Identification in English Auctions with Shill Bidding*

Guillermo Marshall†

August 8, 2023

Abstract

What can we learn from auction data when the seller submits shill bids to inflate the auction price? I study identification in an incomplete model of an English auction with shill bidding, in the context of independent private values. I show that the distribution of valuations is partially identified (as is the optimal reserve price), and I provide bounds for the distribution of valuations that hold even when the seller is not engaging in shill bidding. I apply these results to a sample of eBay auctions.

Keywords: Auctions, partial identification, shill bidding, eBay

1 Introduction

The use of online auctions exploded in the early days of Internet commerce (Einav, Farronato, Levin and Sundaresan 2018). Online auction platforms have reduced matching frictions between buyers and sellers, creating new trade opportunities, but these online marketplaces are far from frictionless. In a survey by the UK's Office of Fair Trading, online auction users reported that shill bidding—i.e., when the seller bids in their own auction to inflate the price—was a common problem on online auction platforms (OFT 2007).[1] A number of eBay users in the UK and US have been prosecuted for using fake identities to drive up auction sales prices, and the penny auctions platform PennyBiddr was even shut down when the

*I thank Günter Hitsch (Editor), Jeff Ely, Álvaro Parra, Rob Porter, two anonymous referees, and workshop participants at Northwestern University for helpful comments and suggestions. I am supported in part by funding from the Social Sciences and Humanities Research Council (Canada). The author has no competing interests to declare that are relevant to the content of this article. All errors are my own.

†University of British Columbia, Sauder School of Business; guillermo.marshall@sauder.ubc.ca

[1] Shill bidding is also considered a criminal fraud in the UK and US (Kauffman and Wood 2005).

platform itself faced allegations of shill bidding.[2] Grether, Porter and Shum (2015) present evidence suggestive of shill bidding on Copart (an online car auctions platform), and a lawsuit alleges the use of shill bidding on platform Auction.com (online real estate auctions).[3]

Why would a seller engage in shill bidding? There are a number of explanations. A first one is that some online auction platforms charge sellers a fee for posting a reserve price (e.g., eBay), which is a fee that the seller can avoid if they use shill bids in lieu of a reserve price (Kauffman and Wood 2005). A second one is that platforms such as eBay have a "second chance offer" feature, which allows a seller to allocate the object to the second highest bidder in the event that the winner fails to pay. This feature of the platform substantially reduces a seller's cost of engaging in shill bidding, as the seller can allocate the good to the second highest bidder in the event that the "winning bid" is a shill bid. A third explanation is that sellers may face uncertainty about primitives that impact the optimal reserve price. For example, a seller may face uncertainty about the exact composition of buyers (if bidders are asymmetric) or the number of bidders that will participate in the auction. As the seller observes the bidding process, the seller can learn about the composition of buyers or the number of bidders and use shill bids to "adjust" the reserve price (Graham, Marshall and Richard 1990, Wang, Hidvégi and Whinston 2001, Andreyanov and Caoui 2020).[4]

I study identification in an English auction (i.e., ascending price auction) with shill bidding in the symmetric independent private values framework.[5] Specifically, I investigate what an econometrician can learn about the distribution of valuations and the optimal reserve price when using auction data that may be contaminated with shill bidding. The auction model is incomplete in that I make weak assumptions about the behavior of bidders and sellers. In particular, I do not make assumptions about the bidding behavior of the seller.

I find that the distribution of valuations is partially identified in the presence of shill bidding. To see why, consider an auction with $n+1$ bidders, where n are legitimate bidders

[2]See, for example, "3 Men Are Charged With Fraud In 1,100 Art Auctions on EBay," *The New York Times*, March 9, 2001, "Phony Bids Pose Difficulties, Putting eBay on the Defensive," *The Wall Street Journal*, May 24, 2000, "Officials Accuse Three in Scam To Drive Up Prices in eBay Bids," *The Wall Street Journal*, February 8, 2002, and "How do you catch online auction cheats?," *BBC News*, July 5, 2010. .

[3]See, e.g., "Lawsuit accuses Auction.com of using 'shill bidder"', *New York Post*, December 25, 2014.

[4]The optimal reserve price can vary with the number of bidders in a number of cases. With independent private values, this may occur whenever the distribution of valuations $F(v)$ is such that $v-(1-F(v))/F'(v)$ is not monotone increasing (see the discussion in Wang et al. (2001)). With affiliated private values, the valuations may depend on a common factor (e.g., market conditions), which may also affect the seller's valuation for the object (i.e., the value from a future sale if the object does not sell in the auction).

[5]See, for example, Hasker and Sickles (2010a) for a survey of the use of eBay data in the economic literature.

and one is a shill bidder. Assume no minimum bid increments and that the game ends when there is only one bidder left. Define the auction price as the lowest price at which only one bidder remains active. The presence of a shill bidder implies that the auction price may not necessarily be the second highest valuation among the n legitimate bidders. If the shill bidder places the highest bid, then the auction price would be the highest valuation among the n legitimate bidders, whereas if a legitimate bidder places the highest bid, then the auction price is the greater between the shill bid and the second highest valuation among the legitimate bidders. As a result, the auction price is bounded between the second and first highest valuations among the n legitimate bidders. This inequality is the basis of the identification region for the distribution of valuations when the econometrician only observes the auction price. I then use these bounds to investigate what an econometrician can learn about the optimal reserve price of an auction—taking the perspective of a seller who wishes to sell an object without engaging in shill bidding.

I argue that even if there is no shill bidding in the data (i.e., all $N + 1$ bidders are legitimate), the true distribution of valuations will still be contained in this identification region, provided that the assumptions of the symmetric independent private values setting hold. That is, the bounds that I derive hold regardless of shill bidding. I discuss how this result can also be used to implement a specification test for a particular complete model in which at least N out of $N+1$ bidders draw their values independently from some distribution.

Given the general concern about shill bidding in online auctions, the methods I develop can be applied to settings as varied as eBay auctions, real estate auctions (auction.com), or car auctions (Copart). I apply these results to a sample of eBay auctions for 3.4-oz bottles of Armani Acqua di Gio perfume (mint condition), which took place between the years 2008 and 2010. I estimate the identification regions for the distribution of valuations and optimal reserve price and discuss the informativeness of these bounds in that particular context. I also compare the estimates of the identification region of the distribution of valuations with estimates of the distribution of valuations using methods that rule out shill bidding.

This paper contributes to the literature on identification in auctions. Athey and Haile (2002) study identification in standard auctions. Haile and Tamer (2003) present identification results in an English auction that deviates from the "button auction" abstraction imposing weak conditions on bidder behavior. Song (2004) and Hickman, Hubbard and Paarsch (2017) propose methods to identify the distribution of valuations in online ascending-price auctions with a potentially unknown number of bidders. Tang (2011) bounds the revenue distributions of an auction under counterfactual formats. The results are derived without imposing parametric restrictions on the model structure and allow for affiliated values and signals. In related work, Coey, Larsen and Sweeney (2019) propose a test of independence of valuations and the number of valuations in ascending button auctions with symmetric independent private values, which can be used to bound counterfactual revenue distributions. Beyond the independent private values framework, Aradillas-López, Gandhi and Quint (2013) provide identification results for ascending price auctions with correlated private values. More broadly, see Athey and Haile (2007) and Hendricks and Porter (2007) for literature surveys.

Among these, Haile and Tamer (2003) is the closest to my work, as I draw from their work in specifying an incomplete model of an English auction. The key difference, however, is that my identification results can handle the potential presence of an active seller. Beyond that, I also provide partial identification results in the case in which the number of potential bidders is unobserved.

This paper is also related to the literature studying collusion in auctions, which is another type of auction fraud. The empirical literature studying collusion in auctions has mostly focused on detection and testing competitive versus collusive bidding rather than on the identification of objects of interest when the data may be contaminated with fraudulent behavior. See for example Feinstein, Block and Nold (1985), Porter and Zona (1993), Baldwin, Marshall and Richard (1997), Porter and Zona (1999), and Marmer, Shneyerov and Kaplan (2016). My contribution is in providing identification results in the presence of fraudulent behavior rather than to provide methods to detect such behavior.

The paper is organized as follows. Section 2 discusses why a seller would want to use shill bidding in the context of the independent private values framework. Section 3 presents results on the identification of the distribution of valuations and discusses estimation. Section 4 presents extensions to the baseline framework. Section 5 investigates whether the optimal reserve price can be identified from auction data with shill bidding. Section 6 presents the empirical application using eBay auction data, and Section 7 concludes.

2 Shill bidding in the independent private values framework

Why may shill bidding arise in an auction in the context of the independent private values framework? For at least two reasons: auction design and platform-specific features.

Consider a seller that values a good in $v_0 \geq 0$ and faces n bidders who draw their valuations for the object independently from the distribution F_v with support $[0, \bar{v}]$ (with $v_0 \leq \bar{v}$). The seller uses an ascending price auction, where the winner pays the second highest bid (i.e., the price at which the second to last bidder drops out of the auction).

Assume that bidders play the weakly-dominant strategy of bidding their valuation. The optimal reserve price is then given by the solution to the problem of maximizing expected revenue:

$$\max_{r \in [\underline{v}, \bar{v}]} v_0 F_v(r)^n + n \int_r^{\bar{v}} \left(v - \frac{1 - F_v(v)}{F_v'(v)} \right) F_v'(v) F_v^{n-1}(v) dv \tag{1}$$

(Riley and Samuelson 1981). The first-order condition is given by

$$v_0 = r - \frac{1 - F_v(r)}{F_v'(r)}. \tag{2}$$

If the right-hand side of equation (2) is monotone increasing, then the first-order condition has one solution and the optimal reserve price does not depend on the number of bidders (Riley and Samuelson 1981). In contrast, if the right-hand side of equation (2) is not monotone, then the first-order condition may have multiple solutions, and the optimal reserve price will generally depend on the number of bidders, n. Wang et al. (2001) provide an example where the valuations distribute normal, $F_v = 0.95 \cdot N(20, 20^2) + 0.05 \cdot N(120, 20^2)$, and $v_0 = 20$. In this example, equation (2) has multiple solutions, and the optimal reserve price is 38 when $n \leq 11$ and 98 if $n > 12$.

What does this imply for shill bidding behavior? If $x - (1 - F_v(x))/F_v'(x)$ is not monotone increasing, then the optimal reserve price will depend on n. If the seller does not know the number of bidders that will participate in the auction (n) at the time of setting the reserve price, the seller can set the reserve price at the lowest of the reserve price candidates, and then engage in shill bidding to "increase the reserve price" as a function of the actual number of bidders that participate in the auction (i.e., if a different reserve price candidate is optimal given n). That is, shill bidding may increase the expected revenue of the seller.

Shill bidding may also arise because of features of the platform. eBay charges the seller for setting a reserve price. Some sellers may prefer to avoid this payment and use shill bids in lieu of a reserve price. eBay also has a "second chance offer" feature, which allows a seller to allocate the object to the second highest bidder if the winner of the auction fails to make their payment. This feature of the platform incentivizes the seller to engage in shill bidding to "increase competition," as the seller can always allocate the good to the second highest bidder in the event that the "winning bid" is the seller's shill bid.

3 Identification of the distribution of valuations

Consider an English auction (open ascending-price auction) with $N+1$ potential bidders, where one of the bidders may be a shill bidder, while all other bidders are *legitimate* bidders. The legitimate bidders have valuations for the object that are independently drawn from a distribution F_v. The valuation and bid of bidder j are given by V_j and B_j, respectively. Let $V_{k:n}$ and $B_{k:n}$ be the k-th highest valuation and bid, respectively, in a sample of n bidders. Similarly, define $F_{k:n}$ as the distribution of the k-th highest valuation in a sample of n bidders.

The auction format is such that players may submit bids that increase the price of the object by no less than the minimum bid increment of $\Delta \geq 0$. Denote the auction price (i.e., the price at which the object is sold) in an auction with $n+1$ bidders by W_{n+1} and assume that the price rule is such that $B_{2:n+1} \leq W_{n+1} \leq B_{2:n+1} + \Delta$, where $B_{2:n+1}$ is the second highest bid.[6] The distribution of auction prices in auctions with $n+1$ bidders is given by $F_{w,(n+1)}$. In what follows, I use the term auction price and winning bid interchangeably.

I follow Haile and Tamer (2003) and make the following two assumptions about bidder behavior: 1) bidders do not bid more than they are willing to pay (i.e., $B_j \leq V_j$, for every player j); 2) bidders do not allow an opponent to win at a price they are willing to beat (i.e., $V_j \leq W + \Delta$ for all runner-ups, where W is the auction price).[7] All other aspects of the model are left unspecified, including the behavior of the shill bidder. I assume that the number of potential bidders equals the number of observed bidders (an assumption I relax in the next section) and that the econometrician only observes the auction price and the number of observed bidders.

The potential presence of a shill bidder in this incomplete model of an English auction has two implications for identification. First, the econometrician does not know whether the shill bidder or a legitimate bidder won the auction, and second, the econometrician does not know whether one of the bidders is a shill bidder.

Let us consider an auction with $n+1$ bidders. Using our assumptions about bidder behavior, we can establish two facts. The first one is that the legitimate bidder with the second highest valuation has a valuation that is less than the auction price plus the minimum bid increment: $V_{2:n} \leq W_{n+1} + \Delta$. If the legitimate bidder with the second highest valuation loses the auction, this inequality holds, since otherwise the bidder would be violating the

[6]These inequalities accommodate the case in which the top two bids differ by less than the minimum bid increment. See Hickman et al. (2017) for a treatment of this case in the context of a complete model of electronic auctions.

[7]Runner-ups are defined as all players except for the one with the highest bid.

assumption that bidders do not allow an opponent to win at a price they are willing to beat. If the legitimate bidder with the second highest valuation wins the auction, it must be because the bidder with the highest valuation is constrained by the minimum bid increment and cannot beat the auction price (i.e., the top two valuations are within Δ dollars of each other): $V_{1:n} \leq W_{n+1} + \Delta$. This implies that $V_{2:n} \leq W_{n+1} + \Delta$ must hold, since $V_{2:n} \leq V_{1:n}$.

The second fact that we can establish is that that the legitimate bidder with the highest valuation has a valuation greater than the auction price minus Δ: $W_{n+1} - \Delta \leq V_{1:n}$. To see this, note that if this bidder places one of the top two bids, then $B_{2:n+1} \leq V_{1:n}$ (where the inequality comes from the assumption that bidders never bid more than their valuation), and since $W_{n+1} - \Delta \leq B_{2:n+1}$ by the price-rule assumption, the inequality holds regardless of who wins the auction. The other case to consider is when the legitimate bidder with the highest valuation places the third lowest bid, which can only happen when the valuations of the top two legitimate bidders are within Δ dollars of each other and the highest-valuation bidder is constrained by the minimum bid increment.[8] In this case, $B_{2:n+1} \leq W_{n+1} \leq B_{2:n+1} + \Delta$ (by the price-rule assumption) and $B_{2:n+1} \leq V_{2:n}$ (by the assumption that bidders do not bid more than they are willing to pay), which combined imply that $W_{n+1} - \Delta \leq B_{2:n+1} \leq V_{2:n} \leq V_{1:n}$, establishing the result.

Combining these inequalities, we have established that $V_{2:n} - \Delta \leq W_{n+1} \leq V_{1:n} + \Delta$. That is, the winning bid in an auction with n legitimate bidders and one shill bidder ($n+1$ bidders in total) is bounded between the highest and second highest valuations among all n legitimate bidders (up to a minor correction due to the minimum bid increment). These sets of inequalities combined allow the econometrician to bound the distribution of valuations, as indicated in the following proposition.[9]

Proposition 1 *Consider the environment described above, and suppose that the econometrician observes the auction price w_i and the total number of bidders $n_i + 1 \in \Omega$ of every auction i, where Ω is the set of unique values of $n+1$ that are observed by the econometrician. Then, the identification region for $F_v(t)$ is given by*

$$H[F_v(t)] = \left[\max_{n+1\in\Omega} \phi_2^{-1}(F_{w,n+1}(t-\Delta)|n), \min_{n+1\in\Omega} \phi_1^{-1}(F_{w,n+1}(t+\Delta)|n) \right] \equiv [L(t), U(t)],$$

[8] The same argument can be used if the legitimate bidders with the top k valuations all have valuations within Δ dollars of each other and the bidder with the highest valuation places a bid that is fourth highest or lower.

[9] The proposition resembles theorems 1 and 2 in Haile and Tamer (2003), but must take into account that the shill bidder may win the auction.

where $W_i \sim F_{w,n+1}(t)$ is auction price distribution when the total number of bidders is $n+1$, and $\phi_1(\cdot|n)$ and $\phi_2(\cdot|n)$ are the distribution functions of the first- and second-order statistics, defined as

$$\phi_1(x|n) = x^n \quad \textit{and} \quad \phi_2(x|n) = n(n-1)\int_0^x u^{n-2}(1-u)du.$$

Proof. From the discussion in the text, we know that $V_{2:n} - \Delta \leq W_{n+1} \leq V_{1:n} + \Delta$, which imply that $F_{w,n+1}(t-\Delta) \leq F_{2:n}(t)$ and $F_{1:n}(t) \leq F_{w,n+1}(t+\Delta)$, where $F_{k:n}$ is the distribution of the k-th highest valuation in a sample of n bidders.

Let us consider first the lower bound of the identification region when using data from auctions with $n+1$ bidders. Applying the inverse of the second-order statistic operator to both sides of $F_{w,n+1}(t-\Delta) \leq F_{2:n}(t)$, we obtain the lower bound

$$\phi_2^{-1}(F_{w,n+1}(t-\Delta)|n) \leq F_v(t),$$

where we use that $\phi_k(F_v(t)|n) = F_{k:n}$ and that $\phi_k(\cdot|n)$ is a strictly increasing function for all $1 \leq k \leq n$.

Similarly, for the upper bound of the distribution of valuations, we apply the inverse of the first-order statistic operator to $F_{1:n}(t) \leq F_{w,n+1}(t+\Delta)$, to obtain

$$F_v(t) \leq \phi_1^{-1}(F_{w,n+1}(t+\Delta)|n).$$

Putting these two bounds together, we obtain

$$\phi_2^{-1}(F_{w,n+1}(t-\Delta)|n) \leq F_v(t) \leq \phi_1^{-1}(F_{w,n+1}(t+\Delta)|n), \ \forall t, \forall n+1.$$

Intersecting these inequalities over all $n+1 \in \Omega$ yields the result. No additional information is available to make the bounds tighter. ■

The bounds for the distribution of valuation in Proposition 1 hold regardless of the behavior of the shill bidder. In particular, the bounds hold if the shill bidder behaves as a legitimate bidder (i.e., they draw a valuation from F_v and behave according to the bidder behavior assumptions discussed above). That is, the bounds Proposition 1 hold whether or not a shill bidder is active, making them informative about the distribution of evaluations in any event.

Corollary 1 *Proposition 1 provides bounds for F_v that are robust to shill bidding.*

Proposition 1 can also be used as the basis of a specification test. Proposition 1 implies that any point estimate of the distribution of valuations derived from a complete model in which the assumption of independent private values holds for at least N of the $N+1$ bidders should lie within the identification bounds. One example of a complete model would be that of a "button auction" with all $N+1$ bidders drawing their valuations independently from some distribution F_v (Milgrom and Weber 1982). Under the null hypothesis of the "button auction" model, one can estimate the distribution of valuations using standard methods, i.e., the winning bid equals the second highest valuation among all $n+1$ bidders (see, for example, the identification results in Athey and Haile (2002)). The estimate $\hat{F}_v$ should then lie within the identification region for F_v, else, the data reject that complete model. Rejection can come from bidder asymmetries (e.g., a shill bidder drawing "valuations" from a distribution that is not F_v or asymmetric bidders more broadly) or the role of minimum bid increments. This specification test does not rely on variation in the number of bidders, as does the specification test in Athey and Haile (2002); the test relies on properties of order statistics, which is a novelty, as the test can be applied even if all the auctions in the sample have the same number of bidders.

If the estimate $\hat{F}_v$ fails to lie within the identification region for F_v, a formal test can be implemented using the Cramer-von Mises criterion to test the null hypothesis that $\hat{F}_v$ equals the lower or upper bound of the distribution of valuations, depending on which bound $\hat{F}_v$ crosses (Anderson 1962).

Corollary 2 *A specification test for a complete model is given by checking whether*

$$\hat{F}_v(t) \in \hat{H}[F_v(t)]$$

holds for all t, where $\hat{F}_v(t)$ is a point estimate of $F_v(t)$ based on the assumptions of the complete model and $\hat{H}[F_v(t)]$ is an estimate of the identification region given in Proposition 1.

Lastly, I discuss two of the assumptions that I have made so far. The first is that the analysis assumes that all auctions are for identical goods. In practice, the set of auctions that are sampled may be for goods that differ in ways that are observable to the econometrician through a vector of covariates X_i for each auction i. Proposition 1 continues to hold if the econometrician conditions on a vector of covariates X, and all distribution functions are replaced by conditional distribution functions.[10]

[10] Depending on the dimensionality of the vector of covariates X, the econometrician may prefer to specify a single-index model so that the distribution of valuations only depend on the covariates through an index that depends on X and some vector of parameters β (e.g., $X'\beta$). See Paarsch, Hong et al. (2006) for a discussion on single-index models.

The second assumption is that the model allows for one source of bidder asymmetry: the presence of a shill bidder. Brendstrup and Paarsch (2006) consider identification in an English auction with asymmetric bidders in the context of the independent private values framework. If the econometrician is able to classify bidders into a set of bidder types, and observes the identity of the winner of each auction, then the authors show that the distributions of valuations of all bidder types are non-parametrically identified in a complete model of an English auction. Whether their ideas can be applied here to establish identification results for the distribution of valuations, allowing for bidder asymmetries beyond the presence of a shill bidder, is left for future research.

3.1 Estimation

Consider a sequence of T independent auctions. Each auction i has $n_i + 1$ bidders, where one of the bidders in each auction may be a shill bidder drawing their "exit point" from some arbitrary distribution. Let Ω be the set of all values of $n_i + 1$.

The estimator for the distribution function of the winning bid among all $N+1$ bids (which include legitimate bids and that of the shill bidder), $F_{w,n+1}(t)$ for every $n + 1 \in \Omega$, is given by

$$\hat{F}_{w,n+1,T}(t) = \frac{1}{T_{n+1}} \sum_{i=1}^{T} 1\{m_i = n + 1; w_i \leq t\},$$

where $T_{n+1} = \sum_{i=1}^{T} 1\{m_i = n+1\}$ and m_i and w_i are the total number of bidders (including a potential shill bidder) and the auction price of auction i.

Using these definitions, an estimator for the the identification region of the distribution of valuations is given by

$$\begin{aligned} \hat{H}_T[F_v(t)] &= \left[\max_{n+1 \in \Omega} \phi_2^{-1}(\hat{F}_{w,n+1,T}(t - \Delta)|n), \min_{n+1 \in \Omega} \phi_1^{-1}(\hat{F}_{w,n+1,T}(t + \Delta)|n) \right] \\ &\equiv [\hat{L}_T(t), \hat{U}_T(t)]. \end{aligned}$$

The following proposition establishes the consistency of this estimator.

Proposition 2 (Consistency) *Consider a sequence of T independent auctions. Each auction i has $n_i + 1 \in \Omega$ bidders, with at least n_i of them drawing their valuations independently from $F_v : [\underline{v}, \bar{v}] \rightarrow [0, 1]$ and no more than one shill bidder drawing their "exit point" from some arbitrary distribution with support $[\underline{v}, \bar{v}]$. Suppose that for each $m \in \Omega$, $T_m \rightarrow \infty$ as $T \rightarrow \infty$. Then, as $T \rightarrow \infty$, $\hat{L}_T(t) \overset{a.s.}{\rightarrow} L(t)$ and $\hat{U}_T(t) \overset{a.s.}{\rightarrow} U(t)$ uniformly in t.*

While these estimators are consistent, the estimators may be biased in small samples because of the concavity (convexity) of the min (max) function, as discussed in Haile and Tamer (2003). To see the problem, consider the estimate for the lower bound of the identification region, which amounts to taking the point-wise maximum of a number of cumulative distribution functions. In small samples, taking the maximum of these estimated cumulative distribution functions will tend to select an estimate with upward estimation error, which will lead to an upward bias of the lower bound. A similar problem arises for the upper bound of the identification region, but with a downward bias.

To alleviate the problem, Haile and Tamer (2003) replace the min (max) function in their estimators with a smooth weighted average of the estimated cumulative distribution functions that approximates the min (max). Specifically, they define the function

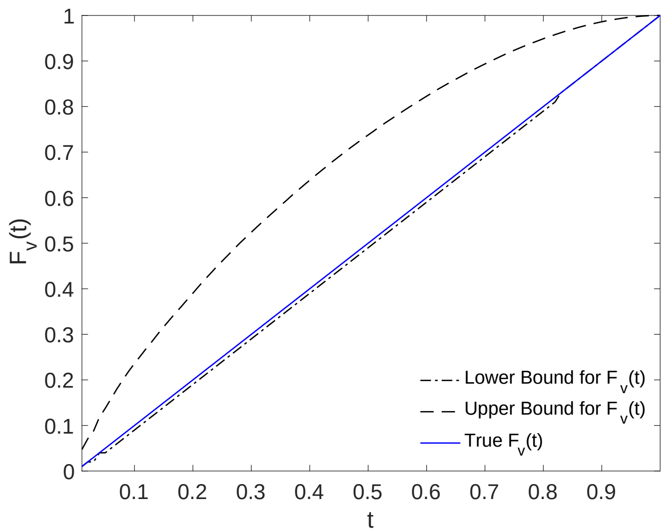
$$\mu(\hat{y}_1, \cdots, \hat{y}_J; \rho_T) = \sum_{j=1}^{J} \hat{y}_j \left[\frac{\exp(\hat{y}_j \rho_T)}{\sum_{k=1}^{J} \exp(\hat{y}_k \rho_T)} \right] \tag{3}$$

for $\rho_T \in \mathbb{R}$. When $\rho_T \to -\infty$, $\mu(\hat{y}_1, \cdots, \hat{y}_J; \rho_T)$ converges to $\min(\hat{y}_1, \cdots, \hat{y}_J)$. Likewise, when $\rho_T \to \infty$, $\mu(\hat{y}_1, \cdots, \hat{y}_J; \rho_T)$ converges to $\max(\hat{y}_1, \cdots, \hat{y}_J)$. For estimation, the authors replace the min (max) functions with the function in equation (3) choosing values of ρ_T that decrease (increase) to minus infinity (infinity) at an appropriate rate as $T \to \infty$ to ensure consistency. Following Haile and Tamer (2003), I make use of these smooth weighted averages in my empirical application to alleviate small sample bias.

3.2 Monte Carlo simulations

Consider a sequence of T=400,000 auctions, half of them with $n = 5$ legitimate bidders and the other half with $n = 4$ legitimate bidders. The valuations of all legitimate bidders are independent draws from a uniform distribution on the interval [0,1]. All auctions have one shill bidder. That is, the total number of players in an each auction i is $n_i + 1 \in \{5, 6\}$. The shill bidder bids independently up to a value S drawn from a distribution with cumulative distribution function $H(s) = s^{1/10}$ with $s \in [0, 1]$. Note that the identification results above hold regardless of whether the shill bids are independent of the behavior of bidders—I make the independence assumption here for simplicity. The minimum bid increment is $\Delta = 0$. Figure 3.2 displays the true distribution of valuations as well as the identification region for the distribution of valuations derived in Proposition 1.

Figure 1: Monte Carlo simulations: Distribution of valuations and estimated identification region using results in Proposition 1



4 Extensions

4.1 Number of potential bidders is unobserved

In this section, I relax two of the assumptions in the analysis above, one at a time. I first consider the case in which the econometrician observes the observed number of bidders, $N+1$, but not the potential number of bidders, $M+1$. In the previous section, I assumed $M = N$, but it can be the case that some potential bidders do not get to place their bids if they enter the auction at a time when the standing price exceeds their valuations. In this latter case, $N \leq M$. I assume that $M + 1 \in \{2, \ldots, \bar{M} + 1\}$, where $\bar{M}$ is known to the econometrician.

Consider an auction with $n + 1$ observed bidders, one of which may be a shill bidder. From our analysis in the previous section, we know that

$$V_{2:m} - \Delta \leq W_{n+1} \leq V_{1:m} + \Delta,$$

where $m+1$ is the number of potential bidders, which is unobserved. That m is unobserved, implies that the econometrician must consider all feasible values of m for bounding the distribution of valuations, i.e., $M \geq N$.

Using the same arguments than in Proposition 1, we can establish that

$$\min_{n \leq m \leq \bar{M}} \phi_2^{-1}(F_{w,n+1}(t - \Delta)|m) \leq F_v(t) \leq \max_{n \leq m \leq \bar{M}} \phi_1^{-1}(F_{w,n+1}(t + \Delta)|m), \tag{4}$$

for every $n+1$, where the minimum and maximum operators are used to take the union over all possible events (i.e., values of m). In Lemma 1 in Appendix B, I show that $\phi_1^{-1}(x|n)$ and $\phi_2^{-1}(x|n)$ are increasing in n for $x \in (0,1)$. Hence, the identification region simplifies to

$$\phi_2^{-1}(F_{w,n+1}(t-\Delta)|n) \leq F_v(t) \leq \phi_1^{-1}(F_{w,n+1}(t+\Delta)|\bar{M}),$$

for every $n+1$. The identification region for $F_v(t)$ is thus given by the intersection of the above inequalities over all $n+1 \in \Omega$:

$$\max_{n+1 \in \Omega} \phi_2^{-1}(F_{w,n+1}(t-\Delta)|n) \leq F_v(t) \leq \min_{n+1 \in \Omega} \phi_1^{-1}(F_{w,n+1}(t+\Delta)|\bar{M}). \tag{5}$$

Note that when $\bar{M}$ is large, the upper bound of the identification region can become uninformative, as $\phi_1^{-1}(x|n) = x^{1/n}$ approaches 1 for large n.

A less conservative approach is to form bounds for F_v that hold in expectation (where the expectation is with respect to $M+1$), which is feasible when the econometrician knows (or is able to estimate) the joint distribution of potential and observed bidders: $\Pr(M+1, N+1)$. Note that in equation (4), the econometrician must take the union over all possible values of M, as M is unobserved. In this other approach, the econometrician instead uses $\Pr(M+1, N+1)$ to take the expected value over all lower and upper bounds of F_v. The tradeoff is that the bounds only hold in expectation (rather than with certainty, as in equation (5)), but the bounds are tighter.

Consider the set of auctions with $n+1$ observed bidders. The econometrician makes use of the bounds that hold for every value of $M+1$, $\phi_2^{-1}(F_{w,n+1}(t-\Delta)|m) \leq F_v(t) \leq \phi_1^{-1}(F_{w,n+1}(t+\Delta)|m)$, and the conditional probabilities, $\Pr(M+1|n+1)$, to form the following bounds that hold in expectation:

$$\underbrace{\sum_m \Pr(m+1|n+1)\phi_2^{-1}(F_{w,n+1}(t-\Delta)|m)}_{\equiv L_{n+1}(t)} \leq F_v(t) \leq \underbrace{\sum_m \Pr(m+1|n+1)\phi_1^{-1}(F_{w,n+1}(t+\Delta)|m)}_{\equiv U_{n+1}(t)}.$$

Lastly, the econometrician can use the marginal probabilities, $\Pr(N+1)$, to combine the inequalities for every observed value of $n+1$:

$$\sum_n \Pr(n+1)L_{n+1}(t) \leq F_v(t) \leq \sum_n \Pr(n+1)U_{n+1}(t). \tag{6}$$

An intermediate approach is to use a probability threshold based on the distribution $\Pr(M+1, N+1)$ to restrict the set of values of M to be considered by the econometrician. Specifically, define the value $M_{\tau,n+1}$ such that $\Pr(M_{\tau,n+1}|n+1) = \tau$ for auctions with $N+1 =$

$n+1$ observed bidders and some critical value τ (e.g., $\tau = 0.9$). Instead of using $\bar{M}$ in the upper bound of equation (5), the econometrician can use $M_{\tau,n+1}$:

$$\max_{n+1\in\Omega} \phi_2^{-1}(F_{w,n+1}(t-\Delta)|n) \leq F_v(t) \leq \min_{n+1\in\Omega} \phi_1^{-1}(F_{w,n+1}(t+\Delta)|M_{\tau,n+1}-1). \tag{7}$$

The benefit of this approach is that it produces a more informative upper bound, but at the cost of being less conservative.

Proposition 3 *Consider the environment described above, and suppose that the econometrician observes the auction price w_i, the total number of observed bidders $n_i + 1$ of every auction i, and the maximum number of potential bidders in each auction, $\bar{M}$.*

- *a) The identification region for $F_v(t)$ is given by equation* (5).
- *b) Assume further that the econometrician knows the joint distribution of potential and observed bidders:* $\Pr(M+1, N+1)$. *The bounds for $F_v(t)$ in equation* (6) *hold in expectation.*

The same techniques discussed above apply for the estimation of the bounds in Proposition 3, although the bounds in equation (6) require knowledge of $\Pr(M+1, N+1)$. Hickman et al. (2017) present non-parametric identification results and an estimation method for $\Pr(M+1, N+1)$ requiring data on the observed number of bidders only (i.e., knowledge of F_v is not required). Their model assumes that the number of potential bidders is unobserved by each bidder and is exogenous from the bidders' perspective. Bidders choose their bids before the auction starts and they submit their bids based on a predetermined order chosen by Nature. If the standing auction price exceeds a bidder's bid when it is their turn, then their bid is not recorded, which gives rise to the discrepancy between the number of potential bidders and the number of observed bidders. As long as equilibrium bidding is monotonic, their method can be implemented without knowledge of F_v (i.e., monotonicity allows the authors work with quantile ranks instead). See Hickman et al. (2017) for details.

4.2 More than one shill bidder

I next consider the case in which two shill bidders are active in an auction with $n+1$ bidders. I assume that the econometrician observes the third highest bid in the auction, $B_{3:n+1}$, as well as the auction price, W_{n+1}. Denote the distribution of $B_{3:n+1}$ in auctions with $n+1$ bidders by $F_{B3,n+1}$.

To derive the bounds of the identification region for $F_v(t)$, I establish two facts. The first one is that $V_{2:n-1} \leq W_{n+1} + \Delta$, which is similar to the observation used for Proposition 1, and can be proven using the same argument.

The second fact is that $B_{3:n+1} \leq V_{1:n-1}$. That this always holds, follows from the fact that there are only two shill bidders, implying that the highest bid by a legitimate bidder is at least $B_{3:n+1}$ (i.e., the third highest overall). By the assumption that bidders do not bid more than they are willing to pay, we know that $B_{3:n+1} \leq V_j$ for the legitimate player placing the highest bid. Since $V_j \leq V_{1:n-1}$ for all player j, $B_{3:n+1} \leq V_{1:n-1}$ always holds.

Combining these inequalities, we have established that $V_{2:n-1} - \Delta \leq W_{n+1}$ and $B_{3:n+1} \leq V_{1:n-1}$. The key difference with the case with only one shill bidder is that highest valuation among legitimate bidders cannot be bounded from below using the auction price, as it is always possible that the two shill bidders place the top two bids. These sets of inequalities combined allow the econometrician to bound the distribution of valuations, as indicated in the following proposition.

Proposition 4 *Consider the environment described above, and suppose that the econometrician observes the auction price w_i, the third highest bid $b_{3:n+1}$, and the total number of bidders $n_i + 1 \in \Omega$ of every auction i, where Ω is the set of unique values of $n + 1$ that are observed by the econometrician. Then, the identification region for $F_v(t)$ is given by*

$$H[F_v(t)] = \left[\max_{n+1 \in \Omega} \phi_2^{-1}(F_{w,n+1}(t - \Delta)|n-1), \min_{n+1 \in \Omega} \phi_3^{-1}(F_{B3,n+1}(t)|n-1) \right],$$

where $W_i \sim F_{w,n+1}(t)$, and $B_{3:n+1} \sim F_{B3,n+1}(t)$ are auction price and third-highest bid distributions when the total number of bidders is $n+1$, and $\phi_1(\cdot|n)$ and $\phi_3(\cdot|n)$ are the distribution functions of the first- and third-order statistics, defined as

$$\phi_k(s|n) = \frac{n!}{(n-k)!(k-1)!} \int_0^s x^{n-k}(1-x)^{k-1} dx.$$

Note that the same analysis can be conducted for more than two shill bidders, with the data requirements increasing with the number of shill bidders (i.e., the econometrician is required to observe more bids). Lastly, the same techniques discussed above apply for the estimation of the bounds in Proposition 4.

5 Identification of the optimal (fixed) reserve price

Consider a seller who wishes to sell an object in an auction with a fixed reserve price. That is, the seller does not wish to engage in shill bidding. The seller has access to auctions data and wishes to compute the optimal reserve price based on these data. Given concerns about shill bidding in the auctions in the sample, the seller uses the identification results discussed above.

What can be learned about the optimal reserve price? To answer this question, I assume that the seller can set the minimum bid increment to zero, $\Delta = 0$, and I make the following regularity assumption.

Assumption 1 *The distribution of valuations, F_v, is continuously differentiable and its support is a compact interval, $[\underline{v}, \bar{v}]$.*

Under these assumptions, the existence of an optimal reserve price is guaranteed for a number of bidders n. The optimal reserve price is given by the solution to the problem of maximizing expected revenue:

$$\max_{r \in [\underline{v}, \bar{v}]} \pi_n(r|v_0) = \max_{r \in [\underline{v}, \bar{v}]} v_0 F_v(r)^n + n \int_r^{\bar{v}} (F_v(v) + vF_v'(v) - 1) F_v^{n-1}(v) dv,$$

where v_0 is the seller's valuation for the object (assumed exogenous). I do not assume that F_v has the property that $x - (1 - F_v(x))/F_v'(x)$ is monotone increasing, which may lead to multiple solutions to the first-order condition of the problem above, with the optimal reserve price depending on the number of bidders (see Section 2). Although my identification analysis here resembles Haile and Tamer (2003) (Theorem 4), it differs in that I do not make assumptions about the shape of F_v (i.e., that $x - (1 - F_v(x))/F_v'(x)$ is monotone increasing).

I assume that the shill bidder's distribution of "exit points" is continuously differentiable. These assumptions together imply that the upper and lower bounds of $H[F_v(\cdot)]$ form continuously differentiable distribution functions.

Assumption 2 *The shill bidder's distribution of "exit points" is continuously differentiable.*

5.1 Identification

Define the following bounds for the seller's expected revenue when the seller values the object at v_0 and faces n bidders,

$$\pi_n^U(r|v_0) = v_0 L(r)^n + n \int_r^{\bar{v}} (L(v) + vL'(v) - 1) L^{n-1}(v) dv \tag{8}$$

and

$$\pi_n^L(r|v_0) = v_0 U(r)^n + n \int_r^{\bar{v}} (U(v) + vU'(v) - 1) U^{n-1}(v) dv, \tag{9}$$

where $L(\cdot)$ and $U(\cdot)$ are given by the lower and upper bounds, respectively, of $H[F_v(\cdot)]$. Here, I restrict to reserve prices that lie above of the seller's valuation for the object, and to distributions $G(\cdot) \in H[F_v(\cdot)]$ that are consistent with assumptions 1 and 2.

To see that $\pi_n^L(r|v_0)$ and $\pi_n^U(r|v_0)$ are in fact bounds for the seller's expected revenue, one can show that for $r \in [v_0, \bar{v}]$,

$$\pi_n(r|v_0, F) \geq \pi_n(r|v_0, G)$$

if $F(t) \leq G(t)$, $\forall t$. Since $L(t) \leq G(t) \leq U(t)$ for all $t \in [\underline{v}, \bar{v}]$ and for every distribution $G(\cdot) \in H[F_v(\cdot)]$ that is consistent with assumptions 1 and 2, the result follows.

The argument behind the identification approach can be illustrated using Figure 2. The dotted line in the figure is the constant function that takes the value given by

$$\sup_{r \in [\underline{v}, \bar{v}]} \pi_n^L(r).$$

Since π_n^L is a lower bound for the true expected revenue function, π_n, we know that the true optimal reserve price(s), r^*, must satisfy

$$\pi_n(r^*) \geq \sup_{r \in [\underline{v}, \bar{v}]} \pi_n^L(r).$$

At the same time, it must be that

$$\pi_n(r^*) \leq \pi_n^U(r^*),$$

since π_n^U is an upper bound for π_n. Note that the peak(s) of the function π_n, that give(s) us the optimal reserve price(s), can be achieved at any point r such that

$$\sup_{a \in [\underline{v}, \bar{v}]} \pi_n^L(a) \leq \pi_n^U(r).$$

This set of points defines the identification region for the optimal reserve price when the seller faces n bidders. When the seller faces uncertainty about the number of bidders that they will face in the auction to be run, this set can be computed for each plausible value of the number of bidders, and the identification region for the optimal reserve price will thus be the union of these sets.

Proposition 5 *Assume that Assumptions 1 and 2 hold. Given v_0, $L(\cdot)$ and $U(\cdot)$ (defined by the lower and upper bounds, respectively, of $H[F_v(\cdot)]$), $\pi_n^U(\cdot|v_0)$ and $\pi_n^L(\cdot|v_0)$ (defined in (8) and (9), respectively), the identification region of the optimal reserve price is given by*

$$H[r^*] = \bigcup_{n \in \aleph} \left\{ r \in [v_0, \bar{v}] : \pi_n^U(r) \geq \sup_{a \in [\underline{v}, \bar{v}]} \pi_n^L(a) \right\},$$

where $\aleph$ is the set of possible number of bidders to be faced by the seller (potentially a singleton).

5.2 Estimation

Note that one can rewrite $\pi_n^j(r|v_0)$, for $j \in \{L, U\}$, as

$$\begin{aligned}\pi_n^j(r|v_0) &= v_0 F_{-j}(r)^n + n\int_r^{\bar{v}} (F_{-j}(v) + vF'_{-j}(v) - 1)F_{-j}^{n-1}(v)dv \\ &= \bar{v} + (v_0 - r)F_{-j}(r)^n - \int_r^{\bar{v}} F_{-j}^{n-1}(v)[n(1 - F_{-j}(v)) + F_{-j}(v)]dv,\end{aligned}$$

where $F_{-U}(\cdot) = L(\cdot)$ and $F_{-L}(\cdot) = U(\cdot)$, i.e., the lower and upper bounds of the identification region for $F_v(t)$. This expression is convenient for estimation, as it saves the econometrician from estimating $F'_{-j}(\cdot)$.

Having estimates for the bounds of the identification region, $\hat{L}_T(\cdot)$ and $\hat{U}_T(\cdot)$, and given v_0, the econometrician can estimate $\pi_n^L(\cdot|v_0)$ and $\pi_n^U(\cdot|v_0)$ using:

$$\begin{aligned}\hat{\pi}_{T,n}^L(r|v_0) &= \bar{v} + (v_0 - r)\hat{U}_T(r)^n - \int_r^{\bar{v}} \hat{U}_T^{n-1}(v)[n(1 - \hat{U}_T(v)) + \hat{U}_T(v)]dv, \\ \hat{\pi}_{T,n}^U(r|v_0) &= \bar{v} + (v_0 - r)\hat{L}_T(r)^n - \int_r^{\bar{v}} \hat{L}_T^{n-1}(v)[n(1 - \hat{L}_T(v)) + \hat{L}_T(v)]dv,\end{aligned}$$

which I prove are uniformly consistent (i.e., $\hat{\pi}_{T,n}^j(r|v_0) \overset{a.s.}{\to} \pi_n^j(r|v_0)$ uniformly in r, for $j \in \{L, U\}$) in Lemma 3 in Appendix B.

In order to estimate the optimal reserve price, I define the following function,

$$Q_n(t) = \max\left\{0, \sup_a \pi_n^L(a) - \pi_n^U(t)\right\}, \tag{10}$$

which is defined for a given value of the number of bidders, n. By the previous discussion, it follows that the identification region for the optimal reserve price is given by

$$\Xi_n \equiv \arg\min_{t \in [\underline{v}, \bar{v}]} Q_n(t).$$

The sample analogue of $Q_n(t)$ can be defined as

$$\hat{Q}_{n,T}(t) = \max\left\{0, \sup_a \hat{\pi}_{T,n}^L(a) - \hat{\pi}_{T,n}^U(t)\right\}, \tag{11}$$

which in turn leads to the sample analogue of $\Xi_n(t)$, which can be defined as

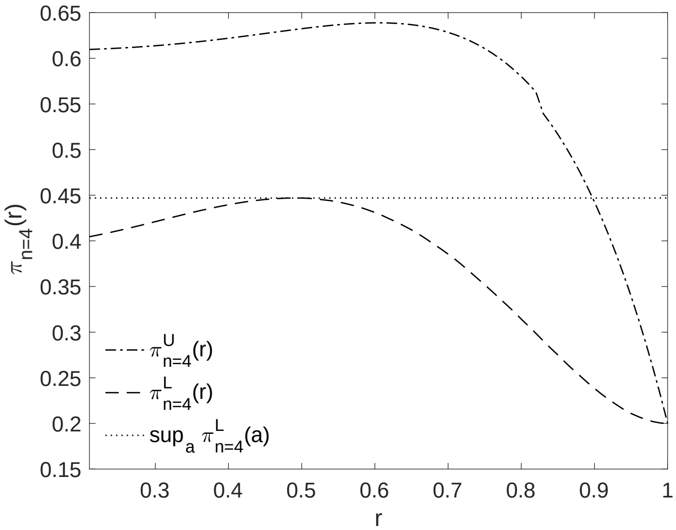
$$\hat{\Xi}_{n,T} \equiv \left\{t \in [\underline{v}, \bar{v}] : Q_{n,T}(t) \leq \inf_s Q_{n,T}(s) + \varepsilon_T\right\},$$

where $\varepsilon_T \to 0$ as $T \to \infty$.

In order to discuss consistency of the estimator $\Xi_{n,T}$, a notion of distance between two sets must be used. Consider two non-empty sets $A, B \subset \mathbb{R}^K$ and define $\rho(A, B) = \sup_{a \in A} \inf_{b \in B} |a - b|$. The Hausdorff distance between both sets is given by

$$d_H(A, B) = \max\{\rho(A, B), \rho(B, A)\}.$$

Figure 2: Estimated identification region for the optimal reserve price: $\hat{\Xi}_{n=4,T} = [0.2, 0.89]$.



Proposition 6 (Consistency) *Suppose the conditions in Proposition 2 hold. Let the set of possible reserve prices be* $[\underline{v}, \bar{v}]$*. Let* $Q_n(t)$ *and* $\hat{Q}_{n,T}(t)$ *be defined as in equations (10) and (11), respectively. Let* $T \to \infty$*, and* $\varepsilon_T \overset{a.s.}{\to} 0$*.*

a) Then $\rho(\hat{\Xi}_{n,T}, \Xi_n) \overset{a.s.}{\to} 0$*.*

b) Let $\sup_{t \in R} |\hat{Q}_{n,T}(t) - Q_n(t)|/\varepsilon_T \overset{a.s.}{\to} 0$*. Then* $\rho(\Xi_n, \hat{\Xi}_{n,T}) \overset{a.s.}{\to} 0$*.*

5.3 Monte Carlo simulations

Following the example in Section 3.2, I estimate the identification region for the optimal reserve price using the bounds for $F_v(t)$. I consider the case in which the number of bidders that the seller expects to face is exactly equal to four, $\aleph = \{4\}$, and that the seller's valuation for the object is $v_0 = 0.2$. Figure 2 depicts $\hat{\pi}^L_{n=4,T}(r)$ and $\hat{\pi}^U_{n=4,T}(r)$. Using the results in Proposition 5, the estimated identification region for the optimal reserve price in this example is given by $\hat{\Xi}_{n=4,T} = [0.2, 0.89]$.

5.4 Alternative methods for computing the optimal reserve price

A series of articles have investigated whether the optimal reserve price of an auction can be identified without knowing the distribution of valuations using data on past auctions (Cesa-Bianchi, Gentile and Mansour 2014, Mohri and Medina 2014, Austin, Seljan, Monello and

Tzeng 2016, Rudolph, Ellis and Blei 2016, Rhuggenaath, Akcay, Zhang and Kaymak 2019, Coey, Larsen, Sweeney and Waisman 2021).

Mohri and Medina (2014) and Coey et al. (2021) consider an ascending price auction in an independent values context, where bidders are assumed to bid their valuation (a weakly dominant strategy). They use the insight that the seller's expected payoff, as a function of reserve price r, can be written as a function of the two highest order statistics:

$$\pi(r) = E[v_0 \cdot 1\{V^1 < r\} + r \cdot 1\{V^2 < r < V^1\} + V^2 \cdot 1\{r < V^2\}],$$

where V^1 and V^2 are the first and second highest valuations among the bidders of the auction, the expectation is with respect to V^1 and V^2, and v_0 is the seller's valuation. If the econometrician observes V^1 and V^2 in their dataset, then $\pi(r)$ can be estimated using the sample analog of $\pi(r)$, and the optimal reserve price $r^* = \arg\max \hat{\pi}(r)$ can be computed.

In ascending price auctions where bidders enter at random times, a bidder may not submit a bid if they see that the standing price (i.e., the second highest bid among those who have submitted a bid) exceeds their valuation. Of course, the bidders with the first and second highest valuations will submit a bid, regardless of their entry time, provided that there is no shill bidder. This proves the usefulness of the method (again, provided that shill bidding is not present), as the econometrician can observe the top two bids on platforms such as eBay.

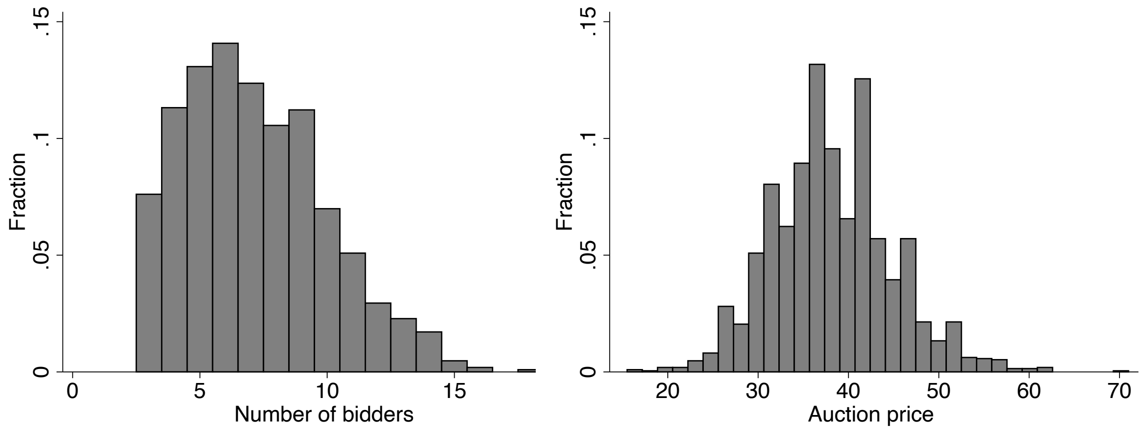
Can this method be used if shill bidding is present in the past auctions dataset? If a shill bidder places a bid that is higher than V^2, then the legitimate bidder with the second highest valuation will not submit a bid if they enter the auction after the shill bid has been submitted. That is, the econometrician will generally not have the necessary data to use this approach for computing the optimal reserve price. Adapting their results to a setting with a shill bidder is left for future research.

6 An application to eBay auctions

I apply my results on a sample of eBay auctions. The data include 2,103 auctions of sealed containers (i.e., mint condition) of Armani Acqua di Gio perfume (3.4 oz), which took place between the years 2008 and 2010. Given the broad claims about the problem of shill bidding in online auctions, this is a suitable setting for an empirical investigation.

The auction is an ascending price auction with a minimum bid increment $\Delta > 0$ that depends on the standing price (e.g., $\Delta = \$0.5$ when the price stands between \$5 and \$24.99, and $\Delta = \$1$ when it stands between \$25 and \$99.99). The auction price, W_{n+1}, satisfies $B_{2:n+1} \leq W_{n+1} \leq B_{2:n+1} + \Delta$, where $B_{2:n+1}$ is the second highest bid. The auctions in the

Figure 3: Summary Statistics: Armani Acqua di Gio (3.4 oz) eBay auctions



A. Observed bidders B. Auction price

Notes: An observation is an auction. The sample includes 2,103 auctions.

sample do not feature the buy-it-now or reserve-price options that are available to sellers on eBay. For more background information on eBay auctions, see, for example, Hasker and Sickles (2010b), Hickman et al. (2017), or Einav et al. (2018).

The independent private values assumption is also plausible in this context, as bidders can acquire the object from other retailers at a fixed (posted) price, but bidders are heterogeneous in how costly (or beneficial) it is for them to shop at a traditional retailer and it is plausible that these differences are independent across bidders.

With respect to the data, for each auction in the sample, I observe the number of observed bidders (i.e., those who placed a bid) and the auction price (i.e., the price paid by the winner of the auction). Figure 3A shows that the number of observed bidders ranges between 3 and 18 in these auctions, with an average of 7 bids per auction. Figure 3B shows the distribution of auction prices, with prices for the object roughly ranging between \$20 and \$60 and averaging \$38.13. In the analysis, I use raw bids because the object can be viewed as a commodity (mint condition, sealed container).

In the empirical analysis, I apply my results considering i) the case in which the number of potential bidders and the number of observed bidders are assumed equal and ii) the case in which the number of potential bidders is assumed unobserved.

As discussed in Section 3, when the number of potential bidders is assumed unobserved, I assume that this number ranges between 2 and $\bar{M}+1 = 100$ bidders. To implement some of my results, I estimate the joint distribution of potential and observed bidders $P(M+1, N+1)$, using the method proposed in Hickman et al. (2017). In their model, the marginal probability distribution of potential bidders, $M+1$, is given by a generalized Poisson with a probability distribution function given by $\Pr(m+1;\lambda) = \lambda_1(\lambda_1+(m+1)\lambda_2)^m \exp\{-(\lambda_1+(m+1)\lambda_2)\}/(m+1)!$, with $\lambda_1 > 0$ and $|\lambda_2| < 1$. The conditional distribution $\Pr(N+1|M+1)$ is simulated using the procedure outlined in Section 3 (see Hickman et al. (2017) for more details). The model parameters are estimated using a nonlinear least squares estimator that seeks to match the empirical distribution of observed bidders with the one predicted by the model. Using the code made available by Hickman et al. (2017), I estimate $\lambda_1 = 5.570$ and $\lambda_2 = 0.889$, with 95-percent bootstrapped confidence intervals given by $[5.220, 5.948]$ and $[0.880, 0.897]$, respectively.[11]

Bounds for the distribution of valuations Figure 4A displays the estimates for the identification region for the distribution of valuations using the results in Proposition 1. Here, I assume that the number of potential bidders equals the number of observed bidders. Given the sample size, I replace the min (max) functions in the estimator with the smooth weighted averages proposed by Haile and Tamer (2003), as discussed in Section 2. Specifically, I set $\rho_T = -\sqrt{\text{sample size}}$ and $\rho_T = \sqrt{\text{sample size}}$ for the lower and upper bounds, respectively. The figure also reports 95-percent bootstrapped confidence intervals for the bounds of the identification region. These confidence intervals are one-sided and were computed using 2,500 replicates.[12]

Figure 4B does the same for the case when the number of potential bidders is assumed unobserved (see Proposition 3). I present two sets of bounds in this figure. The first one, labeled $H[F_v(t)]$, are constructed using the (relatively mild) assumption that $M+1 \in \{2,\ldots,100\}$, as derived in equation (5). The second one, labeled "Bounds with $M_{90,n+1}$", assume that $M+1 \leq M_{90,n+1}$, where $M_{90,n+1}$ is a threshold defined as $\Pr(M+1 \leq M_{90,n+1}|n+1) = 0.9$, which relies on the estimates of $\Pr(M+1, N+1)$ (see equation (7)). The lower bounds are the same in both cases.

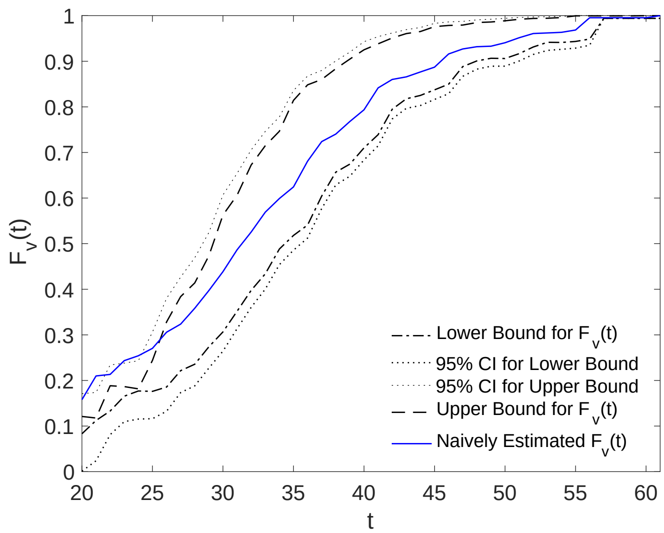
Comparing figures, one can see that the bounds are naturally tighter when the econometrician has more information (or they assume they have more information). As discussed in Section 3, because $\bar{M}+1 = 100$, the upper bound of $H[F_v(t)]$ in Figure 4B is relatively uninformative, and it gets only slightly better when using the assumption that $M+1 \leq M_{90,n+1}$.

Comparison with a method that rules out shill bidding I estimate the distribution of valuations assuming no shill bidding. In particular, I use the "button auction" model

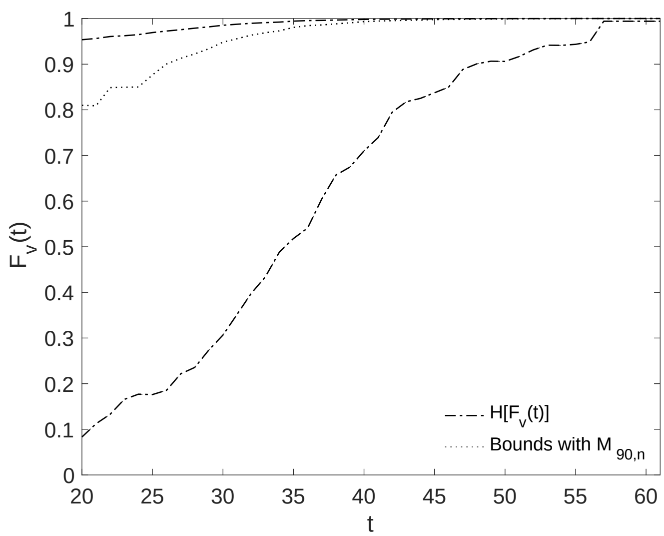
[11]The bootstrapped confidence intervals are based on 2,500 replicates.

[12]Haile and Tamer (2003) discuss consistency of bootstrapped confidence intervals in a similar setting.

Figure 4: Identification region for distribution of valuations of Armani Acqua di Gio perfume



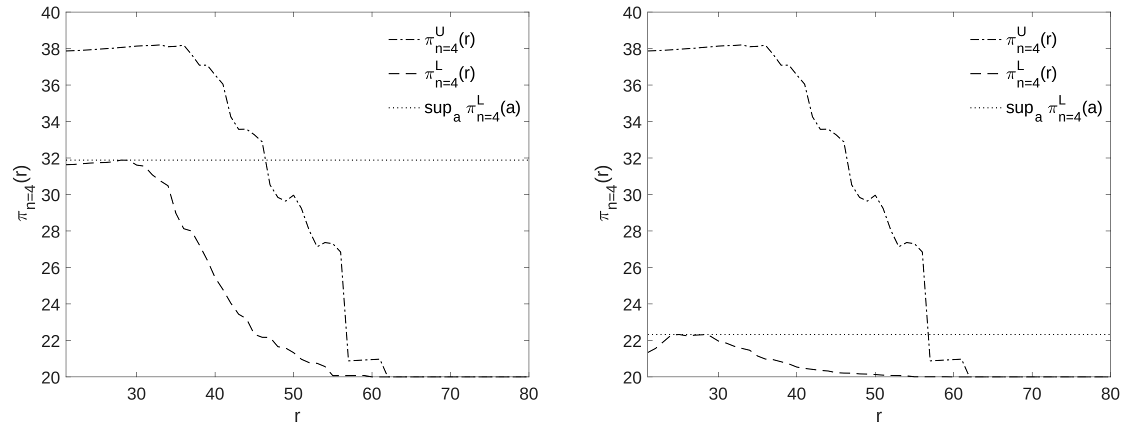
A. Identification region for the distribution of valuations when the number of bidders is assumed known (i.e., $M = N$)



B. Identification region for the distribution of valuations when the number of bidders is assumed unknown

Notes: These estimates are based on the results in Proposition 1 (Panel A) and Proposition 3 (Panel B). Confidence intervals in Panel A are one-sided and were computed using the bootstrap (2,500 replicates).

Figure 5: Identification region for the optimal reserve price



A. Bounds for $F_v(t)$ assuming $M = N$ (Figure 4A)

B. Bounds for $F_v(t)$ assuming $M = M_{90,n+1}$ (Figure 4B)

Notes: These estimates are based on the results in Proposition 5.

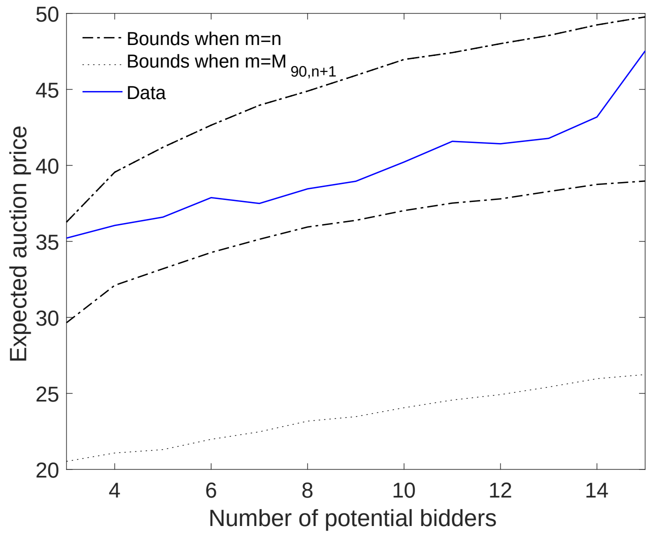
with $\Delta = 0$, independent private valuations, no shill bidding, and that the observed bidders equal the potential bidders. Under these assumptions, the auction price equals the second highest valuation among all bidders. I make use of the identification results in Athey and Haile (2002) to estimate F_v under these assumptions. Following Athey and Haile (2007), I estimate the distribution of valuations separately for each subsample of auctions with $n + 1$ bidders, and then compute an optimally weighted average of these estimators to minimize variance of the estimated distribution of valuations.

The estimate of F_v is displayed in Figure 4A under the label "Naively Estimated $F_v(t)$." As one can see in the figure, the estimated distribution of valuations lies within the bounds (or confidence interval of the bounds) at almost every point.

As discussed in Corollary 2, the bounds in Proposition 1 must hold for any complete model in which at least N of the $N + 1$ bidders draw their valuations independently from some distribution F_v. That is, the distribution of valuations that is estimated based on the assumptions of the complete model should lie within the identification bounds. If this fails to hold, there is evidence of model misspecification, which may for example be due to bidder asymmetries or correlation in valuations.

Bounds for the optimal reserve price Figure 5 plots the bounds for the seller's expected revenue when the seller values the good at $v_0 = \$20$ and expects an auction with 4 bidders.

Figure 6: Bounding the gains of adding an extra bidder



In panel A, the bounds are constructed based on the bounds of the distribution of valuations in Figure 4A (i.e., $M+1 = N+1$), whereas, in panel B the bounds are based on the estimates in Figure 4B when $M + 1 \leq M_{90,n+1}$. Using the results in Proposition 5, the identification regions for the optimal reserve price are $\hat{\Xi}^{M=N}_{n=4} = [20, 44]$ and $\hat{\Xi}^{M<M90}_{n=4} = [20, 57]$ in panels A and B, respectively. Since the distribution of valuations ranges between \$20 and \$60, these bounds only provide some information in the case of panel A.

Bounds for the auction price Figure 6 plots the bounds for the expected auction price as a function of the number of bidders based on i) the bounds of the distribution of valuations in Figure 4A (i.e., $M + 1 = N + 1$) and ii) the bounds in Figure 4B when $M + 1 \leq M_{90,n+1}$. The figure also plots the expected price in the sample of auctions (i.e., raw means assuming that the number of observed bidders equals the potential number of bidders). To compute the expected auction price, I make use of the assumptions about bidder behavior, and simulate 2,500 auctions for each $m + 1$, and average the bounds on the auction price across auctions. Using the bounds when $M = N$, the figure shows that the lower and upper bounds for the expected auction price increase by \$3 when moving from 3 to 4 bidders. Similarly, the upper bound in an auction with 3 bidders is less than the lower bound in an auction with 12 bidders.

7 Conclusion

This paper studies identification in an English auction with shill bidding in an independent private values setting. I show that the distribution of valuations and the optimal reserve price are partially identified when shill bids may be present in the data. Partial identification stems from the fact that the winning bid no longer equals the second highest valuation among the legitimate buyers, as the shill bidder can win the auction. I show that the winning bid will be bounded between the second and first highest valuations among the legitimate buyers when a shill bidder is present (up to minor corrections for the minimum bid increment). This observation can be used to bound the distribution of valuations and optimal reserve price. I apply these results on a sample of eBay auctions.

References

Anderson, Theodore W (1962) "On the distribution of the two-sample Cramer-von Mises criterion," *The Annals of Mathematical Statistics*, pp. 1148–1159.

Andreyanov, Pasha and El Hadi Caoui (2020) "Secret reserve prices by uninformed sellers,"Technical report, Working paper, University of Toronto.

Aradillas-López, Andrés, Amit Gandhi, and Daniel Quint (2013) "Identification and inference in ascending auctions with correlated private values," *Econometrica*, Vol. 81, No. 2, pp. 489–534.

Athey, Susan and Philip A. Haile (2002) "Identification of Standard Auction Models," *Econometrica*, Vol. 70, No. 6, pp. 2107–2140.

Athey, Susan and Philip A Haile (2007) "Nonparametric approaches to auctions," *Handbook of econometrics*, Vol. 6, pp. 3847–3965.

Austin, Daniel, Sam Seljan, Julius Monello, and Stephanie Tzeng (2016) "Reserve price optimization at scale," in *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 528–536, IEEE.

Baldwin, L, R Marshall, and J.F. Richard (1997) "Bidder collusion at forest service timber sales," *Journal of Political Economy*, Vol. 105, No. 4, pp. 657–699.

Brendstrup, Bjarne and Harry J Paarsch (2006) "Identification and estimation in sequential, asymmetric, english auctions," *Journal of Econometrics*, Vol. 134, No. 1, pp. 69–94.

Cesa-Bianchi, Nicolo, Claudio Gentile, and Yishay Mansour (2014) "Regret minimization for reserve prices in second-price auctions," *IEEE Transactions on Information Theory*, Vol. 61, No. 1, pp. 549–564.

Coey, Dominic, Bradley J Larsen, Kane Sweeney, and Caio Waisman (2021) "Scalable optimal online auctions," *Marketing Science*, Vol. 40, No. 4, pp. 593–618.

Coey, Dominic, Bradley Larsen, and Kane Sweeney (2019) "The bidder exclusion effect," *The RAND Journal of Economics*, Vol. 50, No. 1, pp. 93–120.

Einav, Liran, Chiara Farronato, Jonathan Levin, and Neel Sundaresan (2018) "Auctions versus posted prices in online markets," *Journal of Political Economy*, Vol. 126, No. 1, pp. 178–215.

Feinstein, J., M. Block, and F. Nold (1985) "Asymmetric information and collusive behavior in auction markets," *American Economic Review*, Vol. 75, No. 3, pp. 441–460.

Graham, Daniel A., Robert C. Marshall, and Jean-Francois Richard (1990) "Phantom Bidding Against Heterogeneous Bidders," *Economic Letters*, Vol. 32, pp. 13–17.

Grether, David, David Porter, and Matthew Shum (2015) "Cyber-shilling in automobile auctions: evidence from a field experiment," *American Economic Journal: Microeconomics*, Vol. 7, No. 3, pp. 85–103.

Haile, Philip A. and Elie Tamer (2003) "Inference with an Incomplete Model of English Auctions," *Journal of Political Economy*, Vol. 111, No. 1, pp. 1–51.

Hasker, Kevin and Robin Sickles (2010a) "eBay in the Economic Literature: Analysis of an Auction Marketplace," *Review of Industrial Organization*, Vol. 37, pp. 3–42.

——— (2010b) "eBay in the economic literature: Analysis of an auction marketplace," *Review of Industrial Organization*, Vol. 37, No. 1, pp. 3–42.

Hendricks, Ken and Robert H Porter (2007) "An empirical perspective on auctions," *Handbook of Industrial Organization*, Vol. 3, pp. 2073–2143.

Hickman, Brent R, Timothy P Hubbard, and Harry J Paarsch (2017) "Identification and estimation of a bidding model for electronic auctions," *Quantitative Economics*, Vol. 8, No. 2, pp. 505–551.

Kauffman, Robert J. and Charles A. Wood (2005) "The effects of shilling on final bid prices in online auctions," *Electronic Commerce Research and Applications*, Vol. 4, pp. 21–34.

Manski, Charles F. and Elie Tamer (2002) "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, Vol. 70, No. 2, pp. 519–546.

Marmer, Vadim, Artyom A Shneyerov, and Uma Kaplan (2016) "Identifying collusion in english auctions," *Available at SSRN 2738789*.

Milgrom, Paul R and Robert J Weber (1982) "A theory of auctions and competitive bidding," *Econometrica: Journal of the Econometric Society*, pp. 1089–1122.

Mohri, Mehryar and Andres Munoz Medina (2014) "Learning theory and algorithms for revenue optimization in second price auctions with reserve," in *International conference on machine learning*, pp. 262–270, PMLR.

OFT (2007) *Internet Shopping. An OFT market study.*: `http://www.oft.gov.uk/shared_oft/reports/consumer_protection/oft921.pdf`.

Paarsch, Harry J, Han Hong et al. (2006) *An introduction to the structural econometrics of auction data*: The MIT Press.

Porter, Robert and D. Zona (1993) "Detection of bid rigging in procurement auctions," *Journal of Political Economy*, Vol. 101, No. 3, pp. 518–538.

——— (1999) "Ohio school milk markets: An analysis of bidding," *RAND Journal of Economics*, Vol. 30, No. 2, pp. 263–288.

Rhuggenaath, Jason, Alp Akcay, Yingqian Zhang, and Uzay Kaymak (2019) "Optimizing reserve prices for publishers in online ad auctions," in *2019 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr)*, pp. 1–8, IEEE.

Riley, John G and William F Samuelson (1981) "Optimal auctions," *The American Economic Review*, Vol. 71, No. 3, pp. 381–392.

Rudolph, Maja R, Joseph G Ellis, and David M Blei (2016) "Objective variables for probabilistic revenue maximization in second-price auctions with reserve," in *Proceedings of the 25th International Conference on World Wide Web*, pp. 1113–1122.

Song, Unjy (2004) "Nonparametric estimation of an eBay auction model with an unknown number of bidders."

Tang, Xun (2011) "Bounds on revenue distributions in counterfactual auctions with reserve prices," *The RAND Journal of Economics*, Vol. 42, No. 1, pp. 175–203.

Wang, Wenli, Zoltan Hidvégi, and Andrew B. Whinston (2001) "Shill Bidding in English Auctions," *Working Paper*.

Appendix A: Omitted proofs

Proof of Proposition 2

We first have that by the Glivenko-Cantelli theorem,

$$\hat{F}_{w,n+1,T}(t) = \frac{1}{T_{n+1}} \sum_{i=1}^{T} 1\{m_i = n+1; w_i \leq t\} \overset{a.s.}{\to} F_{w,n+1}(t)$$

uniformly in t, for all $n+1 \in \Omega$.

Consider $\hat{L}_T(t)$. Since $\phi_2^{-1} : [0,1] \to [0,1]$ is a uniformly continuous function for all n, it follows from Lemma 2 that

$$\phi_2^{-1}(\hat{F}_{n+1,T}(t-\Delta)|n) \overset{a.s.}{\to} \phi_2^{-1}(F_{n+1}(t-\Delta)|n)$$

uniformly in t, for all $n+1 \in \Omega$. Since the max function is continuous, it follows from the continuous mapping theorem that

$$\hat{L}_T(t) \overset{a.s.}{\to} L(t),\ \forall t.$$

Finally, that the convergence of $L_T(t)$ to $L(t)$ is a.s. uniformly in t, follows from the following inequality

$$\sup_t |\hat{L}_T(t) - L(t)| \leq \sum_n \sup_t |\phi_2^{-1}(\hat{F}_{w,n+1,T}(t-\Delta)|n) - \phi_2^{-1}(F_{w,n+1}(t-\Delta)|n)|.$$

The rest of the proof follows by applying analogous arguments.

Proof of Proposition 3

The proof follows from the arguments provided in the text.

Proof of Proposition 4

The proof follows from arguments that are analogous to those in the proof of Proposition 1.

Proof of Proposition 5

Fix $n \in \aleph$. Define

$$\pi_n(r|v_0) = v_0 F_v(r)^n + n \int_r^{\bar{v}} (F_v(v) + vF_v'(v) - 1) F_v^{n-1}(v) dv,$$

where $F_v(\cdot)$ is the true but unobserved distribution of valuations, and take

$$r_n^* \in \arg\max_r \pi_n(r|v_0).$$

It is true that

$$\pi_n(r_n^*) \geq \sup_{a\in[\underline{v},\bar{v}]} \pi_n^L(a) \quad (12)$$

$$\pi_n^U(r_n^*) \geq \pi_n(r_n^*) \quad (13)$$

since $\pi_n^U(t) \geq \pi_n(t) \geq \pi_n^L(t), \forall t$.

Suppose $r_n^* \notin H[r^*]$. That implies, in particular, that $r_n^* \notin \left\{r : \pi_n^U(r) \geq \sup_{a\in[\underline{v},\bar{v}]} \pi_n^L(a)\right\}$. If $r_n^* \notin \{r : \pi_n^U(r) \geq \pi_n^L(r_n^L)\}$, then

$$\sup_{a\in[\underline{v},\bar{v}]} \pi_n^L(a) > \pi_n^U(r_n^*).$$

But then by making use of (12) and (13), we reach the following contradiction

$$\sup_{a\in[\underline{v},\bar{v}]} \pi_n^L(a) > \pi_n^U(r_n^*) \geq \sup_{a\in[\underline{v},\bar{v}]} \pi_n^L(a).$$

Proof of Proposition 6

Part a) follows from Lemma 4 in Appendix A. Part b) follows from Proposition 5b in Manski and Tamer (2002).

Appendix B: Additional results

Lemma 1 *Let $\phi_1(\cdot|n)$ and $\phi_2(\cdot|n)$ be the distribution functions of the first- and second-order statistics, defined as*

$$\phi_1(x|n) = x^n \quad \text{and} \quad \phi_2(x|n) = n(n-1)\int_0^x u^{n-2}(1-u)du.$$

The inverse functions $\phi_1^{-1}(x|n)$ and $\phi_1^{-2}(x|n)$ are increasing in n for $x \in (0,1)$.

Proof. I first show that $\phi_2^{-1}(x|n) \leq \phi_2^{-1}(x|n+1)$ for $x \in (0,1]$. Call the left-hand side expression, y_n, and the right-hand side expression, y_{n+1}. From the expression for the second-order distribution function, $\phi_2(\cdot|n)$, we note that y_n and y_{n+1} are implicitly defined as

$$\begin{aligned} x &= n(y_n^{n-1} - y_n^n) + y_n^n, \\ x &= ny_{n+1}(y_{n+1}^{n-1} - y_{n+1}^n) + y_{n+1}^n. \end{aligned}$$

By setting these expressions equal, and by using the fact that $x \in [0,1]$, we obtain the following inequality

$$\begin{aligned} n(y_n^{n-1} - y_n^n) + y_n^n &= ny_{n+1}(y_{n+1}^{n-1} - y_{n+1}^n) + y_{n+1}^n \\ &\leq n(y_{n+1}^{n-1} - y_{n+1}^n) + y_{n+1}^n, \end{aligned}$$

where the inequality follows from $y_{n+1} \in (0,1]$. The inequality can be rewritten as

$$\phi_2(y_n|n) \leq \phi_2(y_{n+1}(t)|n).$$

Since $\phi_2(\cdot|n)$ is a strictly increasing function, the result follows.

Consider next $\phi_1^{-1}(x|n)$. By taking the derivative of $\phi_1^{-1}(x|n) = x^{1/n}$, one can show that the function is increasing in n for $x \in (0,1)$. ■

Lemma 2 *Take a sequence of functions $\{g_T(\omega,\theta)\}$, $g_T : X \to Y$, that converges to $g(\theta)$ a.s. uniformly in $\theta \in \Theta$, that is,*

$$\Pr\left[\lim_{T\to\infty} \sup_{\theta\in\Theta} |g_n(\theta) - g(\theta)| = 0\right] = 1.$$

Take a uniformly continuous function $\psi : Y \to Y$. Then $\{\psi(g_T(\omega,\theta))\}$ converges to $\psi(g(\theta))$ a.s. uniformly in $\theta \in \Theta$.

Proof.

Fix any $\varepsilon > 0$. By uniform continuity of ψ, $\exists\delta > 0$ such that for any $x, y \in X$, $|x - y| < \delta$ implies $|\psi(x) - \psi(y)| < \varepsilon$.

By convergence a.s. uniformly of g_T,

$$\lim_{T\to\infty} \sup_{\theta\in\Theta} |g_T(\theta) - g(\theta)| = 0 \quad \text{a.e. },$$

that is, $\exists\, T_\delta$ such that $\forall\, m \geq T_\delta$

$$\sup_{\theta} |g_m(\theta) - g(\theta)| < \delta \quad \text{a.e. }.$$

By uniform continuity of ψ, we conclude that $\forall m \geq T_\delta$

$$\sup_{\theta} |\psi(g_m(\theta)) - \psi(g(\theta))| < \varepsilon \quad \text{a.e. }.$$

Since this holds for any $\varepsilon > 0$,

$$\lim_{T\to\infty} \sup_{\theta\in\Theta} |g_T(\theta) - g(\theta)| = 0 \quad \text{a.e.} \quad \Rightarrow \lim_{T\to\infty} \sup_{\theta\in\Theta} |\psi(g_T(\theta)) - \psi(g(\theta))| = 0 \quad \text{a.e. }.$$

The result follows since

$$1 = \Pr\left[\lim_{T\to\infty} \sup_{\theta\in\Theta} |g_T(\theta) - g(\theta)| = 0\right] \leq \Pr\left[\lim_{T\to\infty} \sup_{\theta\in\Theta} |\psi(g_T(\theta)) - \psi(g(\theta))| = 0\right].$$

■

Lemma 3 $\pi_{T,n}(r) \stackrel{a.s.}{\to} \pi_n(r)$ *uniformly in* r.

Proof.

Note that

$$\begin{aligned}
\sup_r |\pi_{T,n}(r) - \pi_n(r)| &= \sup_r \Big|(v_0 - r)(F_T(r)^n - F(r)^n) \\
&\quad - \int_r^{\bar{v}} \left(F_T(v)^{n-1}[n(1 - F_T(v)) + F_T(v)] - F(v)^{n-1}[n(1 - F(v)) + F(v)]\right) dv\Big| \\
&\leq |K_1| \cdot \sup_r |F_T(r)^n - F(r)^n| \\
&\quad + \sup_r \left| \int_r^{\bar{v}} \left(F_T(v)^{n-1}[n(1 - F_T(v)) + F_T(v)] - F(v)^{n-1}[n(1 - F(v)) + F(v)]\right) dv \right| \\
&\leq |K_1| \cdot \sup_r |F_T(r)^n - F(r)^n| \\
&\quad + |K_2| \cdot \sup_v \left| F_T(v)^{n-1}[n(1 - F_T(v)) + F_T(v)] - F(v)^{n-1}[n(1 - F(v)) + F(v)] \right| \\
&= |K_1| \cdot \sup_r |\psi_1(F_T(r)) - \psi_1(F(r))| + |K_2| \cdot \sup_v |\psi_2(F_T(r)) - \psi_2(F(r))|,
\end{aligned}$$

where K_1 and K_2 are constants, and $\psi_1 : [0,1] \to [0,1]$ and $\psi_2 : [0,1] \to [0,1]$ are uniformly continuous functions. Since $F_T(x) \stackrel{a.s.}{\to} F(x)$ uniformly in x, the result follows from Lemma 2.

■

Lemma 4 $Q_T(t) \stackrel{a.s.}{\to} Q(t)$ *uniformly in* t.

Proof.

Note that

$$\begin{aligned}
\sup_t |Q_{n,T}(t) - Q_n(t)| &= \sup_t \Big| 1\{\sup_a \pi^L_{T,n}(a) - \pi^U_{T,n}(t) > 0\}(\sup_a \pi^L_{T,n}(a) - \pi^U_{T,n}(t)) \\
&\quad - 1\{\sup_a \pi^L_n(a) - \pi^U_n(t) > 0\}(\sup_a \pi^L_n(a) - \pi^U_n(t)) \Big| \\
&\leq \sup_t \left| (\sup_a \pi^L_{T,n}(a) - \pi^U_{T,n}(t)) - (\sup_a \pi^L_n(a) - \pi^U_n(t)) \right| \\
&\leq \left| \sup_a \pi^L_{T,n}(a) - \sup_a \pi^L_n(a) \right| + \sup_t \left| \pi^U_n(t) - \pi^U_{T,n}(t) \right| \\
&\leq \sup_a \left| \pi^L_{T,n}(a) - \pi^L_n(a) \right| + \sup_t \left| \pi^U_n(t) - \pi^U_{T,n}(t) \right|.
\end{aligned}$$

Since $\pi_{T,n}(r) \stackrel{a.s.}{\to} \pi_n(r)$ uniformly in r, the result follows from Lemma 2. ■