

Nonparametric Methods for Ascending Auctions with Unobserved Heterogeneity

Amit Gandhi and Daniel Quint

April 19, 2009

Preliminary and Incomplete

1 Introduction

The combined assumptions that the number of bidders participating in an auction is independent of the characteristics of the auction itself, and that the bidders have symmetric and independently distributed private values, play a critical role in the classical theory of auctions. For example, a simple but powerful policy prescription that arises from the classical theory is that under these assumptions, the optimal selling mechanism for a good is achieved by running any of the standard auction formats with a suitably chosen reserve price.

Starting with Paarsch (1992), empirical researchers have sought to operationalize classical auction theory by using data on bids from a cross section of auctions to identify and estimate the demand curve facing the seller. Using the demand curve, the seller can then simulate the effects on revenue of different auction design parameters and calculate the reserve price that maximizes revenue.¹

¹An early example of this policy exercise is Paarsch (1997).

However, both the execution of this empirical exercise, and the theoretical results above, rely critically on the classical assumptions that bidders enter auctions exogenously (nonstrategically) and receive identical and independently distributed private values. Even if we maintain the hypothesis of private values (and thus ignore the question of common versus private values), introducing any small degree of correlation in valuations among bidders (Cremer and McLean 1988, McAfee and Reny 1992), or endogenizing bidders' participation decisions (Samuelson 1985, McAfee and McMillan 1987, Levin and Smith 1994), significantly alters the design of the optimal mechanism. Thus for the purposes of mechanism design, an even more basic problem facing the seller than identification and estimation of demand is whether the classical assumptions are an appropriate model of demand in the first place.

In this paper we nonparametrically address the empirical problem of mechanism design in the presence of potentially endogenous entry and/or correlated valuations among bidders. Both endogenous entry and correlated values tend to imply an optimal reserve price below that of the classical baseline.² Thus we take the empirical problem facing the seller to be twofold: (1) Is entry endogenous? (2) If entry is exogenous but values are potentially correlated, can the optimal reserve price be identified from the data?

As we show, these two questions, while distinct, are intimately related to one another. They are linked by a third issue facing the seller: the existence of unobserved auction heterogeneity. The existence of such heterogeneity can be thought of (almost without loss) as the sole mechanism that induces correlation among bidders' valuations. The economic interpretation of this

²Under the well known Levin and Smith (1994) model of endogenous entry with independent values, the optimal reserve price is equal to the seller's value, lower than with exogenous entry; for ascending auctions (as we consider in this paper), this holds for correlated private values as well. With exogenous entry but correlated values, the issue facing the auctioneer is not as clear cut because the correlated values model is not empirically identifiable from standard auction data (Athey and Haile (2002)); Quint (2008) shows that when values are *affiliated*, the optimal reserve price is always weakly lower than the independent-values benchmark.

heterogeneity is that it represents an unobservable shock to the demand curve facing the seller. As we show, the essential question underlying the issue of endogenous entry is whether entry decisions are independent of this demand shock. Our first main result shows that auction data can be used to nonparametrically test for the presence of endogenous entry.

Even if entry is exogenous,³ the presence of an unobservable demand shock means the seller faces a potentially correlated demand structure within an auction. When past data is from ascending auctions, this correlated structure of demand cannot be identified from bid data under standard assumptions (Athey and Haile 2002). Nevertheless, we show that knowledge of the full structure of demand is not necessary for the main policy problem of computing the revenue-maximizing reserve price. Rather, only the *marginal* distributions of the *top two* valuations among bidders are needed. We further show that data from ascending auctions does indeed allow for the identification of these two marginal distributions, and thus identification of the optimal reserve price, given exogenous variation in the number of bidders in each auction.

We focus our attention on ascending price, or English, auctions. English auctions are the most empirically prevalent auction format. In addition, bidding behavior at English auctions has a dominance solvability property, which we capture through the individual rationality constraints of Haile and Tamer (2003). Thus following the logic of Haile and Tamer (2003), we not only avoid imposing a full game theoretic structure on the auction (and hence, properly speaking, we work with an incomplete model of English auctions), but we avoid having to impose any assumptions about tastes for risk or strategic sophistication among bidders. Thus the main empirical content of the (incomplete) English auction model is driven by the demand structure.

³As we will further explain, exogenous entry can be driven by the fact that bidders do not enter strategically, or bidders enter strategically but do not have any more information than the seller at the time of the entry decision.

2 Model

2.1 Demand

Consider a single item auction. Since we focus on English auctions throughout, the only mechanism characteristic that may vary across auctions is the reserve price r . For now we exclude any other observable (to the seller) auction-specific covariates – we discuss later how to incorporate such covariates into the empirical analysis.

We will assume that bidders have private values which are *conditionally independent*: there is an unobserved (to the seller) random variable θ , drawn from some probability distribution over some arbitrary space Θ ; and conditional on the realization of θ , bidder values (V_1, \dots, V_n) are independent and identically distributed random variables drawn from a distribution $H(\cdot|\theta)$. Note that we make no assumptions about the dimensionality of Θ – the heterogeneity may be arbitrarily complicated, even infinite dimensional – we only require that it exist. (De Finetti’s theorem (as formalized by Hewitt and Savage (1955)) says that under exchangeability of potential bidder valuations, this is almost without loss of generality.⁴) Note that the stronger assumption of independence of the valuations (V_1, \dots, V_n) would be equivalent to assuming away any variation in the demand shock θ .

2.2 Entry

The number of entrants n at an auction is a random variable whose distribution potentially depends on the auction characteristics (θ, r) . If the number of entrants n is determined independently of (θ, r) , then we will say that

⁴Formally, consider an infinite sequence of random variables $(V_i)_{i=1}^{\infty}$, representing the private valuations of an infinite set of potential bidders. If these potential bidders are ex-ante symmetric (the V_i are exchangeable random variables), then there exists a random variable $\theta \in \Theta$ and a probability distribution over Θ such that any finite sample $(V_{i_1}, \dots, V_{i_n})$ of the potential bidders, V_{i_j} are conditionally independent draws from $H(\cdot|\theta)$.

entry is exogenous. Otherwise, we say entry is endogenous.⁵ The question facing the seller is how to test whether entry is exogenous or endogenous.⁶

For now, we simplify the exposition by “shutting down” the reserve price and assuming that it is 0, or some low level that never binds, in the data. This has the advantage that the number of entrants equals the number of actual bidders, and we do not need to carry around this distinction.⁷ Thus we shall henceforth speak of the number of bidders as synonymous with the number of entrants. The remaining source of information on endogenous entry, then, is to ask whether the demand shock θ is correlated with the number of bidders n . By definition, θ is unobservable to the seller, and thus a direct test of independence is impossible. However, information about θ is transmitted through the bids submitted at the auction; we will use this, along with variation in the number of bidders n across auctions, to develop a test for exogenous versus endogenous entry. We first, however, consider the

⁵The demand shock θ will trivially be independent of n if the bidders do not have any information on θ when making the entry decision, i.e., they face the exact same information set as the seller at the time of making the entry decision.

⁶One model of endogenous entry is that proposed by Levin and Smith (1994). In their model, bidders are assumed to not know their private values at the time of entry, so the participation cost is viewed as the time and resources invested to evaluate the object for sale and learn one’s valuation for it. Ex-ante, before a potential bidder becomes an entrant, the expected surplus to entering is a function of r and the number of entrants n , and the mixed strategy equilibrium determines a distribution of n .

⁷A seemingly straightforward test of endogeneity would be to check whether the reserve price r has an effect on the number of entrants. As raising r decreases a bidder’s ex-ante expected surplus, if entry is endogenous, then higher r should be empirically associated with a smaller number of entrants. The problem with this potential test is that sellers typically only observe the m bidders who actually bid, but not the additional $n-m$ entrants who choose to learn their valuations but do not place bids because the reserve price exceeds their valuations. Thus, even under exogenous entry, this censoring will create a spurious negative correlation between the reserve price and the observed number of bidders.

Our decision to “shut down” the reserve price can be understood in two ways. From the perspective of a seller who is attempting to test for endogenous entry using auction data, setting $r = 0$ and observing bidding behavior is an experiment that can be operationalized. However even if the seller is not intentionally behaving this way, the assumption also matches many data sources, from auctions with low or no reserve, and avoids the problem of whether reserve price variation can be treated as exogenous when the seller was not deliberately running an experiment.

nature of bidding behavior in the auctions we consider.

2.3 Bidding

We focus attention on selling mechanisms μ that lie within the class \mathcal{M} of English auctions. The essential feature of an English auction is its open and ascending price property. That is, the price of the good is bid up in a way that is public to all bidders. As there are a variety of conceivable mechanisms μ that satisfy these requirements, we follow in the spirit of Haile and Tamer (2003) and do not commit to a particular mechanism μ within \mathcal{M} , but rather work with an incomplete model of an English auction. Of course, without conditioning on a particular $\mu \in \mathcal{M}$, one cannot fully map the relationship between an agent’s latent valuation V_i and the bid B_i . Nevertheless there exist robust properties of equilibrium bidding that can reasonably be thought to hold true for all $\mu \in \mathcal{M}$. Two such assumptions are those proposed by Haile and Tamer (2003), who put them forth as essentially the defining assumptions of equilibrium bidding in English auctions:

Assumption 1. *For any $\mu \in \mathcal{M}$, bidders do not bid more than they are willing to pay, and bidders do not allow an opponent to win at a price they are willing to beat.*

Our test will require no stronger assumptions about bidding behavior than these. However, in order to better explain the underlying economics of the test, it will be useful to impose a stronger assumption on bidding behavior than Assumption 1. To motivate the assumption, consider the common alternative approach of modeling ascending auctions as “button auctions” (in which losing bidders bid up to their willingness to pay). In terms of imputing only the second-highest bidder’s willingness to pay, the main generality introduced by the weaker assumptions of Haile and Tamer (2003) is that they allow for jump bidding and/or a discrete bid increment.

Hence a bidder’s bid is not necessarily exactly reflective of the bidder’s value (it will generally be an understatement of value).

However, even in an open outcry auction, which offers the greatest flexibility for jump bidding, when the auction comes down to only two bidders actively competing with each other for the sale of the good, both bidders face a risk of increasing the standing bid by too large an amount. So long as the bid increments towards the end of the auction are small relative to the size of the overall bid, then the bidding behavior towards the end of the auction is well approximated by a button auction. If this approximation were exactly true, it would imply the following strengthening of Assumption 1:⁸

Assumption 2. *For any $\mu \in \mathcal{M}$, the bidder with the second highest value bids up to his value.*

The approximate validity of Assumption 2 can be tested directly in the data by checking that the difference between the top two bids in an open outcry auction is small relative to the size of the bids. While Assumption 2 is almost surely not exactly true in any English auction unless the auction is being literally run as an button auction, it nevertheless provides a useful starting point for us. We will derive our test under Assumption 2, and then show that it generalizes easily to the weaker bidding assumptions of Haile and Tamer (2003) (our Assumption 1).

3 Testing for Endogenous Entry

3.1 Notation

Let $V_{k:n}$ denote the k^{th} lowest valuation in an auction with $n \geq k$ bidders (so $V_{n:n}$ is the highest), and let $F_{k:n}$ denote its distribution function. Let

⁸As Athey and Haile (2002) write, “A plausible alternative hypothesis for many ascending auctions is that bids $B^{n-2:n}$ and below do not always reflect the full willingness to pay of losing bidders, although $B^{n-1:n}$ does (since only two bidders are active when that bid is placed).”

$B_{k:n}$ denote the k^{th} lowest bid in an auction with $n \geq k$ bidders, and $G_{k:n}$ its distribution function.

For each $k \leq n$, the distribution $G_{k:n}$ is identified from data on bids across auctions of size n . Assumption 2 is the assumption that $F_{n-1:n} = G_{n-1:n}$, and so under Assumption 2, $F_{n-1:n}$ is identified for each n . Thus, we will develop our test under that assumption that $F_{n-1:n}$ is known exactly for every n .

As discussed in Haile and Tamer (2003), Assumption 1 implies that $V_{n-1:n} \geq B_{n-1:n}$ and $V_{n-1:n} \leq B_{n:n} + \delta$, where δ is the minimum bid increment. In turn, this implies that for any z , $G_{n:n}(z - \delta) \leq F_{n-1:n}(z) \leq G_{n-1:n}(z)$. Assumption 1 therefore leads to upper and lower bounds on $F_{n-1:n}$ for each n ; after developing the test under Assumption 2, we will show how it extends naturally to Assumption 1.

3.2 Benchmark Case: No Heterogeneity

First, we review the seminal contribution of Athey and Haile (2002), who derive a nonparametric testable implication of English auctions under Assumption 2. The key to their test is that in addition to assuming potential bidders are symmetric (as we do), they also assume that potential bidders' valuations are independent. Under these assumptions, if variation in n is exogenous, the distribution $F_{n-1:n}$ varies in a predictable way with n , leading to a nonparametric test of this model, which we will refer to as (symmetric independent private values + exogenous entry), or (SIPV + exogenous entry).

For the remainder of this section, assume that bidders' valuations V_i are both exchangeable and mutually independent random variables drawn from a parent distribution H . Observe that for any $k \leq n$,

$$\begin{aligned} F_{k:n}(z) &= Pr(\text{at least } k \text{ bidders have a value } \leq z) \\ &= \sum_{j=k}^n \binom{n}{j} H(z)^j (1 - H(z))^{n-j} \end{aligned} \tag{1}$$

For $n \geq 2$ and $k \in \{1, 2, \dots, n\}$, define functions $\psi_{k:n} : [0, 1] \rightarrow [0, 1]$ by

$$\psi_{k:n}(s) = \sum_{j=k}^n \binom{n}{j} s^j (1-s)^{n-j} \quad (2)$$

so that $F_{k:n}(z) = \psi_{k:n}(H(z))$. Though it is not clear from this definition, the functions $\psi_{k:n}$ are strictly increasing and onto, and therefore invertible, so that

$$\psi_{k:n}^{-1}(F_{k:n}(z)) = H(z) \quad (3)$$

As first recognized by Athey and Haile (2002), this implies a nonparametric testable implication of the (SIPV + exogenous entry) model: specifically, for any $n \neq n'$ and any z ,

$$\psi_{n-1:n}^{-1}(F_{n-1:n}(z)) = \psi_{n'-1:n'}^{-1}(F_{n'-1:n'}(z)) \quad (4)$$

It is useful to consider more carefully the underlying economics of the nonparametric test in (4). As n increases, the distribution of $V_{n-1:n}$ shifts upwards (in a first-order stochastic dominance sense), what we will refer to as the *competition effect* – as n grows, $V_{n-1:n}$ is the second-highest of a larger sample, and is therefore stochastically higher. The essence of the test is that independence of valuations exactly determines the size of the competition effect, via (4): for $n > n'$, $F_{n-1:n}(z) < F_{n'-1:n'}(z)$, but

$$\psi_{n'-1:n'}^{-1} \circ \psi_{n-1:n}^{-1}(F_{n-1:n}(z)) = F_{n'-1:n'}(z). \quad (5)$$

(As discussed in Athey and Haile (2002), (3) suggests that when $F_{n-1:n}$ is learned exactly, even without variation in n , the symmetric IPV model is exactly identified. That (4) must hold for every pair (n', n) suggests that the SIPV model with exogenous entry is overidentified. When we remove the assumption of independence, the model becomes underidentified for fixed n ; as we show below, when we add exogenous variation in n , it becomes exactly

identified.)

3.3 Our Case: Unobserved Heterogeneity

We now drop the assumption of independence, which led to the test (4). As explained in Section 2.1, exchangeability without independence implies that there exists an unobserved demand shock $\theta \in \Theta$ from the perspective of the seller that shifts the underlying distribution of valuations $H(\cdot | \theta)$. We replace the assumption of independence with the assumption of conditional independence, and refer to the new set of joint assumptions as (symmetric conditionally independent private values + exogenous entry), or (SCIPV + exogenous entry).

Under the new model (SCIPV + exogenous entry), the test (4) is not valid. To see why, consider the degenerate case where $H(z | \theta)$ puts all its probability mass on the point θ , so that $H(z | \theta) = \mathbf{1}(\theta \leq z)$. Then for any auction with $n \geq 2$ bidders, $V_{n-1:n} = \theta$, so $F_{n-1:n}(z) = E_\theta[\mathbf{1}(\theta \leq z)]$. The competition effect has vanished – since the unconditional joint distribution $F(V_1, \dots, V_n)$ exhibits perfect correlation, adding another bidder to the sample has no effect on the distribution of the winning price. In this case, the test in Athey and Haile (2002) will over-predict the magnitude of the competition effect; the true model in this case satisfies the inequality

$$\psi_{n'-1:n'} \circ \psi_{n-1:n}^{-1}(F_{n-1:n}(z)) \geq F_{n'-1:n'}(z), \quad (6)$$

for $n > n'$. (While still weakly lower, $F_{n-1:n}(z)$ is closer to $F_{n'-1:n'}(z)$ than before, so using the same “correction function” overshoots.)

We now show that this effect holds generally: namely, that (SCIPV + exogenous entry) always “slows down” the size of the competition effect relative to the (SIPV + exogenous entry) benchmark. The key to this result is the following:

Lemma 3.1. *For any n , $\psi_{n-1:n} \circ \psi_{n:n+1}^{-1} : [0, 1] \rightarrow [0, 1]$ is concave.*

Proof. For any two functions f and g ,

$$\frac{d}{ds} (f \circ g^{-1})(s) = f'(g^{-1}(s)) \cdot (g^{-1})'(s) = \frac{f'(g^{-1}(s))}{g'(g^{-1}(s))} \quad (7)$$

$\psi_{n-1:n}(t) = nt^{n-1} - (n-1)t^n$ implies $\psi'_{n-1:n}(t) = n(n-1)t^{n-2}(1-t)$, and so

$$\frac{d}{ds} (\psi_{n-1:n} \circ \psi_{n:n+1}^{-1})(s) = \frac{\psi'_{n-1:n}(t)}{\psi'_{n:n+1}(t)} = \frac{n(n-1)t^{n-2}(1-t)}{(n+1)nt^{n-1}(1-t)} = \frac{n-1}{n+1} \cdot \frac{1}{t} \quad (8)$$

where $t = \psi_{n:n+1}^{-1}(s)$. By inspection, this is decreasing in t ; since $\psi_{n:n+1}$ is increasing, t is increasing in s , and so $\frac{d}{ds} (\psi_{n-1:n} \circ \psi_{n:n+1}^{-1})(s)$ is decreasing in s , making $\psi_{n-1:n} \circ \psi_{n:n+1}^{-1}$ concave. \square

Theorem 3.2. *Under SCIPV and exogenous entry, for any s , $\psi_{n-1:n}^{-1}(F_{n-1:n}(s))$ is weakly increasing in n .*

Proof. It suffices to show that $\psi_{n:n+1}^{-1}(F_{n:n+1}(s)) \geq \psi_{n-1:n}^{-1}(F_{n-1:n}(s))$, or (since $\psi_{n-1:n}$ is strictly increasing) that $(\psi_{n-1:n} \circ \psi_{n:n+1}^{-1})(F_{n:n+1}(s)) \geq F_{n-1:n}(s)$. At a given realization of $\theta \in \Theta$, the distribution of $V_{n-1:n}$ is $\psi_{n-1:n}(H(\cdot|\theta))$; so the unconditional distribution of $V_{n-1:n}$ is

$$F_{n-1:n}(s) = E_{\theta} \{ \psi_{n-1:n}(H(s|\theta)) \} \quad (9)$$

Then

$$\begin{aligned} (\psi_{n-1:n} \circ \psi_{n:n+1}^{-1})(F_{n:n+1}(s)) &= (\psi_{n-1:n} \circ \psi_{n:n+1}^{-1}) E_{\theta} \{ \psi_{n:n+1}(H(s|\theta)) \} \\ &\geq E_{\theta} \{ (\psi_{n-1:n} \circ \psi_{n:n+1}^{-1})(\psi_{n:n+1}(H(s|\theta))) \} \\ &= E_{\theta} \{ \psi_{n-1:n}((\psi_{n:n+1}^{-1} \circ \psi_{n:n+1})(H(s|\theta))) \} \\ &= E_{\theta} \{ \psi_{n-1:n}(H(s|\theta)) \} \\ &= F_{n-1:n}(s) \end{aligned}$$

where the second line is Jensen's Inequality since $\psi_{n-1:n} \circ \psi_{n:n+1}^{-1}$ is concave. \square

Since under Assumption 2, $F_{n-1:n}$ is identified from the data, this gives us a fully nonparametric test of the assumptions (SCIPV + exogenous entry), which is a one-sided version of the test in (4):

Corollary 3.3. *For any $n > n'$ and $z \in \mathbb{R}$,*

$$\psi_{n'-1:n'} \circ \psi_{n-1:n}^{-1}(F_{n-1:n}(z)) \geq F_{n'-1:n'}(z). \quad (10)$$

We believe the test in (10) to be precisely a test of exogenous versus endogenous entry. As we have shown in Corollary 3.3, the (SCIPV + exogenous entry) model must “slow down” the competition effect relative to the prediction of the (SIPV + exogenous entry) model. Hence if we see that the size of the competition effect in the data is *larger* than that predicted under (SIPV + exogenous entry), this can only be due to endogenous entry: values of θ that stochastically increase demand $F(\cdot | \theta)$ induce more entry n . Thus a violation of the inequality (10) provides evidence of correlation between θ and n , despite the fact that the demand shock θ is unobservable to the seller (and econometrician). If the competition effect in the data is *smaller* than that predicted under (SIPV + exogenous entry), we will not be able to distinguish whether this is due to unobserved heterogeneity or endogenous entry; but the type of entry required to generate these observations – bidders selectively entering auctions in greater numbers when the object is likely to be worth less – seems empirically unlikely. Thus, when (10) holds, we feel this provides some support for the assumption of exogenous entry.

3.4 Testing under Weaker Bidding Assumptions

We have predicated our test on Assumption 2. We now show that the test easily generalizes to Assumption 1. Let δ be the minimum bid increment in the Haile and Tamer model, i.e., the smallest amount by which a bidder can raise the standing bid at an English auction. As we mentioned above, Assumption 1 can be shown to imply that $B_{n-1:n} \leq V_{n-1:n}$ and $V_{n-1:n} \leq$

$B_{n:n} + \delta$ (see Haile and Tamer 2003), giving rise to the following more general test:

Corollary 3.4. *Under (SCIPV + exogenous entry) and Assumption 1, for any $n > n'$,*

$$\psi_{n-1:n}^{-1}(G_{n-1:n}(s)) \geq \psi_{n'-1:n'}^{-1}(G_{n':n'}(s - \delta)) \quad (11)$$

Proof. $B_{n-1:n} \leq V_{n-1:n}$ implies $G_{n-1:n}(s) \geq F_{n-1:n}(s)$, and therefore (since $\psi_{n-1:n}$ is strictly increasing) $\psi_{n-1:n}^{-1}(G_{n-1:n}(s)) \geq \psi_{n-1:n}^{-1}(F_{n-1:n}(s))$. Similarly, $V_{n'-1:n'}(s) \leq B_{n':n'} + \delta$ implies $F_{n'-1:n'}(s) \geq G_{n':n'}(s - \delta)$ and therefore $\psi_{n'-1:n'}^{-1}(F_{n'-1:n'}(s)) \geq \psi_{n'-1:n'}^{-1}(G_{n':n'}(s - \delta))$. Along with Corollary 3.3, these give the result. \square

3.5 Testing with Observable Heterogeneity

We have thus far excluded from the model any auction heterogeneity that is observable to the seller. If there existed such observable heterogeneity, we could simply condition on the observed covariates and repeat the entire analysis we have pursued to this point. While such conditioning does not impact the nonparametric validity of the tests in (4), (10), or (11), it does impact the empirical feasibility of carrying out the tests on finite data. This is due to the well known curse of dimensionality problem with nonparametric estimators (see e.g., Silverman (1986)). However, there exists a simple solution to the curse of dimensionality in the context of our tests (10) and (11) of (SCIPV + exogenous entry) that is not available to the test (4) of (SIPV + exogenous entry).

The curse of dimensionality arises because the observed covariates $x \in \mathbb{R}^K$ may be high dimensional (which for nonparametric estimators implemented against a few hundred observations arises once $K \geq 3$). Thus it is not possible to estimate $F_{n-1:n}(\cdot | x)$ precisely in a fully nonparametric way. But conditioning on x is necessary for the Athey and Haile test (4), since other-

wise bidders would not have independent valuations. However in the case of our tests (10) and (11), however, we can simply incorporate x into the unobservable θ and pursue the analysis unconditional on x . Thus we implement the tests by estimating $G_{n-1:n}$ and $G_{n:n}$ without any conditioning whatsoever, freeing ourselves the curse of dimensionality problem. Should we choose to condition on some subset of x , or control for some of the heterogeneity across auctions, our test remains valid; whatever residual variation we do not capture is simply treated as part of θ .

4 Identification under Exogenous Entry

4.1 Identifying $F_{j:k}$ Given Infinite Data

If entry is exogenous in the sense we have defined (confidence in this assumption could be achieved by running our test from last section, for example), we now show that the seller can use variation in the number of bidders n across auctions to identify the unobserved primitives and through them the optimal reserve price. The essential problem is that with the existence of the demand shock $\theta \in \Theta$, the seller faces a correlated demand structure within an auction, which is not identified under standard assumptions (Athey and Haile 2002). However, in this section, we will show that the variation induced by exogenous entry does identify the *marginal* distribution of each order statistic of bidder valuations, $V_{j:k}$. (We will first show that these distributions are precisely identified under Assumption 2 and “infinite data,” and then show that these results lead to bounds on these distributions under “finite data” and the weaker Assumption 1.) We then show that under reasonable assumptions, identification of these distributions is “enough”: that expected revenue in an auction with m bidders, and the reserve price that maximizes it, depend only on the marginal distributions $F_{m-1:m}$ and $F_{m:m}$.

For any fixed number of bidders m , $F_{m-1:m}$ is identified from the data, but $F_{m:m}$ is not. We show now how variation in the number of bidders $n > m$

across auctions (assuming exogenous entry) can be used to identify $F_{m:m}$ for any m . To see the intuition behind the result, let H be short for $H(v|\theta)$, and recall that

$$\begin{aligned} F_{n:n}(v) &= E_\theta \{H^n\} \\ F_{n-1:n}(v) &= E_\theta \{nH^{n-1} - (n-1)H^n\} \end{aligned} \tag{12}$$

Let $m = 3$, that is, we will identify $F_{3:3}$ from $\{F_{n-1:n}\}_{n>3}$. We know that

$$\begin{aligned} F_{3:4}(v) &= E_\theta \{ 4H^3 - 3H^4 \} \\ F_{4:5}(v) &= E_\theta \{ 5H^4 - 4H^5 \} \\ F_{5:6}(v) &= E_\theta \{ 6H^5 - 5H^6 \} \\ F_{6:7}(v) &= E_\theta \{ 7H^6 - 6H^7 \} \end{aligned}$$

We can then construct linear combinations to cancel terms, giving

$$\begin{aligned} \frac{1}{4}F_{3:4}(v) &= E_\theta \{ H^3 - \frac{3}{4}H^4 \} \\ \frac{1}{4}F_{3:4}(v) + \frac{3}{20}F_{4:5}(v) &= E_\theta \{ H^3 - \frac{3}{5}H^5 \} \\ \frac{1}{4}F_{3:4}(v) + \frac{3}{20}F_{4:5}(v) + \frac{1}{10}F_{5:6}(v) &= E_\theta \{ H^3 - \frac{3}{6}H^6 \} \\ \frac{1}{4}F_{3:4}(v) + \frac{3}{20}F_{4:5}(v) + \frac{1}{10}F_{5:6}(v) + \frac{5}{70}F_{6:7}(v) &= E_\theta \{ H^3 - \frac{3}{7}H^7 \} \end{aligned}$$

With each added term, the right-hand side moves toward $E_\theta H^3 = F_{3:3}(v)$.

We now present the general result. In addition to identifying $F_{m:m}$ for each m , we can then use $\{F_{m:m}\}$ to identify the distribution of any other order statistic $V_{j:k}$.

Theorem 4.1. *Suppose bidder values are conditionally independent: in each auction, bidder values are i.i.d. draws from a distribution $H(\cdot|\theta)$, where θ is unobserved. Suppose that n varies exogenously, that is, the number of bidders in each auction is independent of θ . Suppose that $F_{n-1:n}(v)$ is known exactly for every $n \in \{2, 3, \dots\}$. Then $F_{j:k}$ is identified for every j, k with $j \leq k$.*

Proof. First, for any m , we will show how $F_{m:m}(v)$ is uniquely determined by $\{F_{n-1:n}(v)\}_{n>m}$. Let H be short for $H(v|\theta)$. For any $M > m + 1$,

$$\begin{aligned}
& \frac{1}{m-1} \sum_{n=m+1}^M \left(\prod_{i=m}^{n-1} \frac{i-1}{i+1} \right) F_{n-1:n}(v) \\
&= \frac{1}{m-1} \sum_{n=m+1}^M \left(\prod_{i=m}^{n-1} \frac{i-1}{i+1} \right) E_\theta \{nH^{n-1} - (n-1)H^n\} \\
&= \frac{1}{m-1} \sum_{n=m+1}^{m+1} \left(\prod_{i=m}^{n-1} \frac{i-1}{i+1} \right) E_\theta \{nH^{n-1}\} \\
&\quad + \frac{1}{m-1} \sum_{n=m+2}^M \left(\prod_{i=m}^{n-1} \frac{i-1}{i+1} \right) E_\theta \{nH^{n-1}\} \\
&\quad - \frac{1}{m-1} \sum_{n=m+1}^{M-1} \left(\prod_{i=m}^{n-1} \frac{i-1}{i+1} \right) E_\theta \{(n-1)H^n\} \\
&\quad - \frac{1}{m-1} \sum_{n=M}^M \left(\prod_{i=m}^{n-1} \frac{i-1}{i+1} \right) E_\theta \{(n-1)H^n\} \\
&= \frac{1}{m-1} \frac{m-1}{m+1} E_\theta \{(m+1)H^m\} \\
&\quad + \frac{1}{m-1} \sum_{n=m+2}^M \left(\prod_{i=m}^{n-1} \frac{i-1}{i+1} \right) E_\theta \{nH^{n-1}\} \\
&\quad - \frac{1}{m-1} \sum_{n'=m+1}^{M-1} \left(\prod_{i=m}^{(n'-1)-1} \frac{i-1}{i+1} \right) E_\theta \{((n'-1)-1)H^{n'-1}\} \\
&\quad - \frac{1}{m-1} \left(\prod_{i=m}^{M-1} \frac{i-1}{i+1} \right) E_\theta \{(M-1)H^M\} \\
&= E_\theta H^m \\
&\quad + \frac{1}{m-1} \sum_{n=m+2}^M \left(\prod_{i=m}^{n-2} \frac{i-1}{i+1} \right) \frac{n-2}{n} E_\theta \{nH^{n-1}\} \\
&\quad - \frac{1}{m-1} \sum_{n'=m+2}^M \left(\prod_{i=m}^{n'-2} \frac{i-1}{i+1} \right) E_\theta \{(n'-2)H^{n'-1}\} \\
&\quad - \frac{1}{m-1} \left(\prod_{i=m}^{M-1} \frac{i-1}{i+1} \right) E_\theta \{(M-1)H^M\} \\
&= E_\theta H^m - \frac{m}{M} E_\theta H^M
\end{aligned}$$

(To get the third equality, we simply introduced a new index $n' = n + 1$ and substituted $n' - 1$ for n everywhere in the third expression.) Since $H \leq 1$, the second term vanishes as $M \rightarrow \infty$, so

$$F_{m:m}(v) = E_\theta H^m(v|\theta) = \frac{1}{m-1} \sum_{n=m+1}^{\infty} \left(\prod_{i=m}^{n-1} \frac{i-1}{i+1} \right) F_{n-1:n}(v) \quad (13)$$

Next, pick arbitrary j, k with $j \leq k$. We know that

$$F_{j:k}(v) = E_{\theta} \{F_{j:k}(v|\theta)\} = E_{\theta} \left\{ \sum_{i=j}^k \binom{k}{i} H^i (1-H)^{k-i} \right\} \quad (14)$$

If we expand the polynomial $(1-H)^{k-i}$ term and sum over the different i , we can write this as

$$F_{j:k}(v) = E_{\theta} \left\{ \sum_{m=j}^k a_m H^m \right\} \quad (15)$$

where a_m are scalar coefficients. Then given linearity of expectations,

$$F_{j:k}(v) = \sum_{m=j}^k a_m E_{\theta} \{H^m(v|\theta)\} = \sum_{m=j}^k a_m F_{m:m}(v) \quad (16)$$

completing the proof. \square

4.2 Bounds on $F_{m:m}$ Given Finite Data

The above identification results relied on having data from auctions with n bidders for every positive $n \geq 2$. If there is some upper bound M on the number of bidders that appear in auctions in the data, then point identification of the optimal reserve price won't be possible, but we can nevertheless provide bounds on $F_{m:m}$, leading to bounds on π_n and on the optimal reserve price.

First, suppose we want to bound $F_{m:m}$ using only a finite number of terms $F_{m:m+1}, F_{m+1:m+2}, \dots, F_{M-1:M}$. Rearranging our sum from before,

$$F_{m:m}(v) = \frac{1}{m-1} \sum_{n=m+1}^M \left(\prod_{i=m}^{n-1} \frac{i-1}{i+1} \right) F_{n-1:n}(v) + \frac{m}{M} F_{M:M}(v) \quad (17)$$

Conveniently, conditional independence leads to upper and lower bounds on

$F_{M:M}(v)$, as a function of $F_{M-1:M}(v)$:

Lemma 4.2. *For any n , under conditionally independent private values,*

$$\psi_{n:n} \circ \psi_{n-1:n}^{-1}(F_{n-1:n}(v)) \leq F_{n:n}(v) \leq F_{n-1:n}(v) \quad (18)$$

Proof. The upper bound is because by definition, $V^{n:n} \geq V^{n-1:n}$. The lower bound is because

$$\frac{d}{ds} (\psi_{n:n} \circ \psi_{n-1:n}^{-1})(s) = \frac{\psi'_{n:n}(t)}{\psi'_{n-1:n}(t)} = \frac{nt^{n-1}}{n(n-1)t^{n-2}(1-t)} = \frac{t}{(n-1)(1-t)}$$

where $t = \psi_{M-1:m}^{-1}(s)$, which is increasing in t and therefore s , so $\psi_{n:n} \circ \psi_{n-1:n}^{-1}$ is convex. Using Jensen's Inequality, then,

$$\begin{aligned} \psi_{n:n} \circ \psi_{n-1:n}^{-1}(F_{n-1:n}(v)) &= (\psi_{n:n} \circ \psi_{n-1:n}^{-1})(E_{\theta} \psi_{n-1:n}(H(v|\theta))) \\ &\leq E_{\theta} (\psi_{n:n} \circ \psi_{n-1:n}^{-1})(\psi_{n-1:n}(H(v|\theta))) \\ &= E_{\theta} \psi_{n:n}(H(v|\theta)) \\ &= F_{n:n}(v) \end{aligned}$$

completing the proof. □

Plugging the bounds given by (18) into (17) provides upper and lower bounds on $F_{m:m}(v)$:

$$\begin{aligned} \frac{1}{m-1} \sum_{n=m+1}^M \left(\prod_{i=m}^{n-1} \frac{i-1}{i+1} \right) F_{n-1:n}(v) + \frac{m}{M} (\psi_{M:M} \circ \psi_{M-1:M}^{-1})(F_{M-1:M}(v)) \\ \leq F_{m:m}(v) \leq \\ \frac{1}{m-1} \sum_{n=m+1}^M \left(\prod_{i=m}^{n-1} \frac{i-1}{i+1} \right) F_{n-1:n}(v) + \frac{m}{M} F_{M-1:M}(v) \end{aligned} \quad (19)$$

These bounds can be explicitly calculated under Assumption 2. Under the weaker Assumption 1, we can combine (19) with the bounds on $F_{n-1:n}$ implied by Assumption 1, $G_{n:n}(v - \delta) \leq F_{n-1:n}(v) \leq G_{n-1:n}(v)$. (We do this explicitly in the next section.)⁹

⁹Upper and lower bounds on $F_{m:m}(v)$ for each $m \in \{j, \dots, k\}$ would lead to upper and

4.3 Bounding π_n and the Optimal Reserve Price

Next, we consider the relationship between the marginal distributions of $V_{n-1:n}$ and $V_{n:n}$ and the outcome of an ascending auction with n bidders. We will show that, under reasonable assumptions, bounds on $F_{n-1:n}$ and $F_{n:n}$ lead to bounds on π_n , and through them, to bounds on the optimal reserve price.

To get from $(F_{n-1:n}, F_{n:n})$ to expected revenue, Haile and Tamer (2003) appeal to revenue equivalence: that whatever exact mechanism $\mu \in \mathcal{M}$ is being played, as long as the bidder with the highest valuation wins and the bidders are playing a Bayesian Nash equilibrium, any mechanism with a reserve price r must be revenue-equivalent to the second-price sealed-bid auction with the same reserve price. In a second-price sealed-bid auction, the winner pays $\max\{V_{n-1:n}, r\}$; so the expected revenue is

$$\pi_n^{SP}(r) = E_{V_{n-1:n}, V_{n:n}} \{ \mathbf{1}_{V_{n:n} \geq r} (\max\{V_{n-1:n}, r\} - v_0) \} \quad (20)$$

However, the revenue equivalence argument does not automatically transfer to our setting, since revenue equivalence does not hold when bidder valuations are correlated. Thus, how precisely $\pi_n(r)$ is pinned down by $F_{n-1:n}$ and $F_{n:n}$ depends on the exact assumptions we are willing to make about bidder knowledge and behavior. We consider two sets of assumptions:

Assumption 3. *In each auction, bidders each learn θ as well as their own private valuation; and in whatever auction is chosen, bidders play a Bayesian Nash equilibrium in which the bidder with the highest value always wins.*

One natural defense of the assumption that bidders observe θ is that *all* heterogeneity in objects might be observable, but the econometrician might choose to condition only on a subset for reasons of dimensionality or limited

lower bounds on $F_{j:k}(v)$ via (16), but as we show below, only $F_{m-1:m}$ and $F_{m:m}$ turn out to be payoff-relevant.

data. Thus, from the econometrician’s point of view, residual valuations are still correlated, while from the bidders’, knowing θ , they are independent.

Nonetheless, the assumption that bidders play a Bayesian Nash equilibrium, while standard, is much stronger than Assumption 1, so we consider a weaker set of bidding assumptions as well:

Assumption 4. *Bidders may or may not observe θ , and may or may not play equilibrium bidding strategies; but bidders play strategies satisfying Assumption 1, and there are no jump bids (bidders always raise the standing high bid by the minimum increment δ).*

If δ is small, then expected revenue is nearly pinned down under either Assumption 3 or 4:

Lemma 4.3. 1. *Under Assumption 3, $\pi_n(r) = \pi_n^{SP}(r)$.*

2. *Under Assumption 4, $\pi_n(r) \in [\pi_n^{SP}(r) - \delta, \pi_n^{SP}(r) + \delta]$.*

Proof. Part 1 is because, if θ is common knowledge among the bidders, then conditional on a realization of θ , we are back in a symmetric IPV setting and revenue equivalence holds.

Part 2 is because, under Assumption 1 and without jump bids, the winning bid must be at least $V_{n-1:n} - \delta$ and at most $V_{n-1:n} + \delta$ when $V_{n-1:n} \geq r$, and exactly r when $V_{n-1:n} < r$; taking the expectation, π_n must be within δ of the right-hand side of (20). \square

(If bidders do employ jump bids, then under Assumption 1, $\pi_n(r) \geq \pi_n^{SP}(r) - \delta$, but $\pi_n(r)$ could be significantly higher than $\pi_n^{SP}(r)$, and we cannot place bounds on the optimal reserve price without a fuller model of bidders’ bidding strategies.¹⁰)

¹⁰For example, the strategy “increase my bid to 99% of my willingness to pay if I get outbid three times” is consistent with Assumption 1, but would likely favor a much lower reserve price than the analysis below would suggest.

Next, we show how bounds on $F_{n-1:n}$ and $F_{n:n}$ translate to bounds on π_n^{SP} ; combined with Lemma 4.3, this will give bounds on π_n and therefore bounds on $r^* = \arg \max \pi_n(r)$. Note first that (20) can be rewritten as

$$\pi_n^{SP}(r) = (F_{n-1:n}(r) - F_{n:n}(r))(r - v_0) + \int_r^\infty (v - v_0) dF_{n-1:n}(v) \quad (21)$$

which depends only on the marginal distributions $F_{n-1:n}$ and $F_{n:n}$; so once both of these are identified, π_n is identified as well (exactly under Assumption 3, or to within $\pm\delta$ under Assumption 4).¹¹

Next, note that for $r \geq v_0$ (the seller's cost), $\mathbf{1}_{V_{n:n} \geq r} (\max\{V_{n-1:n}, r\} - v_0)$ is increasing in both $V_{n:n}$ and $V_{n-1:n}$; and so $\pi_n^{SP}(r)$, being its expectation, is stochastically increasing in both $F_{n-1:n}$ and $F_{n:n}$.¹² Thus, plugging the upper bounds on $F_{n-1:n}$ and $F_{n:n}$ into (21) (and subtracting δ under Assumption 4) gives the lower bound on $\pi_n(r)$; plugging in the lower bounds (and adding δ) gives the upper bound. (We do this explicitly in the next section.)

Once we have pointwise upper and lower bounds $\bar{\pi}_n(r)$ and $\underline{\pi}_n(r)$, we can use these to bound the revenue-maximizing reserve price r^* , as in Haile and Tamer (2003), via the inequality

$$\bar{\pi}_n(r^*) \geq \max_r \underline{\pi}_n(r) \quad (22)$$

This holds by optimality, because $\bar{\pi}_n(r^*) \geq \pi_n(r^*) \geq \pi_n(r') \geq \underline{\pi}_n(r')$, where

¹¹Similarly, a bidder's surplus in such an auction is (exactly under Assumption 3, or to within δ under Assumption 4) $\mathbf{1}_{V_{n:n} \geq r} (V_{n:n} - \max\{r, V_{n-1:n}\}) = \max\{r, V_{n:n}\} - \max\{r, V_{n-1:n}\}$ if he is the bidder with the highest valuation, and 0 otherwise; the expected value of this expression can be simplified to

$$\frac{1}{n} \int_r^\infty (F_{n-1:n}(v) - F_{n:n}(v)) dv$$

which again depends only on $F_{n:n}$ and $F_{n-1:n}$. Thus, under the Levin and Smith (1994) model of endogenous entry which we discuss later, bounds on $F_{n-1:n}$ and $F_{n:n}$ are similarly sufficient to model the entry problem and therefore the effect of r on participation.

¹²That is, replacing $F_{n-1:n}$ or $F_{n:n}$ with a distribution that first-order stochastically dominates it, increases $\pi_n^{SP}(r)$.

$r' = \arg \max_r \underline{\pi}_n(r)$. This specifies an interval $[\underline{r}, \bar{r}]$ within which r^* must lie; when the bounds $\bar{\pi}_n$ and $\underline{\pi}_n$ get close, this interval becomes narrow.

4.4 What If Entry Is Endogenous

If the data fails the test in Corollary 3.4, forcing us to abandon the assumption of exogenous entry, we can still place useful bounds on $F_{n:n}(r)$, and therefore $\pi_n(r)$, without relying on variation in n . Lemma 4.2 does not rely on variation in n ; so $\psi_{n:n} \circ \psi_{n-1:n}^{-1}(F_{n-1:n}(v)) \leq F_{n:n}(v) \leq F_{n-1:n}(v)$. Under Assumptions 2 and 3, plugging these upper and lower bounds into (21) gives pointwise upper and lower bounds on $\pi_n(r)$. (Under the weaker Assumptions 1 and 4, we use the upper and lower bounds on $F_{n-1:n}$ implied by the bidding assumptions, and widen the bounds on π_n by δ as above.)

Given these pointwise bounds on $\pi_n(r)$, (22) still holds, providing some bounds on the revenue-maximizing reserve price r^* . However, $\underline{\pi}_n(r)$ is low enough now that these bounds are extremely wide. (Like in Quint (2008), which considers only affiliated private values, the lower bound on r^* is the seller's valuation v_0 ; unlike in Quint (2008), however, the upper bound here is *above* the reserve price that would be optimal under independence.)

5 Summing Up

So, you're a seller, with one object to sell which you value at v_0 . You've got data from some past auctions, and you're trying to decide how to set the reserve price for your next auction. What do you do?

Unfortunately, it depends on whether you believe the number of bidders who choose to bid in your auction depends on the reserve price you set, which we do not yet have a test for. However, knowing the answer to that question, the policy conclusions are clear:

Proposition 5.1. *If entry is endogenous, the optimal reserve price is equal to v_0 .*

This is based on the model of entry from Levin and Smith (1994), in which potential bidders explicitly calculate their ex-ante expected surplus from participating in the auction (in expectation over the number of other bidders who may enter), weigh it against the cost of evaluating the object and participating in the auction, and play a symmetric mixed strategy in which some participate and some do not. In this model, raising the reserve price decreases bidder participation; setting $r = v_0$ (and charging no entry fee, even if one were permitted) ends up maximizing expected seller profits. (Levin and Smith (1994) claim this result for $v_0 = 0$ and private values which are independent or affiliated. Their proof, however, holds (with slight modification) for arbitrary v_0 and arbitrarily correlated symmetric private values. For completeness, we reproduce the proof in the appendix.)

Conveniently, this recommendation does not depend on whether potential bidders know θ or do not know θ at the time they decide to participate; in either case, expected profits are maximized by setting $r = v_0$. However, it does assume that bidder participation responds to r , which our test above does not check for. (Our test (10) tests whether participation responds to θ , not to r , which we assumed did not bind in the data.) It may be reasonable to suppose that if participation responds to θ , bidders are somehow considering their expected surplus from the auction, rather than being dropped randomly into auctions; thus, it may seem as well that participation would respond to changes in r . However, an explicit test of this could easily be run by randomly varying r in a series of auctions and seeing how n responded.

On the other hand, when entry is exogenous (n does not respond to either r or θ), our identification results hold:

Proposition 5.2. *If entry is exogenous, the optimal reserve price is identified under Assumptions 2 and 3 and infinite data; and it is bounded under the weaker Assumptions 1 and 4 and finite data.*

To calculate the optimal reserve price, we do the following:

1. For each n , calculate upper and lower bounds on $F_{n-1:n}$ based on observed bid data:

$$\begin{aligned} F_{n-1:n}(v) &\leq \bar{F}_{n-1:n}(v) \equiv G_{n-1:n}(v) \\ F_{n-1:n}(v) &\geq \underline{F}_{n-1:n}(v) \equiv G_{n:n}(v - \delta) \end{aligned} \quad (23)$$

(As discussed above, these follow from the bidding assumptions of Haile and Tamer (2003). Under the stronger Assumption 2, we would instead use $\bar{F}_{n-1:n}(v) = \underline{F}_{n-1:n}(v) = G_{n-1:n}(v)$.)

2. For each m , use these bounds and (19) (and the highest M possible) to calculate upper and lower bounds on $F_{m:m}(v)$:

$$\begin{aligned} F_{m:m}(v) &\leq \bar{F}_{m:m}(v) \equiv \frac{1}{m-1} \sum_{n=m+1}^M \left(\prod_{i=m}^{n-1} \frac{i-1}{i+1} \right) \bar{F}_{n-1:n}(v) \\ &\quad + \frac{m}{M} \bar{F}_{M-1:M}(v) \\ F_{m:m}(v) &\geq \underline{F}_{m:m}(v) \equiv \frac{1}{m-1} \sum_{n=m+1}^M \left(\prod_{i=m}^{n-1} \frac{i-1}{i+1} \right) \underline{F}_{n-1:n}(v) \\ &\quad + \frac{m}{M} (\psi_{M:M} \circ \psi_{M-1:M}^{-1}) (\underline{F}_{M-1:M}(v)) \end{aligned} \quad (24)$$

3. For each n , use these bounds, (21), and Lemma 4.3 to calculate upper and lower bounds on $\pi_n(r)$:

$$\begin{aligned} \pi_n(r) &\leq \bar{\pi}_n(r) \equiv (\underline{F}_{n-1:n}(r) - \underline{F}_{n:n}(r)) (r - v_0) \\ &\quad + \int_r^\infty (v - v_0) d\underline{F}_{n-1:n}(v) + \delta \\ \pi_n(r) &\geq \underline{\pi}_n(r) \equiv (\bar{F}_{n-1:n}(r) - \bar{F}_{n:n}(r)) (r - v_0) \\ &\quad + \int_r^\infty (v - v_0) d\bar{F}_{n-1:n}(v) - \delta \end{aligned} \quad (25)$$

(Under the stronger Assumption 3, we can remove the δ from both bounds. Note also the counterintuitive fact that $\bar{\pi}_n$ is calculated from $\underline{F}_{n-1:n}$ and $\underline{F}_{n:n}$ and vice versa; this is because lower values of a cumu-

lative distribution function indicate stochastically higher values of the random variable.)

4. If n is known (or you are able to vary the reserve price in response to the number of bidders), use $\underline{\pi}_n$ and $\bar{\pi}_n$, and (22), to calculate the optimal reserve price for an auction of size n :

$$\begin{aligned} r_n^* \leq \bar{r}_n^* &\equiv \max \{r : \bar{\pi}_n(r) \geq \max_{r'} \underline{\pi}_n(r')\} \\ r_n^* \geq \underline{r}_n^* &\equiv \min \{r : \bar{\pi}_n(r) \geq \max_{r'} \underline{\pi}_n(r')\} \end{aligned} \quad (26)$$

5. If n is not known, then for each n , calculate the probability p_n of exactly n bidders participating; define

$$\begin{aligned} \bar{\pi}(r) &\equiv \sum_n p_n \bar{\pi}_n(r) \\ \underline{\pi}(r) &\equiv \sum_n p_n \underline{\pi}_n(r) \end{aligned} \quad (27)$$

and use $\bar{\pi}$ and $\underline{\pi}$ to calculate bounds on the optimal reserve price:

$$\begin{aligned} r^* \leq \bar{r}^* &\equiv \max \{r : \bar{\pi}(r) \geq \max_{r'} \underline{\pi}(r')\} \\ r^* \geq \underline{r}^* &\equiv \min \{r : \bar{\pi}(r) \geq \max_{r'} \underline{\pi}(r')\} \end{aligned} \quad (28)$$

As we noted before, there are two different ways in which entry can be endogenous: n can be related to r , and n can be related to θ . Our test in Section 3 attempts to detect when n responds to θ . When n responds to both r and θ , Proposition 5.1 above holds. When n responds to neither r nor θ , Proposition 5.2 holds.

When n responds to r but not to θ , Proposition 5.1 still holds ($r^* = v_0$), but the test in Section 3 will not detect the endogeneity. When n responds to θ but not r , the optimal reserve price will be greater than v_0 , but is not identified from bidding data.¹³ A separate test for whether n responds to

¹³Quint (2008) shows that when bidders have *affiliated* private values, the optimal reserve price is weakly lower than it would be if bidders' values were *i.i.d.* draws from the distribution $\psi_{n-1:n}^{-1}(F_{n-1:n})$. However, we have a counterexample showing that this result

(for example, experimentally varying r and measuring the response) would complement the test in Section 3 and fully distinguish between these four cases.

6 Monte Carlo and Empirical Application

(Coming soon.)

does not always hold for conditionally independent private values.

Appendix. $r^* = v_0$ under Endogenous Entry

This is essentially the proof from Levin and Smith (1994); since they claimed the result only for $v_0 = 0$ and affiliated values, we reproduce the proof here to show that it extends to our setting.

Suppose there are N potential bidders, each of whom face an entry cost $c > 0$ to learn their valuation for the object and participate in the auction. Suppose that N is large enough that all bidders participating in the auction is unprofitable for them. Suppose bidders have exchangeable private values which are otherwise arbitrarily jointly distributed. Assume that the bidders play the unique symmetric equilibrium in which they each enter with the same probability q .

Theorem. *Setting reserve price equal to v_0 (regardless of how many entrants there are) maximizes the seller's expected profit.*

Proof. We will first show that if the seller is allowed to charge a fee to each bidder who participates, then setting $r = v_0$ would be optimal; we will then show that the optimal entry fee is 0.

If an entry fee is available, reserve prices $r \neq v_0$ are never optimal (even when the reserve price is allowed to depend on the number of entrants). This is because any set of (potentially entry-dependent) reserve prices (r_1, r_2, \dots, r_N) , along with an entry fee e , leads to some level of entry q which could be replicated by setting all the reserve prices to v_0 and charging a different entry fee e' . This change would leave entry the same, but it would lead to greater ex-post efficiency. But since (by assumption) bidders mix between entering and not entering, they must get 0 expected profit; so the seller captures all the surplus, which means that this change would lead to higher revenue. So now we only worry about setting the optimal e .

Since we now consider only auctions with reserve price $r = v_0$, we can without loss replace each bidder's private value V_i with $\max\{V_i, v_0\}$. This is without loss because a bidder with value $V_i < v_0$ gets the same ex post surplus (0) as a bidder with value $V_i = v_0$, and a seller gets the same profit from a bidder with value $V_i < v_0$ as from a bidder with $V_i = v_0$ (whether this bidder is the highest, second-highest, or lower). So replace each bidder's value V_i with $\max\{V_i, v_0\}$, and let $\bar{v}^{k:n}$ be the expected value of the k^{th} lowest of these when n bidders participate.

Next, put aside the seller's problem and suppose instead that we were a social planner who could choose q (each bidder's probability of entry) to maximize total surplus. Note that when n bidders arrive and participate in an auction with reserve price $r = v_0$, the expected social surplus is simply $E(\max\{V^{n:n}, v_0\}) = \bar{v}^{n:n}$; so a social planner would maximize

$$S(q) = \sum_{n=0}^N \binom{N}{n} q^n (1-q)^{N-n} \bar{v}^{n:n} - Nqc$$

where N is the number of potential bidders and c is the entry cost. Differentiating with respect to q gives

$$\frac{dS}{dq} = \sum_{n=0}^N \binom{N}{n} [nq^{n-1}(1-q)^{N-n} - (N-n)q^n(1-q)^{N-n-1}] \bar{v}^{n:n} - Nc$$

Dividing by N and setting this equal to 0 gives

$$\sum_{n=0}^N \binom{N}{n} \left[\frac{n}{N} q^{n-1} (1-q)^{N-n} - \frac{N-n}{N} q^n (1-q)^{N-n-1} \right] \bar{v}^{n:n} = c$$

Splitting this into two separate sums, and noting that the $n = 0$ term of the first and the $n = N$ term of the second vanish, gives

$$\sum_{n=1}^N \binom{N}{n} \frac{n}{N} q^{n-1} (1-q)^{N-n} \bar{v}^{n:n} - \sum_{n=0}^{N-1} \binom{N}{n} \frac{N-n}{N} q^n (1-q)^{N-n-1} \bar{v}^{n:n} = c$$

Combining the fractions $\frac{n}{N}$ and $\frac{N-n}{N}$ with the adjacent combinatorial terms gives

$$\sum_{n=1}^N \binom{N-1}{n-1} q^{n-1} (1-q)^{N-n} \bar{v}^{n:n} - \sum_{n=0}^{N-1} \binom{N-1}{n} q^n (1-q)^{N-n-1} \bar{v}^{n:n} = c$$

Substituting a new index $n' = n - 1$ into the first sum gives

$$\sum_{n'=0}^{N-1} \binom{N-1}{n'} q^{n'} (1-q)^{N-n'-1} \bar{v}^{n'+1:n'+1} - \sum_{n=0}^{N-1} \binom{N-1}{n} q^n (1-q)^{N-n-1} \bar{v}^{n:n} = c$$

Finally, replacing n' with n and combining the two sums gives

$$\sum_{n=0}^{N-1} \binom{N-1}{n} q^n (1-q)^{N-n-1} (\bar{v}^{n+1:n+1} - \bar{v}^{n:n}) = c$$

Next, start with a sample of $n+1$ bidders, and randomly select n of them. With probability $\frac{n}{n+1}$, your subsample includes the one with the highest V_i ; with probability $\frac{1}{n+1}$, it contains the second-highest but not the highest. Since these probabilities are independent of $\{V_i\}$,

$$\bar{v}^{n:n} = \frac{n}{n+1} \bar{v}^{n+1:n+1} + \frac{1}{n+1} \bar{v}^{n:n+1}$$

Plugging this into the last expression gives

$$\sum_{n=0}^{N-1} \binom{N-1}{n} q^n (1-q)^{N-n-1} \frac{1}{n+1} (\bar{v}^{n+1:n+1} - \bar{v}^{n:n+1}) = c \quad (29)$$

which defines the q that maximizes social surplus.

Next, consider one bidder's problem. He knows that with probability $\binom{N-1}{n} q^n (1-q)^{N-n-1}$, exactly n of his opponents will enter. Should that happen and he chooses to enter, he'll have the highest value with probability $\frac{1}{n+1}$; and should that happen, his expected surplus will be $\bar{v}^{n+1:n+1} - \bar{v}^{n:n+1}$. So the unique symmetric equilibrium to the entry game involves each bidder entering with probability q , where

$$\sum_{n=0}^{N-1} \binom{N-1}{n} q^n (1-q)^{N-n-1} \frac{1}{n+1} (\bar{v}^{n+1:n+1} - \bar{v}^{n:n+1}) = c + e \quad (30)$$

where e is the entry fee imposed by the seller. Comparing (29) to (30), when $e = 0$, the socially optimal level of entry occurs. Since bidders get expected payoff of 0, social surplus equals seller's expected profit, which is therefore maximized by setting $r = v_0$ and $e = 0$. Since this does not depend on the joint distribution of bidder valuations, it is true whether or not entrants condition entry on the value of θ . \square

References

- Athey, S. and P. Haile (2002). Identification of standard auction models. *Econometrica* 70(6), 2107–2140.
- Cremer, J. and R. McLean (1988). Full extraction of the surplus in Bayesian and dominant strategy auctions. *Econometrica* 56(6), 1247–1257.
- Haile, P. and E. Tamer (2003). Inference with an incomplete model of English auctions. *Journal of Political Economy* 111(1), 1–51.
- Hewitt, E. and L. Savage (1955). Symmetric measures on Cartesian products. *Transactions of the American Mathematical Society* 80(2), 470–501.
- Levin, D. and J. Smith (1994). Equilibrium in auctions with entry. *The American Economic Review* 84(3), 585–599.
- McAfee, R. and J. McMillan (1987). Auctions with entry. *Economics Letters* 23(4), 343–347.
- McAfee, R. and P. Reny (1992). Correlated Information and Mechanism Design. *Econometrica* 60(2), 395–421.
- Paarsch, H. (1997). Deriving an estimate of the optimal reserve price: an application to British Columbian timber sales. *Journal of Econometrics* 78(2), 333–357.
- Paarsch, H. J. (1992). Deciding between the common and private value paradigms in empirical models of auctions. *Journal of Econometrics* 51(1-2), 191–215.
- Quint, D. (2008). Unobserved correlation in private-value ascending auctions. *Economics Letters* 100(3), 432–434.
- Samuelson, W. (1985). Competitive bidding with entry costs. *Economics Letters* 17(1-2), 53–57.
- Silverman, B. (1986). *Density estimation for statistics and data analysis*. Chapman & Hall/CRC.