

# Econometrics of Auctions by Least Squares

Leonardo Rezende\*

Dept. of Economics, University of Illinois  
1206 S Sixth St., Champaign, IL 61820, U.S.A.  
lrezende@uiuc.edu

January 14, 2006

## Abstract

This paper proposes a method to structurally estimate an auction model using a variation of OLS, under commonly held assumptions in both auction theory and econometrics. In spite of its computational simplicity, the method applies to a wide variety of environments, including interdependent values in general, and certain forms of endogenous participation and bidder asymmetry. Furthermore, it can be used for hypotheses testing about the shape of the valuation distribution, valuation interdependence, or existence of bidder asymmetry.

JEL Codes: D44, C14.

Keywords: Auctions, Semiparametric Methods, Private Values, Common Values, Endogenous Participation.

## 1 Introduction

Empirical methods designed to investigate auction data are of interest not only because they can be used to test auction theory, but also because they

---

\*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 and Ali Hortagsu for valuable suggestions. All errors are mine.

provide a rich source of information about the underlying consumer preferences or producer costs that give rise to the bidders' observed behavior. As such they can be of considerable interest to a wide class of empirical economists, not necessarily restricted to those directly involved in auction-theoretic issues.

For that purpose, one would want to use an empirical strategy that is at once well-founded in auction theory, robust to misspecification of the details of the environment, and yet easy to implement, both computationally and intuitively. The purpose of this paper is to propose such a method.

Studies that seek to understand how covariates affect demand in auction markets usually run regressions like that:

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

where  $p$  is the transaction price (or its log). To provide a specific example, several studies 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).

The problem with analyses that rely on such specification is not in the right-hand side of regression — covariates may indeed impact consumer preferences in a linear fashion — but rather in the left-hand side. A superior starting point for an empirical analysis would be a specification like

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

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

Given a precisely defined institutional environment, one can use auction theory to obtain a mapping between valuations and observed bidding behavior and from that mapping develop an estimation strategy. Laffont and

Vuong (1996) have shown that, depending on the auction protocol, the valuation distribution is just identified from bidding data. Several methods have been developed to estimate valuations from auction data: for example, for independent private values, Laffont et al. (1995) have proposed a simulated nonlinear least squares methodology; Donald and Paarsch (1993) have introduced a piecewise pseudo-maximum likelihood estimator; Guerre et al. (2000) have shown how to obtain fully non-parametric estimates of the valuation distribution, and Haile and Tamer (2003) use bounds estimation on an incomplete model of an English auction.

These structural methods have two weaknesses from the point of view of its applicability: they are tailored to a specific type of the auction being played and they are computationally complex. This state of affairs may lead to the perception that applied work with auction data informed by auction theory would necessarily lead to overly complex and specific empirical procedures, and that, for researchers unwilling to adopt them, the only alternative is to rely on misspecified regressions like equation 1.

This is not true. The objective of this paper is to show that, starting from a specification like equation 2, one can obtain an estimation method that is at once structural, justified by auction theory, computationally simple and robust to several types of misspecifications of the auction game being played.

The method proposed involves running an ordinary least squares regression of observed transaction prices in the covariates of interest and an additional regressor involving the number of bidders in each auction. Computationally the method is as straightforward as running an OLS regression, involving variables that are readily observable in an auction dataset. There is no need for example to use information on losing bids, that is often not observable or not reliable from theoretical and practical grounds.<sup>1</sup>

And yet, the method is fully theoretically justified. It also remains valid in a wide variety of auction rules and assumptions about the statistical properties of valuations, including common values and other forms of affiliation. It also robust to several versions of endogenous or hidden entry, and can also accommodate some forms of bidder heterogeneity.

---

<sup>1</sup>Bikhchandani et al. (2002) have shown that In English auctions there is a continuum of symmetric equilibria with losing bids not uniquely determined. Furthermore, in practice the English auction rules do not fully pin down the losing bidders drop-out points (Haile and Tamer, 2003). Even in first-price auctions, in practice bidders sometimes submit “non-serious” bids — bids at very low prices that they do not expect to win (Garcia and Rezende, 2000).

A second product of the methodology is that it provides not only information about the impact of covariates, but also delivers in a computationally cheap way a significant amount of information about the shape of the valuation distribution within each auction. The discussion on how to explore the method to obtain such information is done in section 6.

This work is related to several strands in the empirical literature on auctions. The regression proposed, involving winning bids in the left hand side and the number of bidders in the right hand side, is similar to regressions ran both in the reduced form literature (e.g. Gilley and Karels, 1981) or as an initial step in the structural literature (e.g. Paarsch, 1991). The main difference is that, because here the number of bidders is included in a nonlinear fashion, our regression method is fully justified on theoretical grounds.<sup>2</sup>

The method here will exploit two important properties of auctions: the Revenue Equivalence Theorem and the invariance of the set of equilibria to affine transformations on valuations (see Proposition 2 below). Both ideas have been used elsewhere, but not simultaneously: Laffont et al. (1995) use the Revenue Equivalence Theorem to justify their simulated NLLS methodology; versions of the invariance property are used in Bajari and Hortaçsu (2003), Deltas and Chakraborty (2001) and Krasnokutskaya (2002). However, these papers do not combine both ideas; as a result, their methodologies are either computationally demanding, applicable to a limited class of auctions, or both.

The paper is organized as follows: Section 2 introduces the method in the classic independent private values framework. Sections 3 to 5 discuss the robustness of the method in more general contexts: Interdependent values are discussed in section 3; bidder heterogeneity is discussed in section 4; and various types of endogenous or hidden participation are discussed in section 5. The possibility of exploiting the information obtained from the method for inference about valuations distributions is discussed in section 6, and illustrated by an practical application on a sample of Palm Pilot eBay auctions in section 7. Section 8 concludes.

---

<sup>2</sup>Also, in most cases these regressions include the number of bidders in a linear or quadratic form, which will be shown to be incorrect functional assumptions (see Proposition 1 below).

## 2 The Basic Method

With minor modifications, the method proposed in this paper applies to a variety of symmetric auction models, including auctions with interdependent values and endogenous entry. For the sake of expositional clarity, it is convenient to present the method first in the context of the benchmark model in auction theory: the independent private values model. This is done in this section. Sections 3–5 discuss the necessary changes to accommodate a richer set of assumptions.

Let  $V_{il}$  be the valuation of the  $i$ -th bidder in the  $l$ -th auction. Let  $\mu_l$  be the mean and  $\sigma_l$  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  $\epsilon_{il}$  are *i.i.d.* with distribution  $F$ .

The normalized 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 mere 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.<sup>3</sup>

Independence across bidder valuations within an auction is one of the central characteristics of the benchmark model in auction theory, the independent private values auction model. Imposing independent private values simplifies the analysis somewhat, but is not necessary: see Section 3 for an extension to the much more general interdependent values case.

The first object of interest of the researcher is to evaluate the effect of covariates on  $\mu_l$  and  $\sigma_l$ . Let  $X_l$  be the vector of covariates that affect the expected valuation  $\mu_l$  and  $Z_l$  the vector of covariates that affect  $\sigma_l$ . 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  $\sigma_l = Z_l \alpha$ .

---

<sup>3</sup>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.

We intend to obtain estimates of  $\beta$  and  $\alpha$  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  $p_l$ . The method will exploit the central result in auction theory, the Expected Revenue Equivalence Theorem. Besides independent private values, we impose

**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 —  $\epsilon_{il}$  and  $n_l$  are independent. The number of bidders is also assumed to be publicly known before bidding. Both hypotheses can be relaxed somewhat: See Section 5 for details.

One remarkable advantage of the Independent Private Values framework is that it allows the method to be unaffected by the details of the auction process: The same method applies to any standard auction rule. More precisely, we require that

**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 auction, the second-price auction, and also by the all-pay auction.<sup>4</sup>

Although this is an admittedly long list of assumptions, they are quite conventional. In fact, assumptions 3, 4 and independence and identical distribution of valuations within an auction are the conditions for an independent private values model, the classical auction-theoretic framework introduced by Vickrey (1961). Assumptions 1 and 2 are typical in Econometrics; if for example  $F$  is standard normal and  $Z_l$  included just the intercept we would obtain the classical regression model for  $V_{il}$  (except for the fact that  $V_{il}$  is not observed). A richer  $Z_l$  allows for heteroskedasticity.

The method explores the central result in auction theory, the Expected Revenue Equivalence Theorem (Vickrey, 1961; Myerson, 1981):<sup>5</sup>

---

<sup>4</sup>In the case of an all-pay auction,  $p_l$  should be interpreted as the sum of the prices paid by all bidders, rather than the amount paid by the winner.

<sup>5</sup>In what follows the notation  $x_{(k:n)}$  represents the  $k$ th-highest order statistic from an i.i.d. sample of  $n$  observations of  $x$ .

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

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  $E[p_l|X_l, Z_l, n_l] = E[V_{(2:n_l)}|X_l, Z_l, n_l]$ .

## 2.1 Estimation

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 second is due to the specification of the location and scale parameters, and the third by the definition of conditional expectation. Defining  $a(n) = E[\epsilon_{(2:n)}]$ ,

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

Note that the conditional expectation of the winning bid is *linear* in  $\beta$  and  $\alpha$ . This means that OLS is an unbiased, consistent estimation method to estimate these coefficients. This observation gives rise to two straightforward procedures to estimate  $\beta$  and  $\alpha$ : If the researcher is willing to impose a particular choice for the shape of the distribution of values, the following method is suggested:

**Method 1** Using the normalized 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.

The first step requires the computation of  $a(n) = E[\epsilon_{(2:n)}]$  for the specified distribution of  $\epsilon$ . These calculations are shown for some normalized distributions in Table 1 below. After this calculation is done, unbiased estimates of the parameters can be obtained by OLS.

$n$	Uniform	Normal	Log-normal	Chi-square
2	-0.6188	-0.6074	-0.4304	-0.4720
3	0.0000	0.0000	-0.2030	-0.2227
4	0.3695	0.3168	-0.0237	-0.0068
5	0.6158	0.5279	0.1279	0.1803
6	0.7917	0.6844	0.2606	0.3445
7	0.9237	0.8077	0.3793	0.4902
8	1.0263	0.9088	0.4873	0.6213
9	1.1083	0.9941	0.5865	0.7401
10	1.1755	1.0677	0.6785	0.8489
11	1.2314	1.1322	0.7645	0.9491
12	1.2787	1.1895	0.8453	1.0420
13	1.3192	1.2410	0.9216	1.1286
14	1.3544	1.2876	0.9940	1.2096
15	1.3851	1.3302	1.0628	1.2858
16	1.4122	1.3693	1.1285	1.3576
17	1.4362	1.4055	1.1915	1.4256
18	1.4577	1.4391	1.2518	1.4902
19	1.4771	1.4704	1.3099	1.5516
20	1.4945	1.4997	1.3658	1.6101

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



It is important to notice that method 1 is not the same as introducing  $n_l$  as an additional regressor, as is often done in the literature, but rather it entails introducing  $a(n_l)$ , a nonlinear function of  $n_l$ . In fact, ignoring the nonlinearity is never correct; there is no distributional assumption that would justify that:

**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 6, it will be deferred to appendix B.

Unbiased estimators for  $\alpha$  and  $\beta$  do depend on 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. Often, however, one may be unwilling to commit to a specific assumption about the valuation distribution. In that case, one can still obtain unbiased estimates for  $\alpha$  and  $\beta$  using the following method:

**Method 2** Run OLS of the observed winning bids on  $X_l$  and interactions of  $Z_l$  and dummies for each  $n_l$  in the sample.

Method 2 has the advantage of allowing the researcher to be agnostic about the shape of the valuations distribution. It requires constructing dummies  $d_{kl}$  for the event  $n_l = k$  for every  $k$  in the support of  $n_l$ , and then running an OLS regression of  $p_l$  on  $X_l$  and interactions of  $Z_l$  and  $d_{kl}$ .

An interesting feature of method 2 is that the pattern of the estimated dummy coefficients contains information about the shape of  $F$  — in fact, it will be shown in section 6 that the sequence  $\{a(n)\}_{n=2}^{\infty}$  identifies  $F$ . This gives rise to the possibility of using the dummy estimates for inference about the shape of  $F$ . Section 6 discusses this possibility in detail.

### 3 Interdependent Values

For expositional convenience, the presentation of the method in the last section was done under the assumption of independent private values. However, this assumption is not at all necessary. This section shows that the method works essentially unchanged for virtually any other information structure and most auction rules (including all commonly observed auction rules).

To describe the necessary conditions on the auction rules, let  $b_i$  to be bidder  $i$ 's "bid" in a given auction (The quotation marks here are used because  $b_i$  does not necessarily need to be a bid *per se*; it is simply a way to describe bidder  $i$ 's action). In most auctions,  $b_i$  is taken to be a number (but not necessarily a payment promise); all that will be needed here is that  $b_i$  be a member of a real vector space. Let  $b = (b_1, \dots, b_n)$ . Given an auction rule, let  $W_i(b)$  be  $i$ 's (expected) probability of winning the item given that bidders in the auction played  $b$ , and let  $P_i(b)$  be the expected payment given  $b$ .

Let  $v_i$  be  $i$ 's valuation (not necessarily private or independent), let  $v = (v_1, \dots, v_n)$ , and let  $\mu$  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 an auction rule is such that (i)  $W_i(\mu + \sigma b) = W_i(b)$  and (ii)  $P_i(\mu + \sigma b) = \mu + \sigma P_i(b)$ , for all  $i$ . Then, if  $\beta = (\beta_1, \dots, \beta_n)$  is a Nash equilibrium of an auction when bidders have valuations  $v$ , then  $\mu + \sigma \beta$  is a Nash equilibrium of the auction when valuations are  $\mu + \sigma v$ . As a consequence, if the expected selling price in the first auction is  $\bar{p}$ , then the expected price in the second auction will be  $\mu + \sigma \bar{p}$ .*

**Proof:** For  $\beta$  to be a Nash equilibrium under  $v$ , it must be that it prescribes  $i$  to bid

$$\operatorname{argmax}_{b_i} E[v_i W_i(b) - P_i(b)],$$

where this expectation is taken to be conditional on all information available to  $i$  (including the contingency of winning), and  $b_{-i}$  follows  $\beta_{-i}$ .

In the game with valuations  $\mu + \sigma v$ , by following  $\mu + \sigma \beta$  bidder  $i$  achieves

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

Clearly, any  $b_i$  that solves the first problem solves the second as well.

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

Notice that the two conditions on  $W_i$  and  $P_i$  are satisfied by all common, and many uncommon, auction rules. For example, the condition on  $W_i$  is satisfied by any rule that assigns a winner based on a ordering of the bids. This includes all standard auctions, that assign the good to the highest

bidder, and many non-standard auctions, that for example assign the good to the  $k$ -th highest bidder.<sup>6</sup>

The condition on  $P_i$  is also very easily met; any auction where payments depend linearly on a bidder's own bid (as in a first price auction or the all-pay auction) or somebody else's bid (as in the Vickrey or English auctions) will satisfy it.

For our purposes proposition 2 is very important because it implies that, in any auction equilibrium, we can write something very similar to equation 3, namely,

$$E[p_l|X_l, Z_l, n_l] = X_l\beta + Z_l\alpha\tilde{a}(n_l), \quad (4)$$

where  $\tilde{a}(n_l) = E[\tilde{p}_l|n_l]$  and  $\tilde{p}_l$  is the selling price from an auction where bidders have normalized valuations.

So for estimation purposes the only impact of relaxing the IPV assumption is that the first step in method 1 involves calculating a different  $\tilde{a}$  function, a function that depends not only on the distribution of the normalized values, but also on their statistical dependency, their relation with the bidders' information, *and* the specific auction rule (recall that without IPV revenue equivalence no longer holds). Appendix A illustrates these calculations for several affiliated values models.

Such calculations might be quite hard in particular instances, but fortunately for researchers interested in only estimating  $\beta$  and  $\alpha$  they are not necessary. From equation 4 we learn that method 2 would work here in exactly the same manner without requiring the researcher to commit to a specific informational structure or auction rule.<sup>7</sup>

## 4 Asymmetry

Another important assumption imposed in section 2 is that bidders are symmetric, in the sense that within auction their ex-ante value distribution is the

---

<sup>6</sup>It may be worthwhile to point out that this assumption implicitly imposes some restrictions 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 in line with  $v$  (so that in the second auction the reserve price is  $\mu + \sigma r$ ).

<sup>7</sup>That of course assumes that no changes occur in the auction rules across the sample. If, for example, part of the sample contains sealed-bid auctions and part contains open ascending auctions, one must introduce different sets of dummies for each type of auction rule.

same. On the other hand, as the discussion in section 3 suggests, the driving property of auction models that validates the method is Proposition 2, and it does not depend on symmetry. This suggest that the method can also accomodate asymmetric auctions. This section discusses this possibility.

For concreteness, suppose we consider an auction where bidders come from two different categories (say, “A” and “B”). Bidders from different categories obtain their values from different distributions, and this distinction is publicly known (that is what makes this model fundamentally different from a symmetric one). Let  $n_l^A$  and  $n_l^B$  be the number of bidders from categories A and B that participate in auction  $l$ . Then, in light of Proposition 2, we have

$$E[p_l|X_l, Z_l, n_l^A, n_l^B] = X_l\beta + Z_l\alpha a(n_l^A, n_l^B), \quad (5)$$

where  $a(n_l^A, n_l^B)$  is the expected price of a normalized auction with  $n_l^A$  bidders of category A and  $n_l^B$  of category B.

The same regression strategies would still work, but now the nonlinear function  $\tilde{a}$  is somewhat more complex. Not only it is now a function of two rather than one index, but calculating it would require finding an equilibrium of an asymmetric auction. This can be hard; for example, in first-price auctions that would involve solving a system of ordinary differential equations (Lebrun, 1996; Bajari, 1999).

Another difficulty with the asymmetric setting is in how to exactly interpret  $\mu_l$  and  $\sigma_l$  (or, equivalently, on how to normalize valuations). It would probably be desirable to have a definition based on means and standard deviations from each category’s valuation distributions; while it is clear that they are related with  $\mu_l$  and  $\sigma_l$ , the relationship is far from transparent, and is dependent on the specific nature of the asymmetry.

In any case, here again method 2 is applicable, provided that we allow dummies for both  $n_l^A$  and  $n_l^B$ , and does not require specifying the exact nature of the asymmetry, let alone computing equilibria of asymmetric auctions.

The estimated  $\beta$  and  $\alpha$  would not provide information about the exact nature of the asymmetry (e.g., they would not tell which group tends to have higher values), but they would provide information on how covariates affect valuations in a way that is robust to asymmetry.

Furthermore, the values of the dummy coefficients can be used to test whether symmetry is present (or more precisely, whether it significantly impacts selling prices). To test whether valuation distributions differ across categories A and B, a researcher can run a regression of equation 5 by method 2

and then perform a linear hypothesis test, where the null hypothesis of symmetry corresponds to the restriction  $a(n^A, n^B) = a(n^A + n^B)$  for all values of  $n^A$  and  $n^B$ .

## 5 Endogenous or Hidden Participation

So far the analysis has relied on the hypothesis that the number of bidders in each auction is determined exogenously and is common knowledge among bidders. This is a conventional assumption in the auction theory literature, but may not necessarily hold in practice. Because the number of bidders play such an important role in the estimation strategy, it is appropriate to discuss how alternative hypotheses about bidder entry would affect the estimation method.

This section discusses three possible alternative hypotheses: endogenous entry prior to learning about values, endogenous entry after learning about values, and exogenous hidden entry.

### 5.1 Why the Timing of the Entry Decision is Important

A natural way to formalize the entry decision process is to assume that prior to participating in the market bidders face an entry cost. This cost may be related to the process of searching for the auction and evaluating the item being sold, or can be related to the effort to submit a bid. This distinction is important, because in the first case bidders must decide to enter and incur the cost before they know the value they attach to the item (ex-ante endogenous entry), while in the second case they do so after they know their value (interim endogenous entry).

Statistically, the distinction can be described as following: let  $N$  be the original number of potential bidders in a given auction, that is typically not observed by the econometrician directly. Let as before  $n \leq N$  be the number of bidders that participate. Let  $\lambda_i$  be the probability that bidder  $i$  participates. The crucial distinction between ex ante endogenous entry and interim endogenous entry is that they predict different correlation patterns between  $\lambda_i$  and  $\epsilon_i$ . In ex-ante endogenous entry models,  $\lambda_i$  is independent of  $\epsilon_i$  (and typically assumed to be the same across bidders); in interim endogenous

entry models,  $\lambda_i = \Pr(\epsilon_i > c)$ , where  $c$  is a threshold that depends on the magnitude of the entry cost.

## 5.2 Ex Ante Endogenous Entry

The theory for the case of ex-ante endogenous entry can be found in McAfee and McMillan (1987b) and Levin and Smith (1994). McAfee and McMillan (1987b) studied asymmetric equilibria where a deterministic number of bidders enter; Levin and Smith (1994) focused on the symmetric, mixed strategy equilibrium of the same game. In either case, as long as the number of bidders that eventually enter is common knowledge when bidding is decided, bidding behavior is not affected.

As long as participation is large enough ( $n \geq 2$ ),<sup>8</sup> information about  $N$  is not necessary to estimate  $\beta$  and  $\alpha$ : If the parameterization of the value in Assumption 1 is still valid, the methods work unchanged. The number of bidders  $n$  is endogenous in the game-theoretic sense, but not on the econometric sense, since it is still uncorrelated with the regression residual. Variation in  $n$  can be justified by (unobserved) variation in entry costs or by the realizations of a mixed strategy equilibrium. An econometric endogeneity problem would arise only if  $\epsilon$  is correlated with information observable by bidders prior to entry that would affect their entry decision.

The ex-ante endogenous entry model provides a justification for not worrying about sample selection bias. As long as the values within an auction (and therefore the residuals) are independent from the variables that determine entry, then no sample selection bias exists. In this theory low participation would be explained through either a low number of initial potential bidders ( $N$ ) or high participation costs, and not through an unobserved value component.<sup>9</sup>

## 5.3 Interim Endogenous Entry

Another possibility is that the entry decision happens after bidders learn about their value (Samuelson, 1985). This assumption is appropriate if the

---

<sup>8</sup>Auctions with little or no participation ( $n \leq 1$ ) obviously cannot be included in the regression, since for them the expected price does not obey equation 3.

<sup>9</sup>Here one would need to assume that participation costs are independent of the unobserved value component as well.

entry cost is related to the bidding process itself, rather than finding the auction or evaluating the value of the product.

Consider a situation where there are  $N$  potential bidders with independent private values, but in order to bid they must pay an entry cost. A bidder would enter if the expected payoff in the auction is enough to cover the entry cost. Since the expected payoff is increasing in the bidder's valuation, there will be a cutoff such that bidders with values above it enter and those below it stay out. In this case, the probability of  $i$  entering would be  $\lambda_i = \Pr(\epsilon_i > c)$ , a function of  $\epsilon_i$ . (Notice that this would also happen in the presence of a reserve price.)

So the entry process distorts the valuation distribution of entrants — it truncates it below. However, the expected transaction price is still a second order statistic. So the methods proposed here would still work, provided that the researcher used the number of potential bidders  $N$  rather than the number of actual bidders.

Unfortunately in practice the number of potential bidders is seldom observed by the econometrician. In these circumstances substituting the actual number of bidders would lead to specification bias. The next two sections discuss two ways to account for lack of direct information about  $N$  in the estimation.

### 5.3.1 Nonlinear Error in Variables

One way to proceed, bar direct knowledge of  $N$ , would be to assume the researcher has access to a set  $W$  of proxies for it. Assuming a linear relationship between  $N$  and  $W$ , we could perhaps write

$$N = W\gamma + u, \tag{6}$$

where  $u$  is an error term (possibly assumed to be independent of  $\{\epsilon_i\}$ ). In that case, however, the initial regression would read

$$p = X\beta + Z\alpha a(W\gamma + u) + \eta,$$

something that cannot be estimated in a straightforward fashion given the non-linearity of  $a(\cdot)$ . This is a nonlinear error-in-variables model. Because of the nonlinearity around  $N$ , such a model cannot be estimated by instrumental variables (Amemiya, 1985).

Inference with nonlinear-error-in variables is a topic of active research. The existing methods require further sources of information, such as validation data (e.g. Lee and Sepanski, 1995; Newey, 2001), specific distributional assumptions for the measurement error (Hong and Tamer, 2003), or the availability of separate data about the marginal distribution of the measurement error (Hu and Ridder, 2003). Furthermore, most of the literature cannot be applied to the problem at hand, since they focus on the case of a continuous mismeasured variable. An exception is Hu (2004), that develops a method for a discrete variable with error in the presence of an instrument. Hu’s method may be applied in conjunction to our method on auction data, provided valid instruments are available.

### 5.3.2 Systematic Relationship between $n$ and $N$

An alternative that avoids the non-linear error in variables problem is to assume that the entry process is such that, conditional on observable intrumentes  $W$ , there is a systematic relationship between the observed number of bidders  $n$  and the potential number of bidders  $N$  we are interested in:

$$n = \phi(W, N). \quad (7)$$

If  $\phi$  is strictly monotonic with respect to  $N$ , then one can invert this relationship and write<sup>10</sup>

$$N = \psi(W, n),$$

so that

$$p = X\beta + Z\alpha a(\psi(W, n)) + \eta = X\beta + Z\alpha \tilde{a}(W, n) + \eta. \quad (8)$$

Estimation would be straightforward: all that is needed is to allow  $\tilde{a}$  do depend not only on  $n$ , but on  $W$  as well.

While this idea provides an easy way out of the problem of not observing  $N$ , it is not easily justifiable on theoretical grounds. Natural theories of entry that justify focusing on  $N$  in the first place would predict that the relationship between  $n$  and  $N$  is not deterministic; for example, the model discussed in section 5.3 would predict that  $n$  is a binomial random variable from a sample of size  $N$ . It is not clear whether it is possible to find a reasonable model of endogenous entry that would simultaneously justify using  $N$  in equation 3 and yet have  $N$  and  $n$  be deterministically related as in equation 7.

---

<sup>10</sup>This “trick” is originally due to Olley and Pakes (1996), and has been used otherwise in the empirical auction literature by Campo et al. (2003) and Haile et al. (2003).



## 5.4 Exogenous Hidden Entry

Another aspect of entry that is important in the analysis is whether bidders know the number of competitors before bidding. McAfee and McMillan (1987a) have developed the theory of how lack of knowledge about the number of bidders affects the properties of an auction.<sup>11</sup>

The way hidden entry affects the method will depend on the specific auction rule. In an English or a Vickrey auction bidders do not use information about the number of competitors in their strategies; therefore lack of knowledge about it does not have any effect.

In a first price auction, equilibrium bidding strategies do depend on the number of bidders. So if required to bid without knowing  $n$ , a bidder will have to follow a strategy that maximizes the expected payoff over the range of possible values for  $n$ . That would lead to a different equilibrium bid strategy, and therefore a different expected price, depending on whether  $n$  is hidden or not.

McAfee and McMillan (1987a) show that a version of the Expected Revenue Equivalence Theorem is still valid: the expected price of a first-price auction with hidden  $n$  is the same as the English or Vickrey auction, and therefore also of a first price auction with known  $n$ . However, this is not sufficient for the applicability of the method. Revenue equivalence holds only ex-ante, for the unconditional expected price. For the method to work unchanged we would need instead equivalence of the expected price conditional on  $n$ . If the econometrician uses the methods described here with information about  $n$  that was not available to the bidders in a first price auction, the regression would be misspecified, since compared with predicted behavior bidding would be too aggressive in auctions with few bidders and not aggressive enough in auctions with many bidders.

## 6 Identifying Distributions from Least Squares Coefficients

Besides providing unbiased estimates for parameters that determine the location and scale of the value distributions, the method proposed in this paper also provide indirect information about the underlying shape of the valuation distribution through the estimated pattern of  $a(n)$ 's.

---

<sup>11</sup>Here entry is again assumed to be exogenous.

This section discusses ways to explore this information. It establishes that this information is enough to fully identify the valuation distribution: in other words, in principle it is possible to obtain a nonparametric estimator of the valuation distribution from the estimated coefficients of an ordinary 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 Palm III PDA auctions from eBay.

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

A perhaps surprising fact is that a 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  $n = 2, 3, \dots$  and all  $k = 2, 3, \dots, n$ . 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.□

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

**Proof:** See appendix B.□

Observe that 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.

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.<sup>12</sup>

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

**Proof:** See appendix B.  $\square$

**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  $k_n/n \rightarrow 1 - \alpha$ . Use the recurrence formula to compute  $\omega_{(k_n:n)}$  from the  $a(n)$ 's, and then take the limit. The converse is immediate.  $\square$

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 unbiased and consistent for  $a(n)$ . Since the recurrence formula is linear, the corresponding  $\hat{\omega}_{(k:N)}$  are also unbiased and 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.

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

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

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

---

<sup>12</sup>Another application of the recurrence relation in Economics can be found in Athey and Haile (2002). Using the relation with  $g(x) = \mathbb{I}\{x < t\}$  one obtains a recurrence relations between the distributions of order statistics. Athey and Haile (2002) use this relation to investigate ways of testing for the winner's curse.

So  $\omega_{(k:n)}$  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.

To see that, consider the case where there are no covariates for the the mean and variance of the valuations ( $\mu_l = 0$  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  $\text{Var}(\hat{\omega}_{(k:n)}) = \sum_{j=0}^{k-2} c(j, k, n)^2 \text{Var}(\hat{a}(n-j))$ , where  $c(j, k, n) = \frac{n!}{(n-k)!(k-1)!} \frac{(k-2)!}{j!(k-2-j)!(n-j)(n-j-1)}$ .

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  $c(0, k, n)$ . For  $k = n/2$  we find

$$\begin{aligned} c(0, k, n) &= \frac{n!}{(n-k)!(k-1)!n(n-1)} = \frac{(n-2)!}{(n-k)!(k-1)!} \\ &\simeq \frac{n!}{((n/2)!)^2} \xrightarrow{n \rightarrow \infty} \infty. \end{aligned}$$

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 direct method.

However, that theoretical result 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.

## 6.2 Hypothesis Testing about Distributions

Perhaps the easiest and most useful way to use the information contained on the  $\hat{a}(n)$  estimators is to test for the hypothesis of a specific distribution  $F$ .

Under the null hypothesis that a given  $F$  is the normalized 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. This procedure is illustrated in section 7. In an application to data on Palm III auctions,

this method succeeds in indicating that the distribution of bidder valuations is log-normal — a distribution that is often thought to be appropriate for valuations in auctions.<sup>13</sup>

Of course, this idea is not restricted to the IPV class of distributions; any distribution of values to which one can compute equilibrium bids and therefore the corresponding  $a$  function can be tested in the same way. Section 7.1 provides an illustration by testing and rejecting several common values models for Palm Pilot eBay auctions.

## 7 An Application to the Palm Pilot Market

This section illustrates the capacity of the method to deliver information about the shape of the value distribution using actual data. The method has been applied to a sample of 2299 Palm III personal digital assistants auctions at eBay, from October 14 to November 13, 2000.

We have introduced as covariate for the mean a dummy for the object being described as broken, and no covariate for the variance. Unfortunately there is no data on the number of bidders per auction in this sample, and the number of bids have been used in its place. That potentially overestimates the number of bidders. Although, under the proxy bidding rules used in eBay, bidders do not have an incentive to bid more than once, if valuations are independent.

Estimated coefficients for all regressions are reported in Tables 5 and 6 at the end of the paper. Figure 1 presents graphically the estimated  $a(n)$ 's, obtained with the method 2 and the corresponding estimated expected transaction prices under the assumption that  $\epsilon$  has uniform, normal, log-normal

---

<sup>13</sup>If  $F$  is known up to a family of distributions indexed by a set of parameters  $\theta$ , a similar idea can be used in a two step procedure to estimate  $\theta$ .

Assume that  $\epsilon_{il}$  is drawn from a distribution  $F \in \{F_\theta\}_\theta$ , where  $\{F_\theta\}_\theta$  is a family parameterized by  $\theta$ . We are interested in finding  $\theta$  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 samples.

or chi-square distribution.<sup>14</sup>

Figure 1: Estimated  $a(n)$ 's, Palm Pilot III regression.

The more concave lines are the  $a(n)$ 's of the uniform and normal distribution. The least concave lines, that seem to fit the data better, correspond to the distributions with a thicker upper tail, namely the log-normal and chi-square.

As discussed in Section 6.2, formal hypothesis testing is straightforward, requiring a mere linear restriction test on the coefficients of a linear regression. Table 2 presents the results of this test for each distribution.<sup>15</sup> The conclusions coincide roughly with the ones reached informally with the graphs:

---

<sup>14</sup>Because the number of auctions with many bids is small (cf. Table 7 at end of the paper), it is convenient to introduce dummies for high values of  $n$ . All five regressions include dummies for each  $n$  above 20.

<sup>15</sup>The test statistic is the F-test based on  $R^2$  differences, for the hypothesis that the estimated dummies for  $\hat{a}(2)$  to  $\hat{a}(20)$  is equal to the numbers in each column of table 1, up

the hypotheses of uniform or normal residuals are rejected at conventional significance levels.

$H_0$ : residuals are	$R^2$	test statistic	p-value
Uniform	0.2337	4.0901	0.0000
Normal	0.2379	3.3461	0.0000
Log-normal	0.2487	1.4236	0.1154
Chi-square	0.2490	1.3846	0.1339
Unrestricted	0.2568	—	—

Table 2: Hypothesis testing on the distribution of residuals.

## 7.1 Testing for Gaussian Common Values

Tests have also been run for the hypothesis that the true data generating process is an English common value auction with distributional assumptions as described in appendix A. Tests have been run for  $s^2 = 1/2, 1, 2, 4, 16$  and  $64$ .<sup>16</sup>

$s^2$	$R^2$	test statistic	p-value
1/2	0.2042	9.3214	0
1	0.2105	8.2120	0
2	0.2219	6.1900	0.0000
4	0.2315	4.4920	0.0000
16	0.2455	2.0024	0.0087
64	0.2319	4.4152	0.0000
Unrestricted	0.2568		

Table 3: Hypothesis testing, Gaussian Common Values.

The results reported on table 3 show that all of the affiliated models tested have been rejected. The results therefore do not provide evidence against the independent private values assumption in this sample. One must

---

to scale and location. Here the degrees of freedom are  $57 - 30 = 17$  and  $2299 - 57 = 2242$ .

<sup>16</sup>See appendix A for details.

keep in mind, however, that these are tests on specific affiliated models. They provide no indication on whether some other affiliated model is rejected.

## 8 Concluding Remarks

This paper proposes a method that combines the strengths of two distinct branches of the auction econometrics literature, being at once computationally accessible and theoretically sound. It is also robust in a wide variety of contexts.

In a nutshell, the main finding of the paper is that one can estimate parameters that affect the location and scale of the value distribution in a simple and unbiased way, provided one controls for variables that affect bidding behavior in a flexible way. Under symmetry, the main such variable is the number of participating bidders (equation 4); in more complex asymmetric situations, it is the number of bidders of each given type (equation 5); similarly, the discussion in section 5.3.2 suggest that when entry is a concern, the appropriate regression would include a correction parameter that depends on instruments correlated with the entry decision process (equation 8).

As such, the method provides a simple way to separate the effect of regressors that affect bidding only by affecting the value distribution from those that affect bidding strategically. It provides an useful tool for researchers interested in measuring the former while properly controlling for the latter; conversely, it helps identify the misspecification bias that occurs when the latter are not accounted for.

And because it not only provides a way to control for variation across auctions, but also allows for testing about value distribution assumptions, it also might be a helpful preliminary exploratory tool for researchers interested in applying more ambitious structural econometric models in auction data.

## A Examples with Affiliation

This section provides examples of the methodology applied to the context of the affiliated model of Milgrom and Weber (1982). In this model, the value of the item for each bidder is no longer private and independent; it is rather a function of a number of affiliated random variables.  $V_{il}$ , the value of the



item for bidder  $i$  in auction  $l$ , is

$$V_{il} = V_l(S_i^l, S_{-i}^l, S_0^l),$$

where  $S_i^l$  is the signal of bidder  $i$  in auction  $l$ ,  $S_{-i}^l$  is the vector of signals of the other bidders, and  $S_0^l$  is an additional random variable or vector not observed by any bidder. Each bidder  $j$  observes  $S_j^l$  prior to the auction, but not  $V_{jl}$ .  $S_0^l, \dots, S_n^l$  are affiliated, and  $V_l$  is increasing in all arguments.

We parameterize the model in a way analogous to Assumption 1:

**Assumption 5**  $V_{il} = \mu_l + \sigma_l \nu_{il}$ , where  $\nu_{il} = \nu(S_i^l, S_{-i}^l, S_0^l)$ . The distribution of  $(S_0^l, \dots, S_n^l)$  is *i.i.d.* across auctions, and  $\nu$  is the same across auctions.

As before, all variation across auctions is captured by the mean and the variance of the valuations. Signals and values are allowed to exhibit fairly complex dependencies within an auction, but once one normalizes the values, one obtains a random sample of auctions.

One feature that is lost once we move away from the independent private values framework in the Expected Revenue Equivalence Theorem. Now the expected price for the good will depend on the specific auction format to be used. Following Milgrom and Weber (1982), we discuss three possible auction formats: the second-price or Vickrey auction, the English auction, and the first-price auction. We distinguish bid functions and resulting prices across these auctions using superscripts  $V$ ,  $E$  and  $F$ , respectively. Milgrom and Weber (1982) have provided characterizations for equilibria in each of the three alternative mechanisms.

In the Vickrey auction  $l$ , it is an equilibrium to bid according to

$$b_i^V(s) = E[V_{il} | S_i^l = s, \max(S_{-i}^l) = s],$$

where  $b_i^V(s)$  represents the bid of bidder  $i$  when it observes a signal  $s$ . Assuming  $b^V$  is increasing, and since the affiliated model is symmetric, the winner is the bidder with the highest signal, and the final price will be  $p^V = b^V(S_{(2:n_l)}^l)$ . Using assumptions 2 and 5, we can write the expected price as follows:

$$\begin{aligned} E[p_l^V | X_l, Z_l, n_l] &= E[b_l^V(S_{(2:n_l)}^l) | X_l, Z_l, n_l] \\ &= E[E[V_{il} | S_i^l = S_{(2:n_l)}^l, S_{(1:n_l)}^l = S_{(2:n_l)}^l] | X_l, Z_l, n_l] \\ &= E[E[\mu_l + \sigma_l \nu_i | S_i^l = S_{(2:n_l)}^l, S_{(1:n_l)}^l = S_{(2:n_l)}^l] | X_l, Z_l, n_l] \\ &= \mu_l + \sigma_l E[\nu_i | S_i^l = S_{(1:n_l)}^l = S_{(2:n_l)}^l], \end{aligned}$$

where the second equality comes from the equilibrium bidding function, the third from the specification of  $V_{il}$ , and the fourth from linearity of the conditional expectation operator. Defining  $a^V(n) = E[\nu_i | S_i^l = S_{(2:n)}^l, S_{(1:n)}^l = S_{(2:n)}^l]$ , we have

$$E[p_l^V | X_l, Z_l, n_l] = X_l \beta + a^V(n_l) Z_l \alpha. \quad (9)$$

As in equation 3, the conditional expectation of the price is linear, and OLS provides unbiased and consistent estimates. Both methods 1 and 2 can again be used; the only change in method 1 needed to accommodate affiliation would be to calculate  $a^V(n) = E[\nu_i | S_i^l = S_{(2:n)}^l, S_{(1:n)}^l = S_{(2:n)}^l]$  instead of  $a(n) = E[\epsilon_{(2:n)}]$ . Method 2 does not require any change.

The same result obtains in the English auction. In that case, a bid function must specify a drop-out point for each signal and each history of drop-out points by the opponents. Let  $b_l^E(s, \emptyset)$  is the drop-out point of a bidder with signal  $s$  who has not seen anybody drop out, and  $b_l^E(s, (d_1, \dots, d_k))$  is the one for a bidder that has observed a history  $(d_1, \dots, d_k)$  of drop-out points.

Milgrom and Weber (1982) characterize an equilibrium to this game recursively as follows:

$$\begin{aligned} b_l^E(s, \emptyset) &= E[V_{il} | S_i^l = s, S_{-i(1)}^l = \dots = S_{-i(n-1)}^l = s], \\ b_l^E(s, (d_1)) &= E[V_{il} | S_i^l = s, S_{-i(1)}^l = \dots = S_{-i(n-2)}^l = s, b_l^E(S_{-i(n-1)}^l, \emptyset) = d_1], \\ b_l^E(s, (d_1, d_2)) &= E[V_{il} | S_i^l = s, S_{-i(1)}^l = \dots = S_{-i(n-3)}^l = s, b_l^E(S_{-i(n-2)}^l, (d_1)) = d_2, \\ &\quad b_l^E(S_{-i(n-1)}^l, \emptyset) = d_1], \\ &\vdots \end{aligned}$$

Assuming that these functions are one-to-one on  $s$ , one can infer the signal of each losing bidder from their drop-out points. The auction price, the drop-out point of the final losing bidder, is

$$p^E = E[V_{(2:n)l} | S_{(1)}^l = S_{(2)}^l, S_{(2)}^l, \dots, S_{(n)}^l],$$

that is, the estimated value for this bidder if we know the signals of all losers and assume the winner has the same signal as the last loser.

Since the price is still a conditional expectation, linearity holds in this case as well. We obtain

$$E[p_l^E | X_l, Z_l, n_l] = X_l \beta + a^E(n_l) Z_l \alpha, \quad (10)$$

where this time  $a^E(n) = E[E[\nu_i | S_i^l = S_{(2)}^l, S_{(1:n)}^l = S_{(2:n)}^l, S_{(2)}^l, \dots, S_{(n)}^l]]$ . So again method 1 is applicable using this expression for  $a(n)$ , and method 2 is applicable unchanged.

We now turn to the case of a first-price auction. Milgrom and Weber (1982) show that in a symmetric equilibrium of this game the bid function is

$$b^F(s) = \int^s b^V(t) dL(t|s),$$

where

$$L(t|s) = \exp\left(-\int_t^s h(y|y) dy\right),$$

and  $h(x|y)$  is the hazard rate of the opponents' highest signal, conditional on an individual's signal. Observe that  $L$  is a function of the signal structure of the model only, and not of the value function. So it does not change across auctions.

So we can write

$$\begin{aligned} b^F(s) &= \int^s E[\mu_l + \sigma_l \nu_i | S_i^l = t, \max(S_{-i}^l) = t] dL(t|s) \\ &= \mu_l + \sigma_l \int^s E[\nu_i | S_i^l = t, \max(S_{-i}^l) = t] dL(t|s), \end{aligned}$$

and therefore the expected price is

$$E[p_l^F | X_l, Z_l, n_l] = X_l \beta + a^F(n_l) Z_l \alpha, \quad (11)$$

where this time  $a^F(n) = E\left[\int^{S_{(1)}^l} E[\nu_i | S_i^l = t, \max(S_{-i}^l) = t] dL(t|S_{(1)}^l)\right]$ .

In all three cases the methodology suggested works. Unbiased, consistent estimates of the coefficients that determine the expectation and variance of the valuations can be obtained by OLS even in the more general affiliated framework. In order to apply method 1 the only necessary modification is to redefine the artificial regressor  $a(n)$  in a way that depends on the specific auction rule. For method 2, no changes are necessary.

We now provide artificial regressors  $\tilde{a}(n)$  for a particular example of an affiliated model: a case where agents have a common value for the object, and only differ because they observe different signals of that value.

In order to obtain specific figures for  $\tilde{a}(n)$ , we assume  $V_{il} = V_i \sim \mathcal{N}(\mu_l, \sigma_l^2)$  is common across bidders within an auction. Let  $v_l = (V_l - \mu_l)/\sigma_l$  be the

normalized value. Bidders observe different signals  $S_i^l$ , with  $S_i^l = v_l + \epsilon_{il}$ , with disturbances  $\epsilon_{il} \sim \mathcal{N}(V_l, s^2)$  independent of each other and  $v_l$ .

This model is a special case of the affiliated model, known in the literature as the common-values or mineral rights model. Here we allow for variation in the mean and variance of the valuations across auctions as before. We impose a specific distributional assumption for valuations and signals in order to compute artificial regressors.<sup>17</sup>

Artificial regressors have been computed for the English auction, the Vickrey auction and for the First-price auction, for all numbers of bidders between 2 and 20. They are reported in Table 4. Three values for the variance of the signal error have been used: 0.5, 1 and 2. Since the variance of the normalized value is 1, these choices represent the cases where uncertainty about signal errors is smaller than, equal, or greater than uncertainty about the value.

## B Proofs

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

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

so the recurrence relation holds.  $\square$

**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  $\omega_{(k:n+1)}$ .  $\square$

**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  $F^{-1}(\alpha)$ .

---

<sup>17</sup>We truncate the distributions of normalized values at the interval  $[-10, 10]$ . This is done both for computational convenience and to make the model satisfy the requirement in Milgrom and Weber (1982) that restricts attention to variables in a compact interval. Numerical results do not seem to be sensitive to the truncation point.

$n$	English auction			Vickrey auction			First-price auction		
	$s = 1/2$	$s = 1$	$s = 2$	$s = 1/2$	$s = 1$	$s = 2$	$s = 1/2$	$s = 1$	$s = 2$
2	-0.2508	-0.3761	-0.3761	-0.2508	-0.3761	-0.3761	-0.4522	-0.5443	-0.4390
3	-0.1302	-0.2116	-0.2418	-0.1483	-0.2341	-0.2551	-0.2674	-0.3508	-0.3077
4	-0.0862	-0.1465	-0.1831	-0.1101	-0.1789	-0.2055	-0.2001	-0.2750	-0.2538
5	-0.0636	-0.1113	-0.1484	-0.0897	-0.1487	-0.1772	-0.1646	-0.2331	-0.2231
6	-0.0500	-0.0894	-0.1251	-0.0770	-0.1294	-0.1586	-0.1423	-0.2061	-0.2027
7	-0.0410	-0.0744	-0.1081	-0.0682	-0.1158	-0.1451	-0.1270	-0.1870	-0.1880
8	-0.0346	-0.0635	-0.0952	-0.0617	-0.1057	-0.1349	-0.1157	-0.1726	-0.1767
9	-0.0299	-0.0553	-0.0850	-0.0568	-0.0979	-0.1267	-0.1069	-0.1614	-0.1677
10	-0.0262	-0.0489	-0.0768	-0.0528	-0.0916	-0.1201	-0.1000	-0.1523	-0.1604
11	-0.0233	-0.0437	-0.0699	-0.0495	-0.0864	-0.1146	-0.0942	-0.1447	-0.1542
12	-0.0210	-0.0395	-0.0642	-0.0468	-0.0820	-0.1098	-0.0895	-0.1383	-0.1489
13	-0.0190	-0.0360	-0.0593	-0.0445	-0.0782	-0.1057	-0.0854	-0.1329	-0.1443
14	-0.0174	-0.0330	-0.0550	-0.0425	-0.0750	-0.1021	-0.0819	-0.1281	-0.1403
15	-0.0160	-0.0305	-0.0513	-0.0408	-0.0722	-0.0990	-0.0788	-0.1239	-0.1367
16	-0.0148	-0.0283	-0.0481	-0.0393	-0.0696	-0.0961	-0.0761	-0.1201	-0.1335
17	-0.0138	-0.0264	-0.0452	-0.0379	-0.0674	-0.0936	-0.0736	-0.1168	-0.1306
18	-0.0129	-0.0247	-0.0427	-0.0367	-0.0654	-0.0913	-0.0715	-0.1138	-0.1280
19	-0.0121	-0.0232	-0.0404	-0.0356	-0.0635	-0.0892	-0.0695	-0.1110	-0.1256
20	-0.0114	-0.0219	-0.0383	-0.0346	-0.0619	-0.0873	-0.0677	-0.1085	-0.1234

Table 4: Values of  $a(n)$ , Gaussian common values.

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

$$\begin{aligned} \frac{\int_0^u [t^{\alpha_n}(1-t)^{1-\alpha_n}]^n dt}{\int_0^1 [t^{\alpha_n}(1-t)^{1-\alpha_n}]^n dt} &\leq \frac{\int_0^u [t^{\alpha_n}(1-t)^{1-\alpha_n}]^n dt}{\int_v^{\alpha_n} [t^{\alpha_n}(1-t)^{1-\alpha_n}]^n dt} \leq \\ \frac{\int_0^u [u^{\alpha_n}(1-u)^{1-\alpha_n}]^n dt}{\int_v^{\alpha_n} [v^{\alpha_n}(1-v)^{1-\alpha_n}]^n dt} &= \frac{u}{\alpha_n - v} \left[ \frac{u^{\alpha_n}(1-u)^{1-\alpha_n}}{v^{\alpha_n}(1-v)^{1-\alpha_n}} \right]^n \rightarrow 0. \end{aligned}$$

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

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

as  $n \rightarrow \infty$ , 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

- Amemiya, Y. (1985), Instrumental variable estimator for the nonlinear errors-in-variables model, *Journal of Econometrics* 28, 3, 273–289.
- Athey, S. and Haile, P. A. (2002), Identification of standard auction models, *Econometrica* 70, 6, 2107–2140.
- Bajari, P. (1999), Sealed-bid auction with asymmetric bidders: Theory and computation, Stanford University.
- Bajari, P. and Hortag̃su, A. (2003), The winner’s curse, reserve prices, and endogenous entry: Empirical insights from ebay auctions, *Rand Journal of Economics* 34, 2.
- Bikhchandani, S., Haile, P. A. and Riley, J. G. (2002), Symmetric separating equilibria in english auctions, *Games and Economic Behavior* 38, 19–27.
- Campo, S., Perrigne, I. and Vuong, Q. (2003), Asymmetry in first-price auctions with affiliate private values, *Journal of Applied Econometrics* 18, 2, 179–207.
- Chan, L. K. (1967), On a characterization of distribution by expected values of extreme order statistics, *American Mathematics Monthly* 74, 950–951.
- Deltas, G. and Chakraborty, I. (2001), Robust paratric analysis of first price auctions, University of Illinois and University of Oklahoma.
- Donald, S. G. and Paarsch, H. J. (1993), Piecewise pseudo-maximum likelihood estimation in empirical models of auctions, *International Economic Review* 34, 1, 121–148.
- Garcia, M. and Rezende, L. (2000), Short-term security auctions by brazilian central bank: Study of conditioning factors of dispersion of the bids, *Brazilian Journal of Political Economy* 20, 4, 8–25, (in Portuguese).
- Gilley, O. W. and Karels, G. V. (1981), The competitive effect in bonus bidding: New evidence, *Bell Journal of Economics* 12, 2, 637–648.
- Guerre, E., Perrigne, I. and Vuong, Q. (2000), Optimal nonparametric estimation of first-price auctions, *Econometrica* 68, 3, 525–574.

- Haile, P., Hong, H. and Shum, M. (2003), Nonparametric tests for common values in first-price sealed-bid auctions, Yale, Duke and John Hopkins.
- Haile, P. and Tamer, E. (2003), Inference with an incomplete model of english auctions, *Journal of Political Economy* 111, 1–51.
- Hoeffding, W. (1953), On the distribution of the expected values of the order statistics, *Annals of Mathematical Statistics* 24, 93–100.
- Hong, H. and Tamer, E. (2003), A simple estimator for nonlinear error in variable models, *Journal of Econometrics* 117, 1, 1–19.
- Houser, D. and Wooders, J. (2000), Reputation in auctions: Theory, and evidence from ebay, University of Arizona.
- Hu, Y. (2004), Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution, Texas.
- Hu, Y. and Ridder, G. (2003), Estimation of nonlinear models with measurement error using marginal information, Texas and USC.
- Krasnokutskaya, E. (2002), Identification and estimation in highway procurement auctions under unobserved auction heterogeneity, Yale.
- Laffont, J.-J., Ossard, H. and Vuong, Q. (1995), Econometrics of first-price auctions, *Econometrica* 63, 4, 953–980.
- Laffont, J.-J. and Vuong, Q. (1996), Structural analysis of auction data, *American Economic Review* 86, 2, 414–420.
- Lebrun, B. (1996), Existence of equilibrium in first price auctions, *Economic Theory* 7, 421–443.
- Lee, L.-F. and Sepanski, J. H. (1995), Estimation of linear and nonlinear error-in-variables models using validation data, *Journal of the American Statistical Association* 90, 429, 130–140.
- Levin, D. and Smith, J. L. (1994), Equilibrium in auctions with entry, *American Economic Review* 84, 3, 585–599.
- Lucking-Reiley, D., Bryan, D., Prasad, N. and Reeves, D. (2000), Pennies from ebay: the determinant of price in online auctions, University of Arizona.



- McAfee, R. P. and McMillan, J. (1987a), Auctions with a stochastic number of bidders, *Journal of Economic Theory* 43, 1, 1–19.
- (1987b), Auctions with entry, *Economics Letters* 23, 4, 343–347.
- McDonald, C. G. and Slawson, V. C. (2002), Reputation in an internet auction market, *Economic Inquiry* 40, 4, 633–50.
- Melnik, M. I. and Alm, J. (2002), Does a seller’s ecommerce reputation matter? evidence from ebay auctions, *Journal of Industrial Economics* 50, 3, 337–49.
- Milgrom, P. R. and Weber, R. J. (1982), A theory of auctions and competitive bidding, *Econometrica* 50, 5, 1089–1122.
- Myerson, R. B. (1981), Optimal auction design, *Mathematics of Operations Research* 6, 1, 58–73.
- Newey, W. K. (2001), Flexible simulated moment estimation of nonlinear errors-in-variables models, *Review of Economics and Statistics* 83, 4, 616–627.
- Olley, G. S. and Pakes, A. (1996), The dynamics of productivity in the telecommunications equipment industry, *Econometrica* 64, 6, 1263–1297.
- Paarsch, H. J. (1991), Empirical models of auctions and an application to british columbian timber, Research Report 9212, University of Western Ontario Department of Economics.
- Pollak, M. (1973), On equal distributions, *Annals of Statistics* 1, 180–182.
- Samuelson, W. F. (1985), Competitive bidding with entry costs, *Economics Letters* 17, 53–57.
- Vickrey, W. (1961), Counterspeculation, auctions, and competitive sealed tenders, *Journal of Finance* 16, 1, 8–37.

regressor	coef.	st.dev.	regr.	coef.	st.dev.	regr.	coef.	st.dev.
broken	-106.66	28.70	21	198.53	14.39	42	212.50	102.50
2	58.82	7.93	22	220.01	12.91	43	292.50	72.48
3	59.63	8.63	23	217.47	14.21	44	293.35	59.17
4	91.68	9.81	24	218.39	13.94	45	300.00	102.50
5	109.82	10.68	25	221.85	14.49	46	322.00	102.50
6	109.99	10.68	26	219.85	14.07	47	310.00	102.50
7	136.57	10.74	27	226.79	17.32	48	167.50	102.50
8	145.31	11.32	28	242.82	19.03	53	305.00	102.50
9	182.05	10.93	29	242.71	18.71	58	222.50	102.50
10	139.77	10.64	30	236.44	17.57	64	281.00	102.50
11	160.15	11.46	31	233.96	20.50	67	225.00	102.50
12	162.11	10.06	32	228.80	26.46	90	291.00	102.50
13	184.43	10.63	33	246.58	29.58	93	205.00	102.50
14	199.03	12.08	34	239.61	28.42	106	280.00	102.50
15	198.21	11.99	35	172.84	34.16	126	285.50	102.50
16	187.46	12.61	36	182.14	38.74	128	215.50	102.50
17	214.40	13.23	37	242.10	45.84	134	280.00	102.50
18	217.64	11.68	38	190.00	51.25			
19	228.21	10.92	39	206.25	51.25			
20	204.15	12.92	40	278.50	59.17			

Table 5: Estimated coefficients (and standard deviations) under free distributional assumptions.

regressor	uniform		normal		lognormal		chi-square	
	coef.	st.dev.	coef.	st.dev.	coef.	st.dev.	coef.	st.dev.
broken	-109.98	28.94	-109.45	28.86	-108.00	28.65	-108.50	28.65
intercept	77.72	4.17	79.48	4.05	93.45	3.45	90.65	3.54
slope	78.06	3.83	81.53	3.93	95.45	4.41	81.25	3.75
21	120.94	15.14	119.16	15.07	105.13	14.82	107.95	14.84
22	142.29	13.71	140.53	13.64	126.56	13.38	129.35	13.41
23	139.75	14.97	137.98	14.90	124.02	14.65	126.81	14.67
24	140.67	14.71	138.91	14.64	124.94	14.39	127.73	14.41
25	144.13	15.24	142.37	15.17	128.40	14.92	131.19	14.94
26	142.13	14.84	140.37	14.77	126.40	14.51	129.20	14.53
27	149.07	18.01	147.31	17.94	133.34	17.69	136.14	17.71
28	165.10	19.70	163.34	19.62	149.37	19.37	152.17	19.38
29	164.99	19.38	163.23	19.30	149.26	19.06	152.05	19.07
30	158.71	18.26	156.95	18.19	142.99	17.94	145.78	17.95
31	156.24	21.15	154.47	21.07	140.51	20.82	143.30	20.83
32	151.07	27.09	149.31	27.00	135.35	26.73	138.14	26.74
33	168.86	30.22	167.09	30.12	153.13	29.83	155.92	29.84
34	161.89	29.05	160.13	28.96	146.16	28.68	148.95	28.69
35	95.12	34.81	93.36	34.70	79.39	34.39	82.18	34.40
36	104.42	39.41	102.65	39.29	88.69	38.95	91.48	38.96
37	164.37	46.55	162.61	46.42	148.64	46.04	151.44	46.04
38	112.27	52.01	110.51	51.86	96.54	51.44	99.34	51.44
39	128.53	52.01	126.76	51.86	112.80	51.44	115.59	51.44
40	200.78	60.00	199.01	59.83	185.05	59.37	187.84	59.37

Table 6: Estimated coefficients (and standard deviations) under selected distributional assumptions, linear formulation. (Coefficients for  $n > 40$  were included but not reported.)

bids	auctions	bids	auctions	bids	auctions
2	167	21	51	40	3
3	141	22	63	42	1
4	109	23	52	43	2
5	92	24	54	44	3
6	92	25	50	45	1
7	91	26	53	46	1
8	82	27	35	47	1
9	88	28	29	48	1
10	93	29	30	53	1
11	80	30	34	58	1
12	104	31	25	64	1
13	93	32	15	67	1
14	72	33	12	90	1
15	73	34	13	93	1
16	66	35	9	106	1
17	60	36	7	126	1
18	77	37	5	128	1
19	88	38	4	134	1
20	63	39	4		

Table 7: Number of auctions, by number of bids.