Both Sides Now: Identification of Buyer and Seller Values in Ascending Auctions ⋆

Aaron Barkley[a], David Genesove[b], James Hansen[a]

[a]*Department of Economics The University of Melbourne*
[b]*Department of Economics The Hebrew University of Jerusalem and C.E.P.R.*

Abstract

Auctions are inherently two-sided, yet empirical auction models typically recover bidder values while treating sellers' preferences as known, fixed, or irrelevant. We study identification of buyer and seller primitives in ascending auctions in which a sale occurs only when the highest buyer value exceeds a seller-specific threshold. Using variation in bidder counts, we characterise identification across data environments and modelling assumptions. When sale outcomes are observed but rejected bids are not, buyer value and seller threshold distributions are point identified. When only successful auctions are observed, identification reduces to a one-dimensional fixed-point problem in the ratio of sale probabilities across bidder counts. We show that the set of fixed points is finite, and provide an upper-tail support condition guaranteeing uniqueness. A conduct assumption maps seller thresholds into values; weaker conduct restrictions generate bounds. Monte Carlo experiments validate our approach and quantify biases from ignoring rejected offers.

Keywords: Auctions, identification, set identification
JEL: C57, D44

October 2026
PRELIMINARY
Please do not cite without permission

⋆We thank David P. Byrne, Tong Li, Simon Loertscher, Xiaoyun Qiu, Emanuele Tarantino, and participants at various conferences and seminars for their helpful comments. This paper previously circulated under the title "Auction Identification with Unobserved Rejected Offers."

Email addresses: aaron.barkley@unimelb.edu.au (Aaron Barkley), david.genesove@mail.huji.ac.il (David Genesove), james.hansen@unimelb.edu.au (James Hansen)

1. Introduction

All auctions have two sides: they allocate an object between a seller who may retain it and buyers who compete to acquire it. Yet the empirical auction literature has concentrated on only the bidding side, recovering buyer value distributions via bids or transaction prices. The seller's preferences, in contrast, are commonly represented by a fixed reserve price, or eliminated by assuming that reserves are absent or non-binding. This asymmetry is consequential because sellers often reject the highest bid.[1] Two issues arise: (i) *selection*, because estimation regularly uses transaction records, which condition on sale and bias bidder value estimates upward; and (ii) *model misspecification*, because omitting the seller removes a primitive that determines allocations and prices and that is needed for counterfactuals involving reserve policies, entry, market design, and welfare.

We study the identification of both buyer and seller primitives in ascending auctions. We formulate a second-price auction model in which a sale occurs only if the highest bidder's value exceeds a seller-chosen threshold.[2] The threshold captures the seller's right to refuse bids without requiring a particular theory of how that right is exercised. The allocation and payment rule is consistent with a knockout auction followed by a seller take-it-or-leave-it offer (Bulow and Klemperer, 1996), and it captures the features of many ascending auctions used in practice, including those for automobiles, livestock, art and collectables, and real estate, in which the seller may retain the object when bidding is insufficient. We assume independent private values and that the distributions of buyer values and seller thresholds have discrete support. The main seller-side object is the threshold distribution. A "conduct" assumption on how thresholds are determined by the seller – e.g., an optimal reserve – then provides a mapping between the threshold and seller value distributions.

Our identification results cover a range of model and data environments. Throughout, the researcher observes transaction prices and bidder counts for successful auctions, and identification uses exogenous variation in the number of bidders.[3] We show that identifi-

[1] Sellers exercise the right to reject the highest bid in many settings, including auctions for automobiles (Genesove, 1995, Lacetera et al., 2016), consumer goods (Einav et al., 2015), housing (Genesove and Hansen, 2023), financial products (McConnell and Saretto, 2010), art (Ashenfelter and Graddy, 2006), and mineral rights (Hendricks et al., 1994).

[2] We refer to the bidding side as "buyers" and the side represented by the auctioneer as the "seller" throughout the paper to simplify the exposition, noting that our approach also applies to procurement auctions where these roles are reversed.

[3] We can also accommodate selective entry within the framework of Gentry and Li (2014), as we discuss in Appendix D. Bidder count variation has been extensively employed in the empirical auctions literature, including in Haile et al. (2003), Aradillas-López et al. (2013), and Hernández et al. (2020).

cation reduces to a single scalar: the ratio of sale probabilities between two bidder counts. When the outcomes, but not necessarily the bids, of failed auctions are known, this yields point identification of the model. However, researchers commonly have access only to transacted auction records; in this setting, the fixed points of a mapping from a candidate sales rate ratio to the model-implied ratio determine identification. We show that the identified set is finite. We also provide a simple sufficient condition that guarantees uniqueness: if the buyer value distribution has mass at points above the maximum seller threshold, so that sufficiently high buyer values always induce trade, then the model is point identified. Inferring seller values from thresholds requires a seller conduct mapping; absent a specific conduct assumption, we derive informative bounds using weak restrictions in the spirit of Haile and Tamer (2003). Monte Carlo experiments validate the approach and quantify biases in the estimated buyer distributions when ignoring rejected offers.

Our main contribution is to bring both sides of the market to the analysis of auction transaction data. A large literature beginning with Guerre et al. (2000) and Athey and Haile (2002) is devoted to non-parametrically identifying buyer primitives in auctions.[4] A singular focus on buyers often requires researchers to make additional assumptions about sellers when conducting counterfactuals. For example, assumptions about seller values have been applied when calculating optimal reserve prices (Paarsch, 1997, Hickman et al., 2017), investigating the effects of entry on auction outcomes (Li and Zheng, 2012), and measuring platform-level effects of auction design (Backus and Lewis, 2025). Our results expand this literature by considering when seller primitives can be identified, rather than assumed, in typical auction data environments.

We use the model to demonstrate the consequences of seller rejections for inferred primitives and counterfactuals; these consequences depend jointly on the data available and the assumptions placed on seller behaviour. Transaction-only records select auctions with relatively strong demand, biasing inferred buyer values upward; previous studies have noted this concern (Roberts, 2013, Gorbenko and Malenko, 2014). Richer auction records can mitigate selection bias, but counterfactual conclusions still depend on how the seller is modelled. We illustrate this distinction through simulated mergers and entry and exit episodes, settings in which auction-based pricing is relevant for policy (Miller and Sheu, 2021). Ignoring seller rejections can substantially overstate buyers' gains from reduced competition because it omits the seller's option to retain the object. Credible welfare and

[4] See Hortaçsu and Perrigne (2021) for a recent survey of empirical auction approaches and applications.

policy analysis therefore requires attention to both sample selection and seller behaviour.

Our identification results begin by considering scenarios where at least some information on failed auctions is observed by the researcher. When the highest bid and sale outcome are always observed, including for failed auctions, the model is identified using standard arguments. However, researchers often lack access to complete records for failed auctions. Our first main result pertains to this case and shows that the model is identified up to a single scalar: the ratio of sale rates across two bidder counts. This result provides point identification when the number of bidders and the sale outcome, but no bidding records, for failed auctions supplement the transaction price and bidder count observations for successful ones.

The more challenging case concerns settings in which the researcher has access only to transaction prices and bidder counts for successful auctions, with no information on failed auctions. This data setting is a frequent focus of the empirical auction literature; indeed, the transaction price is "the most commonly available datum" (Athey and Haile, 2006).[5] We exploit the result that the model is identified up to the sales rate ratio to establish a mapping between a candidate sales rate ratio and a model-implied one. Our main result here is to show that the identified set is (i) indexed by the fixed points of this mapping and (ii) finite. We provide examples of both point- and set-identified cases. The two-point distribution case provides the analytical tractability needed to fully characterise identification across all possible transaction price distributions; we demonstrate that set-identified regions lie close to the border of transaction price distributions that the model can rationalise. As identification in the case of general distributions concerns the fixed points of a scalar function over a bounded region, computing the identified set is simple in practice.

Our next result establishes a sufficient condition for point identification of the buyer's value and seller's threshold distributions when only the transaction prices are observed. The condition, which we refer to as upper-tail support separation between the buyer value and seller threshold distribution, is that that the seller threshold distribution's upper support ends two grid points before the buyer value distribution. Substantively, this condition requires that the highest possible buyer values always induce trade. That the maximum of the seller support is at least one grid point below that of the buyer support is intuitive - why offer your good at an auction if it is guaranteed that no buyer will be willing to pay above

[5]The literature studying identification or application of auction models in transaction-price-only settings is voluminous, beginning with the first structural auction application of Paarsch (1992) and continuing through the present (e.g., Bugni and Wang (2026)).

your threshold? - and our two grid point upper-tail support separation condition is the minimal strengthening of the intuitive condition.[6] Relative to the usual approach that ignores the seller rejection option and so implicitly imposes the condition that the seller threshold distribution always equals one, the upper-tail support separation condition is a near-total, but not quite total, relaxation.

Whether the buyer value and seller threshold have common upper or lower supports is an important factor when determining identification. The upper support separation condition is an important identifying restriction, and the results described above are obtained by assuming that buyer values and seller thresholds share the same lower support bound. Common support assumptions are often used in ascending auction settings with heterogeneous participants (Komarova, 2013). However, a shared lower support bound may be less reasonable in some settings, as it places restrictions on both the threshold distribution and seller values, the underlying primitive. For example, suppose buyer and seller values have the same support. If the threshold is determined by an optimal reserve price, the lower support bound of the threshold distribution will lie strictly above the lower support bound of the value distributions. We examine how identification is affected by relaxing the common lower support bound assumption. We first establish a negative result: without further restrictions, a continuum of buyer value and seller threshold distributions can rationalise transaction data if buyer values are truncated by the threshold's lower support bound. More importantly, no model-relevant object is recoverable: neither the truncated distribution of buyer values nor the seller threshold distribution is identified. However, identification over the transaction price grid is restored by the upper-tail support separation assumption.

Our final results concern seller values and the distribution over the number of bidders. We require an additional assumption to translate the threshold value to the seller value, which we label as a conduct assumption to mirror existing approaches to recovering seller costs from transaction data in other markets. For example, assuming that the seller threshold is set by the optimal reserve corresponds to monopoly, while an assumption of an efficient reserve (i.e., a reserve equal to the seller's value) is analogous to perfect competition. If a specific conduct assumption is imposed, the seller value distribution is identified from the threshold distribution.[7] However, to address cases where the imposition of a specific conduct assumption is overly restrictive, we derive bounds for the seller's

[6] At least in stable conditions: the intuitive condition may not hold if say sellers are not yet aware that buyers' values have shifted down.

[7] At least up to a constraint that follows from use of a grid.

value distribution by using assumptions similar to Haile and Tamer (2003): the seller will never set a threshold value that is below their true value or above the optimal reserve price. These assumptions restrict the set of seller values that can rationalise any threshold value. We conclude our discussion of model identification by noting that the distribution over the number of bidders is recoverable from the empirical distribution of bidders conditional on sale (which over-samples high bidder numbers and under-samples low) and the sales rates for each number of bidders, which are identified from the buyer value and threshold distributions.

Next, we present Monte Carlo simulation evidence on the ability to recover the model's primitives in finite samples. We compare two estimators. A nonparametric estimator follows the recursive construction in the identification argument, while a semi-nonparametric estimator uses Hermite distribution approximations. Both estimators can reasonably recover the model primitives with 1,000 observations. The recursive estimator produces tighter simulation bands when the seller plays a larger role in determining prices and allocations, such as with lower bidder counts, but it is sensitive to sampling error over the regions of the grid where transactions are sparse. The Hermite estimator produces wider bands but covers the true distributions over the reported support. The simulations also show the bias in buyer value estimates that results from ignoring rejected offers.

We conclude with an empirical application of our methodology to housing auctions in the Greater Sydney region of New South Wales, Australia. Auctions are a commonly used mechanism to sell residential property in major Australian cities, with approximately one in five properties sold via auction in the Sydney region. The setting provides an ideal showcase for several reasons: the auction format, an ascending, open-cry (English) auction, is strategically similar to our second-price model; bidders must register in advance, giving direct measurement of the number of bidders; sellers may make a single bid on their own behalf during the auction, which can be used to enforce a threshold price; and auction failures are common, with over 26% of auctions not resulting in a sale. Applying a semi-nonparametric estimator for buyer values and seller thresholds, the transaction-only sample can accurately recover similar distributions to those obtained when using the full sample that includes failed auctions. We also use the estimates to accurately estimate sale rates by bidder count, despite the transaction-only data containing no information about auction failure rates. Estimating the model assuming that all auctions transact generates distributions with a higher mean and lower variance than the full-sample estimates, reinforcing that ignoring rejected offers can substantially bias bidder value estimates.

The rest of the paper is structured as follows. The related literature appears at the end of this section. Section 2 describes the model. Section 3 describes identification, first for the two-point distribution, then the general case, and finally for seller values and the distribution of bidder numbers. Section 4 presents estimation methods and the results of our Monte Carlo simulations. Section 5 contains our empirical application to Syndey housing auctions, and Section 6 concludes.

Related literature. Our approach is closely aligned with the empirical IO tradition that identifies supply primitives from transaction data by combining demand shifters with assumptions about supply-side conduct. In homogeneous good supply-demand settings, exogenous demand variation traces out the supply relation, while a conduct specification maps supply into marginal costs; Rosse (1970) is an early example, and Reiss and Wolak (2007) discuss the development of this methodology. We extend this framework to auction settings: variation in the number of bidders serves as a demand shifter, identifying the distribution of seller thresholds, while a conduct assumption translates thresholds into seller values.[8] This yields the seller-side primitives needed for supply-side counterfactuals, such as reserve rules, platform fees, and participation policies, that are inaccessible in buyer-only analyses.

A small but growing empirical literature considers both buyer and seller primitives in auctions.[9] It is commonly assumed that the seller's value is known to identify or bound other objects of interest; for example, Aradillas-López et al. (2013) bounds seller profit and Gimenes et al. (2025) identifies seller risk aversion, both using this assumption. Existing approaches to identifying distributions of seller primitives alongside buyer values in auctions require rich data on the full set of auction results and on the (possibly secret) reserve prices set by sellers across all auctions, including failed ones (Andreyanov and Caoui, 2022, Marra, 2025, Guan and Xu, 2026). We provide the first identification results for both buyers and sellers that do not rely on full auction-result histories, detailed reserve price information for all auctions, or a requirement that sellers' values be fixed or observed.

Finally, our approach solves the single-bidder problem. In practice, data sets often

[8]The same shifter–conduct logic extends to differentiated-product settings (Berry et al. (1995)), where demand is identified using instruments and substitution patterns and supply first-order conditions recover costs under a conduct assumption.

[9]Related two-sided identification issues arise in bargaining. Larsen and Zhang (2018) develop a revealed preference approach for general quasilinear trading games, and Li and Liu (2026) study a one-buyer, one-seller double auction. These environments feature different mechanisms and more direct observations of seller actions than the multi-bidder ascending auction environment considered here.

contain only transacted sales, so researchers are led to treat all auctions as having transacted. When they are additionally confronted with observations with a single bidder (and no binding reserve), researchers tend to then either (i) discard the single-bidder observations altogether (e.g., Aradillas-López et al. (2013), Kong (2021), Hernández et al. (2020), Haile and Tamer (2003)) or (ii) rationalise those observations by modelling the seller as a "bidder of last resort" with an assumed bid distribution (e.g., Barkley (2021), Li and Zheng (2009), Krasnokutskaya and Seim (2011), Athey et al. (2011), Groeger (2014), Xu (2013), Bajari and Hortaçsu (2003)).[10] Thus the first discards these observations, while the second imposes a rigid constraint on seller thresholds. Neither gives the seller a role in price and allocation determination where there is more than one bidder. Our approach, by incorporating a role for sellers automatically allows for single bidder auctions, and so neither discards data nor imposes special restrictions on the seller threshold distribution.

2. Model

We now introduce a second-price auction model. The model features buyers with symmetric independent private values competing against each other and also against the seller, who requires a minimum price from the buyers to proceed with the sale. If the seller's required price, which we refer to as the threshold value, is not met, then no sale occurs. We start by stating the independent private values assumption. On the buyers' side, this assumption takes the standard form and is commonly used in empirical work. We also place a similar independence restriction on the seller's value, assuming the seller's threshold to be independent of the buyers' values. This assumption is consistent with optimal reserve prices or "naive" strategies requiring the highest bid to clear the seller's value.

Assumption 1. *Bidders have independent private values drawn from a shared distribution with cdf F. Seller threshold values are independent of buyers' values and are drawn from a distribution with cdf G.*

Let $N \geq 1$ denote the number of bidders in the auction. The model must describe allocations (i.e., whether the object is sold or retained by the seller) and the price at which a transaction occurs. Instead of deriving the allocation and transfer rules as an equilibrium outcome of the auction, we assume the process by which this occurs.[11] We do this to avoid

[10] For example, the latter approach will often assume that sellers have the same value distribution as buyers, and that they participate in the auction as a bidder if and only if a single buyer shows up.

[11] Aradillas-López et al. (2013) and Haile et al. (2003) follow a similar approach to modeling auction outcomes.

placing restrictions on how the seller's threshold value is determined, a matter which we return to in Section 3.5. Assumption 2 states when sales occur and at what price.

Assumption 2. *The auction results in a sale only if the highest buyer value is greater than or equal to the seller's threshold value. If a sale occurs, the transaction price is the larger of the threshold value and the second-highest buyer value.*

In an auction with N buyers, we let the vector $\mathbf{V}_B \equiv (V_1, V_2, ..., V_N)$ represent the vector of buyer valuations and V_T be the seller's threshold value. The order statistics of buyer values are denoted $V_{N:N} \geq V_{N-1:N} \geq ... \geq V_{1:N}$. The outcome of the auction is represented by $S \in \{0, 1\}$, where $S \equiv \mathbf{1}\{V_{N:N} \geq V_T\}$, and the transaction price in such cases is $P \equiv \max\{V_{N-1:N}, V_T\}$. The main restriction implied by Assumption 2 is to rule out post-auction bargaining except for seller take-it-or-leave-it offers.

Assumption 2 permits an equilibrium interpretation in commonly-modelled settings in which the seller's threshold value is a binding optimal reserve price. In second-price sealed bid auctions, it is the equilibrium outcome resulting from buyers' unique dominant bidding strategy of bidding their private value. It is also the equilibrium outcome for button auctions, the dynamic auction format strategically equivalent to second-price sealed bid auctions in which the price rises exogenously and bidders drop out over time. The optimal auction mechanism of Bulow and Klemperer (1996), where a knockout auction identifies the highest-value buyer and the seller issues an optimal take-it-or-leave-it offer, is also captured by this setting. Finally, it can also provide a reasonable approximation to ascending open-outcry (English) auctions, provided that bidding is frictionless and "jump bids" are not used to determine the transaction price.

The symmetric IPV framework for buyers established in Assumption 1 implies that buyer values do not depend on N, as all buyers draw their value from F regardless of how many competitors there are. However, because of this feature's importance in the identification arguments below, and to make clear that buyers do not receive informative signals about their value before deciding to participate in the auction, we make this assumption explicit now:

Assumption 3. *Buyer values and the seller threshold value are independent of N.*

This assumption on auction models has been used extensively in the literature, being adopted in Guerre et al. (2000), Haile et al. (2003), Aradillas-López et al. (2013), and Hernández et al. (2020), among others. It is consistent with entry procedures that follow Levin and

Smith (1994), in which bidders pay an entry cost before learning their private valuation.[12] More generally, it holds whenever entry signals and valuations are independent across potential bidders (Grundl and Zhu, 2024).

Finally, we also assume that bidder values and the seller threshold value are drawn from a discrete set of possible values.[13] This assumption provides analytical tractability. The support grid can be arbitrarily fine, providing flexibility when implemented in practice. We show in Section 4 that the empirical results of our estimator are not sensitive to this assumption, as we can recover primitives with high accuracy even if the underlying data-generating process uses continuous distributions.

Assumption 4. *Buyer values and seller thresholds are drawn from a finite set of J possible values $\{x_1, x_2, ..., x_J\}$.*

To shorten notation throughout, we will use subscripts on distributions equal to the index on the grid to indicate the cumulative distribution value at the corresponding point on the grid, so, for example, $F_j \equiv F(x_j)$. Where needed, superscripts in parentheses will refer to the number of bidders. Under Assumption 4 we can express the transaction price distribution at x_j for a given number of bidders N, $H_j^{(N)} = \Pr(\max\{V_{N-1:N}, V_T\} \leq x_j \mid V_{N:N} \geq V_T)$, as

$$H_j^{(N)} = \frac{1}{\Omega_N}\left[NF_j^{N-1}(1-F_j)G_j + \sum_{k=1}^{j} G_k[F_k^N - F_{k-1}^N]\right] \tag{1}$$

where

$$\Omega_N \equiv \sum_{k=1}^{J} G_k[F_k^N - F_{k-1}^N]$$

is the overall sales rate for N bidders, and where for notational convenience we set $V_{0:N} \equiv 0$ and $F_0 \equiv 0$. The first term in the brackets of Equation (1) is the probability of exactly one bidder having a value that exceeds x_j while all other bidders and the seller threshold value are at or below x_j, and the second term in brackets is the probability that all bidders and the seller threshold value are at or below x_j but the highest bidder value is at least as large as the seller threshold value. Together, these terms sum to the joint probability of having a

[12] Li and Zheng (2009) find empirical support for this entry model in their procurement auction setting, and Marmer et al. (2013) develop empirical tests for auction entry models.

[13] Discrete values or bid increments have also been used in Hickman et al. (2017), Haile and Tamer (2003), and Barkley et al. (2021), among others.

sale and having a transaction price less than or equal to x_j.[14]

2.1. Model discussion

The model applies directly to second-price sealed-bid auctions and, under the usual assumption of frictionless bidding, to ascending auctions. The model is designed to accommodate three specific features – one-bidder auctions, ambiguous reserve policies, and failed auctions – that existing approaches often struggle to rationalise. Our framework is suited to settings in which bidders do not observe the seller's threshold before bidding and the seller may reject the highest bid, reveal a reserve, or make a take-it-or-leave-it offer after bidding ends. Many real-world auction platforms use such a mechanism; examples include auctions for automobiles, art and collectables, livestock and farm machinery, and real estate.[15] It may also apply to settings without an institutionally defined mechanism that are often modelled as competitive bidding processes, such as takeover auctions (Gorbenko and Malenko, 2014).

Procurement auctions provide a second broad application, with the buyer and seller labels reversed. Open-outcry descending auctions are used in many countries, including Brazil, India, and Russia. An analysis of US Department of Defense descending procurement auctions indicates the empirical relevance of the features our model accommodates: one-bidder auctions are common, the stated reserve price frequently does not bind, and the lowest bidder fails to receive the contract in about one quarter of cases (U.S. Government Accountability Office, 2013). Procurement also illustrates variation in the data available to researchers; variation across South America, where descending auctions are frequently used for procurement, provides examples. Peru reports transaction prices and bidder counts for successful auctions, while Brazil provides full bid histories for federal procurement (Ferraz et al., 2015) but may report only transaction prices at the state level (Dahis et al., 2026). Competition policy often dictates data availability, as evidenced by Colombia's decision to restrict public bid data to deter coordination (Bernasconi et al., 2025). The availability of failed-auction outcomes and losing bids therefore differs both

[14]Note that when sellers accept all offers, $G_1 = ... = G_J = 1$ and equation (1) reduces to, $H_j^{(N)} = NF_j^{N-1}(1 - F_j) + F_j^N$, which is the typical expression relating transaction prices to model primitives in second-price auctions.

[15]For example, eBay Motors indicates to bidders whether a secret reserve has been met; Sotheby's and Christie's allow auctioneers to bid on behalf of the seller up to a pre-specified reserve; and the AuctionsPlus livestock and farm machinery platform may reveal an unmet reserve to the highest bidder as a take-it-or-leave-it offer.

across and within jurisdictions, motivating identification results for several data environments.

2.2. Implications for surplus and counterfactual policies

We use population-level simulations to show that ignoring seller rejections can substantially overstate the effects of changes in bidder competition on buyer surplus and transaction prices. We consider bidder entry and exit, which are common policy counterfactuals in procurement settings where participation is often limited (Kang and Miller, 2022), and mergers between bidders. The latter are relevant to merger reviews of business-to-business markets in which buyers procure through auctions or request-for-proposal processes.[16]

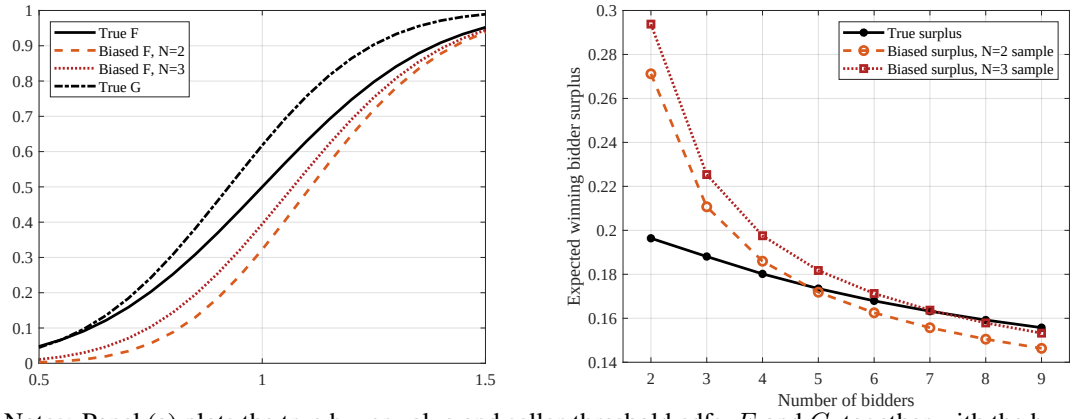
To demonstrate the impact of unobserved rejected offers on competition variation and auction outcomes, we provide a simulation using normal distributions for F and G, as shown in Panel (a) of Figure 1. These distributions generate a sale probability with three bidders of 85%. We also plot the biased distributions that would be inferred for F when rejected offers are ignored, with the implied distribution using $N = 2$ displayed as the orange dashed line and $N = 3$ as the red dotted line. These are generated using the mapping $H_N(x_j) = NF_j^{N-1} - (N-1)F_j^N$ that applies when $G_1 = 1$ and is the standard identifying representation for F when rejected offers are not considered. Both of these first-order stochastically dominate the true distribution F, with the extent of the bias decreasing as the number of bidders increases. This intuitive result follows from the declining probability that the seller either (i) rejects an offer or (ii) is influential in determining the price via the threshold value as the number of bidders increases.

Panel (b) reports expected winning-bidder surplus per auction, assigning zero surplus when no sale occurs. The solid black line uses the true distributions F and G. The dashed and dotted lines impose $G_1 = 1$, so that every offer is accepted, and use the buyer distributions inferred from the $N = 2$ and $N = 3$ transaction-price distributions under this misspecified model. At a bidder count of $N = 2$, these models overstate buyer surplus by 38% and 50%, respectively. The larger error from the $N = 3$ estimate arises despite that estimate being closer to the true F. Although it is less upward-biased and more dispersed than the $N = 2$ estimate, omitting the seller's threshold removes its role in determining

[16]For studies of entry in auctions, see (Li and Zheng, 2009, 2012, Gentry and Li, 2014, Marra, 2025, Xu, 2013). Auction-based merger simulations have been used in several matters before the US Department of Justice and Federal Trade Commission, including *Sysco/US Foods* (2015), *Anthem/Cigna* (2016), and *Wilhelmsen/Drew Marine* (2018); Miller and Sheu (2021) provide further discussion.

Figure 1: Bias in inferred buyer value and surplus

(a) Distributions **(b)** Expected winning bidder surplus



Notes: Panel (a) plots the true buyer-value and seller-threshold cdfs, F and G, together with the buyer-value cdfs inferred from the $N = 2$ and $N = 3$ transaction-price distributions under the misspecified assumption that every auction results in a sale. Panel (b) plots expected winning-bidder surplus per auction by bidder count under the true model and misspecified models.

transaction prices and eliminates the possibility of no trade. When competition falls, the second-highest buyer value falls, but the threshold continues to support the price and prevent low-value trades in the true model. The misspecified model therefore converts too much of the reduction in competition into buyer surplus. At higher bidder counts, the threshold is rarely pivotal and the greater dispersion of the true F becomes dominant, so the curves eventually cross. Thus, reducing bias in the estimated buyer distribution need not improve counterfactual predictions if the seller remains omitted from the model.

Table 1 reports the effects of bidder exit, entry, and merger on expected buyer surplus per auction and expected transaction prices conditional on sale. The correctly specified counterfactuals use the true distributions F and G. The misspecified counterfactuals impose $G_1 = 1$ and use the buyer distribution inferred from the initial market structure: the $N = 3$ distribution for exit and the $N = 2$ distribution for entry. The misspecified model substantially overstates the effects of both changes in competition. When a bidder exits, the resulting decline in the second-highest buyer value tends to reduce the transaction price and increase buyer surplus. In the true model, however, the realised seller threshold may continue to determine the payment or prevent trade altogether. The misspecified model eliminates both channels and therefore attributes too much of the reduction in competition to buyer surplus. Panel B shows that entry generates the corresponding effect in the opposite direction.

Table 1: Counterfactual effects of exit, entry, and a merger

Panel A: Exit ($N: 3 \to 2$)	True F and G	Biased F ($N = 3$), $G = 1$
Pct. change in buyer surplus	4.42%	30.36%
Pct. change in transaction price	-8.43%	-13.09%
Panel B: Entry ($N: 2 \to 3$)	True F and G	Biased F ($N = 2$), $G = 1$
Pct. change in buyer surplus	-4.23%	-22.30%
Pct. change in transaction price	9.20%	12.98%
Panel C: Merger	True F and G	Biased F ($N = 3$), $G = 1$
Pct. change in merging firms' surplus	28.31%	47.77%
Pct. change in transaction price	-3.64%	-6.39%

Notes: Panel A shows the changes in winning bidder surplus and transaction prices when a bidder exits, reducing competition from three bidders to two under the true model and the biased model when F is inferred from the three-bidder sample and G is one. Panel B shows the same outcomes for bidder entry, increasing competition from two bidders to three under the true model and the biased model when F is inferred from the two-bidder sample and G is one. Panel C shows the predicted effects of a merger taking the market from three firms to two, with the merged firm bidding with the higher of the two private value draws of the merging entities; the biased F is inferred from the $N = 3$ sample.

Panel C considers a merger between two of three bidders. Before the merger, all three bidders draw independently from F. After the merger, the combined firm retains two independent draws and values the object at their maximum, while the non-merging bidder continues to draw once from F. The post-merger auction therefore has two asymmetric bidders, and the merging firms' pre-merger surplus is defined as their combined expected surplus. We hold the threshold distribution fixed after the merger, making this a partial-equilibrium counterfactual that isolates the effects of changed bidder competition. Under this specification, the misspecified model overstates both the merged firm's surplus gain and the reduction in transaction prices because it omits the threshold's continuing role in determining payments and trade.

3. Identification

We now discuss when a dataset generated in a manner consistent with Assumptions 1-4 can be utilised to identify the primitives F and G. We consider several data and model environments which vary in what the researcher observes about failed auctions (if anything), whether F and G share the same lower support, and assumptions mapping seller values to thresholds. Our focus is on cases where transaction price distributions are observed for two different realisations of the number of bidders $M > N \geq 1$, possibly supplemented with

additional information on failed auctions, and our goal is to characterise when this variation identifies the model primitives F and G. Proofs of all results are contained in Appendix A unless noted otherwise.

We begin by discussing the case when some information about failed auctions is observed, where we show that the ratio of sales rates across two bidder counts determines identification. Next we consider data consisting of only transacted sales; here, identification is determined by fixed points of a scalar mapping. The two point distribution, when both buyer and seller threshold values can take either a low or high value, is a useful case to demonstrate how point- and set-identified cases arise. We then consider the general case of an arbitrary, but finite, number of support points, providing a sufficient condition under which the distributions F and G can be uniquely identified and discussing approaches to set-identify F and G when this condition does not hold. We conclude by discussing identification of seller values from the threshold distribution and the distribution over the number of bidders when failed auctions are unobserved.

We first make several observations about how the identification of F and G relates to the discrete grid $\{x_1, ..., x_J\}$. First, the minimum and maximum points in the support of F are identified from observed transaction price distributions.[17] Second, absent additional restrictions, we cannot identify the maximum point in the support of G, as realisations for V_T that exceed the maximum bidder value never generate transactions, and hence the data contain no information about this case. Accordingly, we assume that $G(x_J) = 1$, where x_J is the upper limit of the buyer value support. This corresponds to the assumption that a seller will only auction an item if there is a non-zero probability of sale; alternatively, the identified threshold distribution is that of the distribution including the super-seller thresholds as well, truncated from below at the maximum bidder value.[18] Lastly, the minimum of the support of G is identified for the general case only if the transaction price distribution for $N = 1$ is observed. For simplicity, we assume that the lowest support point of G is the same as for F, which we denote x_1.

[17] We can identify the lowest point in the support of F as the lowest value x for which $H^{(N)}(x) \neq 0$ for any N, and the highest point as the lowest value for x such that $H^{(N)}(x) = 1$ for all N.

[18] While there are several possible explanations for why an auction might be held even when the seller's value exceeds x_J, such as inexperienced sellers who are learning about the buyer value distribution, or a recent shock to buyer values, information on which has not yet disseminated to all sellers, identifying the prevalence of these sellers requires outside information on the total proportion of failed auctions, which we assume is unavailable to the econometrician.

3.1. Identification with some failed auction results

Full result histories

We begin by considering identification when the researcher has access to data on failed auctions. The first case assumes that full auction results containing the highest bid and number of bidders for all auctions are known. In this case, identification follows immediately from standard arguments for asymmetric IPV auctions, even absent variation in bidder counts, as formalised below.

Proposition 1 (Theorem 2 (a) of Athey and Haile (2002)). *If bidder counts and auction outcomes are observed for all auctions along with the transaction price, for successful auctions, and the second-highest value among $\{V_1, ..., V_N, V_T\}$ for failed auctions, then F and G are identified.*

Theorem 2 (a) of Athey and Haile (2002) establishes identification for asymmetric IPV ascending auctions when the transaction price and the identity of the winner are known. This is identical to the environment in which the highest bid is observed for all auctions along with the auction outcome, as the winner is whichever individual is allocated the object, which may be the seller when no buyer value exceeds V_T.

Transaction prices and auction outcomes

In other settings, data on transacted sales is readily available, while information on failed auctions is absent or incomplete. For example, auction houses may indicate when lots failed to sell without providing any bid details.[19]

The system of equations generated by evaluating equation (1) for distinct M and N generates a system of $2J-2$ independent equations, one for each $H_j^{(n)}$ for $j = 1, ..., J-1$ and $n \in \{N, M\}$, and $2J-2$ unknowns, the set of $\{F_j, G_j\}_{j=1}^{J-1}$ (recalling that we have set $G_J = 1$). While this establishes a necessary condition for identification, the nonlinearity of the system makes multiple solutions a concern. Our first main result shows that a unique solution is guaranteed when information about how sales rates change with the number of bidders is known.

[19] This also arises in procurement, albeit primarily in first-price settings, where auction results from sources like the Tenders Electronic Daily platform for the EU report tenders that failed to transact but do not report any bids on these contracts.

Theorem 1. *Suppose transaction price distributions $H^{(N)}$ and $H^{(M)}$ are observed for bidder counts $M > N \geq 1$. If the sales-rate ratio*

$$\rho \equiv \frac{\Omega_M}{\Omega_N}$$

is known, then F, G, Ω_N, and Ω_M are identified.

The proof uses an induction argument in three parts. The first is to show that taking the ratio of equation (1) for M and N generates an M-degree polynomial in F_j, taking as given $\{F_k, G_k\}_{k=1}^{j-1}$, for each $j = 1, ..., J$. We then use Descartes' rule of signs to bound the number of positive real roots at or below two, and use the sign change in the polynomial evaluated at F_{j-1} compared to one to establish that the number of real roots in this interval must be odd. Therefore, there is a unique solution for F_j.

An immediate consequence of this result is that if records of a set of auctions is available, with each auction record including both the number of bidders and the final result of the auction, the price, in the case of a sale, or a "Not Sold" result in the case that the auction is unsuccessful, then the distributions of the buyer and seller threshold values are identified.

Corollary 1. *If transaction prices from successful auctions and the overall sale rate are observed for two distinct numbers of bidders, then F and G are identified.*

We adopt the following notation based on Theorem 1. Let $F_k(\rho)$, $G_k(\rho)$ denote the F and G recovered by the theorem given a sales rate ratio ρ, with $\Omega_N(\rho)$ and $\Omega_M(\rho)$ the associated sales rates. The proof of Theorem 1 uses a recovered threshold distribution scaled by the N-bidder sales rate, which we denote by $\tilde{G}_k(\rho) \equiv G_k(\rho)/\Omega_N(\rho)$. The notation for the probability mass functions $f_k(\rho)$, $g_k(\rho)$, and $\tilde{g}_k(\rho)$ follows.

3.2. Identification with no failed auction results

Auction records often contain little or no information about failed auctions. We now consider this case. We assume that the econometrician observes transaction price distributions $H^{(n)}$, which are linked to the model primitives via equation (1), and we consider whether variation in the number of bidders across two distinct values results in point identification of F and G.[20]

[20]Without variation in n, the model is not identified. This can be seen by noting that when sellers accept all offers (i.e, $G_k = 1$ for all k), equation (1) reduces to $H_N(x_j) = NF_j^{N-1} - (N-1)F_j^N$, which is the usual expression for second-price auctions (as in Athey and Haile (2002), for example). This system is just identified; adding another model primitive G generates an under-identified system.

Our next main result shows that reducing identification to the sales rate ratio across two bidder counts allows the fixed points of a mapping from a candidate ratio to the model-implied ratio to characterise identification. Since Theorem 1 shows that any given ratio ρ can be consistent with at most one buyer value and seller threshold distribution pair, the sales rate ratio implied by these recovered distributions must match the original ratio for them to be consistent with the data. The set of identified F and G are indexed by the fixed points of this mapping over the set of admissible sales rate ratios.[21] We also show that the identified set is finite. We formalise these findings in the theorem below.

Theorem 2. *For bidder counts $M > N \geq 1$, suppose that the set of admissible ratios is the compact interval $\mathcal{R} = [1, \overline{\rho}]$, where $\overline{\rho} > 1$. Then*

(i) The sales rate ratio implied by $F(\rho), \tilde{g}(\rho)$ is given by

$$\Upsilon(\rho) \equiv 1 + \sum_{k=1}^{J-1} \left[F_k(\rho)^N - F_k(\rho)^M \right] \tilde{g}_{k+1}(\rho). \tag{2}$$

(ii) The identified set is indexed by $\mathbf{\Omega}^* = \{\rho \in \mathcal{R} : \Upsilon(\rho) = \rho\}$.

(iii) $\mathbf{\Omega}^*$ is finite.

Theorem 2 requires the candidate distributions to remain admissible over a single compact interval of sales rate ratios.[22] Admissibility also requires positive threshold mass above x_1, so that some seller thresholds can prevent trade. A fixed point of (2) equates the model-implied sales rate ratio with the ratio used to generate the candidate distributions and therefore constitutes an internally consistent set of primitives that generate the observed transaction price distributions. If only a single such fixed point exists, the model is point identified. If multiple fixed points exist, then there is set identification.

Theorem 2 establishes that there are at most finitely many distributions F and G that are consistent with a given pair of transaction price distributions $H^{(N)}, H^{(M)}$. However, it does not guarantee a unique fixed point, and the possibility of a non-singleton identified set remains. We now investigate when multiple fixed points are present. We illustrate how

[21] We call a candidate ratio ρ *admissible* if the recursive construction in Theorem 1, together with the normalisation $G_J = 1$, yields well-defined normalising constants in $(0, 1]$ and distributions satisfying $0 = \tilde{F}_0(\rho) < \tilde{F}_1(\rho) < \cdots < \tilde{F}_J(\rho) = 1$, $0 = G_0(\rho) < G_1(\rho) \leq \cdots \leq G_J(\rho) = 1$ and $g_j(\rho) \equiv G_j(\rho) - G_{j-1}(\rho) > 0$ for some $j \geq 2$.

[22] For ρ sufficiently large, no admissible solution for F exists. We show in Appendix C that the upper bound for ρ that could possibly rationalise the data is M/N.

multiplicity arises using the two-point distribution, where the analytical tractability of this simple setting permits a complete characterisation of the identified set. Further examples from distributions with more support points extend this intuition to more general settings.

The two-point distribution

To demonstrate that multiplicity can arise and characterize it exactly, consider buyer values and seller thresholds with support $\{x_1, x_2\}$. We use the two-point distribution to demonstrate that multiplicity can arise and to characterize exactly when it does for adjacent bidder counts. Let buyer values and seller thresholds have support x_1, x_2. Since $F(x_2) = G(x_2) = 1$, the primitives are summarized by F_1 and G_1. For each bidder count N, the researcher observes $H_1^{(N)}$, the fraction of successful auctions transacting at x_1, but does not observe the sale probability Ω_N.

$$\Omega_N = F_1^N G_1 + [1 - F_1^N] \tag{3}$$

$$[NF_1^{N-1} - (N-1)F_1^N]G_1 = \Omega_N H_1^{(N)} \tag{4}$$

For two bidder numbers M and N, with data $\{H_1^{(N)}, H_1^{(M)}\}$, the equations above describe a polynomial system in F_1 and G_1. Identification depends on whether this system has one or two admissible solutions $(F_1, G_1) \in (0,1)^2$. [23] The following proposition, proved in Appendix A, describes the point- and set-identified regions with adjacent bidder counts $N, N+1$.

Proposition 2. *Suppose $H_1^{(N+1)}$ and $H_1^{(N)}$ lie in the interior of the set of observables rationalizable by the two-point model for some $N \geq 1$. Then the two-point distribution is point-identified if and only if*

$$H_1^{(N+1)} \leq \frac{N H_1^{(N)}}{N + 1 - H_1^{(N)}}$$

and is set-identified with exactly two admissible solutions for F and G otherwise. Both regions have non-empty interiors.

The set-identified region arises when adding a bidder causes only a small decline in the conditional probability of a low transaction price. As failed auctions are unobserved, this

[23] For $N = 1$ and $M = 2$, eliminating G_1 produces a quadratic in F_1. For example, if $H_1^{(1)} = 1/2$, any $H_1^{(2)} \in (1/3, 6 - 4\sqrt{2})$ produces two admissible roots and therefore two primitive pairs.

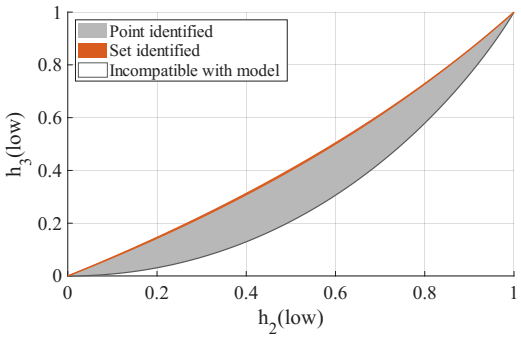
weak response can be rationalized by two different combinations of buyer values and seller thresholds. The data therefore do not reveal which combination generated the observed transactions.

Figure 2 displays the set-identified region of the two point distribution for the $M = 3$, $N = 2$ case, with Panel (a) in the space of the data $(H_1^{(2)}, H_1^{(3)})$, and Panel (b) in the space of the model primitives (F_1, G_1). In Panel (a) of Figure 2, the orange region is the set of data generated by two different pairs of distributions. The grey area below it represents data generated by a unique F and G. The remaining white areas cover data that are inconsistent with the model. For example, the region above the 45-degree line represents transaction price distributions that are inconsistent with any pair of distributions, as this region requires a higher share of transactions occurring at the low price when the number of bidders is larger.

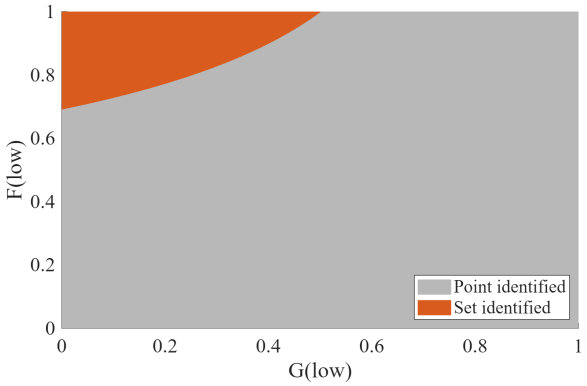
Panel (b) shows that set identification is associated with large F_1 and small G_1: buyers frequently have the low value, while seller thresholds frequently take the high value. These primitives generate low sale probabilities and substantial selection. Although the set-identified region is narrow in observable space, a broader range of primitive configurations generates data in that region.

Figure 2: Identification for the two-point distribution, $M = 3$, $N = 2$

(a) Transaction prices



(b) Distributions



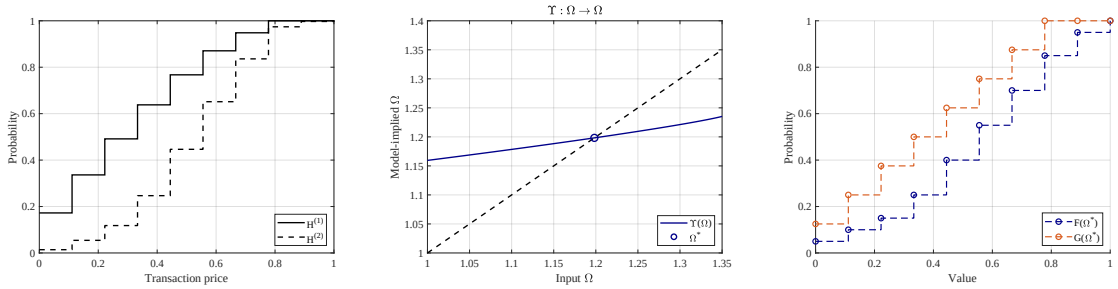
These features also appear in our more general support examples: set identification occurs near the boundary of model-consistent transaction-price distributions, where prices respond weakly to bidder counts and seller thresholds are high relative to buyer values.

3.3. The general case

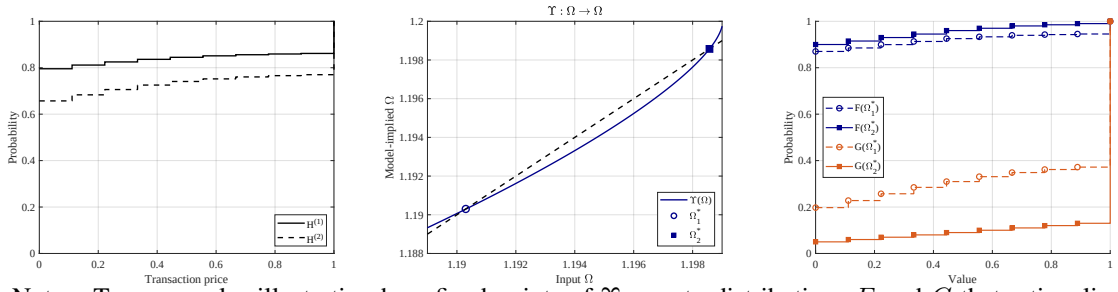
As the mapping defined by equation (2) corresponds to a fixed-point problem in a single variable and calculation of $\tilde{F}$ and $\tilde{g}$ is straightforward, given ρ, in practice it is often possible to fully characterise all fixed points. We show two examples in Figure 3. Each case uses distributions for F and G with support on 10 evenly spaced points between zero and one. The left-most panel shows the transaction price distributions associated with $M = 2$ and $N = 1$, $H^{(2)}$ and $H^{(1)}$, the middle panel shows the Υ mapping generated by these transaction price distributions, and the right-most panel shows all F and G associated with fixed points of Υ. Panels (a-c) show transaction prices corresponding to a single fixed point, and so point identification of F and G. This setting captures the qualitative features associated with identification established for the two-point distribution, in that G does not stochastically dominate F.

Figure 3: Identification examples

(a) Transaction prices **(b)** Unique fixed point **(c)** Point-identified F and G



(d) Transaction prices **(e)** Two fixed points **(f)** Set-identified F and G



Notes: Two examples illustrating how fixed points of Υ map to distributions F and G that rationalise observed transaction price distributions. Panels (a-c) show an example with a single fixed point of Υ and point identification. Panels (d-f) show an alternative with two fixed points and set identification.

In contrast, Panels (d-f) show a set-identified example. The transaction price distributions in Panel (d) show flat distributions for $H^{(2)}$ and $H^{(1)}$. This is qualitatively similar to the setting of the two-point distribution in Figure 2, where transaction price distributions

associated with set-identification of F and G were nearer to the 45-degree line than the point-identified region, implying similar values for the two transaction price distributions. The Υ mapping in this case generates two fixed points, as shown in Panel (e). Note that the x-axis reveals these two points to be very near each other in the $\Omega_{M/N}$ space, and that Υ is convex. This is again consistent with the observations from the two-point case in that set-identification is associated with data that are near the boundary of data sets that can be rationlised by the model: a small change in the transaction price distributions of Panel (d) would generate an Υ mapping that lay completely above the 45-degree line, implying no model solution able to rationalise the data. The two fixed points in Panel (e) correspond to the distributions of Panel (f). For each distribution pair, seller threshold values are much larger on average than buyer values.

Sufficient condition for point identification

Our next main result establishes a sufficient condition for point identification of F and G. This condition is based on the mapping from sales rate ratios to model primitives established in Theorem 1. The sufficient condition imposes a simple upper-tail separation condition on the supports of the two primitives. A sufficient condition for point identification of (F, G) from the observed data is that there exists $\ell \geq 2$ such that $G(x_{J-\ell}) = 1$ and $F(x_{J-\ell}) < 1$, i.e., G places no mass on $\{x_{J-\ell+1}, \ldots, x_J\}$ while F has not yet reached one by $x_{J-\ell}$.[24] This support separation delivers a simplified structure in the mapping from (F, G) to observables: outcomes at the top ℓ grid points can only be attributed to F and Ω_M, Ω_N, allowing one to recover the sales rate ratio using only prices at the top of the distribution, and then use this sales rate ratio to uniquely recover F and G at all lower points.

Assumption 5. *The top two buyer grid points contain positive mass, $F_{J-2} < F_{J-1} < 1$, and the top two threshold grid points contain zero mass, $G_{J-2} = 1$.*

The condition achieves identification without additional shape restrictions on F or G, relying instead on exclusion of G from the upper tail. The intuition is that the condition removes the dependence of Ω_n on the distribution of buyer value mass in the top two grid points. At every other point up to $J - 2$, the buyer value mass partially determines Ω_n. However, Assumption 5 implies that this is not true for f_{J-1} and f_J given F_{J-2}: how the

[24] Equivalently, $\text{supp}(G) \subseteq \{x_1, \ldots, x_{J-\ell}\}$ and F assigns positive probability to at least one of the top ℓ grid points.

total buyer value mass above F_{J-2} is distributed across the top two grid points does not affect Ω_n, as a sale is already assured for these buyer values, but it does have implications for the transaction price mass at these points. Variation in n then allows this to be expressed as a system of two nonlinear equations in F_{J-1} and F_{J-2}, and we show that this system has a unique solution. The result is given below.

Theorem 3. *Suppose $H^{(N)}$ and $H^{(M)}$ are observed for bidder counts $M > N \geq 2$ and Assumption 5 holds. Then Ω_N, Ω_M, F, and G are identified.*

We view Assumption 5 as a mild condition: in freeing up the seller distribution from the usual imposed constraint that sellers sell no matter what the maximum realised bidder value is, we require only that sellers sell if the latter is sufficiently high.[25] That said, it is not innocuous: Figure 3 (f) is an example where the condition fails, and its stringency depends on J. When the grid is fine and J is large, excluding G from the top two cells is relatively weak; conversely, when $J = 4$, for example, it rules out mass for G on $\{x_2, x_3\}$ and is correspondingly strong. The key is that some upper-tail variation in support – G reaching one strictly earlier than F by at least two points – is sufficient to isolate how variation in transaction prices in this upper range is separately determined by F_{J-2}, F_{J-1} and $\Omega^{(M)}, \Omega^{(N)}$. Then, given $\Omega^{(M)}$ and $\Omega^{(N)}$, the distributions at the remaining points are identified from Theorem 1. [26]

3.4. Lower support mismatch

The results above impose a common lower support assumption on F and G – namely, that $F_0 = G_0 = 0$, $F_1 > 0$, and $G_1 > 0$. This insures that the transaction price distribution has positive support over the entire set of possible buyer values and assumes that any bidder, even one with the lowest possible value, has a chance to win the auction. Similar common support assumptions have been employed elsewhere when establishing identification in auctions with heterogeneous participants (Athey and Haile, 2002, Komarova, 2013). However, in our model, some settings may be inconsistent with this assumption. Restrictions on the support of G affect both the threshold distribution itself and the distribution of seller values that can rationalise the threshold. For example, if seller values have the same distribution of buyer values, but the threshold is determined by an optimal reserve

[25] This condition is also substantially less restrictive than the $G(x_1) = 1$ assumption required to estimate the model assuming that all auctions transact.

[26] The lowest-dimension case permits a closed-form solution; Appendix D contains the derivation.

price, then the lower support of the threshold distribution will exceed that of the buyer value distribution. We examine how identification is affected by this change to the model environment.

Our first result is negative: if there is positive buyer value mass below the lower threshold support, a continuum of solutions exists for F and G. More importantly, even the truncation of F to the region where the seller threshold has positive support is not identified. This result contrasts with previous identification results focusing on fixed reserve prices (Guerre et al., 2000) in which the buyer value distribution truncated to the region above the reserve is identified, and it demonstrates that seller bid rejections interfere with typical identification intuition in such settings. We formalise this result below.

Proposition 3. *The class of models satisfying $F_0 > 0$, $G_0 = 0$, and $G_1 > 0$ is not globally point identified. In particular, a continuum of pairs (F, G) can generate the same observed transaction price distributions. The conditional buyer-value distribution*

$$F_j^+ = \frac{F_j - F_0}{1 - F_0}, \qquad j \geq 1,$$

also need not be identified.

We establish the result using a counterexample with buyer values on $0, 1, 2$, seller thresholds on $1, 2$, and transaction price distributions for $N = 1$ and $M = 2$. Buyer mass below x_1 affects the bidder-count-specific sales rates, which are central to identification, but is not observed directly because transaction prices are recorded only conditional on sale. Although this mass can affect conditional prices indirectly through the composition of bidder values, changes in it can be offset by changes in the buyer distribution above x_1 and in G. The counterexample constructs a continuum of such primitives that generate the same two transaction price distributions.

While lower support mismatch can result in non-identification, support mismatch at the upper end of the distribution restores it. The upper-tail support separation condition of Assumption 5 permits the buyer value and seller threshold distributions to be identified over the transaction price grid. The intuition is that identification is achieved via a top-down approach, where first the sales rates are identified from the top grid points and then an iterative procedure similar to the construction of $\tilde{F}$ and $\tilde{G}$ contained in the proof of Theorem 1 sequentially identifies F and G over the transaction price grid.

Proposition 4. *Suppose transaction prices for bidder counts $M > N \geq 2$ are known and that Assumption 5 holds. Let $x_0 < x_1$ be a lower boundary with $H_0^{(n)} = 0$ for all observed bidder counts n.*

(i) If $G_0 = 0$, $F_0 > 0$, and $G_1 > 0$, then the aggregate buyer mass below the first observed transaction price grid point is identified. The distributional shape of buyer values below x_1 is not identified from transaction prices alone.

(ii) If $F_0 = 0$, $G_0 > 0$, and only bidder counts $n \geq 2$ are observed, then transaction prices identify G_1, but not the decomposition of seller threshold mass below and at the first buyer-support point.

(iii) If $N = 1$ transaction prices are observed, then seller-threshold increments below the buyer support are identified wherever $F_{j-1} < 1$.

Propositions 3 and 4 demonstrate the importance of the relative supports of F and G when establishing identification. At the lower support, positive buyer value mass below the lowest seller threshold is problematic because it is directly relevant for the sales rates at each bidder count, and the sales rates are a key component of the identification approach. The upper support separation condition reverses this, because the overall sales rates cannot be affected by the distribution of the buyer value mass above the point where the seller threshold equals one.

3.5. Seller values

We have thus far not specified how the seller threshold relates to seller values. As discussed previously, an additional assumption is required to relate the seller threshold to the value underlying it; we refer to this assumption as a conduct assumption. Depending on the theoretical context, different assumptions about this mapping may be reasonable. Assuming that the threshold is set by an optimal (static) reserve price may be sensible in settings where the seller can make a binding take-it-or-leave-it offer to the highest bidder at the end of the auction (Bulow and Klemperer, 1996). In marketplace settings with competing sellers, the equilibrium mechanism may be for the seller to have a reserve price equal to his or her own value (Albrecht et al., 2014). Empirically, some settings suggest that sellers may set the reserve price sub-optimally (Paarsch, 1997).

Assumption 6 imposes two restrictions on seller thresholds that are consistent with many interpretations of the value-threshold mapping. These restrictions, which are similar to those used in Haile and Tamer (2003), require that a seller's threshold should never be below her value, as this could generate a sale when she would prefer to retain the item, and that the threshold should never exceed the optimal reserve price. Note that this formulation rules out reserve prices that exceed the optimal reserve price. While this is not strictly necessary, the range between efficient and seller-optimal provides a direct link to theoretical models of seller behaviour.

Assumption 6. *The seller threshold obeys the following conditions:*

1. *Sellers never sell below their own value.*
2. *Sellers never refuse a sale at a price above the optimal reserve.*

Let $R(V_S)$ be the optimal reserve price given V_S.[27] We define a conduct assumption as a parameter $\alpha \in [0, 1]$ such that $V_T = \alpha V_S + (1 - \alpha)R(V_S)$; that is, a convex combination of the efficient reserve V_S and the seller-optimal reserve $R(V_S)$. Each value of α is consistent with Assumption 6. As this is a known transformation when α is known, identification of F and G allows the econometrician to recover seller values under a specific conduct assumption.[28]

When a specific value for α is not imposed, Assumption 6 places bounds on the set of seller values that are consistent with an observed threshold, permitting set identification of seller values absent a more restrictive conduct assumption.

3.6. Identification of the distribution over the number of bidders

We conclude the discussion of identification by mentioning how the distribution over the number of bidders can be recovered using the identification of F and G combined with an empirical distribution of the number of bidders conditional on sale. The observed distribution over auction entrants is often a central object in auction studies, informing market design policies such as bidder preference programs (Krasnokutskaya and Seim (2011)) and revenue maximization via the optimal reserve. To recover entry costs or other model primitives related to the empirical distribution over the number of auction participants, the model must first obtain unbiased estimates of this distribution.

We define π_k as the probability of observing k bidders conditional on a sale. Following a similar logic to why the distribution of F is biased in favor of higher bidder values when using only transacted sales data, the distribution over the number of bidders will be biased in favor higher bidder numbers. Denoting by p_k the true distribution that there are k bidders in an auction, we would then expect $\pi_k < p_k$ when k is low and $\pi_k > p_k$ when k is high.

[27] The standard optimal reserve price in the IPV setting with continuous buyer value distribution (density) F (f) is the solution r^* to $r^* = V_S + [1 - F(r^*)]/f(r^*)$.

[28] To be precise, in the discrete case, the optimal reserve mapping will associate each grid point x_j with an interval on the real line of seller values for which x_j is the optimal threshold. Recovering a discrete seller value distribution requires these intervals to be associated with a grid of points for seller values, some of which may lie below x_1.

The identification of F and G, and the associated winning probabilities conditional on N, Ω_N, permits straightforward identification the overall distribution of bidders without conditioning on sale (but conditioning on $N \geq 1$). Using the fact that sales rates are identified given F and G, p_N is identified from π_N, a result formalised below:

Proposition 5. *Assume that* $\{\pi_n, \Omega_n\}_{n=1}^{\overline{N}}$ *are known. Then* p_N *is identified for each* $n \in \{1, ..., \overline{N}\}$

The proof is immediate based on the representation of the probability of observing N bidders conditional on the auction resulting in a sale as $p_N = (\pi_N/\Omega_N)/[\sum_{m=1}^{\overline{N}} \pi_m/\Omega_m]$.

3.7. Selective entry

Our identification results assume exogenous variation in the number of bidders. In some auction environments, bidders might receive a signal that is correlated with their value before making a (costly) auction entry decision. We discuss how our results can be extended to accommodate this environment; details are contained in Appendix D. We make two main assumptions on the entry process. The first is that each bidder's entry signal might be correlated with her own value, but entry signals and valuations are not correlated across bidders. The second is that we assume that sellers always enter: the threshold distribution is always G.[29] Under these assumptions, our identification results apply to the distribution of *entrants'* values. For welfare calculations or counterfactuals that do not rely on changes to entry, these model components may be sufficient.

When the distribution of potential, and not just realised, entrants' values are required, the approach of Gentry and Li (2014) to recover this distribution extends to our setting whenever we have point identification of the auction model. This approach assumes that there are observables that exogenously shift entry decisions, such as by changing entry costs without affecting valuation draws. When combined with an auction mechanism in which the distribution of values associated with submitted bids is identified, the distribution of potential bidders' values is identified.[30] The set of auction mechanisms they consider nests our model. Thus, when this model is point identified given the submitted bids, the distribution of potential entrants' values is identified; we provide a more detailed description in

[29] In settings in which sellers can also choose whether to enter the auction (versus an alternative, such as a negotiation listing), a similar instrumental variables approach as Gentry and Li (2014) might be applied to identifying seller selective entry primitives; we leave this for future work.

[30] Whether this is point or partial identification depends on whether the entry variation is continuous or discrete.

Appendix D. Our framework therefore accommodates existing empirical approaches to selective entry.

4. Monte Carlo Simulations

4.1. Estimation

The data available to the econometrician are $\{b_t, N_t\}_{t=1}^{T}$, where b_t is the transaction price and N_t is the number of bidders in successful auction t. We compare two estimators. The first implements the constructive identification argument and is nonparametric apart from discretising the price support. The second approximates the buyer and seller densities by finite Hermite series and estimates their coefficients jointly by minimum distance. Both estimators use only successful auctions.

The Omega-based estimator has an inner loop that computes buyer and seller distributions consistent with a candidate sales-rate ratio $\rho \equiv \Omega_M/\Omega_N$ and an outer loop that searches for a fixed point of the mapping Υ in equation (2). The inner loop is unique under the conditions of Theorem 1. If the model-implied ratio is within a pre-specified tolerance of the input ratio, the algorithm stops; otherwise, it replaces the input by the model-implied ratio and repeats. We implement the recursion as follows.

0. Estimate $H^{(n)}(x_j)$ for each $n \in M, N$ and $j = 1, ..., J$.

$$\hat{H}_j^{(n)} = \frac{\sum_{t=1}^{T} \mathbf{1}\{b_t \leq x_j,\ N_t = n\}}{\sum_{t=1}^{T} \mathbf{1}\{N_t = n\}}$$

1. Fix an input sales rate ratio $\rho \equiv \Omega_M/\Omega_N$.

2. At $j = 1$, $F_1(\rho)$ is the solution to

$$\lambda_1(x) \equiv (M-1)x^{M-N+1} - Mx^{M-N} - (N-1)\rho\frac{\hat{H}_1^{(M)}}{\hat{H}_1^{(N)}}x + N\rho\frac{\hat{H}_1^{(M)}}{\hat{H}_1^{(N)}} = 0$$

3. At each $j > 1$, $F_j(\rho)$ is the unique root in $(F_{j-1}(\rho), 1)$ of

$$\lambda_j(x) \equiv (M-1)x^M - Mx^{M-1} - (N-1)\phi_j x^N + N\phi_j x^{N-1} + F_{j-1}^M(\rho) - \phi_j F_{j-1}^N(\rho)$$

where

$$\phi_j = \frac{\rho H_j^{(M)} - \sum_{k=1}^{j-1} \tilde{G}_k(\rho)(F_k^M(\rho) - F_{k-1}^M(\rho))}{H_j^{(N)} - \sum_{k=1}^{j-1} \tilde{G}_k(\rho)(F_k^N(\rho) - F_{k-1}^N(\rho))}$$

5. At each j, calculate $\tilde{G}_j$ as

$$\tilde{G}_j = \frac{h_j^{(N)} + \tilde{G}_{j-1}(\rho)[NF_{j-1}^{N-1}(\rho) - NF_{j-1}^{N}(\rho)]}{N\tilde{F}_j^{N-1} - (N-1)F_j^N(\rho) - F_{j-1}^N(\rho)}$$

6. Evaluate the sales rate ratio implied by ρ using equation (2):

$$\Upsilon(\rho) = 1 + \sum_{k=1}^{J-1}[F_k^N(\rho) - F_k^M(\rho)]\tilde{g}_{k+1}(\rho)$$

The outer loop repeats steps 2 through 6 to minimise $|\Upsilon(\rho) - \rho|$.[31] Conditional on ρ, the main nonlinear computation is the solution of a sequence of sparse polynomial equations.

We also consider a standard semi-nonparametric estimator. We approximate each primitive distribution by integrating a squared polynomial multiplying a Normal reference density, following the construction of Gallant and Nychka (1987).[32] We demonstrate the construction for the buyer value distribution F; the seller threshold is analogous. Let r_F be the reference density restricted to an interval $\mathcal{I}_F$, and let $\psi_F(x) \equiv (\psi_{F,0}(x), \psi_{F,1}(x), ..., \psi_{F,K}(x))$ collect the Hermite polynomials of orders zero through K, evaluated at $(x - \mu_F)/\sigma_F$ and divided by the square roots of their factorials. The reference location μ_F and scale σ_F are obtained from preliminary CDF estimates and held fixed during joint estimation. For a unit coefficient vector a_F, the density is

$$\overline{f}_D(x; a_F) = r_F(x)[a_F'\psi_F(x)]^2, \tag{5}$$

Numerical integration gives the CDF $\overline{F}(x; a_D)$. We set $K = 4$ and parameterise the column vector $a_F = (1, \eta_F)/\sqrt{1 + \|\eta_F\|^2}$, with $\eta_F \in \mathbb{R}^4$, which fixes the coefficient norm and the sign of the constant coefficient and leaves four free parameters.

Let η_G denote the analogous parameters governing the seller threshold distribution, define $\eta \equiv (\eta_F, \eta_G)$, and let $H^{(n)}(x; \eta)$ denote the transaction-price CDF implied by the resulting F and G through equation (1). For the bidder-count set $\mathcal{N} = \{M, N\}$, we min-

[31] Our implementation uses contraction mapping iterations to find the fixed point. We establish conditions under which equation (2) defines a contraction in Appendix D.

[32] Other auction applications of this approach include Hernández et al. (2020), Freyberger and Larsen (2020), and Foster (2022).

imise

$$Q(\eta) = \frac{1}{J|\mathcal{N}|} \sum_{n \in \mathcal{N}} w_n \sum_{j=1}^{J} \left[\widehat{H}_j^{(n)} - H^{(n)}(x_j; \eta) \right]^2 \tag{6}$$

where $w_n = T_n / [\max_{m \in \mathcal{N}} T_m]$ and T_n is the number of successful auctions with n bidders.

4.2. Simulation specifications and results

We consider three distribution pairs. Case (i) is the "low sales rate" setting, with buyer values given by an exponential distribution with parameter 2.5 (shifted and truncated to the interval [3.5, 7]), and the seller threshold following a Normal distribution with mean $\mu_s = 5$ and standard deviation $\sigma_s = 0.85$, truncated to the interval [3.5,6].[33] Case (ii) is the "medium sales rate" setting, with buyer values following a Normal distribution with $\mu_b = 5$ and $\sigma_b = 0.55$ and the seller threshold also following a Normal distribution with $\mu_s = 4.75$ and $\sigma_s = 0.35$. Finally, case (iii) is the "high sales rate" setting, with the buyer value distribution given by a two-component Normal mixture with means $(4.5, 5.75)$, component variances $(0.05, 0.05)$, and weights $(0.25, 0.75)$; the seller threshold has the same Normal distribution as case (ii). The probability that the seller is influential in allocations and prices varies across the three settings. In case (i), sales probabilities are 70.5% with two bidders and 88.3% with four bidders; they are 83.6% with two bidders and 94.5% with four bidders in case (ii); and 95.7% with two bidders and 99.7% with four bidders in case (iii). This variation in distribution shape and seller influence captures a range of auction settings that empirical researchers might encounter.

For each pair, we conduct 1000 replications. Candidate auctions are generated in equal-sized bidder-count blocks, and $T = 1000$ successful auctions are sampled without replacement in each replication. We use the same simulated samples for both estimators. Figure 4 reports the pointwise 2.5th and 97.5th percentiles for the distributions of buyer values and the seller threshold. Appendix C reports the corresponding results for different bidder count pairs $(M, N) = (2, 1)$ and sample size $T = 5000$.

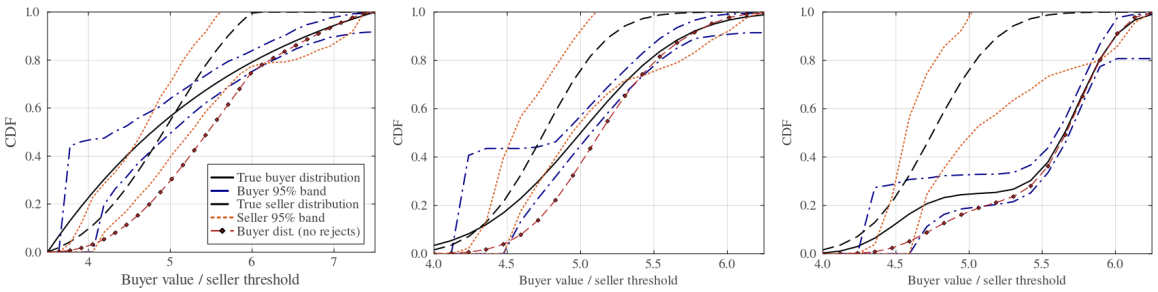
The distribution pairs show finite-sample estimation performance under varying conditions. When sellers are frequently influential in determining whether a sale occurs and at what price, as in case (i), their threshold distribution is recovered precisely at most grid points; this is demonstrated in panels (a) and (d). As the seller becomes less important for auction outcomes, the recovered seller threshold is less precise. However, even in the high

[33] "Low" is a relative term; the two-bidder sales rate for this case is substantially higher than in two-bidder housing auctions in our empirical application.

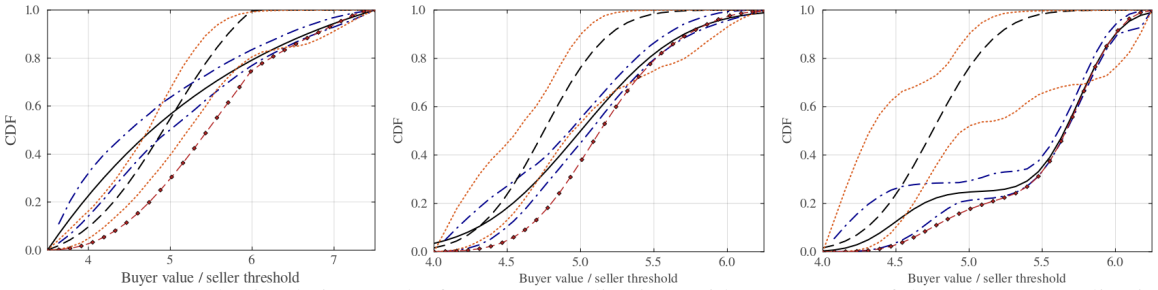
sales rate setting of case (iii), where the two- and four-bidder auctions' sales rates exceed 95% and 99%, panels (c) and (f) demonstrate that it remains possible to extract meaningful information about the seller threshold distribution.

Figure 4: Monte Carlo simulations, $(M, N) = (4, 2)$, 1000 auctions

(a) Low sales rate, Omega **(b)** Med sales rate, Omega **(c)** High sales rate, Omega



(d) Low sales rate, Hermite **(e)** Med sales rate, Hermite **(f)** High sales rate, Hermite



Notes: Monte Carlo simulation results from 1,000 replications with 1,000 successful auctions per replication. The top panels show the Omega-based estimator, while the bottom panels show the Hermite polynomial-based estimator.

Comparison of the top and bottom panels of Figure 4, together with the corresponding appendix figures, shows the tradeoffs associated with the non-parametric Omega-based estimator versus the semiparametric Hermite polynomial approximation. At lower transaction prices, the Omega bands can exclude the true CDFs, as there is too little transaction price mass to support a stable recursive update. This is more apparent with higher bidder counts and lower sample sizes. However, when the underlying data contain stronger identifying variation, as in the $(M, N) = (2, 1)$ cases shown in Appendix C, the non-parametric Omega estimator's performance generally exceeds the Hermite polynomial approximation.

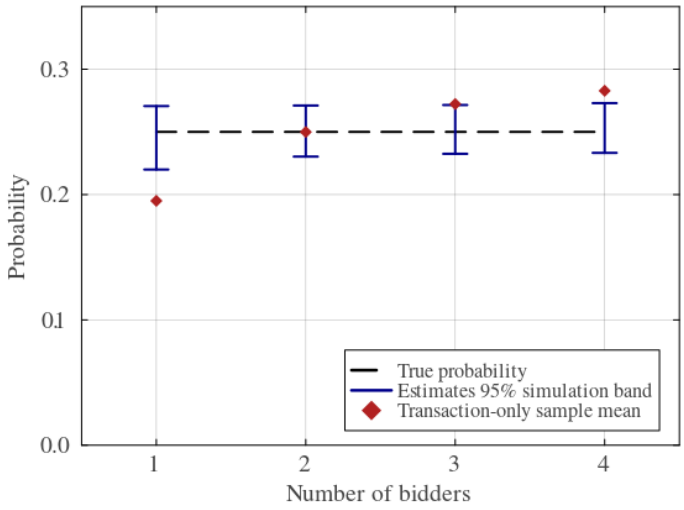
For comparison, the red diamonds show the mean buyer CDF estimate obtained when rejected offers are ignored. This estimate uses only the higher-bidder-count observations. The resulting CDF lies below the true buyer CDF over most of the support, corresponding to an estimated buyer-value distribution that first-order stochastically dominates the truth. The magnitude and location of the bias depend on the frequency of rejection and on the

relative shapes of the buyer and seller distributions.

4.3. Estimating the number of bidders

Figure 5 shows simulation results for estimating the number of bidders using the estimated $\hat{\Omega}_n$ constructed from the distributions of F and G and the observed probability of n bidders conditional on sale. The simulations use an estimation sample of 1000 auctions of the case (ii) specification (Normal buyer values and Normal seller threshold) described above. The F and G estimates used to construct the estimated sales rates are generated using the Hermite polynomial estimator. The mean estimated from the raw data, i.e., the distribution over N conditional on sale occurring, is shown as orange dots, while the 95% simulation band associated with the Monte Carlo simulations are shown in blue. These intervals all contain the true value of 1/4 for each $n = 1, 2, 3, 4$ used to generate the data, while the raw transaction data show significant bias in the number of bidders, underestimating $n = 1$ observations and over-estimating $n = 4$.

Figure 5: Estimating the distribution of N



Notes: Monte Carlo simulation results for the bidder count distribution using the Normal Buyer/Normal Seller specification of case (ii). The sample is generated by drawing n uniformly from $\{1, 2, 3, 4\}$, simulating auction outcomes, then sampling transaction prices and bidder counts from 1,000 successful auctions.

5. Application to Housing Auctions

We apply our methodology to housing auctions in the Greater Sydney region of New South Wales, Australia. Housing is one of the largest assets on household balance sheets,

and the Sydney housing market has an estimated total value of over AUD$4 trillion, with annual sales of more than AUD$240 billion. In Sydney, as in several other major cities in Australia, auctions are often used to sell residential properties and account for approximately 20% of all residential home sales.[34]

All auctions are ascending, open-outcry English auctions. The seller enlists the services of a third party who holds the auction in person with potential buyers. Per NSW regulations, buyers must register in advance to participate in the auction. Bidding proceeds as in many other English auctions, with bidders placing progressively higher bids until no party is willing to place another bid.

During the auction, the seller has the opportunity to place a single bid on her own behalf. Called the vendor bid, this bid is declared as such by the auctioneer when it is placed and is binding, in that regulations require the auction to be declared as unsuccessful (or "passed in") if the vendor bid is not surpassed. Auctions also specify a price at which the seller will face a monetary penalty if they do not sell to the highest bidder; these prices are required by the auctioneer to discourage sellers from "testing the market" with no intention of selling, among other behaviours.[35] We term this the commitment price and interpret it as the auctioneer's estimate of the market value of the home – for example, in our sample, its distribution is similar to the highest bid distribution. The commitment price is never revealed to bidders during the auction (except, possibly, when it has been exceeded), and sellers often choose to sell at high bids below it.

Our data set comes from a large Sydney-based firm that specialises in third-party auctioneer services and contains both successful and unsuccessful auctions. The data contain the date of the auction, the number of registered bidders, commitment price, highest bid, property postcode, the auctions result, and whether the auction was passed in on a vendor bid.

The institutional setting and data are useful for an application of our method for several reasons. The auction process is similar to our model: bidders must register in advance, they bid in an ascending open-cry auction that is strategically equivalent to a second-price auction (with the assumption of frictionless bidding or no winning "jump bids"), and the seller can make a binding take-it-or-leave-it offer a single time during the auction, which can

[34] Auctions are used to sell a wide range of properties; see Genesove and Hansen (2023).

[35] While the seller can refuse to sell at bids exceeding this price, they are obligated to pay the sales commission and other fees to the auctioneer and listing agent if they do not sell; in practice, refusals above this price are extremely rare.

serve as the seller's threshold. We also observe failed auctions, which allows us to compare how our method is able to recover the buyer value and seller threshold distributions to the distributions estimated using the full sample that includes failed auctions.[36]

Table 2 shows summary statistics for the auction data, separated by whether the auction was successful or not. We use all auctions with 10 or fewer bidders; the transaction only sample is the set of these auctions that results in a sale. About 28% of auctions with at least one registered bidder do not result in a transaction. Successful auctions are more likely to have higher bidder counts, with an average of 4.34 bidders compared to 2.31 for unsuccessful auctions, and are also more likely to have higher bids relative to the commitment price. However, the commitment price alone is not directly indicative of the sale price: 43% of successful auctions have a transaction price below the commitment price.

Table 2: Summary statistics

	Mean	Std. Dev.	p10	Median	p90
Panel A. Transaction sample					
Highest bid	1.29	0.69	0.65	1.10	2.20
Number of bidders	4.34	2.33	2.000	4.000	8.000
Commitment price	1.26	0.68	0.65	1.05	2.20
Num. Obs.	9,766				
Panel B. Full sample					
Price	1.31	0.70	0.65	1.11	2.28
Number of bidders	3.76	2.31	1.000	3.000	7.000
Commitment price	1.33	0.72	0.65	1.10	2.34
Sale	0.72	0.45	0.000	1.000	1.000
Num. Obs.	13,592				

5.1. Estimation

Our estimator follows the Hermite polynomial approximation minimum distance estimator described in Section 4, adapted to condition on the commitment price as an observable covariate. A semi-parametric approach allows us to circumvent practical issues that may arise in the application of the non-parametric Ω-based estimator.[37]

[36] However, even in our setting the failed auction records are incomplete, as it is not always clear for failed auctions whether the recorded price represents a vendor bid or a bidder's bid.

[37] For example, imprecise transaction price distributions estimate may violate the model's cross-bidder-count restrictions in small samples, leading to a failure to find an admissible solution for F and G using the algorithm described in Section 4.1. A continuous covariate further reduces the precision of the empirical

We extend the semi-parametric construction to allow the shapes of both primitive distributions to vary with commitment price c. Let $X_t = b_t/c_t$ denote the transaction-to-commitment price ratio, and let $F(x|c)$ and $G(x|c)$ be the conditional CDFs of buyer values and seller thresholds expressed relative to commitment price. We assume that, conditional on c, the primitive distributions do not otherwise vary with the number of bidders. To accommodate the institutional features of the commitment price (i.e., that sellers are effectively obligated to sell at bids above this price), we assume that the seller threshold always lies below the commitment price, so that seller threshold ratios are bounded above by one.

For each primitive, we follow equation (5) and orthonormalise its degree-four polynomial basis under the reference density on its support. We allow the coefficient vector to depend on commitment price through

$$z(c) = \frac{\log c - \log \underline{c}}{\log \overline{c} - \log \underline{c}}, \quad v(c) = \Big(1, z(c) - \tfrac{1}{2}, [z(c) - \tfrac{1}{2}]^2\Big), \tag{7}$$

where $\underline{c}$ and $\overline{c}$ are the lower and upper endpoints of the commitment price grid. For $D \in \{F, G\}$, the polynomial coefficients $a_D = (\beta_D v(c)')/\|\beta_D v(c)'\|$ for a matrix $\beta_D \in \mathbb{R}^{5\times 3}$. Each matrix β_D has a common scale and sign normalisation, leaving 14 parameters per primitive. Integration gives $F(x|c)$ and $G(x|c)$, and equation (1) gives the corresponding $H^{(n)}(x|c)$ and sale probabilities $\Omega_n(c)$.

Let $\overline{H}_j^{(n)}(c_\ell)$ denote the kernel estimate of the conditional transaction price CDF. We estimate the two coefficient matrices by minimising

$$\mathcal{Q}(\beta_F, \beta_G) = \frac{1}{JL|\mathcal{N}|} \sum_{\ell=1}^{L} \sum_{n\in\mathcal{N}} w_{n\ell} \sum_{j=1}^{J} \left[\overline{H}_j^{(n)}(c_\ell) - H^{(n)}(x_j \mid c_\ell; \beta_F, \beta_G)\right]^2. \tag{8}$$

Here $\mathcal{N} = \{1, \ldots, 10\}$, $J = 101$, $L = 15$, and $w_{n\ell}$ is the local kernel effective sample size divided by its largest value across bidder counts and commitment prices.

We also estimate the model using the full sample, including failed auctions. We retain the degree four conditional Hermite representation of $F(x|c)$ and $G(x|c)$, with quadratic coefficient functions in normalised log commitment price with coefficients represented by β. Let U denote the seller threshold ratio and Y_n the highest buyer-value ratio in an auction with n bidders. An auction fails when $Y_n < U$. For auctions recorded as failing on a

transaction price distributions, increasing the probability that some (M, N) pair and covariate value c^0 will have estimate distributions $H^{(M)}(x|c^0), H^{(N)}(x|c^0)$ that do not generate valid F, G solutions.

vendor bid, we interpret the price (i.e., vendor bid) ratio as U; we interpret price records from failures without a vendor bid as Y_n.[38] The model distributions matched to the two failed auction groups are therefore $H^{(n)}_{\mathrm{v},j}(c) = \Pr(U \leq x_j \mid Y_n < U, n, c)$ and $H^{(n)}_{\mathrm{b},j}(c) = \Pr(Y_n \leq x_j \mid Y_n < U, n, c)$. The full-sample criterion is

$$Q_{\text{full}}(\beta) = Q_{\mathrm{s}}(\beta) + Q_{\mathrm{v}}(\beta) + Q_{\mathrm{b}}(\beta) + Q_{\Omega}(\beta), \tag{9}$$

where the first three terms match the empirical CDFs for successful transactions, vendor-bid failures and non-vendor failures, respectively, and the fourth matches observed sale probabilities. The components are weighted by their counts and averaged over commitment price nodes and bidder counts. Appendix B.4 provides additional implementation details.

5.2. Results

Figures 6 and 7 show the estimation results. The estimated distributions for buyer values and the seller threshold at the median commitment price are displayed in Figure 6; results for the 15th and 85th percentiles of the commitment price are contained in Appendix B. Following 4, the shape of $F(\cdot|c)$ is unidentified below the lower bound of the transaction price support. The lowest conditional transaction price quantile used in estimation is the one percent level, corresponding to approximately a price-to-commitment ratio of 0.9, which we set to the lower x-axis support in Figure (a); the shape of F below this value is not identified. Over the support of x, the transaction only estimate provides very similar values to the full sample estimate obtained that uses the information from the full sample, including failed auctions. The estimates for the seller threshold distributions across the transaction only and full sample cases also have similar shapes.

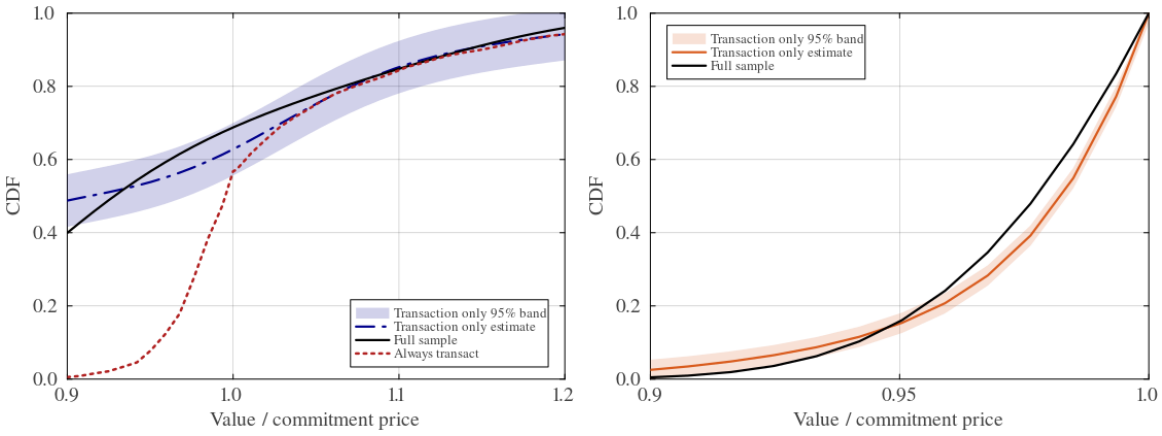
As a comparison, we also estimate and plot the buyer value distribution inferred from assuming that all auctions transact using the $n \geq 2$ sample of transactions. As expected, this distribution implies substantially higher values, dominating the transaction-only sample at values lower than the commitment price. As bids below the commitment price account for over 40% of sales in our sample, this bias is economically meaningful for predicting both outcomes and buyer surplus.

Both the full sample and transaction only estimates indicate that a large fraction of

[38] This follows our earlier work in Barkley et al. (2025), where we found this specification fit best fit the data in that full sample estimation. We assume that, conditional on failure, bidder count and commitment price, the recorded prices/bids of each type represent the corresponding latent distribution among failed auctions.

Figure 6: Estimated buyer and seller distributions for median c

(a) Buyer values **(b)** Seller threshold



buyer values lie below the lower support of the seller threshold. This finding is related to the Australian housing policy debate on "underquoting," the practice of sellers (or the listing agents operating on their behalf) intentionally advertising homes at prices lower than what the seller is prepared to accept in order to attract a larger pool of interested buyers.[39] Our estimates are consistent with underquoting by sellers successfully attracting to the auction buyers who have no realistic chance of purchasing the property.

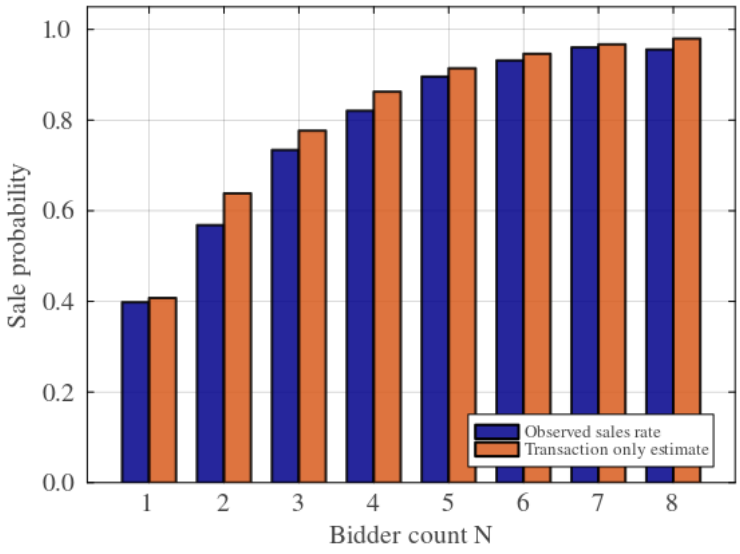
We also compare the transaction only estimates' ability to recover the empirical sales rates. These are untargeted moments, as the transaction only sample does not contain any information about the empirical sales rates. The results are displayed in Figure 7. While the model estimates tend to over-predict sale probabilities, the model-implied sales rates nevertheless provide a close approximation to the empirical sales rates despite using only transaction data to obtain them.

6. Conclusion

We study identification of buyer and seller primitives in second-price auctions when rejected offers are unobserved and only successful transactions are recorded. With variation in the number of bidders and discrete supports, we reduce identification to a fixed-point problem in the ratio of sale probabilities across bidder counts; uniqueness delivers point identification and multiple fixed points imply set identification. We provide a sim-

[39]For example, Victoria, another Australian state in which residential housing auctions are common, has recently implemented a policy requiring auction sellers to publicly reveal the commitment price one week before the auction is held to curtail this practice.

Figure 7: Estimated sales rates



ple sufficient condition of upper-tail support separation which ensures point identification. Thresholds map into seller values under a conduct assumption. Monte Carlo experiments validate our approach and quantify the upward bias from ignoring rejected offers, and an application to Sydney housing auctions illustrates the model's performance in a real-world setting.

These results expand the set of feasible counterfactuals in auction marketplace environments when only transaction data is available. By recovering the distribution of seller thresholds – and, with an additional assumption, seller values – researchers can evaluate policies that hinge on supply-side responses: reserve-rule design, platform fee changes and revenue sharing, seller participation, and entry/eligibility rules that shift bidder counts. Since the approach identifies sale probabilities by bidder count and decomposes prices into bidder- and seller-side primitives, it supports welfare decompositions and comparative statics for market-design interventions that are inaccessible with buyer-only analysis.

Although our non-identification results are non-parametric, they also inform parametric approaches. Our dual results on identification given the sales rates ratio and upper and lower bounds on the ratio ensures that an econometrician can find all roots in a non-parametric environment, and so conclude in any given case whether the primitives are identified or not. When the curse of dimensionality forces the econometrician to impose parametric restrictions, such certainty is lost. Our results establishing sufficient conditions for identification and examples of non-identification are then a sufficient guide for when the

researcher can be confident that underlying non-identification, perhaps masked by the parametric restrictions, is a problem and when not. Specifically, estimates that display seller distributions substantially to the right of buyer distributions should be viewed as potentially non-reliable.

Our independent private values setting abstracts from many other features of auction environments studied in the empirical literature, including bidder risk aversion (Guerre et al., 2009) and unobserved heterogeneity (Krasnokutskaya, 2011). Our model differs in assigning an active role to both buyers and sellers in determining auction outcomes. The selection and model misspecification issues arising from ignoring seller decisions may affect the implementation of many of the aforementioned models. For example, identifying unobserved heterogeneity may rely on the joint distribution of bid pairs (Krasnokutskaya, 2011) or pairs of bid order statistics (Luo and Xiao, 2023). If seller actions determine prices or allocations but are not modelled or directly observable in the data, the bid distributions used to estimate these models are based on selected samples, possibly leading researchers to make incorrect inferences about the data generating process. Utilising the information contained in over-identifying restrictions, such as when more than two bidder counts are observed, may provide an avenue to identifying these model components. We leave detailed consideration of these issues to future work.

References

J. Albrecht, P. A. Gautier, and S. Vroman. Efficient entry in competing auctions. *American Economic Review*, 104(10):3288–3296, 2014.

P. Andreyanov and E. H. Caoui. Secret reserve prices by uninformed sellers. *Quantitative Economics*, 13(3):1203–1256, 2022.

A. Aradillas-López, A. Gandhi, and D. Quint. Identification and inference in ascending auctions with correlated private values. *Econometrica*, 81(2):489–534, 2013.

O. Ashenfelter and K. Graddy. Art auctions. *Handbook of the Economics of Art and Culture*, 1:909–945, 2006.

S. Athey and P. A. Haile. Identification of standard auction models. *Econometrica*, 70(6): 2107–2140, 2002.

S. Athey and P. A. Haile. Empirical models of auctions. In R. Blundell, W. K. Newey, and T. Persson, editors, *Advances in Economics and Econometrics: Theory and Applications,*

Ninth World Congress, Volume II, chapter 1, pages 1–45. Cambridge University Press, Cambridge, 2006. doi: 10.1017/CBO9781139052276.003.

S. Athey, J. Levin, and E. Seira. Comparing open and sealed bid auctions: Evidence from timber auctions. *The Quarterly Journal of Economics*, 126(1):207–257, 2011.

M. Backus and G. Lewis. Dynamic demand estimation in auction markets. *Review of Economic Studies*, 92(2):837–872, 2025.

P. Bajari and A. Hortaçsu. The winner's curse, reserve prices, and endogenous entry: Empirical insights from ebay auctions. *RAND Journal of Economics*, pages 329–355, 2003.

A. Barkley. Cost and efficiency in government outsourcing: Evidence from the dredging industry. *American Economic Journal: Microeconomics*, 13(4):514–547, 2021.

A. Barkley, J. R. Groeger, and R. A. Miller. Bidding frictions in ascending auctions. *Journal of Econometrics*, 223(2):376–400, 2021.

A. Barkley, D. Genesove, and J. Hansen. Haggle or hammer? dual-mechanism housing search. Technical report, SSRN, 2025.

M. Bernasconi, M. Espinosa, R. Macchiavello, and C. Suarez. Anticipatory responses and coordinated bidding in the colombia wholesale energy market. Technical report, 2025.

S. Berry, J. Levinsohn, and A. Pakes. Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890, 1995.

F. A. Bugni and Y. Wang. Inference in auctions with many bidders using transaction prices. *arXiv preprint arXiv:2311.09972*, 2026.

J. Bulow and P. Klemperer. Auctions Versus Negotiations. *The American Economic Review*, 86(1):180–194, 1996.

R. Dahis, B. Ricca, T. Scot, N. Sales, and L. Nascimento. Fiscal capacity and execution at the local level: New evidence from brazil. *The World Bank Economic Review*, page lhag006, 2026.

L. Einav, T. Kuchler, J. Levin, and N. Sundaresan. Assessing sale strategies in online markets using matched listings. *American Economic Journal: Microeconomics*, 7(2): 215–247, 2015.

C. Ferraz, F. Finan, and D. Szerman. Procuring firm growth: the effects of government purchases on firm dynamics. Technical report, National bureau of economic research, 2015.

J. Foster. Semi-nonparametric estimation of secret reserve prices in auctions. *Economics Letters*, 220:110843, 2022. doi: 10.1016/j.econlet.2022.110843.

J. Freyberger and B. J. Larsen. Identification in ascending auctions, with an application to digital rights management. *Quantitative Economics*, Forthcoming, 2020.

A. R. Gallant and D. W. Nychka. Semi-nonparametric maximum likelihood estimation. *Econometrica: Journal of the Econometric Society*, pages 363–390, 1987.

D. Genesove. Search at wholesale auto auctions. *The Quarterly Journal of Economics*, 110 (1):23–49, 1995.

D. Genesove and J. Hansen. Auctions and negotiations in housing price dynamics. *Review of Economics and Statistics*, pages 1–39, 2023.

M. Gentry and T. Li. Identification in auctions with selective entry. *Econometrica*, 82(1): 315–344, 2014.

N. Gimenes, T. Qi, and S. Srisuma. Identification and estimation of seller risk aversion in ascending auctions. *arXiv preprint arXiv:2509.19945*, 2025.

A. S. Gorbenko and A. Malenko. Strategic and financial bidders in takeover auctions. *The Journal of Finance*, 69(6):2513–2555, 2014.

J. R. Groeger. A study of participation in dynamic auctions. *International Economic Review*, 55(4):1129–1154, 2014.

S. Grundl and Y. Zhu. Two results on auctions with endogenous entry. *Economics Letters*, 234:111500, 2024.

J. Guan and B. Xu. Reserve price signaling with public information: Evidence from online auto auctions. *The RAND Journal of Economics*, 57(1):245–266, 2026.

E. Guerre, I. Perrigne, and Q. Vuong. Optimal nonparametric estimation of first-price auctions. *Econometrica*, 68(3):525–574, 2000.

E. Guerre, I. Perrigne, and Q. Vuong. Nonparametric identification of risk aversion in first-price auctions under exclusion restrictions. *Econometrica*, 77(4):1193–1227, 2009.

P. Haile, H. Hong, and M. Shum. Nonparametric tests for common values at first-price sealed-bid auctions, 2003.

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

K. Hendricks, R. H. Porter, and C. A. Wilson. Auctions for oil and gas leases with an informed bidder and a random reservation price. *Econometrica: Journal of the Econometric Society*, pages 1415–1444, 1994.

C. Hernández, D. Quint, and C. Turansick. Estimation in english auctions with unobserved heterogeneity. *The RAND Journal of Economics*, 51(3):868–904, 2020.

B. R. Hickman, T. P. Hubbard, and H. J. Paarsch. Identification and estimation of a bidding model for electronic auctions. *Quantitative Economics*, 8(2):505–551, 2017.

A. Hortaçsu and I. Perrigne. Empirical perspectives on auctions. In *Handbook of Industrial Organization*, volume 5, pages 81–175. Elsevier, 2021.

K. Kang and R. A. Miller. Winning by default: Why is there so little competition in government procurement? *The Review of Economic Studies*, 89(3):1495–1556, 2022.

T. Komarova. A new approach to identifying generalized competing risks models with application to second-price auctions. *Quantitative Economics*, 4(2):269–328, 2013.

Y. Kong. Sequential auctions with synergy and affiliation across auctions. *Journal of Political Economy*, 129(1):148–181, 2021.

S. G. Krantz and H. R. Parks. *A primer of real analytic functions*. Springer Science & Business Media, 2002.

E. Krasnokutskaya. Identification and estimation of auction models with unobserved heterogeneity. *The Review of Economic Studies*, 78(1):293–327, 2011.

E. Krasnokutskaya and K. Seim. Bid preference programs and participation in highway procurement auctions. *American Economic Review*, 101(6):2653–2686, 2011.

N. Lacetera, B. J. Larsen, D. G. Pope, and J. R. Sydnor. Bid takers or market makers? the effect of auctioneers on auction outcome. *American Economic Journal: Microeconomics*, 8(4):195–229, 2016.

B. Larsen and A. L. Zhang. A mechanism design approach to identification and estimation. Technical report, National Bureau of Economic Research, 2018.

D. Levin and J. L. Smith. Equilibrium in auctions with entry. *The American Economic Review*, pages 585–599, 1994.

H. Li and N. Liu. Nonparametric identification and estimation of double auctions with bargaining. *Econometric Theory*, page 1–36, 2026. doi: 10.1017/S0266466626100413.

T. Li and X. Zheng. Entry and competition effects in first-price auctions: Theory and evidence from procurement auctions. *The Review of Economic Studies*, 76(4):1397–1429, 2009.

T. Li and X. Zheng. Information acquisition and/or bid preparation: A structural analysis of entry and bidding in timber sale auctions. *Journal of Econometrics*, 168(1):29–46, 2012.

Y. Luo and R. Xiao. Identification of auction models using order statistics. *Journal of Econometrics*, 236(1):105457, 2023.

V. Marmer, A. Shneyerov, and P. Xu. What model for entry in first-price auctions? a nonparametric approach. *Journal of Econometrics*, 176(1):46–58, 2013.

M. Marra. Estimating an auction platform game with two-sided entry. *Journal of Political Economy*, 133(6):1932–1974, 2025.

J. J. McConnell and A. Saretto. Auction failures and the market for auction rate securities. *Journal of Financial Economics*, 97(3):451–469, 2010.

N. H. Miller and G. Sheu. Quantitative methods for evaluating the unilateral effects of mergers. *Review of Industrial Organization*, 58(1):143–177, 2021.

H. J. Paarsch. Deciding between the common and private value paradigms in empirical models of auctions. *Journal of econometrics*, 51(1-2):191–215, 1992.

H. J. Paarsch. Deriving an estimate of the optimal reserve price: an application to british columbian timber sales. *Journal of Econometrics*, 78(1):333–357, 1997.

P. C. Reiss and F. A. Wolak. Structural econometric modeling: Rationales and examples from industrial organization. *Handbook of econometrics*, 6:4277–4415, 2007.

J. W. Roberts. Unobserved heterogeneity and reserve prices in auctions. *The RAND Journal of Economics*, 44(4):712–732, 2013.

J. N. Rosse. Estimating cost function parameters without using cost data: Illustrated methodology. *Econometrica: Journal of the Econometric Society*, pages 256–275, 1970.

U.S. Government Accountability Office. Reverse auctions: Guidance is needed to maximize competition and achieve cost savings. Technical Report GAO-14-108, U.S. Government Accountability Office, Washington, DC, Dec. 2013. URL https://www.gao.gov/products/gao-14-108.

P. Xu. Nonparametric estimation of entry cost in first-price procurement auctions. *Journal of Applied Econometrics*, 28(6):1046–1065, 2013.

Appendix A. Proofs

Several equations are used in the proofs that are not present in the main text. We first define the function A_n as

$$A_n(x) \equiv nx^{n-1} - (n-1)x^n.$$

For $n \geq 2$, $A_n(F_j)$ is the probability that the second highest buyer value is at most x_j. Next, the transaction-price probability mass function at point j satisfies

$$\Omega_n h_j^{(n)} = A_n(F_j)G_j - A_n(F_{j-1})G_{j-1} - F_{j-1}^n(G_j - G_{j-1}). \tag{A.1}$$

Appendix A.1. Derivations for the two-point distribution

We first record the ordering of the transaction-price distributions across bidder counts.

Lemma 1. *In the two-point model, $H_1^{(M)} < H_1^{(N)}$ for $M > N$.*

Proof. It suffices to compare adjacent bidder counts. After cancelling common positive factors, $H_1^{(N+1)} < H_1^{(N)}$ is equivalent to

$$(1 - F_1)[N(1 - F_1) + F_1^{N+1}(1 - G_1)] > 0$$

■

Proof of Proposition 2. Letting x represent a candidate F_1 value, the threshold probability implied by the n-bidder transaction price equation is

$$g_n(x) = \frac{H_1^{(n)}(1 - x^n)}{A_n(x) - H_1^{(n)}x^n}.$$

Equating $g_N(x)$ and $g_{N+1}(x)$ yields $p(x) = 0$, where

$$\begin{aligned} p(x) &= [H_1^{(N)}(1 - x^N)][A_{N+1}(x) - H_1^{(N+1)}x^{N+1}] \\ &\quad -[H_1^{(N+1)}(1 - x^{N+1})][A_N(x) - H_1^{(N)}x^N] \end{aligned}$$

Note that $x = 1$ is always a root. Write $p(x) = (x-1)q(x)$; polynomial division gives

$$\begin{aligned} q(x) &= [NH_1^{(N)} - (N-1)H_1^{(N+1)}]x^{N+1} + [H_1^{(N+1)} - H_1^{(N)}]\sum_{k=2}^{N} x^k \\ &\quad +[H_1^{(N+1)} - (N+1)H_1^{(N)} - H_1^{(N)}H_1^{(N+1)}]x + NH_1^{(N+1)} \end{aligned}$$

where the sum is zero when $N = 1$. Lemma 1 implies that the first coefficient is positive, the second coefficient is negative, the third coefficient is negative, and the fourth coefficient is positive. The coefficient sequence of q therefore has two sign changes. Because the data are rationalisable, q has at least one root in $(0, 1)$, so Descartes' rule implies that it has exactly two positive roots, counted with multiplicity.

Moreover, $q(0) = NH_1^{(N+1)} > 0$, its leading coefficient is positive, and

$$q(1) = p'(1) = (N+1)H_1^{(N+1)} - H_1^{(N)}[N + H_1^{(N+1)}]$$

If $q(1) < 0$, one positive root lies in $(0, 1)$ and the other exceeds one. If $q(1) = 0$, the second root is one. If $q(1) > 0$, q has the same sign at zero and one. The number of roots in $(0, 1)$ is then even and, because the interval contains the true value of F_1, both positive roots lie in that interval. To verify admissibility, let $\underline{x}$ solve $A_N(\underline{x}) = H_1^{(N)}$. Then $g_N(x) < 1$ if and only if $x > \underline{x}$. Interior rationalisability implies $H_1^{(N+1)} > A_{N+1}(\underline{x})$, and hence $g_{N+1}(\underline{x}) > g_N(\underline{x}) = 1$. Because p has the sign of $g_N - g_{N+1}$, $q(\underline{x}) > 0$. One root is the true value of F_1 and exceeds $\underline{x}$; when $q(1) > 0$, the other root must therefore also exceed $\underline{x}$. Both roots then generate $g_n(x) \in (0, 1)$. They are distinct away from the boundary of the rationalisable set. Finally,

$$q(1) \leq 0 \quad \Longleftrightarrow \quad H_1^{(N+1)} \leq \frac{NH_1^{(N)}}{N + 1 - H_1^{(N)}},$$

which establishes the result. ■

Appendix A.2. Proofs for general distributions case

The proof of Theorem 1 makes use of the following lemma:

Lemma 2. *Let $M > N \geq 1$, $\phi > 0$, and $w \in [0, 1)$. Define*

$$\lambda(x) \equiv (M-1)x^M - Mx^{M-1} - (N-1)\phi x^N + N\phi x^{N-1} + w^M - \phi w^N.$$

Then:

(i) if $\lambda(w) < 0$, then λ has no root in $[w, 1]$;

(ii) if $\lambda(w) > 0$, then λ has at most one root in $[w, 1]$;

(iii) if $w > 0$ and $\lambda(w) = 0$, then λ has no root in $(w, 1]$;

(iv) if $w = 0$, then λ has at most one positive root in $(0, 1]$.

Moreover, any root in $(w, 1)$ *under (ii) or (iv) is simple and satisfies* $\lambda'(x) < 0$*.*

Proof. First note that

$$\lambda(w) = (1 - w)w^{N-1}\left(N\phi - Mw^{M-N}\right), \tag{A.2}$$

with $w^0 = 1$ when $N = 1$. If $N = 1$, then λ is strictly decreasing on $(0, 1)$:

$$\lambda'(x) = M(M - 1)x^{M-2}(x - 1) < 0$$

for $x \in (0, 1)$. All four claims follow immediately.

Now suppose $N \geq 2$. For $x \in (0, 1)$,

$$\lambda'(x) = (1 - x)x^{N-2}\left[N(N - 1)\phi - M(M - 1)x^{M-N}\right]. \tag{A.3}$$

The bracketed term is strictly decreasing in x, and the term outside the brackets is strictly positive. Hence, once λ is decreasing at some point in $(0, 1)$, it is decreasing at all higher points in $(0, 1)$.

If $\lambda(w) < 0$, equation (A.2) implies

$$N\phi < Mw^{M-N}.$$

Since $M > N$, this gives

$$N(N - 1)\phi < M(M - 1)x^{M-N}$$

for all $x \in [w, 1]$. Equation (A.3) then implies that $\lambda'(x) < 0$ throughout $[w, 1)$. Thus λ is strictly decreasing on $[w, 1]$, and because it starts below zero it has no root in that interval, proving (i).

If $\lambda(w) = 0$ and $w > 0$, equation (A.2) gives

$$N\phi = Mw^{M-N}.$$

Again $M > N$, so $\lambda'(x) < 0$ for all $x \in [w, 1)$. Therefore w is the only possible root in $[w, 1]$, and there is no root in $(w, 1]$, proving (iii).

If $\lambda(w) > 0$, the derivative can switch sign at most once, from positive to negative. Therefore λ can cross zero at most once after w: if a first root exists, the crossing must

occur while λ is strictly decreasing, and after that point λ remains decreasing. This proves (ii) and the simplicity claim. Finally, if $w = 0$, then $\lambda(x) > 0$ for all sufficiently small positive x; the same single-crossing argument gives at most one positive root in $(0, 1]$, and any such root lies on the strictly decreasing branch. This proves (iv) and completes the proof. ■

Proof of Theorem 1. The proof proceeds by induction on the grid index j. At $j = 1$, equation (1) gives

$$\Omega_n H_1^{(n)} = G_1 A_n(F_1) = G_1 \left[n F_1^{n-1} - (n-1) F_1^n \right].$$

Taking the ratio for bidder counts M and N, and writing

$$\phi_1 \equiv \frac{\Omega_M H_1^{(M)}}{\Omega_N H_1^{(N)}} = \rho \frac{H_1^{(M)}}{H_1^{(N)}},$$

we obtain

$$\phi_1 = \frac{M F_1^{M-1} - (M-1) F_1^M}{N F_1^{N-1} - (N-1) F_1^N}.$$

For positive roots this is equivalent to

$$(M-1)x^M - Mx^{M-1} - (N-1)\phi_1 x^N + N\phi_1 x^{N-1} = 0.$$

This is the polynomial in Lemma 2 with $w = 0$ (case (iv)). Since the model generates the data, an admissible positive root exists; Lemma 2 implies that it is unique. Hence F_1 is identified as a function of ρ. Then

$$G_1 = \frac{\Omega_N H_1^{(N)}}{A_N(F_1)} = \Omega_N \tilde{G}_1(\rho), \qquad \tilde{G}_1(\rho) \equiv \frac{H_1^{(N)}}{A_N(F_1(\rho))}.$$

Now assume that $F_1, \ldots, F_{j-1}$ are identified as functions of ρ, and that

$$G_k = \Omega_N \tilde{G}_k(\rho), \qquad k = 1, \ldots, j-1.$$

Define

$$Q_j^{(n)} \equiv \Omega_n H_j^{(n)} - \sum_{k=1}^{j-1} G_k (F_k^n - F_{k-1}^n).$$

Equation (1) implies

$$Q_j^{(n)} = G_j \left[nF_j^{n-1} - (n-1)F_j^n - F_{j-1}^n \right].$$

Taking the ratio for M and N, let

$$\phi_j \equiv \frac{Q_j^{(M)}}{Q_j^{(N)}}.$$

Then F_j must solve

$$(M-1)x^M - Mx^{M-1} - (N-1)\phi_j x^N + N\phi_j x^{N-1} + F_{j-1}^M - \phi_j F_{j-1}^N = 0$$

on $[F_{j-1}, 1]$. This is the polynomial in Lemma 2 with $w = F_{j-1}$ and $\phi = \phi_j$. Consistency of the model and data imply existence of an admissible root. The lemma implies that there is at most one admissible root, so F_j is identified.

It remains to verify that ϕ_j depends only on the sales-rate ratio ρ, not on the levels Ω_N and Ω_M separately. Using the induction hypothesis,

$$\frac{Q_j^{(M)}}{Q_j^{(N)}} = \frac{\rho H_j^{(M)} - \sum_{k=1}^{j-1} \tilde{G}_k(\rho)(F_k^M - F_{k-1}^M)}{H_j^{(N)} - \sum_{k=1}^{j-1} \tilde{G}_k(\rho)(F_k^N - F_{k-1}^N)}.$$

Thus F_j is a function of ρ. Finally,

$$G_j = \frac{Q_j^{(N)}}{NF_j^{N-1} - (N-1)F_j^N - F_{j-1}^N} = \Omega_N \tilde{G}_j(\rho)$$

for a known function $\tilde{G}_j(\rho)$. This completes the induction.

At $j = J$, imposing $G_J = 1$ gives

$$1 = G_J = \Omega_N \tilde{G}_J(\rho),$$

so $\Omega_N = 1/\tilde{G}_J(\rho)$, $\Omega_M = \rho\Omega_N$, and $G_j = \Omega_N \tilde{G}_j(\rho)$ are identified. ■

Lemma 3. *For every $\rho_0 \in \mathcal{R}$, the functions $\tilde{F}_j(\rho)$, $j = 1, \ldots, J-1$, and $\tilde{G}_j(\rho)$, $j = 1, \ldots, J$, admit real analytic extensions to an open interval in $\mathbb{R}$ containing ρ_0. Consequently, Υ admits such an extension at every $\rho_0 \in \mathcal{R}$ and is continuous on $\mathcal{R}$.*

Proof. Fix $\rho_0 \in \mathcal{R}$. At $j = 1$, $\phi_1(\rho) = \rho H_1^{(M)}/H_1^{(N)}$ is real analytic. The admissible root $\tilde{F}_1(\rho_0) \in (0,1)$ of the polynomial in the proof of Theorem 1 lies on its strictly decreasing

branch by Lemma 2, and hence the derivative of the polynomial with respect to its root is nonzero. The analytic implicit function theorem (Krantz and Parks, 2002, Theorem 2.3.5) therefore gives an open interval $I \subset \mathbb{R}$ containing ρ_0 and a real analytic extension of $\tilde{F}_1$ to I. After shrinking I, the extension remains on the strictly decreasing branch and in $(0,1)$, so uniqueness of the admissible root implies that it agrees with the recovered $\tilde{F}_1$ on $I \cap \mathcal{R}$. The expression

$$\tilde{G}_1(\rho) = \frac{H_1^{(N)}}{A_N(\tilde{F}_1(\rho))}$$

then gives a real analytic extension of $\tilde{G}_1$.

After shrinking I as necessary, suppose that $\tilde{F}_k$ and $\tilde{G}_k$ are real analytic on I for $k < j$. The ratio used to recover $\tilde{F}_j$ is

$$\phi_j(\rho) = \frac{\rho H_j^{(M)} - \sum_{k=1}^{j-1} \tilde{G}_k(\rho) \left[\tilde{F}_k(\rho)^M - \tilde{F}_{k-1}(\rho)^M\right]}{H_j^{(N)} - \sum_{k=1}^{j-1} \tilde{G}_k(\rho) \left[\tilde{F}_k(\rho)^N - \tilde{F}_{k-1}(\rho)^N\right]}.$$

At ρ_0, the denominator equals

$$\frac{Q_j^{(N)}}{\Omega_N} = \tilde{G}_j \left[A_N(F_j(\rho_0)) - F_{j-1}^N(\rho_0)\right] > 0,$$

where positivity follows from admissibility. Thus ϕ_j is real analytic near ρ_0. The polynomial defining $F_j(\rho)$ has analytic coefficients and, by the single-crossing argument, a simple admissible root in $(F_{j-1}(\rho), 1)$. The analytic implicit function theorem therefore gives a real analytic extension of $F_j(\rho)$ which, after shrinking I, remains in $(F_{j-1}(\rho), 1)$ and agrees with the unique admissible root on $I \cap \mathcal{R}$. Finally,

$$\tilde{G}_j(\rho) = \frac{H_j^{(N)} - \sum_{k=1}^{j-1} \tilde{G}_k(\rho) \left[F_k(\rho)^N - F_{k-1}(\rho)^N\right]}{A_N(F_j(\rho)) - F_{j-1}(\rho)^N}$$

is real analytic after shrinking I, because its denominator is strictly positive at ρ_0:

$$A_N(F_j(\rho_0)) - F_{j-1}^N(\rho_0) = N F_j^{N-1}(\rho_0)(1 - F_j(\rho_0)) + F_j^N(\rho_0) - F_{j-1}^N(\rho_0) > 0.$$

This completes the induction through $j = J - 1$. Setting $F_J(\rho_0) = 1$ in the preceding expression gives a real analytic extension of $\tilde{G}_J$, since its denominator is $1 - F_{J-1}^N(\rho_0) > 0$ at ρ_0. Equation (2) then gives a real analytic extension of Υ to an open interval in $\mathbb{R}$

containing ρ_0. Since ρ_0 was arbitrary, Υ is continuous on $\mathcal{R}$. ■

Proof of Theorem 2. Fix $\rho \in \mathcal{R}$ and suppress dependence on ρ. Let $\Omega_n^I(\rho)$ denote the sale probability implied by the candidate primitives. Summation by parts in equation (1) at the upper support point, where $F_J = G_J = 1$, gives

$$\Omega_n(\rho) = 1 - \sum_{k=1}^{J-1} F_k^n (G_{k+1} - G_k).$$

This follows from $\sum_{j=1}^{J} G_j (F_j^n - F_{j-1}^n) = 1 - \sum_{k=1}^{J-1} F_k^n g_{k+1}$. Subtract the N-bidder expression from the M-bidder expression:

$$\Omega_M^I(\rho) - \Omega_N^I(\rho) = \sum_{k=1}^{J-1} \Big[F_k(\rho)^N - F_k(\rho)^M \Big] g_{k+1}(\rho)$$

Now we use that $g_{k+1} = \Omega_N(\rho)\tilde{g}_{k+1}$ and $\Omega_N^I(\rho) = \Omega_N(\rho)$, with the latter equality arising from using $\Omega_N(\rho)$ as the normalisation for $\tilde{G}$. Dividing by $\Omega_N(\rho)$ gives

$$\frac{\Omega_M^I(\rho)}{\Omega_N^I(\rho)} = 1 + \sum_{k=1}^{J-1} \Big[F_k(\rho)^N - F_k(\rho)^M \Big] \tilde{g}_{k+1}(\rho) = \Upsilon(\rho).$$

Internal consistency requires the input ratio to equal the model-implied ratio, so $\rho = \Upsilon(\rho)$. Conversely, any admissible fixed point generates primitives whose implied sales-rate ratio equals the input ratio and therefore rationalises the observed transaction-price distributions. This proves (i) and (ii).

To prove (iii), suppose that Ω^* is infinite. Compactness of $\mathcal{R}$ implies that its zeros have an accumulation point $\rho_0 \in \mathcal{R}$. By Lemma 3, $\Upsilon(\rho) - \rho$ admits a real analytic extension to an open interval in $\mathbb{R}$ containing ρ_0. The identity theorem for real analytic functions implies that this extension vanishes on that interval. Hence $\Upsilon(\rho) - \rho$ vanishes on a nonempty open subinterval of $(1, \overline{\rho})$. The function $\Upsilon(\rho) - \rho$ is real analytic on the connected open interval $(1, \overline{\rho})$, so a second application of the identity theorem gives $\Upsilon(\rho) - \rho = 0$ throughout $(1, \overline{\rho})$. Continuity then gives $\Upsilon(1) - 1 = 0$.

This conclusion would make $\rho = 1$ a fixed point. However, for the candidate primitives recovered at $\rho = 1$,

$$\Omega_M^I(1) - \Omega_N^I(1) = \sum_{k=1}^{J} g_k(1) F_{k-1}(1)^N \left[1 - F_{k-1}(1)^{M-N} \right] > 0.$$

The inequality follows from admissibility: $g_k(1) > 0$ for some $k \geq 2$ and $0 < F_{k-1}(1) < 1$. Part (i) therefore gives $\Upsilon(1) = \Omega_M^I(1)/\Omega_N^I(1) > 1$, a contradiction. Thus $\mathbf{\Omega}^*$ is finite. ■

Proof of Theorem 3. Let $x \equiv F_{J-1}, \quad y \equiv F_{J-2}$. Since $G_{J-2} = G_{J-1} = G_J = 1$, the pmf equation (A.1) at the top two grid points imply

$$\Omega_n h_J^{(n)} = 1 - A_n(x),$$

and

$$\Omega_n h_{J-1}^{(n)} = A_n(x) - A_n(y).$$

Therefore the ratios

$$a \equiv \frac{h_J^{(N)}}{h_{J-1}^{(N)}}, \qquad b \equiv \frac{h_J^{(M)}}{h_{J-1}^{(M)}}$$

satisfy

$$a = \frac{1 - A_N(x)}{A_N(x) - A_N(y)}, \qquad b = \frac{1 - A_M(x)}{A_M(x) - A_M(y)}.$$

Equivalently, (x, y) solves

$$\Phi(x; y) \equiv (a+1)A_N(x) - 1 - aA_N(y) = 0, \tag{A.4}$$

$$\Psi(x; y) \equiv (b+1)A_M(x) - 1 - bA_M(y) = 0. \tag{A.5}$$

We first show that, for each $y \in (0, 1)$, equation (A.4) has a unique solution $x(y) \in (y, 1)$. Since

$$\Phi(y; y) = A_N(y) - 1 < 0,$$

and

$$\Phi(1; y) = a[1 - A_N(y)] > 0,$$

existence follows from continuity. Moreover,

$$\frac{\partial \Phi(x; y)}{\partial x} = (a+1)N(N-1)x^{N-2}(1-x) > 0$$

for $x \in (y, 1)$, so the solution is unique.

Next, $x(y)/y$ is strictly decreasing. Implicit differentiation gives

$$x'(y) = \frac{a}{a+1} \frac{y^{N-2}(1-y)}{x(y)^{N-2}(1-x(y))}. \tag{A.6}$$

Using equation (A.4),

$$\frac{a}{a+1} = \frac{1 - A_N(x)}{1 - A_N(y)}.$$

Thus $x'(y) < x(y)/y$ is equivalent to

$$[1 - A_N(x)]y^{N-1}(1-y) < [1 - A_N(y)]x^{N-1}(1-x).$$

After rearranging, this inequality is

$$C(x, y) > 0,$$

where

$$C(x, y) = -(1 - y^{N-1})x^N + (1 - y^N)x^{N-1} - y^{N-1}(1-y).$$

As a polynomial in x, C has at most two positive roots by Descartes' rule of signs. Since $C(y, y) = 0$ and $C(1, y) = 0$, these are its only positive roots. Finally,

$$C_x(1, y) = -(1 - A_N(y)) < 0,$$

so $C(x, y) > 0$ for all $x \in (y, 1)$. Therefore $x'(y) < x(y)/y$, and

$$\frac{d}{dy}\left(\frac{x(y)}{y}\right) < 0.$$

It remains to show that $\Psi(x(y); y) = 0$ has a unique solution in $(0, 1)$. Let

$$\widetilde{\Psi}(y) \equiv \Psi(x(y); y).$$

The function extends continuously to $y = 1$ with $x(1) = 1$, and $\widetilde{\Psi}(1) = 0$. Differentiating and using (A.6),

$$\widetilde{\Psi}'(y) = M(M-1)by^{M-2}(1-y)\left[\frac{b+1}{b}\frac{a}{a+1}\left(\frac{x(y)}{y}\right)^{M-N} - 1\right].$$

Because $M > N$ and $x(y)/y$ is strictly decreasing, the bracketed term is strictly decreasing in y. Hence $\widetilde{\Psi}$ has at most one turning point. The model is assumed to generate the data, so there exists a true $y_0 \in (0, 1)$ such that $\widetilde{\Psi}(y_0) = 0$. If there were another interior root, then the two interior roots together with the endpoint root at 1 would require at least two turning

points by Rolle's theorem, a contradiction. Thus y_0, and therefore $x(y_0)$, are uniquely identified.

With $x = F_{J-1}$ identified, the top grid point equations identify the sales rates:

$$\Omega_n = \frac{1 - A_n(F_{J-1})}{h_J^{(n)}}, \qquad n \in \{N, M\}.$$

Theorem 1 then identifies the remaining values of F and G. Equivalently, Lemma 4 recovers the lower grid points recursively once the top of the distribution and the sales rates are known. ■

Proof of Proposition 3. It suffices to construct a continuum of observationally equivalent primitives. Let buyer values have support $\{0, 1, 2\}$, with

$$\Pr(V = 0) = 1 - p, \qquad \Pr(V = 1) = pa, \qquad \Pr(V = 2) = p(1 - a),$$

and let the seller threshold equal 1 with probability r and 2 with probability $1 - r$. Thus, $1 - p$ is the buyer mass below the lower threshold support and $a = \Pr(V = 1 \mid V \geq 1)$.

Write $h_n = \Pr(P = 1 \mid \text{sale}, N = n)$. For one bidder,

$$h_1 = \frac{r}{1 - a(1 - r)}.$$

For two bidders, enumeration of the events producing transaction prices 1 and 2 gives

$$h_2 = \frac{[2 - p - p(1 - a)^2]r}{[2 - p - p(1 - a)^2]r + p(1 - a)^2 + 2(1 - a)[1 - p(1 - a)](1 - r)}.$$

Solving the first equation for r and substituting into the second yields

$$0 = h_1 a^2 + (h_2 - h_1 h_2 - 2h_1)a + (2h_1 - h_2) - \frac{2(h_1 - h_2)}{p}.$$

Consider the observable probabilities $h_1 = 2/3$ and $h_2 = 11/19$. The preceding equation becomes

$$38a^2 - 65a + 43 - \frac{10}{p} = 0.$$

For every $a \in [1/3, 2/3]$, define

$$p(a) = \frac{10}{38a^2 - 65a + 43}, \qquad r(a) = \frac{2(1 - a)}{3 - 2a}.$$

These quantities lie strictly between zero and one, and substitution into the original observation equations gives

$$h_1 = \frac{2}{3}, \qquad h_2 = \frac{11}{19}$$

for every a in the stated interval. Since prices have support $\{1, 2\}$, these probabilities determine the complete transaction-price distributions conditional on sale. Along this family, $F_0 = 1 - p(a)$, $F^+(1) = a$, and $G_1 = r(a)$. Each therefore varies while the observables remain fixed. Hence F, G, and F^+ are not point identified. ■

Before proving Proposition 4, we establish the following lemma that shows how top-down recursion recovers F and G given the sales rates. For example, if Ω_M and Ω_N are know via Theorem 3, top-down iteration recovers the primitives in a similar fashion to Theorem 1.

Lemma 4. *Suppose Ω_N and Ω_M are known for bidder counts $M > N$, and suppose (F_j, G_j) has been identified. Define*

$$R_j^{(n)} \equiv A_n(F_j)G_j - \Omega_n h_j^{(n)}.$$

If there is an admissible solution (F_{j-1}, G_{j-1}) with

$$0 < F_{j-1} < F_j, \qquad 0 \leq G_{j-1} \leq G_j,$$

then it is unique. In particular, F_{j-1} is the unique admissible root of

$$P_j(x) \equiv (M - N)G_j x^M - MR_j^{(N)} x^{M-N} + NR_j^{(M)} = 0, \tag{A.7}$$

and G_{j-1} is then recovered from

$$G_{j-1} = \frac{R_j^{(N)} - G_j F_{j-1}^N}{NF_{j-1}^{N-1}(1 - F_{j-1})}. \tag{A.8}$$

Proof. Let $x = F_{j-1}$ and $y = G_{j-1}$. Rearranging the mass equation (A.1) gives

$$R_j^{(n)} = G_j x^n + nx^{n-1}(1 - x)y.$$

For $n = N$,

$$y(x) = \frac{R_j^{(N)} - G_j x^N}{Nx^{N-1}(1 - x)}.$$

Substituting this expression into the equation for bidder count M yields (A.7). On the admissible region, $y(x) \geq 0$, so

$$R_j^{(N)} \geq G_j x^N.$$

Now

$$P_j'(x) = M(M-N)x^{M-N-1}\left(G_j x^N - R_j^{(N)}\right) \leq 0$$

on the admissible region. Thus P_j is monotone on the admissible region and has at most one admissible root. Once F_{j-1} is known, equation (A.8) pins down G_{j-1}. ■

Proof of Proposition 4. At the first grid point, equation (A.1) gives

$$\Omega_n h_1^{(n)} = A_n(F_1)G_1 - A_n(F_0)G_0 - F_0^n(G_1 - G_0).$$

If $G_0 = 0$, this reduces to

$$\Omega_n h_1^{(n)} = G_1\left[A_n(F_1) - F_0^n\right],$$

which gives

$$F_0 = \left[A_n(F_1) - \frac{\Omega_n h_1^{(n)}}{G_1}\right]^{1/n}.$$

This formula identifies only the total mass F_0, establishing (i).

If $F_0 = 0$ and $n \geq 2$, then $A_n(F_0) = 0$ and $F_0^n = 0$, so

$$\Omega_n h_1^{(n)} = A_n(F_1)G_1.$$

Thus G_0 drops out of the transaction price equation. The data can identify the cumulative threshold mass G_1, but not how much of that mass lies below the first buyer-support point, proving (ii).

Finally, when $n = 1$, $A_1(x) = 1$, and the mass equation becomes

$$\Omega_1 h_j^{(1)} = (1 - F_{j-1})(G_j - G_{j-1}).$$

This proves (iii). ■

Appendix B. Hermite approximation and implementation

Appendix B.1. Density normalisation and the transaction-price mapping

For each primitive $D \in \{F, G\}$, the reference density is proportional to $\phi((x-\mu_D)/\sigma_D)/\sigma_D$ on its integration interval. The overall scale of this reference cancels from equation (5). We divide by a normalising constant $C_D(a) = \int_{\mathcal{I}_D} r_D(u)\{[a'\psi_D(u)]^2 + \epsilon_D\}\, du$, adding a lower bound ϵ_D to the squared polynomial term in implementation. The implementation evaluates the integral and its cumulative counterpart by the trapezoidal rule, and divides the cumulative integral by its terminal value. The resulting CDF takes values zero and one at the endpoints of the interval.

To evaluate the transaction CDF, write F_j and G_j for the primitive CDFs on the grid used for the auction mapping, and set $A_j^{(n)} = nF_j^{n-1} - (n-1)F_j^n$. With $A_0^{(n)} = F_0 = G_0 = 0$, define the unnormalized transaction masses

$$m_j^{(n)} = A_j^{(n)} G_j - A_{j-1}^{(n)} G_{j-1} - F_{j-1}^n (G_j - G_{j-1}), \qquad \Omega_n = \sum_j m_j^{(n)}. \tag{B.1}$$

The implied transaction CDF is $H_j^{(n)} = \sum_{k \le j} m_k^{(n)} / \Omega_n$, which implements equation (1).

Appendix B.2. Monte Carlo implementation

For both bidder-count pairs used in the simulations, the preliminary buyer CDF is $F_j^0 = 1 - \sqrt{1 - \widehat{H}_j^{(2)}}$, obtained by ignoring seller rejections in two-bidder auctions. The preliminary seller CDF is $G_j^0 = \max\{\widehat{H}_j^{(\min \mathcal{N})}, F_j^0\}$. For each primitive, the reference location is the median of its preliminary CDF and its scale is the difference between its 90th and 10th percentiles divided by $\Phi^{-1}(0.9) - \Phi^{-1}(0.1)$. Quantiles are interpolated on the objective grid after enforcing monotonicity. These reference parameters remain fixed during joint estimation.

The basis uses the probabilists' Hermite polynomials, with $\mathrm{He}_0(z) = 1$, $\mathrm{He}_1(z) = z$, and $\mathrm{He}_{k+1}(z) = z\,\mathrm{He}_k(z) - k\,\mathrm{He}_{k-1}(z)$, scaled as $\mathrm{He}_k(z)/\sqrt{k!}$. We set $K = 4$ and $\epsilon_D = 10^{-10}$. All three distribution cases use 30 objective nodes.

Appendix B.3. Housing implementation

The first stage uses successful auctions with observed transaction-to-commitment price ratios $X_t = b_t/c_t$ and bidder counts from one to ten. For bidder count n, the raw conditional

CDF estimate is

$$\widehat{H}^{(n)}(x \mid c) = \frac{\sum_{t=1}^{T} \mathbf{1}\{N_t = n\}\kappa((c_t - c)/b_n)\mathbf{1}\{X_t \le x\}}{\sum_{t=1}^{T} \mathbf{1}\{N_t = n\}\kappa((c_t - c)/b_n)}, \tag{B.2}$$

where κ is the standard Normal kernel and b_n is the bidder-count-specific bandwidth. The smoothing variable is commitment price in levels. The bandwidth rule is $b_n = 1.06 \min\{s_n, \mathrm{IQR}_n / 1.349\} T_n^{-1}$ where s_n and IQR_n are computed within the bidder count n. The commitment grid has 15 quantile-based nodes. The ratio grid consists of 100 equally spaced points with the value one included.

Reference densities are calculated once from preliminary primitive CDFs following the description above. First average $\hat{H}$ equally across commitment nodes. The preliminary buyer CDF uses the average ten-bidder transaction CDF. The preliminary seller CDF is the pointwise maximum of that buyer CDF and the average one-bidder transaction CDF, set to one at ratios of one and above.

We orthonormalise each polynomial basis with respect to the reference density over the corresponding support. Let $p_D(u)$ represent the set of K standardised Hermite polynomials divided by the square roots of their factorials. At integration nodes u_m with trapezoidal weights q_m, normalise r_D so that $\sum_m q_m r_D(u_m) = 1$, and form

$$M_D = \sum_m q_m r_D(u_m) p_D(u_m) p_D(u_m)' = R_D' R_D, \qquad \psi_D(u) = R_D^{-\prime} p_D(u), \tag{B.3}$$

where R_D is the upper-triangular Cholesky factor. Then $\sum_m q_m r_D(u_m) \psi_D(u_m) \psi_D(u_m)' = I$. With $\|a_D(c)\| = 1$, the density integrates to one, and the basis transformation preserves the degree-four polynomial span.

The coefficient matrices in equation (7) are parameterised by

$$\mathrm{vec}(\beta_D) = \frac{(1, B_D)}{\sqrt{1 + \|B_D\|^2}}, \qquad B_D \in \mathbb{R}^{14}, \tag{B.4}$$

using columnwise vectorisation. Thus the norm of β_D is one and the coefficient on the constant basis term at the constant covariate term is positive. This fixes the matrices' common scale and sign.

Appendix B.4. Full sample moments

For a given bidder count n and commitment price c, write $F_j = F(x_j|c)$ and $G_j = G(x_j|c)$. The joint masses of failure and a threshold or highest buyer value at x_j are $v_j^{(n)}(c) = F_{j-1}^n(G_j - G_{j-1})$ and $b_j^{(n)}(c) = (F_j^n - F_{j-1}^n)(1 - G_j)$, with $F_0 = G_0 = 0$ (noting that a threshold tied with a buyer value results in a sale). The first expression requires all buyers to lie strictly below the threshold. The second requires the threshold to exceed the highest buyer value. Each mass sequence sums separately to $1 - \Omega_n(c)$, where $\Omega_n(c) = \sum_j G_j(F_j^n - F_{j-1}^n)$. Thus

$$\begin{aligned} H_{\mathrm{v},j}^{(n)}(c;\beta) &= \frac{\sum_{k=1}^{j} v_k^{(n)}(c;\beta)}{1 - \Omega_n(c;\beta)}, \\ H_{\mathrm{b},j}^{(n)}(c;\beta) &= \frac{\sum_{k=1}^{j} b_k^{(n)}(c;\beta)}{1 - \Omega_n(c;\beta)}. \end{aligned}$$

For $r \in \{\mathrm{s}, \mathrm{v}, \mathrm{b}\}$, define the CDF distance at (n, ℓ) by

$$D_{n\ell}^r(\beta) = \frac{1}{J} \sum_{j=1}^{J} \left[\widehat{H}_{r,j}^{(n)}(c_\ell) - H_{r,j}^{(n)}(c_\ell;\beta) \right]^2,$$

where $H_{\mathrm{s}}^{(n)}$ is the transaction CDF as already defined. Let $\mathcal{A}$ denote the (n, ℓ) nodes with both failure CDFs available. For each bidder count, let T_n^{a}, T_n^{s}, T_n^{v} and T_n^{b} be the counts of all auctions, sales, vendor failures and non-vendor failures. Set $T_{\max} = \max_n T_n^{\mathrm{a}}$ and $w_n^r = T_n^r / T_{\max}$. The objective function in equation (9) is

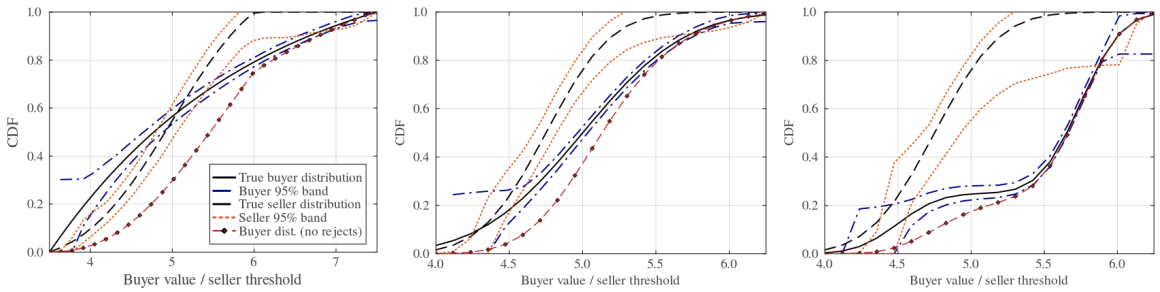
$$\begin{aligned} Q_{\mathrm{full}}(\beta) = \frac{1}{L|\mathcal{N}|} \Bigg\{ & \sum_{n \in \mathcal{N}} \sum_{\ell=1}^{L} \left[w_n^{\mathrm{s}} D_{n\ell}^{\mathrm{s}}(\beta) + w_n^{\mathrm{a}} \left\{ \widehat{\Omega}_n(c_\ell) - \Omega_n(c_\ell;\beta) \right\}^2 \right] \\ & + \sum_{(n,\ell) \in \mathcal{A}} \left[w_n^{\mathrm{v}} D_{n\ell}^{\mathrm{v}}(\beta) + w_n^{\mathrm{b}} D_{n\ell}^{\mathrm{b}}(\beta) \right] \Bigg\}. \end{aligned}$$

Appendix C. Additional simulations

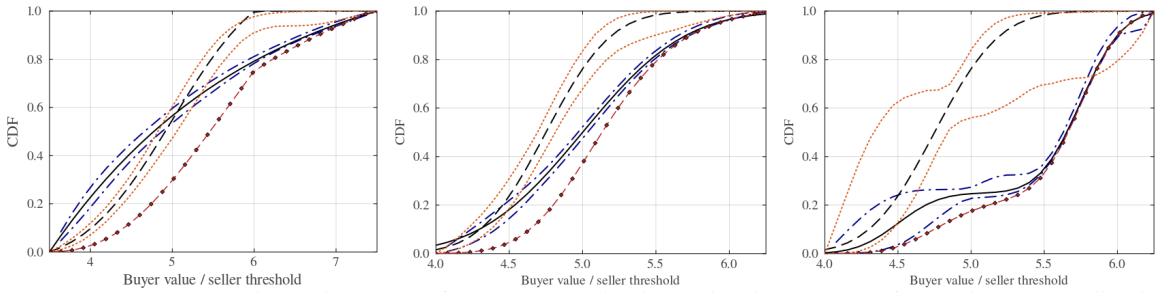
Figures C.8 and C.9 repeat the main Monte Carlo comparison for various bidder count combinations and sample sizes.

Figure C.8: Monte Carlo simulations, $(M, N) = (4, 2)$, 5000 auctions

(a) Low sales rate, Omega **(b)** Med sales rate, Omega **(c)** High sales rate, Omega



(d) Low sales rate, Hermite **(e)** Med sales rate, Hermite **(f)** High sales rate, Hermite

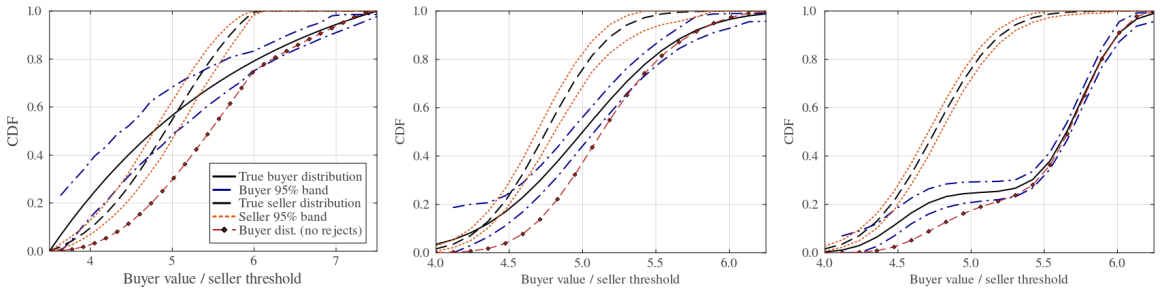


Notes: Monte Carlo simulation results from 1,000 replications with 5,000 successful auctions per replication. The top panels show the Omega-based estimator, while the bottom panels show the Hermite polynomial-based estimator.

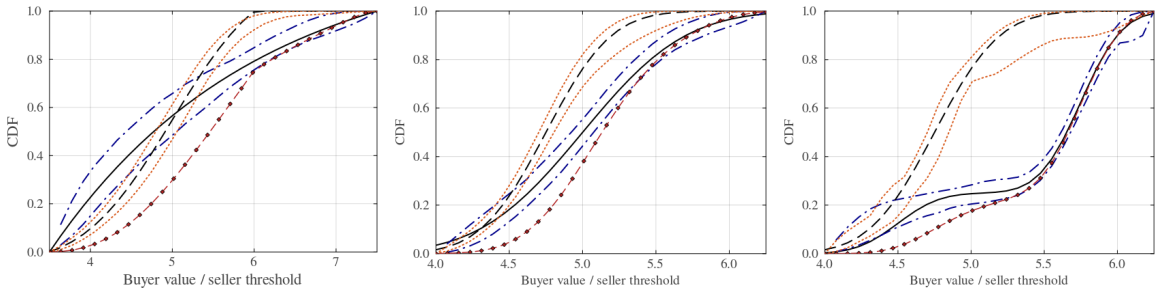
Figure C.9: Monte Carlo simulations, $(M, N) = (2, 1)$

T = 1,000

(a) Low sales rate, Omega **(b)** Med sales rate, Omega **(c)** High sales rate, Omega

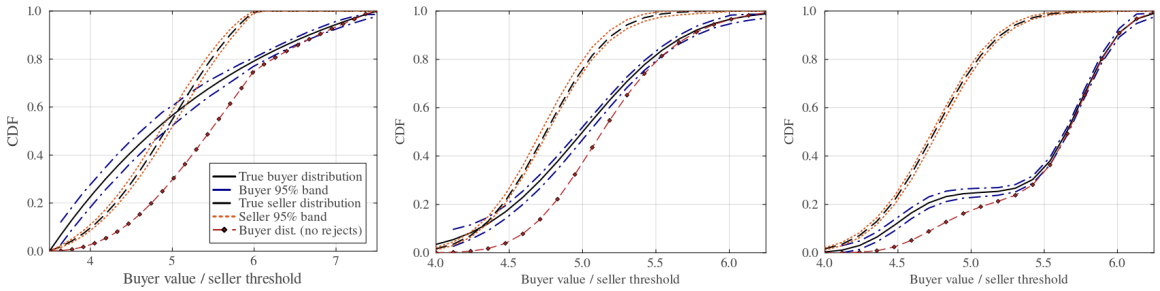


(d) Low sales rate, Hermite **(e)** Med sales rate, Hermite **(f)** High sales rate, Hermite

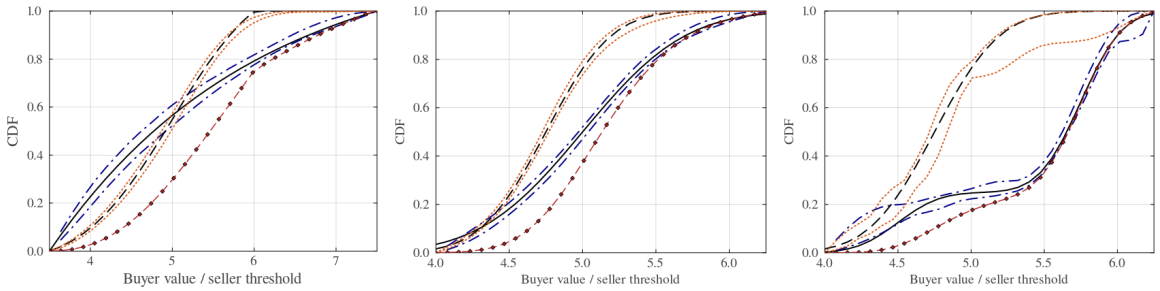


T = 5,000

(g) Low sales rate, Omega **(h)** Med sales rate, Omega **(i)** High sales rate, Omega



(j) Low sales rate, Hermite **(k)** Med sales rate, Hermite **(l)** High sales rate, Hermite



Notes: Monte Carlo simulation results for $(M, N) = (2, 1)$.

Appendix D. Selective entry and other results

Appendix D.1. Selective entry

Condition throughout on a known number $K \geq 2$ of potential bidders. Potential bidder i has pre-entry signal Q_i and value V_i; bidder i learns her value only after entry. The realised number of entrants is N. Let Z be an observed participation shifter, V_T the seller threshold drawn from distribution G assumed independent of all bidder value-signal pairs, and S the sale indicator. Auction outcomes follow Assumption 2. Conditional on a sale, the econometrician observes N, Z, and the transaction price. For each z, define

$$\pi_n(z) \equiv \Pr(N = n \mid S = 1, Z = z, K), \quad \Omega_n(z) \equiv \Pr(S = 1 \mid N = n, Z = z, K).$$

The potential bidder conditional value CDF and the value CDF selected by a cutoff s are, respectively,

$$F(v \mid q) \equiv \Pr(V_i \leq v \mid Q_i = q, K), \quad F^{\text{ent}}(v; s) \equiv \Pr(V_i \leq v \mid Q_i \geq s, K).$$

We now formalise the entry stage assumptions; these largely follow Gentry and Li (2014), with two main differences. The first is that, rather than derive an equilibrium entry threshold, we assume that entry follows a threshold cutoff rule for simplicity. The second is that we assume Z shifts participation probabilities directly, rather than operating through an entry cost. These reduced-form simplifications of their framework enable a more concise description of the potential bidder value distribution identification result, but preclude identification of entry costs.

Assumption 7. *Conditional on a known number of potential bidders K:*

(i) conditional on $(Z = z, K)$, $\{(V_i, Q_i)\}_{i=1}^K$ are independently and identically distributed, their common joint distribution does not depend on z, and Q_i is uniform on $[0, 1]$;

(ii) conditional on $(Z = z, K)$, V_T is independent of all bidder value-signal pairs and has CDF G, which does not depend on z;

(iii) for every $q' \geq q$, $F(v|q') \leq F(v|q)$ for all v;

(iv) potential bidder i observes Q_i and Z, but not V_i, before entry, and entry follows a threshold cutoff rule: for each z there exists $s_K(z)$ such that i enters if and only if $Q_i \geq s_K(z)$; and

(v) the set of cutoffs $\mathcal{S}_K \equiv \{s_K(z) : z \in \text{supp}(Z \mid K), s_K(z) \in (0,1)\}$ *satisfies cl*$(\mathcal{S}_K) = [0,1]$.

The last condition requires that the instrument must induce cutoff entry values whose closure covers the signal support. This allows pointwise identification almost everywhere; otherwise, bounding approaches are needed.

Next we assume that we can identify the F^{ent} and G from the transaction only data (e.g., the sufficient condition of Assumption 5 holds for each z):

Assumption 8. *For every* z *such that* $s_K(z) \in (0,1)$*, transaction prices and bidder counts* $n < m$ *are observed, and* $F^{ent}(\cdot; s_K(z))$*,* G*, and* $\Omega_n(z), \Omega_m(z)$ *are point identified.*

Under Assumption 7, the independence implication used by Gentry and Li (2014) applies directly. At cutoff $s_K(z)$, a representative entrant has value CDF $F^{\text{ent}}(\cdot; s_K(z))$. Because value-signal pairs are conditionally independent across potential bidders given $Z = z$ and entry depends only on a bidder's own signal, conditional on $N = n$ the n entrant values are independent draws from this CDF. Under the seller threshold independence assumption, V_T is also independent of entrant values. The conditional transaction price equation (1) can therefore be applied with buyer CDF $F^{\text{ent}}(\cdot; s_K(z))$. The next proposition recovers the entry cutoff from the transaction data.

Lemma 5. *Suppose Assumptions 7 and 8 hold. Fix* z *and choose any* $1 \leq n < m \leq K$ *such that* $\pi_n(z), \pi_m(z) > 0$*. Then the transaction data identify the entry probability* $e(z)$*, the entry cutoff* $s_K(z) = 1 - e(z)$*, and the cutoff set* $\mathcal{S}_K$*.*

Proof. Let $p_n(z) \equiv \Pr(N = n \mid Z = z, K)$ denote the entrant-count distribution before conditioning on sale. Proposition 5 gives $\pi_n(z) = (p_n(z)\Omega_n(z))/[\sum_{\ell=1}^{K} p_\ell(z)\Omega_\ell(z)]$. The common denominator cancels when taking the ratio across m and n, so $p_m(z)/p_n(z) = [\pi_m(z)/\Omega_m(z)]/[\pi_n(z)/\Omega_n(z)]$. Independent cutoff entry and uniform signals imply that N is binomial with K trials and success probability $e(z) = 1 - s_K(z)$. Hence

$$\frac{\pi_m(z)/\Omega_m(z)}{\pi_n(z)/\Omega_n(z)} = \frac{\binom{K}{m}}{\binom{K}{n}} \left(\frac{e(z)}{1-e(z)}\right)^{m-n}.$$

Solving yields $e(z)$. Repeating the argument over the support of Z identifies $\mathcal{S}_K$. ■

We can now apply the identification argument of Proposition 2 in Gentry and Li (2014).

Corollary 2. *Suppose Assumptions 7 and 8 hold. Then* $F(x_j \mid q) = \Pr(V_i \leq x_j \mid Q_i = q, K)$ *is point identified for every* $j = 1, \ldots, J$ *and almost every* $q \in (0,1)$*.*

Proof. Lemma 5 and Assumption 8 identify $\{s, F^{\text{ent}}(\cdot; s)\}_{s \in \mathcal{S}_K}$. Assumption 7 gives the affiliated signal structure and $\text{cl}(\mathcal{S}_K) = [0, 1]$. The result therefore follows from their argument: take the selection identity $(1 - s)F^{ent}(x; s) = \int_s^1 F(x|q)dq$, noting that the left-hand side is identified for every $s \in \mathcal{S}_K$ and continuous in s, and $\text{cl}(\mathcal{S}_K) = [0, 1]$ implies it is identified on $[0, 1]$. Differentiating with respect to s gives $F(x|s) = -\frac{\partial}{\partial s}[(1 - s)F^{ent}(x; s)]$. Because the selection inversion is pointwise in values, finite support of V_i does not affect the argument. ■

Appendix D.2. Other results

Proposition 6. *For any $M > N \geq 1$ with $\Omega_N > 0$,*

$$1 \leq \frac{\Omega_M}{\Omega_N} \leq \frac{M}{N}.$$

These bounds are tight. The lower bound is sharp if all auctions result in sale.

Proof of Proposition 6. Write $\Omega_n = \sum_{j=1}^{J} g_j(1 - F_{j-1}^n)$. Since $x^M \leq x^N$ on $[0, 1]$, $\Omega_M \geq \Omega_N$. For the upper bound, the average of the first n terms of the non-increasing sequence $1, x, x^2, \ldots$ is non-increasing in n. The geometric sum identity therefore gives

$$\frac{1 - x^M}{M} \leq \frac{1 - x^N}{N}, \qquad 0 \leq x \leq 1.$$

Substitution of $x = F_{j-1}$ and summation with weights g_j yield $\Omega_M/M \leq \Omega_N/N$, proving the upper bound.

The lower bound is attained when $G_1 = 1$. To establish upper-bound tightness while retaining positive threshold mass at both points, take $F_1 = 1 - \epsilon$ and $G_1 = \epsilon^2$ in the two-point model. Then

$$\frac{\Omega_n}{\epsilon} = \sum_{r=0}^{n-1}(1 - \epsilon)^r + \epsilon(1 - \epsilon)^n \longrightarrow n$$

so $\Omega_M/\Omega_N \to M/N$ as $\epsilon \downarrow 0$. ■

Proposition 7. *If the transaction price distributions for two distinct bidder levels $M > N \geq 1$ are observed, then Υ is a contraction mapping if there exists $C \in (0, 1)$ such that*

$$\left|\sum_{k=1}^{J-1}\left\{g_k(\rho)\tilde{F}_k(\rho)'\left[N\tilde{F}_k^{N-1}(\rho) - M\tilde{F}_k^{M-1}(\rho)\right] + g_k'(\rho)\left[\tilde{F}_k^N(\rho) - \tilde{F}_k^M(\rho)\right]\right\}\right| \leq C \quad \text{(D.1)}$$

for all $\rho \in [\underline{\Omega}, \overline{\Omega}]$ *where* $\{g_k(\rho), F_k(\rho)\}$ *are known functions of* ρ *given the transaction price distributions* $\{H_k^{(M)}, H_k^{(M)}\}_{k=1}^J$.

Proof of Proposition 7. Starting with the sales rate ratio mapping (2)

$$\Upsilon(\rho) = 1 + \sum_{k=1}^{J-1} \left[F_k^N(\rho) - F_k^M(\rho)\right] g_{k+1}(\rho)$$

A sufficient condition for Υ to be a contraction is that there exists some positive $C < 1$ such that the absolute value of the derivative does not exceed C. Taking the derivative with respect to ρ gives

$$\Upsilon'(\rho) = \sum_{k=1}^{J-1} \left\{\tilde{g}_k(\rho)\tilde{F}_k'(\rho)\left[N\tilde{F}_k^{N-1}(\rho) - M\tilde{F}_k^{M-1}(\rho)\right] + \tilde{g}_k'(\rho)\left[\tilde{F}_k^N(\rho) - \tilde{F}_k^M(\rho)\right]\right\}$$

■

If $N = 1$ transactions are observed, the sufficient condition can be substantially simplified, which we demonstrate below for the $M = 2$, $N = 1$ case.

Corollary 3. *If the transaction price distributions for* $M = 2$ *and* $N = 1$ *are known, then* F *is point identified if*

$$\left|\sum_{k=0}^{J-1} h_1^{k+1} \frac{\partial F_k(\rho)}{\partial \rho}\right| \leq C$$

for some positive $C < 1$ *for all* $\rho \in (1, 2)$*, where* $F(\rho)$ *is the buyer value distribution associated with sales rate ratio of* ρ

Proof. Using $g_{k+1}(\rho) = \frac{h_{k+1}^{(1)}}{1-F_k(\rho)}$ gives

$$\begin{aligned}
\Upsilon(\rho) &= \sum_{k=0}^{J-1} \frac{h_{k+1}^{(1)}}{1 - F_k(\rho)} - \frac{F_k^2(\rho) h_{k+1}^{(1)}}{1 - F_k(\rho)} \\
&= \sum_{k=0}^{J-1} \frac{h_{k+1}^{(1)}(1 + F_k(\rho))(1 - F_k(\rho))}{1 - F_k(\rho)} \\
&= 1 + \sum_{k=0}^{J-1} h_{k+1}^{(1)} F_k(\rho)
\end{aligned}$$

Taking the absolute value of the derivative with respect to ρ gives the result. ■

Closed-form expression for Theorem 3 when $M = 3, N = 2$. The system of equations in

the proof of Theorem 3 becomes

$$\begin{aligned} a\,(A_2(x) - A_2(y)) &= 1 - A_2(x), \\ b\,(A_3(x) - A_3(y)) &= 1 - A_3(x), \end{aligned}$$

Rearranging the first equation gives $a(1-y^2) = (a+1)(1-x)^2$. Taking the positive square root gives $y = 1 - c(1-x)$ where $c \equiv \sqrt{1 + (1/a)}$.

Next, we factor $1 - A_3(x) = 1 - 3x^2 + 2x^3 = (1-x)^2(1+2x)$. Using this in the second equation gives

$$b(1-y)^2(1+2y) = (b+1)(1-x)^2(1+2x)$$

Substituting $1 - y = c(1-x)$ and cancelling the $(1-x)^2$ factor gives $bc^2(1+2y) = (b+1)(1+2x)$. Substituting $y = 1 - c(1-x)$ gives

$$bc^2(3 - 2c + 2cx) = (b+1)(1+2x),$$

a linear equation in x. Solving gives

$$x = 1 - \frac{3(1 - b/a)}{2[1 - b(c^3 - 1)]}, \quad y = 1 - c(1-x)$$