Reserve Prices in Art Auctions

Kirill Ponomarev, Bruce Wen*

March 4, 2024

Abstract

This paper introduces a new nonparametric approach to the identification of ascending auctions and applies it to a large novel dataset of live art auctions that we constructed. We apply the approach, which requires only a lower bound on the number of bidders and the top two bids, to our dataset comprising complete bids from more than 2500 live auctions by Sotheby's and Christie's. New reserve prices for modern art sold in New York City are proposed, at approximately the high estimate. Compared to setting the reserve at the low estimate as is common practice today, our proposed reserve increases average expected profits by up to 13.0% of the high estimate, equivalent to US$26,900,000 per auction.

*Kenneth C. Griffin Department of Economics, University of Chicago.

1 Introduction

1.1 Motivation

The auction market for art has large economic significance. In 2023, the two largest art auction houses, Sotheby's and Christie's, reported revenues of US$7.9 billion and USD$6.2 billion respectively. On November 9 and 10, 2022 alone, the 155-work collection of deceased technology billionaire Paul Allen sold for US$1,622,249,500 towards philanthropy, with the most expensive piece *Georges Seurat, Les Poseuses, Ensemble (Petite version), 1888-1890* selling for US$149,240,000.

However, art auctions present a unique empirical challenge to researchers because they are notoriously secretive (see Akbarpour and Li (2020), Marra (2020)). On the websites of major art auction houses, one can only find very basic information including the low and high estimate, and the final transaction price. Important information for analysis such as the trajectory of bids and the number of bidders are kept secret. Even if a bidder were to be physically present during a live auction, she would not be able to have a full grasp on the number of bidders because besides live bidding, alternate forms of bidding exist today in the live auction room including telephone bidding, online bidding, and absentee bidding, all of which are opaque to the live bidder.

Therefore, the literature thus far focuses only on small samples of limited auctions, and few discuss the optimal reserve price for maximizing profits. Ashenfelter and Graddy (2003), Ashenfelter (1989), McAndrew, Smith, and Thompson (2012), Beggs and Graddy (2009) all discuss wine and art auctions, their mechanics, and pricing effects. Hortaçsu and Perrigne (2021) reviews empirical auctions and discuss online eBay and wine auctions. Ashenfelter and Graddy (2011) find that the confidential reserve price is commonly thought to be related to an auctioneer's pre-sale estimates, and that the convention is for it to be at or below the auctioneer's low estimate. Such a common rule gives rise to the suspicion that very likely, this blanket rule to setting reserve prices is not optimal towards maximizing expected profit in certain auction categories. On this note, Marra (2020) analyzed Sotheby's auction data to find an improved reserve price 110% of the estimate that increases revenue by 2.5%, but the data was restricted to a small sample of 884 wine lots from only 1 day of online wine auctions. All of the above works are novel in their data collection and application. However, they do not discuss optimal reserve prices for auctions held in the live art auction room, where most of the high value objects are sold[1].

To overcome the limited data, we adopt a novel approach to data collection by analzying

[1] According to Christie's 2023 press release, $4.6b of their 2023 total $6.2b sales came from live auction rooms.

livestream video data that Sotheby's and Christie's have uploaded to their YouTube channel over the past few years. By applying suitable algorithms to the frame-by-frame video image data on bids and the audio data in the videos, we are able to generate complete bidding trajectories and a lower bound on the number of bidders systematically across a large number of live auctions. We applied such an approach to 65 Sotheby's live auctions and 39 Christie's auctions, generating data on more than 2500 auction lots. These are the first publicly available data collected on live ascending auctions that include complete bidding trajectories and a lower bound on the number of bidders.

A second major challenge in the analysis of art auctions is an uncertain number of bidders. In addressing an unknown number of bidders, the current literature has formulated several approaches to estimate optimal reserves in ascending auctions. Hernández, Quint, and Turansick (2020) develop a method of point-identification for English auctions that allows for unobserved heterogeneity and no need to observe more than the number of bidders and the winning bid in each auction. Freyberger and Larsen (2022) use the (known) reserve price and two order statistics of bids to estimate optimal reserve prices. Marra (2020) uses the stochastic difference between adjacent order statistics, and requires two losing order statistics besides the winning bid. These papers are all innovative in their econometric use of bid data, but do not apply to the data situation in this paper because only the top two order statistics of bids and a lower bound on the number of bidders are observed in our case.

Our econometric approach to nonparametric identification of ascending auctions overcomes the problem of uncertain number of bidders but requires minimal assumptions that are all intuitively obvious in the context of a Sotheby's or Christie's ascending art auction: bidder symmetry, bidders will not bid more than their willingness to pay, a bidder will outbid an opponent if her valuation is higher than the opponent's bid, and the dependence among bidder values is nonnegative. Our paper extends on the econometric works by Haile and Tamer (2003), Quint (2008), Aradillas-López, Gandhi, and Quint (2013), Chesher and Rosen (2017), Ponomarev (2022) by providing sharp bounds on profits when the econometrician only observes the top two bids and a lower bound on the number of bidders. We additionally show how to partially identify the profit difference when a new reserve is chosen relative to a previous reserve price. Even though our bounds are relatively wider with these minimal assumptions, they are sufficiently informative to provide insightful suggestions on a better choice of reserve price. When applied to our data subset of Impressionist and Modern Art in New York City, our proposed reserve increases average expected profits relative to the auction house's typical reserve by up to 13.0% of the high estimate, equivalent to US$26,900,000 per auction.

1.2 Relationship to Literature

This paper contributes to literature in three ways: the empirical literature on art auctions, the methodological literature on nonparametric estimation of English auctions, and the theoretical literature on minimax regret.

The empirical contribution is the construction of a large novel dataset from Christie's and Sotheby's live auctions. Our dataset contains complete bids and a lower bound on the number of bidders from more than 2500 live auction lots covering the largest live auctions run by Christie's and Sotheby's in the last few years. While there exists datasets of final transaction prices (for example in Ashenfelter and Graddy (2011)), or small samples of bids from singular auctions (for example online Sotheby's wine auctions in Marra (2020)), this is the first paper to propose a comprehensive and non-tedious method to collect large amounts of auction data. This is also the first paper to collect bids from Christie's and Sotheby's *live* auctions. We further show that adjusting the reserve price can improve expected profits by the order of millions of dollars per auction in certain subgroups.

The methodological contribution is a novel nonparametric identification approach that builds from Aradillas-López, Gandhi, and Quint (2013)'s and Haile and Tamer (2003)'s methods. In our approach, the econometrician no longer requires an exact number of bidders in the data as was in Aradillas-López, Gandhi, and Quint (2013) but instead only a lower bound on the number of bidders. Even though our bounds are wider in comparison, they are still sufficiently informative in our application to provide insightful reserve price recommendations. We also develop an accompanying linear programming approach to partially identify the change in profit when a new reserve price is chosen relative to a previously set reserve price − this is highly useful in practice since auctioneers would be interested in the profit change if they choose to change their current reserve price.

The theoretical contribution is an analytical solution to the choice of a single optimal reserve price given profit bounds via the minimax regret criterion as first formulated by Savage (1951) and most recently suggested by Manski (2022). Such an approach offers an alternate decision criteria to Aryal and Kim (2013)'s and Jun and Pinkse (2019)'s papers, which provide methods to choose a single optimal reserve price as well. It turns out that the solution to the minimax-regret criterion is exactly identical to Aryal and Kim (2013)'s maxmin approach, which provides further justification to using either method.

The rest of the paper proceeds as follows. In Section II, data construction is discussed in detail. Section III describes the identification argument and estimation strategy, and analytically solves for the choice of a single optimal reserve price using the minimax-regret criterion. Section IV discusses the results. Section V concludes.

2 Data Construction

2.1 Data Source and Collection of Bids

Despite the large number of public auctions that Christie's and Sotheby's have run, they reveal very little information about their auctions. In particular, their public websites only provide data on lot details (e.g. artist name, period, provenance, condition report), low and high estimates, and the final sale price after the buyer's premium. Information such as bids, bidder identities, and number of bidders in past auctions are kept private to the firms.

To overcome this, we adopt a novel data collection process using live-stream auction videos. While not encompassing all auctions, Sotheby's and Christie's have been posting YouTube live-stream videos of their largest live auctions for the past few years, perhaps to increase publicity on their auctions. As of February 2023, we found 65 Sotheby's past live-stream auctions on Sotheby's YouTube channel and 39 Christie's auctions on Christie's YouTube channel. The total objects sold in these videos exceed 3000.

The videos present a large trove of data on bids. They range in length, from short hour-long auctions to marathon 6-hour auctions. Typically, a recorded live-stream video starts with a filler waiting period. Then it proceeds with one to three consecutive auctions, all of which are related to some common theme. For example, as shown in Figure 1, Sotheby's *New York — Monet, Warhol and Basquiat Lead Marquee Evening Sales* auction video comprises of three auctions held in immediate succession at the same Sotheby's New York auction room, in total making up 83 auction lots in a single video alone.

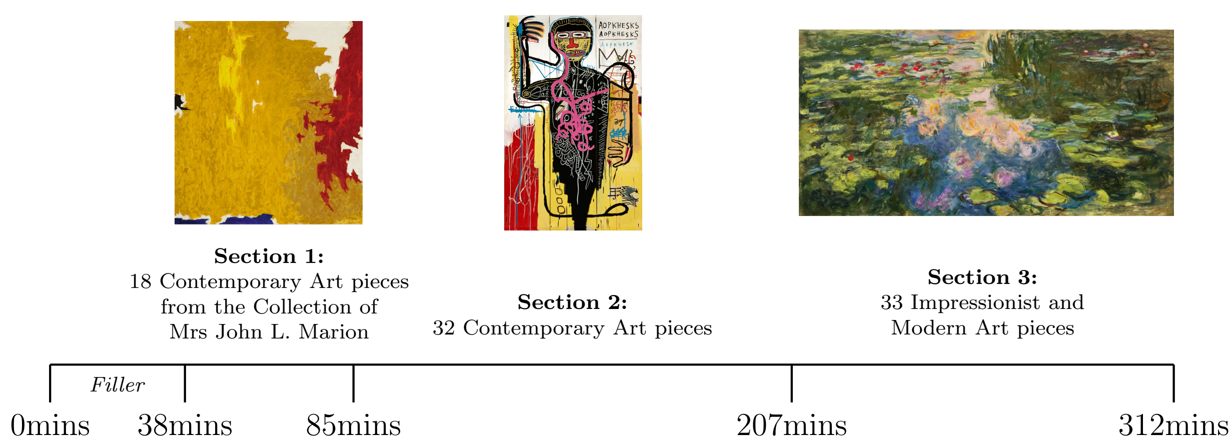


Figure 1: Timeline for Sotheby's *New York — Monet, Warhol and Basquiat Lead Marquee Evening Sales* auction livestream video. The auction video is split into three sections. The images are representative items from each section: *PH-125 (1948-No. 1)* by Clyfford Still (Hammer Price $30.7M), *Versus Medici* by Jean-Michel Basquiat (Hammer Price $50.8M), and *Le Bassin aux nymphéas* by Claude Monet (Hammer Price $70.4M).

We generate complete bidding trajectory with the possible exception of a few starting bids[2] for each auction lot using computer vision techniques. We do so by identifying text content boxes from the videos, such as in Figure 2, at a high frequency of two frames per second.[3] These bid data are paired with scraped data from the auction houses' websites to match the lot and its characteristics such as the estimate and description, ensuring accuracy.

Lot 38, Christie's Hong Kong, Nov 2022

Lot 117, Sotheby's New York, Nov 2022

Figure 2: Example screenshots taken from Christie's and Sotheby's YouTube live stream auctions. Red boxes enclose image areas where important data can be collected. The current bids are clearly seen to be $\$26,000,000$HKD and $\$2,800,000$USD respectively.

2.2 A Close Lower Bound on the Number of Bidders

Unfortunately, the bids collected from the live-stream videos do not have bidder identity mappings. There is also no additional information available on the auction websites that may provide direct knowledge on the number of bidders. Therefore, it is impossible to know the exact number of unique bidders in each auction lot, without proprietary data provided from the auction houses themselves.

However, we can attempt to produce an informative and close lower bound on the number of bidders. Let the number of bidders be denoted by the random variable $N \in \{2, 3, ...\}$. In a live Christie's or Sotheby's auction, there are exactly four sources of bids: (i) telephone bids, (ii) live bids, (iii) absentee bids, and (iv) online bids. A close lower bound to the true number of bidders, $\underline{N} \geq 2$, can be constructed by accurately capturing a certain unique number of bidders in each of these four sources of bids using the audio from the videos.

[2]This is due to a slight lag in the video live-stream and a typically large number of bids around the start of the auction. Thus, some initial bids are missed out and not displayed by the auction house.

[3]This is chosen to exceed the maximum frequency of bid updates in any video.

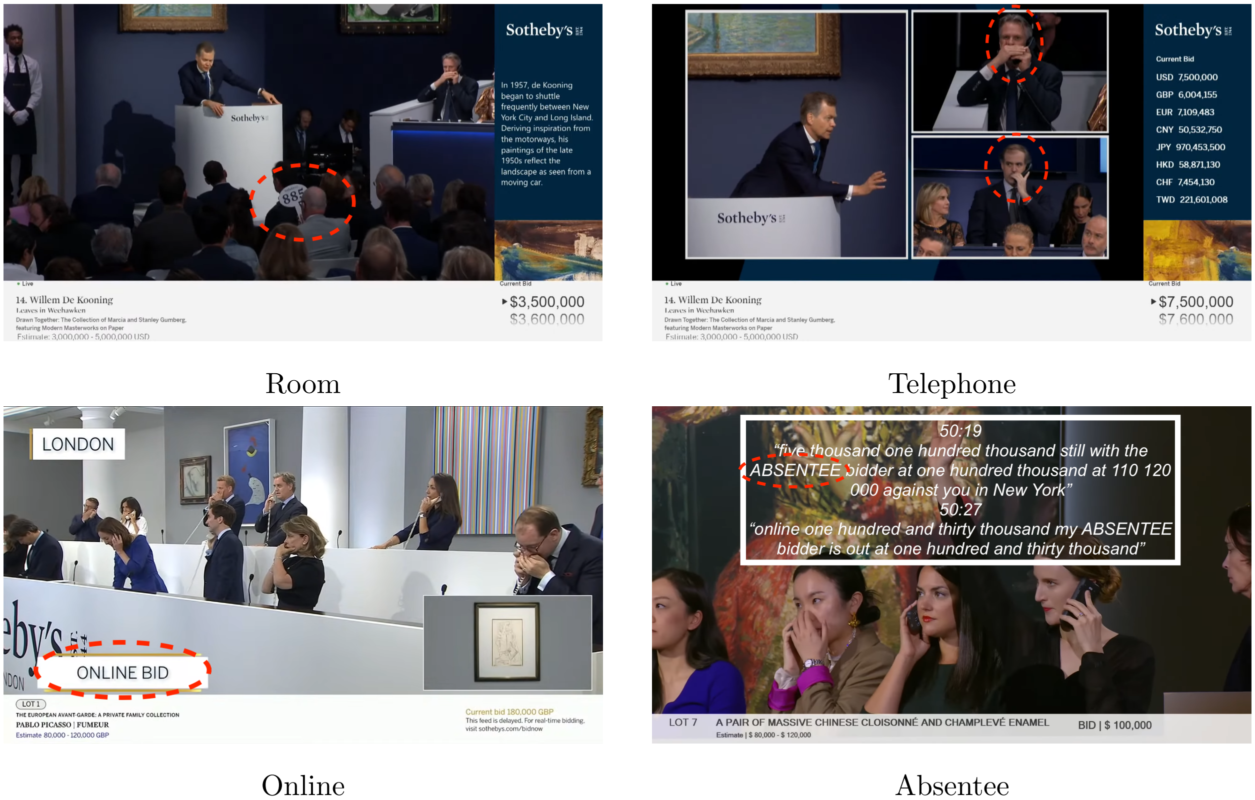


Figure 3: The four types of bidders possible in a Sotheby's or Christie's auction. Dotted red circles indicate the bidder's location. The absentee bidder is captured through the audio transcript.

For (i) telephone bids, our method allows us to capture any number of unique telephone bidders. Auctioneers typically call out specific names of colleagues when they submit bids from some unknown bidder through the telephone. Telephone bids commonly involve an auctioneer calling out a fellow employee working in the same firm. As an example, in figure 4, the auctioneer references Olivier Camu, another employee at Christie's who is putting in telephone bids on behalf of an unknown buyer. These employees are connected on the phone with only one buyer at a time, and there are many employees on the phone with different buyers at an auction. Thus we can construct a lower bound for the number of unique telephone bidders by finding the unique names called out by the auctioneer during an auction lot.

Olivier Camu places a Telephone Bid, \$5m Auctioneer receives Olivier's \$5m bid

Figure 4: Example screenshots, taken within seconds apart at $\approx$ 52min from Christie's *Visionary: The Paul G. Allen Collection Part I, Nov 9 2022* YouTube live stream auction video. Red boxes enclose Olivier Camu, Deputy Chairman Christie's Impressionist & Modern Art Department. On the left, Olivier bids USD \$5m on behalf of an unknown buyer. On the right, auctioneer Adrien Meyer, Christie's Global Head of Private Sales, receives the bid from Olivier, pointing at Olivier and calling out that the current highest bid is *"back with Olivier Camu"*.

For (ii) live bids, it is possible to capture up to 5 unique bidders that submit live bids from the physical auction room. Again, this can be done by exploiting the specific terminology that auctioneers use to address these live bidders. Unlike receiving telephone bids from collegues, whom the auctioneer knows the names of, the auctioneer does not know the names of live bidders sitting in the room. Therefore, the auctioneer uses terms such as "Sir" or "Madam" to address bidders from the room, and positional references such as "back of the room", "to the left", "to the right", etc. By parsing through a list of transcripts, we are able to summarize 5 uniquely identifiable groups of bidders: a) terminology referencing a female bidder, b) terminology referencing a male bidder, c) terminology referencing the back of the auction room, d) terminology referencing the right of the room, e) terminology referencing the left of the room. In order to ensure that the positional references and male/female references do not double count, we only count additional unique bidders if these references are in separate sentences.

For (iii) absentee bids and (iv) online bids, it is possible to capture a single unique online bidder identity and a single unique absentee bidder via the video audio. An online bid or absentee bid is often referenced verbally by the auctioneer (auctioneer would say something such as "(bid)...from online bidder..."); thus when the auctioneer calls out an online/absentee bid at any time, we would know that there is at least one such bidder. Unfortunately, because the auctioneer himself does not know the identity of this bidder [4], it is impossible to distinguish between the different bidders. Thus we are only able to capture

[4]The auctioneer takes in online/absentee bids from a television screen, which does not disclose the identity of the bidder.

the identity of at most one online bidder and one absentee bidder.

Bid Type	**Description**	**Keywords from Audio to Identify Bidder Identities**	**Number of Unique Bidders Identified**
Telephone bid	Telephone bids are live bids placed remotely through an auctioneer attending in person. These are extremely common.	Names of auctioneer's colleagues	No upper bound.
Live bid	Live bids are bids placed in-person by the buyer herself within the auction room on-site.	”sir”/”gentleman”, ”madam”/”lady”, ”back of the room”, ”to the right”/”on my right”, ”to the left”/”on my left”	≤ 5
Absentee bid	Absentee bids are pre-placed bids, at least 24 hours before the auction, that the auctioneer will prioritize over any other equivalent value bids placed during the live auction itself.	”absentee”	≤ 1
Online bid	Online bids are live bids placed on the auction house's website and relayed to the auctioneer in real time via a television screen.	”online”	≤ 1

Table 1: Types of Bids in Auctions and our Upper Bound on Number of Bidders for Each.

In summary, we are able to capture as many unique telephone bidders, one unique online bidder, one unique absentee bidder, and up to 5 unique in-person bidders. These are summarized in Table 1. However, it is unclear how close the observed lower bound $\underline{N}$ is to the true N. In 2021, telephone bids made up 42% of winning bids in Christie's live auctions versus 7% in the saleroom[5]. While the winning bid is not symmetric to our requirement of a conception for all bids, it tells us that our lower bound is informative but still likely slightly less than the true number of bidders.

We now discuss the methodology to capture the bidder identities from video audio. The audio in each video is downloaded then split into per-lot smaller audioclips. To transcribe audio files into text, Google Cloud Platform Speech-to-Text Recognition is chosen due to its competitive results with modern speech recognizers, and its continuous updates using the latest transformers machine learning architecture. Figure 5 shows an example transcript

[5]Source: Christie's website at this link.

(with individual phrases joined together) produced by Google Cloud Speech Recognition from a single auction lot.

> "against you ***calvin*** seven is with my ***online bidder*** here on the screen at seven thousand seven five thank you at seven five at seven thousand five hundred coming in over here at seven thousand five hundred eight thousand at eight thousand dollars at eight thousand dollars eight thousand give me eight five uh eight thousand dollars is bid at eight thousand five eight five thank you at eight five is here at eight five nine thousand ***kelvin*** at nine thousand dollars bit is with ***calvin harvey*** at nine thousand and nine five thank you at nine five should we try one more nine thousand five hundred bit then ***parmigiano*** nine thousand five hundred here on my right and ten back with ***calvin*** at ten thousand dollars try one more eleven thousand thank you at eleven now and twelve and twelve thousand coming in there online twelve thousand is here on my left and thirteen at thirteen thousand dollars with ***alex*** beautiful drawing thirteen thousand dollars what should we do ***calvin*** one more and fourteen yes why not fourteen thousand and fifteen fifteen not with you online here's the bid the front left here with ***calvin*** 15 000 i'm looking for 16 next ***alex*** 15 000 is here 16. thank you 16 with your client there that's 16 000 shall we go on 16 thousand i have at sixteen thousand against the two telephones here and our online bidders and our room sixteen thousand is bid with ***alex*** and i can sell it what should we do at sixteen thousand dollars fair warning to you all last chance at sixteen thousand here it is and i'm selling yours thank you very much paddle number two one one though 102 painted terracotta here in the manor of antonio begarrelli and i will open the bidding for this one at 13 000."

Figure 5: Audio Transcript from a single lot in a Sotheby's Auction. ***Red*** words highlight potential bidders involved in this auction lot.

Consider the example audio transcript, in figure 5, from a section of a live-stream video transcribed via Google Cloud Speech Recognition. Assuming that names can be identified correctly, the example transcript tells us that there are at least 4 bidders in this auction: Calvin=Kelvin, Alex, some online bidder, and "Parmigiano".

In order to capture the unique identities of the telephone bidders, we use a combination of named entity recognition language models to identify names. Such models are able to incorporate the grammatical structure surrounding a name and thus more accurately capture bidder names compared to matching words with a database of names. Two state-of-the-art named entity recognition models are used. The first is RoBERTa by Liu, Ott, Goyal, Du, Joshi, Chen, Levy, Lewis, Zettlemoyer, and Stoyanov (2019), in which the authors robustly optimize the original BERT (Devlin, Chang, Lee, and Toutanova (2019)) model training to achieve state-of-the-art results on GLUE, RACE and SQuAD datasets. The second uses Contextual String Embeddings from Akbik, Blythe, and Vollgraf (2018), in which the authors leverage the internal states of a trained character language model to produce a type of word embeddings. Their trained model reports state-of-the-art F1 scores on the CoNLL03 shared task.

Empirical testing on our dataset of transcripts showed that the length of the string used as the input into these models results in different output names from both models. Therefore, in addition to the list of phrases returned by Google Cloud Speech Recognition for each auction lot audio, we also join the phrases together as a single long paragraph and pass this into both models. The results from these different inputs are union together, improving the models' ability to capture all names from each lot.

In order to capture the unique identities of the live bidders, absentee bidders, and online bidders, we match the strings to a list of keywords as tabulated in Table 1, {"online bidder", "online", "telephone", "sir", "madam", "gentleman", "lady", "phone", "back of the room", "to the right", "to the left", "starting with", "absentee"}.

Finally, we remove names that are similar or that are subsets of one another. For example, the word "online" is a subset of "online bidder", and they should refer to the same individual. As another example, the names "Kelvin" and "Calvin" (from the transcript in figure 5) should refer to the same individual. To identify these duplicates, the Ratcliff and Metzener (1988) Gestalt string matching algorithm is used. It identifies the largest common substring plus recursively the number of matching characters in the non-matching regions on both sides of the longest common substring. The similarity metric can be written as,

$$D = \frac{2K}{|S_1| + |S_2|} \tag{1}$$

where $0 \leq D \leq 1$ is the similarity metric, K is the number of matching characters, and S_1 and S_2 are the two strings. Empirical testing suggests setting $D = 0.6$ to our use case.

We detail the entire algorithm to produce the estimated number of bidders as in Algorithm 1.

Algorithm 1: A CLOSE LOWER BOUND ON N

Inputs: Transcript Segments, Identity Strings
Output: Lower Bound on the Number of bidders, $\underline{N}$

1. Initialize an empty array. For each Transcript Segment:
   - Apply ***RoBERTa*** and ***ContextualStringEmbeddings*** models to extract names.
   - Add strings in the segment that are part of Identity Strings if they are not in the same phrase/sentence as any name.
   - Append these names and strings to the array.
2. Join the Transcript Segments into a single string. Apply ***RoBERTa*** and ***ContextualStringEmbeddings*** on the concatenated string. Append these names into the array.
3. Use the ***Ratcliff-Obershelp*** algorithm on the array to remove similar name duplicates, with a similarity metric $D = 0.6$.
4. Return the length of the final array of names.

The distribution of number of bidders generated from this algorithm is shown in Figure 6.

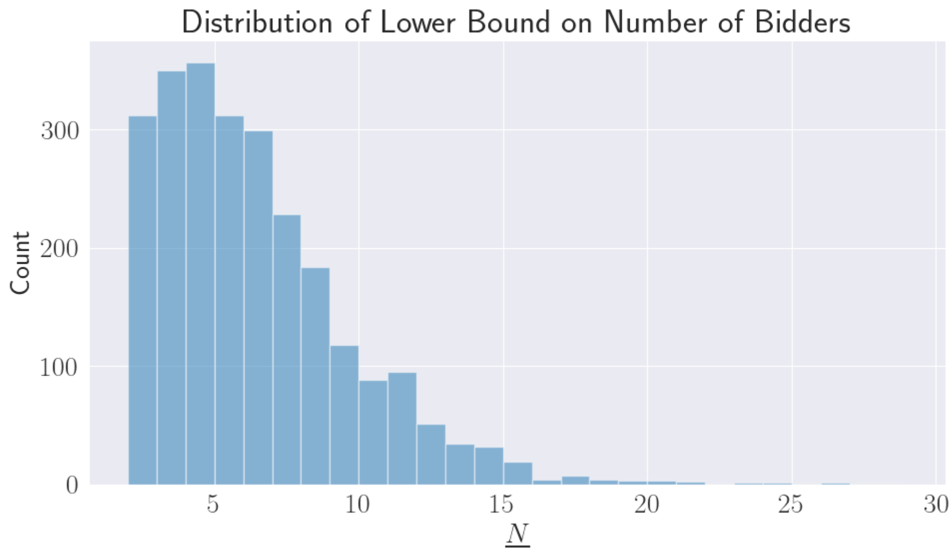


Figure 6: Distribution of lower bound on number of bidders across all auctions.

2.3 Data Cleaning and Overview

Christie's and Sotheby's charge substantial buyer's premiums, which is payable by the successful buyer of an item at an auction based on the hammer price of a lot sold. Table 9 summarizes the Buyer's Premium Schedule, for both Christie's and Sotheby's, as of March 2023. Each recorded bid from the video is thus scaled according to the buyer's premium.

To clean the data, auction lots with only one bid, estimated number of bidders > number of bids, number of bidders less than two, and high estimate < low estimate are removed. Auction lots with insensible bid increments (i.e. > 10× increment) are also removed. Each auction lot is matched with its good category by scraping the auction websites. Table 2 reports the counts of items in each category.

After data cleaning, we produce a combined dataset consisting of 2505 auction lots. The summary statistics are shown in Table 3 and Table 4. The distribution of transaction prices is shown in Figure 7.

Next we bin the data into smaller sub-sample groups to take into account how auction characteristics might affect the price. Consider the following functional determinants of price,

$$P = f(V, N, \Xi)$$

where V is the vector of bidder valuations, N is the number of bidders, and Ξ is the vector of covariates. Here, Ξ can include the following:

1. **Category of Art**, e.g. Impressionist vs. Contemporary vs. Chinese Art.

Category	Subcategory	Location	Count
Art	Chinese Art	New York	112
Art	Impressionist/20th/21st Century Art	Hong Kong	398
Art	Impressionist/20th/21st Century Art	Las Vegas	9
Art	Impressionist/20th/21st Century Art	London	341
Art	Impressionist/20th/21st Century Art	New York	537
Art	Impressionist/20th/21st Century Art	Paris	215
Art	Impressionist/20th/21st Century Art	Shanghai	55
Art	Old Masters	London	47
Art	Old Masters	New York	72
Designer Furniture	n/a	New York	41
Designer Furniture	n/a	Paris	184
Luxury Goods	Jewelry	New York	125
Luxury Goods	Watches	New York	192
Treasures	n/a	London	22
Wines and Spirits	Whisky	Edinburgh	16
Others			139
Total			2505

Table 2: Table of Category Counts.

Variable	Median	Mean	Std	Min-Max
Transaction Price	1.33	2.14	3.88	0.45 - 76.38
2nd-Highest Bid	1.26	1.96	3.60	0.42 - 75.80
Number of Bidders	6.00	6.45	3.39	2.00 - 23.00
Low Est. Relative to High Est.	0.67	0.67	0.07	0.44 - 1.00
Number of Bids	11.00	12.72	8.78	2.00 - 64.00
Number of Auction Lots		660		

Table 3: Summary Statistics for Christie's Auctions

Variable	Median	Mean	Std	Min-Max
Transaction Price	1.36	2.22	4.37	0.34 - 134.40
2nd-Highest Bid	1.27	1.96	3.79	0.09 - 126.00
Number of Bidders	5.00	5.71	3.23	2.00 - 26.00
Low Est. Relative to High Est.	0.67	0.67	0.07	0.33 - 0.89
Number of Bids	9.00	11.36	8.55	2.00 - 88.00
Number of Auction Lots		1,845		

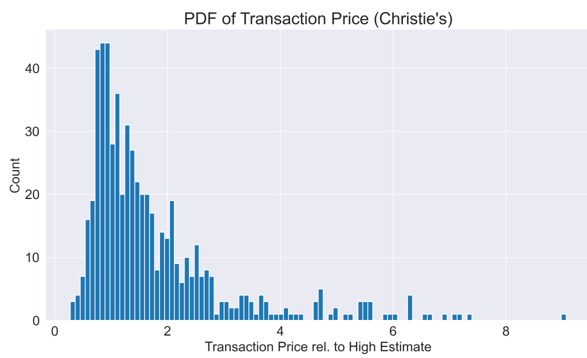
Table 4: Summary Statistics for Sotheby's Auctions

2. **Location of Auction**, e.g. New York, Paris, London, Shanghai.

3. **Degree of Competition**, organized into three percentile groups.

4. **High Estimates** as provided by Auction House.

Based on the possible realizations of Ξ, we create sub-sample groups of auctions. We first group by category and location (as in table 2). The degree of competition is determined by the number of bids, and is filtered by three groups of percentiles. The same is done for the high estimates provided by the auction house to produce low, middle, and high groups. Sample groups with small sizes are removed.



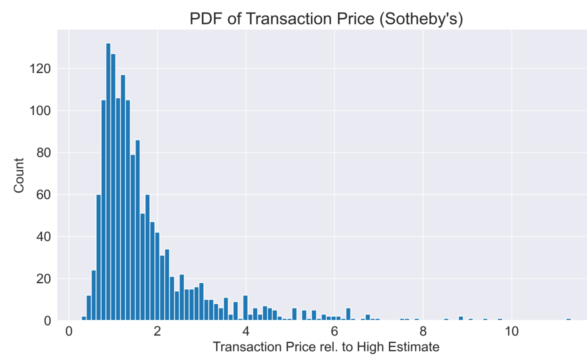


Figure 7: Distribution of Final Bids, with Buyer's Premium included. Prices $\geq$ 3 standard deviations from the mean are removed for graph visibility.

3 Identification and Estimation

3.1 Environment and Bidding Behavior

Here, the theoretical framework is introduced. Let N (a random variable) denote the number of bidders in an auction and let n denote a value in the support of N. The following assumptions 1, 2, 3 are due to Haile and Tamer (2003) and Aradillas-López, Gandhi, and Quint (2013) and are maintained throughout the paper.

Assumption 1 *Bidders behave as follows:*

a. Bidders do not bid more than they are willing to pay.

b. Bidders do not allow an opponent to win at a price they are willing to beat.

This assumption is taken from Haile and Tamer (2003). Assumption 1a says that no bidder should bid above his valuation, i.e. $b_{i:n} \leq v_{i:n}$. The motivation for this is clear, since bidding above one's valuation would be a strictly negative expected utility play, with the negative utility realized when the bid actually wins that auction lot. Meanwhile, 1b implies that $v_{n-1:n} \leq b_{n:n}$. The motivation for this assumption is also intuitively clear; that bidders do not pass up an opportunity to make profit.

Assumption 2 *Bidders have symmetric private values.*

This assumption is very standard in the literature. Suppose that in a n-bidder auction, $(V_1, V_2, ..., V_n)$ are the private values of the bidders and $\boldsymbol{F}^n$ represent the bidders' joint probability distribution. Then, any rearrangement of the valuations $\sigma : \{1, ..., n\} \rightarrow \{1, ..., n\}$ preserves $\boldsymbol{F}^n(v_1, ..., v_n) = \boldsymbol{F}^n\left(v_{\sigma(1)}, ..., v_{\sigma(n)}\right)$.

Assumption 3 *For each n, the joint distribution $\boldsymbol{F}^n$ is such that for any v and i, the probability $\mathbb{P}(V_i < v \mid N = n, |j \neq i : V_j < v| = k)$ is nondecreasing in k.*

This assumption is due to Aradillas-López, Gandhi, and Quint (2013). It is sufficiently general to nest all standard models of correlated private values, including (i) *symmetric, affiliated private values*, (ii) *symmetric, conditionally independent private values*, or (iii) *symmetric, independent private values with unobserved heterogeneity.* This assumption is particularly apt in our application to art auctions because bidders' valuations are likely correlated due to the highly transparent nature of live English auction and the ability to resell.

3.2 Identification

In this section, we detail a partially identified model on expected profit that extends on the works by Haile and Tamer (2003) and Aradillas-López, Gandhi, and Quint (2013). In particular, our proposed model requires only a lower bound on the number of bidders instead of the exact number of bidders as was in either paper. We will also introduce a partially identified model on expected profit difference between any new reserve price and a previously set reserve.

3.2.1 Partial Identification of Profit

Let N be the number of bidders, a random variable. Define the random variables $B_{1:n}, B_{2:n}..., B_{n:n}$ as the order statistics of the maximum bid by each of bidders conditional on there being $N = n$ bidders, with $b_{i:n}$ denoting the realization of the i^{th} lowest of the n bidders' maximum bids. Let $G_{i:n}$ denote the distribution of $B_{i:n}$. Similarly, let $V_{1:n}, V_{2:n}, ..., V_{n:n}$ denote the ordered valuations of the bidders (again conditional on there being n bidders), with each

$V_{i:n} \sim F_{i:n}$, the cumulative distribution function. Note that $b_{i:n}$ need not be the bid made by the bidder with valuation $v_{i:n}$.

Letting r denote a reserve price, and v_0 denote the value of the unsold lot to the seller, the profit is given by:

$$\pi(r) = (r - v_0) \cdot \mathbb{1}(V_{N-1:N} \leq r, V_{N:N} > r) + (V_{N-1:N} - v_0) \cdot \mathbb{1}(V_{N-1:N} \geq r). \tag{2}$$

Taking expectations conditional on $N = n$ and rearranging:

$$\mathbb{E}[\pi(r)|N = n] = \int_0^{+\infty} \max\{r, v\} dF_{n-1:n}(v) - v_0 - F_{n:n}(r)(r - v_0). \tag{3}$$

Therefore, to study optimal reserve prices, it suffices to identify or bound the distributions $F_{n:n}$ and $F_{n-1:n}$.

Identification of $F_{n-1:n}$ follows from assumption 1. By assumption 1a, we get $b_{n-1:n} \leq v_{n-1:n}$, so it follows that $F_{n-1:n}(v) \leq G_{n-1:n}(v)$. By assumption 1b, we get $v_{n-1:n} \leq b_{n:n}$, so it follows that $G_{n:n}(v) \leq F_{n-1:n}(v)$. Combining these two inequalities, the pointwise bounds for $F_{n-1:n}(v)$ can thus be identified:

$$G_{n:n}(v) \leq F_{n-1:n}(v) \leq G_{n-1:n}(v). \tag{4}$$

For identification of $F_{n:n}$, define the strictly increasing differentiable function $\phi_{i:n}(H) : [0, 1] \to [0, 1]$ as the implicit solution to $H = \frac{n!}{(n-i)!(i-1)!} \int_0^{\phi} s^{i-1}(1-s)^{n-i} ds$. Then by assumptions 2 and 3 (see Aradillas-López, Gandhi, and Quint (2013) for derivation), the bounds for $F_{n:n}$ are given by:

$$\phi_{n-1:n}(G_{n:n}(v))^n \leq F_{n:n}(v) \leq G_{n:n}(v). \tag{5}$$

Ponomarev (2022) showed that the bounds for $F_{n-1:n}(v)$ (see 4) and $F_{n:n}(v)$ (see 5) are sharp. Thus, the sharp bounds for the profit function conditional on $N = n$ are:

$$\mathbb{E}[\pi(r)|N = n] \geq \int_0^{\infty} \max\{r, v\} dG_{n-1:n}(v) - v_0 - G_{n:n}(r)(r - v_0), \tag{6}$$

$$\mathbb{E}[\pi(r)|N = n] \leq \int_0^{\infty} \max\{r, v\} dG_{n:n}(v) - v_0 - \phi_{n-1:n}(G_{n:n}(r))^n(r - v_0). \tag{7}$$

We now generalize the bounds so that they are unconditional on the number of bidders. Let $F_{\mathcal{I}}$ and $F_{\mathcal{II}}$ denote the unconditional CDFs of $V_{n:n}$ and $V_{n-1:n}$ correspondingly. Then, the general form of the profit equation in equation 3 unconditional on N is:

$$\mathbb{E}[\pi(r)] = \int_0^{+\infty} \max\{r, v\} dF_{\mathcal{II}}(v) - v_0 - F_{\mathcal{I}}(r)(r - v_0). \tag{8}$$

Now we introduce the following theorem to bound 8 for a scenario where we only observe a lower bound on the number of bidders.

Theorem 1 (Bounds on Expected Profit Unconditional on N) *Let $\underline{n}$ be the minimum number of bidders in the entire auction population. Further let $G_{\mathcal{I}}$ and $G_{\mathcal{II}}$ denote the unconditional CDFs of $B_{n:n}$ and $B_{n-1:n}$ correspondingly. Under assumptions 1, 2, 3, the sharp bounds for the expected profit unconditional on N are:*

$$\mathbb{E}[\pi(r)] \geq \int_0^{+\infty} \max\{r, v\} dG_{\mathcal{II}}(v) - v_0 - G_{\mathcal{I}}(r)(r - v_0), \tag{9}$$

$$\mathbb{E}[\pi(r)] \leq \int_0^{+\infty} \max\{r, v\} dG_{\mathcal{I}}(v) - v_0 - \phi_{\underline{n}-1:\underline{n}}(G_{\mathcal{I}}(r))^{\underline{n}}(r - v_0). \tag{10}$$

Proof. By our assumptions, for each n and for all v,

$$\begin{aligned} G_{n:n}(v) \leq F_{n-1:n}(v) &\leq G_{n-1:n}(v), \\ \phi_{n-1:n}(G_{n:n}(v))^n \leq F_{n:n}(v) &\leq G_{n:n}(v). \end{aligned}$$

Taking expectations in the first line, we obtain:

$$G_{\mathcal{I}}(v) \leq F_{\mathcal{II}}(v) \leq G_{\mathcal{II}}(v).$$

For the second line, note that the function $f(t, n) = \phi_{n-1:n}(t)^n$ is increasing in n for all t and is convex in t for all n. Therefore, for each $n \geq \underline{n}$,

$$\phi_{\underline{n}-1:\underline{n}}(G_{n:n}(v))^{\underline{n}} \leq F_{n:n}(v) \leq G_{n:n}(v),$$

and by Jensen's inequality,

$$\phi_{\underline{n}-1:\underline{n}}(G_{\mathcal{I}}(v))^{\underline{n}} \leq F_{\mathcal{I}}(v) \leq G_{\mathcal{I}}(v).$$

■

Theorem 1 is useful because it allows us to bound the profit function while only knowing a lower bound on the number of bidders and not the exact number. An accompanying benefit is that it also pools together a larger amount of data, that is, all auctions with at least $\underline{n}$ bidders. Note that in fact, only the lower bounds changes; the expression for the upper bound does not depend on the number of bidders. Theorem 1 is especially relevant to our empirical application because we only observe a very close lower bound to the true number of bidders rather than an exact number of bidders from the auction videos.

3.2.2 Partial Identification of Profit Difference

Consider the change in profit when a new reserve price r is chosen relative to a previously set reserve price r_0. Using equation 3, the expected profit difference is given by:

$$\mathbb{E}[\pi(r)-\pi(r_0)|N=n] = \mathbb{E}\left[\max\{r, V_{n-1:n}\} - \max\{r_0, V_{n-1:n}\}\right] - (r-v_0)F_{n:n}(r) + (r_0-v_0)F_{n:n}(r_0). \tag{11}$$

In order to bound the above expression, it might be tempting to directly apply the bounds derived in inequalities 4 and 5 on $F_{n:n}$ and $F_{n-1:n}$. However, this is insufficient because not all tuples $(F_{n:n}(r_0), F_{n:n}(r))$ are permissible given some tuple $(r_0, r) \in [0, \bar{v}]^2$. In particular, we also require the monotone non-decreasing property of cumulative distribution functions. In Figure 8, we give visual examples of both permissible and impermissible $F_{n:n}$ cumulative distribution functions that would otherwise satisfy inequalities 4 and 5.

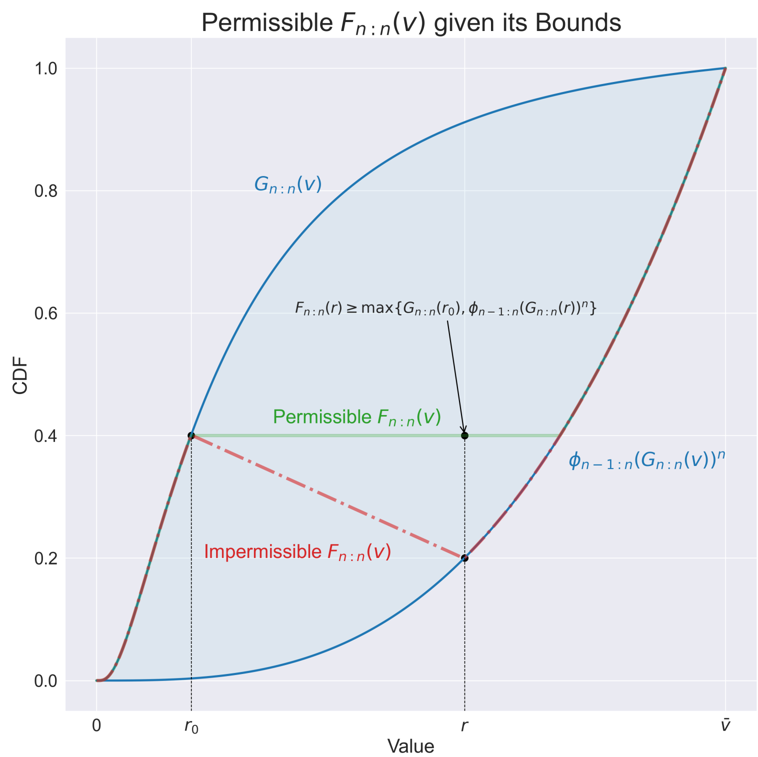


Figure 8: Given the bounds on $F_{n:n}$ from inequality 5, not all tuples $(F_{n:n}(r_0), F_{n:n}(r))$ are permissible. For example, the red CDF is impermissible. In this example where $r > r_0$, we need to additionally impose that $F_{n:n}(r) \geq F_{n:n}(r_0)$.

We solve the issue by imposing an additional inequality through linear programming. Then, the profit difference is bounded by:

$$\mathbb{E}[\pi(r)-\pi(r_0)|N=n] \geq \mathbb{E}\left[\max\{r, B_{n-1:n}\}\right] - \mathbb{E}\left[\max\{r_0, B_{n:n}\}\right] + \underline{\Lambda}^\star(r, r_0); \tag{12}$$

$$\mathbb{E}[\pi(r)-\pi(r_0)|N=n] \leq \mathbb{E}\left[\max\{r, B_{n:n}\}\right] - \mathbb{E}\left[\max\{r_0, B_{n-1:n}\}\right] + \bar{\Lambda}^\star(r, r_0); \tag{13}$$

where $\underline{\Lambda}^\star(r, r_0), \bar{\Lambda}^\star(r, r_0)$ solve the linear programs in Table 5.

Case	**Solution to** $\underline{\Lambda}^{\star}(r, r_0)$	**Solution to** $\bar{\Lambda}^{\star}(r, r_0)$
$r \geq r_0$	$\min_{\alpha,\beta} \ -(r - v_0)\alpha + (r_0 - v_0)\beta$ s.t. $\phi_{n-1:n}\left(G_{n:n}(r)\right)^n \leq \alpha \leq G_{n:n}(r)$ $\phi_{n-1:n}\left(G_{n:n}(r_0)\right)^n \leq \beta \leq G_{n:n}(r_0)$ $\alpha \geq \beta$	$\max_{\alpha,\beta} \ -(r - v_0)\alpha + (r_0 - v_0)\beta$ s.t. $\phi_{n-1:n}\left(G_{n:n}(r)\right)^n \leq \alpha \leq G_{n:n}(r)$ $\phi_{n-1:n}\left(G_{n:n}(r_0)\right)^n \leq \beta \leq G_{n:n}(r_0)$ $\alpha \geq \beta$
$r \leq r_0$	$\min_{\alpha,\beta} \ -(r - v_0)\alpha + (r_0 - v_0)\beta$ s.t. $\phi_{n-1:n}\left(G_{n:n}(r)\right)^n \leq \alpha \leq G_{n:n}(r)$ $\phi_{n-1:n}\left(G_{n:n}(r_0)\right)^n \leq \beta \leq G_{n:n}(r_0)$ $\beta \geq \alpha$	$\max_{\alpha,\beta} \ -(r - v_0)\alpha + (r_0 - v_0)\beta$ s.t. $\phi_{n-1:n}\left(G_{n:n}(r)\right)^n \leq \alpha \leq G_{n:n}(r)$ $\phi_{n-1:n}\left(G_{n:n}(r_0)\right)^n \leq \beta \leq G_{n:n}(r_0)$ $\beta \geq \alpha$

Table 5: Linear Programs for the Point-wise Bounds on Profit Difference.

3.3 Estimation

Let $T_{\underline{n}}$ be the total number of auctions in the sample with at least $\underline{n}$ bidders, h the smoothing parameter for a Gaussian kernel $\mathcal{K}$, and $b^i_{\mathcal{II},\underline{n}}$, $b^i_{\mathcal{I},\underline{n}}$ the second highest bid and highest bid in the i^{th} auction with at least $\underline{n}$ bidders respectively. Further let $c : \mathbb{R}_+ \to \mathbb{R}_+$ be a scaling function of bids, as defined in Table 9, to account for the auction house buyer's premium.

We estimate the profit bounds using a Kernel Density Estimator, with the probability densities of the top two order statistics being:

$$\hat{g}_{\mathcal{II},\underline{n}}(v) = \frac{1}{hT_{\underline{n}}} \sum_{i=1}^{T_{\underline{n}}} \mathcal{K}\left(\frac{v - c\left(b^i_{\mathcal{II},\underline{n}}\right)}{h}\right), \tag{14}$$

$$\hat{g}_{\mathcal{I},\underline{n}}(v) = \frac{1}{hT_{\underline{n}}} \sum_{i=1}^{T_{\underline{n}}} \mathcal{K}\left(\frac{v - c\left(b^i_{\mathcal{I},\underline{n}}\right)}{h}\right). \tag{15}$$

Further define the CDF $\hat{G}_{\mathcal{I},\underline{n}}$ corresponding to the probability density $\hat{g}_{\mathcal{I},\underline{n}}$. Then, the pointwise estimated bounds are,

$$\hat{\pi}_{\underline{n}}(r) \geq \int_0^{\infty} \max\{r, v\} \hat{g}_{\mathcal{II},\underline{n}}(v) dv - v_0 - \hat{G}_{\mathcal{I},\underline{n}}(r)(r - v_0) \tag{16}$$

$$\hat{\pi}_{\underline{n}}(r) \leq \int_0^{\infty} \max\{r, v\} \hat{g}_{\mathcal{I},\underline{n}}(v) dv - v_0 - \phi_{\underline{n}-1:\underline{n}}\left(\hat{G}_{\mathcal{I},\underline{n}}(r)\right)^{\underline{n}}(r - v_0) \tag{17}$$

For kernel density estimation, the bandwidth h is determined by the Improved Sheather-Jones (ISJ) method as in Botev, Grotowski, and Kroese (2010) to reduce out of sample prediction error. It is chosen for its ability to fit multimodal data much better than the original Silverman rule. ϕ is computed using the Powell (1964) conjugate direction numerical optimization method.

95% pointwise confidence intervals (CI) are constructed analytically and applied to the kernel density estimates. Let $\hat{\underline{\pi}}_{\underline{n}}(r)$ and $\hat{\bar{\pi}}_{\underline{n}}(r)$ be the estimated lower and upper bounds on profits respectively. Let $\hat{\underline{\sigma}}_{\underline{n}}(r)$ and $\hat{\bar{\sigma}}_{\underline{n}}(r)$ be the standard deviation of the lower and upper bounds on profit computed using the Delta method. This involves the random vectors $\left(\begin{smallmatrix} \max\{r, B_{\mathcal{I}\mathcal{I},\underline{n}}\} \\ \mathbb{1}(B_{\mathcal{I},\underline{n}} \leq r) \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} \max\{r, B_{\mathcal{I},\underline{n}}\} \\ \mathbb{1}(B_{\mathcal{I},\underline{n}} \leq r) \end{smallmatrix}\right)$ for the lower and upper bounds on profit respectively.

Following Imbens and Manski (2004) and Stoye (2009), the CI is computed using:

$$\mathrm{CI}_{1-\alpha}(\pi_{\underline{n}}(r)) = \left[\hat{\underline{\pi}}_{\underline{n}}(r) - c_\alpha \cdot \frac{\hat{\underline{\sigma}}_{\underline{n}}(r)}{\sqrt{T_{\underline{n}}}}, \hat{\bar{\pi}}_{\underline{n}}(r) + c_\alpha \cdot \frac{\hat{\bar{\sigma}}_{\underline{n}}(r)}{\sqrt{T_{\underline{n}}}}\right], \tag{18}$$

where c_α solves

$$\Phi\left(c_\alpha + \frac{\sqrt{T_{\underline{n}}}\left(\hat{\bar{\pi}}_{\underline{n}}(r) - \hat{\underline{\pi}}_{\underline{n}}(r)\right)}{\max\{\hat{\underline{\sigma}}_{\underline{n}}(r), \hat{\bar{\sigma}}_{\underline{n}}(r)\}}\right) - \Phi(-c_\alpha) = 1 - \alpha, \tag{19}$$

and Φ is the standard normal cumulative distribution function, $T_{\underline{n}}$ is the number of auctions in the sample, α is the significance level.

3.4 Selecting a Single Optimal Reserve Price

While the approach so far allows us to generate informative bounds, an auction house must make a decision on a single reserve price for an auction. We will now setup a minimax regret problem for profit and proceed to solve it analytically.

Such decision problems applied to partially identified English auction models have received substantial research. One approach is the max-min solution, which is directly the max of the lower bound as in Aryal and Kim (2013). Another maximum entropy approach was suggested by Jun and Pinkse (2019), which additionally uses information about the upper bound as well. Our approach follows from Manski (2022)'s formulation of the minimax regret criterion, but applied to the partially identified English auction scenario. We will show that the solution is equivalent to the maxmin approach, thus giving more justification to either approach.

Suppose $v_0 \in \mathbb{R}$ is an auctioneer's valuation of a good, and $\bar{v} \in \mathbb{R}$ is some arbitrarily high number exceeding all possible valuations of bidders. Let:

- $\mathcal{C} = [v_0, \bar{v}]$ be the choice set of reserve prices;

- $\Pi = \{\pi : \pi_L \leq \pi \leq \pi_U\}$ be the space of possible profit functions $\pi(r) : \mathcal{C} \to \mathbb{R}$, bounded by $\pi_L : \mathcal{C} \to \mathbb{R}$ and $\pi_U : \mathcal{C} \to \mathbb{R}$.

Now we introduce an assumption on the shape of the profit functions.

Assumption 4 (Continuity of Profit Function and Bounds) *$\pi(\cdot)$, π_L and π_U are continuous on $[v_0, \bar{v}]$.*

This assumption is very standard.

Given the above problem set-up, with no additional information on the probabilities of the profit function being at a certain region within the bounds, the minimax-regret criterion can be stated as:

$$\min_{r \in \mathcal{C}} \max_{\pi \in \Pi} \left\{ \max_{d \in \mathcal{C}} \{\pi(d)\} - \pi(r) \right\}; \tag{20}$$

where d represents a decision within the choice set $\mathcal{C}$. The term $\max_{d \in \mathcal{C}} \{\pi(d)\} - \pi(r)$ represents the regret given a particular choice of a reserve and a profit function within the profit bounds.

The goal is to find the set of reserve prices $\mathcal{R}^*$ that solves equation 20,

$$\mathcal{R}^* = \operatorname*{argmin}_{r \in \mathcal{C}} \max_{\pi \in \Pi} \left\{ \max_{d \in \mathcal{C}} \{\pi(d)\} - \pi(r) \right\}$$

i.e. the reserve prices that minimizes maximum regret on profit that could have been attained given the bounds π_L, π_U and assumption 4.

One natural way to approach solving for $\mathcal{R}^*$ is as follows. Given a particular fixed $r \in \mathcal{C}$, we can attempt to characterize a regret-maximizing profit function that attains its lowest point at r and attains its maximum point at $\operatorname{argmax}_{\mathcal{C}} \pi_U$. The following Lemma helps make the feasibility of such an idea concrete, albeit with a restriction.

Lemma 1 (Feasibility of π^*) *Fix some $p, q \in \mathcal{C}$, where $p \neq q$. $\exists$ a continuous function $\pi^*(r) : \mathcal{C} \to \mathbb{R}$ such that,*

1. $\pi_L \leq \pi^*(r) \leq \pi_U \; \forall r \in \mathcal{C}$,

2. $\pi^*(p) = \pi_L(p)$, *and,*

3. $\pi^*(q) = \pi_U(q)$.

Proof. Use a linear combination of π_L and π_U to construct π^*, with a convex combination of π_L and π_U in between p and q. ■

The construction in Lemma 1 can be visualized in the following example graph, where it is the case that $p < q = \operatorname{argmax} \pi_U$. It is intuitive that such a construction of π^* indeed maximizes the regret given the choice of a reserve price p, since we attained the lowest possible profit at p despite the best possible profit of $\max \pi_U$.

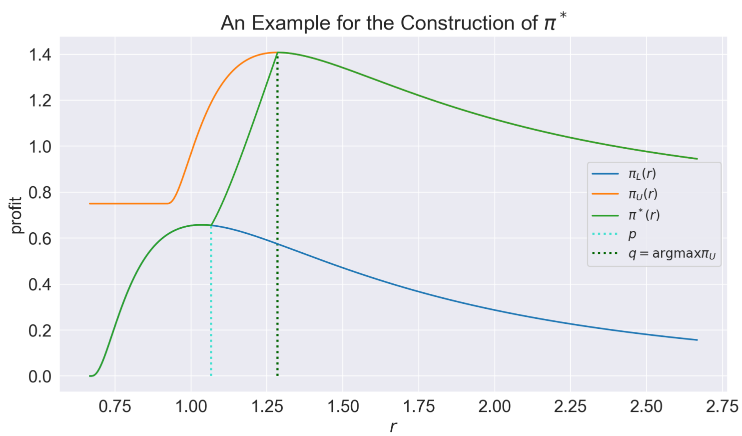


Figure 9: An example of a feasible π^* in Lemma 1

It is not difficult to generalize this idea for all $p, q \in \mathcal{C}$ even if they are equal, which would happen when we attempt to compute the maximum regret at $\operatorname{argmax} \pi_U$. The following theorem characterizes the reserve price solution that minimizes maximum regret, and the proof extends the ideas from above.

Theorem 2 (Minimax-Regret Solution for Profit) *Let $\mathcal{C} = [v_0, \bar{v}]$ and assume that π_U and π_L are continuous on $\mathcal{C}$. Let $\Pi = \{\pi : \pi_L \leq \pi \leq \pi_U\}$ be the space of possible profit functions $\pi(r) : \mathcal{C} \to \mathbb{R}$. Then the solution to the minimax-regret problem is $\operatorname*{argmax}_{r \in \mathcal{C}} \pi_L(r)$, i.e.*

$$\operatorname*{argmin}_{r \in \mathcal{C}} \max_{\pi \in \Pi} \left\{ \max_{d \in \mathcal{C}} \{\pi(d)\} - \pi(r) \right\} = \operatorname*{argmax}_{r \in \mathcal{C}} \pi_L(r)$$

Proof. See Appendix A. ■

4 Results

The collected data spans many categories of auctions, but we focus on the category of Impressionist/20th/21st Century Art sold in New York City. This is due to the high monetary significance of these pieces, averaging a transaction price of USD$6.4m, and the large number of observations, 537.

We estimate the bounds on expected seller profit, $\mathbb{E}[\pi(r)]$, and the bounds on expected profit difference, $\mathbb{E}[\pi(r) - \pi(r_0)]$ relative to setting the reserve r_0 at the low estimate as is

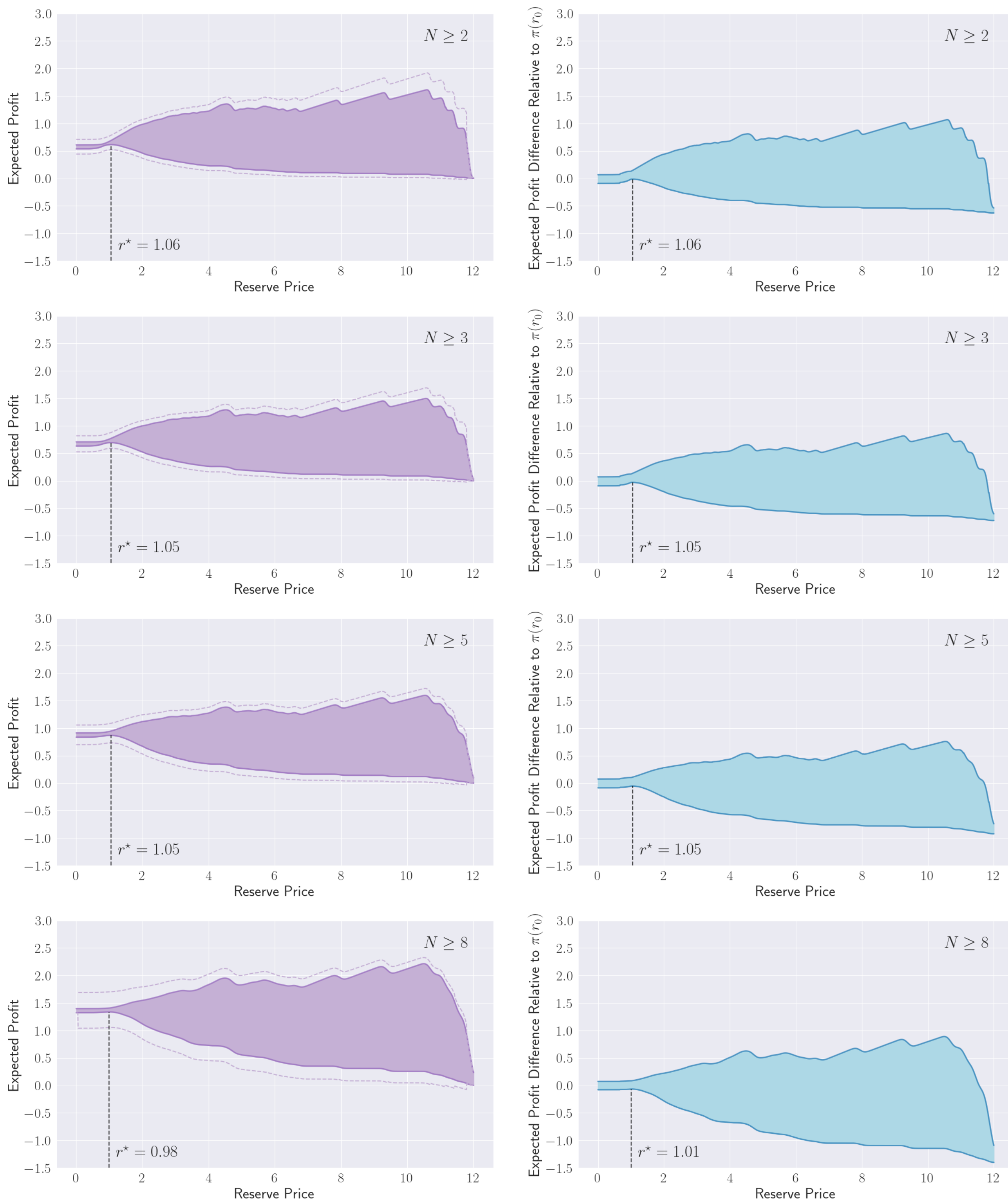


Figure 10: Bounds on expected profit (left) and profit difference (right) against reserve price for Impressionist/20th/21st Century Art sold in New York City. The number of bidders is estimated to be at least 2, 3, 5, and 8.

common practice by Christie's and Sotheby's. We assume that v_0, the seller's valuation, is equal to the high estimate provided by the auction house.

Figure 10 shows our estimates of the bounds on profit and profit difference in solid lines, as well as the pointwise 95% confidence interval in dashed lines, plotted against reserve price for different values on the lower bound on the number of bidders. The y-axes (expected profit and expected profit difference) are scaled by the average high estimate for the respective auction samples. We display the graphs for at least 2, 3, 5, and 8 bidders.

The following patterns are apparent from these graphs:

- When there are at least a low number of bidders, such as $N \geq 2$ or $N \geq 3$, the minimax-regret choice of reserve *almost guarantees* a profit increase relative to the original reserve r_0 at the low estimate. This is in the sense that the lower bound on profit difference at the minimax-regret choice is very close to zero: -0.013 and -0.030 of the average high estimate for $N \geq 2$ and $N \geq 3$ respectively. Meanwhile, the upper bound on profit difference is positive and large: 0.149 and 0.134 of the average high estimate respectively.

- Both the upper and lower bounds on expected profit increase as the number of bidders increases.

- The minimax-regret choice on the optimal reserve price decreases as the number of bidders increases.

We can further quantify the policy significance of setting the reserve price at our proposed minimax-regret choice and compare it with the auctioneer's reserve price, which is typically set at the low estimate[6]. Table 6 shows our proposed minimax-regret reserve for each sample, the expected profit range, expected profit difference range relative to r_0, and their 95% confidence intervals. The profit values are scaled by the average transaction price for the respective auction samples.

Consider the case of $N \geq 2$ (the first column of the table) as an example and notice the following facts:

- The first row shows the average high estimate across all lots in the sample.

- The second row shows that the minimax-regret choice of reserve price is 1.057 times the high estimate, which is $1.057 * \frac{3}{2} = 1.59$ times the auctioneer's typical reserve price.

- The third row shows the bounds on the average expected profit when selecting the minimax-regret reserve price. Here, the expected profit range is \$3.49m to \$3.93m.

[6] This is almost always $\frac{2}{3}$ the high estimate, which we verified with our data.

Number of Bidders	$N \geq 2$	$N \geq 3$	$N \geq 4$	$N \geq 5$
Avg High Estimate	\$5.67m	\$5.90m	\$5.82m	\$5.58m
Suggested reserve $r^\star$	1.057	1.045	1.045	1.045
Bounds on $\pi(r^\star)$	[\$3.49m, \$3.93m]	[\$4.10m, \$4.55m]	[\$4.51m, \$4.98m]	[\$4.85m, \$5.29m]
95% CI	[\$2.98m, \$4.47m]	[\$3.49m, \$5.18m]	[\$3.84m, \$5.68m]	[\$4.09m, \$6.07m]
Bounds on $\pi(r^\star) - \pi(r_0)$	[-\$0.07m, \$0.84m]	[-\$0.17m, \$0.79m]	[-\$0.27m, \$0.72m]	[-\$0.30m, \$0.60m]
95% CI for $\pi(r^\star) - \pi(r_0)$	?	?	?	?

Number of Bidders	$N \geq 6$	$N \geq 7$	$N \geq 8$	$N \geq 9$
Avg High Estimate	\$5.81m	\$6.68m	\$7.47m	\$7.18m
Suggested reserve $r^\star$	1.021	1.021	1.009	1.033
Bounds on $\pi(r^\star)$	[\$5.94m, \$6.33m]	[\$7.93m, \$8.42m]	[\$10.0m, \$10.53m]	[\$10.64m, \$11.12m]
95% CI	[\$4.94m, \$7.37m]	[\$6.43m, \$9.98m]	[\$7.88m, \$12.72m]	[\$7.90m, \$13.95m]
Bounds on $\pi(r^\star) - \pi(r_0)$	[-\$0.32m, \$0.50m]	[-\$0.42m, \$0.56m]	[-\$0.48m, \$0.63m]	[-\$0.45m, \$0.57m]
95% CI for $\pi(r^\star) - \pi(r_0)$	?	?	?	?

Table 6: Bounds on expected profit and expected profit increase when selecting a higher reserve for Impressionist and Modern Art sold in New York City. The minimax regret choice of reserve price in the second column is computed by taking the argmax of the lower bound, and is scaled by the median high estimate. Note that the auctioneer's assumed reserve, r_0, is defined as the median of the auctioneer's low estimates for that sample and is $\frac{2}{3}$ of the high estimate for this sample. All monetary figures are in US$.

- The fourth row shows the 95% confidence interval for the expected profit range when selecting the minimax-regret reserve price. The confidence interval is [2.98, 4.47].
- The fifth row shows the bounds on the average expected profit difference when selecting the minimax-regret reserve price. Here, the expected profit difference range is -\$0.07m to \$0.84m.
- The sixth row shows the 95% confidence interval for the expected profit difference range when selecting the minimax-regret reserve price.
- Using our suggested reserve price, the average change in profit in an entire Impressionist and Modern Art auction in New York City, relative to Christie's and Sotheby's typical reserve set at the low estimate, is at worst -\$2.24m and at most +\$26.88m.[7]

The remaining rows of the table show similar results for different lower bounds on the number of bidders. The results suggest that the minimax-regret choice of reserve price is

[7]Here we scale the bounds on expected profit difference by the average number of lots in an Impressionist and Modern art auction in New York Cirty, which is 32.

significantly higher than the auctioneer's reserve price, and that the expected profit difference is positive and large especially for low values of the lower bound on the number of bidders.

As a note on the tightness of our bounds, relative to previous works such as by Haile and Tamer (2003) and Aradillas-López, Gandhi, and Quint (2013) who applied their identification and estimation approach on timber auctions, our bounds on expected profit are relatively wider, resulting in wider optimal reserve price intervals. While this is a limitation of our approach, our approach requires a much weaker set of assumptions relative to Aradillas-López, Gandhi, and Quint (2013) including not knowing the exact number of bidders and not assuming that the number of bidders and bidders' valuations are independent. Such a tradeoff between the degree of assumptions and width of bounds is typical in partial identifcation. These assumptions would have been otherwise difficult to justify in our context of ascending auctions conducted live in an auction room, and where our data only provides a lower bound on the number of bidders. Despite the wider bounds, in certain scenarios such as when there are at least 2 or 3 bidders, our bounds are sufficiently informative to suggest a choice of reserve that almost guarantees an increase in expected profit.

5 Conclusion

Our paper constructs a large novel dataset on live art auctions from the two largest auction houses in the world and provides a close lower bound on the number of bidders. We use the top two bids and the lower bound on the number of bidders to estimate profit bounds non-parametrically, and solve the minimax-regret decision problem applied to this partially-identified model. Using this approach, we find that the auction houses' practice of setting the reserve price at the low estimate is suboptimal in Impressionist and Modern Art auctions in New York City, and propose the minimax-regret choice of reserve price which can significantly increase profit. For the sample of at least 2 bidders, we expect a profit change of between -\$2.24m and +\$26.88m when choosing the minimax-regret reserve price instead of the auctioneer's typical reserve price of the low estimate.

A key innovation in our econometric analysis was freeing the need to know the exact number of bidders, instead only relying on a lower bound on the number of bidders. In our data where we do not have information on the exact number of bidders, this weaker form of partial identification was critical to provide a justified estimate on profit bounds. We also rigorously tightened the bounds on profit difference by imposing the monotonicity condition of cumulative distribution functions.

One possible fruitful area to extend the existing results is to consider the effects of a secret versus public reserve price on the number of bidders and the auctioneer's profit. Typically, Christie's and Sotheby's auctions have secret reserve prices, and it would be interesting to see how the number of bidders and the auctioneer's profit would change if the reserve price was made public.

Tables and Figures

Auction Title	URL
20th21st Century Shanghai to London	link
20th Century Christie's	link
21st Century Evening Sale	link
20th Century Christie's	link
20th Century London to Paris Christie's	link
20th Century Evening Sale	link
20th 21st Century Art Auctions Christies Hong Kong	link
20th21st Century Evening Sale Including Thinking Italian London Christie's	link
21st Century Evening Sale New York	link
The Cox Collection and 20th Century Evening Sale New York	link
20th 21st Century Art Evening Sales Christies Hong Kong	link
The Collection of Thomas and Doris Ammann Evening Sale — Christie's New York	link
21st Century Evening Sale — Christie's New York	link
The Collection of Anne H. Bass and 20th Century Evening Sale — Christie's New York	link
20th / 21st Century Art Evening Sales — Christie's Hong Kong	link
Hubert de Givenchy – Collectionneur: Chefs-d'oeuvre — Christie's Paris	link
20th/21st Century: London to Paris Evening Sales	link
20th/21st Century: London	link
The Ann & Gordon Getty Evening Sale	link

Table 7: Christie's YouTube Data Sources

Auction Title	URL
Hong Kong — Modern, De Beers Blue Diamond & Contemporary Auctions	link
Paris — Surrealism and Its Legacy	link
Hong Kong: Jay Chou x Sotheby's — Evening Sale	link
London — The Now and Modern & Contemporary Evening Auctions	link
New York — Master to Master: The Nelson Shanks Collection	link
New York — Master Paintings and Sculpture Part I	link
New York — Important Watches	link
London — Old Masters Evening Sale	link
New York — PROUVÉ x BASQUIAT: The Collection of Peter M. Brant and Stephanie Seymour	link

New York — Magnificent Jewels	link
London — Treasures	link
Monaco — KARL, Karl Lagerfeld's Estate Part I	link
Edinburgh — The Distillers One of One Whisky Auction	link
Paris — Art Contemporain Evening Sale	link
New York — The Now & Contemporary Evening Auctions With U.S. Constitution Sale	link
New York — Modern Evening Auction	link
New York — The Macklowe Collection	link
Paris — Past/Forward and Modernités	link
Las Vegas: Icons of Excellence & Haute Luxury	link
Las Vegas — Picasso: Masterworks from the MGM Resorts Fine Art Collection	link
New York — Collector, Dealer, Connoisseur: The Vision of Richard L. Feigen	link
London — Richter, Banksy and Twombley lead the Contemporary Art Evening Auction	link
Hong Kong — Modern and Contemporary Art Evening Sales	link
London: British Art + Modern & Contemporary Auctions	link
New York — Important Watches	link
New York — Magnificent Jewels	link
Paris — Important Design: from Noguchi to Lalanne	link
New York — Monet, Warhol and Basquiat Lead Marquee Evening Sales	link
Hong Kong — Contemporary Art Evening Sale	link
Hong Kong — Icons and Beyond Legends: Modern Art Evening Sale	link
Impressionist & Modern Art + Modern Renaissance Auctions	link
Sales of Important Chinese Art and Chinese Art from the Brooklyn Museum	link
The Collection of Hester Diamond Auction in New York	link
Master Paintings & Sculpture Auction in New York	link
London Old Masters Evening Sale	link
Marquee Evening Sales of Contemporary and Impressionist & Modern Art	link
Hong Kong Contemporary Art Evening Sale (LIVE)	link
LIVE from Sotheby's Hong Kong	link
'Rembrandt to Richter' London Evening Sale	link
New York — Now & Contemporary Evening Auctions	link
New York — The David M. Solinger Collection & Modern Evening Auctions	link
Paris — Modernités	link
London — The Now & Contemporary Evening Auctions	link
Paris — Hôtel Lambert, The Illustrious Collection, Volume I: Chefs-d'oeuvre	link
London — Old Masters Evening Auction	link

Hong Kong — Modern, Williamson Pink Star & Contemporary Auctions	link
London — The Jubilee Auction and Modern & Contemporary Evening Auction	link
Paris — Art Contemporain Evening Auction	link
New York — The Now & Contemporary Evening Auctions	link
New York — Modern Evening Auction	link
New York — The Macklowe Collection	link

Table 8: Sotheby's YouTube Data Sources

Saleroom Location	Christie's		Sotheby's	
	Threshold	Rate	Threshold	Rate
Hong Kong	≤HK$7.5M	26.0%	≤HK$7,500,000	26.0%
	>HK$7.5M and ≤ HK$50M	20.0%	>HK$7.5M and ≤HK$40M	20.0%
	>HK$50M	14.5%	>HK$40M	13.9%
London	≤£700k	26.0%	≤£800k	26.0%
	>£700,000 and ≤£4.5M	20.0%	>£800k and ≤£3.8M	20.0%
	>£4.5M	14.5%	>£3.8M	13.9%
Paris	≤€700k	26.0%	≤€800k	26.0%
	>€700,000 and ≤€4M	20.0%	>€800k and ≤€3.5M	20.0%
	>€4M	14.5%	>€3.5M	13.9%
New York	≤$1M	26.0%	≤$1M	26.0%
	>$1M and ≤$6M	20.0%	>$1M and ≤$4.5M	20.0%
	>$6M	14.5%	>$4.5M	13.9%
Shanghai	≤¥6M	26.0%	-	-
	>¥6M and ≤¥40M	20.0%	-	-
	>¥40M	14.5%	-	-

Note: This table is accurate as of February 7 2022 for Christie's and February 1 2023 for Sotheby's. In the last 10 years, there are only minor changes to the base rate (i.e. lowest threshold category). These buyer premium thresholds are additive, so final transaction amounts are strictly increasing. *Source:* Christie's and Sotheby's Websites.

Table 9: Buyer's Premiums in Christie's and Sotheby's Auctions

Appendix A. Proofs

Proof of Theorem 2.

First (Scenario 1), we will show that taking any $\hat{r} \in \underset{r \in \mathcal{C}}{\operatorname{argmax}}\ \pi_U(r)$, for some given $p \in \mathcal{C}$, where $p \neq \hat{r}$, the function $\pi^*(r) : \mathcal{C} \to \mathbb{R}$ with the properties (1)-(3) as stated in Lemma 1 maximizes regret, i.e.

$$\pi^*(r) = \underset{\pi \in \Pi}{\operatorname{argmax}} \left\{ \max_{d \in \mathcal{C}} \{\pi(d)\} - \pi(p) \right\}$$

To see why this is true, suppose that we construct a function π^* which at p is $\pi^*(p) = \pi_L(p)$. Given the choice of p, the best one can do is achieved at some other point than p which gives the highest profit, i.e. $\pi_U(\hat{r})$. From Lemma 1, such a π^* which is $\pi_L(p)$ at p and $\pi_U(\hat{r})$ at $\hat{r}$ is feasible (i.e. the constructed π^* is continuous), and the regret is $\pi_U(\hat{r}) - \pi_L(p)$. If we do not pick $\pi^*(p) = \pi_L(p)$, the other possibilities involve picking some $\pi^*(p)$ such that

$$\pi_L(p) < \pi^*(p) \leq \pi_U(p)$$

In this situation, the best one can do at some point other than p is still $\hat{r}$ with the profit $\pi_U(\hat{r})$. However, any one of these regrets is strictly less than $\pi_U(\hat{r}) - \pi_L(p)$.

It then follows that the max regret for some $p \in \mathcal{C}$ where $p \neq \hat{r}$ is,

$$\max_{\pi \in \Pi} \left\{ \max_{d \in \mathcal{C}} \{\pi(d)\} - \pi(p) \right\} = \pi_U(\hat{r}) - \pi_L(p).$$

Next (Scenario 2), we will consider the scenario of choosing a point $p \in \mathcal{C}$ such that $p = \hat{r} \in \operatorname{argmax}_{\mathcal{C}} \pi_U(r)$. In this scenario, the sup regret at $\hat{r}$ is

$$\sup_{\pi \in \Pi} \left\{ \max_{d \in \mathcal{C}} \{\pi(d)\} - \pi(\hat{r}) \right\} = \pi_U\left(\hat{r}\right) - \pi_L\left(\hat{r}\right).$$

To see this, consider the following 2 cases:

1. **Case 1** is when there is only one max point for both π_U and π_L (i.e. $|\operatorname{argmax}_{r \in \mathcal{C}} \pi_L(r)| = |\operatorname{argmax}_{r \in \mathcal{C}} \pi_U(r)| = 1$) and their argument $\hat{r}$ happens to be the same. First we show that $\forall \epsilon > 0$, $\exists \pi^* \in \Pi$ such that $\text{regret}_{\pi^*}(\hat{r}) \equiv \underset{d \in \mathcal{C}}{\max} \pi^*(d) - \pi^*(\hat{r}) > \pi_U\left(\hat{r}\right) - \pi_L\left(\hat{r}\right) - \epsilon$. Pick any $\epsilon > 0$. By the continuity of $\pi_U(.)$, there exists a $\delta > 0$ such that for some $t \in \mathcal{C}$, $|\hat{r} - t| < \delta \implies \pi_U(\hat{r}) - \pi_U(t) < \epsilon$. Pick such a δ and a corresponding $t \neq \hat{r}$

where $|\hat{r} - t| < \delta$. By Lemma 1, it is possible to construct a $\pi^* : \mathcal{C} \to \mathbb{R}$ such that $\pi^*(\hat{r}) = \pi_L(\hat{r})$ and $\pi^*(t) = \pi_U(t)$. Then the regret at $\hat{r}$ is,

$$\begin{aligned}\text{regret}_{\pi^*}(\hat{r}) &= \pi_U(t) - \pi_L(\hat{r}) \\ &> \pi_U(\hat{r}) - \pi_L(\hat{r}) - \epsilon\end{aligned}$$

At the same time, it is impossible for the regret to be $\geq \pi_U(\hat{r}) - \pi_L(\hat{r})$ since it is impossible to construct a continuous π^* to satisfy this. So the sup regret at $\hat{r}$ is $\pi_U(\hat{r}) - \pi_L(\hat{r})$.

2. **Case 2** is any other situation than Case 1. An alternate way to interpret this is that for any point $r \in \mathcal{C}$, it is always possible to pick another $\hat{r} \neq r$, where $\hat{r} \in \text{argmax}_{r \in \mathcal{C}} \pi_U(r)$. Then, as shown previously, the max regret follows to be $\pi_U(\hat{r}) - \pi_L(\hat{r})$.

For this scenario, see that at any $p \in \mathcal{C}$ where $p \neq \hat{r}$, the max regret is $\pi_U(\hat{r}) - \pi_L(p)$. But $\pi_L(p) \leq \pi_L(\hat{r})$, so

$$\begin{aligned}\max_{\pi \in \Pi}\left\{\max_{d \in \mathcal{C}}\{\pi(d)\} - \pi(p)\right\} &= \pi_U(\hat{r}) - \pi_L(p) \\ &\geq \pi_U(\hat{r}) - \pi_L(\hat{r})\end{aligned}$$

Finally, it is clear that in either scenario, the maximum regret is minimized at some $p \in \mathcal{C}$ where p minimizes $\pi_U(\hat{r}) - \pi_L(p)$. This is precisely at $\underset{r \in \mathcal{C}}{\text{argmax}}\, \pi_L(r)$.

■

References

Akbarpour, M., and S. Li (2020): "Credible Auctions: A Trilemma," *Econometrica*, 88(2), 425–467.

Akbik, A., D. Blythe, and R. Vollgraf (2018): "Contextual String Embeddings for Sequence Labeling," in *COLING 2018, 27th International Conference on Computational Linguistics*, pp. 1638–1649.

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

Aryal, G., and D.-H. Kim (2013): "A Point Decision for Partially Identified Auction Models," *Journal of Business and Economic Statistics*, 31(4), 384–397.

Ashenfelter, O. (1989): "How Auctions Work for Wine and Art," *Journal of Economic Perspectives*, 3(3), 23–36.

Ashenfelter, O., and K. Graddy (2003): "Auctions and the Price of Art," *Journal of Economic Literature*, 41(3), 763–787.

——— (2011): "Sale Rates and Price Movements in Art Auctions," *American Economic Review*, 101(3), 212–16.

Beggs, A., and K. Graddy (2009): "Anchoring Effects: Evidence from Art Auctions," *American Economic Review*, 99(3), 1027–39.

Botev, Z. I., J. F. Grotowski, and D. P. Kroese (2010): "Kernel density estimation via diffusion," *The Annals of Statistics*, 38(5), 2916 – 2957.

Chesher, A., and A. M. Rosen (2017): "Generalized Instrumental Variable Models," *Econometrica*, 85(3), 959–989.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova (2019): "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," .

Freyberger, J., and B. J. Larsen (2022): "Identification in ascending auctions, with an application to digital rights management," *Quantitative Economics*, 13(2), 505–543.

Haile, P. A., and E. Tamer (2003): "Inference with an Incomplete Model of English Auctions," *Journal of Political Economy*, 111(1), 1–51.

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

Hortaçsu, A., and I. Perrigne (2021): "Empirical Perspectives on Auctions," Working Paper 29511, National Bureau of Economic Research.

Imbens, G. W., and C. F. Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72(6), 1845–1857.

Jun, S. J., and J. Pinkse (2019): "An Information-Theoretic Approach to Partially Identified Auction Models," .

Liu, Y., M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov (2019): "RoBERTa: A Robustly Optimized BERT Pretraining Approach," .

Manski, C. F. (2022): "Identification and Statistical Decision Theory," .

Marra, M. (2020): "Sample Spacings for Identification: The Case of English Auctions with Absentee Bidding," Working Papers hal-03878412, Hal.

McAndrew, C., J. L. Smith, and R. Thompson (2012): "The impact of reserve prices on the perceived bias of expert appraisals of fine art," *Journal of Applied Econometrics*, 27(2), 235–252.

Ponomarev, K. (2022): "Essays in Econometrics," *UCLA Electronic Theses and Dissertations*.

Powell, M. J. D. (1964): "An efficient method for finding the minimum of a function of several variables without calculating derivatives," *The Computer Journal*, 7(2), 155–162.

Quint, D. (2008): "Unobserved correlation in private-value ascending auctions," *Economics Letters*, 100(3), 432–434.

Ratcliff, J. W., and D. Metzener (1988): "Pattern Matching: The Gestalt Approach," *Dr. Dobb's Journal*, (46).

Savage, L. J. (1951): "The Theory of Statistical Decision," *Journal of the American Statistical Association*, 46(253), 55–67.

Stoye, J. (2009): "More on Confidence Intervals for Partially Identified Parameters," *Econometrica*, 77(4), 1299–1315.