# Sample spacings for identification: the case of English auctions with absentee bidding

Marleen Marra*

April 14, 2021

**Abstract**

This paper presents new nonparametric identification results for IPV ascending auctions when the number of bidders is unknown. The setting is motivated by the limited information observable in English auctions with absentee bidding, but results apply generally when besides the transaction price also the highest bids of two losing bidders is known. A particularly novel feature of the paper is that it exploits insights from the statistics literature about the stochastic ordering of sample spacings. In an incomplete model with weak shape restrictions on the latent value distribution, such spacings are shown to set-identify expected bidder surplus and seller revenue. Applying the method to a small sample of wine auctions, informative bounds on policy-relevant counterfactuals are estimated. It turns out that Sotheby's restricts full exploitation of the exclusion principle of optimal reserve prices, so that sellers set sub-optimally low reserves. They would also benefit between 2.5-9 (up to 13) percent from adopting a common reserve price rule equal to 110 (120) percent of the Wine Department's pre-auction value estimate. (JEL codes: D44, C01, C46, C57)

KEYWORDS: Nonparametric set-identification, English auctions, order statistics, shape restrictions, optimal reserve price

**Department of Economics, Sciences Po, 28 Rue des Saints-Pères, 75007 Paris, France. E-mail: marleen.marra@sciencespo.fr* I would like to thank Adam Rosen, Andrew Chesher, Phil Haile, Suehyun Kwon, Lars Nesheim, Áureo de Paula, anonymous reviewers, an editor, and conference and seminar participants for helpful comments. Thanks also to Sebastian Fahey for going over practicalities of Sotheby's auctions and to Alan Crawford for introducing us. All errors are my own. This work was supported by a PhD Studentship from the Economic and Social Research Council.

# 1 Introduction

English auctions can be notoriously secretive.[1] Motivated by the limited information observable in English auctions with absentee bidding, I show how so-called *sample spacings* (Pyke (1965, 1972)), i.e. the difference between adjacent order statistics, deliver nonparametric set-identification of structural features of interest. The structural analysis of IPV English auction data typically relies on knowing the number of bidders and at least one bid order statistic (Athey and Haile (2002)). When bid data does not contain the number of bidders, the known mapping of the distribution of an order statistic from an i.i.d. sample *of known size* and its parent distribution cannot be applied. In this paper, I propose a new method that relies on the stochastic difference between adjacent order statistics, which contain previously unexplored identifying information. Results are presented for an incomplete IPV English auction model without unobserved heterogeneity, as in Haile and Tamer (2003), but due to the more limited information setting focuses on directly identifying structural features of interest.

The new set-identification method can be summarized in one paragraph. Observing both the transaction price and the highest bids of two losing bidders is informative about the stochastic spacing between the second- and third-highest valuation (e.g. the second-to-last spacing) even without knowing the number of bidders. Identification relies on facts from the statistics literature: that properly normalized spacings from distribution functions with increasing failure rates are stochastically decreasing (Barlow and Proschan (1966)). This bounds the last spacing, and also set-identifies (counterfactual) surplus and revenue. Providing an upper bound on the last spacing is crucial because of the well-known issue in English auctions that the auction stops when the second-highest bidder drops out, so that the willingness to pay of the winning bidder is never revealed.

The case of English auctions with absentee bidding is a fitting example of a setting where information revealed by sample spacings can benefit structural analysis: the number of bidders is typically unknown but highest bids of two distinct losing bidders can be identified. Absentee bidders report their maximum willingness to pay to the auctioneer, who then bids on their behalf during the auction.[2] The proxy bidding

[1] See e.g. Cassady (1967) and Akbarpour and Li (2019).

[2] In my empirical setting, non-absentee ("live") bidders do not have to be present in the auction room but can also bid online or by phone. Absentee bidding in English auctions is discussed pre-

system on eBay and other online auction environments is essentially an absentee bidding tool as well. Typical of English auctions is that bidders may not all place just one bid at their maximum valuation, and some may choose not to bid at all. Hence, one cannot simply recover all bidders' highest bids nor the number of bidders from a vector of observed bids.[3] As bidder identities are unknown it is furthermore impossible to determine the highest bids of all bidders who do submit a bid. However, as in the incomplete English auction model of Haile and Tamer (2003), the transaction price (plus increment) can be used to bound the highest two valuations. Bids placed by absentee bidders reveal another distinct highest bid needed to construct bounds on adjacent sample spacings, set-identifying structural features of interest.

The limited information content described above diverges from what is assumed known in previous ascending auction studies, including Paarsch (1997), Quint (2008), and Platt (2017), relying on the button-auction format and thus using all bidders' drop-out values and the number of bidders, Haile and Tamer (2003) and Chesher and Rosen (2015, 2017) using all bidders' highest bids and the number of bidders, Song (2004), Mbakop (2017), Freyberger and Larsen (2017), and Luo and Xiao (2019) using a vector of bids that includes at least the second and third-highest drop-out values, and Aradillas-López, Gandhi, and Quint (2013) and Coey, Larsen, Sweeney, and Waisman (2017) using the second highest drop-out value and number of bidders. These papers are all innovative in their econometric use of bid data from ascending auctions and some are derived for more general auction settings, but their identification strategies cannot be applied to the limited data central to this paper.

For the analysis developed here, one needs to observe only: 1) a vector of bids and 2) which bids are submitted by absentee bidders. Results apply in general when besides the transaction price also the highest bids from two losing bidders is observed. Bids are not assumed to equal the maximum willingness to pay of the bidders who submit them.

To underscore its practical use in overcoming data limitations, I apply the method to an original dataset of fine wine auctions with absentee bidding collected at Sotheby's.

viously in: Rothkopf, Teisberg, and Kahn (1990), Thiel and Petry (1995), Ginsburgh (1998), and Akbarpour and Li (2019). Lucking-reiley (2000) finds that this practice has been used since at least 1878 for stamp auctions.

[3] Different from the highest bid in an auction, a "bidder's highest bid" is quite literally the highest bid placed by a given bidder in that auction. I reserve the term "drop-out value" for when a bid is assumed to equal that bidder's valuation - as in the button-auction format.

Leveraging information from bids placed by absentee bidders and the second-highest bidder, expected bidder surplus is estimated without knowing the number of bidders. Empirical results show that expected winning bidder surplus is higher for high-end wines, and lies between 75 to 125 percent of the average highest bid in the full sample. More fundamentally, this empirical application shows that the sample spacings method delivers informative bounds in small samples with sizeable bidding increments.

This setting is also interesting from a policy angle as Sotheby's allows sellers to set a secret reserve price, but restricts it to be less than the low bound of the Wine Department's estimated value bracket. My estimates show that this results in suboptimally low reserves, which precludes sellers from leveraging the *exclusion principle* of optimal reserve prices (see Krishna (2009), based on Myerson (1981) and Riley and Samuelson (1981)). In other words, too much weight is placed on the sale probability and too little on revenues conditional on a sale. I consider a simple counterfactual policy where the reserve price is automatically set at $\alpha$ times the pre-auction value estimate, and show that expected seller revenue increases by 2.5-9 percent when setting $\alpha = 1.1$ and up to 13 percent when setting $\alpha = 1.2$. The latter upper bound corresponds to roughly 25.000 pounds of additional seller revenue for a single day of wine auctions, of which the auction house holds about 20 per year in its London branch alone. Also the lower bound is non-negative for any $\alpha \in [0.75, 1.2]$, indicating that it is not only the binding value bracket constraint that delivers low sale revenues. Even automatic reserve prices at a value of $\alpha \leq 1$ increase expected seller revenue compared to the current policy where sellers set reserves individually.

**Related literature.** The results in this paper build on the Haile and Tamer (2003) model but relax that the number of bidders and their highest bids are observed, contributing to the structural analysis of bid data with incomplete auction models. Within that literature, Quint (2008), Aradillas-López et al. (2013), and Coey et al. (2017), and for first-price auctions Tang (2011) are among the most related. They also provide methods to directly (set-)identify structural features of interest rather than the latent value distribution. Aradillas-López et al. (2013) relaxes the usual IPV ascending auction model to allow for correlated private values, and uses exogenous variation in the number of bidders to set-identify expected bidder and seller surplus. They build on Quint (2008), who develops similar results for a less general ascending auction model with symmetric affiliated private values. Coey et al. (2017) extend the Aradillas-López et al. (2013) model to one with potentially unobserved bidder asymmetries, and show that observing bidder types is especially useful to narrow estimated bounds that otherwise require significant variation in the number of bidders to shrink. In the case where valuations are increasing in the number of bidders, only upper bounds on expected seller and bidder surplus can be identified. Furthermore, Coey, Larsen, Sweeney, and Waisman (2018) focus on online ascending auctions and assume that the highest two values are revealed from the highest two bids. This allows them to directly estimate the optimal reserve price, without needing to know the number of bidders and allowing for asymmetric and correlated private values.

Song (2004) also studies a case with an unknown number of bidders. Similar to how absentee bidding reveals additional information in traditional English auctions, she exploits a feature of eBay auctions with fixed closing times to pin down two adjacent value order statistics. My model is more general as it requires only that besides the transaction price the highest bid two losing bidders is observed, while the assumption that the second- and third-highest values are known exactly delivers point-identification in Song (2004). Other auction papers assuming that (at least) two value order statistics are known, to address unobserved heterogeneity, include Mbakop (2017), Freyberger and Larsen (2017), and Luo and Xiao (2019). Only a lower bound on the third-highest and second-highest value is needed to set-identify expected bidder surplus and seller revenue in this paper, but I show how sample spacings deliver an additional bound for the case where the third-highest valuation is known exactly.

While this is not the first paper to rely on multiple order statistics, or on non-ordered measurements (e.g. Li and Vuong (1998) and Krasnokutskaya and Seim (2011)) to support identification in (auction) models, the innovation of my approach is that it relies on shape restrictions and their relation to the way adjacent order statistics are spaced out. This adds previously unexploited identifying information that helps facilitate structural analysis. Also related is recent work that uses shape properties to aid identification and estimation in auctions, notably Larsen and Zhang (2018) and Pinkse and Schurter (2019). Furthermore, Coey, Larsen, and Sweeney (2019) use the difference between the expected values of the second-highest and third-highest valuation to test whether the number of bidders varies exogenously.

The paper proceeds as follows. After introducing the auction model in section 2, the motivating setting with absentee bidding is described in section 3. Section 4 sets out the main results. An application in section 5 illustrates the method's relevance for structural analysis of incomplete bid data. Section 6 concludes.

## 2 Auction model

The mechanism is a standard English auction with a flexible closing rule ("going once, going twice"), fixed bidding increments, and a secret reserve price.[4] Ex-ante symmetric, risk neutral bidders face negligible entry and bidding cost. I follow the convention to denote random variables in upper case and their realizations in lower case. The empirical analysis controls for auction-level heterogeneity, so values should be interpreted as conditional values, and the analysis assumes no unobserved heterogeneity. The unobserved number of bidders $n$ is constant across auctions, and crucially the incomplete model does not require everyone to bid.[5] Let $i = 1, ..., n$ index bidders, and $V_i$ bidder $i$'s valuation with CDF $F_{V_i}$ and PDF $f_{V_i}$, and joint CDF $F_{\mathbf{V}} = F_{V_1,...,V_n}$. Throughout, it is maintained that:

**Assumption 1.** *All $n$ bidders symmetrically and independently draw values from a common value distribution, such that:*
*i) $F_{V_i} = F_V$, $\forall i = \{1, .., n\}$ (exchangeability)*
*ii) $F_{\mathbf{V}} = F_V^n$ (independence)*
*iii) $F_V$ is absolutely continuous and is defined on bounded support $[0, \bar{v}]$ (regularity conditions)*

This is the symmetric IPV assumption that is the main tenet in the structural analysis of English auctions (Paarsch and Hong (2006)). Exchangeability allows me to drop the $i$ subscript to let $V \sim F_V$ denote the valuation of any bidder. In addition, I impose a shape restriction that rules out distribution functions that, from a reliability theory perspective, age weakly slower than the exponential distribution:

[4] In English auctions with fixed bidding increments the auctioneer raises the price in increments $\Delta \geq 0$, the highest bidder wins the auction and the transaction price is the second-highest bid $+\Delta$. If the transaction price is less than the secret reserve price, the item remains unsold.

[5] This assumption is justified by the absence of entry cost and the secret (and hence ex-ante non-binding) reserve price. The method proposed by Song (2004) instead allows the number of bidders to vary exogenously across auctions but requires the three highest values to be pinned down by bid data.

**Assumption 2.** $\frac{f_V(v)}{1-F_V(v)}$ *weakly increases in* $v$*,* $\forall v \in [0, \bar{v}]$ *(increasing failure rate)*

The increasing failure rate (IFR) assumption is already commonplace in the empirical auction literature, as it is necessary to derive a unique optimal reserve price (Myerson (1981) and Riley and Samuelson (1981)). Haile and Tamer (2003) for instance use it to derive bounds on the optimal reserve price based on their (fully nonparametric) bounds on the latent value distribution.[6] IFR also captures most parametric distribution functions found realistic in applications, including the Normal, Exponential, Logistic, Extreme Value, Weibull (shape parameter $\geq 1$), Gamma (shape parameter $\geq 1$), and Beta (shape parameter $\geq 1$).[7]

The main identification result of the paper (Lemma 3) relies on IFR. Lemma 4 furthermore shows how to apply a second shape restriction to provide a richer illustration of the identifying power of sample spacings. It rules out distribution functions that age slower than the uniform distribution.

I furthermore impose the intuitive behavioral assumptions that define bidding strategies in the incomplete model of Haile and Tamer (2003):

**Assumption 3.** *Bidders do not bid more than they are willing to pay, and do not let an opponent win at a price they are willing to beat.*

This is satisfied in all symmetric separating equilibria of the Milgrom and Weber (1982) button auction model (Bikhchandani, Haile, and Riley (2002)), but also allows for alternative behavior including not bidding at all or bidding less than one's valuation. However, I allow the number of bidders to be unknown, relaxing the informational requirement of the two benchmark models. In Milgrom and Weber (1982) all bids are assumed to be a different bidder's maximum willingness to pay. But even

[6] They rely on the shape restriction of strict quasi-concavity of the seller's expected revenue function, which is slightly weaker than IFR (see also Bagnoli and Bergstrom (2005)).

[7] See Bagnoli and Bergstrom (2005), also for an overview of the many papers that use the slightly stronger log-concavity shape restriction in the economics literature. IFR can be tested nonparametrically when the data reveals more information than strictly required for the identification approach presented here, although none of the auction papers studying optimal reserve prices or imposing IFR explicitly (for instance Paarsch (1997), Haile and Tamer (2003), Roberts (2013)) has done so (to my knowledge). Barlow and Proschan (1966) show that that IFR results in decreasing normalized sample spacings (defined in Section 4), which can be tested with two sample spacings / three adjacent value order statistics. An (1995, Proposition 9) shows that order statistics of a log-concave distribution themselves have a log-concave density. Hence, pinning down one value order statistic allows one to check a necessary but not sufficient condition for IFR. Under the restriction that $\Delta = 0$, I show that this necessary condition is met in my empirical application.

when relaxing the behavioral restriction that all bidders bid their maximum willingness to pay, identification of the latent value distribution using the incomplete model of Haile and Tamer (2003) requires observing the number of bidders.[8] Instead, the econometrician must only observe the highest bid of at at least two losing bidders and the winning bidder (the transaction price). The next section presents English auctions with absentee bidding as a motivating example for using sample spacings for structural analysis of bidding data, and illustrates how informative bid order statistics are obtained in that setting.

# 3 English auctions with absentee bidding

This section introduces the motivating example of the paper, where the practice of absentee bidding in traditional English auctions both obscures the number of bidders and provides additional identifying information. After describing the setting and deriving equilibrium bidding strategies, Section 3.3 obtains (bounds on) value order statistics from the bid data. Specifically, these include the Haile and Tamer (2003) bounds on the second-highest valuation and lower bound on the highest-valuation, as well as a lower bound on the third-highest valuation. The novel identification method based on sample spacings set out in Section 4 is generally applicable under the restrictions of the auction model and the availability of these (bounds on) value order statistics, but does not rely on the absentee bidding environment *per se*.

## 3.1 Absentee bidding environment

In English auctions with absentee bidding, bidders have the option to place a sealed bid ahead of the auction. Exchangeability of $V_i$ for all bidders in the auction and the absence of entry or bidding cost guarantees that bidders' preference for live or absentee bidding is independent of their valuation. Suitability of this assumption depends on the empirical setting. In my application live bidding not only includes being present in the room but also bidding by phone or in the online bidding environment, so live

[8] The identification method in Haile and Tamer (2003) is presented for the case where the econometrician observes the number of bidders and all bidders' highest bids. However, it can also be applied when knowing the number of bidders and only one bid order statistic. In that case, the minimum in Theorem 1 of Haile and Tamer (2003) is only taken over the number of bidders and not over the different bid order statistics, resulting in weakly wider bounds on the distribution of valuations.

bidding is not necessarily more costly in terms of travelling to the auction house. The exchangeability assumption implies that when estimating bounds on $F_V$ from a subset of auctions, no selection is introduced.

Absentee bidding is a widely adopted practice in English auctions, as also documented by Akbarpour and Li (2019). For the data in my empirical application, Sotheby's provides the following guidance: *Absentee bids are to be executed as cheaply as permitted by other bids or reserves and in an amount up to but not exceeding the specified amounts. Bids will be rounded down to the nearest amount consistent with the bidding increment. In the case of identical (rounded) bids, the earliest submitted form will take precedence.*[9]

Auction mechanisms with absentee bidding are designed in a way that does not disadvantage absentee bidders.[10] Hence, if only one absentee bid is submitted for an item, the opening bid must start below that value. The rest of the bidding sequence will vary by institution. The following stylized sequence is based on my empirical observations that: i) the data contains various auctions with multiple absentee bids before live bidding starts and ii) after initial absentee bids the live bidding alternates with absentee bids:

**Bidding sequence.** *With at most one absentee bid, the opening bid equals a fixed share of the reserve price. With multiple absentee bids, the opening bid equals the minimum of the absentee bids. Subsequent bids follow fixed bidding increments until all but one absentee bidder drops out. Afterwards, if any live bids are placed, they are alternated against the last remaining absentee bidder unless he has dropped out. If there are still multiple live bidders willing to place bids, their bids are alternated as usual.*

Crucially, the bidding sequence formalizes that absentee bidders are not (dis)advantaged relative to live bidders, which is maintained throughout. It is unnecessary for my approach to specify the bidding sequence further. For instance, if multiple bidders have a valuation within two increments of the second-highest bid it is irrelevant who among them is recorded to place the highest bid. A bidder's highest

[9]Source: http://www.sothebys.com/en/auctions/2014/finest-rarest-wines-l14711.html (last accessed June 11 2020).

[10]Otherwise, rational bidders would not find it optimal to place absentee bids. An auctioneer not representing (absentee) bidders truthfully would not go undetected as the distribution of winning bids for absentee bidders would be stochastically dominated by the winning bid distribution for live bidders.

bid quite literally refers to the highest recorded bid of a given bidder, and the above makes clear that this highest bid might be strictly less than that bidder's valuation. Further specifying the general informational requirement (observing the highest bid of at least two losing bidders plus the transaction price) to the case with absentee bidders, it is assumed that:

**Informational requirement.** *The econometrician observes: i) a vector of bids, ii) which ones are absentee bids.*

## 3.2 Equilibrium bidding strategies

This section describes the equilibrium bid strategies as a function of a bidder's valuation, $\beta^k(v)$, for $k \in \{abs, live\}$ for respectively absentee and live bidders. I restrict attention to type-symmetric Bayes Nash equilibria in weakly undominated strategies. Recall that bidders are symmetric up to their bidding mode and valuation draw. However, there is a crucial difference in the set of bidding strategies available to them. Live bidders might bid early (squat) or raise the price by more than necessary (jump), bid only once or bid many times incrementally.[11] The strategy available to absentee bidders is limited to the height of the bid submitted.

**Lemma 1.** *It is optimal for an absentee bidder with $V = v$ to bid: $\beta^{abs}(v) = v$.*

*Proof.* For absentee bidders, the auction is strategically equivalent to an IPV second-price sealed bid auction. It therefore follows directly from Vickrey (1961) that truthful revelation is a weakly undominated strategy for absentee bidders. □

**Lemma 2.** *It is optimal for a live bidder with $V = v$ to bid: $\beta^{live}(v) \leq v$, with the added constraint that a final bid is placed if $v \geq$ standing price plus bidding increment.*

*Proof.* For live bidders, the auction is strategically equivalent to the English auction setting in Haile and Tamer (2003), where bidding up to one's valuation is optimal and where bidders won't let an opponent win at a price they are willing to beat. □

## 3.3 Resulting bounds on valuation order statistics

In this section, I leverage additional information revealed by absentee bids to bound an additional order statistic of the valuation distribution. Recall that $V_i$ denotes

[11] See Hasker and Sickles (2010) and the references therein.

bidder $i$'s value, for $i = 1, ..., n$, and $\Delta \geq 0$ the bidding increment. To work with order statistics, let $X_{i:n}$ denote the $(n-i+1)$th largest value in a sample of $n$ values of random variable $X$. As such, $V_{n:n}$ ($V_{1:n}$) denotes the highest (lowest) value from the sample $\{V_1, V_2, ..., V_n\}$. The observed vector of bids in each auction is used to pin down (bounds) on value order statistics. As not all bidders submit one and only one bid, let $B_j$ denote an observed bid for $j = 1, ..., \nu$ with $\nu$ the total number of bids in the auction. Specifically $B_{\nu:\nu}$ denotes the highest observed bid (e.g. the "transaction price") and $B_{\nu-1:\nu}$ the second-highest bid.[12] The + (−) superscript is consistently used to indicate that a value constitutes an upper (lower) bound on the relevant object.

By Assumption 3, e.g. the behavioral restrictions of Haile and Tamer (2003), the second-highest valuation is bounded between $B_{\nu-1:\nu}$ and $B_{\nu:\nu} + \Delta$, and the highest valuation must exceed $B_{\nu:\nu}$:

$$V_{n:n} \geq B_{\nu:\nu} \equiv V^-_{n:n} \tag{1}$$

$$V_{n-1:n} \geq B_{\nu-1:\nu} \equiv V^-_{n-1:n} \tag{2}$$

$$V_{n-1:n} < B_{\nu:\nu} + \Delta \equiv V^+_{n-1:n} \tag{3}$$

I add a fourth relationship using information revealed by absentee bids to bound the third-highest valuation. Lemma 1 established that absentee bidders bid truthfully. However, besides absentee bidders' highest bids, the bid vector also includes intermediate bids as the auctioneer is to determine the lowest price given competing bids. Bidder identities are unknown, but the stylized bidding sequence can be used to determine which bids correspond to an absentee bidders' highest bid rather than such intermediate values.

To do so, let $I(B_j) = \{abs, live\}$ indicate whether bid $B_j$ is submitted by an absentee or live bidder. I define $B_\mu$ as the highest bid that is certainly placed by a different bidder than who has bid $B_{\nu:\nu}$ or $B_{\nu-1:\nu}$, and it is identified from the bid vector as follows:

$$B_\mu = \max\{B_j < B_{\nu-1:\nu} : I(B_j) = abs \wedge I(B_{\nu:\nu}) = live \wedge I(B_{\nu-1:\nu}) = live\}$$

In words, $B_\mu$ is the highest absentee bid in auctions where the highest two bids are

[12] In ascending auction settings where bidding does not follow fixed increments, one must observe both $B_{\nu-1:\nu}$ and another distinct highest bid to apply the spacings method.

placed by live bidders. To illustrate, consider the following examples, where the pair of $(B_{j:\nu}, \{abs, live\})$ describe the value and type of bidder of the $(\nu - j + 1)$th highest bid.

**Characterizing $B_\mu$: examples.**
Auction 1: $(B_{1:5}$,abs), $(B_{2:5}$,abs), $(B_{3:5}$,abs), $(B_{4:5}$,live), $(B_{5:5}$,live)
Auction 2: $(B_{1:5}$,abs), $(B_{2:5}$,live), $(B_{3:5}$,abs), $(B_{4:5}$,live), $(B_{5:5}$,live)
Auction 3: $(B_{1:3}$,abs), $(B_{2:3}$,live), $(B_{3:3}$,abs)
In Auction 1 and Auction 2, $B_{4:5}$ and $B_{5:5}$ must be placed by different live bidders because nobody outbids himself, so both are (lower bounds) on values of different live bidders. $B_{3:5}$ is the highest other bid identified, from an absentee bidder, and hence equal to $B_\mu$.[13] In Auction 3, $B_{3:3}$ and $B_{2:3}$ must be by different bidders, but no $B_\mu$ is identified as $B_{1:3}$ might be placed by the same absentee bidder as $B_{3:3}$.

We therefore know that, since the highest two bids must be submitted by different bidders, and $B_\mu$ must be by a third bidder by definition, this identifies a lower bound on the third-highest value:

$$V_{n-2:n} \geq B_\mu \equiv V^-_{n-2:n} \tag{4}$$

Note how in the general case, the highest bid of the losing bidder who bids less than $B_{\nu-1:\nu}$ identifies $B_\mu \equiv V^-_{n-2:n}$. Econometric application of the method requires restricting the data to auctions in which $B_\mu$ is identified.

# 4 The identifying power of sample spacings

In this section, I provide a novel method to identify structural features of interest in an incomplete IPV ascending auction model. It is relevant to note that even when assuming that bids equal values, such auctions pose an identification challenge due to the censoring problem: the auction ends as soon as the second-highest bidder drops out and hence before the willingness to pay of the winning bidder is revealed.[14] On top

[13] Also note that in Auction 1 there must be more than one absentee bidder as there are multiple absentee bids placed before live bidding starts, and any other explanation would violate that the auctioneer acts in the best interest of the absentee bidder. According to the stylized bidding sequence, $B_{1:5}$ and $B_{3:5}$ are highest bids of different absentee bidders.
[14] Haile, Hong, and Shum (2003) refer to this as a "missing data problem".

of that, the presented auction model relaxes the behavioral restriction of all bidders bidding their willingness to pay, following Haile and Tamer (2003), and furthermore relaxes the informational requirement that the number of bidders is known to the econometrician. Two separate results are presented, of which Lemma 3 is particularly important in the light of the censoring problem as it can be used to derive an upper bound on expected winning bidder surplus and seller revenue.

**Definition: sample spacings.** Following Pyke (1965, 1972), spacings $D_{i:n}$ between two adjacent order statistics are defined as: $D_{i:n} = V_{i:n} - V_{i-1:n}$, $\forall i = 2,..,n$. Normalized spacings are defined as: $\tilde{D}_{i:n} = (n-i+1)(V_{i:n} - V_{i-1:n})$, $\forall i = 2,..,n$. Both $D_i$ and $\tilde{D}_i$ are random variables with CDF $F_{D_{i:n}}$ and $F_{\tilde{D}_{i:n}}$ $\forall i = 2,...,n$.

To understand the usefulness of the last spacing, consider that it contains same information as the highest two order statistics combined. Specifically, the density function of $D_{i:n}$ based on $V \sim^{i.i.d.} F_V$ (Pyke (1965)):

$$f_{D_{i:n}}(d) = \frac{n!}{(i-2)!(n-i!)} \int_0^{\bar{v}} F_V(x)^{i-2}[1 - F_V(x+d)]^{n-i} F_V(x) F_V(x+d) dx, \quad (5)$$

which equals the density of $V_{i:n}$ conditional on the realization of $V_{i-1:n}$ in expectation over all such realizations. Expected seller revenue ($\pi_S$) and expected winning bidder surplus ($\pi_B$) at counterfactual reserve prices are of primary interest in structural auction studies. They can be expressed in terms of the last spacing and the marginal distribution of the second-highest valuation:

$$\pi_S(r) = \int_0^{\bar{v}} [1 - F_{D_{n:n}}(r - v_{n-1})] \max(r, v_{n-1}) dF_{V_{n-1:n}}(v_{n-1}) \quad (6)$$

$$\pi_B(r) = \int_0^{\bar{v}} [1 - F_{D_{n:n}}(r - v_{n-1})] \{ \int_{\max(r, v_{n-1})}^{\bar{v}} v_n dF_{D_{n:n}}(v_n - v_{n-1}) \} dF_{V_{v-1:n}}(v_{n-1}) \quad (7)$$

where $[1 - F_{D_{n:n}}(r - v_{n-1})]$ denotes the sale probability when $V_{n-1:n} = v_{n-1}$ and with reserve price $r$, equal to: $1 - F_{V_{n:n}|V_{n-1:n}}(r|v_{n-1})$. As the distribution of the last spacing is not point-identified, let $F^-_{D_{n:n}}$ and $F^+_{D_{n:n}}$ respectively denote the lower and upper bound on the distribution of $D_{n:n}$, such that: $F^-_{D_{n:n}}(d) \leq F_{D_{n:n}}(d) \leq F^+_{D_{n:n}}(d)$, $\forall d \in [0, \bar{v}]$. To better illustrate the role of $F_{D_{n:n}}$, the bounds on $\pi_S$ and $\pi_B$ are presented as if $F_{V_{n-1:n}}$ is known rather than bounded.[15] By stochastic dominance,

[15] This is without loss of generality. With $\Delta$ strictly greater than 0, $F_{V_{n-1:n}}$ is bounded by the

bounds on $\pi_B$ and $\pi_S$ are defined as:

$$\int_0^{\bar{v}} [1 - F^-_{D_{n:n}}(r - v_{n-1})] \max(r, v_{n-1}) dF_{V_{n-1:n}}(v_{n-1}) \geq \pi_S(r) \quad (8)$$
$$\geq \int_0^{\bar{v}} [1 - F^+_{D_{n:n}}(r - v_{n-1})] \max(r, v_{n-1}) dF_{V_{n-1:n}}(v_{n-1})$$

$$\int_0^{\bar{v}} [1 - F^-_{D_{n:n}}(r - v_{n-1})] \{ \int_{\max(r, v_{n-1})}^{\bar{v}} v_n dF^-_{D_{n:n}}(v_n - v_{n-1}) \} dF_{V_{n-1:n}}(v_{n-1}) \geq \pi_B(r) \quad (9)$$
$$\geq \int_0^{\bar{v}} [1 - F^+_{D_{n:n}}(r - v_{n-1})] \{ \int_{\max(r, v_{n-1})}^{\bar{v}} v_n dF^+_{D_{n:n}}(v_n - v_{n-1}) \} dF_{V_{n-1:n}}(v_{n-1})$$

For example, consider the upper bound on $\pi_B$. It is based on $F^-_{D_{n:n}}$ that delivers both an upper bound on the sale probability for a given value of the second-highest valuation $(1 - F^-_{D_{n:n}}(r - v_{n-1}))$ and on the winning bidder surplus conditional on a sale (the inner integral in equation 9). This is because a stochastically larger last spacing translates directly into a larger $V_{n:n} - V_{n-1:n}$, and hence a larger probability that $V_{n:n}$ exceeds $r$ for any value of $v_{n-1}$ as well as placing more weight on higher values of $V_{n:n}$ conditional on a sale.

Note that if $n$ were known and $F_V$ were point-identified, one could work with the marginal distribution $F_{V_{n:n}}$ by taking it out of the (outer) integrals in (6)-(7) to point-identify $\pi_B$ and $\pi_S$. In an incomplete English auction model that allows for correlated private values, Aradillas-López et al. (2013) and Coey et al. (2017) instead exploit observed exogenous variation in the number of bidders to bound these structural features of interest. The novelty of my identification approach is that it exploits the relation between ageing properties of $F_V$ and the stochastic ordering of its sample spacings, in a setting where $n$ is unknown.

distributions of $V^-_{n-1:n}$ and $V^+_{n-1:n}$. The stochastic dominance argument that defines bounds on $\pi_B$ and $\pi_S$ in equations (8) and (9) also applies when using the relevant bounds on $F_{V_{n-1:n}}$. Specifically, the sample spacings method does not rely on the assumption made in the majority of the ascending auction literature that $F_{V_{n-1:n}}$ is point-identified from the distribution of the transaction price, as in e.g. Aradillas-López et al. (2013), Coey et al. (2017, 2018), Song (2004), Freyberger and Larsen (2017), Platt (2017), Larsen (2021), Quint (2008), and Hernández, Quint, and Turansick (2020).

## 4.1 Identification of $F^{-}_{D_{n:n}}$

**Lemma 3.** *$V^{+}_{n-1:n}$ and $V^{-}_{n-2:n}$ identify $F^{-}_{D_{(n:n)}}$ without knowing $n$:*

$$F^{-}_{D_{n:n}}(d) \equiv P[2(V^{+}_{n-1:n} - V^{-}_{n-2:n}) \leq d] \ , \forall d \in [0, \bar{v}] \tag{10}$$

*Proof.* Normalized spacings $\tilde{D}_{i:n}$ from distribution functions with increasing failure rates are stochastically decreasing in $i$ for $i = 2, ..., n$ (Barlow and Proschan (1966, Corollary 5.2)). Hence, by assumption 2: $P[D_{n:n} \leq d] \geq P[2D_{n-1:n} \leq d]$ for all d on its support. Substituting $D_{n-1:n}$ with the relevant value order statistics, we get that: $P[D_{n:n} \leq d] \geq P[2(V_{n-1:n} - V_{n-2:n}) \leq d]$ and by definition of their bounds: $P[D_{n:n} \leq d] \geq P[2(V^{+}_{n-1:n} - V^{-}_{n-2:n}) \leq d]$. This completes the proof as, $\forall d \in [0, \bar{v}]$, $F_{D_{n:n}}(d) = P[D_{n:n} \leq d] \geq P[2(V^{+}_{n-1:n} - V^{-}_{n-2:n}) \leq d]$. □

In addition, in the special case that $\Delta = 0$, it is straightforward to see that the sample spacings method identifies a tighter lower bound using the exact second-highest value pinned down by the transaction price.[16] Further intuition behind the proof is that *normalized spacings* of i.i.d. draws from an exponential distribution are exchangeable random variables (Pyke (1965)), and that by assumption 2 that $F_V$ satisfies IFR it ages faster than the exponential in the reliability theory sense. The resulting $F^{-}_{D_{n:n}}$ is sufficient to bound $\pi_B$ and $\pi_S$ from above, as evident in (8) and (9), successfully overcoming the censoring problem.

## 4.2 Identification of $F^{+}_{D_{n:n}}$

I next show how to apply a similar idea to identify $F^{+}_{D_{n:n}}$ without knowing the number of bidders, which can be used to derive lower bounds on the structural features of interest. Different from Lemma 3, the result is based on $V_{n-2:n}$ being pinned down exactly from bidding data, such as in Song (2004), Adams (2007), Kim and Lee (2014), Mbakop (2017), Freyberger and Larsen (2017), Luo and Xiao (2019) and any paper that adopts the button-auction format, although the result does not require that $V_{n-1:n}$ is point-identified. The model presented in this paper clearly allows for additional incompleteness of the observed bid vector, so the result in Lemma 4 is

[16] In the case that $\Delta = 0$, one still cannot point-identify the third-highest valuation in the presented model as there might be multiple bidders with a valuation exceeding $B_\mu$. The bidder submitting the third-highest bid is not necessarily the one with the third-highest valuation.

presented to provide a richer perspective on the usefulness of sample spacings in the structural analysis of bid data. It relies on the following shape restriction defined in terms of dispersive ordering:

**Definition: Dispersive ordering.**
Random variable $X \sim F_X$ is defined to be less dispersed than random variable $Y \sim F_Y$ iff $[F_X^{-1}(b) - F_X^{-1}(a)] \leq [F_Y^{-1}(b) - F_Y^{-1}(a)]$, $\forall 0 < a \leq b < 1$, denoted by $X \leq^{disp} Y$.

**Assumption 4.** *$V \leq^{disp} U$, where $U \sim^{i.i.d.} Unif[0, \bar{v}]$ (less dispersed than uniform)*

In words, the assumption requires that the difference between any two quantiles of the distribution of valuations is less than the corresponding difference in the uniform distribution on the same support. Jongwoo, Kochar, and Park (2006) provide further statistical properties and relations to other partial orderings. It is indeed plausible to assume that as extreme a dispersion as generated by the Uniform distribution, where all values on the support are equally likely, is unsuitable to describe latent values in auction data. This is especially true for the commonly selected parametric distribution functions (e.g. Log-Normal, Exponential, Logistic, Weibull, Gamma) that are defined on $[0, \infty)$.[17] With the upper bound of the support being very large it is certainly persuasive that the density function of $V$ tapers off quicker than the Uniform density.[18]

**Lemma 4.** *$V^-_{n-1:n}$ and $V_{n-2:n}$ identify $F^+_{D_{n:n}}$ without knowing $n$:*

$$F^+_{D_{n:n}}(d) \equiv P[(V^-_{n-1:n} - V_{n-2:n}) \leq d] \ , \forall d \in [0, \bar{v}] \tag{11}$$

*Proof.* First, spacings from a uniform distribution are exchangeable random variables (Pyke (1965)). Letting $D^U_{i:n}$ denote the $i$th spacing of a sample of $n$ draws from $G_U = Unif[0, \bar{v}]$, we know that $P[D^U_{n-1:n} \leq d] = P[D^U_{n:n} \leq d]$. Second, $V \leq^{disp} U$

[17] Note that technically all commonly applied parametric distribution functions, except notably the Beta distribution, violate the bounded support regularity condition. While not overly elegant, I do not consider this to be problematic given the way the literature deals with this discordance. For example, Paarsch (1997) imposes bounded support and specifies a Weibull distribution to estimate latent values. Haile and Tamer (2003) impose bounded support and assume Lognormal, $\mathcal{X}^2$, and Beta distributions in Monte Carlo simulations. Song (2004) imposes bounded support and assumes a Gamma distribution in simulations. In general, any paper that solves for a unique optimal (counterfactual) reserve price but uses an unbounded parametric distribution function to estimate it is culpable of this imperfection.

[18] Moreover, assumption 4 only rules out distribution functions that are *more* dispersed than the Uniform, such as the U-quadratic(a,b) on support $[a, b]$ with inverted-U shaped PDF $a(x - b)^2$.

by Assumption 4 and Bartoszewicz (1986) proves that this implies that any spacings $D_{i:n}$ from $F_V$ are stochastically dominated by corresponding spacings of a sample of the same size drawn from $F_U$ (Kochar (2012, Theorem 4.2)). Hence, with $\leq^{st}$ denoting the stochastic ordering, we know that: $D_{i:n} \leq^{st} D^U_{i:n}$ $\forall i = 2, .., n$. Supposing $V \sim G_U$ would imply: $P[D_{n:n} \leq d] = P[D_{n-1:n} \leq d]$, so that under assumption 4: $P[D_{n:n} \leq d] \leq P[D_{n-1:n} \leq d]$ for all $d$ on its support.

Substituting $D_{n-1:n}$ for the relevant value order statistics, we get that: $P[D_{n:n} \leq d] \leq P[(V_{n-1:n} - V_{n-2:n}) \leq d]$ and by definition of the bound: $P[D_{n:n} \leq d] \leq P[(V^-_{n-1:n} - V_{n-2:n}) \leq d]$. This completes the proof as, $\forall d \in [0, \bar{v}]$, $F_{D_{n:n}}(d) = P[D_{n:n} \leq d] \leq P[(V^-_{n-1:n} - V_{n-2:n}) \leq d]$. □

In the special case that $\Delta = 0$, it is straightforward to see that the sample spacings method identifies a tighter upper bound and if only $V^+_{n-2:n} \leq V_{n-2:n}$ were identifiable from data a weaker upper bound would be identified. As mentioned above, I don't assume that $V_{n-2:n}$ is point-identified in my empirical application. Instead, I verify a necessary condition for applicability of $F^+_{D_{n:n}}$ based on a comparison between empirical and simulated sale probabilities. I also provide a comparison of estimated lower bounds on $\pi_B$ and $\pi_S$ with those obtainable from the distribution of $V^-_{n:n}$ that is trivially identified from $B_{\nu:\nu}$ and that does not rely on Assumption 4.[19] This again goes to show that it is really the upper bound on the highest valuation, or a lower bound on the distribution of the last spacing, that is more urgent to pin down from bid data in light of the censoring problem.

## 4.3 Simulations

Figure 1 shows how Lemma 3 and 4 apply to familiar distribution functions. Plots compare the true distribution of the spacing between the highest two order statistics with its bounds implied by the difference between $V_{n-1:n}$ and $V_{n-2:n}$. Plotted are simple empirical CDF's based on 10.000 simulated sets of values, abstracting from bidding increments.

[19]Specifically:

$$\pi_S(r) \geq \int_0^{\bar{v}} [1 - F^+_{V_{n:n}|V_{n-1:n}}(r|v_{n-1})] \max(r, v_{n-1}) dF_{V_{n-1:n}}(v_{n-1}) \tag{12}$$

$$\pi_B(r) \geq \int_0^{\bar{v}} [1 - F^+_{V_{n:n}|V_{n-1:n}}(r|v_{n-1})]\{\int_{\max(r,v_{n-1})}^{\bar{v}} v_n dF^+_{V_{n:n}|V_{n-1:n}}(v_n|v_{n-1})\} dF_{V_{n-1:n}}(v_{n-1}) \tag{13}$$

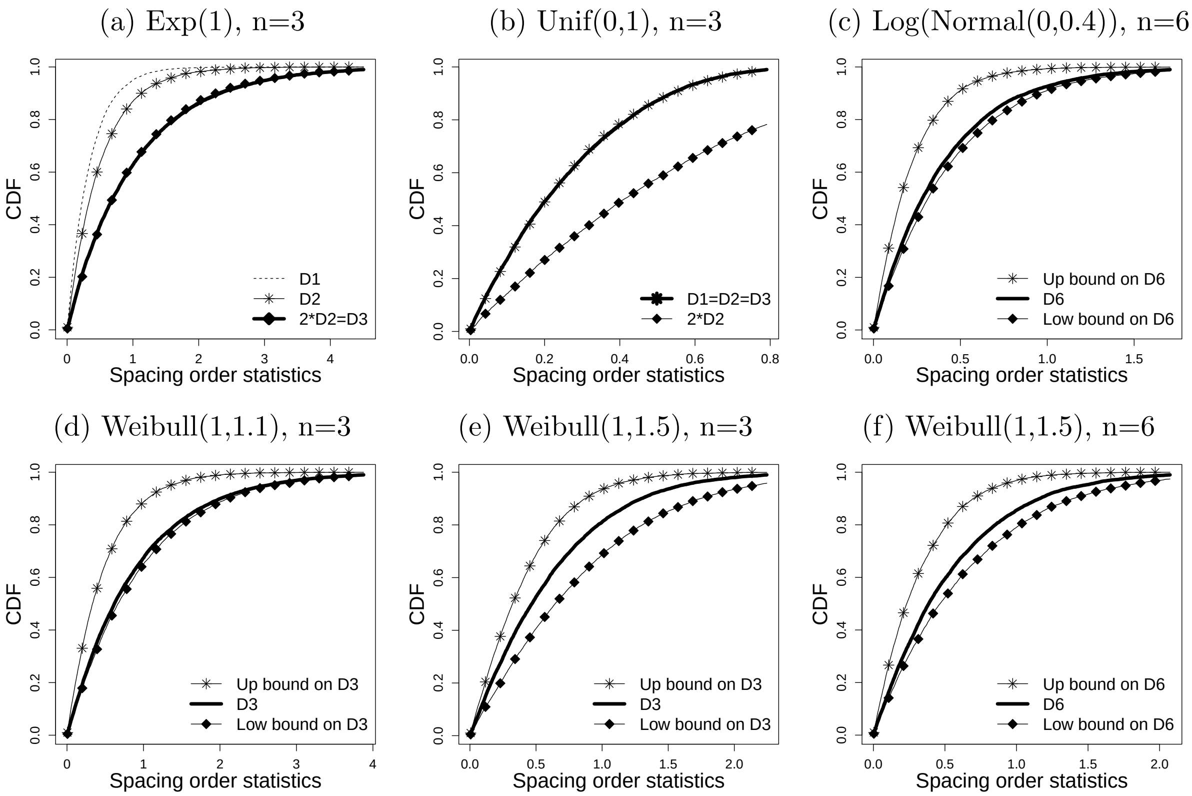


Figure 1: Informativeness of bounds on $F_{D_{n:n}}$

Based on Lemma 3 and 4 applied to different data generating processes, plotted for spacings between the 1st percentile of $D_{n-1:n}$ and the 99th percentile of $D_{n:n}$ (x-axis). $Di$ in legends refers to $D_{i:n}$.

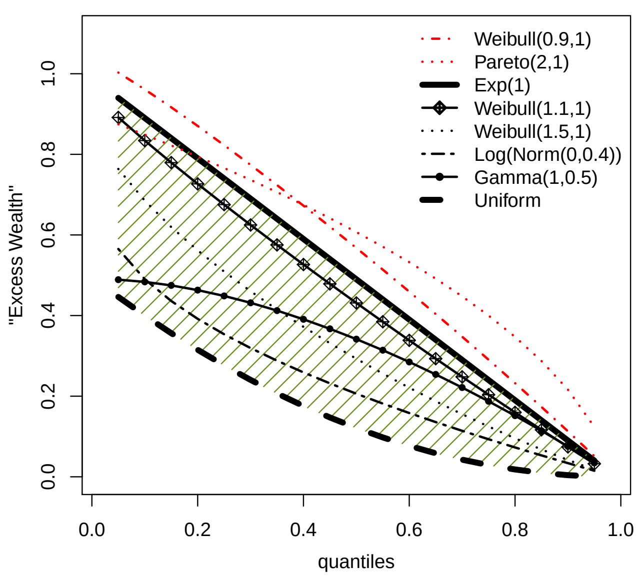


Figure 2: Illustration of excess wealth order to summarize shape restrictions

Plots a and b illustrate the simple basis for the presented identification results. Plot a is based on the Exponential distribution. $F^-_{D_{n:n}}$ as derived in Lemma 3 collapses to the true $F_{D_{n:n}}$. For any distribution function that ages faster than the Exponential, the last spacing will therefore be stochastically dominated by $F^-_{D_{n:n}}$. Plot b is based on the Uniform distribution. $F^+_{D_{n:n}}$ as derived in Lemma 4 collapses to the true $F_{D_{n:n}}$. For any distribution that ages slower than the Uniform, the last spacing will therefore be stochastically dominating $F^+_{D_{n:n}}$.

Four other points are important to take away from these simulations. First, leveraging information revealed by sample spacings can indeed result in highly informative bounds. Second, $F^-_{D_{n:n}}$ is tighter the less IFR the underlying distribution function is (plot d versus e). This simultaneously means that the lower bound is less tight in those instances. Third, the larger the number of (unobserved) bidders, the tighter $F^-_{D_{n:n}}$ (plot f versus e). And fourth, IFR is sufficient but not necessary for the sample spacing method to identify $F^-_{D_{n:n}}$ (plot c).[20] While the method results in valid bounds for some value distributions that are not IFR, it is chosen as a shape restriction commonly used in the economics literature. Bagnoli and Bergstrom (2005) lists many well-known papers using the slightly stronger log-concavity restriction.

I find the excess wealth order intuitive to interpret the two shape restrictions in Assumptions 2 and 4, especially because identification focuses on the upper tail of the distribution of $F_V$.

**Definition: Excess Wealth Order.** Random variable $V$ is larger than $Y$ in the excess wealth order ($V \geq^{EW} Y$) if and only if: $\int_{F_V^{-1}(p)}^{\infty} 1 - F_V(t)dt \geq \int_{F_Y^{-1}(p)}^{\infty} 1 - F_Y(t)dt, \forall p$. For IFR distribution functions it holds that the dispersive ordering implies the excess wealth order (Kochar (2012)), so that all IFR distributions satisfying $V \geq^{EW} Y \sim Unif[a, b]$ also satisfy $V \geq^{disp} Y$.

It is easy to show that the excess wealth of $E \sim Exp(\lambda)$ at quantile $p$ equals: $\frac{p}{\lambda}$, e.g. in figure 2 the thick solid line with slope $-1$ (exponential with rate $\lambda = 1$). IFR distributions (with $F(0) = 0$) have lower excess wealth at each $q$. So distribution functions with excess wealth in between of the $Unif[0, 1]$ and $Exp(1)$ distributions

[20] Plot $c$ in Figure 1 is based on the Lognormal distribution, parameterized to replicate the estimated idiosyncratic timber logging cost distribution in Haile (2001), with the shape parameter of 0.4 taken from Table 7 for Region 5. The failure rate of the Lognormal distribution is non-monotonic, only increasing on the part of its support $\in (0, 1)$, but the sample spacing method delivers a valid upper bound on $F_{D_{n:n}}$ with 6 bidders.

satisfy the two shape assumptions of this paper. This corresponds to the shaded area in the figure.

# 5 Application: wine auctions at Sotheby's

I apply the sample spacing method to a unique dataset covering the 884 lots from the "Finest and Rarest Wines & Vintage Port" auction on November 19th 2014 at Sotheby's London.[21] This provides a good and conservative test case as it is a relatively small dataset, with large bidding increments, and it does not contain the number of bidders. The dispersion of highest bids is large, even when normalizing by the number of bottles in the lot as is done throughout the application, and especially the upper tail is long. To reduce the impact of extremal values I therefore exclude auctions for which $V^-_{n:n}$ or $V^+_{n:n}$ exceed their 95th percentile without conditioning on observables.[22] Descriptive statistics of the remaining sample are provided in table 1. As common in traditional English auctions: increments are high at between 2 - 16 percent of the winning bid.

The presence of absentee bidding requires bidders to be willing to announce their bids before others do, which is by itself a strong indication that bidders do not anticipate a "winners curse" and that the assumption of private values is justified. It is plausible that observables, including especially the estimated market value estimated by Sotheby's Wine Department, capture well all auction-level heterogeneity. It is furthermore reasonable to treat the remaining variation in valuations (both within and across auctions) as driven by idiosyncratic private valuations of bidders. The data supports this idea. The high predictive power of Sotheby's pre-auction value estimate is highlighted in figure 4 a, showing its relation with the realized winning bid. The remainder of this paper explicitly accounts for auction-level observed heterogeneity $Z$, with $Z$ being Sotheby's pre-auction lower bound on the estimated value of the item. The (conditional) IPV assumption implies that bidders independently draw values $v \sim^{i.i.d.} F_{V|Z}$, with $Z \perp V$.

A necessary condition for $F_{V|Z}$ having an increasing failure rate is that the den-

[21] I collected the data by simply registering as an online bidder, recording the complete auction, translating the video material into a dataset of bids, and adding lot descriptors from the catalogue.

[22] The 99th (95th) percentiles of unconditional per-bottle $V^-_{n:n}$ and $V^+_{n:n}$ are respectively: 945.83 (3356.67) and 2419.50 (6540.00) pounds.

Table 1: Descriptive statistics for sample of fine wine auctions

| | N | Mean | St. Dev. | Min | 25th pct. | Median | 75th pct. | Max |
|---|---|---|---|---|---|---|---|---|
| Opening bid (pounds) | 697 | 846.255 | 847.841 | 50 | 280 | 520 | 1,150 | 9,000 |
| Highest bid (pounds) | 697 | 999.857 | 993.776 | 90 | 340 | 620 | 1,300 | 9,200 |
| Number bottles per lot | 697 | 8.624 | 4.721 | 1 | 6 | 8 | 12 | 36 |
| Highest bid per bottle (pounds) | 697 | 166.434 | 183.561 | 5 | 40 | 87.50 | 225 | 933.33 |
| Increment at highest bid (%) | 697 | 5.623 | 1.949 | 2.041 | 4.348 | 5.263 | 6.667 | 16.000 |
| Is sold | 697 | 0.902 | 0.297 | 0 | 1 | 1 | 1 | 1 |
| Number of bids | 697 | 4.184 | 2.936 | 2 | 2 | 3 | 5 | 25 |
| Sotheby's low estimate (pounds) | 697 | 917.805 | 980.857 | 80 | 300 | 550 | 1,200 | 13,500 |
| Sotheby's high estimate (% above low) | 697 | 27.418 | 6.582 | 9.091 | 23.077 | 27.273 | 30.769 | 62.500 |

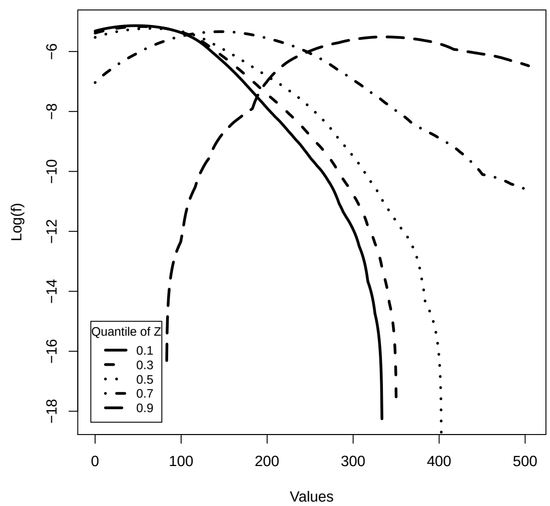


Figure 3: Log-concavity

Plotting logged nonparametric conditional density estimates of the second-highest bid $log(\hat{f}_{B_{\nu-1:\nu}|Z})$ by quantiles of $Z$, with bids normalized by the number of bottles.

sities of its order statistics are log-concave.[23] I verify this condition in the data by plotting the estimated density of the second-highest bid, conditional on quantiles of $Z$, estimated with nonparametric conditional Kernels and optimal cross-validated bandwidths as detailed in the next section. The check is therefore based on the restriction that $V_{n-1:n} = B_{\nu-1:\nu}$ (e.g. imposing $\Delta = 0$). With this caveat in mind, eyeballing figure 3 does suggest that $\hat{f}_{B_{\nu-1:\nu}|Z}$ is log-concave across the distribution of $Z$, thereby not invalidating the weaker IFR shape restriction on $F_{V|Z}$.

[23]Log-concavity of $f_{V|Z}$ implies log-concavity of $F_{V|Z}$ and IFR (Bagnoli and Bergstrom (2005, Theorem1, Theorem3)), and log-concavity of $F_{V|Z}$ (and therefore $f_{V|Z}$) also implies log-concavity of $f_{V_{i:n}|Z}$ for any order statistic (An (1995, Proposition 9)).

(a) For predicting winning bids (b) For (secret) reserve prices

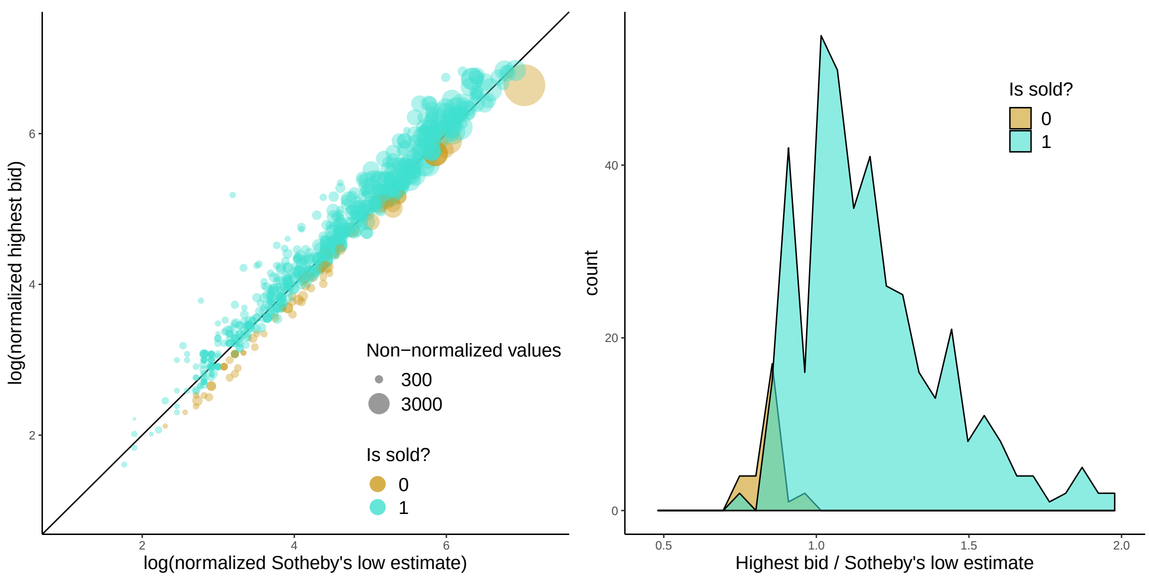


Figure 4: The role of Sotheby's pre-auction value estimate.

Plot a illustrates the strong correlation between Sotheby's pre-auction value estimate (the lower bound of the value bracket) and the realized winning bid. Plotted are log normalized values, which are amounts divided by the number of bottles in the lot, and dot sizes reflect non-normalized values. Plot b illustrates that secret reserve prices are at most equal to Sotheby's low value estimate, with the highest bid as a share of the value estimate always being between 0.75-1 for unsold items.

## 5.1 Nonparametric estimation

This section shows how to estimate bounds on $\pi_S$ and $\pi_B$ (equations 8 and 9) by applying the spacings identification approach. $\hat{\pi}_S$ and $\hat{\pi}_B$ are based on product Kernel estimators of conditional spacings density functions ($dF_{V_{n-1:n}|Z}$, $dF^+_{D_{n:n}|Z}$, $dF^-_{D_{n:n}|Z}$) and conditional spacings cumulative distribution functions ($F^+_{D_{n:n}|Z}$, $F^-_{D_{n:n}|Z}$), such as:

$$d\hat{F}^+_{D_{n:n}|Z}(d|z) = \frac{\frac{1}{h^{D-}}\sum_{t\in\mathcal{T}^-}^{K} L\left(\frac{D^{t-}_{n:n}-d}{h^{D-}}\right) K\left(\frac{Z^{t-}-z}{h^{Z-}}\right)}{\sum_{t\in\mathcal{T}^-} K\left(\frac{Z^{t-}-z}{h^{Z-}}\right)} \tag{14}$$

$$\hat{F}^-_{D_{n:n}|Z}(d|z) = \frac{\sum_{t\in\mathcal{T}^+} L\left(\frac{D^{t+}_{n:n}-d}{h^{D+}}\right) K\left(\frac{Z^{t+}-z}{h^{Z+}}\right)}{\sum_{t\in\mathcal{T}^+} K\left(\frac{Z^{t+}-z}{h^{Z+}}\right)} \tag{15}$$

$$L(x) = \int_{-\infty}^{x} K(u)\, du$$

$\forall(d,z)$ on their supports. The minus (plus) superscripts on $\mathcal{T}$, $D_{n:n}$, $Z$, and $h$ indicate that they relate to the lower (upper) bound on the last spacing and the therefore relevant observations. $K$ indicates the Epanechnikov kernel function, $h$ the cross-validated variable-specific bandwidths as functions of the relevant sample sizes. Moreover, let $T^- = |\mathcal{T}^-|$ ($T^+ = |\mathcal{T}^+|$) denote the total number of auctions in which $D^-_{n:n}$ ($D^+_{n:n}$) is identified, with the caligraphic script denoting sets of such auctions.[24]

For a meaningful analysis of these estimators, estimated spacings CDF's are first applied to bound consumer surplus in the data ($CS^t$):

$$\hat{CS}^t \in [\int (b^t+x)d\hat{F}^-_{D_{n:n}|Z}(x|z)dx-(b^t+\Delta), \int (b^t+x)d\hat{F}^+_{D_{n:n}|Z}(x|z)dx-(b^t+\Delta)], \quad (16)$$

for all $t$ sold, and 0 otherwise, and with $B_{\nu-1:\nu} = b_t$ and $B_{\nu:\nu} = b_t+\Delta$ respectively the observed second-highest and highest bid. $\hat{CS}^t$ differs from counterfactual surplus as defined in (7) as it conditions on the realized second-highest bid and does not involve the sale probability: it conditions on the event of a sale given the current (unknown) reserve price. The upper bound on $CS^t$ is also compared against one derived from the distribution of $B_{\nu:\nu} = V^-_{n:n}$ itself, as the alternative lower bound described on page 17:

$$\hat{CS}^t \geq \int x d\hat{F}^+_{V_{n:n}|Z}(x|z)dx - (b^t+\Delta) \quad (17)$$

After estimating the three conditional densities, $\hat{CS}^t$ is approximated numerically on a fine grid of $x$, for all sold auctions. Table 2 reports estimated bounds on $\mathbb{E}[CS^t]$, both by tertile of the conditioning variable and for the whole sample. Results highlight how informative the estimated bounds are, even in this "worst-case scenario": a small sample with large bidding increments and without knowing the number of bidders.

The (point estimate of the) lower bound on expected consumer surplus equals 75 percent of the average winning bid (87 percent when based on $V^-_{n:n}$ instead of $D^-_{n:n}$), not far removed from the upper bound at 125 percent. Estimated bounds are wider at the tertile level, based on only a third of the sample and about 200 observations. The middle tertile for instance has estimated bounds between 121 and 193 percent of the average winning bid, and the first tertile between 137 and 218 percent of the winning bid. Another clear result is that consumer surplus is higher for higher-end wines, with

[24] Uniform consistency of the PDF requires $h^{D-} \to 0$, $h^{Z-} \to 0$, and $T^- h^{D-} h^{Z-} \to \infty$ as $T^- \to \infty$, and for the CDF that $h^{Z+} \to 0$ and $T^+ h^{Z+} \to \infty$ as $T^+ \to \infty$.

## Table 2: Estimated bounds on $\mathbb{E}[CS^t]$, by tertile of Z

| | 1st tertile | 2nd tertile | 3rd tertile | all auctions |
|---|---|---|---|---|
| Number observations (sold) | 225 | 195 | 209 | 629 |
| Z, mean | 29.981 | 96.546 | 341.194 | 154.025 |
| Winning bid | 35.087 | 111.218 | 385.726 | 175.197 |
| Estimated bounds on $\mathbb{E}[CS^t]$: | | | | |
| Lower bound (based on $V^-_{n:n}$) | 55.392 | 118.029 | 276.335 | 151.592 |
| 95% Confidence interval | [50.471,61.768] | [112.477,123.637] | [254.843,291.649] | [141.767,162.380] |
| Lower bound (based on $D^-_{n:n}$) | 47.830 | 134.683 | 234.089 | 130.792 |
| 95% Confidence interval | [43.626,52.584] | [125.689,143.423] | [218.889,248.296] | [120.669,139.681] |
| Upper bound (based on $D^+_{n:n}$) | 76.460 | 214.824 | 382.647 | 218.339 |
| 95% Confidence interval | [68.655,85.092] | [191.840,237.115] | [342.632,432.390] | [201.082,233.435] |

Estimates are in pounds per bottle, based on the 629 auctions in which $B_\mu$ is identified. Conservative confidence intervals are calculated across 100 nonparametric bootstrap samples. Z = Sotheby's low value estimate.

(a) Bounds (b) Difference up-low bound

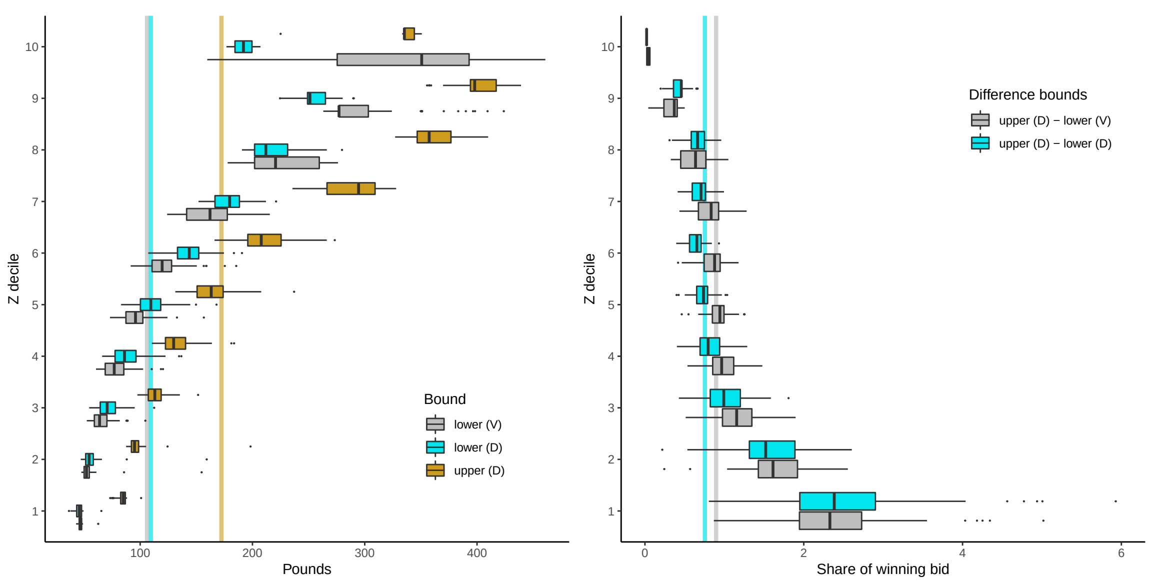


Figure 5: Heterogeneity of auction-level surplus $CS^t$

Plot a: point estimates of bounds, by decile of Z=Sotheby's pre-auction value estimate. Plot b: differences in point estimates bounds as a share of the winning bid, by decile of Z. Vertical lines indicate sample medians.

even the upper bound of the estimated surplus for the first tertile being less than the lower bound for the second tertile, and similarly for comparing the second to the third tertile. This goes through even when taking the 95% bootstrapped confidence intervals into account.[25]

Figure 5 explores further auction-level heterogeneity of $\hat{CS}^t$. Plot a displays heterogeneity of estimated bounds by decile of the conditioning variable. The previous result that higher-value lots deliver more consumer surplus (which is not trivially so) is reinforced. The boxplots also reveal that there is more heterogeneity for higher-end wines, insofar as this is reflected by Sotheby's pre-auction value estimate. Plot b plots the *difference* between the point estimate of high and low bound on $CS^t$, in this case reported as a share of the winning bid, and plotted by decile of the winning bid. There is again remarkable auction-level variation, with the sample spacing method resulting in bounds being more informative in terms of this outcome for higher-end wines.

## 5.2 Policy simulation: optimal reserve price

It is an official policy of Sotheby's to allow only reserve prices at or below the low bound of the value estimate.[26] Also empirically, figure 4 plot b shows that secret reserve prices are set at a fixed share of Sotheby's low value estimate. The highest bids are below this estimate for all unsold lots (between 75-100%). Together with the high sale probability of 90 percent, this begs the question whether and how much sellers would benefit from adopting an optimal reserve price; a counterfactual of primary interest in empirical auction studies.[27]

To bound the increase in expected seller revenue from adopting a higher reserve price, I run the following simulation exercise. I consider counterfactual reserve prices $\tilde{r}_t = \alpha Z_t$ for a range of $\alpha \in (1, 1.5)$, thereby relaxing the constraint on reserve prices insofar as the current policy is binding. To evaluate the impact on sellers, I estimate bounds on $\pi_S$ by applying the results from Lemma's 3 and 4 to (8). With reserve

[25] I rely on simple pointwise inference as it is beyond the scope of this paper to develop results for a uniform inference approach. As in Coey et al. (2017), I report conservative confidence intervals based on a nonparametric bootstrap method. Pointwise bounds are estimated in each of 100 bootstrap samples, and the 2.5th and 97.5th percentile across these samples is reported.

[26] Source: https://www.sothebys.com/en/glossary (last accessed June 11 2020).

[27] For example, optimal reserve prices are of central policy interest in e.g. Paarsch (1997), Haile and Tamer (2003), Tang (2011), Aradillas-López et al. (2013), Coey et al. (2017, 2018, 2019), who either first estimate the latent value distribution or also (set-)identify this structural feature directly.

prices being secret, the highest bid of the bid vector is the floor of the counterfactual bid realizations. The trade-off between increasing the price conditional on a sale and having a lower sale probability drives the results.[28]

Specifically, the expectation over $V_{n-1:n}$ in (8) is done over realizations of $B_{\nu:\nu}$ in the data (not excluding unsold lots). Suggestive evidence that the upper bound $F^+_{D_{n:n}}$ in Lemma 4 applies is that the simulated lower bound on the sale probability indeed is lower than the empirical sale probability, at 93.7 percent versus 94.6 percent. All counterfactual results are simulated only for auctions in which $D^+_{n:n}$ is identified (e.g. where $B_\mu$ is identified), to guarantee that the lower and upper bounds relate to the same primitive. $\hat{F}^+_{D_{n:n}}$ and $F^-_{D_{n:n}}$ are estimated as described in the previous section, with the addition that the larger cross-validated bandwidths from the upper bound are used in both estimators to guarantee that the bounds don't cross in the even smaller sample (167 observations).

The benchmark policy is that Sotheby's determines $Z$ and sellers choose a reserve price less than $Z$, so this counterfactual sheds light on two questions: 1) whether sellers on the whole would be better off when adopting the simple policy of a standardized reserve price rule, and 2) whether sellers and Sotheby's gain from allowing the reserve price (rule) to exceed $Z$.

Resulting $\hat{\pi}_S(r)$ are plotted in figure 6. Despite not point-identifying counterfactual seller revenue, the simulations reveal interesting facts. Moving from a system where sellers set their secret reserve individually between 75-100% of Z to one where there is a common reserve price rule *at any level* between 75-100% of Z is clearly beneficial to sellers. This suggests that even conditional on the reserve price constraint $r \leq Z$, individual sellers set suboptimal reserve prices. They would gain at least 2.5 percent by setting a common reserve price rule $\alpha \in [0.75, 1]$, and the upper bound for such a policy is estimated at 6 percent gain.

Additional benefits are associated with $\alpha > 1$ meaning that reserve prices are set sub-optimally low, and that the reserve price constraint is indeed binding. Setting $\alpha$ equal to 1.05 or 1.1 leads respectively to $\hat{\pi}_S$ increasing between 2.9-7.4 percent and

[28] As pointed out by an anonymous reviewer, switching to a fixed reserve price rule implicitly makes the secret reserve price public. This does not affect outcomes in the IPV ascending auction model presented here, but it might make it easier for (say) a bidding ring to extract surplus in a world with collusion. The use of secret reserves remains a puzzle in the literature. Ashenfelter and Graddy (2006), Jehiel and Lamy (2015), and articles reviewed in Hasker and Sickles (2010) provide various explanations for when it might be optimal.

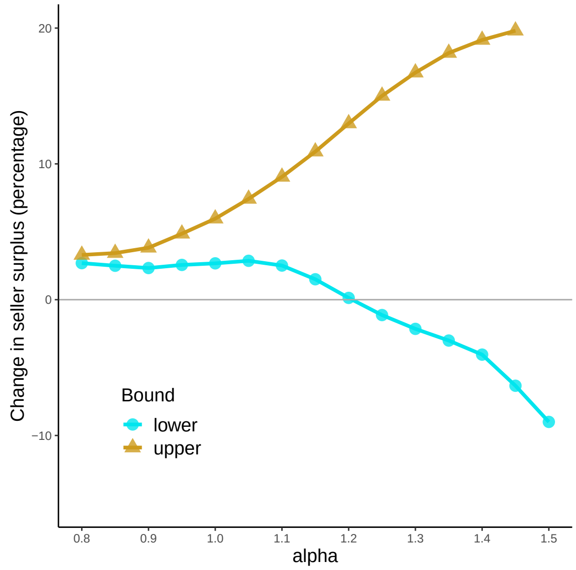


Figure 6: Policy simulation: $\pi_S^{\Delta}(\tilde{r})$ with reserve price rule $r_t = \alpha z_t$

2.5-9.0 percent. The policy to set $\tilde{r}_t = 1.2z_t$ increases seller revenue up to 13 percent (and at least 0.13 percent), but for higher values of $\alpha$ the lower bound becomes negative so it becomes unclear whether sellers would be better off. At higher values of $\alpha$, the lower bound on $\hat{\pi}_S(r)$ becomes negative so can't be said with certainty that higher levels are beneficial.

Recall that Sotheby's provides a value *bracket* for each lot, and it is notable that $\alpha = 1.2$ would set the secret reserve near the upper end of this bracket (the median upper bound is 27% higher than $Z$). In other words, by restricting $\alpha \leq 1$, Sotheby's does not fully allow for reserve prices to satisfy the *exclusion principle* of optimal reserve prices (Krishna (2009), based on Myerson (1981) and Riley and Samuelson (1981)). It places too much weight on the sale probability and too little on the expected revenue conditional on a sale. It is even suboptimal for Sotheby's themselves as most of their income comes from commissions.

In terms of monetary values and extrapolating to the whole sample, the estimated upper bound implies that these wine auctions would generate up to 25,694 pounds in additional seller revenue when setting a common reserve price rule of $\tilde{r}_t = 1.2z_t$ and the lower bound implies that they would in any case not lose out from this policy. The same back of the envelope calculation suggests an increase in seller revenue between 5,732 and 14,626 pounds when setting $\tilde{r}_t = 1.05z_t$. This is for just one day of auctions; gains add up with Sotheby's holding over 20 of these "Finest and Rarest Wine auctions" each year.

# 6 Conclusion

This paper proposes a new approach to set-identify policy counterfactuals in a non-button IPV English auction setting, without knowing the number of bidders. The identification method exploits information contained in the *spacing* of order statistics in combination with weak shape restrictions. Simulations further illustrate the simple identification approach and relate it to the ageing properties of the extremal cases of exponentially and uniformly distributed latent values. A particular benefit of the approach is that it provides a feasible solution to not knowing the number of bidders, as in the motivating example of ascending auctions with absentee bidding. Results apply generally when the transaction price and two other highest bids are observed. Applying insights from the statistics literature about the stochastic ordering of adjacent spacings, the paper shows that this delivers an upper bound on the last spacing and hence overcomes the censuring problem of English auctions.

The method is applied to a new dataset of fine wine auctions in which the number of bidders is unknown. Results highlight that even in small samples with large bidding increments, sample spacings allow for the estimation of informative bounds on structural features of interest. For example, consumer surplus is estimated to be between 75-125 percent of the average highest bid. It also turns out that surplus is higher for higher-end wines. Structural estimates are used to evaluate the benefit of relaxing Sotheby's policy that the reserve price has to be lower than the pre-auction value estimate. Results show that the restriction is indeed binding: sellers benefit between 2.5 and 9 percent from setting a reserve equal to 110 percent of the value estimate, and up to 13 percent from setting the reserve to 120 percent of the value estimate. This suggests that the current policy limits sellers to fully leverage the exclusion principle of optimal reserve prices.

# References


Adams, Christopher P. Estimating demand from eBay prices. *International Journal of Industrial Organization*, 25(6):1213–1232, 2007.

Akbarpour, Mohammad and Shengwu Li. Credible auctions: A trilemma. *Econometrica*, 88(2):425–467, 2019.

An, Mark Yuying. Log-concave probability distributions: theory and statistical testing. *Duke University Dept of Economics Working Paper No. 95-03*, 1995.

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

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

Athey, B. Y. Susan and Philip A. Haile. Identification of standard auction models. *Econometrica*, 70(6):2107–2140, 2002.

Bagnoli, Mark and Ted Bergstrom. Log-concave probability and its applications. *Economic Theory*, 26:445–469, 2005.

Barlow, Richard E. and Frank Proschan. Inequalities for linear combinations of order statistics from restricted families. *The Annals of Mathematical Statistics*, 37(6): 1574–1592, 1966.

Bartoszewicz, Jarosław. Dispersive ordering and the total time on test transformation. *Statistics and Probability Letters*, 4(6):285–288, 1986.

Bikhchandani, Sushil, Philip A. Haile, and John G. Riley. Symmetric separating equilibria in English auctions. *Games and Economic Behavior*, 38(1):19–27, 2002.

Cassady, Ralph. *Auctions and auctioneering*. University of California Press, 1967.

Chesher, Andrew and Adam M. Rosen. Identification of the distribution of valuations in an incomplete model of English auctions. *Cemmap Working Paper CWP30/15*, 2015.

Chesher, Andrew and Adam M. Rosen. Generalized instrumental variable models. *Econometrica*, 85(3):959–989, 2017.

Coey, Dominic, Bradley J. Larsen, Kane Sweeney, and Caio Waisman. Ascending auctions with bidder asymmetries. *Quantitative Economics*, 8(1):181–200, 2017.

Coey, Dominic, Bradley Larsen, Kane Sweeney, and Caio Waisman. The simple empirics of optimal online auctions. *NBER Working Paper No. 24698*, 2018.

Coey, Dominic, Bradley Larsen, and Kane Sweeney. The bidder exclusion effect. *RAND Journal of Economics*, (1):93–120, 2019.

Freyberger, Joachim and Bradley J. Larsen. Identification in ascending auctions, with an application to digital rights management. *NBER Working Paper No. 23569*, 2017.

Ginsburgh, Victor. Absentee bidders and the declining price anomaly in wine auctions. *Journal of Political Economy*, 106(6):1302–1319, 1998.

Haile, Philip A. Auctions with resale markets: An application to U.S. forest service timber sales. *The American Economic Review*, 91(3):399–427, 2001.

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

Haile, Philip A., Han Hong, and Matthew Shum. Nonparametric tests for common values at first-price sealed-bid auctions. *NBER Working Paper No. 10105*, 2003.

Hasker, Kevin and Robin Sickles. eBay in the economic literature: Analysis of an auction marketplace. *Review of Industrial Organization*, 37(1):3–42, 2010.

Hernández, Cristián, Daniel Quint, and Christopher Turansick. Estimation in English auctions with unobserved heterogeneity. *RAND Journal of Economics*, 51(3):868–904, 2020.

Jehiel, Philippe and Laurent Lamy. On absolute auctions and secret reserve prices. *RAND Journal of Economics*, 46(2):241–270, 2015.

Jongwoo, Jeon, Subhash Kochar, and Chul Gyu Park. Dispersive ordering - Some applications and examples. *Statistical Papers*, 47(2):227–247, 2006.

Kim, Kyoo and Joonsuk Lee. Nonparametric estimation and testing of the symmetric IPV framework with unknown number of bidders. *Working Paper*, 2014.

Kochar, Subhash. Stochastic comparisons of order statistics and spacings: A review. *International Scholarly Research Network, Article ID 839473*, 2012.

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

Krishna, Vijay. *Auction theory*. Academic Press, 2009.

Larsen, Bradley and Anthony Lee Zhang. A mechanism design approach to identification and estimation. *NBER Working Paper No. 24837*, 2018.

Larsen, Bradley J. The efficiency of real-world bargaining: Evidence from wholesale used-auto auctions. *Review of Economic Studies*, 88(2):851–882, 2021.

Li, Tong and Quang Vuong. Nonparametric estimation of the measurement error model using multiple indicators. *Journal of Multivariate Analysis*, 65(2):139–165, 1998.

Lucking-reiley, David. Vickrey auctions in practice: From nineteenth-century philately to twenty-first-century e-commerce. *The Journal of Economic Perspectives*, 14(3):183–192, 2000.

Luo, Yao and Ruli Xiao. Identification of Auction Models Using Order Statistics. *Working Paper*, 2019.

Mbakop, Eric. Identification of auctions with incomplete bid data in the presence of unobserved heterogeneity. *Working Paper*, 2017.

Milgrom, Paul R and Robert J. Weber. A theory of auctions and competitive bidding. *Econometrica*, 50(5):1089–1122, 1982.

Myerson, Roger B. Optimal auction design. *Mathematics of Operations Research*, 6 (1):58–73, 1981.

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

Paarsch, Harry J. and Han Hong. *An introduction to the structural econometrics of auction data*. MIT press, 2006.

Pinkse, Joris and Karl Schurter. Estimation of auction models with shape restrictions. *Working Paper, arXiv:1912.07466v1*, 2019.

Platt, Brennan C. Inferring ascending auction participation from observed bidders. *International Journal of Industrial Organization*, 54:65–88, 2017.

Pyke, Ronald. Spacings. *Journal of the Royal Statistical Society. Series B (Methodological)*, 27(3):395–449, 1965.

Pyke, Ronald. Spacings revisited. *Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Theory of Statistics*, pages 417–427, 1972.

Quint, Daniel. Unobserved correlation in private-value ascending auctions. *Economic Letters*, 100:432–434, 2008.

Riley, John G. and William F. Samuelson. Optimal Auctions. *Optimal auctions*, 71 (3):381–392, 1981.

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

Rothkopf, Michael H., Thomas J. Teisberg, and Edward P. Kahn. Why are Vickrey auctions rare? *Journal of Political Economy*, 98(1):94–109, 1990.

Song, Unjy. Nonparametric estimation of an eBay auction model with an unknown number of bidders. *Working Paper*, 2004.

Tang, Xun. Bounds on revenue distributions in counterfactual auctions with reserve prices. *RAND Journal of Economics*, 42(1):175–203, 2011.

Thiel, Stuart E. and Glenn H. Petry. Bidding behaviour in second-price auctions: Rare stamp sales, 1923-1937. *Applied Economics*, 27(1):11–16, 1995.

Vickrey, William. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of Finance*, 16(1):8–37, 1961.