THE UNIVERSITY OF CHICAGO

RESERVE PRICES IN ART AUCTIONS

A BACHELOR THESIS SUBMITTED TO

THE FACULTY OF THE DEPARTMENT OF ECONOMICS

FOR HONORS WITH THE DEGREE OF

BACHELOR OF THE ARTS IN ECONOMICS

BY BRUCE WEN

CHICAGO, ILLINOIS

MAY 2023

Reserve Prices in Art Auctions*

Bruce Wen†

April 28, 2023

Abstract

A large dataset of complete bids from the two largest auction houses in the world, Sotheby's and Christie's, is constructed via their YouTube livestream videos. Informative bounds on the number of bidders in each auction lot are also generated using audio transcripts from these videos. Using the top two bids in each auction lot and the estimated number of bidders, the profit bounds against reserve prices are estimated nonparametrically. To decide on a single best reserve price, the minimax-regret criterion is solved analytically and applied.

In some subsamples, we propose new reserve prices that significantly increase profits by at least 7.6% to at most 17.7%, equivalent to hundreds of thousands of dollars per individual auction lot, showing that significant profits are lost when setting reserve prices at the low estimate.

*I am beyond grateful to my adviser Kirill Ponomarev for his tireless help and discussions on the data, econometrics and theory for this project. I would like to thank Michael Dinerstein, Ryan Fang, Joseph Hardwick and Alexander Torgovitsky for their advice on the project, as well as Kotaro Yoshida and Victor Lima and fellow thesis writer students for the engaging feedback from the Honors Workshop.

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

1 Introduction

1.1 Motivation

The auction market for art and luxury goods has large economic significance. In 2021, the two largest art auction houses, Sotheby's and Christie's, reported revenues of USD$7.3 billion and USD$7.1 billion respectively. Recently on November 9 and 10, 2022 alone, the 155-work collection of deceased technology billionaire Paul Allen sold for USD$1,622,249,500 towards philanthropy, with the most expensive piece *Georges Seurat, Les Poseuses, Ensemble (Petite version), 1888-1890* selling for USD$149,240,000.

There has been considerable research in high value auctions in the private market. Hortaçsu and Perrigne (2021) review empirical auctions and discuss online eBay and wine auctions. Ashenfelter and Graddy (2003) review the art auction system, discussing in detail the mechanics of an art auction, price movements over time of art, and the specifics of the English Auction mechanism for Christie's and Sotheby's auctions. They find that art experts provide extremely accurate predictions of market prices, and explain the high reserve prices leading to high unsold rates as optimal search in the face of stochastic demand. Ashenfelter (1989) collected data on impressionist and contemporary paintings in Sotheby's and Christie's auctions and suggested that auction houses go to considerable efforts to estimate the price of an item accurately, and discuss that reserve prices are likely to be low in the optimal auction sense because the auctioneer will receive neither a buyer's premium nor a seller's commission if the good is unsold. McAndrew, Smith, and Thompson (2012) study the low and high estimates provided by auction houses and find no evidence that they are biased. Beggs and Graddy (2009) provide evidence that current art prices are influenced by previous prices, an effect known as "anchoring".

However, a pertinent question remains: Are the reserve prices set by these auction houses optimal to maximize revenue? As studied by Ashenfelter and Graddy (2011), 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.

To this end, recently, Marra (2020) developed a new nonparametric identification method to identify IPV ascending auctions with unknown number of bidders using the stochastic difference between adjacent order statistics, and applied her method to determine an optimal reserve price rule using a small sample of wine auctions by Sotheby's. This results in a more optimal common reserve price rule of 110% of the estimate that increases revenue by 2.5%.

There has also been a considerable amount of literature in estimating empirical English auctions. Haile and Tamer (2003) introduce a groundbreaking method to partially identify English auctions using order statistics transformations in the independent private values setting. Quint (2008) writes a new equation to compute expected revenue which allows for estimation in the setting of symmetric and affiliated private values. Aradillas-López, Gandhi, and Quint (2013) further extend this result using an assumption to allow for a wide class of correlated values. Chesher and Rosen (2017) develop sharp bounds on the distribution of valuations in Haile and Tamer (2003)'s incomplete model of English auctions.

In this paper, we also face the problem of an uncertain number of bidders. Several approaches have been formulated to estimate optimal reserves in English auctions under the restriction of not knowing the number of bidders. 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 bounds on the number of bidders are observed.

To the best of my knowledge, this is the first paper which extensively collected complete bid data of large quantities of *live* English auctions run by Sotheby's and Christie's, the two largest auction houses in the world. Such live auctions are also where Christie's and

Sotheby's sell their most high value items; for example, in 2021, Christie's online auctions recorded only \$445m out of a total of their \$7.1b in sales[1]. Excluding \$1.7b in private sales, the remaining \$5b comes from live auction rooms. Using these data, we are able to find numerous subsamples where a change to the reserve price can significantly increase expected profit.

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 dataset of bids from Christie's and Sotheby's auctions via an innovative method. 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 in Marra (2020)), this is the first paper to propose a comprehensive and non-tedious method to collect large amounts of private 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 for certain subgroups of auctions can improve expected profits for these auction houses by the order of millions of dollars per auction.

The methodological contribution builds from Aradillas-López, Gandhi, and Quint (2013)'s and Haile and Tamer (2003)'s papers. We justify Aradillas-López, Gandhi, and Quint (2013)'s choice to select the highest number of bidders when pooling auctions with higher number of bidders under the varying number of bidders assumption.

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,

[1] See Christie's 2021 press release.

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. Section IV analytically solves for the choice of a single optimal reserve price using the minimax-regret criterion. Section V discusses the results. Section VI 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 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 buyer's premium. Information such as bids and bidders in past auctions are kept private to the firms.

I adopt a novel data collection process using livestream auction videos. While not encompassing all auctions, Sotheby's and Christie's have been posting YouTube livestream videos of their largest live auctions for the past few years, perhaps to increase publicity on their auctions and attract more bidders. As of February 2023, one can find 65 Sotheby's past livestream auctions on Sotheby's YouTube channel and 39 Christie's past auctions on Christie's YouTube channel. The total number of objects sold in these videos sum to the order of thousands.

These videos range in length, from short hour-long auctions to marathon 6-hour auctions. They present a huge trove of unmined data on bids. Typically, a recorded livestream 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, consider the Sotheby's *New*

York — Monet, Warhol and Basquiat Lead Marquee Evening Sales auction as in Table 6. The first section of the livestream, from about $0hr : 38min$ to $1hr : 25mins$ in the video, corresponds to 18 contemporary art pieces from the collection of Mrs John L Marion, the wife of an auctioneer and philanthropist. The second section of the livestream, from about $1hr : 38mins$ to $3hr : 27mins$, corresponds to 32 pieces of contemporary art. The third and last section of the livestream, from about $3hr : 49mins$ to $5hr : 12mins$, corresponds to 33 art pieces from the Impressionist and Modern era. These three auctions are held in immediate succession in the same Sotheby's New York auction room, each consisting of art from the modern period, and in total make up 73 auction lots in a single video alone.

To extract the data from these videos, I use computer vision techniques to read bids. Complete bidding trajectory is generated with the possible exception of a few starting bids[2] for each auction lot. These bids are paired with scraped data from the auction houses' websites to match the auction lot and its characteristics such as the estimate and lot description, ensuring accuracy.

The data generation process for bid data from the YouTube videos is described as follows:

1. **Obtain Video data.** From the URLs of the videos download the full videos into a hard drive. Filter the video into the parts that are actually the auction.[3]

2. **Convert Videos to Images.** Split the videos into image frames. The heuristic is an image every half a second. This frequency is sufficient to capture all bids.

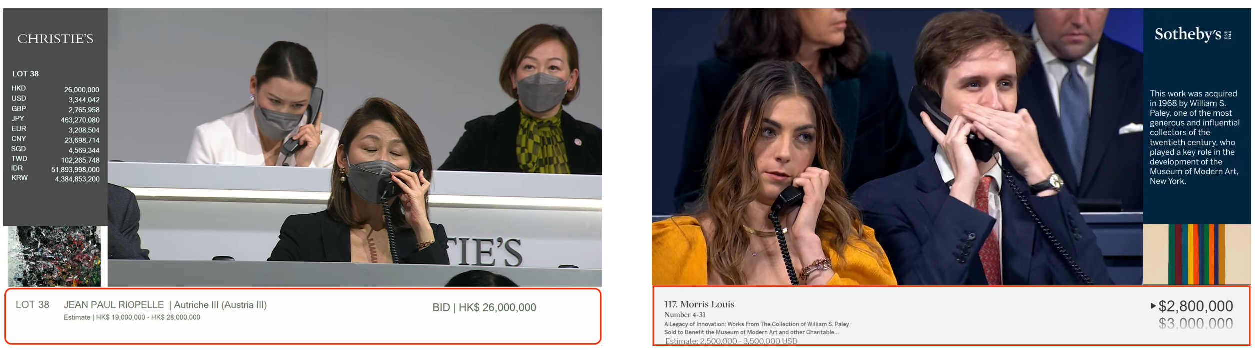
3. **Crop and Filter.** Based on the video, crop the images so that only the section of the images that contain bid data remain (see Figure 1). Since these bid data images are usually black and white in color, a majority color detector script is used to filter for images that are black and white and thus contain bid data. The remaining are deleted.

[2]This is due to a slight lag in the video livestream 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]Usually, the videos are live streams and therefore a large portion of the front of the video is just a blank filler page waiting for the auction to start.

4. **OCR.** Using Tesseract Optical Character Recognition (OCR), maintained (Smith (2007)) by Google [4], read the bid data from images into text. Produce the dataset which contains all bids and times of bids for each auction lot.

To speed up the above process, the computations are done in parallel with appropriate mutexes on file writing. 93.64% of the total time is spent on cropping the image, with about 75.45% on reading of the original image, and 18.18% saving the cropped image. It is noted that extraction of text using the Tesseract command line only takes less then 1% of the total time. The runtime for the entire process was about 10 days, on a 9^{th}-generation 6-Core Intel Core i7 Processor with 16GB RAM.



Lot 38, Christie's Hong Kong, Nov 2022 Lot 117, Sotheby's New York, Nov 2022

Figure 1: 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 bid is clearly seen to be $\$26,000,000$HKD and $\$2,800,000$USD respectively.

2.2 Bounds on Number of Bidders

Unfortunately, the bids collected from the livestream 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.

[4]The latest Tesseract can be found on GitHub at this link.

However, we can attempt to produce informative bounds on the true number of bidders, denoted by the random variable $N \in \{2, 3, ...\}$. The first step involves disregarding the bidders that have valuations much below the seller's low estimate and thus do not submit any bids during the auction.

Assumption 1 *The bidders that do not submit bids in the auction are irrelevant to the results of the auction.*

Using this assumption, it is possible to obtain bounds on the true number of bidders, N:

$$\underline{N} \leq N \leq \bar{N} \equiv \text{total number of bids} \tag{1}$$

The upper bound on N can be characterized by the total number of bids, because of the obvious fact that given k bids, the bids are at most submitted by k different bidders. Without any additional information, it is impossible to improve this upper bound. Unfortunately, from the available data in auction websites and videos, beyond tedious manual work checking for repeat bids by same bidders in a video, it is difficult to automate the tightening of this upper bound with good accuracy.

An informative lower bound that is very close to the true number of bidders, $\underline{N} \geq 2$, can be constructed using the fact that the auctioneer often calls out bidders using specific terminology. To see how such a close lower bound on the number of bids can be obtained, consider the fact that in a typical Christie's or Sotheby's auction, there are four sources of bids:

1. **Live bid from the room**. Live bids are bids placed in-person by the buyer herself within the auction room on-site.

2. **Absentee bid**. Absentee bids are pre-placed bids, at least 24 hours before the auction, that the auctioneer will prioritize over any other bids placed during the live auction itself.

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

Figure 2: 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"*.

3. **Telephone bid**. Telephone bids are live bids placed remotely through an auctioneer attending in person. They are extremely common. See Figure 2 for an example of a telephone bid.

4. **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.

Auctioneers typically call out specific names of colleagues when they submit bids from some unknown bidder through the telephone. Obtaining the lower bound estimate thus relies largely on the audio data from these telephone bids. The uniqueness of the sources of telephone bids can be at least partially captured and at most fully captured through the audio of each auction video. Telephone bids commonly involve an auctioneer calling out a fellow employee working in the same firm. 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. As an example, in figure 2, the auctioneer references Olivier Camu, another employee at Christie's who is putting in telephone bids on the behalf of an unknown buyer. This knowledge thus allows to 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.

Beyond the telephone 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 [5], it is impossible to distinguish between the different bidders. Thus we are only able to capture the identity of at most one online bidder and one absentee bidder.

Finally, it is possible to capture some 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. These are necessarily distinct from an online bidder. It is also necessarily distinct from the telephone bidders, because the auctioneer knows the names of his colleagues.

In summary, we are able to capture as many unique telephone bidders, one unique online bidder, one unique absentee bidder, and some unique in-person bidders. It is however unclear what proportion of the total number of bidders is captured, i.e. 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[6]. 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 less than the true number of bidders.

We now discuss the methodology to capture the bidder identities from video audio. The

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

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

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 3 shows an example transcript (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 3: 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 3, from a section of a livestream video transcribed via Google Cloud. 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”. This motivates the next step, which is to identify the unique bidders from the transcript.

One option to identify names from a piece of text is to reference a large database of all names and do string matching with the obtained transcript. However, this method does not apply well in this context because of the large variety of names that the databases will not sufficiently capture, or that selecting a large database of names would overwhelmingly capture names that are not actually people’s names, e.g. single letter names like ”a”.

Instead, 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. 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 they name as ”contextual string 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. We also include a list of manual matching strings, {”online bidder”, ”online”, ”telephone”, ”sir”, ”madam”, ”gentleman”, ”lady”, ”phone”, ”back of the room”, ”to the right”, ”to the left”, ”starting with”, ”absentee”} to capture the online, absentee, and live bidders.

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. Another example is that the names ”Kelvin” and ”Calvin” (from the transcript in figure 3) 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,

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

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. The algorithm is $O(n^3)$ time. 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: Transcript to $\underline{N}$, lower bound on number of bidders

---

Inputs: $X \equiv$ [transcript segments from 1 auction lot], manualMatches
Output: Lower bound on number of bidders, $\underline{N}$
joinedTexts $\leftarrow$ join(X) ; `/* Join strings into single long chunk */`
for $i = 1$ *to len(X)* **do**
  namesR $\leftarrow$ RoBERTa on $X[i]$;
  namesC $\leftarrow$ ContextualStringEmbeddings on $X[i]$;
  manualNames $\leftarrow$ [x in $X[i]$ if $x \in$ manualMatches]
end
namesJoinedR $\leftarrow$ RoBERTa on joinedTexts;
namesJoinedC $\leftarrow$ ContextualStringEmbeddings on joinedTexts;
names $\leftarrow$ Ratcliff-Obershelp algorithm on (namesR, namesJoinedR, namesC, namesJoinedC, manualNames) ; `/* Remove similar name duplicates */`
return length(names)

The distribution of number of bidders generated from this algorithm is shown in Figure 4.

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 7 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 1

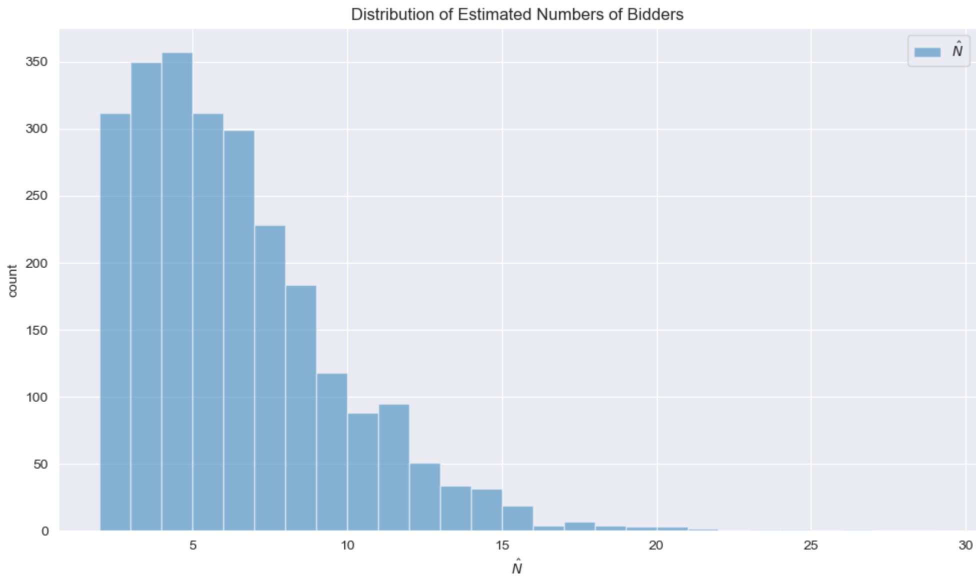


Figure 4: Distribution of Estimated Number of Bidders across all auctions.

reports the counts of items in each category.

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 1: Table of Category Counts.

After data cleaning, we produce a combined dataset consisting of 2505 auction lots. The summary statistics are shown in Table 2 and Table 3. The distribution of transaction prices is shown in Figure 5.

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 2: 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 3: Summary Statistics for Sotheby's Auctions

Next we construct smaller sub-sample groups to take into account how other factors 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.

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 1). 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.

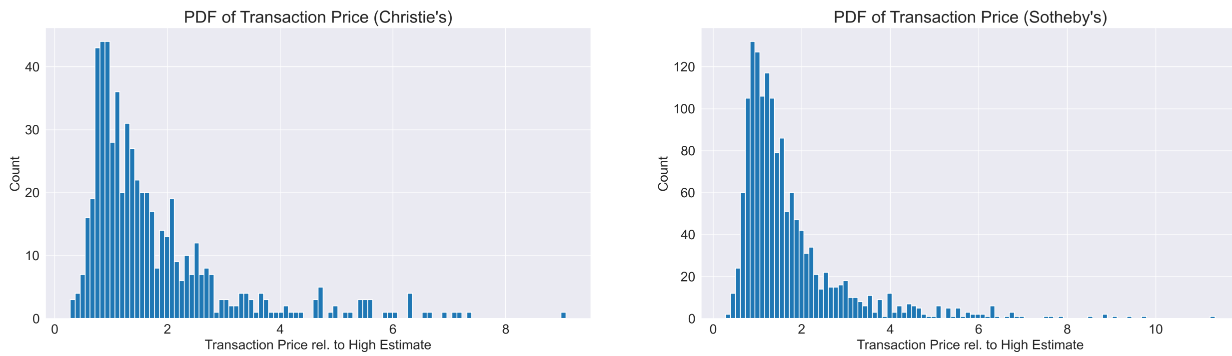


Figure 5: Distribution of Final Bids, with Buyer's Premium included. Prices ≥ 3 standard deviations from the mean are removed for graph visibility.

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 2, 3, 4 are due to Haile and Tamer (2003) and Aradillas-López, Gandhi, and Quint (2013) and are maintained throughout the paper. Assumption 5 is a statement on the measurement error from the data.

Assumption 2 *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 2a 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, 2b 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 3 *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 4 *For each n, the joint distribution $\boldsymbol{F}^n$ is such that for any v and i, the probability $\mathbb{P}(V_i < v | N = n, ||j \neq i : V_j < v|| = k)$ is nondecreasing in k.*

This assumption is due to Aradillas-López, Gandhi, and Quint (2013). In our application to art auctions, it is likely that there is some form of correlation between bidders' valuations due to the nature of the English auction and the ability to resell. This assumption 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.*

Assumption 5 *The estimated lower bound on the true number of bidders $\underline{N}$ is equivalent to the true number of bidders n.*

We justified in Algorithm 1 why our lower bound is a very close estimate to the true number of bidders. While there might be some measurement error here, we have done our best to reduce these errors in the data collection process. From now on, we will take $N = \underline{N}$ as given.

3.2 Identification

Identification builds on the approaches proposed by Haile and Tamer (2003) and Aradillas-López, Gandhi, and Quint (2013). 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 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

$$\pi(r) = (r - v_0) \cdot \mathbb{1}(V_{N-1:N} \leq r, V_{N:N} > r) + (V_{N-1:N} - v_0) \cdot \mathbb{1}(V_{N-1:N} \geq r) \tag{3}$$

Taking expectations conditional on N and rearranging,

$$\mathbb{E}[\pi(r)|N] = \int_0^{+\infty} \max\{r, v\} dF_{N-1:N}(v) - v_0 - F_{N:N}(r)(r - v_0) \tag{4}$$

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 2. By assumption 2a, 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 2b, 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,

$$G_{N:N}(v) \leq F_{N-1:N}(v) \leq G_{N-1:N}(v) \tag{5}$$

Identification of $F_{N:N}$ follows from assumptions 3 and 4. Define the strictly increasing differentiable function $\phi_{i:N}(H) : [0,1] \to [0,1]$ as the implicit solution to

$$H = \frac{N!}{(N-i)!(i-1)!} \int_0^{\phi} s^{i-1}(1-s)^{N-i} ds \tag{6}$$

Then we have,

$$\phi_{N-1:N}(G_{N:N}(v))^N \leq F_{N:N}(v) \leq G_{N:N}(v) \tag{7}$$

Ponomarev (2022) showed that the bounds for $F_{N-1:N}(v)$ (eq. 5) and $F_{N:N}(v)$ (eq. 7) are sharp. Thus, the sharp bounds for the profit function conditional on N are:

$$\mathbb{E}[\pi(r)|N] \geq \int_0^{\infty} \max\{r, v\} dG_{N-1:N}(v) - v_0 - G_{N:N}(r)(r - v_0) \tag{8}$$

$$\mathbb{E}[\pi(r)|N] \leq \int_0^{\infty} \max\{r, v\} dG_{N:N}(v) - v_0 - \phi_{N-1:N}(G_{N:N}(r))^N (r - v_0) \tag{9}$$

Unfortunately, the lower bound can be shown to be decreasing most of the time at v_0 by taking the derivative with respect to r and noticing that it is only when there are numerous observations of objects sold below the auctioneer's valuation that it is positive. In order to generate more informative bounds, we require the following additional assumption.

Assumption 6 (Valuations are independent of N) *Let $\boldsymbol{F}_m^N$ be the joint distribution of m randomly chosen bidders in an N-bidder auction. Then, $\boldsymbol{F}_m^N = \boldsymbol{F}_m^{N'}$ for any $m \leq N, N'$.*

This assumption allows for the relationship $F_{N:N} = F_{N:N}^{N+1} = \frac{1}{N+1} F_{N:N+1}(v) + \frac{N}{N+1} F_{N+1:N+1}(v)$ to hold [7]. Then, successive application of this relationship results in the following lemma.

Lemma 1 *Fix N and $\bar{N} > N$. If valuations are independent of N, then for any v,*

$$F_{N:N}(v) \leq \bar{F}_{N:N}(v) \equiv \sum_{m=N+1}^{\bar{N}} \frac{N}{(m-1)m} F_{m-1:m}(v) + \frac{N}{\bar{N}} F_{\bar{N}-1:\bar{N}}(v) \tag{10}$$

$$F_{N:N}(v) \geq \underline{F}_{N:N}(v) \equiv \sum_{m=N+1}^{\bar{N}} \frac{N}{(m-1)m} F_{m-1:m}(v) + \frac{N}{\bar{N}} \left(\phi_{\bar{N}}\left(F_{\bar{N}-1:\bar{N}}(v)\right)\right)^{\bar{N}} \tag{11}$$

[7] See Aradillas-López, Gandhi, and Quint (2013) equation 4 for details.

Proof. See Lemmas 3 and 4 in Aradillas-López, Gandhi, and Quint (2013). ∎

Lemma 1 provides a set of inequalities that characterize the two-sided bounds for $F_{N:N}$. For example, if $N = 3$ and there are also $N = 4, 5, ..., 20$ in the data sample, there will be 18 inequalities on each side to bound $F_{N:N}$, since $\bar{N}$ can be any of $S \equiv \{3, 4, ..., 20\}$. Thankfully, instead of having to intersect these bounds, it is possible to show that the highest $\bar{N}$ is the best to pick.

Theorem 1 (Tightest Inequalities for $F_{N:N}$) *Let $S = \{N+1, N+2, ..., \bar{N}\}$ be the set of all possible upper limits of summation. Let*

$$\bar{F}_{N:N,s}(v) \equiv \sum_{m=N+1}^{s} \frac{N}{(m-1)m} F_{m-1:m}(v) + \frac{N}{s} F_{s-1:s}(v)$$
$$\underline{F}_{N:N,s}(v) \equiv \sum_{m=N+1}^{s} \frac{N}{(m-1)m} F_{m-1:m}(v) + \frac{N}{s} \left(\phi_s \left(F_{s-1:s}(v)\right)\right)^s$$

Then, the upper bound $\bar{F}_{N:N,\bar{N}}(v) \leq \bar{F}_{N:N,s}(v)$ for all $s \in S$. Furthermore, if bidders' valuations are i.i.d. or common, the lower bound $\underline{F}_{N:N,\bar{N}}(v) \geq \underline{F}_{N:N,s}(v)$ for all $s \in S$.

Proof. See Appendix A. ∎

Theorem 1 tells us that picking $\bar{N} = \max(S)$ guarantees the tightest upper bound in the set of inequalities specified by 10, regardless of the dependence among bidders' valuations. For the lower bound, even though we are only able to formally prove that picking $\bar{N} = \max(S)$ guarantees the tightest lower bound for the case of *i.i.d.* or common values, simulations show that it is likely true that this can be generalized to any dependence among bidders' valuations. The following conjecture captures this.

Conjecture 1 *Suppose that assumptions 4 and 6 hold, and no additional assumptions on the bidders' valuations. Then, the lower bound $\underline{F}_{N:N,\bar{N}}(v) \geq \underline{F}_{N:N,s}(v)$ for all $s \in S$.*

To see why Conjecture 1 may hold, consider the following simulation setup. Bidder valuations are drawn from $\log V_i \sim Normal(2.5, 0.5)$. A covariance matrix $\Sigma_{N\times N}$ is created where,

$$\Sigma_{ij} = \begin{cases} \sigma^2 & \text{i=j} \\ \sigma^2\rho & \text{i} \neq j \end{cases}$$

with $\sigma^2 = 0.5$ and ρ the degree of correlation among bidder valuations. $V_1, V_2, ..., V_N$ are drawn from this multinomial lognormal distribution with covariance matrix $\Sigma_{N\times N}$. For the simulations, set $N = [2, 3, ..., 10]$, and compute equation 11 for the case of $N = 3$ with all $\bar{N} \in \{3, 4, ..., 10\}$. $10,000$ simulations are conducted. Results are shown in figure 6.

The plots in fig. 6a and 6e represent the cases of i.i.d. and common values respectively and help visually validate that theorem 1 holds. For all degrees of correlation in between (i.e. figures 6b, 6c, 6d), it is visible that the lower bound for $F_{3:3}$ weakly increases in $\bar{N}$ across the entire support. This supports the conjecture.

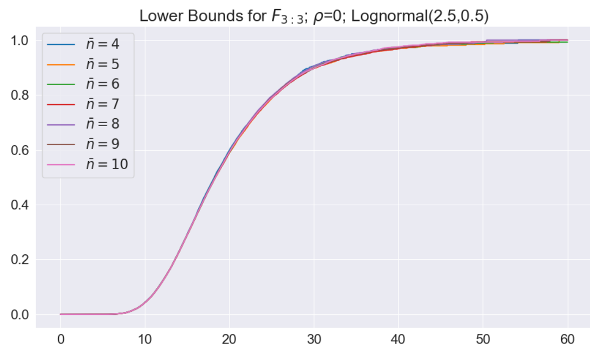
3.3 Estimation

Estimation on the profit bounds was done using a Kernel Density Estimator, with the probability densities of the top two order statistics being

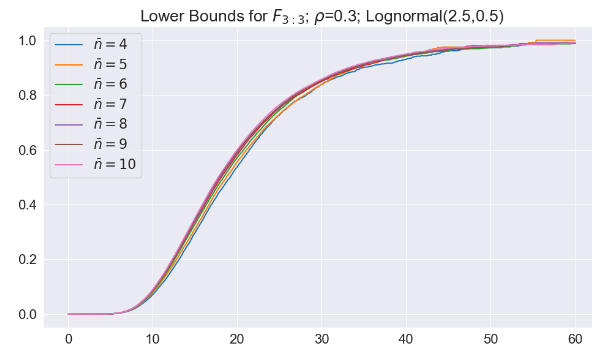
$$\hat{g}_{N-1:N}(v) = \frac{1}{Th}\sum_{i=1}^{T}\mathcal{K}\left(\frac{v - c\left(b^i_{N-1:N}\right)}{h}\right) \tag{12}$$

$$\hat{g}_{N:N}(v) = \frac{1}{Th}\sum_{i=1}^{T}\mathcal{K}\left(\frac{v - c\left(b^i_{N:N}\right)}{h}\right) \tag{13}$$

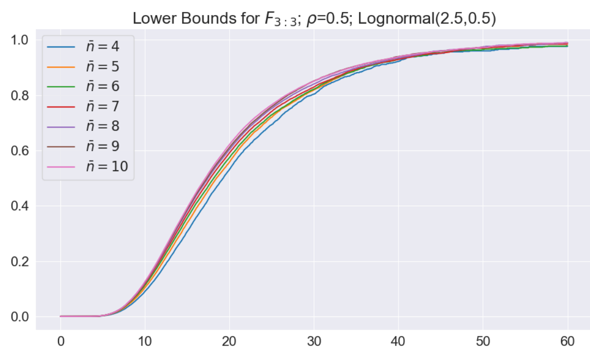
where T is the total number of auctions in the sample, h is the smoothing parameter for a Gaussian kernel $\mathcal{K}$, and $b^i_{n-1:n}$, $b^i_{n:n}$ are the second highest bid and the highest bid in the i^{th} auction respectively. It is reasonable here to assume that these two bids are by different bidders, since no bidder would rationally attempt to out-bid herself. $c : \mathbb{R}_+ \to \mathbb{R}_+$ is a scaling function of bids, as defined in Table 7, to account for the fact that buyers need to pay the auction house buyer's premium for the lot if a bid wins the auction. Furthermore,



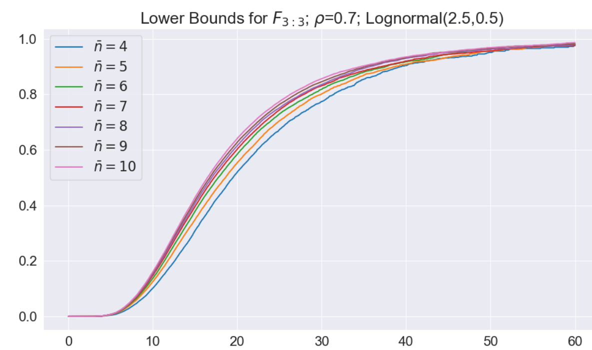
(a) $\rho = 0.0$



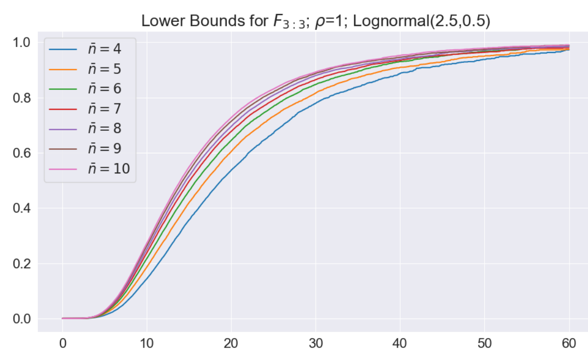
(b) $\rho = 0.3$



(c) $\rho = 0.5$



(d) $\rho = 0.7$



(e) $\rho = 1.0$

Figure 6: Simulations on lower bounds for $F_{3:3}$ in equation 11 for different degrees of correlation among bidder valuations. (a) i.i.d. values, (b) $\rho = 0.3$, (c) $\rho = 0.5$, (d) $\rho = 0.7$, and (e) common values.

set the CDF, $\hat{G}_{N:N}(r) = \int_0^r \hat{g}_{N:N}(v)dv$. Then, the pointwise estimated bounds are,

$$\hat{\pi}_N(r) \geq \int_0^\infty \max\{r, v\}\hat{g}_{N-1:N}(v)dv - v_0 - \hat{G}_{N:N}(r)(r - v_0) \tag{14}$$

$$\hat{\pi}_N(r) \leq \int_0^\infty \max\{r, v\}\hat{g}_{N:N}(v)dv - v_0 - \phi_{N-1:N}\left(\hat{G}_{N:N}(r)\right)^N (r - v_0) \tag{15}$$

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. Speedup is conducted using the Fast Fourier Transform. ϕ is computed using the Powell (1964) conjugate direction numerical optimization method which does not require derivatives.

95% pointwise confidence intervals (CI) are constructed analytically and applied to the kernel density estimates. Let $\underline{\hat{\pi}}_N(r)$ and $\hat{\bar{\pi}}_N(r)$ be the estimated lower and upper bounds on profits respectively. Let $\underline{\hat{\sigma}}(r)$ and $\hat{\bar{\sigma}}(r)$ be the standard deviation of the lower and upper bounds on profit computed using the Delta method. For the case of not pooling together higher number of bidders as in equations 14 and 15, the Delta method involves the two random variables $\max\{r, B^t_{N-1:N}\}$ and $\mathbb{1}(B^t_{N:N} \leq r)$ for the lower bound on profit and the two random variables $\max\{r, B^t_{N:N}\}$ and $\mathbb{1}(B^t_{N:N} \leq r)$ for the upper bound on profit. When pooling together higher number of bidders using Theorem 1, the Delta method involves additionally creating two indicator variables $\mathbb{1}(N = m)$ and $\mathbb{1}(B^t_{N-1:N} \leq r) \times \mathbb{1}(N = m)$ for each realization of number of bidders above N. The correct expressions can then be fully calculated by taking the respective partial derivatives.

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

$$\text{CI}_{1-\alpha}(\pi_N(r)) = \left[\underline{\hat{\pi}}_N(r) - c_\alpha \cdot \frac{\underline{\hat{\sigma}}(r)}{\sqrt{T_N}}, \hat{\bar{\pi}}_N(r) + c_\alpha \cdot \frac{\hat{\bar{\sigma}}(r)}{\sqrt{T_N}}\right] \tag{16}$$

where c_α solves

$$\Phi\left(c_\alpha + \frac{\sqrt{T_N}\left(\hat{\bar{\pi}}_N(r) - \hat{\underline{\pi}}_N(r)\right)}{\max\{\hat{\underline{\sigma}}(r), \hat{\bar{\sigma}}(r)\}}\right) - \Phi(-c_\alpha) = 1 - \alpha \tag{17}$$

where Φ is the standard normal cumulative distribution function, T_N is the number of auctions in the sample, and α is the significance level.

4 Selecting a Single Optimal Reserve Price

While the approach so far allows us to generate relatively informative bounds, it remains that an auction house must make a decision on a single reserve price to choose 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 has 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 is suggested by Jun and Pinkse (2019), which uses information about the upper bound on profit 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:

- $C = [v_0, \bar{v}]$ be the choice set of reserve prices.
- $\Pi = \{\pi : \pi_L \leq \pi \leq \pi_U\}$ be the space of possible profit functions $\pi(r) : C \to \mathbb{R}$, bounded by $\pi_L : C \to \mathbb{R}$ and $\pi_U : C \to \mathbb{R}$.

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

Assumption 7 (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 in literature.

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:

$$\min_{r \in C} \max_{\pi \in \Pi} \left\{ \max_{d \in C} \{\pi(d)\} - \pi(r) \right\} \tag{18}$$

where d represents a decision within the choice set C. The term $\max_{d \in C} \{\pi(d)\} - \pi(r)$ 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 R^* that solves equation 18,

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

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 7.

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

Lemma 2 (Feasibility of π^*) *Fix some $p, q \in C$, where $p \neq q$. $\exists$ a continuous function $\pi^*(r) : C \to \mathbb{R}$ such that,*

1. $\pi_L \leq \pi^*(r) \leq \pi_U \;\forall r \in C$,
2. $\pi^*(p) = \pi_L(p)$, *and,*
3. $\pi^*(q) = \pi_U(q)$.

Proof. Some straightforward continuity arguments. See Appendix A. ∎

The construction in Lemma 2 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 $\max \pi_U$.

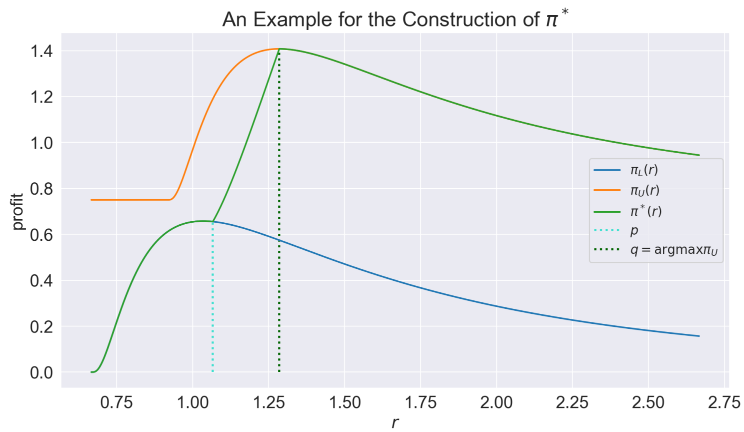


Figure 7: An example of a feasible π^* in Lemma 2

It is not difficult to generalize this idea for all $p, q \in C$ 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 $C = [v_0, \bar{v}]$ and assume that π_U and π_L are continuous on C. Let $\Pi = \{\pi : \pi_L \leq \pi \leq \pi_U\}$ be the space of possible profit functions $\pi(r) : C \to \mathbb{R}$. Then the solution to the minimax-regret problem is $\underset{r \in C}{\operatorname{argmax}}\, \pi_L(r)$, i.e.*

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

Proof. See Appendix A. ■

5 Results

The collected data spans many categories of auctions, but we will focus on a few subsamples where clear reserve prices can be recommended to increase expected profits. In particular, these are the samples (1) all art, (2) impressionist/20th/21st century art, (3) impressionist/20th/21st century art in New York City, (4) old master art, and (5) luxury goods (jewelry and watches). We also give an example of a fully binned sample conditioned on all covariates in ξ, (6) impressionist/20th/21st century art in New York City in the middle one third percentile of high estimates with middle one third percentile level of competition as measured by the number of bids.

We estimate the bounds on seller profit, $\pi_N(r)$, for various values of N. When varying the number of bidders, we set $\bar{N}$ to be the highest number of bidders observed in that subsample such that the number of auctions is greater than or equal to ten. This heuristic allows for numerical optimization methods to converge better. We also assume that v_0, the seller's valuation, is equal to the low estimate provided by the auction house.

Plotted bounds on profits are shown in Figures 8, 9, 10, 11, 12, and 13. We display graphs with two values on the number of bidders for visibility, leaving the rest of the graphs in the Appendix.

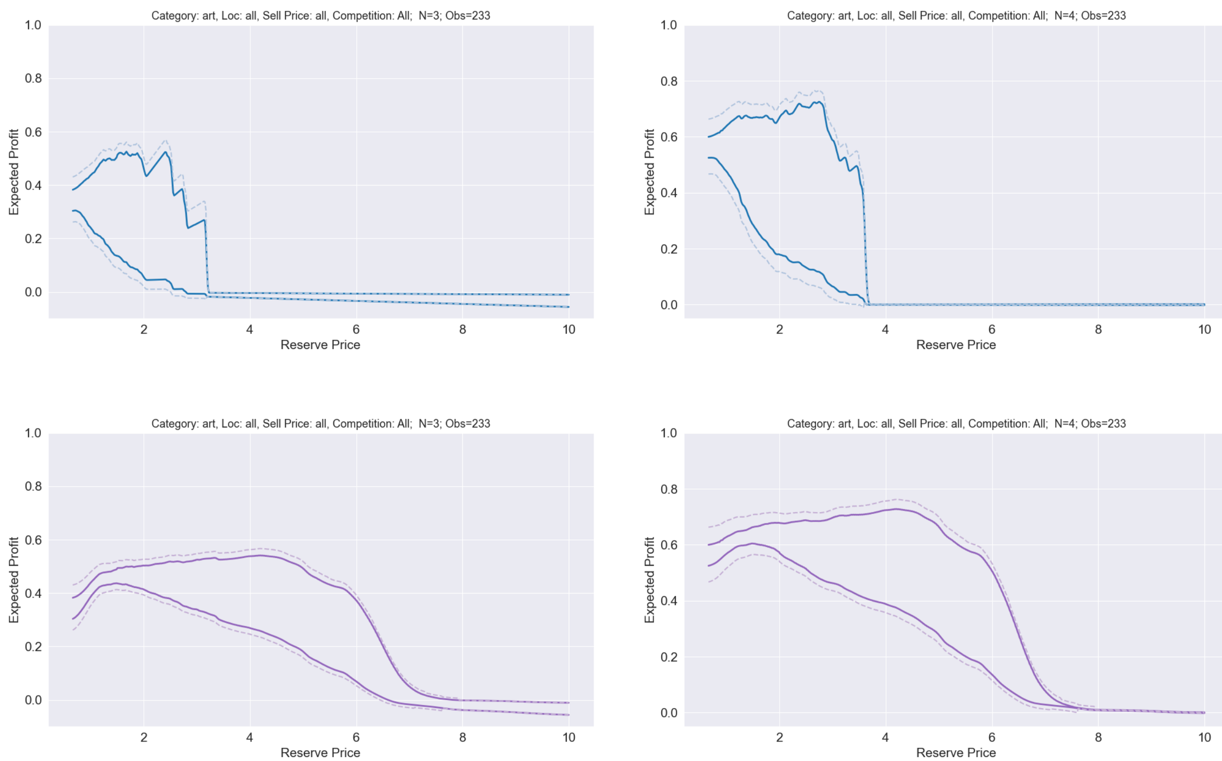
Generally, as the number of estimated bidders increase, the profit increases across reserve prices. When varying the number of bidders, the lower bound shows a significant hump that proves to be informative. This is especially the case when the estimated number of bidders is low, for example at three or four, where the lower bound increases at some value of r to above the upper bound at v_0, thus informing that there exists a better reserve price than the low estimate as set by auction houses.

We can further quantify the policy significance of departure from the reserve price set at the low estimate. Given the bounds on profit, the interval of the optimal reserve price can be defined as,

$$r_L \equiv \sup\{r < \arg\max \pi_L(r) : \pi_U(r) \leq \max \pi_L(r)\} \tag{19}$$

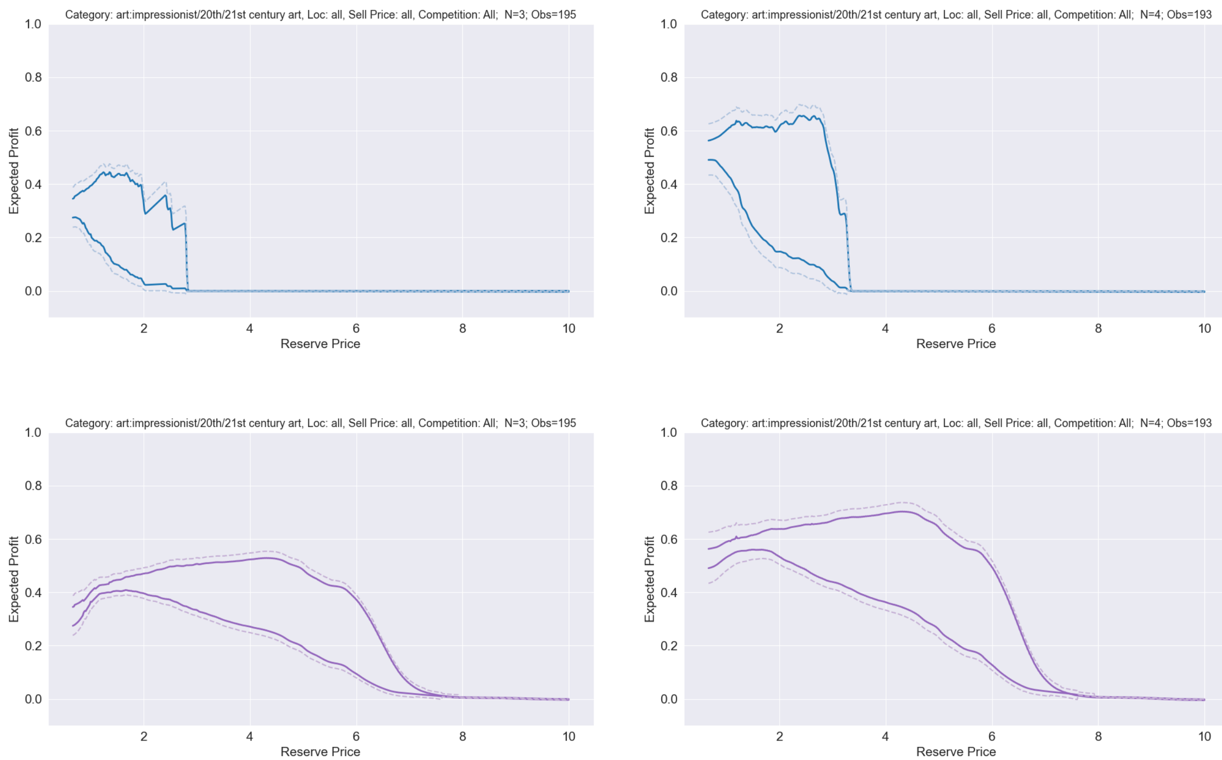
$$r_U \equiv \inf\{r > \arg\max \pi_L(r) : \pi_U(r) \leq \max \pi_L(r)\} \tag{20}$$

Table 5 shows the optimal reserve bound implied for each model for each of the 5 samples



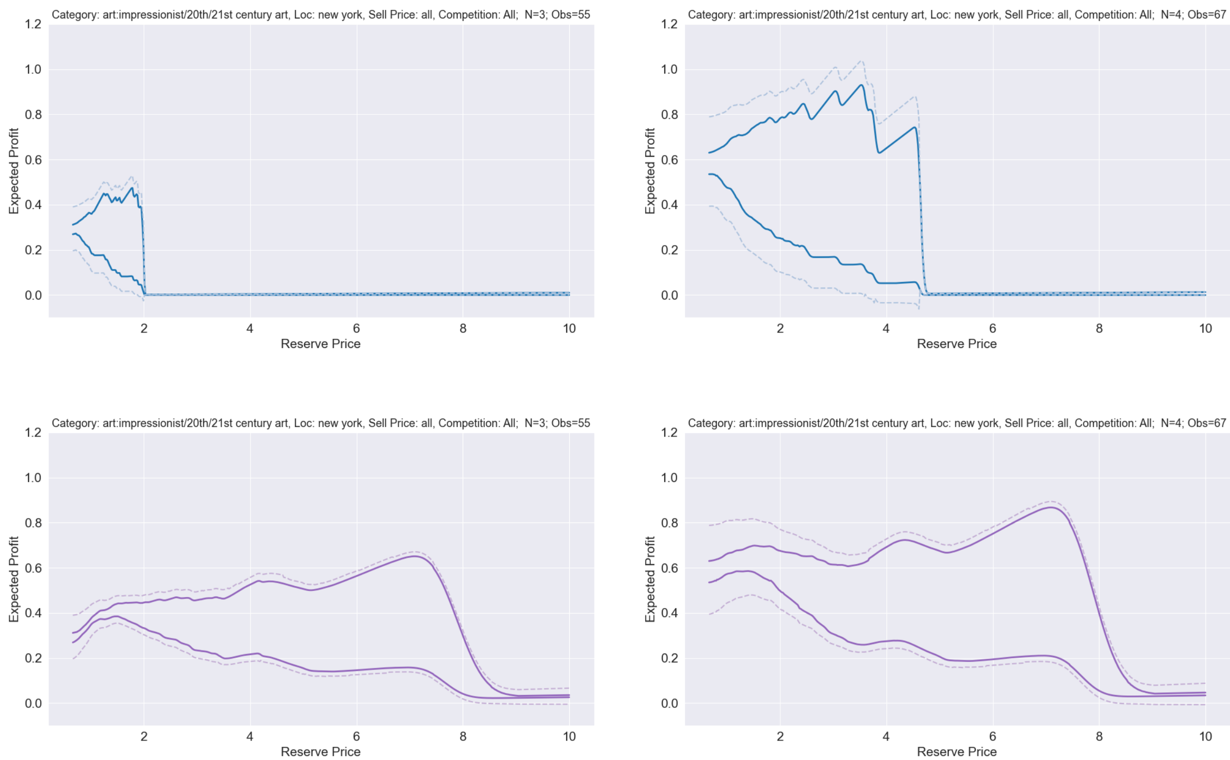
$N = 3$ $N = 4$

Figure 8: Bounds on the expected profit against reserve price, for all 1712 auction lots in the Art category. N ranges from 2 to 15, but we display $N = 3, 4$ only here for brevity. The top two graphs in blue are from equations 14 and 15, while the bottom two graphs in purple incorporate higher numbers of bidders using the assumption that valuations are independent of N. Light blue and light purple lines are the 95% confidence intervals for the respective bounds.



$N = 3$ $N = 4$

Figure 9: Bounds on the expected profit against reserve price, for all 1477 auction lots belonging to the Impressionist/20th/21st century art category. N ranges from 2 to 15, but we display $N = 3, 4$ only here for brevity. The top two graphs in blue are from equations 14 and 15, while the bottom two graphs in purple incorporate higher numbers of bidders using the assumption that valuations are independent of N. Light blue and light purple lines are the 95% confidence intervals for the respective bounds. Both reserve price and expected profits are scaled by the median of high estimates of the lots in the sample.

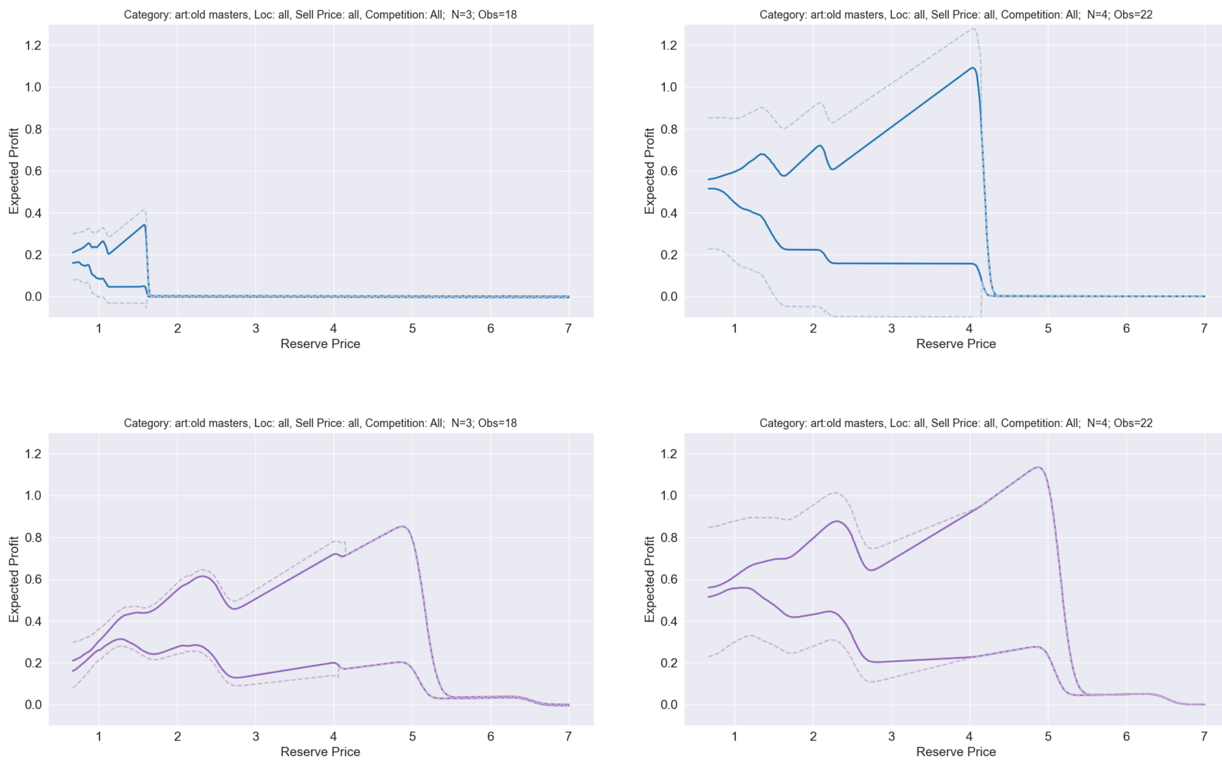


$N = 3$ $\qquad\qquad$ $N = 4$

Figure 10: Bounds on the expected profit against reserve price, for all 497 auction lots belonging to the Impressionist/20th/21st century art category, sold in New York City. N ranges from 2 to 12, but we display $N = 3, 4$ only here for brevity. The top two graphs in blue are from equations 14 and 15, while the bottom two graphs in purple incorporate higher numbers of bidders using the assumption that valuations are independent of N. Light blue and light purple lines are the 95% confidence intervals for the respective bounds. Both reserve price and expected profits are scaled by the median of high estimates of the lots in the sample.

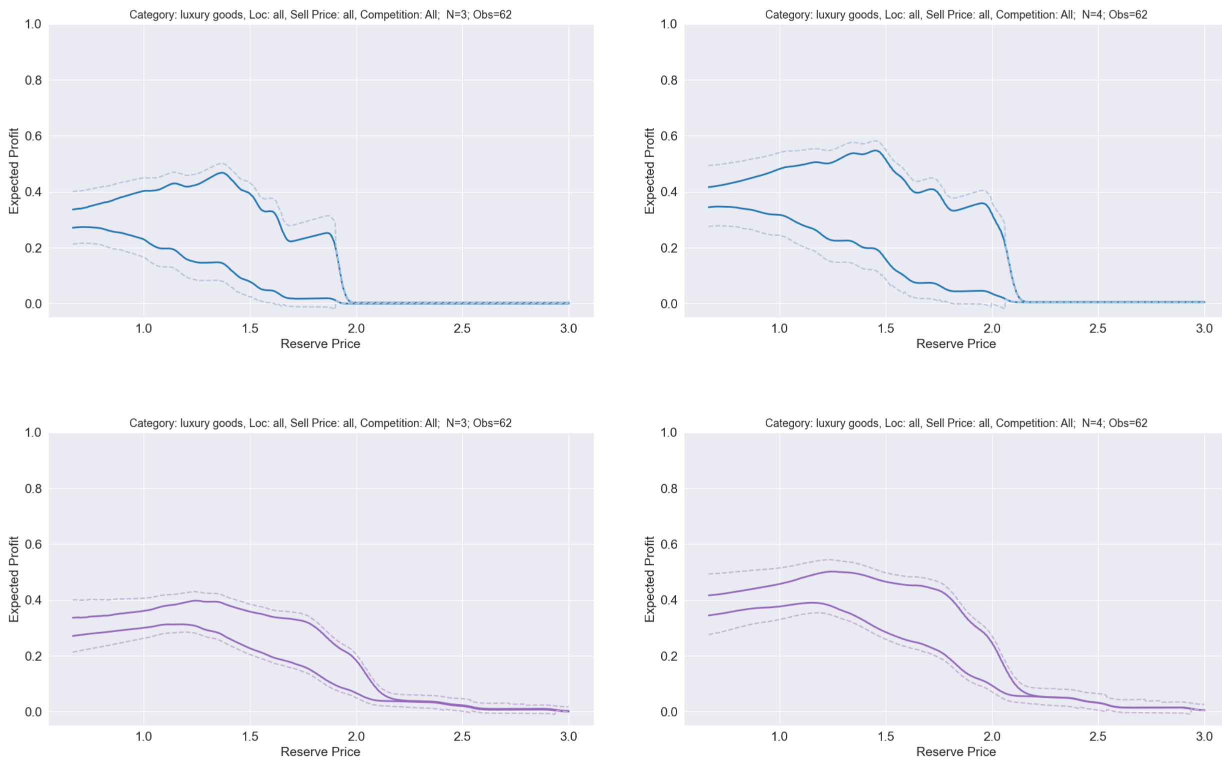
$N = 4$ $\qquad\qquad\qquad\qquad$ $N = 5$

Figure 11: Bounds on the expected profit against reserve price, for all 48 auction lots belonging to the Impressionist/20th/21st century art category, sold in New York City, and in the middle third of high estimates and middle third of competition as measured by number of bids. N ranges from 4 to 6, but we display $N = 4, 5$ only here for brevity. The top two graphs in blue are from equations 14 and 15, while the bottom two graphs in purple incorporate higher numbers of bidders using the assumption that valuations are independent of N. Dotted blue and light purple lines are the 95% confidence intervals for the respective bounds. Both reserve price and expected profits are scaled by the median of high estimates of the lots in the sample.



$N = 3$ $\qquad\qquad$ $N = 4$

Figure 12: Bounds on the expected profit against reserve price, for all 91 auction lots belonging to the Old Master art category, sold in New York City. N ranges from 2 to 6, but we display $N = 3, 4$ only here for brevity. The top two graphs in blue are from equations 14 and 15, while the bottom two graphs in purple incorporate higher numbers of bidders using the assumption that valuations are independent of N. Light blue and light purple lines are the 95% confidence intervals for the respective bounds. Both reserve price and expected profits are scaled by the median of high estimates of the lots in the sample.



$N = 3$ $\qquad$ $N = 4$

Figure 13: Bounds on the expected profit against reserve price, for all 308 auction lots belonging to the Luxury Goods category. These are watches and jewelry. N ranges from 2 to 8, but we display $N = 3, 4$ only here for brevity. The top two graphs in blue are from equations 14 and 15, while the bottom two graphs in purple incorporate higher numbers of bidders using the assumption that valuations are independent of N. Light blue and light purple lines are the 95% confidence intervals for the respective bounds. Both reserve price and expected profits are scaled by the median of high estimates of the lots in the sample.

of auctions. Furthermore, the minimax-regret solution can be directly applied to the derived profit bounds by picking the arg max of the lower bound, and the point choices to set the optimal reserve price are shown in the table as well.

Using these estimates, bounds on the increase to expected profit as compared to when setting the reserve at the low estimate can be stated. For each subsample, we scale the suggested optimal reserve by the median price of lots from that subsample, and show the expected minimum and maximum profit increase. The minimum increase is computed by taking the lower bound on profit at the minimax-regret choice and subtracting the upper bound on profit at v_0, i.e. the auctioneer's low estimate which is where the reserve is usually set. Similarly, the maximum increase is computed by taking the upper bound on profit at the minimax-regret choice and subtracting the lower bound on profit at v_0.

Notably, we find that there are numerous categories where setting a higher reserve price significantly increases profits. For example, in the auctions for Impressionist and Modern Art in New York City, setting the reserve at our proposed 1.51 times the high estimate (including Buyer's Premium) increases profits by at least USD$479k to at most USD$1.12M per auction lot, given that there are 3 bidders. This is a 7.6%-17.7% increase in profits. Given that a typical Impressionist/Modern Art auction has 50 lots with about 10 having 3 bidders, the profit increase per auction would be on average at least USD$4.8M.

	All Art		Imp. & Modern Art		Imp. & Modern Art, NYC	
	$N = 3$	$N = 4$	$N = 3$	$N = 4$	$N = 3$	$N = 4$
Interval	[1.06, 5.55]	[0.80, 5.31]	[1.02, 5.83]	[0.67, 5.70]	[1.04, 7.87]	[0.67, 7.79]
MMR Choice	1.49	1.67	1.66	1.66	1.51	1.40
Avg. Transact Price	$2.50M	$3.59M	$2.79M	$4.08M	$6.31M	$9.26M
π **Increase per Lot**	[$210k, $420k]	[$19k, $369k]	[$226k, $504k]	[$-11k, $436k]	[$479k, $1.12M]	[$-319k, $1.12M]
95%CI	[$-8k, $588k]	[$-273k, $651k]	[$9k, $665k]	[$-333k, $745k]	[$-225k, $1.83M]	[$-2.22M, $3.02M]

	Old Master Art		Luxury Goods		Imp. & Modern Art, NYC, mid/mid	
	$N = 3$	$N = 4$	$N = 3$	$N = 4$	$N = 4$	$N = 5$
Interval	[1.02, 5.23]	[0.67, 5.19]	[0.67, 1.79]	[0.67, 1.83]	[0.71, 1.20]	[0.71, 1.41]
MMR Choice	1.30	1.12	1.18	1.15	0.79	1.00
Avg. Transact Price	$433k	$1.01M	$271k	$176k	$3.34M	$3.09M
π **Increase per Lot**	[$43k, $119k]	[$-2k, $90k]	[$-6k, $32k]	[$-5k, $28k]	[$-131k, $130k]	[$-120k, $179k]
95%CI	[$-7k, $173k]	[$-302k, $453k]	[$-31k, $58k]	[$-23k, $43k]	[$-694k, $718k]	[$-727k, $789k]

Table 4: Profit Increases from selecting a higher reserve for the 6 samples. The optimal reserve interval is calculated from 19 and 20. The minimax regret (MMR) choice is computed by taking the argmax of the lower bound, and includes the Buyer's Premium. Interval and minimax-regret choice are scaled by the median high estimate from the sample. Note that the auctioneer's assumed reserve, v_0, is defined as the median of the auctioneer's low estimates for that sample and is 0.67 of the high estimate usually. 95% CI are computed from the equation 16. All monetary figures are in USD and converted from foreign currency if necessary.

6 Conclusion

We constructed a large novel dataset on live art auctions from the two largest auction houses in the world and provided close approximations on the number of bidders. We use the top two bids and the approximate number of bidders to estimate profit bounds non-parametrically, and solve the minimax-regret decision problem applied to this partially-identified model.

The results show that there are numerous subsamples of auctions where setting a higher reserve than at the low estimate as currently set on default will increase profits significantly, to the order of at least millions of dollars per auction.

For future work, our estimated number of bidders is likely a lower bound on the true number of bidders, rather than an accurate true estimation. Finding a way to generate informative bounds using this weaker assumption would lead to more confidence in the results.

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 5: 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

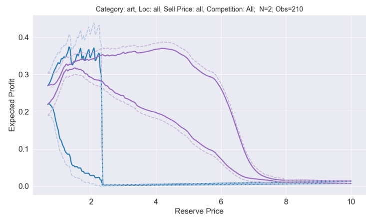
Hong Kong — Modern, Williamson Pink Star & Contemporary Auctions	link
London — The Jubilee Auction and Modern & Contemporary Evening Auction	link
Paris — Art Contemporain Evening Auction	link
New York — The Now & Contemporary Evening Auctions	link
New York — Modern Evening Auction	link
New York — The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

Table 6: Sotheby's YouTube Data Sources

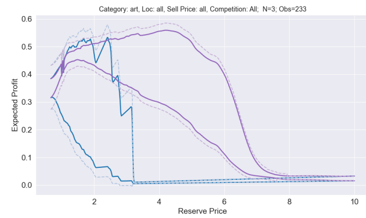
Saleroom Location	Christie's		Sotheby's	
	Threshold	Rate	Threshold	Rate
Hong Kong	≤HK$7.5M	26.0%	≤HK$7,500,000	26.0%
	>HK$7.5M and ≤ HK$50M	20.0%	>HK$7.5M and ≤HK$40M	20.0%
	>HK$50M	14.5%	>HK$40M	13.9%
London	≤£700k	26.0%	≤£800k	26.0%
	>£700,000 and ≤£4.5M	20.0%	>£800k and ≤£3.8M	20.0%
	>£4.5M	14.5%	>£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.

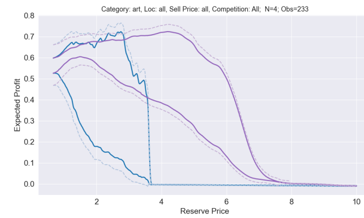
Table 7: Buyer's Premiums in Christie's and Sotheby's Auctions



$N = 2$

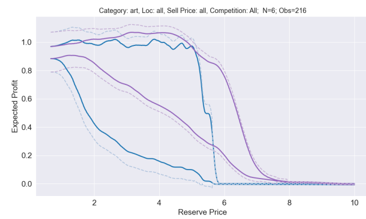


$N = 3$

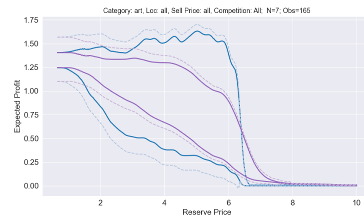


$N = 4$

$N = 5$

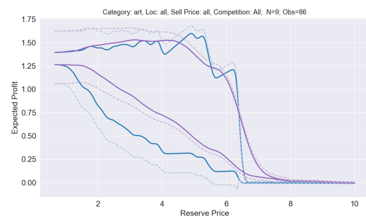


$N = 6$



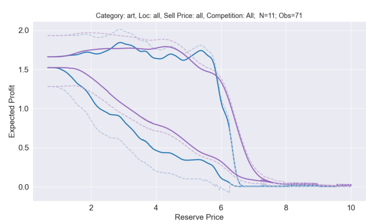
$N = 7$

$N = 8$

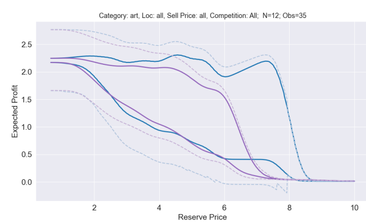


$N = 9$

$N = 10$

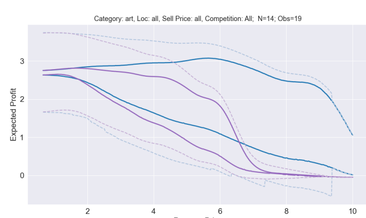


$N = 11$

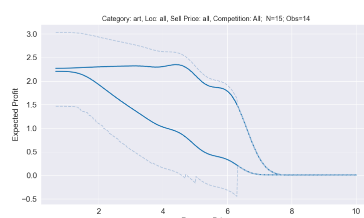


$N = 12$

$N = 13$

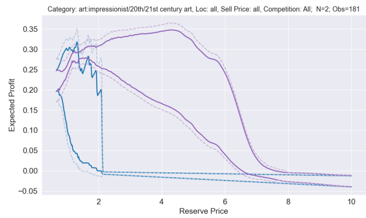


$N = 14$

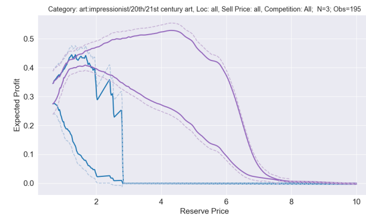


$N = 15$

Figure 14: Bounds on the expected profit against reserve price, for all 1712 auction lots belonging to the Art category. N ranges from 2 to 15, with higher N removed due to low sample counts. Blue bounds are estimated from 14 and 15, while the purple bounds incorporate higher numbers of bidders using the assumption that valuations are independent of N. Light blue and light purple lines are the 95% confidence intervals for the respective bounds.

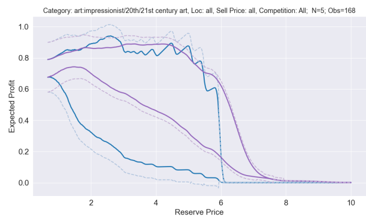


$N = 2$

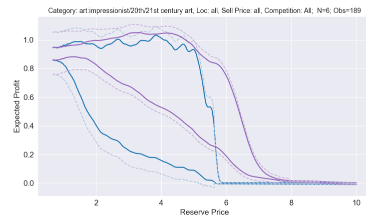


$N = 3$

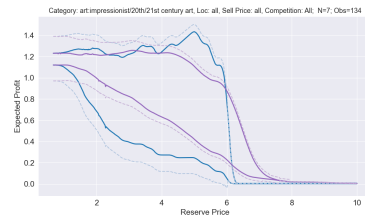
$N = 4$



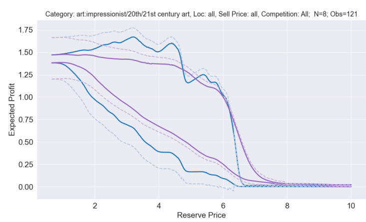
$N = 5$



$N = 6$



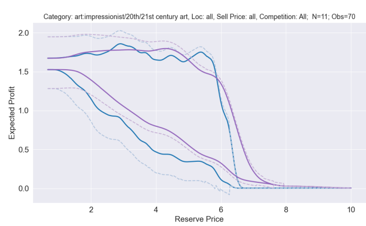
$N = 7$



$N = 8$

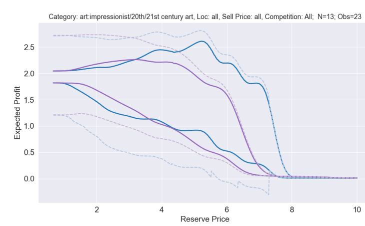
$N = 9$

$N = 10$

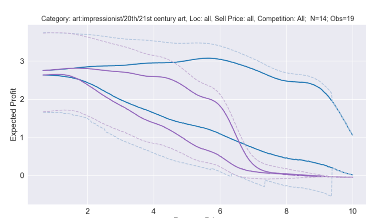


$N = 11$

$N = 12$



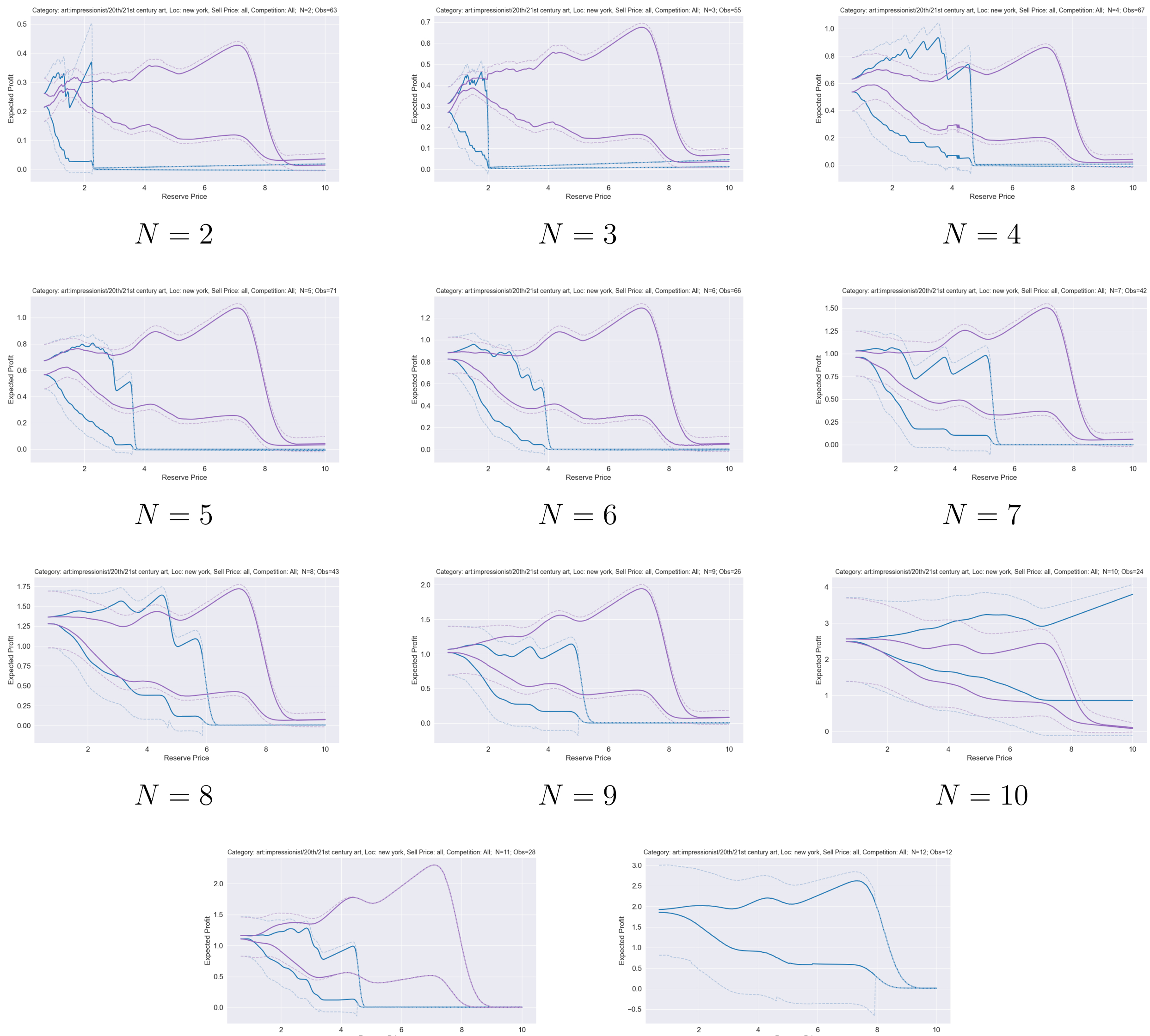
$N = 13$



$N = 14$

$N = 15$

Figure 15: Bounds on the expected profit against reserve price, for all 1477 auction lots belonging to the Impressionist/20th Century/21st Century Art category. N ranges from 2 to 15, with higher N removed due to low sample counts. Blue bounds are estimated from 14 and 15, while the purple bounds incorporate higher numbers of bidders using the assumption that valuations are independent of N. Light blue and light purple lines are the 95% confidence intervals for the respective bounds.



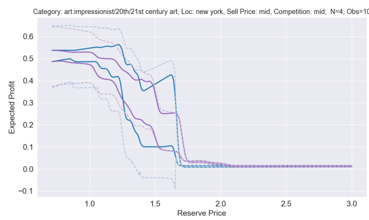
$N = 2$ $N = 3$ $N = 4$

$N = 5$ $N = 6$ $N = 7$

$N = 8$ $N = 9$ $N = 10$

$N = 11$ $N = 12$

Figure 16: Bounds on the expected profit against reserve price, for all 497 auction lots belonging to the Impressionist/20th Century/21st Century Art sold in New York City category. N ranges from 2 to 12, with higher N removed due to low sample counts. Blue bounds are estimated from 14 and 15, while the purple bounds incorporate higher numbers of bidders using the assumption that valuations are independent of N. Light blue and light purple lines are the 95% confidence intervals for the respective bounds.



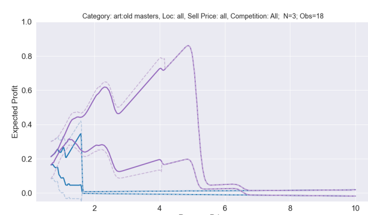
$N = 4$

$N = 5$

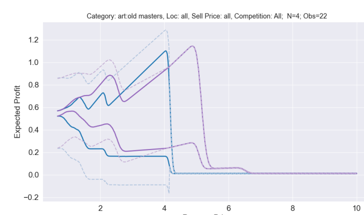
$N = 6$

Figure 17: Bounds on the expected profit against reserve price, for all 48 auction lots belonging to the Impressionist/20th Century/21st Century Art sold in New York City category, with middle one third percentile of high estimates and middle one third percentile of competition as measured by the number of bids. N ranges from 4 to 6, with higher N removed due to low sample counts. Blue bounds are estimated from 14 and 15, while the purple bounds incorporate higher numbers of bidders using the assumption that valuations are independent of N. Light blue and light purple lines are the 95% confidence intervals for the respective bounds.

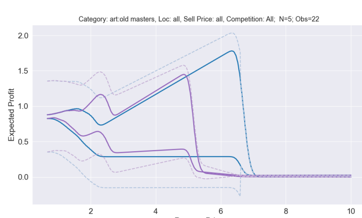
$N = 2$



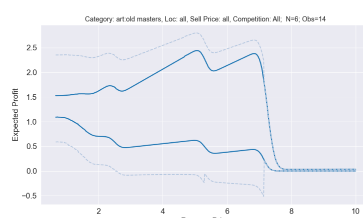
$N = 3$



$N = 4$



$N = 5$



$N = 6$

Figure 18: Bounds on the expected profit against reserve price, for all 91 auction lots belonging to the Old Master Art category. N ranges from 2 to 6, with higher N removed due to low sample counts. Blue bounds are estimated from 14 and 15, while the purple bounds incorporate higher numbers of bidders using the assumption that valuations are independent of N. Light blue and light purple lines are the 95% confidence intervals for the respective bounds.

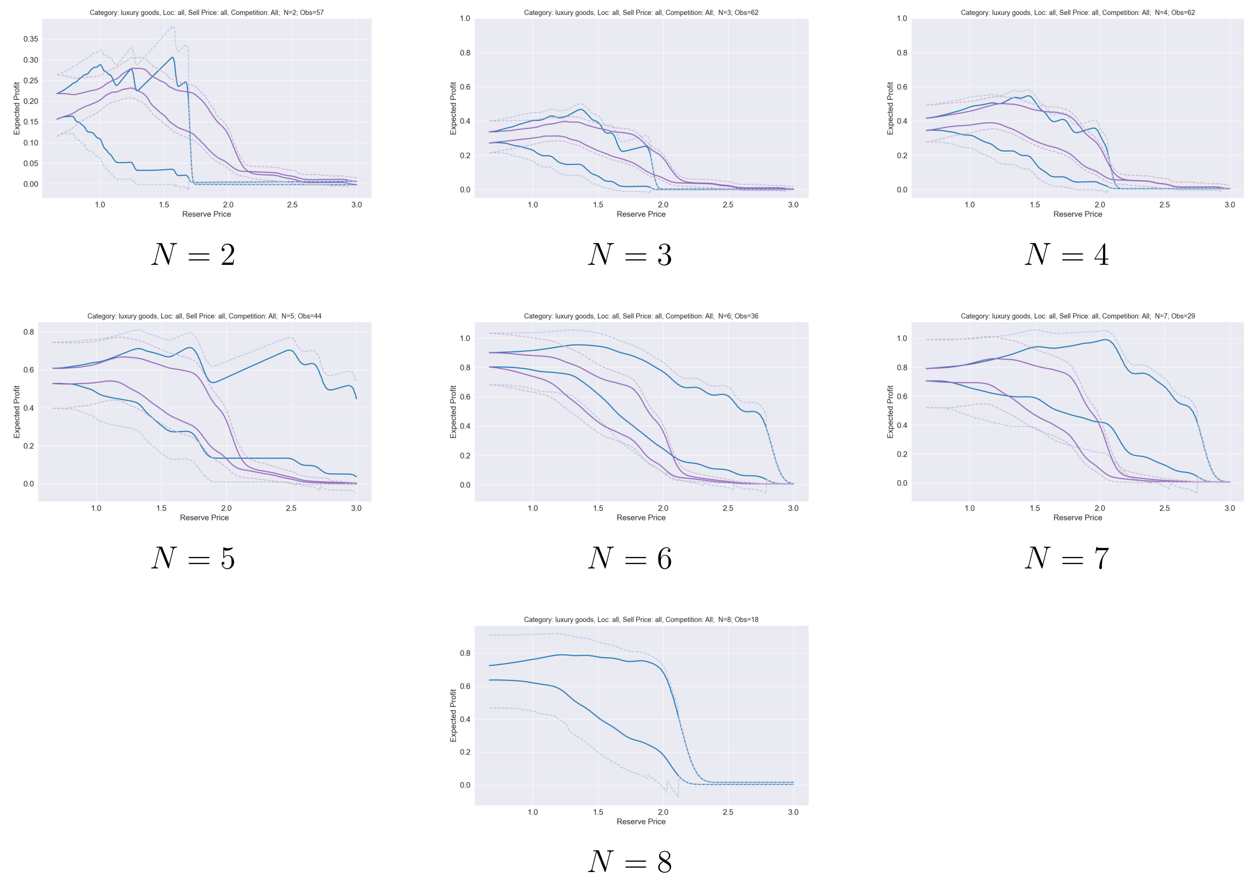


Figure 19: Bounds on the expected profit against reserve price, for all 308 auction lots belonging to the Luxury Goods category. N ranges from 2 to 8, with higher N removed due to low sample counts. Blue bounds are estimated from 14 and 15, while the purple bounds incorporate higher numbers of bidders using the assumption that valuations are independent of N. Light blue and light purple lines are the 95% confidence intervals for the respective bounds.

Appendix A. Proofs

Proof of Theorem 1.

First, I show that the upper bound generated with $\max(S) = \bar{n}$ is the tightest upper bound for $F_{n:n,s}$ from all of $s \in S$.

Define $\bar{n}_2, \bar{n}_1 \in \mathbb{N}$ and let $n < \bar{n}_1 = \bar{n}_2 - 1 \leq \bar{n}$. Take the difference of these upper bounds,

$$\begin{aligned}
F_{n:n,\bar{n}_1}(v) - F_{n:n,\bar{n}_2}(v) &= \frac{n}{\bar{n}_1} F_{\bar{n}_1-1:\bar{n}_1}(v) - \frac{n}{\bar{n}_2} F_{\bar{n}_2-1:\bar{n}_2}(v) - \frac{n}{\bar{n}_2(\bar{n}_2-1)} F_{\bar{n}_2-1:\bar{n}_2}(v) \\
&= \frac{n}{\bar{n}_1} F_{\bar{n}_1-1:\bar{n}_1}(v) - F_{\bar{n}_2-1:\bar{n}_2}(v) * \left(\frac{n + n\,(\bar{n}_2 - 1)}{\bar{n}_2\,(\bar{n}_2-1)} \right) \\
&= \frac{n}{\bar{n}_1} F_{\bar{n}_1-1:\bar{n}_1}(v) - \frac{n}{\bar{n}_2 - 1} F_{\bar{n}_2-1:\bar{n}_2}(v) \\
&= \frac{n}{\bar{n}_1} \left(F_{\bar{n}_1-1:\bar{n}_1}(v) - F_{\bar{n}_2-1:\bar{n}_2}(v) \right)
\end{aligned}$$

Recall that the CDF of the second highest order statistic is $F_{k-1:k}(v) = kF(v)^{k-1} - (k-1)F(v)^k$. So,

$$\begin{aligned}
F_{\bar{n}_1-1:\bar{n}_1}(v) - F_{\bar{n}_2-1:\bar{n}_2}(v) &= \bar{n}_1 F^{\bar{n}_1-1}(v) - (\bar{n}_1 - 1)\, F^{\bar{n}_1}(v) - \left[(\bar{n}_1+1) F^{\bar{n}_1}(v) - \bar{n}_1 F^{\bar{n}_1+1}(v) \right] \\
&= \bar{n}_1 F^{\bar{n}_1-1}(v) - 2\bar{n}_1 F^{\bar{n}_1}(v) + \bar{n}_1 F^{\bar{n}_1+1}(v) \\
&= \bar{n}_1 \left[F^{\bar{n}_1-1}(v) - 2F^{\bar{n}_1}(v) + F^{\bar{n}_1+1}(v) \right] \\
&= \bar{n}_1 F(v)^{\bar{n}_1-1} \left[1 - 2F(v) + F(v)^2 \right] \\
&= \bar{n}_1 F(v)^{\bar{n}_1-1} \left(1 - F(v) \right) \left(1 - F(v) \right) \\
&\geq 0
\end{aligned}$$

So $F_{n:n,\bar{n}_1}(v) \geq F_{n:n,\bar{n}_2}(v)$ for all v. It follows by an inductive argument that $F_{n:n,\bar{n}}(v)$ is indeed the lowest upper bound.

Now, let us consider the lower bounds. In the first scenario, assume that valuations are *i.i.d.* Then the lower bounds $F_{n:n,s}$ from $s \in S$ are all equally tight, as shown below. Define

$\bar{n}_2, \bar{n}_1 \in \mathbb{N}$ and let $n < \bar{n}_1 = \bar{n}_2 - 1 \leq \bar{n}$. Take the difference of these upper bounds,

$$\begin{aligned}
F_{n:n,\bar{n}_2}(v) - F_{n:n,\bar{n}_1}(v) &= \frac{n}{\bar{n}_2(\bar{n}_2-1)} F_{\bar{n}_2-1:\bar{n}_2}(v) + \frac{n}{\bar{n}_2}\left[\phi_{\bar{n}_2-1:\bar{n}_2}\left(F_{\bar{n}_2-1:\bar{n}_2}(v)\right)\right]^{\bar{n}_2} \\
&\quad - \frac{n}{\bar{n}_1}\left[\phi_{\bar{n}_1-1:\bar{n}_1}\left(F_{\bar{n}_1-1:\bar{n}_1}(v)\right)\right]^{\bar{n}_1} \\
&= \frac{n}{\bar{n}_2(\bar{n}_2-1)}\left\{F_{\bar{n}_2-1:\bar{n}_2}(v) + (\bar{n}_2-1)\left[\phi_{\bar{n}_2-1:\bar{n}_2}\left(F_{\bar{n}_2-1:\bar{n}_2}(v)\right)\right]^{\bar{n}_2}\right\} \\
&\quad - \frac{n}{\bar{n}_2(\bar{n}_2-1)}\bar{n}_2\left[\phi_{\bar{n}_1-1:\bar{n}_1}\left(F_{\bar{n}_1-1:\bar{n}_1}(v)\right)\right]^{\bar{n}_1}
\end{aligned}$$

Since valuations are *i.i.d.*, $\phi_{k-1:k}\left(F_{k-1:k}(v)\right) = F(v)$, so we have,

$$\begin{aligned}
F_{n:n,\bar{n}_2}(v) - F_{n:n,\bar{n}_1}(v) &= \frac{n}{\bar{n}_2(\bar{n}_2-1)}\left\{F_{\bar{n}_2-1:\bar{n}_2}(v) + \bar{n}_1 F(v)^{\bar{n}_2} - \bar{n}_2 F(v)^{\bar{n}_1}\right\} \\
&= \frac{n}{\bar{n}_2(\bar{n}_2-1)}\left\{F_{\bar{n}_2-1:\bar{n}_2}(v) - \left[\bar{n}_2 F(v)^{\bar{n}_1} - \bar{n}_1 F(v)^{\bar{n}_2}\right]\right\} \\
&= \frac{n}{\bar{n}_2(\bar{n}_2-1)}\left\{F_{\bar{n}_2-1:\bar{n}_2}(v) - F_{\bar{n}_2-1:\bar{n}_2}(v)\right\} \\
&= 0
\end{aligned}$$

So we have $F_{n:n,\bar{n}_2}(v) = F_{n:n,\bar{n}_1}(v)$. Then, again following an inductive argument, the bounds for $F_{n:n}$ from S are all equally tight.

Now suppose that valuations are common. In this case, I show that the lower bounds for $F_{n:n,s}$ from $s \in S$ are increasing in in s. Define $\bar{n}_2, \bar{n}_1 \in \mathbb{N}$ and let $n < \bar{n}_1 = \bar{n}_2 - 1 \leq \bar{n}$. If valuations are common, $F_{\bar{n}_2-1:\bar{n}_2}(v) = F_{\bar{n}_1-1:\bar{n}_1}(v) \equiv H$. So write,

$$\begin{aligned}
F_{n:n,\bar{n}_2}(v) - F_{n:n,\bar{n}_1}(v) &= \frac{n}{\bar{n}_2(\bar{n}_2-1)}\left\{H + \bar{n}_1\left[\phi_{\bar{n}_2-1:\bar{n}_2}(H)\right]^{\bar{n}_2} - \bar{n}_2\left[\phi_{\bar{n}_1-1:\bar{n}_1}(H)\right]^{\bar{n}_1}\right\} \\
&= \frac{n}{\bar{n}_2(\bar{n}_2-1)}\left\{H + \bar{n}_1\left[\phi_{\bar{n}_1:\bar{n}_1+1}(H)\right]^{\bar{n}_1+1} - (\bar{n}_1+1)\left[\phi_{\bar{n}_1-1:\bar{n}_1}(H)\right]^{\bar{n}_1}\right\}
\end{aligned}$$

So we want to show that $H + \bar{n}_1\left[\phi_{\bar{n}_1:\bar{n}_1+1}(H)\right]^{\bar{n}_1+1} - (\bar{n}_1+1)\left[\phi_{\bar{n}_1-1:\bar{n}_1}(H)\right]^{\bar{n}_1} \geq 0$. Now, again recall that $H = k\phi_{k-1:k}^{k-1} - (k-1)\phi_{k-1:k}^{k}$. Substitute H where $k = \bar{n}_1 + 1$, then

$$\begin{aligned}
&H + \bar{n}_1\left[\phi_{\bar{n}_1:\bar{n}_1+1}(H)\right]^{\bar{n}_1+1} - (\bar{n}_1+1)\left[\phi_{\bar{n}_1-1:\bar{n}_1}(H)\right]^{\bar{n}_1} \\
&= (\bar{n}_1+1)\left[\phi_{\bar{n}_1:\bar{n}_1+1}(H)\right]^{\bar{n}_1} - (\bar{n}_1+1)\left[\phi_{\bar{n}_1-1:\bar{n}_1}(H)\right]^{\bar{n}_1} \\
&= (\bar{n}_1+1)\left[\phi_{\bar{n}_1:\bar{n}_1+1}(H)^{\bar{n}_1} - \phi_{\bar{n}_1-1:\bar{n}_1}(H)^{\bar{n}_1}\right]
\end{aligned}$$

From Lemma 3, $\phi_{\bar{n}_1:\bar{n}_1+1}(H) \geq \phi_{\bar{n}_1-1:\bar{n}_1}(H)$. So $F_{n:n,\bar{n}_2}(v) - F_{n:n,\bar{n}_1}(v) \geq 0$. It follows by an inductive argument that the lower bounds for $F_{n:n,s}$ from $s \in S$ are increasing in in s, and thus $F_{n:n,\bar{n}}(v)$ is indeed the greatest lower bound. ■

Lemma 3 (Ordering of $\phi_{k:k+1}$ with respect to k) *Define $\phi_{i:n} : [0,1] \to [0,1]$ as in equation 6. Then,*

$$\phi_{n:n+1}(H) \geq \phi_{n-1:n}(H)$$

for all $n \geq 2$ and $H \in [0,1]$.

Proof. For reduced notation let $\phi \equiv \phi_{n-1:n}$ in this proof. Recall that $H = n\phi^{n-1} - (n-1)\phi^n$. Now differentiate this with respect to n. To do so, we need to use the identity $\left((\alpha_n)^{\beta(n)}\right)' = \alpha(n)^{\beta(n)}\beta'(n)\log\alpha(n) + \alpha(n)^{\beta(n)-1}$. So,

$$\begin{aligned}
\left(n\phi^{n-1}\right)'_n &= \phi^{n-1} + n\left[\phi^{n-1}\log\phi + \phi^{n-2}(n-1)\phi'\right] \\
\left((n-1)\phi^{n}\right)'_n &= \phi^{n} + (n-1)\left[\phi^{n}\log\phi + \phi^{n-1}n\phi'\right]
\end{aligned}$$

Then it follows that,

$$\phi^{n-1} + n\left[\phi^{n-1}\log\phi + \phi^{n-2}(n-1)\phi'\right] = \phi^{n} + (n-1)\left[\phi^{n}\log\phi + \phi^{n-1}n\phi'\right]$$

Rearranging,

$$\phi' = \frac{\phi - \phi^2 + n\phi\log\phi - (n-1)\phi^2\log\phi}{n(n-1)(\phi-1)}$$

Recall that $\phi \in [0,1]$, so to show this derivative is ≥ 0, we only need to show that the numerator ≤ 0. We can write,

$$\begin{aligned}
1 - \phi + n\log\phi - (n-1)\phi\log\phi &= 1 - \phi + [n - (n-1)\phi]\log\phi \\
&\leq 1 - \phi + [n - (n-1)\phi](\phi - 1) \qquad \text{(i)} \\
&= (1-\phi)^2[1-n] \\
&\leq 0
\end{aligned}$$

(i): This follows from the Taylor expansion centered at 1: $\log x = (x-1) - \frac{(x-1)^2}{2} + \frac{(x-1)^3}{3} - \frac{(x-1)^4}{4} + \ldots$. So $\log(x) \leq x - 1$ for all $x \in [0,1]$. ■

Proof of Lemma 2.

A desired π^* can be constructed as follows. Without loss of generality, suppose that

$p < q$. We can write what is in between p, q as a convex combination of the bounds,

$$\pi^*(r) = \begin{cases} \pi_L(r) & r \leq p \\ \frac{q-r}{q-p}\pi_L(r) + \frac{r-p}{q-p}\pi_U(r) & p < r < q \\ \pi_U(r) & r \geq q \end{cases}$$

First, it immediately follows that π^* is continuous at $C \setminus \{p, q\}$ since π_L, π_U are continuous and any linear combination of continuous functions is continuous. Next, π^* is also continuous at p and q. To see this for p, first, the left limit is $\lim_{r \to p^-} \pi^*(r) = \lim_{r \to p^-} \pi_L(r) = \pi_L(p)$. Meanwhile, the right limit is, $\lim_{r \to p^+} \pi^*(r) = \lim_{r \to p^+} \frac{q-r}{q-p}\pi_L(r) + \frac{r-p}{q-p}\pi_U(r) = \pi_L(p)$. So $\lim_{r \to p} \pi^*(r)$ exists and is equal to $\pi_L(p) = \pi^*(p)$. By a similar argument, $\lim_{r \to q} \pi^*(r)$ exists and is equal to $\pi_U(q) = \pi^*(q)$. ∎

Proof of Theorem 2.

First (Scenario 1), I will show that taking any $\hat{r} \in \underset{r \in C}{\operatorname{argmax}}\, \pi_U(r)$, for some given $p \in C$, where $p \neq \hat{r}$, the function $\pi^*(r) : C \to \mathbb{R}$ with the properties (1)-(3) as stated in Lemma 2 maximizes regret, i.e.

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

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 2, 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 $p \in C$ where $p \neq \hat{r}$ is,

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

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

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

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. $|\operatorname{argmax}_{r \in C} \pi_L(r)| = |\operatorname{argmax}_{r \in C} \pi_U(r)| = 1$) and their argument $\hat{r}$ happens to be the same. First I show that $\forall \epsilon > 0$, $\exists \pi^* \in \Pi$ such that $\text{regret}_{\pi^*}(\hat{r}) \equiv \max_{d \in C} \pi^*(d) - \pi^*(\hat{r}) > \pi_U(\hat{r}) - \pi_L(\hat{r}) - \epsilon$. Pick any $\epsilon > 0$. By the continuity of $\pi_U(.)$, there exists a $\delta > 0$ such that for some $t \in C$, $|\hat{r} - t| < \delta \implies \pi_U(\hat{r}) - \pi_U(t) < \epsilon$. Pick such a δ and a corresponding $t \neq \hat{r}$ where $|\hat{r} - t| < \delta$. By Lemma 2, it is possible to construct a $\pi^* : C \to \mathbb{R}$ 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 $r \in C$, it is always possible to pick another $\hat{r} \neq r$, where $\hat{r} \in \operatorname{argmax}_{r \in C} \pi_U(r)$. 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 $p \in C$ 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

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

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

■

References

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 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.

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.

Smith, R. (2007): "An Overview of the Tesseract OCR Engine," in *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, vol. 2, pp. 629–633.

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