings. In the example reported in appendix A, the coefficient of a exogenous regressor on μ_l can be estimated, but the coefficients of the "auction controls" are very difficult to interpret, and it is not obvious how to use this information for inference on the nature of the asymmetry between bidders. It would also be impossible to apply the method without information on the number of bidders of each type.

6 Illustration: iPod mini auctions in eBay

This section illustrates the methods proposed in this paper using a sample of online auctions for Apple iPod mini players. The iPod is a popular portable device that stores and plays music files, as well as store other electronic data.[15] This product is a convenient choice for this illustration for two reasons: First,

[14] This also implies that reservation prices move along with valuations as regressors change.

[15] Newer models of iPod can also display pictures and play video.

because of its success and size the online market for iPods is very liquid; currently in eBay alone there are several thousand listings of Apple iPods. Second, since the design of all iPods is under the control of the same company, the amount of product heterogeneity is limited.

Out of the several different types of iPods, the analysis here will focus on the iPod mini. Compared to the original iPods, the minis provide the same capability, but are smaller and come in several different colors.[16] At the time the data was collected iPod minis were no longer actively promoted by Apple, but were still commonly found in stores and in online markets such as eBay. It was a mature product in the technology adoption life cycle, in the sense that its characteristics were already well-known by the public, but it was not yet perceived as obsolete.

There is both a vertical and a horizontal dimension of differentiation among versions of the iPod mini. The vertical element (among new devices) is memory size, that can be 4 Gb or 6 Gb. The horizontal element is color (Silver, Blue, Green or Pink). In eBay, a significant fraction of the market is second-hand; even non-working devices are sold, often to be harvested for parts or used as a learning tool by technicians. This introduces an obvious additional source of differentiation. There may also be differentiation among auctions due to the seller characteristics, such as reputation.

6.1 Data

Data were collected from 1225 completed eBay auctions from June 27 to July 18, 2006. The data includes only auctions listed under the Apple iPod mini category in eBay. Auctions for accessories or parts as well as lots of devices were not included. However, the sample does include devices described as defective. 225 auctions with one single bidder we also excluded, since the methodology

[16] Apple later released smaller devices, the iPod shuffle and the iPod nano.

does not apply in this case.[17]

While the majority of auctions in the sample has a symbolic posted reserve price,[18] some sellers elect to place a nontrivial reserve price. Analysis of the latter auctions is problematic, since we do not observe neither the bids nor the number of bidders below the reservation price.[19] We elected to focus attention only on auctions with a posted reserve price of US$ 17 or less. As figure 1 shows, this avoids auctions with reserve prices around $20 where the truncation effect starts to be visible. Fortunately, the restricted sample still contains the majority of the observations, with 654 auctions.

For each auction we collected the final price, the number of bidders (defined as the number of different buyers that placed bids), product characteristics such as color, memory size and condition (New, Used, Refurbished or other), and seller feedback statistics. In addition to stated condition, a dummy variable BROKEN was created for items described as either non-operational or with a serious defect. Usually such defect is short battery life, but includes issues such as a broken screen or headphone jack (which makes the device useless for its intended purpose). We did not classify as broken devices with scratches and other defects of cosmetic nature.

In eBay, after every transaction the parties involved are asked to provide feedback on each other. Two statistics are widely reported: Feedback score, which is the number of positive feedback ratings minus the negatives, and the positive feedback percentage. Instead of using these figures directly as regres-

[17]In eBay, completed auctions with a single bid are almost always due to "Buy It Now" bids. A popular auction rule in eBay allows sellers to post a "Buy It Now" price, at which any buyer can purchase the item *before* bidding starts. Reynolds and Wooders (2003) provide a theoretical analysis of this auction rule. The "Buy It Now" price is not a reserve price; if taken, it immediately ends the auction. Furthermore, once regular bidding starts, the "Buy It Now" option is no longer available, and thus does not affect bidding in the auctions in the sample.

[18]Confusingly, in eBay open reserve prices are called "starting bids".

[19]Note that the same problem does not exist for hidden reserve prices, since for those participation below the reserve is public information.

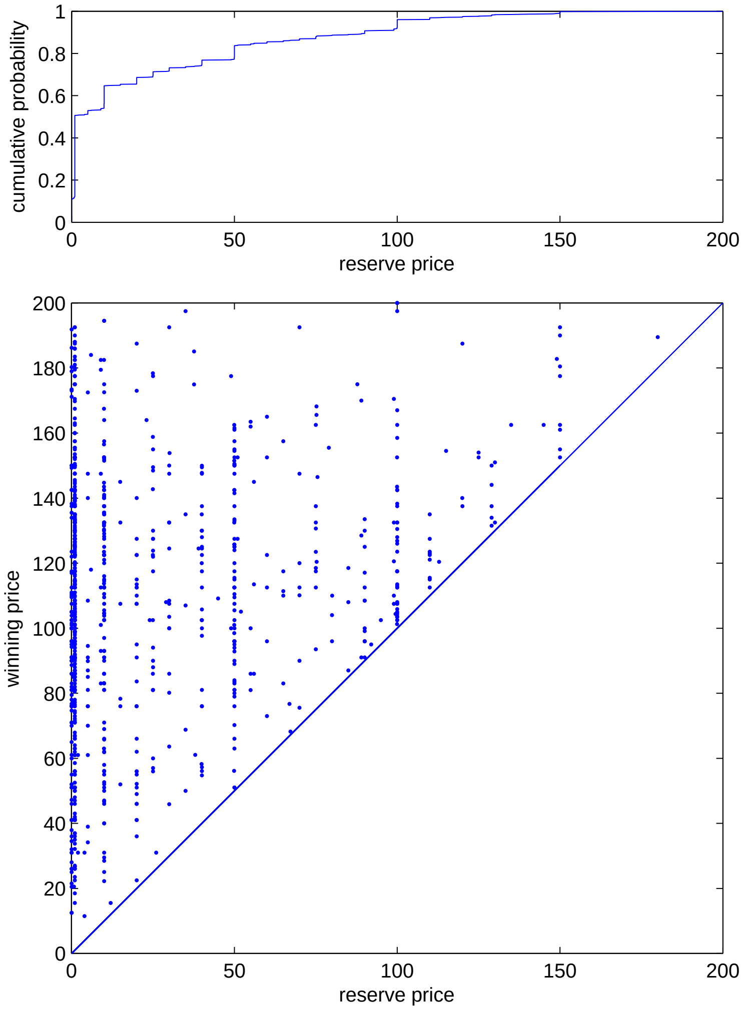

Figure 1: Effect of reservation price.

sors, we used POS $= \log_2(P+1)$ and NEG $= \log_2(N+1)$, where P and N are the numbers of positive and negative ratings not withdrawn.

6.2 Results

We report four specifications involving both indicators of product characteristics and seller reputation. Specification I simply assumes that both μ and σ are constant. Specification II allows μ to be a function of several characteristics that might affect willingness to pay, namely BROKEN, dummies for used, new and refurbished condition,[20], dummies for color[21] and the number of Gb of memory. In addition, POS and NEG have been included in order to evaluate the impact of seller reputation in the bidders' average willingness to pay. In specification III the variables found to be insignificant in specification II are dropped. Finally, in specification IV BROKEN, NEW, POS and NEG are allowed to affect both μ and σ.

For each of the specifications, regressions were ran using method 2 and method 1 for each of the five distributions used in table 1. Table 2 reports the results of tests on the shape of F. In all cases, the Gumbel distribution provides a better fit. There is a theoretical reason to favor this distributional choice as well, as it is one of three possible limiting distributions of maxima of random variables. Suppose valuations in the general population are independent and identically distributed with an arbitrary distribution that satisfies a right tail condition;[22] split the general population into several independent subgroups of the same size, and let bidders that go to eBay be those with the highest value within each group. Then according to the the Fisher-Tippett-Gnedenko Theorem, the distribution of the bidders' valuations will be approximately Gumbel.

[20] The omitted dummy variable is no condition reported.

[21] Namely, silver, blue, pink and green. Again, the omitted category is no color reported.

[22] Namely, let $G(x)$ be the distribution in question; then a function $b(x)$ should exist such that, for all $t > 0$, $\lim_{x\to\infty} \frac{1-F(x+tb(x))}{1-F(x)} = e^{-t}$.

For both reasons, we select the Gumbel distribution and report the coefficients obtained for each specification under this assumption in table 3. The coefficients obtained under other distributional assumptions are similar and have been omitted for brevity. The figures in parentheses are heteroskedasticity-consistent standard error estimates.

Coefficients on the mean valuation have the expected sign. In specification II, we find that broken iPods are worth US$ 39 less, while iPods reported to be new are worth US$ 60 more. Other condition variables and color are not significant, and were dropped in specification III.

Seller feedback has a significant effect on μ: Roughly every doubling of positive ratings increases the buyers' willingness to pay by US$ 2. A doubling of both positive and negative ratings leads to a *decrease* of US$ 1.20.

Why would that be? It is obvious why buyers are willing to pay more for sellers with positive feedback and less for sellers with negative feedback, but it is less clear why the latter effect is greater. In a simple framework of Bayesian updating, increasing both POS and NEG amount to more precise information about the odds of a bad deal, and this is not expected to make buyers less willing to buy.

Specification IV suggests an explanation for this puzzle. In it, POS and NEG (as well as BROKEN and NEW) are allowed to affect valuation dispersion as well as location. Point estimates for *both* POS and NEG are negative, albeit insignificant. This suggests an explanation for why longer feedback histories seem to be worse: as POS and NEG grow, bidders obtain more information about the prospect of a good deal, and as a result their valuations become less dispersed (presumably because their heterogeneous beliefs converge). As a result, the transaction price decreases. Omitting this effect introduces a bias in the estimated mean effect. When this effect is accounted for, we find virtually identical effects of POS and NEG in μ, which is compatible with a theory where bidders care only about the odds of a bad deal.

Similar biases have been found for the BROKEN and NEW variables. There is more uncertainty about the value of broken iPods and the there is less uncertainty about iPods classified as new. This leads to substantial biases on the coefficients on the mean when this effect is omitted; in specification IV we find that broken iPods are worth US$ 5 less and new iPods are worth US$ 13 more than previously estimated.

While suggestive, point estimates of coefficients in specification IV are for the most part not significant. This is probably due to near multicollinearity for regressors used both for μ and σ. This combined with the findings from the Monte Carlo experiments suggests that the method may lead to imprecise estimates for rich specifications of σ.

As discussed in section 3.1, theory predicts heteroskedasticity of a form that depends on the F distribution. In the case of a standardized Gumbel distribution, the variance of the $\epsilon_{(2:n)}$ is $1 - \frac{6}{\pi^2} n(n-1)[\log(n) - \log(n-1)]^2$, and the square root of this expression can be used as a weight to improve efficiency if $\sigma_l = \sigma$. Table 4 reports weighted least squares estimates for the specifications where this requirement is met. Here, conventional standard errors are reported. Results are similar to those reported in the previous table.

7 Concluding Remarks

This paper attempts to combine the strengths of two distinct branches of the empirical literature on auctions with the goal of obtaining a method at once computationally accessible and theoretically sound.

In a nutshell, the main finding of the paper is that one can estimate parameters that impact the location and scale of the value distribution of all bidders in a simple and unbiased way, provided one controls for variables that affect bidding behavior in a flexible way.

Under the framework investigated in this paper, the method provides a way to separate the effect of regressors that affect bidding through the value distribution from those that affect bidding strategically. It can be an useful tool for researchers interested in measuring the former while properly controlling for the latter; conversely, it helps identifying the misspecification bias that occurs when the latter are not accounted for.

The method is not designed for investigation of deeper aspects of the strategic interaction of players in general auctions. It is however an acceptable way to identify covariates that affect the location or scale of all bidder valuations. It can therefore be used as a tool to control for variation across auctions, such as product characteristics. As such, we hope the method might play a complementary role to existing structural methods, as a preliminary exploratory tool for researchers interested in applying more ambitious structural econometric models to auction data.

A Monte Carlo Experiments

We report three Monte Carlo experiments for the estimation method proposed in this paper. The first two have the objective of comparing its small sample performance with the maximum likelihood estimator in simple settings; the third experiment illustrates the applicability of method 2 to the context of asymmetric auctions.

In experiment one, we consider second-price independent private value auctions, where valuations are normally distributed with mean $\mu = 3$ and standard deviation $\sigma = 1$. Participation is exogenous and random: in each auction, the number of bidders n_l is drawn from the uniform distribution over $\{2, 3, 4, 5, 6\}$.

There is no reserve price.

We seek to estimate both μ and σ from samples with 50, 100 or 200 auctions where we observe the selling price p_l and number of bidders n_l. As discussed in Donald and Paarsch (1996), because this is a Vickrey auction the likelihood of the winning price is $h_B(p_l; n_l) = n_l(n_l - 1)\Phi\left(\frac{p_l-\mu}{\sigma}\right)^{n_l-2}\left(1 - \Phi\left(\frac{p_l-\mu}{\sigma}\right)\right)\frac{1}{\sigma}\phi\left(\frac{p_l-\mu}{\sigma}\right)$, where Φ and ϕ are the standard normal distribution and density. One can obtain (efficient) estimates maximizing $\Pi_l h_B(p_l, n_l)$ for μ and σ; alternative, one can run OLS of p_l on a constant and the artificial regressor for the normal distribution reported in table 1.

Table 5 reports statistics from 1000 replications of this experiment. Both estimators for μ show no bias and the variance of the least square estimator is around 50 to 60% larger. The loss of efficiency in the estimator for σ is much larger; this seems to be a general weakness of the method. On the other hand, $\hat{\sigma}_{LS}$ does not show the downward bias that is present in the maximum likelihood estimator.

Experiment two compares the least squares and maximum likelihood estimator in the context of a first-price auction. As discussed in Donald and Paarsch (1996), using maximum likelihood with first-price auction data is more complex for two reasons: first, computing the maximum likelihood is harder, since it may require solving a differential equation for each observation and each parameter evaluation. Second, the maximum likelihood estimators are not asymptotically normally distributed, and thus inference is not straightforward.[23]

In experiment two, we consider first-price auctions with independent private values drawn from a uniform distribution with mean $\mu = 3$ and standard deviation $\sigma = 1$.[24] As above, n_l is uniform between 2 and 6, sample size is 50, 100 or 200 and the number of replications is 1000.

[23] Donald and Paarsch (1993) introduce a technique, piecewise pseudo-maximum likelihood, that leads to asymptotically normal estimators for auction models.

[24] We used the uniform distribution to speed up the computation of equilibrium bids.

Results are reported in table 6. As above, least squares estimators are unbiased but less efficient than maximum likelihood with the efficiency loss being more pronounced in the estimator of σ. The proportional loss of efficiency between the OLS and the ML estimators becomes larger as the sample size grows; this may be due to ML being super-consistent in this context, converging at rate L rather than $\sqrt{L}$ (Donald and Paarsch, 1996). Results suggest that the ML estimator for the mean is biased downward. As predicted by Donald and Paarsch (1996), we find that the Jarque-Bera test strongly rejects normality of either maximum likelihood estimator, while the least squares estimators are not rejected to be normally distributed for a sample size of 50.

The third experiment illustrates the application of method 2 in the context of asymmetric auctions. We consider second-price independent private value auctions with three different types of bidders: bidders of the high type have values drawn independently from distribution $N(12 + X_l\beta, 3)$, bidders of the medium type from $N(11 + X_l\beta, 2)$, and bidders of the low type from $N(10 + X_l\beta, 1)$. X_l is a product characteristic that has the *same* effect β on all bidders. X_l is drawn i.i.d. $U[0, 5]$. The number of bidders is exogenous and random, with one or two of each bidder type per auction. The number of bidders of each type is observed, as it is necessary to apply the method. We seek to estimate $\beta = 1$.

We run method 2 in 1000 replications of samples of size 100. We added a dummy for each configuration of participation; in the table $\hat{\delta}_{ijk}$ is the coefficient for participation of i bidders of low type, j bidders of medium type and k bidders of high type. Results are shown in table 7. We find the method can readily estimate β. The estimated patterns for the control coefficients does not lend itself very easily to interpretation (for example, on average we obtain $\hat{\delta}_{112} > \hat{\delta}_{121} < \hat{\delta}_{211}$).

B Proofs

Proof of lemma 1: Since $\omega_{(k:n)} = \frac{n!}{(k-1)!(n-k)!} \int vF(v)^{n-k}(1-F(v))^{k-1}dF(v)$, we have that

$$\begin{aligned}
n\omega_{(k:n-1)} &= \tfrac{n!}{(k-1)!(n-1-k)!} \int vF^{n-1-k}(1-F)^{k-1}dF(v) \\
&= \tfrac{n!}{(k-1)!(n-1-k)!} \int vF^{n-1-k}(1-F)^{k-1}[F+(1-F)]dF(v) \\
&= \tfrac{n!}{(k-1)!(n-1-k)!} \int vF^{n-k}(1-F)^{k-1}dF(v) \\
&\quad + \tfrac{n!}{(k-1)!(n-1-k)!} \int vF^{n-1-k}(1-F)^{k}dF(v) \\
&= (n-k)\omega_{(k:n)} + k\omega_{(k+1:n)}
\end{aligned}$$

so the recurrence relation holds. □

Proof of corollary 1: By induction, since it is immediate that with all $\omega_{(k:n)}$ and $a(n+1)$, one can directly compute all remaining $\omega_{(k:n+1)}$. □

Proof of lemma 2: The argument roughly follows Hoeffding (1953). We will show that the distribution of $\epsilon_{(k_n,n)}$ converges to the constant $F^{-1}(\alpha)$.

Given a quantile u, $\Pr(\epsilon_{(k_n,n)} < F^{-1}(u))$ can be written as

$$\frac{\int_0^u (1-t)^{k_n} t^{n-k_n} dt}{\int_0^1 (1-t)^{k_n} t^{n-k_n} dt}.$$

We must show that this goes to 0 for $u < \alpha$ and to 1 for $u > \alpha$.

Take $u < \alpha$. Fix $v \in (u, \alpha)$. For a sufficiently high n, $\alpha_n = 1 - k_n/n > v$. The function $t^{\alpha_n}(1-t)^{1-\alpha_n}$ is increasing for $t < \alpha_n$; so

$$\frac{\int_0^u [t^{\alpha_n}(1-t)^{1-\alpha_n}]^n dt}{\int_0^1 [t^{\alpha_n}(1-t)^{1-\alpha_n}]^n dt} \leq \frac{\int_0^u [t^{\alpha_n}(1-t)^{1-\alpha_n}]^n dt}{\int_v^{\alpha_n} [t^{\alpha_n}(1-t)^{1-\alpha_n}]^n dt} \leq$$
$$\frac{\int_0^u [u^{\alpha_n}(1-u)^{1-\alpha_n}]^n dt}{\int_v^{\alpha_n} [v^{\alpha_n}(1-v)^{1-\alpha_n}]^n dt} = \frac{u}{\alpha_n - v}\left[\frac{u^{\alpha_n}(1-u)^{1-\alpha_n}}{v^{\alpha_n}(1-v)^{1-\alpha_n}}\right]^n \quad \to \quad 0.$$

The argument for $u > \alpha$ is analogous, since the function $t^{\alpha_n}(1-t)^{1-\alpha_n}$ is decreasing for $t > \alpha_n$. $\square$

Proof of proposition 1: Suppose there was such distribution. Let $a(n) = \omega_{(2:n)} = c + bn$ for some constants c and b. We must necessarily have $b \geq 0$. If $b = 0$, by the mapping from Theorem 2 it is easy to verify that the distribution is degenerate. So we must have $b > 0$.

Now, successively apply the recursion formula to the case where $k = 1$. We obtain

$$\begin{aligned}
\omega_{(1:n)} &= \frac{n}{n-1}\omega_{(1:n-1)} - \frac{c+nb}{n-1} \\
&= \frac{n}{n-1}\left[\frac{n-1}{n-2}\omega_{(1:n-2)} - \frac{c+(n-1)b}{n-2}\right] - \frac{c+nb}{n-1} \\
&= n\omega_{(1:1)} - \frac{n}{n(n-1)}(c+nb) - \frac{n}{(n-1)(n-2)}(c+(n-1)b) + \cdots \\
&= n\left[\omega_{(1:1)} - c\left(\frac{1}{n(n-1)} + \frac{1}{(n-1)(n-2)} + \cdots\right)\right. \\
&\quad \left. - b\left(\frac{1}{n-1} + \frac{1}{n-2} + \cdots\right)\right] \\
&\to -\infty,
\end{aligned}$$

as $n \to \infty$, since the sum that multiplies b diverges (while the one that multiplies c does not). This contradicts the fact that $\omega_{(1:n)}$ should be increasing in n. $\square$

References

Athey S, Haile PA (2002). Identification of standard auction models. *Econometrica* **70**:2107–2140.

Bajari P, Hortaçsu A (2003). The winner's curse, reserve prices, and endogenous entry: Empirical insights from ebay auctions. *Rand Journal of Economics* **34**.

Chan LK (1967). On a characterization of distribution by expected values of extreme order statistics. *American Mathematics Monthly* **74**:950–951.

Deltas G, Chakraborty I (2001). Robust parametric analysis of first price auctions, Illinois and Oklahoma.

Donald SG, Paarsch HJ (1993). Piecewise pseudo-maximum likelihood estimation in empirical models of auctions. *International Economic Review* **34**:121–148.

Donald SG, Paarsch HJ (1996). Identification, estimation, and testing in parametric empirical models of auctions within the private values paradigm. *Econometric Theory* **12**:517–567.

Donald SG, Paarsch HJ (2002). Superconsistent estimation and inference in structural econometric models using extreme order statistics. *Journal of Econometrics* **109**:305–340.

Godwin H (1949). Some Low Moments of Order Statistics. *The Annals of Mathematical Statistics* **20**:279–285.

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

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

Harter H (1961). Expected Values of Normal Order Statistics. *Biometrika* **48**:151–165.

Hoeffding W (1953). On the distribution of the expected values of the order statistics. *Annals of Mathematical Statistics* **24**:93–100.

Houser D, Wooders J (2000). Reputation in auctions: Theory, and evidence from ebay, Arizona.

Jones H (1948). Exact Lower Moments of Order Statistics in Small Samples from a Normal Distribution. *The Annals of Mathematical Statistics* **19**:270–273.

Krasnokutskaya E (2002). Identification and estimation in highway procurement auctions under unobserved auction heteogeneity, Yale.

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

Levin D, Smith J (1991). Some Evidence on the Winner's Curse: Comment. *American Economic Review* **81**:370–375.

Lucking-Reiley D, Bryan D, Prasad N, Reeves D (2000). Pennies from eBay: the determinant of price in online auctions, University of Arizona.

McDonald CG, Slawson VC (2002). Reputation in an internet auction market. *Economic Inquiry* **40**:633–50.

Melnik MI, Alm J (2002). Does a seller's ecommerce reputation matter? evidence from ebay auctions. *Journal of Industrial Economics* **50**:337–49.

Myerson RB (1981). Optimal auction design. *Mathematics of Operations Research* **6**:58–73.

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

Paarsch HJ, Hong H, Haley MR (2006). *An introduction to the structural econometrics of auction data*. MIT Press.

Parrish R (1992). Computing expected values of normal order statistics. *Communications in Statistics-Simulation and Computation* **21**:57–70.

Pollak M (1973). On equal distributions. *Annals of Statistics* **1**:180–182.

Reynolds SS, Wooders J (2003). Auctions with a buy price, department of Economics, University of Arizona, Tucson.

Thiel S (1988). Some Evidence on the Winner's Curse. *American Economic Review* **78**:884–895.

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

Yamauti Z, Kitagawa T, Masuyama M, Kyōkai N (1972). *Statistical tables and formulas with computer applications: JSA: 1972*. Japanese Standards Association.

	Statistic	Unrestricted	Uniform	Normal	Logistic	Laplace	Gumbel
I	R^2	0.2126	0.1654	0.1735	0.1755	0.1792	0.1822
	F-statistic		2.3811	1.9743	1.8736	1.6827	1.5327
	p-value		0.9982	0.9872	0.9799	0.9545	0.9174
II	R^2	0.4731	0.4402	0.4445	0.4456	0.4479	0.4490
	F-statistic		2.4349	2.1200	2.0346	1.8683	1.7824
	p-value		0.9986	0.9935	0.9903	0.9794	0.9701
III	R^2	0.4502	0.4200	0.4240	0.4251	0.4272	0.4281
	F-statistic		2.1646	1.8761	1.7982	1.6489	1.5850
	p-value		0.9947	0.9801	0.9720	0.9478	0.9326
IV	R^2	0.5074	0.4228	0.4273	0.4286	0.4311	0.4322
	F-statistic		1.4052	1.3305	1.3094	1.2671	1.2500
	p-value		0.9788	0.9554	0.9456	0.9205	0.9079

Table 2: Functional form specification tests

Regressor	I	II	III	IV
on μ:				
constant	65.9728	67.8343	72.6487	68.6269
	(20.2600)	(61.6944)	(48.1675)	(222.9034)
BROKEN		-39.3721	-40.9110	-45.8312
		(15.1692)	(13.9647)	(43.4235)
POS		1.9572	1.9924	2.1551
		(0.7003)	(0.7031)	(4.6552)
NEG		-3.2092	-3.1656	-2.1885
		(0.7145)	(0.7192)	(4.3013)
New		59.8633	61.5885	74.0562
		(36.4966)	(27.7271)	(167.2685)
Used		-1.0429		
		(10.3430)		
Refurbished		8.9332		
		(19.7835)		
Silver		5.9684		
		(21.7494)		
Blue		-4.0321		
		(20.4925)		
Pink		6.6170		
		(27.6143)		
Green		15.7220		
		(28.2819)		
# Gb		0.9592	0.8072	0.7739
		(0.1522)	(0.1738)	(0.1695)
on σ:				
constant	55.0320	34.2180	33.8169	39.7211
	(27.4887)	(24.0819)	(24.1540)	(337.1709)
BROKEN				12.4026
				(99.6003)
POS				-0.1737
				(7.2029)
NEG				-1.4749
				(6.7277)
New				-14.0496
				(174.1814)

Table 3: Estimated coefficients, F Gumbel

Regressor	I	II	III
on μ:			
constant	65.8326	67.6936	72.5584
	(14.5745)	(55.0027)	(42.9749)
BROKEN		-39.2975	-40.8381
		(18.9874)	(18.5294)
POS		1.9562	1.9924
		(0.6954)	(0.6761)
NEG		-3.2107	-3.1688
		(0.7354)	(0.7455)
New		59.8172	61.5575
		(34.1179)	(29.0913)
Used		-1.0479	
		(11.7372)	
Refurbished		8.9111	
		(30.8950)	
Silver		6.0242	
		(17.2327)	
Blue		-4.0017	
		(19.5332)	
Pink		6.6749	
		(21.9380)	
Green		15.7465	
		(26.5944)	
# Gb		0.9547	0.8027
		(0.2544)	(0.2595)
on σ:			
constant	55.1977	34.3841	33.9580
	(20.9938)	(17.0459)	(17.4974)

Table 4: Weighted least squares estimated coefficients, F Gumbel

	$\hat{\mu}_{LS}$	$\hat{\sigma}_{LS}$	$\hat{\mu}_{ML}$	$\hat{\sigma}_{ML}$
true value	3	1	3	1
	sample size = 50			
mean	2.9978	1.0087	3.0055	0.9859
variance	0.0125	0.0579	0.0079	0.0080
mean squared error	0.0125	0.0580	0.0079	0.0082
lower quartile	2.9261	0.8527	2.9460	0.9258
median	3.0000	1.0140	3.0032	0.9868
upper quartile	3.0718	1.1736	3.0657	1.0463
skewness	0.0104	-0.0561	0.0497	0.0443
kurtosis	3.2395	2.9989	3.0816	2.9373
Jarque-Bera	2.4072	0.5250	0.6900	0.4913
p-value	0.6999	0.2309	0.2918	0.2178
	sample size = 100			
mean	3.0015	0.9907	3.0031	0.9872
variance	0.0063	0.0284	0.0042	0.0047
mean squared error	0.0063	0.0285	0.0042	0.0048
lower quartile	2.9453	0.8845	2.9587	0.9400
median	3.0029	0.9900	3.0036	0.9883
upper quartile	3.0551	1.0970	3.0482	1.0319
skewness	-0.0542	-0.0173	0.0548	0.0482
kurtosis	2.9197	3.2266	2.9559	2.8821
Jarque-Bera	0.7573	2.1903	0.5810	0.9665
p-value	0.3152	0.6655	0.2521	0.3832
	sample size = 200			
mean	3.0006	0.9963	3.0022	0.9927
variance	0.0031	0.0151	0.0021	0.0021
mean squared error	0.0031	0.0151	0.0021	0.0021
lower quartile	2.9624	0.9177	2.9713	0.9622
median	3.0010	0.9964	3.0021	0.9923
upper quartile	3.0365	1.0829	3.0337	1.0221
skewness	0.1098	-0.1208	-0.0344	0.0473
kurtosis	2.9902	3.0826	3.2539	3.0364
Jarque-Bera	2.0148	2.7182	2.8831	0.4287
p-value	0.6348	0.7431	0.7634	0.1929

Table 5: Monte Carlo experiment: Least Squares v Maximum Likelihood, Second-Price auctions, valuations normally distributed

	$\hat{\mu}_{LS}$	$\hat{\sigma}_{LS}$	$\hat{\mu}_{ML}$	$\hat{\sigma}_{ML}$
true value	3	1	3	1
	sample size= 50			
mean	2.9984	0.9999	2.9745	1.0067
variance	0.0043	0.0149	0.0022	0.0023
mean squared error	0.0043	0.0149	0.0028	0.0023
lower quartile	2.9553	0.9182	2.9562	0.9838
median	3.0007	0.9985	2.9859	0.9997
upper quartile	3.0433	1.0801	2.9999	1.0281
skewness	-0.0759	0.0601	-1.1046	0.4069
kurtosis	2.8948	2.8713	5.4469	5.0040
Jarque-Bera	1.4223	1.2927	452.8485	194.9273
p-value	0.5089	0.4760	1.0000	1.0000
	sample size= 100			
mean	3.0011	0.999	2.9872	1.0041
variance	0.0021004	0.0076885	0.00064033	0.00065485
mean squared error	0.0021017	0.0076895	0.00080441	0.00067165
lower quartile	2.9707	0.94184	2.9791	0.99624
median	3.0027	1.0002	2.9967	0.9994
upper quartile	3.033	1.0534	2.9998	1.0098
skewness	-0.23027	-0.056137	-2.3094	1.4645
kurtosis	2.8778	3.2961	13.922	11.439
Jarque-Bera	9.4598	4.1793	5858.9	3324.6
p-value	0.99117	0.87627	1	1
	sample size= 200			
mean	2.9978	1.0016	2.9959	0.99959
variance	0.0010318	0.00363	0.00010281	0.00011701
mean squared error	0.0010365	0.0036326	0.00011948	0.00011718
lower quartile	2.9779	0.96149	2.9941	0.99797
median	2.9986	1.0017	2.9996	0.9992
upper quartile	3.0203	1.0437	2.9999	0.99998
skewness	-0.16304	-0.066983	-1.7388	0.45748
kurtosis	2.9733	3.1163	9.6588	11.456
Jarque-Bera	4.4603	1.3113	2351.4	3014.3
p-value	0.89249	0.48089	1	1

Table 6: Monte Carlo experiment: Least Squares *v* Maximum Likelihood, First-Price auctions, valuations uniformly distributed

	$\hat{\beta}$	$\hat{\delta}_{111}$	$\hat{\delta}_{112}$	$\hat{\delta}_{121}$	$\hat{\delta}_{211}$
mean	1.0025	10.7906	11.8132	11.4697	12.2569
variance	0.0088	0.1992	0.2603	0.2069	0.2412
lower quartile	0.9401	10.4955	11.4431	11.1636	11.9314
median	1.0016	10.7968	11.8048	11.4783	12.2489
upper quartile	1.0662	11.0889	12.1405	11.7651	12.5655
skewness	-0.0670	-0.0371	0.1838	0.0073	0.1217
kurtosis	2.9229	3.0511	3.1565	3.3518	3.5363
Jarque-Bera	0.9957	0.3386	6.6543	5.1656	14.4529
p-value	0.3922	0.1557	0.9641	0.9244	0.9993
		$\hat{\delta}_{122}$	$\hat{\delta}_{212}$	$\hat{\delta}_{221}$	$\hat{\delta}_{222}$
mean		11.0366	11.9127	11.5863	12.3214
variance		0.1613	0.2346	0.1802	0.2095
lower quartile		10.7652	11.6049	11.2957	12.0200
median		11.0298	11.9005	11.5647	12.2987
upper quartile		11.2792	12.2198	11.8623	12.6183
skewness		0.1265	0.2060	0.1278	0.1243
kurtosis		3.1537	3.3281	2.8512	3.1200
Jarque-Bera		3.6535	11.5586	3.6440	3.1769
p-value		0.8391	0.9969	0.8383	0.7958

Table 7: Monte Carlo experiment with an asymmetric model