Three Essays on Collusion in English Auctions: Theory and Application

Uma Kaplan

A Thesis

in

The Department

of

Economics

Presented in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy (Economics) at

Concordia University

Montréal, Québec, Canada

November 2017

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Uma Kaplan**

Entitled: **Three Essays on Collusion in English Auctions: Theory and Application**

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Economics)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the Final Examining Committee:

______________________ Chair
Dr. Amy Swiffen

______________________ External Examiner
Dr. Decio Coviello

______________________ External to Program Examiner
Dr. Mélina Mailhot

______________________ Examiner
Dr. Dipjyoti Majumdar

______________________ Examiner
Dr. Huan Xie

______________________ Supervisor
Dr. Szilvia Pápai

Approved by ______________________
Dr. Greg LeBlanc, Chair
Department of Economics

December 14, 2017

____
Dr. André Roy, Dean
Faculty of Arts and Science

Abstract

**Three Essays on Collusion in English Auctions:
Theory and Application**

Uma Kaplan, Ph.D.

Concordia University, 2017

This thesis is composed of three chapters that examine topics related to collusion in English auctions. In the first chapter, we develop a fully nonparametric identification framework and a test of collusion in ascending bid auctions. Assuming efficient collusion, we show that the underlying distributions of values can be identified despite collusive behavior when there is at least one known competitive bidder. We propose a nonparametric estimation procedure for the distributions of values and a bootstrap test of the null hypothesis of competitive behavior against the alternative of collusion. In the second chapter, we adopt a copula-based approach to identification. We succeed in showing that joint distribution function of private valuations is identifiable under certain conditions. Finally, we propose a semiparametric strategy, based on Archimedean copulas, to identify and estimate the model primitives and analyze the dependence relation between bids in English auctions. One advantage this approach has is that it allows us to separate the estimation of the marginal distribution from the estimation of the joint distribution of underlying bidder values. The third chapter is an empirical study of the municipal GIC auctions, motivated by the theoretical frameworks developed in the first two chapters.

Acknowledgments

I dedicate this thesis work to my family, and in memory of my late supervisor, Professor Art Shneyerov.

First and foremost, deepest gratitude goes to my late supervisor, Professor Shneyerov, for his constant guidance, patience, encouragement and support throughout the development of my doctoral research. His enthusiasm for research and teaching were always highly motivating. Professor Shneyerov has been the best academic supervisor and mentor that a Ph.D. student could ever hope for, and a role model from whom I have learned immensely during my academic journey. I would like to thank my supervisor, the late Prof. Art Shneyerov, and Prof. Vadim Marmer of University of British Columbia, for collaborating on my thesis research. The completion of this thesis would not have been possible without their help, encouragement and guidance.

I owe my profound gratitude to Prof. Szilvia Pápai (the Graduate Program Director), who has always been there for me with support and guidance throughout my studies and as my current supervisor, as well as to Prof. Greg LeBlanc (the Chair of the department) for his encouragement and advice.

To the members of my dissertation committee, Professors Huan Xie, Dipjyoti Majumdar, Decio Coviello, Mélina Mailhot and Amy Swiffen, for their invaluable contributions of time and expertise.

I want to extend my thanks to Professors Frank Muller, Tatyana Koreshkova, Nikolay Gospodinov, Faye Diamantoudi, Ming Li, Moshe Lander, Bryan Campbell and Gordon Fisher for their encouragement and support. Many thanks to my friends Irakli Japaridze, Sarah Lerner, Hans Garçon, Yves Tehou, Arezoo Banihashem and Sheridon Elliston for their support and encouragement. Special thanks goes to Elise Melançon for her help and assistance.

Most importantly, I want to thank my parents and my brother for their unconditional love, encouragement, patience and support throughout this journey. I also wish to acknowledge the financial assistance provided by the FRQSC[1].

[1]Fonds de recherche du Québec - Société et culture

Contribution of Authors

The first chapter of this thesis is based primarily on the manuscript currently circulated under the title ”Identifying Collusion in English Auctions”, based on the results of a joint research with Professor Art Shneyerov, and Professor Vadim Marmer of University of British Columbia. Dr. Shneyerov proposed the research idea. We conceived the research questions and collected the related literature. Dr. Shneyerov and Dr. Marmer developed the research design and methodology with my input and suggestions. We performed the data analysis, estimation and interpretation of results.

The second chapter of this thesis is joint work with Prof. Shneyerov, and Prof. Marmer. We conceived the research idea, collected the related literatures, and developed the research design and methodology together. Professor Shneyerov and Professor Marmer revised it critically for important intellectual content. I implemented the methodology, performed the estimation, interpreted and proved the results. Prof. Shneyerov commented it, and I revised it.

Contents

List of Figures

List of Tables

Chapter 1

De-censoring Approach to Identification and Estimation in English Auctions

1.1 Introduction

Collusion in auctions is an antitrust violation, but is nevertheless a pervasive phenomenon. It has been subject to many empirical studies. However, much of the research has focused on the sealed-bid, first-price auction format. For example, Porter and Zona (1993) and Bajari and Ye (2003) have studied collusion in highway procurement, while Porter and Zona (1999) and Pesendorfer (2000) have studied collusion in school milk procurement.[1]

There has been relatively less empirical or econometric work on collusion in *open* (or *English*) auctions, partly because of the dominance of the sealed-bid format in public procurement and sales.[2] The arrival of the Internet has greatly reduced the costs of bringing buyers and sellers together, and thus contributed to the increase in popularity of open auctions.

In this paper, we provide a structural nonparametric identification, estimation and testing framework for collusion in open auctions. The analysis focuses on the commonly accepted theoretic model of such auctions, namely the button (or thermometer) model, where the price is risen continuously and bidders

[1] See a survey by Harrington (2008) for more examples.

[2] One exception is Baldwin, Marshall, and Richard (1997), who have studied collusion is timber auctions.

drop out irrevocably. This model is becoming increasingly relevant for the auctions conducted over the Internet. The reason for this is the availability (and popularity) of electronic bidding agents that update bids continuously on bidders' behalf, which effectively implements the button model.

We make the most often exploited assumption: bidders draw their values independently (the IPV framework), however, allowing for *bidder asymmetries*. As the benchmark, and also the first step in our approach, we consider a model where there is no collusion. It is assumed that all the losing bids are observable. The main difficulty with identification and estimation of value distributions is the *censoring problem*: while the losing bids reveal bidder values, the winning value is censored. Our approach to de-censoring is based on the Nelson-Aalen estimator originally developed in the competing risks literature. We derive a simple formula that allows one to identify the value distribution of a particular bidder using only its losing bids and the losing bids of its highest rival.

Our main contribution is to extend this de-censoring technique to potentially colluding bidders. We restrict attention to collusion through cover (or phantom) bidding, a commonly used form of collusion in auctions.[3] In open auctions, the gains from collusion are maximal when the cartel members do not bid higher than the highest dropout price of their competitive rivals. In reality, they still may bid higher in order to conceal collusion. The exact nature of cover bidding is not needed for our analysis as only the leading cartel bid is used. For example, we allow non-participation, where instead of submitting a low bid, the cover bidder does not bid at all.[4]

Our result relies on several identifying assumptions. First, we assume that values are drawn independently, however, allowing for nonidentical distributions. The latter is important because the cartel is usually stronger on average than any of the non-cartel bidders.

Second, it is assumed that only one serious bid is submitted by the cartel, by a bidder that we call the *cartel leader*. The cartel leader is assumed to be selected *efficiently*, i.e. as the bidder with the highest valuation. This efficiency assumption is commonly made in the empirical literature on auctions, and is also supported by auction theory, as we explain in the next section.

[3]Collusion in auctions can take other forms, notably a market division agreement. See Hendricks and Porter (1989). Pesendorfer (2000) presents evidence that collusion takes different forms in highway procurement auctions in Florida and Texas.

[4]See, e.g. Porter and Zona (1993) and Baldwin et al. (1997).

Third, it is assumed that there is at least one competitive firm bidding against the cartel. This is often the case empirically, as e.g. in Porter and Zona (1993), Porter and Zona (1999), and Baldwin et al. (1997). Apart from this, the composition of the cartel does not need to be known. It is only important that the cartel leader bids competitively against the non-cartel firms.[5]

The cartel leader's value is censored from above by the competitive bid. At the same time, being the maximal value among the cartel bidders, it is censored from *below* by the second-highest cartel value. So unlike the competitive setup, here we have a joint censoring of the value both from above and below. Nevertheless, we show that the value distribution can be de-censored for each bidder in the cartel. This is because, as we show, the selection mechanism is identifiable under efficient collusion. This identification result is constrictive in that it gives a closed-form formula for the de-censored distribution of the values of the cartel members that is simple to estimate nonparametrically.

In our analysis, the cartel set should be understood as a *suspect* set. If competitive firms are mistakenly included in the cartel, the identification of the values of the colluders is unaffected as long as there is at least one competitive firm outside the cartel. Empirical studies often provide direct evidence as to who might be a potential colluder. This evidence often allows to plausibly argue that certain firms are "clean", i.e. did not participate in the conspiracy. Sometimes the cartel composition is known, as the defendant in an antitrust case as in Porter and Zona (1993) and Porter and Zona (1999). However, the strength of our approach is that it works under minimal knowledge concerning the composition of the cartel.

As we have argued, regardless of whether a bidder is competitive or not, its value distribution is identifiable through our de-censoring approach. This allows us to construct the counterfactual distribution of its bids under competition, even if the bidder's actual behaviour is collusive. If the bidder is competitive, then the counterfactual and actual distributions will coincide. However, if the bidder is collusive, we show that the counterfactual competitive bid distribution stochastically dominates the actual collusive one. This allows us to design a formal statistical test of the null hypothesis of competitive bidding against the alternative of collusive bidding. The test can be applied individually bidder by bidder, or can be applied jointly to a group of bidders.

[5]It is also permissible that the fringe firms collude among themselves.

Our test is initially developed at the individual bidder level. However, in combination with Bonferroni-type sequential hypothesis testing such as Holm (1979), it leads to a simple estimator of the composition of the cartel.[6] In our setting, the Holm-Bonferroni procedure works as follows. First, each bidder in the suspect set is tested and the p-value of the test recorded. Second, the p-values are ordered from smallest to highest. The bidders are then tested sequentially at appropriately adjusted levels of significance. If the competitive behaviour of the suspect bidder with the smallest p-value is not rejected, then the procedure terminates with no collusion found. If not, then this bidder is classified as a colluder, and the procedure moves to the next bidder in the order. This bidder is tested at a higher level of significance, and is included in the cartel following rejection. If no rejection occurs, then the test finds no presence of a cartel, as it is impossible to have a single-firm cartel. Continuing in this fashion until termination, the procedure results in an estimated cartel set with at least two bidders. The probability of one or more false bidder inclusions in the cartel is controlled overall at a predetermined level of significance, e.g. 5%. Moreover, the estimator of the cartel set is consistent.[7]

Once the collusive set has been estimated, we can proceed to estimate the collusive damages. For each colluding bidder, we can estimate its value distribution, which determines its dropout prices under competition. This allows us to recover the distribution of the auction price if all bidders were competitive, and to compare this counterfactual distribution with the actual distribution of the prices. For example, one could estimate the average loss of revenue due to collusion, and other statistics of the loss' distribution.

We are not aware of any previous research on nonparametric identification of collusion in open auctions. We believe our paper is the first one to investigate this issue. Our parallel contribution is that we propose full identification of model primitives under collusion. This can be used to address other important policy questions such as, for example, the optimal reserve price under collusion.

[6]This approach is also adopted in Schurter (2017) to estimate the composition of the cartel in a first-price auction.

[7]Recently, Coey, Larsen, and Sweeney (2014) considered placing bounds on collusive damages and proposed an approach based on bidder exclusion.

Relation to the Existing Literature

A common approach in the empirical literature on collusion in auctions is to use different bid responses to exogenous variation under collusion and competition. Porter and Zona (1993) study collusion in first-price highway procurement auctions conducted by the New York State Department of Transportation. They use measures of capacity and utilization rates as explanatory variables, and develop a likelihood-based model stability test across low and high bid ranks. The cartel composition is known in their case as they have access to court records. They find that parameter estimates are stable for the competitive group, but not for the cartel, which provides strong reduced-form evidence for collusion in the form of phantom bidding.

In another influential paper, Porter and Zona (1999) consider collusion in Ohio school milk auctions. They find that while the probability of submitting a bid falls with distance for non-defendant diaries, it increases for the defendants. Also, bid levels increase with distance for the non-defendants, but decrease for the defendants. These reduced-form finding convincingly point to collusion among the defendants, in the form territorial allocation.

Bajari and Ye (2003) adopt a structural approach in their study of collusion in highway procurement. The essence of their approach is to derive high-level testable predictions of the competitive model such as conditional independence and exchangeability, and build a statistical test based on these predictions. The main structural assumption is that the cartel is efficient, as in our paper. An extension of Bajari and Ye's approach to English auctions is difficult because censoring of the highest valuation implies that the dropout prices are correlated even under competition.

Aryal and Gabrielli (2013) consider a different test of collusion in first-price auctions. They exploit the variation in the number of bidders to argue that only the true model (competition or collusion) results in an invariant distributions of bidder values. Also for first-price auctions, Schurter (2017) exploits the potential presence of an exogenous shifter in the level of competition, and develops a test of collusion in first-price auctions based on the independence between the valuations and the shifter if the bidder is competitive.

There is very little work on collusion in open English auctions. Baldwin et al. (1997) considered collusion in US Forest Service Timber auctions. They consider a symmetric setting where bidders draw values from the same parametric distribution, and assume that the cartel is efficient. Within their parametric specifications, they compare likelihoods of competitive and collusive models and find support for collusion.

Asker (2010) estimates damages from collusion in a structural model of a knockout auction of stamp dealer cartel. Athey and Haile (2002) is a fundamental paper on identification in auctions, and provides a proper perspective on our identification results. Without collusion, and in the independent private values (IPV) framework as in our paper, it is known that the asymmetric ascending bid auction is identifiable even if only the winning bids are observable. This has been established in Athey and Haile (2002), building on the results for competing risks in Meilijson (1981). This approach has been recently extended by Komarova (2013). However, feasible nonparametric estimators have not been developed due to the complex nature of the identification arguments.[8] It is not known if the model is identifiable from the winning bids in the presence of collusion.

Our estimator *in the absence of collusion* is based on a well-known Nelson-Aalen estimator for models with random censoring.[9] However, its application to auctions is novel as is our approach to the identification and estimation of the value distributions under collusion.

We adopt the button model of the English auction. Haile and Tamer (2003) emphasize that losing bids do not necessarily reflect true values because of jump bidding in many real-world open auctions. Be this as it may in the traditional open auctions, the arrival of the Internet has opened door to new ascending-bid auctions that, as we have argued, conform more closely to the original "button" model considered in the theoretical literature.

Our main structural identification assumption is that the cartel is efficient. This assumption is commonly used in the empirical literature on auctions, e.g. Bajari and Ye (2003), Baldwin et al. (1997). Auction theory supports it as well: Graham and Marshall (1987) show that, if the bidding cartel is able to distribute the spoils of collusion ex ante, it can efficiently select the cartel leader using an open

[8]The identification using winning bids only relies on Pfaffian integral equations, which are very difficult to solve even numerically. See Brendstrup and Paarsch (n.d.), who instead appeal to parametric flexible-form maximum likelihood estimation. We should also mention that outside the IPV framework, the model is not identifiable even under symmetry. A recent paper by Aradillas-López, Gandhi, and Quint (2011) addresses partial identification of this model.

[9]See e.g. the discussion in Section 20.15 in van der Vaart (1998).

knockout auction. In addition, Mailath and Zemsky (1991) show that efficient collusion can be sustained through appropriate ex-post side payments between the cartel members if the values are independent, while Hendricks, Porter, and Tan (2008) show that this continues to be true if values are affiliated.[10] When cartel bidders are symmetric, a simple knockout auction exists that selects the leader efficiently and balances the budget ex post.

1.2 Identification under Competition

In the baseline competitive model, we consider a standard independent private values (IPV) setting where there are N bidders participating at an auction. The set of bidders is denoted as $\mathcal{N} = \{1, ..., N\}$.

Assumption 1 (IPV). *Each bidder $i \in \mathcal{N}$ draws its value independently from a cumulative distribution $F_i(\cdot)$ supported on $[0, \overline{v}]$.*

We allow the distributions F_i to be different across bidders, but assume that the support $[0, \overline{v}]$ is the same for all bidders. The density of F_i is denoted as f_i.

In an ascending button auction, only the dropout prices of the losing bidders are equal to valuations in a dominant strategy equilibrium. The valuation of the winner is censored from below by the highest dropout price among the losing bidders. For any bidder i, let V_i denote its value, and let V_{-i} denote the maximum value of its rivals, $V_{-i} = \max_{j \neq i} V_j$. The distribution of V_{-i} is denoted as $F_{-i}(\cdot)$. The indicator variable $w_i \in \{0, 1\}$ is equal to 1 if bidder i wins the auction, and is equal to 0 if he loses. If $w_i = 0$, V_i is observable, while V_i is censored from above by V_{-i} when $w_i = 1$. Let $g_i(v|w_i = 0)$ be the density of i's bids, or equivalently, the values conditional on *losing* the auction. It is directly identifiable from the data.

We now show how to recover F_i. Since V_i and V_{-i} are assumed to be independent, the Bayes rule

[10] However not if values are common. See Hendricks et al. (2008).

yields

$$g_i(v|w_i = 0) = \frac{f_i(v)(1 - F_{-i}(v))}{\mathbb{P}(w_i = 0)}$$
$$\implies f_i(v) = \frac{g_i(v|w_i = 0)\mathbb{P}(w_i = 0)}{1 - F_{-i}(v)}.$$

Dividing both sides of the last equation by $1 - F_i(v)$, we obtain

$$\frac{f_i(v)}{1 - F_i(v)} = \frac{g_i(v|w_i = 0)\mathbb{P}(w_i = 0)}{(1 - F_i(v))(1 - F_{-i}(v))}. \tag{1}$$

Our *key insight* is that the function that appears on the right-hand side in the denominator of (1) is directly identifiable. The independence between V_i and V_{-i} implies that

$$(1 - F_i(v))(1 - F_{-i}(v)) = \mathbb{P}(\min\{V_i, V_{-i}\} \geq v).$$

However,

$$B_i = \min\{V_i, V_{-i}\} = w_i V_{-i} + (1 - w_i)V_i$$

is in fact equal to bidder i's actual bid (whether losing or winning), and is directly observable. Its distribution,

$$G_i(v) \equiv \mathbb{P}(B_i \leq v),$$

is therefore directly identifiable from the data. Thus, the result in equation (1) can be equivalently stated as

$$\frac{f_i(v)}{1 - F_i(v)} = \frac{g_i(v|w_i = 0)\mathbb{P}(w_i = 0)}{1 - G_i(v)}, \tag{2}$$

where the expression on the right-hand side involves only terms that can be directly estimated from the data.

It will prove convenient to define

$$G_i^0(b) \equiv P(B_i \leq b, w_i = 0) = G_i(b|w_i = 0)\mathbb{P}(w_i = 0),$$

and its derivative

$$g_i^0(b) \equiv \frac{dG_i^0(b)}{db}.$$

We can now re-state the identification result in (2) as

$$-\frac{d\log(1-F_i(v))}{dv} = \frac{g_i^0(v)}{1-G_i(v)}.$$

The left-hand side of this equation can be recognized as a *full derivative*, so we can integrate this equation and recover the distribution of i's values $F_i(\cdot)$. The result is given in the proposition below.

Proposition 1 (Identification under competition)**.** *Under Assumption 1, we have*

$$F_i(v) = 1 - \exp\left(-\int_0^v \frac{dG_i^0(u)}{1-G_i(u)}\right). \tag{3}$$

This result can be viewed as an adaptation of the well-known Nelson-Aalen estimator originally developed for cumulative hazard functions (Aalen, 1978; Nelson, 1969, 1972) to ascending auctions. The functional that appears on the right-hand side of (3) will be used repeatedly in the sequel. It is defined, for any two functions $H_1(\cdot)$ and $H_2(\cdot)$, as[11]

$$\psi(H_1, H_2)(v) \equiv 1 - \exp\left(-\int_0^v \frac{dH_1(u)}{1-H_2(u)}\right). \tag{4}$$

Note that using the definition in (4), the result in (3) can be stated as $F_i(v) = \psi(G_i^0, G_i)(v)$.

1.3 Collusion

In this section, we show that the distributions of bidder valuations are identifiable even in the presence of collusion. We assume that a subset of bidders potentially forms a *bidding cartel*. The identification is shown under a number of assumptions.

First, we assume that the cartel is not all inclusive. That is, it is known to the researcher that at least

[11] This functional is well-defined when H_1 has bounded variation.

one bidder behaves competitively, i.e. bids up to its true value.[12] Denote the set of known competitive bidders as $\mathcal{N}_{com}$.

Assumption 2 (Competitive bidder)**.** *There is at least one known competitive bidder, i.e. the set* $\mathcal{N}_{com}$ *is non-empty.*

We assume that some bidders *may be* colluding. The colluding bidders are necessarily contained in

$$\mathcal{N}_{col} = \mathcal{N} \backslash \mathcal{N}_{com}.$$

We shall sometimes refer to $\mathcal{N}_{col}$ as the *suspect set*, as this set may also include some firms that are in fact competitive. It is important to note that the set of actually colluding bidders $\mathcal{C} \subseteq \mathcal{N}_{col}$ is not a priory known. We also allow for no collusion at all, in which case $\mathcal{C} = \emptyset$. Our identification approach is based on the idea that a cartel firm still behaves competitively if it is the *cartel leader*, i.e. the designated highest bidder from the cartel.

Second, we restrict attention to *efficient collusion*, where the ring (cartel) leader is the bidder with the highest valuation of the item.[13]

Assumption 3 (Efficient collusion)**.** *Cartel leader's valuation is equal to* $\max_{k \in \mathcal{C}} V_k$.

Let $\ell_i = 1$ indicate the event that bidder i has the leading (maximum) value in the suspect set $\mathcal{N}_{col}$, otherwise $\ell_i = 0$. This obviously includes the event when bidder i is the cartel leader under efficient collusion, however also requires i's value to be higher than any of the competitive bidders' values in $\mathcal{N}_{col}$. Note that together with our assumption that the distributions $F_i(\cdot)$ have the same support, efficient collusion implies that each suspect member has a positive probability of being the leader, i.e. $\mathbb{P}(\ell_i = 1) > 0$. By the Bayes rule,

$$\begin{aligned} f_i(v|\ell_i = 1) &= \frac{\mathbb{P}(\ell_i = 1|V_i = v) f_i(v)}{\mathbb{P}(\ell_i = 1)} \\ \implies f_i(v) &= \frac{\mathbb{P}(\ell_i = 1) f_i(v|\ell_i = 1)}{\mathbb{P}(\ell_i = 1|V_i = v)}. \end{aligned} \tag{5}$$

[12]This assumption can be relaxed, as we remark in the sequel.

[13]This assumption is plausible in empirical applications and frequently made in the literature. See e.g. Bajari and Ye (2003).

Conditional on being a leader, i bids competitively against the competitive fringe $\mathcal{N}_{com}$. This implies that the density $f_i(v|\ell_i = 1)$ is identifiable using the results in the previous section, i.e. by considering i's bids that are both leading ($\ell_i = 1$) and losing in the action ($w_i = 0$) against the competitive fringe. Let

$$V_{com} \equiv \max_{k \in \mathcal{N}_{com}} V_k$$

be the maximum value in the competitive fringe $\mathcal{N}_{com}$. In parallel to (3) in the previous section, the distribution of i's values conditional on leading the cartel,

$$F_i^{\ell}(v) \equiv F_i(v|\ell_i = 1),$$

is identifiable through the de-censoring formula

$$F_i^{\ell}(v) = \psi\big(G_i^{0,\ell}, G_i^{\ell}\big)(v), \tag{6}$$

where the distributions $G_i^{0,\ell}(b)$ and $G_i^{\ell}(b)$ are now conditional on being the cartel leader,

$$G_i^{0,\ell}(b) = \mathbb{P}(b_i \leq b, w_i = 0|\ell_i = 1), \qquad G_i^{\ell}(b) = \mathbb{P}(b_i \leq b|\ell_i = 1).$$

Note that both $G_i^{0,\ell}(b)$ and $G_i^{\ell}(b)$ are identifiable from the data.

Continuing the identification argument, the selection probability $\mathbb{P}(\ell_i = 1|V_i = v)$ that appears in (5) is not directly identifiable. In order to apply the above result, we propose a transformation that does not involve $\mathbb{P}(\ell_i = 1|V_i = v)$. Dividing both sides of (5) by $F_i(v)$, we obtain

$$\frac{F_i'(v)}{F_i(v)} = \frac{\mathbb{P}(\ell_i = 1) f_i(v|\ell_i = 1)}{\mathbb{P}(\ell_i = 1|V_i = v) F_i(v)}. \tag{7}$$

Under independence and efficient collusion, the leader selection probability is simply the product of the

CDFs of bidders in $\mathcal{N}_{col} \backslash \{i\}$,

$$\mathbb{P}(\ell_i = 1 | V_i = v) = \prod_{j \in \mathcal{N}_{col} \backslash \{i\}} F_j(v) \tag{8}$$

$$\implies P(\ell_i = 1 | V_i = v) F_i(v) = \prod_{j \in \mathcal{N}_{col}} F_j(v) \equiv F_{col}(v) \tag{9}$$

where $F_{col}(v)$ is the distribution of the maximum value V_{col} in the suspect set,

$$V_{col} \equiv \max_{k \in \mathcal{N}_{col}} V_k.$$

Since the bidder with valuation V_{col} bids competitively against the maximum value V_{com} in competitive fringe $\mathcal{N}_{com}$, the distribution $F_{col}(v)$ is identifiable by de-censoring in parallel to (3) from the previous section:

$$F_{col}(v) = \psi(G^0_{col}, G_{col})(v), \tag{10}$$

where

$$G^0_{col}(u) = \mathbb{P}\{\min\{V_{com}, V_{col}\} \leq u; w_{col} = 0\}, \qquad G_{col}(u) = \mathbb{P}\{\min\{V_{com}, V_{col}\} \leq u\}.$$

Here $w_{col} \in \{0, 1\}$ indicates whether or not the suspect leader wins the auction. Note that both G^0_{col} and G_{col} are identifiable because $\min\{V_{com}, V_{col}\}$ is observable.

Substituting (9) into (7), we obtain a differential equation for $F_i(v)$ that only involves identifiable objects,

$$\frac{dF_i(v)}{F_i(v)} = \frac{dF^\ell_i(v)}{F_{col}(v)}. \tag{11}$$

This differential equation can be integrated backwards using the boundary condition $F_i(\overline{v}) = 1$ to yield a unique solution given in the proposition below, which is our main result in this section.

Proposition 2 (Identification under efficient collusion)**.** *Under Assumptions 1–3, the distributions $F_i(\cdot)$*

are identifiable. The identification of $F_i(\cdot)$ for the known competitive bidders is unaffected and proceeds according to (3), as before. The identification of $\{F_i(\cdot) : i \in \mathcal{N}_{col}\}$ can be performed according to

$$F_i(v) = \exp\left(-\int_v^\infty \frac{dF_i^\ell(u)}{F_{col}(u)}\right), \tag{12}$$

where the distributions $F_i^\ell(v)$ and $F_{col}(v)$ are identifiable from the previous step according to (6) and (10) respectively.

The intuition behind this identification result can be summarized as follows. First, even though bidders in the cartel may submit noncompetitive "cover" bids, the cartel leader bids competitively against any competitive bidder (i.e. any bidder in the set $\mathcal{N}_{com}$). In particular, we use the fact that it bids competitively against the highest bidder in $\mathcal{N}_{com}$. The implication of this observation is that, *conditionally on being a cartel leader*, the bidder's behavior in the auction is in fact competitive. The de-censoring approach can be used to identify, for any suspect bidder, the distribution of valuations conditionally on leading the cartel.

Second, under our assumption that the cartel is efficient, the valuation of the cartel leader is censored from below. We have shown that the de-censoring approach can be suitably extended to uncover the marginal distribution of bidder values even in this case.

Assumption 2, which requires that there is at least one known competitive bidder, can be relaxed. If the seller is an active participant in the auction, then the seller's bid can be used instead of the maximum competitive bid for the purposes of identification, as long as it is independent of the maximum cartel value. The seller may or may not know that it is facing a cartel, and may or may not bid optimally. It would only be required that the seller's bids have support $[\underline{b}, \infty)$ for some $\underline{b} \geq 0$.

1.3.1 Identifying Collusion

The result in Proposition 2 can be used as a basis for a test of collusion. Regardless of whether bidder $i \in \mathcal{N}_{col}$ is colluding or not, and regardless of the potential presence of an unknown (but efficient) cartel, we can identify the *predicted* distribution of bidder i's bids if i were competitive. It is assumed that, if there is a cartel, it continues to operate with bidder i excluded. This (potentially counterfactual) distribution is denoted as $G_i^{pred}(v)$. As V_i, V_{-i} are independent if bidder i is competitive, the upper CDF of i's bid $B_i = \min\{V_i, V_{-i}\}$ is given by the product

$$\begin{aligned} 1 - G_i^{pred}(v) &= (1 - F_i(v))(1 - F_{-i}(v)), \\ \implies G_i^{pred}(v) &= 1 - (1 - F_i(v))(1 - F_{-i}(v)). \end{aligned} \tag{13}$$

In this formula, $F_i(v)$ is identifiable according to (12), and $F_{-i}(v)$, the distribution of the maximum of all bidder values excluding bidder i, is identifiable by an analogue to (3):

$$F_{-i}(v) = \psi(G_i(\cdot|w_i = 1)\mathbb{P}(w_i = 1), G_i(\cdot))(v). \tag{14}$$

Alternatively, since all the individual CDFs have been identified, one can take

$$F_{-i}(v) = \prod_{j \neq i} F_j(v). \tag{15}$$

It will be more convenient to use the latter expression for F_{-i}.

The actual behavior of bidder $i \in \mathcal{N}_{col}$ may be collusive. We now detail the assumptions on the bidding strategy of a *cover bidder*, i.e. a cartel member who is not the cartel leader. Let $h = ((i_1, p_1), ..., (i_k, p_k))$ be a dropout history, where $i_1, ..., i_k \in \mathcal{N}$ indicate the identities of those k bidders that have dropped out, and $p_1, ..., p_k$ denote their respective dropout prices. If there are no dropouts yet, we let $h = \emptyset$. A bidding strategy $B_i^*(v_{\mathcal{C}}, h)$ of a cover bidder specifies the maximum price up to which the bidder is willing to stay in the auction, given the history h, as a function of the realization of all the cartel valuations $v_{\mathcal{C}} \equiv (v_j)_{j \in \mathcal{C}}$.

For the cartel to maximize the gains from collusion, it must be the case that whenever the cartel leader wins the auction, i.e. $V_{com} < V_{col}$, the cover bidders drop out at or below V_{com}. In addition, it is reasonable to assume that the cartel members never drop out above their valuations. This way, should the cartel leader renege on its promise to bid up to its valuation and drop out earlier, the cover bidders will not suffer a loss from buying at prices higher than their valuations. We therefore make the following assumption concerning the bidding strategy of a cover bidder. Other than this assumption, a cover bidder's strategy is unrestricted.

Assumption 4. *For any cover bidder $i \in \mathcal{C}$, (i) its dropout price never exceeds its valuation, $B_i^*(v_{\mathcal{C}}, h) \leq v_i$, and (ii) whenever the dropout history h involves the last dropout by the highest competitive bidder, the cover bidder also drops out at that price: $B_i^*(v_{\mathcal{C}}, h) = V_{com}$.*

Since a cover bidder never wins auctions, its actual final bid will be given by $B_i^*(v_{\mathcal{C}}, h^*)$ for the realized history h^* after which it drops out before any other bidder does. The actual bid of a cartel member, as a random variable, is then given by

$$\tilde{B}_i = B_i^*(V_{\mathcal{C}}, h^*)\mathbb{1}\{\ell_i = 0\} + \min\{V_i, V_{com}\}\mathbb{1}\{\ell_i = 1\}.$$

For any bidder $i \in \mathcal{C}$, $V_{com} \leq V_{-i}$ and therefore

$$\min\{V_i, V_{com}\} \leq \min\{V_i, V_{-i}\}, \tag{16}$$

while

$$B_i^*(V_{\mathcal{C}}, h^*) \leq \min\{V_i, V_{-i}\}$$

by (i) and (ii) in Assumption 4. It follows that for cartel members, the counterfactual competitive distribution of bids $G_i^{pred}(b) = \mathbb{P}\{\min\{V_i, V_{-i}\} \leq b\}$ weakly stochastically dominates the actual distribution of bids $G_i(b) = \mathbb{P}\{\tilde{B}_i \leq b\}$. We show below that stochastic dominance holds in the strict sense.

Proposition 3 (Testable prediction for collusion)**.** *Under Assumptions 1-4, the predicted competitive distribution of i's bids is identified. Moreover, it strictly stochastically dominates the distribution of i's bids if bidder i is collusive: $G_i(b) \geq G_i^{pred}(b)$ for all b's, with strict inequalities for some b's.*

Proof. In order to prove that the inequality is strict for some b, by Theorem 1 in Hanoch and Levy

(1969),[14] it is sufficient to verify

$$\mathbb{E}[\tilde{B}_i] < \mathbb{E}[\min\{V_i, V_{-i}\}]. \tag{17}$$

For bidder $i \in \mathcal{C}$, define an event $\mathcal{A}_i = \{V_{com} < V_i < V_{col}\}$, and note that $\mathbb{P}(\mathcal{A}_i) > 0$, which holds since the distributions $\{F_j, j \in \mathcal{N}\}$ have the same support.[15] Write $\mathbb{E}[\tilde{B}_i] = \mathbb{E}[\tilde{B}_i \cdot \mathbb{1}(\mathcal{A}_i)] + \mathbb{E}[\tilde{B}_i \cdot \mathbb{1}(\mathcal{A}_i^c)]$. For the first term by Assumption 4, we have $\mathbb{E}[\tilde{B}_i \cdot \mathbb{1}(\mathcal{A}_i)] = \mathbb{E}[V_{com} \cdot \mathbb{1}(\mathcal{A}_i)] < \mathbb{E}[V_i \cdot \mathbb{1}(\mathcal{A}_i)] = \mathbb{E}[\min\{V_i, V_{-i}\} \cdot \mathbb{1}(\mathcal{A}_i)]$. Moreover, (16) implies that $\mathbb{E}[\tilde{B}_i \cdot \mathbb{1}(\mathcal{A}_i^c)] \leq \mathbb{E}[\min\{V_i, V_{-i}\} \cdot \mathbb{1}(\mathcal{A}_i^c)]$, and (17) follows.

□

1.4 Estimation

We consider an i.i.d. sample of L auctions, with each individual auction indexed by $l = 1, ..., L$. For simplicity, we assume that all N bidders participate.

The bids are denoted as b_{il}. For each bidder $i \in \mathcal{N}$, the maximal bid of its rival is denoted as $b_{-il} = \max\{b_{jl} : j \in \mathcal{N} \backslash \{i\}\}$. For $i \in \mathcal{N}$, $w_{il} \in \{0, 1\}$ denotes whether bidder i wins auction l: $w_{il} = 1$ if $b_{il} > b_{-il}$, and $w_{il} = 0$ if $b_{il} < b_{-il}$. In equilibrium, ties will have zero probability, so the allocation rule adopted for tied bids is immaterial. Conditional on losing, i.e. on $w_{il} = 0$, the bidder's valuation v_{il} is revealed and equal to its bid, while for a winning bid, it is only known that the valuation is at or above the bid: $v_{il} = b_{il}$ if $w_{il} = 0$, and $v_{il} \geq b_{il}$ if $w_{il} = 1$.

Our estimation strategy will be based on the plug-in approach, where the distributions appearing in the decensoring formulae are replaced by their empirical analogues. The distributions G_i^0, G_i can be consistently estimated as

$$\hat{G}_i(b) = \frac{1}{L} \sum_{l=1}^{L} \mathbb{1}(b_{il} \leq b), \qquad \hat{G}_i^0(b) = \frac{1}{L} \sum_{l=1}^{L} \mathbb{1}(b_{il} \leq b, w_{il} = 0). \tag{18}$$

[14] This theorem states that if $\int H(b) dG_0(b) \geq \int H(b) dG(b)$ for any non-decreasing function $H(\cdot)$, with a strict inequality for at least one such function, then $G_0(b) < G(b)$ for some b. In our case, we pick the identity function, $H(b) = b$, which leads to the comparison of the expected values.

[15] We use the notation $\mathbb{1}(\mathcal{A})$ for the indicator function of an event $\mathcal{A}$.

Plugging these estimators into (3), we obtain an estimator for the distribution of valuations of a competitive bidder i:

$$\hat{F}_i(v) = \psi(\hat{G}_i, \hat{G}_i^0)(v). \tag{19}$$

It can be shown, as an application of the Continuous Mapping Theorem, that the estimator $\hat{F}_i$ is consistent on the entire support $[0, \overline{v}]$. The rate of convergence can also be established by standard methods. However, we do not pursue this, as weak convergence results and the bootstrap approach discussed below will be our main tools for inference and testing.

Our main tool for deriving the asymptotic distributions of the estimators and their bootstrap approximations will be the Functional Delta Method (FDM).[16] Using the definition of the functional ψ in (4), its functional derivative, at $H_1 = G_i^0$ and $H_2 = G_i$, can be computed as

$$\psi'(h_1, h_2)(v) = (1 - F_i(v)) \left(\int_0^v \frac{dh_1(u)}{1 - G_i(u)} + \int_0^v \frac{h_2(u) dG_i^0(u)}{(1 - G_i(u))^2} \right). \tag{20}$$

Standard results for weak convergence of empirical processes imply, jointly for all i's,

$$\sqrt{L}\big(\hat{G}_i - G_i, \hat{G}_i^0 - G_i^0\big) \rightsquigarrow \big(\mathbb{G}_i, \mathbb{G}_i^0\big), \tag{21}$$

where $\rightsquigarrow$ denotes weak convergence, and $\mathbb{G}_i$ and $\mathbb{G}_i^0$ are (correlated) tight mean-zero Gaussian processes on $[0, \overline{v}]$.[17] The covariance functions of these processes can be computed as

$$\begin{aligned} \mathbb{E}\mathbb{G}_i(v_1)\mathbb{G}_i(v_2) &= G_i(v_1 \wedge v_2) - G_i(v_1)G_i(v_2), \\ \mathbb{E}\mathbb{G}_i^0(v_1)\mathbb{G}_i^0(v_2) &= G_i^0(v_1 \wedge v_2) - G_i^0(v_1)G_i^0(v_2), \quad \text{and} \\ \mathbb{E}\mathbb{G}_i(v_1)\mathbb{G}_i^0(v_2) &= G_i^0(v_1 \wedge v_2) - G_i(v_1)G_i^0(v_2). \end{aligned} \tag{22}$$

Consider any proper sub-interval $[0, \overline{v}_0] \subset [0, \overline{v})$. The functional ψ can be shown to be Hadamard differentiable on the space of bounded, right-continuous, left-limit (cadlag) functions on $[0, \overline{v}_0]$ (with the

[16] See e.g. Chapter 20 of van der Vaart (1998).
[17] See also Lemma H.3 in the Appendix.

derivative given by (20)). The FDM then implies weak convergence of the process $\sqrt{L}(\hat{F}_i(v) - F_i(v))$, to a tight Gaussian process on $[0, \overline{v}_0]$,

$$\begin{aligned}\sqrt{L}(\hat{F}_i(v) - F_i(v)) &\rightsquigarrow \psi'(\mathbb{G}_i, \mathbb{G}_i^0)(v) \\ &= (1 - F_i(v)) \left(\int_0^v \frac{d\mathbb{G}_i^0(u)}{1 - G_i(u)} + \int_0^v \frac{\mathbb{G}_i(u) dG_i^0(u)}{(1 - G_i(u))^2} \right). \end{aligned} \tag{23}$$

The estimator $\hat{F}_i$, together with some other estimators defined later using ψ, will be used as inputs for construction of estimators using the de-censoring formula under collusion in (12). Because in (12) the integral under the exponent extends up to the upper boundary of the support $\overline{v}$, this requires that the input estimators weakly converge on the entire support $[0, \overline{v}]$. However, the main difficulty in obtaining such results is that the denominator $1 - G_i(u)$ in (3) tends to 0 as u approaches $\overline{v}$, and consequently, the functional ψ is not Hadamard differentiable on the space of functions defined on the entire support $[0, \overline{v}]$.

In order to overcome this difficulty, we propose a trimmed version of the estimator. The trimmed estimator is denoted as $\tilde{F}_i(v)$ and is defined as

$$\tilde{F}_i(v) \equiv \hat{F}_i(v \wedge \bar{v}_{i,L}),$$

where $\bar{v}_{i,L} \uparrow \overline{v}$ is the trimming sequence, and the convergence of $\bar{v}_{i,L}$ is in probability. We define $\bar{v}_{i,L}$ through the quantile transformation $\hat{G}_i^{-1}(t_L)$,[18] where $t_L \uparrow 1$:

$$\bar{v}_{i,L} \equiv \hat{G}_i^{-1}(t_L).$$

In other words, we trim values v using a sequence of extreme quantiles of the estimated distribution of bids. Such a trimming scheme is convenient as it does not require estimation of the upper bound of the support of the distribution of valuations. The trimming parameter $\bar{v}_{i,L}$ has to approach the upper bound of the support at a rate faster than $L^{-1/2}$ to avoid an asymptotic bias. At the same time the rate has to

[18] We use the standard definition of quantile transformations: For a CDF H, $H^{-1}(t) = \inf\{v : H(v) \geq t\}$, where $t \in (0, 1)$. In fact, since we considering distributions with compact supports, $(0, 1)$ can be changed to $[0, 1]$.

be sufficiently slow to (uniformly) control the approximation error in the FDM. The assumption below prescribes sufficient bounds on the rate.

Assumption 5. *The trimming parameter t_L satisfies $t_L = 1 - L^{-\beta}$ with $1/2 < \beta < 3/4$.*

We also make the following smoothness assumption.

Assumption 6. *The CDFs F_i's have densities f_i's, which are smooth (belong to C^∞) and bounded away from zero on the support $[0, \bar{v}]$.*

With these assumptions, the result in (23) can be strengthened to hold over the entire support $[0, \overline{v}]$.

Proposition 4 (Weak convergence under competition)**.** *Under Assumptions 1–6, the following weak convergence holds for the trimmed estimators $\tilde{F}_i$ jointly for all i, over the entire support $[0, \overline{v}]$,*

$$\sqrt{L}(\tilde{F}_i - F_i) \rightsquigarrow \psi'(\mathbb{G}_i, \mathbb{G}_i^0).$$

We now turn to estimation of the distribution of bidder valuations under collusion. Our estimation strategy again follows the plug-in approach. It is convenient to define the expression appearing on the right-hand side of collusion de-censoring formula (12) as a functional:

$$\psi_{col}(H_1, H_2)(v) \equiv \exp\left(-\int_v^\infty \frac{dH_1(u)}{H_2(u)}\right).$$

The identification result in Proposition 2 can now be stated as a functional of $F_i^\ell(\cdot)$ and $F_{col}(\cdot)$:

$$F_i(v) = \psi_{col}(F_i^\ell, F_{col})(v),$$

where $F_i^\ell \equiv \psi(G_{col}^{0,\ell}, G_i^\ell)$ and $F_{col} \equiv \psi(G_{col}^0, G_{col})$.

The distributions F_i^ℓ and F_{col} are estimated as follows. First, we estimate the distributions G_i^ℓ and $G_i^{0,\ell}$ as the empirical averages in parallel to (85), however, conditional on the event that i is the leader, $\ell_i = 1$:

$$\hat{G}_i^\ell(b) = \frac{\sum_{l=1}^L \mathbb{1}(b_{il} \le b, \ell_{il} = 1)}{\sum_{l=1}^L \mathbb{1}(\ell_{il} = 1)}, \qquad \hat{G}_i^{0,\ell}(b) = \frac{\sum_{l=1}^L \mathbb{1}(b_{il} \le b, w_{il} = 0, \ell_{il} = 1)}{\sum_{l=1}^L \mathbb{1}(\ell_{il} = 1)}. \tag{24}$$

We similarly estimate the distributions for the maximum bid $b_l^* \equiv \max_{i \in \mathcal{N}_{col}} b_{il}$ in the group of suspects $\mathcal{N}_{col}$:

$$\hat{G}_{col}(b) = \frac{1}{L} \sum_{l=1}^{L} \mathbb{1}(b_l^* \leq b), \qquad \hat{G}_{col}^0(b) = \frac{1}{L} \sum_{l=1}^{L} \mathbb{1}(b_l^* \leq b, w_l = 0). \tag{25}$$

These estimators are then plugged in to obtain the consistent estimators $\hat{F}_{i,L}^{\ell}$ and $\hat{F}_{col}$:

$$\hat{F}_i^{\ell} = \psi(\hat{G}_i^{0,\ell}, \hat{G}_{i,L}^{\ell}), \qquad \hat{F}_{col} = \psi(\hat{G}_{col}^0, \hat{G}_{col}). \tag{26}$$

Using the trimmed estimators

$$\tilde{F}_i^{\ell}(v) \equiv \hat{F}_i^{\ell}(v \wedge \bar{v}_{i,L}), \qquad \tilde{F}_{col} \equiv \hat{F}_{col}(v \wedge \bar{v}_{col,L}), \tag{27}$$

where $\bar{v}_{col,L} \equiv \hat{G}_{col}^{-1}(t_L)$, the estimator of F_i under collusion is defined by the plug-in approach as

$$\tilde{F}_i^{col} = \psi_{col}(\tilde{F}_i^{\ell}, \tilde{F}_{col}). \tag{28}$$

In parallel to the result in Proposition 4, one can show weak convergence on the entire support $[0, \overline{v}]$ of the empirical processes for $\tilde{F}_i^{\ell}$ and $\tilde{F}_{col}$ to tight Gaussian processes, denoted respectively as $\mathbb{F}_i^{\ell}$ and $\mathbb{F}_{col}$:

$$\sqrt{L}(\tilde{F}_i^{\ell} - F_i^{\ell}) \rightsquigarrow \mathbb{F}_i^{\ell} \equiv \psi'(\mathbb{G}_i^{0,\ell}, \mathbb{G}_i^{\ell}), \qquad \sqrt{L}(\tilde{F}_{col} - F_{col}) \rightsquigarrow \mathbb{F}_{col} \equiv \psi'(\mathbb{G}_{col}^0, \mathbb{G}_{col}), \tag{29}$$

where $\left(\mathbb{G}_i^{0,\ell}, \mathbb{G}_i^{\ell}, \mathbb{G}_{col}^0, \mathbb{G}_{col}\right)$ are (correlated) Gaussian processes that arise in the weak convergence of the corresponding estimators:

$$\sqrt{L}(\hat{G}_i^{0,\ell} - G_i^{0,\ell}, \hat{G}_i^{\ell} - G_i^{\ell}, \hat{G}_{col}^0 - G_{col}^0, \hat{G}_{col} - G_{col}) \rightsquigarrow \left(\mathbb{G}_i^{0,\ell}, \mathbb{G}_i^{\ell}, \mathbb{G}_{col}^0, \mathbb{G}_{col}\right), \tag{30}$$

and the weak convergence holds jointly with that in (21) and across i's. The corresponding covariances are defined similarly to those in (22). The functional derivative of ψ_{col}, at $H_1 = F_i^{\ell}$ and $H_2 = F_{col}$, can be computed as

$$\psi'_{col}(h_1, h_2)(v) = F_i(v) \left(-\int_v^{\overline{v}} \frac{dh_1(u)}{F_{col}(u)} + \int_v^{\overline{v}} \frac{h_2(u) dF_i^{\ell}(u)}{F_{col}^2(u)} \right).$$

The following proposition establishes a result analogous to that in Proposition 4, but under collusion.

Proposition 5 (Weak convergence under collusion)**.** *Under Assumptions 1–6, the following weak convergence holds jointly for all i's, over any proper sub-interval $[\underline{v}_0, \overline{v}] \subset (0, \overline{v}]$.*

$$\sqrt{L}(\tilde{F}_i^{col} - F_i) \rightsquigarrow \psi'_{col}(\mathbb{F}_i^{\ell}, \mathbb{F}_{col}),$$

where $\mathbb{F}_i^{\ell}$ and $\mathbb{F}_{col}$ are defined in (29)*.*

Remark 1. The weak convergence in Proposition 5 is over any compact interval that excludes 0, the lower boundary of the support. The reason for this is that $F_{col}(u) \to 0$ as $u \downarrow 0$, which creates a "small denominator" problem: the functional ψ_{col} is not Hadamard differentiable on the space of functions defined on the entire support $[0, \overline{v}]$. However, it is Hadamard differentiable on any sub-interval with a strictly positive lower bound. This is the same difficulty encountered for the estimator $\hat{F}_i$ under competition, which we resolved by trimming the support of valuations from above. We conjecture that a similar trimming approach, now from below, would work here as well, but we do not pursue such an extension.

In finite samples, it is unlikely to observe a cartel leader with a very small valuation. Therefore, the estimator $\hat{F}_i^{col}$ will suffer from a substantial small sample bias for valuations v near zero. Thus, extending Proposition 5 to the lower bound of the support is not practical.

Similarly, one can expect a substantial small sample bias for valuations v near $\bar{v}$: in finite samples, it is unlikely to observe a cartel leader with a very large valuation near the upper boundary of the support losing to the competitive fringe. Hence, for testing purposes, we will focus below on proper sub-intervals $[\underline{v}_0, \overline{v}_0] \subset (0, \overline{v})$.

1.4.1 Econometric Test of Collusion

We begin by testing the null hypothesis that bidder i bids competitively. The null can be stated as $H_{0,i} : G_i(b) = G_i^{pred}(b)$ for all b. The corresponding alternative hypothesis is collusive behavior of bidder i, which can be stated as $H_{1,i} : G_i(b) \geq G_i^{pred}(b)$ with strict inequalities for some b's.

The basis of the test will be the deviation of the actual CDF of bids submitted in the auction $G_i(b)$ from the predicted competitive CDF of i's bids $G_i^{pred}(b)$. Pick a compact proper sub-interval $[\underline{v}_0, \overline{v}_0] \subset (0, \overline{v})$, and consider a maximum deviation statistic

$$\hat{T}_i = \max_{b \in [\underline{v}_0, \overline{v}_0]} \left[\hat{\Delta}_i(b) \right]_+, \tag{31}$$

where

$$\hat{\Delta}_i(b) \equiv \hat{G}_i(b) - \hat{G}_i^{pred}(b)$$

denotes the difference between the estimated distribution of bids of bidder i and the estimated predicted distribution of bids for bidder i under competition, and

$$[x]_+ = \begin{cases} x & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Large values of this statistic will be indicative of collusion.

Using (13) and (15), we can express the predicted (or counterfactual) CDF of bids for *suspect* bidder i under competition as a functional

$$\begin{aligned} G_i^{pred} =& \psi_{i,pred}\big(F_i, \{F_j\}_{j \in \mathcal{N}_{col} \setminus \{i\}}, \{F_j\}_{j \in \mathcal{N}_{com}}\big) \\ \equiv& 1 - (1 - F_i) \left(1 - \prod_{j \in \mathcal{N}_{col} \setminus \{i\}} F_j \prod_{j \in \mathcal{N}_{com}} F_j \right). \end{aligned} \tag{32}$$

The functional $\psi_{i,pred}$ involves only products of CDFs and, consequently, is Hadamard differentiable. We denote its Hadamard derivative by $\psi'_{i,pred}\big(h_i, \{h_j\}_{j \in \mathcal{N}_{col} \setminus \{i\}}, \{h_j\}_{j \in \mathcal{N}_{com}}\big)$. Note that for $j \in \mathcal{N}_{col}$,

$F_j = \psi_{col}(F_j^\ell, F_{col})$. Similarly for $j \in \mathcal{N}_{com}$, $F_j = \psi(G_j^0, G_j)$. Therefore, under the null of competition, a repeated application of the FDM together with Propositions 4 and 5 implies that the difference between the estimated distributions $\hat{G}_i$ and $\hat{G}_i^{pred}$ converges weakly to a mean-zero Gaussian process on $[\underline{v}_0, \overline{v}_0]$:

$$\sqrt{L}\hat{\Delta}_i(b) = \sqrt{L}(\hat{G}_i - \hat{G}_i^{pred}) \rightsquigarrow \mathbb{G}_i - \mathbb{G}_i^{pred},$$

where

$$\mathbb{G}_i^{pred} = \psi'_{i,pred}\left(\psi'_{col}(\mathbb{F}_i^\ell, \mathbb{F}_{col}), \left\{\psi'_{col}(\mathbb{F}_j^\ell, \mathbb{F}_{col})\right\}_{j \in \mathcal{N}_{col} \setminus \{i\}}, \left\{\psi'(\mathbb{G}_j^0, \mathbb{G}_j)\right\}_{j \in \mathcal{N}_{com}}\right). \tag{33}$$

The Continuous Mapping Theorem then implies that under the null of competition, the statistic $\sqrt{L}\hat{T}_i$ also converges weakly:

$$\sqrt{L}\hat{T}_i \rightsquigarrow \max_{b \in [\underline{v}_0, \overline{v}_0]} [\mathbb{G}_i(b) - \mathbb{G}_i^{pred}(b)]_+. \tag{34}$$

At the same time according to Assumption 4, the statistic $\sqrt{L}\hat{T}_i$ is divergent if bidder i participates in the cartel.

In principle, the limiting distribution of $\sqrt{L}\hat{T}_i$ that appears above could be computed through the simulation of the Gaussian processes $\mathbb{G}_i(b)$ and $\mathbb{G}_i^{pred}(b)$. However, since the covariance structure of the limiting process is complicated due to the multi-step nature of our estimator, we propose to approximate the null distribution of our test statistic by the bootstrap.

The bootstrap samples are generated by drawing randomly with replacement L auctions from the original sample of L auctions. Let $\{(b_{1l}^\dagger, \ldots, b_{Nl}^\dagger) : l = 1, \ldots, L\}$ be a bootstrap sample, and M be the number of bootstrap samples. In each bootstrap sample, we construct $\hat{G}_i^\dagger$ and $\hat{G}_i^{0,\dagger}$, which are the bootstrap analogues of $\hat{G}_i$ and $\hat{G}_i^0$ respectively. The bootstrap version of the trimmed estimator $\tilde{F}_i$ is given by

$$\tilde{F}_i^\dagger(v) = \psi(\hat{G}_i^{0,\dagger}, \hat{G}_i^\dagger)(v \wedge \bar{v}_{i,L}^\dagger),$$

where $\bar{v}_{i,L}^{\dagger} \equiv (\hat{G}_i^{\dagger})^{-1}(t_L)$, and the trimming parameter t_L is defined in Assumption 5.

We can similarly define the bootstrap estimators corresponding to the decensoring formula under collusion. Our functional notation allows to define those estimators conveniently as follows. Let $\hat{G}_i^{\ell,\dagger}$, $\hat{G}_i^{0,\ell,\dagger}$, $\hat{G}_{col}^{\dagger}$, and $\hat{G}_{col}^{0,\dagger}$ be the bootstrap analogues of $\hat{G}_i^{\ell}$, $\hat{G}_i^{0,\ell}$, $\hat{G}_{col}$, and $\hat{G}_{col}^{0}$ respectively, see equations (24) and (25). As in equations (26) and (27), we have $\tilde{F}_i^{\ell,\dagger}(v) = \psi(\hat{G}_i^{0,\ell,\dagger}, \hat{G}_i^{\ell,\dagger})(v \wedge \bar{v}_{i,L}^{\dagger})$, and $\tilde{F}_{col}^{\dagger}(v) = \psi(\hat{G}_{col}^{0,\dagger}, \hat{G}_{col}^{\dagger})(v \wedge \bar{v}_{col,L}^{\dagger})$ with $\bar{v}_{col,L}^{\dagger} \equiv (\hat{G}_{col}^{\dagger})^{-1}(t_L)$. Moreover, following equation (28), the bootstrap estimator of the distribution F_i under potential collusion is $\tilde{F}_i^{col,\dagger} = \psi_{col}(\tilde{F}_i^{\ell,\dagger}, \tilde{F}_{col}^{\dagger})$. We can now define the bootstrap analogue of the counterfactual (predicted) distribution of bids of bidder i:

$$\hat{G}_i^{pred,\dagger} = \psi_{i,pred}\big(\tilde{F}_i^{col,\dagger}, \{\tilde{F}_j^{col,\dagger}\}_{j \in \mathcal{N}_{col} \setminus \{i\}}, \{\tilde{F}_j^{\dagger}\}_{j \in \mathcal{N}_{com}}\big).$$

Lastly, we construct the bootstrap analogue of $\hat{T}_i$:

$$\hat{T}_i^{\dagger} = \max_{b \in [\underline{v}_0, \bar{v}_0]} \Big[\hat{\Delta}_i^{\dagger}(b) - \hat{\Delta}_i(b)\Big]_+,$$

where

$$\hat{\Delta}_i^{\dagger}(b) = \hat{G}_i^{\dagger}(b) - \hat{G}_i^{pred,\dagger}(b)$$

is the bootstrap analogue of $\hat{\Delta}_i(b)$.[19]

Let $\{\hat{T}_{i,m}^{\dagger} : m = 1, \dots M\}$ be the collection of the bootstrap test statistics computed in bootstrap samples 1 through M. The critical value $\hat{c}_{i,1-\alpha}$ is the $(1-\alpha)$-th sample quantile of $\{\hat{T}_{i,m}^{\dagger} : m = 1, \dots M\}$, where α is the desired asymptotic significance level. The null hypothesis of competitive behaviour for bidder i is rejected when $\hat{T}_i > \hat{c}_{i,1-\alpha}$.

Our next proposition establishes the validity of the bootstrap procedures.

[19] Note that to ensure a valid bootstrap approximation, we must re-center $\hat{\Delta}_i^{\dagger}(b)$ by $\hat{\Delta}_i(b)$. The re-centering is needed to ensure that the bootstrap version of the test statistic is generated under the null.

Proposition 6. *Under Assumptions 1–6, the following results hold jointly:*

$$\sqrt{L}(\tilde{F}_i^{\dagger} - \hat{F}_i) \rightsquigarrow \psi'(\mathbb{G}_i, \mathbb{G}_i^0), \qquad v \in [0, \overline{v}], \tag{35}$$

$$\sqrt{L}(\tilde{F}_i^{col,\dagger} - \tilde{F}_i^{col}) \rightsquigarrow \psi'_{col}(\mathbb{F}_i^{\ell}, \mathbb{F}_{col}), \qquad v \in [\underline{v}_0, \overline{v}], \tag{36}$$

$$\sqrt{L}(\hat{\Delta}_i^{\dagger} - \hat{\Delta}_i) \rightsquigarrow \mathbb{G}_i - \mathbb{G}_i^{pred}, \qquad b \in [\underline{v}_0, \overline{v}]. \tag{37}$$

Moreover, the results also hold jointly across i's.

Remark 2. The proof of Proposition 6 relies on the strong approximation results for the bootstrap in Chen and Lo (1997). The Gaussian processes $\mathbb{G}_i$, $\mathbb{G}_i^0$, $\mathbb{F}_i^{\ell}$, $\mathbb{F}_{col}$, and $\mathbb{G}_i^{pred}$ in Proposition 6 should be viewed as independent copies of the corresponding processes appearing in Propositions 4, 5, and equation (33).

The validity of the bootstrap test now follows from (37) as an application of the Continuous Mapping Theorem.

Corollary 1. *Under Assumptions 1–6,*

$$\sqrt{L}\hat{T}_i^{\dagger} \rightsquigarrow \max_{b \in [\underline{v}_0, \overline{v}_0]} [\mathbb{G}_i(b) - \mathbb{G}_i^{pred}(b)]_+. \tag{38}$$

Remark 3. Consistency of the bootstrap testing procedure follows from (34) and (38) by Polýa's Theorem, i.e. $\mathbb{P}(\sqrt{L}\hat{T}_i > \hat{c}_{i,1-\alpha}) \to \alpha$ when $H_{0,i} : G_i(b) = G_i^{pred}(b)$ is true.

Our collusion test can be applied bidder by bidder to construct an estimated set of colluders (a cartel set). However, due to the multiple hypothesis nature of this procedure, it is necessary to control the overall probability of falsely implicating a competitive firm. This can be achieved, for example, by using the Holm-Bonferroni sequential testing procedure that we now describe. Let α denote the overall significance level. The procedure is performed by ordering the individual p-values from smallest to largest,

$$p_{(1)} \leq \dots \leq p_{(K)},$$

where K is the number of suspects, i.e. the number of bidders in $\mathcal{N}_{col}$.

Step 1 The firm with the smallest p-value is included in the cartel set if

$$p_{(1)} < \alpha/K,$$

after which one proceeds to Step 2. Otherwise the procedure stops and none of the firms are included in the cartel.

Step 2 The firm with the second-smallest p-value is tested next. It is included in the cartel if

$$p_{(2)} < \alpha/(K-1),$$

after which one proceeds to the next step. Otherwise the procedure stops and none of the firms are included in the cartel. (The first firm that was included is now excluded as there can never be a single-firm cartel.)

Step 3 The firm with the third-lowest p-value is tested and is included in the cartel if

$$p_{(3)} < \alpha/(K-3),$$

after which one proceeds to the next step. Otherwise, the procedure stops with the two-firm cartel (firms 1 and 2).

And so on until termination.

Once the composition of the cartel $\mathcal{C}$ has been estimated, we can investigate the damage caused by collusion. The predicted auction price under competition is distributed as the second-order statistic:

$$G^{pred}(p) \equiv \sum_{j\in\mathcal{N}} \prod_{i\in\mathcal{N}\setminus\{j\}} F_i(p)(1-F_j(p)) + \prod_{i\in\mathcal{N}} F_i(p).$$

This distribution can be estimated by the plug-in approach using the estimates of $F_i(p)$ under competition for $i \in \mathcal{N}\setminus\hat{\mathcal{C}}$, and the estimates under collusion for $i \in \hat{\mathcal{C}}$ under collusion, where $\hat{\mathcal{C}}$ denotes the estimated cartel set.

Remark 4 (Heterogeneity)**.** We have focused on the case where the same object is auctioned. In many applications, auction-specific heterogeneity is important. Following Haile, Hong and Shum (2003), the standard approach in the literature is to control for heterogeneity through a first-step regression,

$$b_{il} = m(x_l; \theta) + \varepsilon_{il},$$

where the error terms ε_{il} are independent of the object characteristics x_l (and are also independent across bidders). This regression can be estimated parametrically as in Haile, Hong, and Shum (2003). Our estimators can be applied to the homogenized bids $\hat{\varepsilon}_{il}$ resulting from this regression, and our bootstrap test of collusion can be similarly performed with the homogenized bids.

1.5 Monte Carlo Experiment

In this section, we investigate the small-sample performance of our individual test in a Monte Carlo experiment. We consider a setting with 3 bidders who draw values independently from the same distribution, specified as log-normal, $\log V_i \sim N(0, 1)$. Bidder 1 is always competitive, while bidders 2 and 3 may collude. We assume that collusion takes the following form: bidders 2 and 3 are aware of the presence of the competitive bidder, and do not compete with each other if the competitive bidder has dropped out. Thus, if the maximal cartel valuation $\max\{V_2, V_3\} > V_1$, the bidding stops at the price equal to the competitive bidder's valuation V_1 even if $\min\{V_2, V_3\} > V_1$ and the price under competition would be V_2. Otherwise, if $\max\{V_2, V_3\} \leq V_1$, then the competitive bidder wins the auction at the price equal to the cartel leader's valuation $\max\{V_2, V_3\}$.

The estimated predicted competitive distribution when the data are generated under *collusion* is reported in Figures 1.1 and 1.2. All figures contain the plots of the estimated actual bid distribution, the true predicted competitive bid distribution, and the estimated predicted competitive bid distributions. For the smaller sample size $L = 100$, both small sample bias and sample variation are clearly present. Still, even though the estimated predicted bid distribution is not too close to the true one, for most values it is below the actual bid distribution (i.e. shifted towards higher bids). This suggests that even in small samples, collusion might be detectable. The situation improves dramatically for the larger sample, $L = 400$ auctions. Indeed, it is remarkable how close the estimated predicted distribution is to the true population distribution. If the data instead are generated under competition, then the three curves are very close to each other for the sample of $L = 400$ auctions; see Figure 1.3.

To evaluate size properties of our testing procedure, we simulated bids data under competition, i.e. for all three bidders their bids are generated as

$$B_i = \min\{V_i, \max_{j \neq i}\{V_j\}\}, \quad i = 1, 2, 3.$$

However, when applying the de-censoring formulas and computing the test statistics in the original and bootstrap samples, we proceeded under the assumption that bidders 2 and 3 were collusive. We expect that in this case there should not be any significant differences between the CDF of bids for a suspected cartel member ($\hat{G}$) and the predicted CDF of bids under competition ($\hat{G}^{pred}$).

For power computations, bids for cartel members (bidders 2 and 3) were generated as described in the beginning of the section:

$$B_i = \min\{V_i, V_1\}, \quad i = 2, 3.$$

In this case, we expect to see the CDF of bids for a suspected cartel member ($\hat{G}$) to be positioned *above* the predicted CDF of bids under competition ($\hat{G}^{pred}$), i.e. our test should reject the null of competitive behaviour for bidders 2 and 3 with high probability.

The results of our Monte Carlo study are summarized in Table 1.1. The table reports average rejection rates for 1,000 Monte Carlo repetitions. To compute bootstrap critical values, we used 1,000 bootstrap samples (at each Monte Carlo replication).

The test is slightly undersized in small samples of 100 auctions. However, in moderate size samples of 400 auctions, the rejection rates under the null of competitive behaviour are very close to the nominal levels. The test also has very good power properties. For example in the case of collusive behaviour for bidders 2 and 3, the 5% test rejects the null with probabilities exceeding 60% in small samples and 98% in moderate samples.[20]

[20] The test was performed for bidder 2.

1.6 Tables

Table 1.1: Average rejection rates of the bootstrap test for collusion for different significance levels and sample sizes (L)

significance level	$L = 100$	$L = 400$	$L = 100$	$L = 400$
	Competition (H_0)		Collusion (H_1)	
0.01	0.009	0.008	0.403	0.934
0.05	0.030	0.043	0.626	0.981
0.10	0.067	0.080	0.732	0.994

1.7 Figures

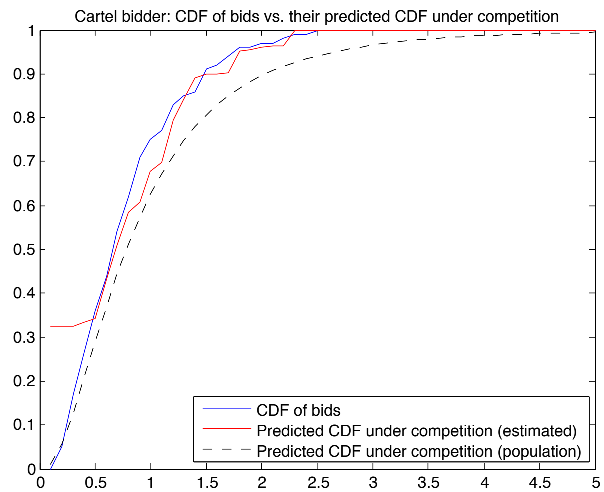

Figure 1.1: Suspect cartel bidder; the data are generated under collusion. The sample size is 100 auctions.

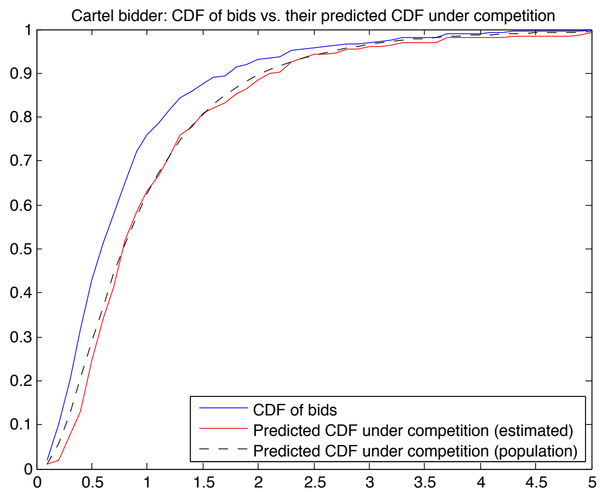

Figure 1.2: Suspect cartel bidder; the data are generated under collusion. The sample size is 400 auctions.

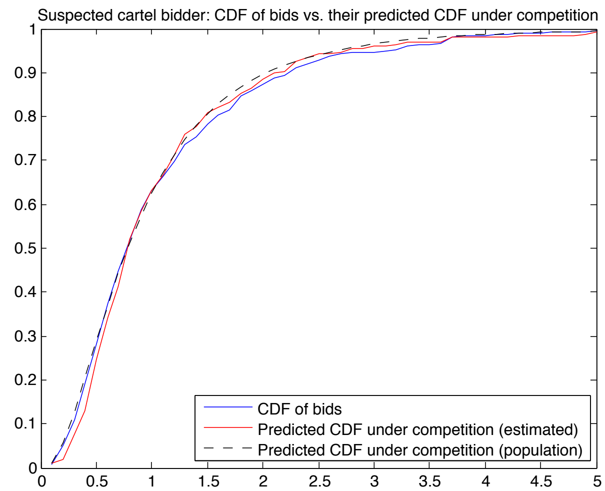

Figure 1.3: Suspect cartel bidder; the data are generated under competition. The sample size is 400 auctions.

1.8 Appendix : Proofs

1.8.1 Extended Functional Delta Method

The following lemma is an extension of the FDM (van der Vaart, 1998, Theorem 20.8) and allows for functionals that depend on the sample size L. This includes functionals with sample-size-dependent trimming.

Lemma H.1 (Extended Functional Delta Method)**.** *Let $\mathbb{D}$ and $\mathbb{E}$ be normed linear spaces. Suppose that:*

(i) $r_L\|\phi_L(F) - \phi(F)\| \to 0$, *where* $r_L \to \infty$ *as* $L \to \infty$, *and* $\phi_L, \phi : \mathbb{D} \to \mathbb{E}$.

(ii) There is a continuous linear map $\phi'_{F,L} : \mathbb{D} \to \mathbb{E}$ *such that, for every compact* $D \in \mathbb{D}_0 \subset \mathbb{D}$,

$$\sup_{h \in D : F + h/r_L \in \mathbb{D}} \left\| \frac{\phi_L(F + h/r_L) - \phi_L(F)}{1/r_L} - \phi'_{F,L}(h) \right\| \to 0.$$

(iii) $\|\phi'_{F,L}(h_L) - \phi'_F(h)\| \to 0$ *for all* h_L *such that* $\|h_L - h\| \to 0$ *with* $h \in \mathbb{D}_0$, *where* $\phi'_F : \mathbb{D}_0 \to \mathbb{E}$ *is a continuous linear map.*

(iv) $\mathbb{G}_L = r_L(F_L - F) \rightsquigarrow \mathbb{G}$, *where* $P(\mathbb{G} \in \mathbb{D}_0) = 1$.

Then, $r_L(\phi_L(F_L) - \phi(F)) \rightsquigarrow \phi'_F(\mathbb{G})$.

Proof. First, $r_L(\phi_L(F_L)-\phi(F)) = r_L(\phi_L(F_L)-\phi_L(F))+r_L(\phi_L(F)-\phi(F))$, where the second term is $o(1)$ by Condition (i) of the Lemma. Next, $r_L(\phi_L(F_L)-\phi_L(F)) = r_L(\phi_L(F+\mathbb{G}_L/r_L)-\phi_L(F)) = (\phi_L(F + \mathbb{G}_L/r_L) - \phi_L(F))/(1/r_L) - \phi'_{F,L}(\mathbb{G}_L) + \phi'_{F,L}(\mathbb{G}_L) = o_p(1) + \phi'_{F,L}(\mathbb{G}_L)$, where the last equality is by (ii), and the $o_p(1)$ term converges in outer probability. The result now follows by (iii), (iv) and the Extended Continuous Mapping Theorem (van der Vaart, 1998, Theorem 18.11(i)).

□

1.8.2 Proofs of the Main Results

For the reasons that will be explained shortly, it will prove convenient to re-state our de-censoring formulas using quantile transformations. For a CDF function $G(\cdot)$, let $G^{-1}(\tau)$ denote its quantile function, $\tau \in (0, 1)$. We introduce the following additional notation. Given a value v, we define

$$t = G_i(v),$$
$$S_i(t) = F_i(G_i^{-1}(t)) \tag{39}$$
$$\implies F_i(v) = S_i(G_i(v)). \tag{40}$$

In addition, we define the following quantile transformation of $G_i^0(v) = P(B_{il} \leq v, w_{il} = 0)$:

$$\mu_i(\tau) = G_i^0(G_i^{-1}(\tau)). \tag{41}$$

Using those definitions, equation (3) implies the following expression for the quantile transformation $S_i(t)$:

$$S_i(t) = 1 - \exp\left(-\int_0^t \frac{d\mu_i(\tau)}{1-\tau}\right). \tag{42}$$

The estimated version of $S_i(t)$ can be stated analogously. With $\hat{G}_i$ and $\hat{G}_i^0$ denoting the estimated versions G_i and G_i^0 respectively, we define $\hat{\mu}_i(\tau) = \hat{G}_i^0(\hat{G}_i^{-1}(\tau))$. We have now

$$\hat{S}_i(t) = 1 - \exp\left(-\int_0^t \frac{d\hat{\mu}_i(\tau)}{1-\tau}\right),$$

where $\hat{S}_i$ is the estimated version of S_i. Thus, our quantile transformation eliminates the random denominator in the integral expression for the estimated CDF. Note that the estimator $\hat{F}_i(v)$ in (19) can be equivalently written via (40), as $\hat{F}_i(v) = \hat{S}_i(\hat{G}_i(v))$. Moreover, one can define the trimmed version of the estimator $\hat{S}_i(t)$, where in view of Assumption 5, the trimming is applied using the sequence t_L:

$$\begin{aligned}\tilde{S}_i(t) &= \hat{S}_i(t \wedge t_L)\\ &= \hat{F}_i(\hat{G}_i^{-1}(t) \wedge \hat{G}_i^{-1}(t_L))\\ &= \hat{F}_i(\hat{G}_i^{-1}(t \wedge t_L))\\ &= 1 - \exp\left(-\int_0^{t\wedge t_L} \frac{d\hat{\mu}_i(\tau)}{1-\tau}\right).\end{aligned}$$

The following notion of continuity plays an important role in the proofs:

Definition 1. *A real-valued function h is α-Hölder continuous, denoted $h \in \mathscr{H}_\alpha$, if there are constants $C > 0$ and $\alpha > 0$ such that $|h(x) - h(y)| \leq C|x-y|^\alpha$ for all x and y in the domain of h.*

The following lemma shows that the derivative of the measure μ_i is α-Hölder continuous with $\alpha = 1/2$.

Lemma H.2. *Suppose that Assumption 1 holds. The function*

$$\mu_i'(t) = \frac{g_i^0(G_i^{-1}(t))}{g_i(G_i^{-1}(t))}$$

is bounded from above and away from zero, continuously differentiable on $[0,1)$, and α-Hölder continuous at $t = 1$ with $\alpha = 1/2$.

Proof of Lemma H.2. It is convenient to write

$$\mu_i'(t) = r_i(G_i^{-1}(t)),$$

where

$$r_i(v) \equiv \frac{g_i^0(v)}{g_i(v)}. \tag{43}$$

We first show that $r_i(\cdot)$ is continuously differentiable on the entire support $[0, \overline{v}]$, including the upper boundary $\overline{v}$. We have

$$r_i(v) = \frac{f_i(v)(1-F_{-i}(v))}{f_i(v)(1-F_{-i}(v)) + f_{-i}(v)(1-F_i(v))}$$
$$= \frac{f_i(v)\frac{1-F_{-i}(v)}{\overline{v}-v}}{f_i(v)\frac{1-F_{-i}(v)}{\overline{v}-v} + f_{-i}(v)\frac{1-F_i(v)}{\overline{v}-v}}$$
$$= \frac{f_i(v)h_{-i}(v)}{f_i(v)h_{-i}(v) + f_{-i}(v)h_i(v)},$$

where we denoted

$$h_i(v) = \frac{1-F_i(v)}{\overline{v}-v}, \qquad h_{-i}(v) = \frac{1-F_{-i}(v)}{\overline{v}-v}.$$

Our assumption that the distributions $F_i(\cdot)$ have densities $f_i(\cdot)$, smooth (C^∞) and bounded away from 0 on the support $[0, \overline{v}]$, implies that $h_i(\cdot)$ and $h_{-i}(\cdot)$ are also smooth and positive on $[0, \overline{v}]$. It follows that $r_i(\cdot)$ is smooth on $[0, \overline{v}]$ (including the upper boundary $\overline{v}$).

Next, we show that $G_i^{-1}(t)$ is Hölder α-continuous with $\alpha = 1/2$. Since

$$1 - G_i(v) = (1 - F_i(v)(1 - F_{-i}(v)) = h_i(v)h_{-i}(v)(\overline{v} - v)^2,$$

it follows that $G_i'(\overline{v}) = 0$ and $G_i''(\overline{v}) = -2h_i(\overline{v})h_{-i}(\overline{v}) < 0$. Using our assumption that the densities $f_i(\cdot)$ are C^∞ on $[0, \overline{v}]$, the Morse Lemma[21] implies that there exists a diffeomorphism $q : [0, \overline{v}] \to [0, 1]$ (a smooth function with a smooth inverse) such that

$$1 - G_i(v) = q(\overline{v} - v)^2.$$

Inverting this relationship yields

$$G_i^{-1}(t) = \overline{v} - q^{-1}(\sqrt{1-t}),$$

which implies that $G_i^{-1}(t)$ is Hölder α-continuous with $\alpha = 1/2$ as a composition of a smooth function and $\sqrt{1-t}$. Finally, $\mu_i'(t) = r_i(G_i^{-1}(t))$ is also Hölder $1/2$-continuous as a composition of a

[21] See Guillemin and Pollack (1974), p. 42.

continuous $r_i(\cdot)$ and Hölder $1/2$-continuous $G_i^{-1}(t)$. □

The population functions F_i, S_i, G_i, G_i^0, and μ_i as well as their estimators can be viewed as elements of the metric space $\mathbb{D}$ of *cadlag* functions equipped with the uniform norm $\|\cdot\|$. Our estimation procedure is driven by $\hat{G}_i$, $\hat{G}_i^0$, and other empirical distributions involving the bids $\{B_{il}\}$. The following lemma presents important properties of those estimators, as well as those of $\hat{\mu}_i$. Let $\rightsquigarrow$ denote the weak convergence.

Lemma H.3. *The following results hold jointly for all i's.*

(a) $\left(\sqrt{L}(\hat{G}_{i,L}-G_i), \sqrt{L}(\hat{G}_{i,L}^0-G_i^0)\right) \rightsquigarrow \left(\mathbb{G}_i, \mathbb{G}_i^0\right)$*, where $\mathbb{G}_i$ and $\mathbb{G}_i^0$ are two correlated Gaussian processes on $[0,\overline{v}]$.*

(b) Under Assumption 1, $\sqrt{L}(\hat{\mu}_{i,L}-\mu_i) \rightsquigarrow \mathbb{M}_i$*, where for $t \in [0,1]$,*

$$\mathbb{M}_i(t) = \mathbb{G}_i^0(G_i^{-1}(t)) - \mathbb{G}_i(G_i^{-1}(t))\mu_i'(t).$$

Furthermore, $P(\mathbb{M}_i(\cdot) \in \mathscr{H}_\alpha) = 1$ for any $\alpha < 1/2$.

(c) Under Assumption 1, there exists a version of the Gaussian process $\mathbb{M}_i$ such that for any $\alpha < 1/2$,

$$\limsup_{L\to\infty} L^{\alpha/2} \left\|\sqrt{L}(\hat{\mu}_{i,L}-\mu_i) - \mathbb{M}_i\right\| < \infty \quad \textit{a.s.}$$

Remark 5.

(1) Part (a) of Lemma H.3 is a standard Functional CLT result for Empirical Processes, see van der Vaart (1998), Theorem 19.5. In fact, the result holds jointly with the weak convergence in (30) for other empirical distributions involving the bids $\{B_{il}\}$.

(2) The first claim in part (b) of the lemma follows from part (a) by the FDM, see van der Vaart (1998), Lemma 20.10 and Lemma 21.3 for quantile functions. Note that Lemma H.2 implies that μ_i' is a bounded function. The α-Hölder continuity result holds by (i) the α-Hölder continuity for of μ_i' with $\alpha = 1/2$ shown in Lemma H.2, and (ii) because the sample paths of $\mathbb{G}_i$ and $\mathbb{G}_i^0$

are α-Hölder continuous with probability one for any $\alpha < 1/2$, see for example Revuz and Yor (1999), Theorem 2.2.

(3) Part (c) uses a point-wise approximation of empirical processes by Gaussian processes, see van der Vaart (1998), page 268, and Hölder continuity of μ_i' in Lemma H.2.

Proof of Lemma H.3. To simplify the notation, we omit bidder's index i in whenever there is no risk of confusion.

To show part (b), for a CDF G, let $q(G) = G^{-1}$ be the quantile transformation. By Lemma 21.3 in van der Vaart (1998), the Hadamard derivative of q (tangentially to the set of continuous functions h), is $q_G'(h) = -h(G^{-1})/g(G^{-1})$, where g is the PDF of G. We have:

$$\begin{aligned} & \frac{1}{\delta_L}\left((G^0 + \delta_L h_L^0)(q(G+\delta_L h_L)) - G^0(q(G))\right) \\ = \quad & h_L^0(q(G+\delta_L h_L)) + \frac{1}{\delta_L}\left(G^0(q(G+\delta_L h_L)) - G^0(q(G))\right) \\ \to \quad & h^0(q(G)) + g^0(q(G))q_G'(h) \\ = \quad & h^0(G^{-1}) - \frac{g^0(G^{-1})}{g(G^{-1})}h(G^{-1}), \end{aligned}$$

where the convergence holds in the uniform norm for all $(h_L^0, h_L) \to (h^0, h)$ as $\delta_L \to 0$ tangentially to the set of continuous functions h. This concludes the proof of the first claim in part (b).

To show the α-Hölder continuity result in (b), write $\mathbb{M}(t+\delta) - \mathbb{M}(t) = \mathbb{G}^0(G^{-1}(t+\delta)) - \mathbb{G}^0(G^{-1}(t)) + \mathbb{G}(G^{-1}(t))(\mu'(t) - \mu'(t+\delta)) - (\mathbb{G}(G^{-1}(t+\delta)) - \mathbb{G}(G^{-1}(t)))\mu'(t+\delta)$. For any $\alpha < 1/2$,

$$|\mathbb{G}^0(G^{-1}(t+\delta)) - \mathbb{G}^0(G^{-1}(t))| \quad \leq \quad C_1|G^{-1}(t+\delta) - G^{-1}(t+\delta)|^\alpha \leq C_1 C_2^\alpha |\delta|^\alpha,$$

where the first inequality follows because $\mathbb{G}^0 \in \mathscr{H}_\alpha$ for any $\alpha < 1/2$ by Theorem 2.2 in Revuz and Yor (1999), and the second inequality holds because G^{-1} is continuously differentiable and, therefore, Lipschitz. By Lemma H.2,

$$|\mathbb{G}(G^{-1}(t))(\mu'(t+\delta) - \mu'(t))| \quad \leq C|\delta|^{1/2} \sup_{v\in[0,\overline{v}]} |\mathbb{G}(v)|.$$

Lastly, for any $\alpha < 1/2$,

$$|\mu'(t+\delta)(\mathbb{G}(G^{-1}(t+\delta)) - \mathbb{G}(G^{-1}(t)))| \leq C|\delta|^{\alpha} \sup_{t\in[0,1]} |\mu'(t)|,$$

where $\sup_{t\in[0,1]} |\mu'(t)| < \infty$ by Lemma H.2.

To show part (c), recall that both $\hat{G}_i$ and $\hat{G}_i^0$ are driven by the same random variable B_{il}. Let $\delta_L = 1/\sqrt{L}$, and $\rho_L = \delta_L(\log L)^2$. By the last result on page 268 in van der Vaart (1998), there are Gaussian processes $\mathbb{G}$ and $\mathbb{G}^0$ such that:

$$\limsup_{L\to\infty} \rho_L^{-1} \left\| \sqrt{L}(\hat{G} - G) - \mathbb{G} \right\| < \infty \quad \text{a.s.,} \tag{44}$$

$$\limsup_{L\to\infty} \rho_L^{-1} \left\| \sqrt{L}(\hat{G}^0 - G^0) - \mathbb{G}^0 \right\| < \infty \quad \text{a.s..} \tag{45}$$

Define $\hat{\mathbb{G}} = \sqrt{L}(\hat{G}_L - G)$, and $\hat{\mathbb{G}}^0 = \sqrt{L}(\hat{G}_L^0 - G^0)$.

$$\begin{aligned}
\sqrt{L}(\hat{\mu} - \mu) &= \sqrt{L}(\hat{G}^0(\hat{G}^{-1}) - G^0(G^{-1})) \\
&= \frac{1}{\delta_L}\Big((G^0 + \delta_L\hat{\mathbb{G}}^0)(q(G + \delta_L\hat{\mathbb{G}}) - G^0(q(G))\Big) \\
&= \hat{\mathbb{G}}^0(q(G + \delta_L\hat{\mathbb{G}})) + g^0(q(G + \delta_L^*\hat{\mathbb{G}}))\frac{1}{\delta_L}(q(G + \delta_L\hat{\mathbb{G}}) - q(G)),
\end{aligned} \tag{46}$$

where $0 \leq \delta_L^* \leq \delta_L$ denotes a generic mean value.

For $0 < \alpha < 1/2$, pick $\epsilon_L = O(\rho_L^{1/\alpha})$. As in the proof of Lemma 21.3 in van der Vaart (1998),

$$(G + \delta_L\hat{\mathbb{G}})(q(G + \delta_L\hat{\mathbb{G}}) - \epsilon_L)) \leq G(q(G)) \leq (G + \delta_L\hat{\mathbb{G}})(q(G + \delta_L\hat{\mathbb{G}})).$$

Moreover,

$$
\begin{aligned}
& \left\|\hat{\mathbb{G}}(q(G+\delta_L\hat{\mathbb{G}})-\epsilon_L))-\hat{\mathbb{G}}(q(G+\delta_L\hat{\mathbb{G}}))\right\| \\
\leq\ & 2\left\|\hat{\mathbb{G}}-\mathbb{G}\right\|+\left\|\mathbb{G}(q(G+\delta_L\hat{\mathbb{G}})-\epsilon_L))-\mathbb{G}(q(G+\delta_L\hat{\mathbb{G}}))\right\| \\
=\ & O_p(\rho_L)+C\left\|q(G+\delta_L\hat{\mathbb{G}})-\epsilon_L)-q(G+\delta_L\hat{\mathbb{G}}))\right\|^{\alpha} \\
=\ & O_p(\rho_L),
\end{aligned}
$$

where the equality in the line before the last holds by the definition of ϵ_L, (44) and α-Hölder continuity of the Gaussian process and because q is Lipschitz. Therefore,

$$
\hat{\mathbb{G}}(q(G+\delta_L\hat{\mathbb{G}}))+O_{a.s.}(\rho_L)\leq \frac{G(q(G))-G(q(G+\delta_L\hat{\mathbb{G}}))}{\delta_L}\leq \hat{\mathbb{G}}(q(G+\delta_L\hat{\mathbb{G}})),
$$

or

$$
\frac{q(G+\delta_L\hat{\mathbb{G}})-q(G)}{\delta_L} = -\frac{\hat{\mathbb{G}}(q(G+\delta_L\hat{\mathbb{G}}))}{g(q(G+\delta_L^*\hat{\mathbb{G}}))}+O_p(\rho_L). \tag{47}
$$

Let $r(\cdot)$ be as in (43). Using (46) and (47), we obtain:

$$
\begin{aligned}
\left\|\sqrt{L}(\hat{\mu}-\mu)-\mathbb{M}\right\| \leq\ & \left\|\hat{\mathbb{G}}^0(q(G+\delta_L\hat{\mathbb{G}}))-\mathbb{G}^0(q(G))\right\| \\
& +\left\|r(q(G+\delta_L^*\hat{\mathbb{G}}))\hat{\mathbb{G}}(q(G+\delta_L\hat{\mathbb{G}}))-r(q(G))\mathbb{G}(q(G))\right\|.
\end{aligned}
$$

The first term on the right-hand side can be bounded by

$$
\left\|\mathbb{G}^0(q(G+\delta_L\hat{\mathbb{G}}))-\mathbb{G}^0(q(G))\right\|+O_p(\rho_L) = O_p(\delta_L^{\alpha}+\rho_L),
$$

for any $\alpha<1/2$, where we used $\|\hat{\mathbb{G}}\|\leq\|\mathbb{G}\|+O_p(\rho_L)$. The second term can be bounded by

$$
\begin{aligned}
& \left\|r(q(G+\delta_L^*\hat{\mathbb{G}}))-r(q(G))\right\|\|\hat{\mathbb{G}}\|+\left\|\hat{\mathbb{G}}(q(G+\delta_L\hat{\mathbb{G}}))-\mathbb{G}(q(G))\right\|\|r\| \\
& = O_p\left(\delta_L^{1/2}+\delta_L^{\alpha}\right).
\end{aligned}
$$

The result in part (c) follows from the last three displays.

□

The following lemma establishes the weak convergence of the trimmed quantile-transformed estimator $\tilde{S}_i(t) = \hat{S}_i(t \wedge t_L)$.

Lemma H.4. *For $t \in [0,1]$, let*

$$\phi(\mu_i)(t) = 1 - \exp\left(-\int_0^t \frac{d\mu_i(\tau)}{1-\tau}\right),$$
$$\phi'(h)(t) = (1 - S_i(t)) \int_0^t \frac{dh(\tau)}{1-\tau},$$

where ϕ' is the functional (Hadamard) derivative of ϕ corresponding to μ_i. Define further $\phi_L(\mu_i)(t) = \phi(\mu_i)(t \wedge t_L)$, $\phi'_L(h)(t) = \phi'(h)(t \wedge t_L)$. Lastly, let

$$\mathbb{D}_0 = \{h \in \mathbb{D}[0,1] : h \in \mathscr{H}_\alpha \text{ for any } \alpha < 1/2, h(0) = 0\}. \tag{48}$$

The following results hold jointly for all i's:

(a) For all sequences h_L such that $\|h_L - h\| = O(\delta_L^\alpha)$ for some $h \in \mathbb{D}_0$ and $0 < \alpha < 1/2$,

$$\left\| \frac{\phi_L(\mu_i + \delta_L h_L) - \phi_L(\mu_i)}{\delta_L} - \phi'_L(h_L) \right\| \to 0, \tag{49}$$

provided that as $\delta_L \to 0$ and $1 - t_L \to 0$,

$$\frac{\delta_L^{1+\alpha}}{1-t_L} = O(1), \quad \frac{\delta_L}{(1-t_L)^{1-\alpha}} = O(1). \tag{50}$$

(b) Under Assumption 1, $\|\tilde{S}_i - S_i\| \to_p 0$ and $\sqrt{L}(\tilde{S}_i - S_i) \rightsquigarrow \phi'(\mathbb{M}_i)$, provided that t_L satisfies the conditions in (50) with $\delta_L = 1/\sqrt{L}$, and $(1-t_L)\sqrt{L} \to 0$.

Remarks.

(1) The modulus of continuity condition for h in the definition of $\mathbb{D}_0$ in (48) can be imposed by part (b) of Lemma H.3.

(2) The result in part (a) of Lemma H.4 is Hadamard differentiability tangentially to $\mathbb{D}_0$ for trimmed functionals with a sample-dependent trimming. In this result, the linearization error is effectively controlled and negligible on the expanding interval $[0, t_L]$. Furthermore, unlike the standard tangential Hadamard differentiability, we require that the sequences h_L converge to elements of $\mathbb{D}_0$ at a sufficiently fast rate, which is justified by the strong approximation rate in Lemma H.3 (c).

(3) The results in parts (b) of Lemma H.4 are the uniform consistency of the trimmed estimator of S_i for its untrimmed population counterpart, and the weak convergence of the trimmed estimator of S_i. Note that, in the weak convergence result, we use the untrimmed population object for re-centering. Similarly, the limiting process involves the untrimmed functional ϕ'. Thus, the trimming has no asymptotic effect on estimation. This is in part due to the condition $\sqrt{L}(1 - t_L) \to 0$, which implies that the trimming parameter t_L must approach 1 at a rate faster than $\sqrt{L}$.

(4) The conditions on the trimming parameter t_L in part (b) ensure that the approximation error in the definition of Hadamard differentiability in (49) is negligible. The rate in the first condition is determined by the approximation of the empirical process by $\mathbb{M}_i$ in Lemma H.3(c). The second rate is driven by the α-Hölder continuity of the limiting process $\mathbb{M}_i$.

(5) All the conditions imposed on t_L in Lemma H.4 can be satisfied, for example, by choosing

$$1 - t_L = L^{-\beta}, \text{ with } 1/2 < \beta < 3/4.$$

as in Assumption 5. With such a choice, $(1 - t_L)\sqrt{L} = L^{-\beta+1/2} \to 0$. The first condition in (50) holds as $L^{-1/2(1+\alpha)+\beta} \to 0$ or $\beta \le (1 + \alpha)/2$, since α can be chosen arbitrarily close to $1/2$. The second condition in (50) implies $\beta \le 1/(2(1 - \alpha)) < 1$, where the last inequality is again due to the fact that α can be chosen arbitrarily close to $1/2$. Hence, the second condition in (50) is non-binding. Thus, the rate of convergence on the trimming parameter is driven mainly by the approximation in Lemma H.3(c).

Proof of Lemma H.4. To simplify the notation, we omit bidder's index i.

For part (a), direct calculations show:

$$\begin{aligned}
&\frac{1}{\delta_L}\left(\phi_L(\mu+\delta_L h_L)(t)-\phi_L(\mu)(t)\right)\\
=\;& \exp\left(-\int_0^{t\wedge t_L}\frac{d\mu(\tau)}{1-\tau}\right)\frac{1}{\delta_L}\left(1-\exp\left(-\delta_L\int_0^{t\wedge t_L}\frac{dh_L(\tau)}{1-\tau}\right)\right)\\
=\;& (1-S(t\wedge t_L))\int_0^{t\wedge t_L}\frac{dh_L(\tau)}{1-\tau}\\
&+0.5\,(1-S(t\wedge t_L))\,\delta_L\left(\int_0^{t\wedge t_L}\frac{dh_L(\tau)}{1-\tau}\right)^2\exp\left(-\delta_L^*\int_0^{t\wedge t_L}\frac{dh_L(\tau)}{1-\tau}\right), \qquad (51)
\end{aligned}$$

where the second equality follows by the mean-value expansion of $1-\exp(-sx)$ around $s=0$, and δ_L^* is the mean-value: $0\le\delta_L^*\le\delta_L$.

Using integration by parts,

$$\begin{aligned}
\int_0^{t\wedge t_L}\frac{dh_L(\tau)}{1-\tau} &= \frac{h_L(t\wedge t_L)}{1-t\wedge t_L}-\int_0^{t\wedge t_L}h_L(\tau)d\left(\frac{1}{1-\tau}\right)\\
&= \frac{h(t\wedge t_L)}{1-t\wedge t_L}-\int_0^{t\wedge t_L}h(\tau)d\left(\frac{1}{1-\tau}\right)+O\left(\frac{\delta_L^\alpha}{1-t\wedge t_L}\right)\\
&= \int_0^{t\wedge t_L}\frac{dh(\tau)}{1-\tau}+O\left(\frac{\delta_L^\alpha}{1-t\wedge t_L}\right), \qquad (52)
\end{aligned}$$

where the big-O term is uniform in t and we used the condition $\|h_L-h\|=O(\delta_L^\alpha)$. Moreover, since $h\in\mathscr{H}_\alpha$ for any $\alpha<1/2$ and $h(1)=0$,

$$\begin{aligned}
\int_0^{t\wedge t_L}\frac{dh(\tau)}{1-\tau} &= -\frac{h(1)-h(t\wedge t_L)}{1-t\wedge t_L}+\int_0^{t\wedge t_L}(1-\tau)^\alpha\frac{h(1)-h(\tau)}{(1-\tau)^\alpha}d\left(\frac{1}{1-\tau}\right)\\
&\quad +h(1)\left(\frac{1}{1-t\wedge t_L}-\int_0^{t\wedge t_L}d\left(\frac{1}{1-\tau}\right)\right)\\
&= O\left(\frac{1}{(1-t\wedge t_L)^{1-\alpha}}\right)+O(1)\int_0^{t\wedge t_L}(1-\tau)^{\alpha-2}d\tau+h(1)\\
&= O\left(1+\frac{1}{(1-t\wedge t_L)^{1-\alpha}}\right), \qquad (53)
\end{aligned}$$

where the $O(1)$ terms are uniform in t. Also, since S is differentiable,

$$\sup_{t\in[0,1]} \left| \frac{1 - S(t \wedge t_L)}{1 - t \wedge t_L} \right| = O(1). \tag{54}$$

By (52), (53), and (54),

$$\begin{aligned} (1 - S(t \wedge t_L))\delta_L \left(\int_0^{t\wedge t_L} \frac{dh_L(\tau)}{1-\tau} \right)^2 &= \delta_L O(1 - t \wedge t_L) O \left(1 + \frac{\delta_L^\alpha}{1 - t \wedge t_L} + \frac{1}{(1 - t \wedge t_L)^{1-\alpha}} \right)^2 \\ &= O \left(\frac{\delta_L^{1/2+\alpha}}{(1 - t \wedge t_L)^{1/2}} + \frac{\delta_L^{1/2}}{(1 - t \wedge t_L)^{1/2-\alpha}} \right)^2, \end{aligned}$$

and, since $1 - t_L \to 0$,

$$\sup_{t\in[0,1]} \left| (1 - S(t \wedge t_L))\delta_L \left(\int_0^{t\wedge t_L} \frac{dh_L(\tau)}{1-\tau} \right)^2 \right| = O \left(\frac{\delta_L^{1/2+\alpha}}{(1 - t_L)^{1/2}} + \frac{\delta_L^{1/2}}{(1 - t_L)^{1/2-\alpha}} \right)^2, \tag{55}$$

where the first term in the O-expression is due to approximation of the empirical process by a Gaussian process, and the second term is due to the α-Hölder continuity of the limiting process. Next, consider the exponential term in (51). By (52) and (53),

$$\sup_{t\in[0,1]} \left| \delta_L \int_0^{t\wedge t_L} \frac{dh_L(\tau)}{1-\tau} \right| = O \left(\delta_L \left(1 + \frac{1}{(1 - t_L)^{1-\alpha}} \right) + \frac{\delta_L^{1+\alpha}}{1 - t_L} \right). \tag{56}$$

Here, the first term in the O-expression is due to α-Hölder continuity of the limiting process, and the second term is due to the approximation of h_L by a Gaussian process. Lastly, by (51), (55), and (56), for h_L's such that $\|h_L - h\| = O(\delta_L^\alpha)$ and $h \in \mathbb{D}_0$,

$$\begin{aligned} &\left\| \frac{1}{\delta_L} \Big(\phi_L(\mu + \delta_L h_L)(t) - \phi_L(\mu)(t) \Big) - (1 - S(t \wedge t_L)) \int_0^{t\wedge t_L} \frac{dh_L(\tau)}{1-\tau} \right\| \\ &= O \left(\frac{\delta_L^{1/2+\alpha}}{(1 - t_L)^{1/2}} + \frac{\delta_L^{1/2}}{(1 - t_L)^{1/2-\alpha}} \right)^2 \exp \left(O \left(\delta_L + \frac{\delta_L}{(1 - t_L)^{1-\alpha}} + \frac{\delta_L^{1+\alpha}}{1 - t_L} \right) \right) \\ &= o(1) \exp\left(O(1)\right), \end{aligned}$$

where the last equality holds by (50).

To show the uniform consistency in part (b), in place of h_L we use $\hat{\mathbb{M}} = \sqrt{L}(\hat{\mu} - \mu)$, which satisfies the conditions imposed on h_L in part (a) of the lemma.

$$\begin{aligned} \left\| \tilde{S}_L - S \right\| &= \left\| \phi_L(\mu + L^{-1/2}\hat{\mathbb{M}}) - \phi(\mu) \right\| \\ &\leq \left\| \phi_L(\mu + L^{-1/2}\hat{\mathbb{M}}) - \phi_L(\mu) \right\| + \| \phi_L(\mu) - \phi(\mu) \| \\ &\leq L^{-1/2} \left\| \phi_L'(\hat{\mathbb{M}}) \right\| + \sup_{t \in [t_L, 1]} (S(t) - S(t_L)), \end{aligned} \tag{57}$$

where the inequality in the last line holds by part (a) of the lemma (for the first term) and because $\phi_L(t) = \phi(t)$ for $t \leq t_L$ (for the second term). Since S is differentiable with a bounded derivative, and because for $t \geq t_L$ we have $t - t_L \leq 1 - t_L$, the second term in (57) is of order

$$\sup_{t \in [t_L, 1]} (S(t) - S(t_L)) = O(1 - t_L) = o(1). \tag{58}$$

Moreover, for h_L that satisfies the conditions from part (a) of the lemma, by (52) and (54) we have

$$\sup_{t \in [0,1]} \left| (1 - S(t \wedge t_L)) \left(\int_0^{t \wedge t_L} \frac{dh_L(\tau)}{1 - \tau} - \int_0^{t \wedge t_L} \frac{dh(\tau)}{1 - \tau} \right) \right| = O\left(\delta_L^\alpha\right). \tag{59}$$

It follows from (53), (54), and (59) that $\phi_L'(\hat{\mathbb{M}})(t)$ in (57) of order

$$\delta_L O(1 - t \wedge t_L) O_p \left(\frac{1}{1 - t \wedge t_L} \right)^{1-\alpha} = o_p(1)$$

uniformly in t, which concludes the proof of the uniform consistency in part (b).

To show the weak convergence result in part (b), we verify the conditions of Lemma H.1 with $r_L = 1/\delta_L = \sqrt{L}$. For condition (i), as in (57) and (58), $\sqrt{L}\|\phi_L(\mu) - \phi(\mu)\| = O(\sqrt{L}(1 - t_L)) = o(1)$, where the second equality is by the conditions imposed on t_L in part (b) of Lemma H.4. Condition (ii) of Lemma H.1 has been established in part (a) of Lemma H.4. Condition (iv) holds by Lemma H.3(b).

To show that condition (iii) of Lemma H.1 holds, first note that $\|\phi_L'(h_L) - \phi_L'(h)\| \to 0$ for $\|h_L - h\| = O(\delta_L^\alpha)$, where the latter condition is satisfied by $\hat{\mathbb{M}}$ with probability approaching one due to

Lemma H.3(c) with $\delta_L = 1/\sqrt{L}$:

$$\begin{aligned}\left\|\phi_L'(h_L) - \phi_L'(h)\right\| &= \sup_{t\in[0,t_L]}\left|(1-S(t\wedge t_L))\int_0^{t\wedge t_L}\frac{d(h_L(\tau)-h(\tau))}{1-\tau}\right| \\ &= O(\delta_L^\alpha),\end{aligned}$$

where the equality in the second line holds by (54). Next, $\phi_L'(h)(t) - \phi'(h)(t) = 0$ for $t \leq t_L$. For $t \geq t_L$,

$$\begin{aligned}\phi_L'(h)(t) - \phi'(h)(t) &= (1-S(t_L))\int_0^{t_L}\frac{dh(\tau)}{1-\tau} - (1-S(t))\int_0^{t}\frac{dh(\tau)}{1-\tau} \\ &= (S(t)-S(t_L))\int_0^{t_L}\frac{dh(\tau)}{1-\tau} - (1-S(t))\int_{t_L}^{t}\frac{dh(\tau)}{1-\tau} \\ &= O\left(1-t_L\right)^\alpha - (1-S(t))\int_{t_L}^{t}\frac{dh(\tau)}{1-\tau},\end{aligned}$$

where the equality in the last line holds by (53) and (58), and the big-O term is uniform in t. For the second term in the last display, consider

$$\sup_{t\in[t_L,1]}\left|(1-S(t))\int_{t_L}^{t}\frac{dh(\tau)}{1-\tau}\right| = \left|(1-S(t_L^*))\int_{t_L}^{t_L^*}\frac{dh(\tau)}{1-\tau}\right|$$

for some t_L^* such that $t_L \leq t_L^* \leq 1$. If $t_L^* < 1$,

$$\begin{aligned}\left|(1-S(t_L^*))\int_{t_L}^{t_L^*}\frac{dh(\tau)}{1-\tau}\right| &\leq \left|(1-S(t_L^*))\int_0^{t_L^*}\frac{dh(\tau)}{1-\tau}\right| + \left|(1-S(t_L^*))\int_0^{t_L}\frac{dh(\tau)}{1-\tau}\right| && (60)\\ &= O\left(1-t_L^*\right)^\alpha + O\left(1-t_L\right)^\alpha. && (61)\end{aligned}$$

If $t_L^* = 1$, take the limit of the expression in (60) as $t_L^* \to 1$ to obtain convergence to zero due to (61), which concludes the proof of part (b). □

We can now state the proof of Proposition 4.

Proof of Proposition 4. Again, to simplify the notation, we omit bidder's index i.

Write $F(v) = \varphi(S,G)(v) \equiv S(G(v))$. The functional φ is Hadamard differentiable, and its

Hadamard derivative is equal to

$$\varphi'_{S,G}(h_S, h_G)(v) = h_S(G(v)) + S'(G(v))h_G(v),$$

where $S'(t)$ denotes the derivative (density) of S at t. Therefore,

$$\begin{aligned}
\sqrt{L}(\tilde{F} - F)(\cdot) &= \sqrt{L}(\varphi(\tilde{S}, \hat{G}) - \varphi(S, G)) \\
&\rightsquigarrow \phi'(\mathbb{M})(G(\cdot)) - S'(G(\cdot))\mathbb{G}(\cdot) \\
&= \phi'(\mathbb{M})(G(\cdot)) - \frac{f(\cdot)\mathbb{G}(\cdot)}{g(\cdot)}, \qquad (62)
\end{aligned}$$

where the result in the second line holds by Lemma H.4(b) and Lemma H.3(a). The result in the last line holds since $S(t) = F(G^{-1}(t))$ and therefore $S'(G(v)) = f(v)/g(v)$ Next,

$$\begin{aligned}
\phi'(\mathbb{M})(G(v)) &= (1 - S(G(v))\int_0^{G(v)} \frac{d\mathbb{M}(\tau)}{1-\tau} \\
&= (1 - F(v))\int_0^v \frac{d\mathbb{M}(G(u))}{1 - G(u)}, \qquad (63)
\end{aligned}$$

where the equality in the second line holds by a change of variable $u = G^{-1}(\tau)$. By the definition of $\mathbb{M}_i$ in Lemma H.3(b),

$$\begin{aligned}
\int_0^v \frac{d\mathbb{M}(G(u))}{1 - G(u)} &= \int_0^v \frac{d\mathbb{G}^0(u)}{1 - G(u)} - \int_0^v \frac{d\left(\mathbb{G}(u)\mu'(G(u))\right)}{1 - G(u)} \\
&= \int_0^v \frac{d\mathbb{G}^0(u)}{1 - G(u)} - \frac{\mathbb{G}(v)\mu'(G(v))}{1 - G(v)} + \int_0^v \frac{\mathbb{G}(u)\mu'(G(u))dG(u)}{(1 - G(u))^2}, \qquad (64)
\end{aligned}$$

where the equality in the second line holds by integration by parts. Since $\mu(t) = G^0(G(u))$, $\mu'(G(u)) = g^0(u)/g(u)$ and therefore,

$$\mu'(G(u))dG(u) = dG^0(u). \qquad (65)$$

Lastly, by our basic decensoring formula (2),

$$\frac{\mu'(G(v))}{1 - G(v)} = \frac{g^0(v)}{g(v)(1 - G(v))} = \frac{f(v)}{(1 - F(v))g(v)}. \qquad (66)$$

The result of the proposition now follows from (62)–(66). □

Proof of Proposition 6. We omit bidder's index i when there is no risk of confusion.

We show (35) first. Following the definition of μ in (41), we define

$$\hat{\mu}^{\dagger}(t) = \hat{G}^{0,\dagger}((\hat{G}^{\dagger})^{-1}(t)).$$

Following the definition of S in (39) and (42), we also define

$$\hat{S}^{\dagger}(t) = \hat{F}^{\dagger}((\hat{G}^{\dagger})^{-1}(t)) = 1 - \exp\left(-\int_0^t \frac{d\hat{\mu}^{\dagger}(\tau)}{1-\tau}\right),$$

and a trimmed bootstrap estimator

$$\tilde{S}^{\dagger}(t) = \hat{S}^{\dagger}(t \wedge t_L) = 1 - \exp\left(-\int_0^{t\wedge t_L} \frac{d\hat{\mu}^{\dagger}(\tau)}{1-\tau}\right).$$

By adapting the proof of Lemma 21.3 in van der Vaart (1998) and as in the proof of Lemma H.3(b), we can write

$$\begin{aligned}
\sqrt{L}(\hat{\mu}-\mu) &= \sqrt{L}\Big(\hat{G}^0(G^{-1}) - G^0(G^{-1})\Big) - \frac{g^0(G^{-1})}{g(G^{-1})}\sqrt{L}\Big(\hat{G}(G^{-1}) - \tau\Big) \\
&\quad + o_p\Big(\sqrt{L}\Big(\hat{G}^0(G^{-1}) - G^0(G^{-1})\Big) + \sqrt{L}\Big(\hat{G}(G^{-1}) - \tau\Big)\Big), \qquad (67) \\
\sqrt{L}(\hat{\mu}^{\dagger}-\mu) &= \sqrt{L}\Big(\hat{G}^{0,\dagger}(G^{-1}) - G^0(G^{-1})\Big) - \frac{g^0(G^{-1})}{g(G^{-1})}\sqrt{L}\Big(\hat{G}^{\dagger}(G^{-1}) - \tau\Big) \\
&\quad + o_p\Big(\sqrt{L}\Big(\hat{G}^{0,\dagger}(G^{-1}) - G^0(G^{-1})\Big) + \sqrt{L}\Big(\hat{G}^{\dagger}(G^{-1}) - \tau\Big)\Big),
\end{aligned}$$

where the o_p term is uniform in τ, and therefore,

$$\begin{aligned}
\sqrt{L}(\hat{\mu}^{\dagger}-\hat{\mu}) &= \sqrt{L}\Big(\hat{G}^{0,\dagger}(G^{-1}) - \hat{G}^0(G^{-1})\Big) - \frac{g^0(G^{-1})}{g(G^{-1})}\sqrt{L}\Big(\hat{G}^{\dagger}(G^{-1}) - \hat{G}(G^{-1})\Big) \\
&\quad + o_p\Big(\sqrt{L}\Big(\hat{G}^{0,\dagger}(G^{-1}) - \hat{G}^0(G^{-1})\Big) + \sqrt{L}\Big(\hat{G}^{\dagger}(G^{-1}) - \hat{G}(G^{-1})\Big)\Big). \qquad (68)
\end{aligned}$$

Let $\tilde{G}$ and $\tilde{G}^0$ denote estimators constructed using independent copies of the original data. By Proposition 3.1 in Chen and Lo (1997),

$$\begin{aligned}\|\hat{G}^{\dagger} - \hat{G} - \tilde{G} + G\| &= O_{a.s.}(L^{-3/4}(\log L)^{3/4}), \\ \|\hat{G}^{0,\dagger} - \hat{G}^0 - \tilde{G}^0 + G^0\| &= O_{a.s.}(L^{-3/4}(\log L)^{5/4}).\end{aligned}$$

Let $\tilde{\mu} = \tilde{G}^0(\tilde{G}^{-1})$, and note that $\tilde{\mu}$ is an independent copy of $\hat{\mu}$. By taking the difference between (68) and the same expansion as in (67) applied to $\tilde{\mu}$, and applying the result of Chen and Lo (1997), we obtain that

$$\sqrt{L}\|\hat{\mu}^{\dagger} - \hat{\mu} - \tilde{\mu} + \mu\| = O_{a.s.}(L^{-1/4}(\log L)^{5/4}). \tag{69}$$

Let $h_L^{\dagger} = \sqrt{L}(\hat{\mu}^{\dagger} - \hat{\mu})$. As in equation (51) in the proof of Lemma H.4(a),

$$\begin{aligned}&\sqrt{L}(\tilde{S}^{\dagger}(t) - \tilde{S}(t)) \\ = \ &(1 - \tilde{S}(t)) \int_0^{t\wedge t_L} \frac{dh_L^{\dagger}(\tau)}{1-\tau} \\ &+0.5\left(1 - \tilde{S}(t)\right)\delta_L\left(\int_0^{t\wedge t_L} \frac{dh_L^{\dagger}(\tau)}{1-\tau}\right)^2 \exp\left(-\delta_L^* \int_0^{t\wedge t_L} \frac{dh_L^{\dagger}(\tau)}{1-\tau}\right).\end{aligned} \tag{70}$$

Next, let $\tilde{h}_L = \sqrt{L}(\tilde{\mu} - \mu)$ and $\epsilon_L = h_L^{\dagger} - \tilde{h}_L$.We have:

$$\begin{aligned}\int_0^{t\wedge t_L} \frac{dh_L^{\dagger}(\tau)}{1-\tau} &= \frac{h_L^{\dagger}(t\wedge t_L)}{1 - t\wedge t_L} - \int_0^{t\wedge t_L} \frac{h_L^{\dagger}(\tau)d\tau}{(1-\tau)^2} \\ &= \int_0^{t\wedge t_L} \frac{d\tilde{h}_L(\tau)}{1-\tau} + \frac{\epsilon_L(t\wedge t_L)}{1 - t\wedge t_L} - \int_0^{t\wedge t_L} \frac{\epsilon_L(\tau)d\tau}{(1-\tau)^2} \\ &= \int_0^{t\wedge t_L} \frac{d\tilde{h}_L(\tau)}{1-\tau} + O_{a.s.}\left(\frac{(\log L)^{5/4}}{L^{1/4}(1 - t\wedge t_L)}\right),\end{aligned} \tag{71}$$

where the equality in the last line is due to the definition of ϵ_L and by (69), and the $O_{a.s.}$ term is uniform in t.

Since $\sqrt{L}(\tilde{S} - S) \rightsquigarrow \phi'(\mathbb{M})$ by Lemma H.4(b), ϕ' is linear, $\mathbb{M}$ is Gaussian and α-Hölder-continuous for $\alpha < 1/2$, and $\mathbb{M}(1) = 0$, it follows that $\sqrt{L}(\tilde{S}(t) - S(t))/(1 - t\wedge t_L)^{\alpha} = O_p(1)$ uniformly in t

for $\alpha < 1/2$, and

$$\begin{aligned} 1-\tilde{S}(t) &= (1-S(t\wedge t_L))\left(1-\frac{\sqrt{L}(\tilde{S}(t)-S(t\wedge t_L))}{\sqrt{L}(1-S(t\wedge t_L))}\right) \\ &= (1-S(t\wedge t_L))\left(1+O_p\left(\frac{1}{\sqrt{L}(1-t\wedge t_L)^{1-\alpha}}\right)\right) \\ &= (1-S(t\wedge t_L))(1+o_p(1)). \end{aligned}$$

The equality in the last line holds by $1-t_L = L^{-\beta}$ with $\beta < 3/4$ and since α can be chosen arbitrarily close to 1/2; moreover the o_p term is uniform in t. Hence, by (71),

$$\begin{aligned} (1-\tilde{S}(\cdot))\int_0^{\cdot\wedge t_L}\frac{dh_L^{\dagger}(\tau)}{1-\tau} &= (1+o_p(1))(1-S(\cdot\wedge t_L))\int_0^{\cdot\wedge t_L}\frac{d\tilde{h}_L(\tau)}{1-\tau}+O_p\left(L^{-1/4}(\log L)^{5/4}\right) \\ &\rightsquigarrow \phi'(\mathbb{M}^{\dagger}(\cdot)), \end{aligned} \tag{72}$$

where $\mathbb{M}^{\dagger}$ is an independent copy of $\mathbb{M}$ since $\tilde{\mu}$ is an independent copy of $\hat{\mu}$.

Similarly to (55) in the proof of Lemma H.4(b), since $\delta_L = 1/\sqrt{L}$, and by (71),

$$\begin{aligned} &\sup_{t\in[0,1]}\left|(1-S(t\wedge t_L))\delta_L\left(\int_0^{t\wedge t_L}\frac{dh_L^{\dagger}(\tau)}{1-\tau}\right)^2\right| \\ &= O_p\left(\frac{\delta_L^{1/2+\alpha}}{(1-t_L)^{1/2}}+\frac{\delta_L^{1/2}}{(1-t_L)^{1/2-\alpha}}+\frac{(\log L)^{5/4}}{L^{1/2}(1-t_L)^{1/2}}\right)^2 \\ &= o_p(1). \end{aligned} \tag{73}$$

Similarly to (56) in the proof of Lemma H.4(b) and by (71),

$$\begin{aligned} \sup_{t\in[0,1]}\left|\delta_L\int_0^{t\wedge t_L}\frac{dh_L^{\dagger}(\tau)}{1-\tau}\right| &= O_p\left(\frac{\delta_L}{(1-t_L)^{1-\alpha}}+\frac{\delta_L^{1+\alpha}}{1-t_L}+\frac{(\log L)^{5/4}}{L^{3/4}(1-t_L)}\right) \\ &= o_p(1), \end{aligned} \tag{74}$$

where the equality in the last line holds since $1-t_L = L^{-\beta}$ with $\beta < 3/4$.

By (70), (72), (73), and (74) we have that

$$\sqrt{L}(\tilde{S}^{\dagger}(t) - \tilde{S}(t)) \rightsquigarrow \phi'(\mathbb{M}^{\dagger}(\cdot)).$$

The result in (35) now follows by the FDM for the bootstrap (van der Vaart, 1998, Theorem 23.5) and the same arguments as in the proof of Proposition 4, since $\tilde{F}^{\dagger} = \tilde{S}^{\dagger}(\hat{G}^{\dagger})$.

The result in (36) holds by the bootstrap FDM, Proposition 3.1 in Chen and Lo, (29), and since the functional ψ_{col} is Hadamard differentiable on $[\underline{v}_0, \overline{v}] \subset (0, \overline{v}]$.

To show (37), write

$$\sqrt{L}(\hat{\Delta}_i^{\dagger}(b) - \hat{\Delta}_i(b)) = \sqrt{L}(\hat{G}_i^{\dagger}(b) - \hat{G}_i(b)) - \sqrt{L}(\hat{G}_i^{pred,\dagger}(b) - \hat{G}_i^{pred}(b)).$$

The result in (37) follows by the bootstrap FDM and the previous results of the proposition as the functional $\psi_{i,pred}$ defined in (32) is Hadamard differentiable. $\square$

Chapter 2

Copula-based Approach to Identification and Estimation in English Auctions

2.1 Introduction

In this chapter we propose a copula-based approach to establish identification and estimation of model primitives within English auctions under the absence of independence. Rising interest in copula-based methods for empirical research has given impetus to an increasing amount of literature focused on the subject. Since then, it has been extensively used to recover and estimate linear and non-linear association between the variables of interest since it can parametrize the link between the variables in a joint distribution. Hence, we develop a simple approach to test for correlation among bids. Furthermore, we succeed in showing that joint distribution function of private valuations is identifiable under certain conditions. We will look at model specification and estimation in this setup. For estimation, we will use a two-stage approach, first estimating the marginals, then estimating the copula function. we assume that the copula is parametric, but the marginal distributions are nonparametric. Finally, we propose a semiparametric strategy, based on Archimedean copulas, to identify and estimate the model primitives and analyze the correlation between bids in English auctions. One advantage that this approach has is that it allows us to separate the estimation of the marginal distribution from the estimation of the joint distribution of underlying bidder values.

2.2 Literature Review

In his seminal paper, Sklar (1959) proved that there exists a mapping between the joint distribution and its marginals. One advantage of this approach is that there is no need of imposing any distributional assumptions on the marginal distribution, allowing us for flexibility to model various scenarios where a different distribution assumption is needed for each marginal in question. Moreover, it allows us to analyze the asymptotic properties of the dependence structure of the variables. Nelsen (2006) and Joe (1997) present a comprehensive introduction and overview of copula theory. Most of the literature on copulas focuses on the parametric approach. Brendstrup and Paarsch (n.d.) deploy a semiparametric approach to establish identification and estimation of in a multi-object English auction. in Hubbard, Li, and Paarsch (2012), the authors study first-price sealed bid auctions. There is less amount of literature focused on the nonparametric methodology, such as Sancetta and Satchell (2004), Fermanian and Scaillet (2003), Genest and Rivest (1993).

There are numerous parametric families of bivariate copulas that have been used in a variety of applications, refer to Joe (1997), Nelsen (2006) for an overview. There is less amount of research done in multidimensional copulas, including those for constructing higher-dimensional Archimedean copulas (refer to Hofert and Scherer (2011)). Other studies include Oh and Patton (2012), where the authors propose a new class of factor copulas and show that they have some desirable features in high-dimensional applications. The Fréchet problem in higher dimensions has also been studied in the recent years - refer to Embrechts (2009) for an overview.

In this paper, we will address the identification of English auctions, with asymmetric bidders and complete set of bid information, attempting to model the correlation between bids within auctions. More specifically, we show the identification of the true copula generator function of the joint distribution of private values nonparametrically in the Archimedean class. Statistical uses of Archimedean copulas were studied in the works of Genest and MacKay (1986), Marshall and Olkin (1988), Oakes (1989).

This chapter draws partially on Fan and Liu (2013), which adopted the methodology used in Braekers and Veraverbeke (2005). In Fan and Liu (2013), the authors provide identification results for a linear quantile regression model with dependent censoring, providing the identified set and inference procedure for the quantile regression coefficient where Archimedean copula were deployed to capture the dependent censoring. Their identified set for the quantile regression coefficient is obtained by allowing the copula function to vary within a parametric Archimedean family.

In the classical competing risks literature, Braekers and Veraverbeke (2005) demonstrate that if the copula function is Archimedean with a known generator function, then the marginal distribution of each potential risk can be identified from data on the failure time and cause of the failure. We will be using a well known result in the competing risks literature, initially pioneered by Cox (1959) for the bivariate case with independent risk, and subsequently extended by Tsiatis (1975). Hence, we will deploy the proposed methodology in the auction paradigm, with an empirical application to data from municipal GIC auctions in the US. Hence, we contribute to the structural auction literature by establishing novel identification results using results from the competing risks literature.

In the next section, we will review the basics of copula functions, focusing our attention on the scenario where the joint distribution of bidders' private values is characterized by the family of Archimedean copulas, and subsequently propose a model and methodology to recover the dependence structure of bids in English auctions.

2.3 Copula Theory

Copula functions are used to recover the link or association between variables, both linear and non-linear. According to the seminal work by Sklar (1959), any multidimensional distribution may be represented by a superposition of marginal distributions and a copula function. Thus, the joint distribution of valuations can be decomposed into its constituent marginal distributions and the structure of dependence between valuations (which we assume is given by a copula). This copula is a multivariate distribution function with uniform marginals on [0,1], which would allow us to separate the estimation of the marginal distributions from the estimation of the joint distribution.

Consider a random vector of two random variables (X, Y). Denote a bivariate distribution function $H(x, y)$, with marginal distribution functions $F(x)$ and $G(y)$. According to Sklar's theorem, there exists a copula C :$[0, 1]^2 \to [0, 1]$ such that

$$H(x, y) = C(F(x), G(y)), \forall (x, y) \in \mathbb{R}^2.$$

Moreover, if we assume continuity for G and F, then the copula C turns out to be unique. Otherwise, the copula is uniquely determined only on range of G and F. In general, every continuous distribution on $\mathbb{R}^d$ with marginals $F_1(x_1), ..., F_d(x_d)$ can be represented by a d-dimensional copula,

$$H(x_1, ..., x_d) = C(F_1(x_1), ..., F_d(x_d)).$$

This fact can be seen by letting $C(u_1, ..., u_n) = H(F_1^{-1}(u_1), ..., F_d^{-1}(u_d))$. In order to identify the marginal distributions, we specify a copula family for the underlying joint distribution of valuations. We will assume that the copula belongs to the Archimedean class, which allows for non-linear dependence.[1]

Definition 2. *: A copula C is said to be Archimedean if a generator $\varphi : (0, 1] \to (\infty, 0)$, with* $\lim_{t\to 0} \varphi(t) = +\infty$*, exists such that*

$$C(u_d) = \varphi^{-1}(\varphi(u_1) + \varphi(u_2) + ... + \varphi(u_d)), u_d \in I^d.$$

[1] Among the most popular Archimedean copulas are Frank, Clayton and Ali-Mikhail-Haq.

Note that our assumption

$$\lim_{t \to 0} \varphi(t) = +\infty$$

ensures the existence of the inverse φ^{-1} on the domain $[0, +\infty)$.[2] The function φ: $[0,1] \to [0,\infty]$ is called the *generator* of the copula C. It is a continuous, convex, strictly decreasing function with $\varphi(1) = 0$.

2.4 Model

We consider a standard independent private value setting where there are N bidders participating at an auction for a single object, where $N \geq 2$. The set of bidders is denoted by $\mathcal{N} = (1, ..., N)$. Bidders' valuations of the object are private information. Each bidder $i \in \mathcal{N}$ draws its valuation V_i independently from a cumulative distribution $F_i(\cdot)$, with a probability density function denoted by $f_i(\cdot)$, on the support $[0, \overline{v}]^N$. For any bidder i, let V_{-i} denote the maximum value of its rivals, $V_{-i} = \max_{j \neq i} V_j$.
In an ascending-bid auction, the dominant strategy for all bidders is to bid truthfully in the auction. Hence, if the bidder loses in the auction, the losing bid would reveal the true value. However, should the bidder win in the auction, the winning bid will only reveal the lower bound of their actual private value. Therefore, the winner's private value is not directly observed, which is equivalent to not observing the highest bid in second price auctions.

Furthermore, we will assume that bidders know their own value distribution, but they don't have information about the value distribution of other bidders in the auction. Each player makes a bid based on its beliefs about the types of other players. Each player is only given information about the status of the bid, whether he is losing or winning; hence, each bidder can update these beliefs based on the actions taken by the other players. The equilibrium would be reached when the bidders no longer have an incentive to change their bid given their information about others.

Let bidders values V_1 and V_2, be continuous random variables. Assume that the random vector (V_1, V_2) has the joint cumulative distribution function denoted by $H(v_1, v_2)$, with marginals $F_1(v_1)$

[2] Thus, the generator function φ is assumed to be *strict*; see e.g. the discussion in Nelsen's (2006) book.

and $F_2(v_2)$, with corresponding probability distribution functions denoted by $h(v_1, v_2)$ for the joint distribution, and $f_1(v_1)$ and $f_2(v_2)$ for the marginals, respectively. We assume that each V_i is continuously distributed on $[v, v]^n$. Our objective is to recover the underlying latent valuations using the observable bid data.

2.4.1 Copulas in the N-bidder Case

We adopt a copula-based approach to model the joint distributions, assuming that the copula belongs to the Archimedean family.

$$F(V_1, ..., V_N) = \varphi^{-1}\left(\sum_{i=1}^{N} \varphi(F_i(v))\right),$$

where $F_i(v)$ is the (marginal) distribution of i's values. There are additional restrictions on $\varphi(\cdot)$ to ensure that in fact we have a well-behaved joint distribution. Since $\varphi(1) = 0$, we also have the Archimedean copula for the joint cumulative distribution function for a subset of bidders. In particular, if we fix bidder *i*, then the joint distribution for the other bidders is also given by the same Archimedean copula,

$$F_{-i}(v_{-i}) = \varphi^{-1}(\sum_{j \neq i} \varphi(F_j(v)))$$

and the joint distribution of v_i, $Y_i = \max_{j \neq i} V_j$ is given by

$$F_{V_i,Y_i}(v, y) = \varphi^{-1}(\varphi(F_i(v)) + \sum_{j \neq i} \varphi(F_j(y))$$

$$\varphi^{-1}(\sum_{j \neq\ i} \varphi(F_j(y)) = F_{Y_i}(y),$$

so we get the same copula representation

$$F_i(v, y) = \varphi^{-1}(\varphi(F_i(v)) + \varphi(F_{Y_i}(y)))$$

This means that our formula applies intact, simply replacing the other valuation by $Y = \max\{V_j\}$.

2.4.2 The Identification of Marginal Distributions with Known $\varphi(\cdot)$

We use a result from Tsiatis (1975) that links the cumulative distribution function of $Z = \min\{X, Y\}$ and the joint survival function

$$S(t_1, t_2) = \mathbb{P}(X \geq t_1, Y \geq t_2).$$

Let δ_X indicate that $X = \min\{X, Y\}$, and let $H^u(t) = \mathbb{P}(Z \leq t, \delta_x = 1)$.

Lemma D.5 (Tsiatis (1975))**.** We have

$$\frac{dH^u(t)}{dt} = -\frac{\partial S(t_1, t_2)}{\partial t_1}\bigg|_{t_1=t_2=t}. \tag{75}$$

In our case, for a fixed bidder i, we let $X = V_i$ and $Y = Y_i = \max_{j \neq i} V_j$, so that $Z = B_i$ corresponds to the actual bid submitted by bidder i in the auction (winning or losing). We let $G_i(t)$ be the CDF of $Z = \min\{V_i, Y_i\}$, and let $\overline{G}_i(t)$ denote the corresponding survival function. $H^u(t) = G_i^0(t)$ is equal to the distribution of i's bids conditional on losing, times the probability that i loses.

Proposition 7 (Identification of F_i under competition)**.** We have

$$F_i(v) = 1 - \varphi^{-1}\left(-\int_0^v \varphi'(\overline{G_i}(t))dG_i^0(t)\right) \tag{76}$$

Proof. The Archimedean copula representation for the survival function in (75) implies

$$\frac{dG_i^0(t)}{dt} = -\frac{\partial S(t_1, t_2)}{\partial t_1}\bigg|_{t_1=t_2=t} = \frac{\varphi'(\overline{F}_i(t))F_i'(t)}{\varphi'(\overline{G_i}(t))},$$

where $\varphi(\cdot)$ is the copula generator, and bars denote survival functions. Therefore

$$\begin{aligned}
-\varphi'(\overline{G_i}(t))\frac{dG_i^0(t)}{dt} &= \varphi'(\overline{F}_i(t))F_i'(t) \\
&= \frac{d\varphi(\overline{F}_i(t))}{dt}
\end{aligned}$$

$$\begin{aligned}
\int_0^\infty -\varphi'(\overline{G_i}(t))\frac{dG_i^0(t)}{dt}dt &= \varphi(\overline{F}_i(t)) - \varphi\overline{F}_i'(0) \\
&= \varphi(\overline{F}_i(t)) - \varphi(1) \\
&= \varphi(\overline{F}_i(t))
\end{aligned}$$

$$1 - F_i(t) = \varphi^{-1}\left(-\int_0^t \overline{\varphi}'(G_i(t))dG_i^0(t)\right)$$

and we get (76). □

2.4.3 Identification of a Cartel Member Distribution

By extending an argument in $LemmaD.5$, we first show the following result for $Z = \max\{X, Y\}$. Now let

$$F_z^u(t) = \mathbb{P}(Z \le t, \delta_x = 1),$$

where δ_x indicates that $X = \max\{X, Y\}$.

Lemma D.6. We have

$$\frac{\partial F_z^u(t)}{\partial t} = \left.\frac{\partial F_{x,y}(t_1, t_2)}{\partial t_1}\right|_{t_1=t_2=t}$$

Proof.

$$\frac{\partial F_z^u(t)}{\partial t} = \frac{\partial F_{x,y}(t,t)}{\partial t}$$

$$
\begin{aligned}
F_{x,y}(t+\epsilon, t) &= P(X \leq t+\epsilon, Y \leq t) \\
&= P(X \leq t, Y \leq t) \\
&= F_{x,y}(t,t) + P(t \leq z \leq t+\epsilon, Y \leq t)
\end{aligned}
\tag{77}
$$

since if $Y \leq$ and $X \leq t \rightarrow z = \max\{X, Y\} = X$

Also

$$F_{x,y}(t \leq X \leq t+\epsilon, Y \leq t) = F_{z,\delta_x}(t \leq z \leq t+\epsilon, \delta_x = 1)$$

because the events (t $\leq X \leq t+\epsilon, Y \leq$ t) and (t $\leq Z \leq t+\epsilon, \delta_x$ = t) are equivalent.

Moreover, since by definition

$$F_{z,\delta_x} = (t \leq z \leq t+\epsilon, \delta_x = 1) = F_z^u(t+\epsilon) - F_z^u(t)$$

we get after passing to the limit $\epsilon \rightarrow 0$,

$$\frac{\partial F_z^u(t)}{\partial t} = \frac{\partial F_{x,y}(t,t)}{\partial t_1}$$

□

This result can be used to identify the marginal distribution of a cartel member's valuations F_i from (i) F_i^{ℓ}, the distribution of its valuations conditional on being the leader among the suspects, and (ii) the distribution of the maximum of the valuations among the suspects. As these both distributions correspond to the competitive case, they are identifiable by the method developed in the previous section.

Proposition 8. We have

$$F_i(t) = \varphi^{-1}\left(-\int_t^{\infty} \varphi'(F_{col}(t))dF_i^{\ell}(t)\right) \tag{78}$$

Proof. Let $X = V_i$ and $Y = \max_{j \in \mathcal{N}_{col} \setminus \{i\}} V_j$. Then the copula representation implies

$$\varphi(F_{x,y}(t_1, t_2)) = \varphi(F_i(t_1)) + \varphi(F_y(t_2))$$

and

$$\frac{\partial \varphi(F_{x,y}(t_1, t_2))}{\partial t_1} = \frac{d\varphi(F_i(t_1))}{dt_1}.$$

At $t_1 = t_2 = t$, $F_{x,y}(t, t) = F_{col}(t)$ and by Lemma D.6, $\frac{\partial F_{x,y}(t,t)}{\partial t_1} = dF_i^{\ell}(t)/dt$. Therefore

$$\varphi'(F_{col}(t)) \frac{dF_i^{\ell}(t)}{dt} = \frac{d\varphi(F_i(t))}{dt}$$

Then we get by integration,

$$\int_t^{\infty} \varphi'(F_{col}(t)) dF_i^{\ell}(t) = \varphi(1) - \varphi(F_i(t))$$

and (78) follows since $\varphi(1) = 0$. □

2.4.4 Identification of $\varphi(\cdot)$

Invoking theorem 4.1.5(3) in Nelson (2006), we know that if $c \geq 0$, then $c\varphi$ is also a generator of φ. Hence the generator φ is not unique, and we normalize $\varphi'(1) = 1$ to achieve point identification. Consider (V_i, V_j) and (V_i, V_k) under competition

$$Y_{ij} = \min \{V_i, V_j\}$$

$$Y_{ik} = \min \{V_i, V_k\}$$

are observable (as a losing bid). Letting the CDFs of Y_{ij}, Y_{ik} be

$$H_{ij} = P\left\{Y_{ij} \leq t\right\}$$
$$H_{ik} = P\left\{Y_{ik} \leq t\right\}$$

and

$$H_{ij}^{l} = P\left\{V_{ij} \leq t, \delta_{ij} = 0\right\}$$
$$H_{ik}^{l} = P\left\{V_{ik} \leq t, \delta_{ij} = 0\right\},$$

where

$$\delta_{ij} = \begin{cases} 0, & V_i \leq V_j \\ 1, & otherwise \end{cases}$$

(similarly for δ_{ik}). We then have

$$1 - F_i(t) = \varphi^{-1}\left(-\int_0^t \varphi'(\overline{H}_{ij}(t))dH_{ij}^{l}(t)\right) = \varphi^{-1}\left(-\int_0^t \varphi'(\overline{H}_{ik}(t))dH_{ik}^{l}(t)\right)$$
$$\int_0^t \varphi'(\overline{H}_{ik}(t))dH_{ik}^{l}(t) = \int_0^t \varphi'(\overline{H}_{ij}(t))dH_{ij}^{l}(t)$$

By differentiating, we get $\varphi'(\overline{H}_{ik}(t))h_{ij}^{l}(t) = \varphi'(\overline{H}_{ik}(t))h_{ik}^{l}(t)$.

Denote $\overline{H}_{ij}(t) = u$, $\overline{H}_{ij}^{-1}(t) = t$,and denote $\overline{H}_{ij} \circ \overline{H}_{ik} = Q$

$$\varphi'(u)h_{ij}(\overline{H}_{ij}^{-1}(u)) = \varphi'(Q(u))h_{ik}^{l}(\overline{H}_{ij}^{-1}(u))$$

Denote $h^l_{ik}(\overline{H}^{-1}_{ij}(u)) = S_{ik}(u)$ and $h^l_{ik}(\overline{H}^{-1}_{ik}(u)) = S_{ik}(u)$

Hence $\varphi'(u)S_{ij}(u) = \varphi'(Q(u))S_{ik}(u)$

$$Q^{(2)}(u) = Q(Q(u))$$
$$Q^{(k)} = Q(\underbrace{Q(...Q(u))}_{\text{k times}}$$

$$\begin{aligned}
\varphi'(u) &= \varphi'(Q(u))\underbrace{\frac{S_{ik}(u)}{S_{ij}(u)}}_{\lambda(u)} \\
&= \varphi'(Q^{(2)}(u))\lambda(u)\lambda(Q(u)) \\
&= ... \\
&= \varphi'(1)\prod_{k=1}^{\infty}\lambda(Q^{(k)}(u))
\end{aligned} \tag{79}$$

2.4.5 Semiparametric Estimation of Copula under Competition

In this section, we construct a semiparametric estimator of the generator function φ assuming that it belongs to a parametric family of the Archimedean class $\varphi_\theta(u)$, $\theta \in \Theta \subset \mathbb{R}$. At the same time, we continue to treat the marginal distributions nonparametrically. Our approach is based on the identification arguments in Section 2.4.4.

Pick $i \in \mathcal{N}$, and define $\mathcal{N}_{-i} = \mathcal{N}\backslash\{i\}$. With $V_{-i} = \max_{j\neq i}\{V_j\}$, $B_i = \min\{V_i, V_{-i}\}$, $w_i = \mathbb{1}\{V_i > V_{-1}\}$, $G_i(b) = P(B_i \leq b)$, and $G^0_i(b) = P(B_i \leq b, w_i = 0)$, we have by the results in Section

2.4.2 that the distribution of values satisfies

$$F_i(v) = 1 - \varphi_\theta^{-1}\left(-\int_0^v \varphi'_\theta(1 - G_i(b))dG_i^0(b)\right), \tag{80}$$

where $\varphi'_\theta(u) = \partial\varphi_\theta(u)/\partial u$.

Now pick any strict subset $K \subset \mathcal{N}\backslash\{i\}$ of the competitive rivals of bidder i, and let $V_{-iK} = \max_{k\in K} V_k$ be the maximum value in this subset. Let F_{-iK} be the cumulative distribution of V_{-iK}. As in section 2.4.1, the joint distribution of V_i and V_{iK} is given by a copula with the same generator function $\varphi_\theta(\cdot)$:

$$F(v_i, v_{-iK}) = \varphi_\theta^{-1}\Big(\varphi_\theta(F_i(v_i)) + \varphi_\theta(F_{-iK}(v_{-iK}))\Big).$$

Note that the minimum between V_i and V_{-iK} is observable, and denote it as $B_{i,K} = \min\{V_i, V_{-iK}\}$. Define further $G_{i,K}(b) = P(B_{i,K} \leq b)$, $w_{i,K} = 1\{V_i > V_{-iK}\}$, and $G^0_{i,K}(b) = P(B_{i,K} \leq b, w_{i,K} = 0)$. Since the generator function remains the same, we can also write the CDF of values of bidder i as

$$F_i(v) = 1 - \varphi_\theta^{-1}\left(-\int_0^v \varphi'_\theta(1 - G_{i,K}(b))dG^0_{i,K}(b)\right). \tag{81}$$

Since equations (80) and (81) identify the same CDF function $F_i(v)$ and use the same generator φ_θ, it follows that for any rival set K of bidder i, we have

$$\int_0^{\tilde{v}} \varphi'_\theta(1 - G_i(b))dG_i^0(b) = \int_0^{\tilde{v}} \varphi'_\theta(1 - G_{i,K}(b))dG^0_{i,K}(b). \tag{82}$$

Note that the functions G_i,$G_{i,K}$ (and G_i^0,$G^0_{i,K}$) are different. Thus, equation (82) imposes a restriction on φ_θ, and this restriction can be used to identify and estimate θ. In order to use this formula for a consistent estimator of θ, we integrate up to a certain trimming threshold $\tilde{v}$ below the upper boundary of support in order to avoid large bids, as $G_{i,k}(b)$ is close to one for those bids, and $\varphi'(\cdot)$ is very large. So these large bids would otherwise dominate the integrals on both sides, potentially causing inconsistency in the estimation of θ.

For a given θ, the left-hand and right-hand sides of (82) can be estimated by replacing the unknown distributions with their empirical analogues. We allow exogenous participation in the sense that $\mathcal{N}_l$ and K_l are allowed to vary with auction l. Thus we obtain our main estimating equation for θ,

$$\frac{1}{L}\sum_{l=1}^{L}\varphi_\theta'(\hat{H}_i(b_{il}))\mathbb{1}[b_{il}\le b_{-il}, b_{il}<\tilde{v}] = \frac{1}{L}\sum_{l=1}^{L}\varphi_\theta'(\hat{H}_{iK}(b_{il}))\mathbb{1}[b_{il}\le y_l, b_{il}<\tilde{v}], \tag{83}$$

where

$$b_{-il} = \max_{j\in\mathcal{N}_l\setminus\{i\}} b_{jl}, \quad y_l = \max_{k\in K_l} b_{kl},$$
$$\hat{H}_i(u) = \frac{1}{L}\sum_{l=1}^{L}\mathbb{1}(\min\{b_{il}, b_{-il}\}\ge u),$$
$$\hat{H}_{iK}(u) = \frac{1}{L}\sum_{l=1}^{L}\mathbb{1}(\min\{b_{il}, y_l\}\ge u).$$

This estimation equation is stated at the bidder level, so the resulting estimator $\hat{\theta}_i$ will depend on the bidder's identity. In order to obtain a bidder-invariant estimator $\hat{\theta}$, we obtain the average of the estimators θ_i across the bidders, using the empirical fractions of the number of bids submitted by bidder i as the weights.

2.4.6 Estimation of $F_i(v)$ under Competition and Collusion

In this second step, we use the estimator $\hat{\theta}$ obtained in the first step to replace the unknown θ in the generator function. Under competition, we estimate $F_i(v)$, $i\in N_{\text{com}}$ (for the competitive bidders),and $F_{col}(v)$, for the maximum value among the suspects, by replacing the distributions that appear in (80) with their empirical analogues. Hence, we have (80)

$$\hat{F}_i(v) = 1-\varphi_{\hat{\theta}}^{-1}\Big(-\int_0^v \varphi_{\hat{\theta}}'(1-\hat{G}_i(b))d\hat{G}_i^0(b)\Big) \tag{84}$$

with

$$\hat{G}_i(b) = \frac{1}{L}\sum_{l=1}^{L}\mathbb{1}(b_{il}\le b), \qquad \hat{G}_i^0(b) = \frac{1}{L}\sum_{l=1}^{L}\mathbb{1}(b_{il}\le b, w_{il}=0). \tag{85}$$

Now, for a suspect colluder i, we use the identification formula (78) from Proposition 8

$$\hat{F}_i(v) = \varphi_{\hat{\theta}}^{-1}\Big(- \int_v^{\infty} \varphi'_{\hat{\theta}}(\hat{F}_{col}(t)) d\hat{F}_i^l(t)\Big) \tag{86}$$

where the marginal distributions of F_{col} (the maximum valuation among the suspects) and F_i^l (the valuation of $i \in \mathcal{N}_{col}$ conditional on being the leader in the set of suspects) are estimated in parallel to (84):

$$\hat{F}_{col}(v) = 1 - \varphi_{\hat{\theta}}^{-1}\Big(- \int_0^{v} \varphi'_{\hat{\theta}}(1 - \hat{G}_{col}(b)) d\hat{G}^0_{col}(b)\Big),$$
$$\hat{F}_i^l(v) = \varphi_{\hat{\theta}}^{-1}\Big(- \int_0^{v} \varphi'_{\hat{\theta}}(1 - \hat{G}_i^{\ell}(b)) d\hat{G}_i^{0,\ell}(b)\Big)\Big)$$

where

$$\hat{G}_{col}(b) = \frac{1}{L}\sum_{l=1}^{L} \mathbb{1}(b_l^* \leq b), \qquad \hat{G}^0_{col}(b) = \frac{1}{L}\sum_{l=1}^{L} \mathbb{1}(b_l^* \leq b, w_l = 0), \tag{87}$$

and

$$\hat{G}_i^{\ell}(b) = \frac{\sum_{l=1}^{L} \mathbb{1}(b_{il} \leq b, \ell_{il} = 1)}{\sum_{l=1}^{L} \mathbb{1}(\ell_{il} = 1)}, \qquad \hat{G}_i^{0,\ell}(b) = \frac{\sum_{l=1}^{L} \mathbb{1}(b_{il} \leq b, w_{il} = 0, \ell_{il} = 1)}{\sum_{l=1}^{L} \mathbb{1}(\ell_{il} = 1)}. \tag{88}$$

2.4.7 Predicting the Competitive Bid of Bidder *i*

We need the CDF:

$$P(min\{v_{il}, Y_{il} \leq v\})$$

$$P(v_{il} \leq v) + P(Y_{il} \leq v) - P(v_{il} \leq v, Y_{il} \leq v) = F_i(v) + F_{-i}(v) - C(F_i(v), F_{-i})$$

$$= F_i(v) + F_{-i}(v) - \varphi^{-1}(\varphi(F_i) + \varphi(F_{-i})),$$

We can also write it as

$$\varphi(F_i) + \varphi(F_{-i}) = \sum_{j=1}^{N} \varphi(F_j(v)),$$

$$F_i(v) = P(Y_{il} \leq v)$$

where Y_i is the maximum of all other values.

$$F_i(v) = C(F_1(v), ..., F_N(v)) = \varphi^{-1}(\sum_{j \neq i} \varphi(F_j(v))).$$

2.5 Econometric Test of Collusion

As in Chapter 1, we follow a similar procedure to test for exhibition of collusive behaviour. As before, we estimate the latent distribution of bidder valuations from the dataset, and simulate counterfactuals of interest based on the estimated distribution. The null hypothesis, stated as $H_{0,i} : G_i(b) = G_i^{pred}(b)$ for all b, is that bidder i is competitive. The corresponding alternative hypothesis, stated as $H_{1,i} : G_i(b) \geq G_i^{pred}(b)$ with strict inequalities for some b's, is collusive behavior the bidder in question. As before, we obtain the deviation statistic by looking at the the actual CDF of bids submitted in the auction $G_i(b)$ and versus the predicted competitive CDF of i's bids $G_i^{pred}(b)$. Given a compact proper sub-interval $[\underline{v}_0, \overline{v}_0] \subset (0, \overline{v})$, consider a maximum deviation statistic

$$\hat{T}_i = \max_{b \in [\underline{v}_0, \overline{v}_0]} \left[\hat{\Delta}_i(b) \right]_+, \tag{89}$$

where

$$\hat{\Delta}_i(b) \equiv \hat{G}_i(b) - \hat{G}_i^{pred}(b)$$

denotes the difference between the estimated distribution of bids of bidder i and the estimated predicted distribution of bids for bidder i under competition, and

$$[x]_+ = \begin{cases} x & \text{if } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Large values of this statistic will be indicative of collusion.

Using (13) and (15), we can express the predicted (or counterfactual) CDF of bids for *suspect* bidder i under competition as a functional

$$\begin{aligned} G_i^{pred} =& \psi_{i,pred}\big(F_i, \{F_j\}_{j\in\mathcal{N}_{col}\setminus\{i\}}, \{F_j\}_{j\in\mathcal{N}_{com}}\big) \\ \equiv& 1 - (1 - F_i)\left(1 - \prod_{j\in\mathcal{N}_{col}\setminus\{i\}} F_j \prod_{j\in\mathcal{N}_{com}} F_j\right). \end{aligned} \tag{90}$$

The functional $\psi_{i,pred}$ involves only products of CDFs and, consequently, is Hadamard differentiable. We denote its Hadamard derivative by $\psi'_{i,pred}\big(h_i, \{h_j\}_{j\in\mathcal{N}_{col}\setminus\{i\}}, \{h_j\}_{j\in\mathcal{N}_{com}}\big)$. Note that for $j \in \mathcal{N}_{col}$, $F_j = \psi_{col}(F_j^\ell, F_{col})$. Similarly for $j \in \mathcal{N}_{com}$, $F_j = \psi(G_j^0, G_j)$. Therefore, under the null of competition, a repeated application of the FDM together with Propositions 4 and 5 implies that the difference between the estimated distributions $\hat{G}_i$ and $\hat{G}_i^{pred}$ converges weakly to a mean-zero Gaussian process on $[\underline{v}_0, \overline{v}_0]$:

$$\sqrt{L}\hat{\Delta}_i(b) = \sqrt{L}(\hat{G}_i - \hat{G}_i^{pred}) \rightsquigarrow \mathbb{G}_i - \mathbb{G}_i^{pred},$$

where

$$\mathbb{G}_i^{pred} = \psi'_{i,pred}\left(\psi'_{col}(\mathbb{F}_i^\ell, \mathbb{F}_{col}), \big\{\psi'_{col}(\mathbb{F}_j^\ell, \mathbb{F}_{col})\big\}_{j\in\mathcal{N}_{col}\setminus\{i\}}, \big\{\psi'(\mathbb{G}_j^0, \mathbb{G}_j)\big\}_{j\in\mathcal{N}_{com}}\right). \tag{91}$$

The Continuous Mapping Theorem then implies that under the null of competition, the statistic $\sqrt{L}\hat{T}_i$

also converges weakly:

$$\sqrt{L}\hat{T}_i \rightsquigarrow \max_{b\in[\underline{v}_0,\overline{v}_0]}[\mathbb{G}_i(b) - \mathbb{G}_i^{pred}(b)]_+. \tag{92}$$

At the same time according to Assumption 4, the statistic $\sqrt{L}\hat{T}_i$ is divergent if bidder i participates in the cartel.

In principle, the limiting distribution of $\sqrt{L}\hat{T}_i$ that appears above could be computed through the simulation of the Gaussian processes $\mathbb{G}_i(b)$ and $\mathbb{G}_i^{pred}(b)$. However, since the covariance structure of the limiting process is complicated due to the multi-step nature of our estimator, we propose to approximate the null distribution of our test statistic by the bootstrap.

We generate the bootstrap samples by drawing randomly with replacement L auctions from the original sample of L auctions. Let $\{(b_{1l}^\dagger, \ldots, b_{Nl}^\dagger) : l = 1, \ldots, L\}$ be a bootstrap sample, and M be the number of bootstrap samples. In each bootstrap sample, we construct $\hat{G}_i^\dagger$ and $\hat{G}_i^{0,\dagger}$, which are the bootstrap analogues of $\hat{G}_i$ and $\hat{G}_i^0$ respectively. The bootstrap version of the trimmed estimator $\tilde{F}_i$ is given by

$$\tilde{F}_i^\dagger(v) = \psi(\hat{G}_i^{0,\dagger}, \hat{G}_i^\dagger)(v \wedge \bar{v}_{i,L}^\dagger),$$

where $\bar{v}_{i,L}^\dagger \equiv (\hat{G}_i^\dagger)^{-1}(t_L)$, and the trimming parameter t_L is defined in Assumption 5.

We can similarly define the bootstrap estimators corresponding to the decensoring formula under collusion. Our functional notation allows to define those estimators conveniently as follows. Let $\hat{G}_i^{\ell,\dagger}$, $\hat{G}_i^{0,\ell,\dagger}$, $\hat{G}_{col}^\dagger$, and $\hat{G}_{col}^{0,\dagger}$ be the bootstrap analogues of $\hat{G}_i^\ell$, $\hat{G}_i^{0,\ell}$, $\hat{G}_{col}$, and $\hat{G}_{col}^0$ respectively, see equations (88) and (87). As in equations (26) and (27), we have $\tilde{F}_i^{\ell,\dagger}(v) = \psi(\hat{G}_i^{0,\ell,\dagger}, \hat{G}_i^{\ell,\dagger})(v\wedge\bar{v}_{i,L}^\dagger)$, and $\tilde{F}_{col}^\dagger(v) = \psi(\hat{G}_{col}^{0,\dagger}, \hat{G}_{col}^\dagger)(v \wedge \bar{v}_{col,L}^\dagger)$ with $\bar{v}_{col,L}^\dagger \equiv (\hat{G}_{col}^\dagger)^{-1}(t_L)$. Moreover, following equation (28), the bootstrap estimator of the distribution F_i under potential collusion is $\tilde{F}_i^{col,\dagger} = \psi_{col}(\tilde{F}_i^{\ell,\dagger}, \tilde{F}_{col}^\dagger)$. We can now define the bootstrap analogue of the counterfactual (predicted) distribution of bids of bidder i:

$$\hat{G}_i^{pred,\dagger} = \psi_{i,pred}\big(\tilde{F}_i^{col,\dagger}, \{\tilde{F}_j^{col,\dagger}\}_{j\in\mathcal{N}_{col}\setminus\{i\}}, \{\tilde{F}_j^\dagger\}_{j\in\mathcal{N}_{com}}\big).$$

Lastly, we construct the bootstrap analogue of $\hat{T}_i$:

$$\hat{T}_i^{\dagger} = \max_{b \in [\underline{v}_0, \overline{v}_0]} \left[\hat{\Delta}_i^{\dagger}(b) - \hat{\Delta}_i(b) \right]_+,$$

where

$$\hat{\Delta}_i^{\dagger}(b) = \hat{G}_i^{\dagger}(b) - \hat{G}_i^{pred,\dagger}(b)$$

is the bootstrap analogue of $\hat{\Delta}_i(b)$.[3]

Let $\{\hat{T}_{i,m}^{\dagger} : m = 1, \ldots M\}$ be the collection of the bootstrap test statistics computed in bootstrap samples 1 through M. The critical value $\hat{c}_{i,1-\alpha}$ is the $(1-\alpha)$-th sample quantile of $\{\hat{T}_{i,m}^{\dagger} : m = 1, \ldots M\}$, where α is the desired asymptotic significance level. The null hypothesis of competitive behaviour for bidder i is rejected when $\hat{T}_i > \hat{c}_{i,1-\alpha}$.

We have already established the validity and consistency of the bootstrap procedures in Proposition 6 of the first chapter, which holds here as well.

Our collusion test can be applied bidder by bidder to construct an estimated set of colluders (a cartel set). However, due to the multiple hypothesis nature of this procedure, it is necessary to control the overall probability of falsely implicating a competitive firm. This can be achieved, for example, by using the Holm-Bonferroni sequential testing procedure that we now describe. Let α denote the overall significance level. The procedure is performed by ordering the individual p-values from smallest to largest,

$$p_{(1)} \leq \cdots \leq p_{(K)},$$

where K is the number of suspects, i.e. the number of bidders in $\mathcal{N}_{col}$.

Step 1 The firm with the smallest p-value is included in the cartel set if

$$p_{(1)} < \alpha/K,$$

after which one proceeds to Step 2. Otherwise the procedure stops and none of the firms are

[3]Note that to ensure a valid bootstrap approximation, we must re-center $\hat{\Delta}_i^{\dagger}(b)$ by $\hat{\Delta}_i(b)$. The re-centering is needed to ensure that the bootstrap version of the test statistic is generated under the null.

included in the cartel.

Step 2 The firm with the second-smallest p-value is tested next. It is included in the cartel if

$$p_{(2)} < \alpha/(K-1),$$

after which one proceeds to the next step. Otherwise the procedure stops and none of the firms are included in the cartel. (The first firm that was included is now excluded as there can never be a single-firm cartel.)

Step 3 The firm with the third-lowest p-value is tested and is included in the cartel if

$$p_{(3)} < \alpha/(K-3),$$

after which one proceeds to the next step. Otherwise, the procedure stops with the two-firm cartel (firms 1 and 2).

And so on until termination.

Remark 6 (Heterogeneity)**.** We have focused on the case where the same object is auctioned. In many applications, auction-specific heterogeneity is important. Following Haile et al. (2003), the standard approach in the literature is to control for heterogeneity through a first-step regression,

$$b_{il} = m(x_l;\theta) + \varepsilon_{il},$$

where the error terms ε_{il} are independent of the object characteristics x_l (and are also independent across bidders). This regression can be estimated parametrically as in Haile et al. (2003). Our estimators can be applied to the homogenized bids $\hat{\varepsilon}_{il}$ resulting from this regression, and our bootstrap test of collusion can be similarly performed with the homogenized bids.

Chapter 3

Empirical Application: Internet GIC Auctions

3.1 Introduction

Federal investigations of collusion in the municipal derivatives market have commenced in the early 1990s. However, only after the Internal Revenue Service (IRS) has found evidence of collusion while pursuing other illegal behaviours in the industry, such as "yield-burning" and "black box" deals, a full-fledged investigation of collusion in the municipal bond industry began. This investigation ultimately exposed extensive collusive behaviour in the municipal derivative market. At that time, IRS conducted over twenty investigations, which revealed pervasive collusion in the industry. In December of 2006, Charles Anderson of the IRS stated that regulators "think [they] have evidence of bid rigging". Anderson went on to say that, "[p]eople were winning GICs at below fair market values and there were obviously deliberate losing bids by the losing bidders, thereby allowing the winner to win a sweetheart deal".[1] Since then, there have been numerous alleged complaints and subsequent investigations that the competitive bidding process is rigged as firms colluded to manipulate the bidding process in violation of antitrust laws. Banks and firms allegedly took part in an illegal conspiracy to pay state and local

[1] See an article on the website of bloomberg.com published on December 7, 2006 and available at http://www.bloomberg.com/apps/news?pid=newsarchive&sid=awq77C8cUwZA.

governments below-market rates on GICs purchased with municipal bond proceeds, to illegally obtain excessive profits.

Hence, in this chapter, we propose an empirical application of the methodologies outlined in the previous two chapters. Initially, we will take a look at the general institutional framework of the municipal Guaranteed Investment Contract (GIC) auctions. After obtaining an understanding of the market environment, we will look at the methodology proposed in the first chapter, which constitutes the independent case by assuming that latent valuations are independent. Then, we propose and implement a nonparametric estimation procedure for the distributions of values and a bootstrap test of the null hypothesis of competitive behaviour against the alternative of collusion and apply it to our dataset. Our framework allows for asymmetric bidders, and the test can be performed on individual bidders. While the focus of our research is to provide a structural framework for English auctions in the presence of collusion allegations, it is useful to compare the performance of the estimation strategies when we relax the independence assumption and analyze the underlying dependence structure. Hence, we deploy the copula-based approach proposed in the second chapter within which dependence between valuations is identified (constituting the dependent case), and apply it once again to the dataset. The two approaches developed in the previous two chapters are applied to data from Internet municipal GIC auctions. These two frameworks would provides a benchmark to contrast the performance of a given strategy when dependence is introduced. Based on the results obtained, we do not find evidence of collusion in our dataset. Finally, the application of these approaches allows us to test for collusion, with the possibility of ultimately assessing and quantifying the damages incurred as a result of collusive behavior.

3.2 The Municipal Derivatives Market

Over the past decade, Grant Street Group Inc. (GSG) has successfully provided municipalities with an Internet auction platform. This platform has been used for bond sales, foreclosure sales, and GIC auctions. Our dataset contains GIC auctions conducted over the Internet by GSG. The rules of the auctions involve closed exit (as defined by Milgrom (2004)), in that bidders do not observe exit by other bidders. The design adopted by GSG allows bidders employ electronic bidding agents, which upgrade their bids in small increments up to the maximum value specified by the bidder. The only information disclosed to the bidders at any time is the status of their bid: winning or not winning. The bidders are not provided any information about actions of other bidders participating in the auction. This makes it a dominant strategy for a bidder to bid up to its value under competition, so that the auction conforms closely to the button model.

It is well known that open auctions may be prone to collusion, as bidders may signal their intentions through their behaviour in the auction. This has been documented for example in spectrum auctions, see vivid discussions in Klemperer (2002). Marshall and Marx (2009) have recently argued that by restricting the information flow in the open auction, the seller can inhibit collusion. This can be more easily achieved on the Internet, as the communication protocol could be programmatically enforced. Marshall and Marx (2009) formally show that first-best collusion cannot be achieved at an open auction if the identities of the registrants as well as of the current highest bidder are not disclosed.[2]

The GSG open auction platform is marketed as a transparent mechanism that may help municipalities combat bidder collusion, which had been a pervasive problem in the municipal derivative market. In our empirical application, we employ a new dataset of auctions for municipal guaranteed investment contracts (GICs for short). These contracts arise as a result of municipal bond issuance. Municipalities auction off cash from bond sales to financial institutions, awarding it in whole to the bidder that offers the highest interest rate on the investment.

Governments, states, municipalities and para-governmental organizations regularly issue municipal bonds to fund diverse capital projects, such as construction of roads, power plants, bridges, schools, or other public facilities. According to the Securities Industry and Financial Markets Association, approximately $670 billion worth of municipal bonds were issued in 2010. The total US municipal bond market is currently valued at approximately $3.7 trillion, being one of the world's largest security markets.

[2] This is true even if the auctioneer reveals the winner's identity. See Proposition 3 in Marshall and Marx (2009).

Municipal bonds are initially sold either through negotiated sales, or through auctions.[3] The municipal bond's issuer gets a cash inflow at the time of issuance, while agreeing in exchange to pay back the principal plus the accrued interest to the bond holders over time. When these bonds are issued, the respective funds are obtained immediately and are deposited into three types of funds: (i) project fund, used to pay for the actual construction or repair work; (ii) sinking fund, used for making principal and interest payments to bond holders; (iii) debt service reserve fund, used to pay debt obligations in case of unforeseen contingencies.

However, the development of a project cannot always be timed perfectly with the expenditure plan (for instance, there may be unpredicted delays in the construction due to external factors). As a solution, once government obtains the proceeds from bonds, it will typically invest it in municipal derivatives until the proceeds are needed to be expensed or paid out to bond holders.

The most common type of instrument is called a guaranteed investment contract. A GIC is comparable to a hybrid of a certificate of deposit and a savings account. As a result, the issuer can earn returns on bond proceeds (which are higher than if the funds were placed in a traditional savings account), and maintain liquidity required for the repayment of the bond's principal and interest accrued.

GICs are usually provided by large financial institutions such as AIG, Citicorp, UBS, Morgan Stanley, Bank of America, MBIA, Goldman Sachs, and others. The government requires each bidder to submit a bid, offering an interest rate – the highest interest rate bid gets to be the winner and acquires the funds in the course of competitive bidding in an auction. In addition, GIC bids are thoroughly analyzed either internally or by external advisors, to be certain that the complex terms of the contracts are suitable to issuer's specifications and requirements. Finally, the winning GIC bidder should get the contracts issued timely and in conformity with the bid proposal.

3.2.1 Collusion in GIC Auctions

Following the IRS investigation, several US municipalities filed individual antitrust complaints with the Department of Justice (DoJ). The leading complaint was filed by the City of Los Angeles, and contained allegations against 37 provider defendants and 9 broker defendants, including CDR, IMAGE

[3] Municipal bond auctions have been studied within the structural paradigm by Shneyerov (2006) and Tang (2011).

and Sound Capital. Since the allegations in these complaints were similar in nature, many of these complaints were later integrated in a single Class Action Complaint (CAC), that was filed in August 2008 against more than 40 corporate defendants. The complaint was dismissed by the court, however, citing insufficient factual evidence.[4] Subsequently, a Second Class Action Complaint (SCAC) was filed against a smaller list of defendants.[5].

Up to date, 20 individuals and several corporate defendants have been indicted, including the executives of CDR, the largest broker. These indictments resulted in significant recent settlements, by the defendants Bank of America ($137 million), UBS AG ($160 million), JP Morgan Chase ($228 million), and GE Funding Capital Market Services ($70 million). See Table 3.1 for the timeline of the investigations.

Several court documents describe the alleged bid rigging schemes in more detail. For example, in the complaint filed by the SEC against J.P. Morgan Securities LLC (JPMS) in a district court in July, 2011, the plaintiff alleges that JPMS, over an eight-year period, "rigged at least 93 transactions concerning the reinvestment of proceeds from the sale of over $14.3 billion of underlying municipal securities, generating millions of dollars in ill-gotten gams".[6] This rigging allegedly took several forms. First, JPMS was able to win some of these auctions because it obtained advance information from a bidding agent on the bids placed by other participants (the so-called "last looks" allegation). In one transaction,

> "Municipality C, a New Jersey entity, issued $690,000,000 of municipal bonds for the purpose of, among other things, funding a portion of the state transportation system costs. In connection with the temporary investment of the proceeds from these bonds, Municipality C also retained the services of Bidding Agent B to bid out the FPA [forward purchase agreement] for a project fund. JPMS — with the help of Bidding Agent B — won this tainted

[4]The original CAC complaint relied heavily on statistical analyses of bidding patterns, in particular using the IRS shortcut that a bid may be a sham bid if it falls below 100 to 150 basis points below the winning bid. Evidently, the court adopted a more stringent standard in its investigation, putting more emphasis on documented communication between the conspirators.

[5]See a recent article in Bond Buyer, available here http://www.bondbuyer.com/issues/122_1/will-market-see-more-big-rigging-cases-in-2013-1047224-1.html?zkPrintable=true, summarizes the state of the investigations and the resulting trials and convictions as of December 31, 2012.

[6]This complaint can be accessed on the SEC website, at http://www.sec.gov/litigation/complaints/2011/comp22031.pdf.

bid through Last Looks. [...] On the morning of the bid date, a telephoned discussion ensued between a Bidding Agent B representative and a JPMS Marketer, in which the JPMS Marketer asked the representative if he had heard ”anything in terms of a rate?” Bidding Agent B's representative responded that he hoped it would be 2.5% or better and that ”I will give you as much help as I can with this trade.” [...] Bidding Agent B's representative stated that the highest bid that he had received to date was 2.7%.”

Second, JPMS participated in an arrangement where it was pre-selected as the auction winner, and the bidding agent solicited non-winning, or courtesy, bids from some other GIC providers in order to make the process appear competitive. Citing another transaction in the SEC complaint,

"In the fall of 2001, Municipality B sought a new FPA for the debt service reserve fund, which its board decided would be awarded through the competitive bidding process to the Provider submitting the bid with the highest upfront payment. JPMS, however, acting both as agent for the Provider and essentially as the de facto Bidding Agent, rigged this bid so that it would win the FPA, by, among other things, limiting the bid list to potential Providers who agreed in advance to submit purposely non-winning bids. JPMS, in order to rig this bid for itself, took advantage of the fact that the Municipality B's chief financial officer (”CFO”) did not want to pay fees to a Bidding Agent and instead preferred that the prospective Providers submit their bids directly to him. However, JPMS — with the aid of Bidding Agent B —surreptitiously assumed the role of the Bidding Agent. Indeed, JPMS drafted the bid specifications and with the help of Bidding Agent B, created a list of prospective Providers who agreed, in advance, to submit purposely non-winning bids.”

In addition, JPMS itself allegedly participated in submitting courtesy bids for bidding agents, thereby allowing other providers to win:

"Transaction F was a purposely non-winning bid. A certain firm underwrote a $145,000,000 offering of revenue bonds and, on October 23, 2001, arranged for its related commercial bank to win, through the mechanism of a fraudulent set-up, the bid for one of the instruments in which the offering proceeds would be invested. To facilitate the rigging of this

> transaction, Bidding Agent A secured a purposely non-winning bid from JPMS. JPMS knew it was being asked to submit a non-winning bid, and, a JPMS Marketer needed Bidding Agent A's help to formulate its bid not only to ensure
>
> A bidding agent, or broker, acts on the behalf of the municipality and administers the auction process. In particular, in order to preserve a tax-its bid was in an appropriate range, but also to ensure its bid would not win."

exempt status of the investment income, IRS regulations require that the investment be purchased at a fair market value. The role of the bidding agent is to ensure that this is in fact the case. In particular, the aforementioned regulations stipulate that, in order for the bidding process to be deemed competitive, at least three serious bids should be available. Moreover, the solicitation should be made in good faith. But in the allegations, the bidding agents sometimes facilitated collusion rather than enhanced competition. In the JPSM case, the above mentioned SEC complaint alleged that:

> "In July 2000, JPMS underwrote a $55,000,000 offering of revenue bonds and caused Municipality A, a California entity, to select Bidding Agent A as its Bidding Agent. As agreed upon with JPMS, in return for this business, Bidding Agent A restricted the list of prospective bidders and afforded JPMS Last Looks with respect to two bids for the temporary investment of proceeds of the aforementioned bonds.[...] In addition, in October 2000, after the responsible JPMS banker had left JPMS's employ, Bidding Agent A paid him approximately $19,600 in cash for causing Municipality A to select Bidding Agent A as the Bidding Agent."

The evidence contained in the court documents indicates that the pattern of collusion is consistent with the operation of a cartel, but that the cartel was probably not all-inclusive, with competitive bids also playing a large role. It is reasonable to conjecture that GIC brokers played a major role in coordinating the cartel, and they themselves might have been pre-selected by the cartel taking into account preferences of the municipalities. If the competitive "fringe" were small and unimportant, while the cartel had had overwhelming market power, there would be little need to resort to tactics such as last bid lookups. It may have been sufficient for the cartel to pre-select the winner, and then force the minimal interest rate acceptable to the municipality by soliciting several courtesy bids in order for the solicitation be deemed competitive. The presence of the last lookup in the cases identified in the court documents indicates that, at least in some cases, the solicitations also involved competitive bids. In other words, both competition and collusion likely played an important role.

In the collusive schemes identified in the court documents, the auction was (or should have been) conducted according to the first-price, sealed-bid format. The winning bidder always paid its bid. The alleged coordination of bids by brokers, in the presence of some competitive (non-cartel) bids, required at times frequent updating of cartel bids to ensure that the cartel would win at the lowest possible rate, while the courtesy bids would remain within a certain range (within 100 basis points would provide a safe harbour to the issuers).

Our dataset, described in more details in the next section, involves *open* auctions conducted over the Internet, rather than sealed-bid auctions coordinated through a broker. The open nature of such auctions is meant to attract more competition. In some allegations[7], brokers actively sought to restrict competition by artificially raising the cost of entry.

As any prospective bidder may simply register through the website, rather than through the broker, the barriers to entry are likely to be lower in Internet open auctions. Also, Grant Street Group implements the *closed-exit* rule, according to which the bidders only see the status of their bid (winning or not), but not the bids of other bidders. This rule makes operating the cartel more difficult since should a deviation occur, it would not be detectable in the current auction.

It could still be possible to collude in open Internet GIC auctions. First, all bids are publicly disclosed after the auction. So a deviation could be detected after the auction and the deviator could be punished in a repeated game. Second, even though bidders might not be able to coordinate their bids on the auction website, they could still use other means of communication such as telephone or email. Ultimately, whether or not the open Internet auctions have succeeded in overcoming potential collusion is an empirical question. In the next section, we show how our tests can be used to answer this important question.

[7] cite the CDR case

3.3 Dataset

We employ a dataset of 215 Internet GIC auctions conducted over the period October 2000 - December 2008. Thus the dataset covers both the pre-investigation period and two years into the investigation. This dataset was obtained from the website of Grant Street Group that administers these auctions on behalf of bond issuers. For each auction, our data include the following information: Issuer's name, brief description of the contract, auction date, bidder name/ticker, bid rate offered, principal amount. In addition, we have also extracted data on yields for two long-term US T-bills, the 10- and 20-year bills, matched with the time that each auction took place. In our empirical exercise, however, we control for the heterogeneity by estimating a fixed-effects regression, and only use the data on bids and bidder identities. In order to control for auction heterogeneity, we have also collected data on yields for two long-term US T-bills, the 10- and 20-year bills. The yields were matched with auction dates to control for the market conditions on the day of the auction.[8]

The auctions are conducted as ascending-bid and closed-exit. This means that the participants only observe the current status of their bids, either winning or losing. A losing bid is automatically rejected, but can be updated to a higher bid at any future instance. If a bidder enters a bid higher than the current winning bid, then this bidder becomes the current winner and is informed of this fact. However, other bidders do not observe the current winning bid, nor are they informed about the identity of the current winner.

The data indicate that bidders in the GIC auctions are cognizant of their incentives. Indeed, the wide majority of these auctions result in tight races where bids are raised by the smallest allowable increment. See Figure 3.1 for one example of such a race.[9] Three bidders participated: Aegon NV, a major Dutch financial services company, Rabobank, a major Dutch-based international bank, and Trinity LLC, owned by the financial arm of General Electric Inc. The auction was won by Trinity LLC. Two facts are notable. First, not all bids are submitted in the smallest increments. The initial bids by

[8]Unfortunately, our data does not include contract durations that would allow a more precise matching with the corresponding treasury yield. However, given the relatively long-term nature of the underlying municipal bonds, it is reasonable to expect that the aforementioned Treasury yields would provide a decent approximation.

[9]The image containing the figure was downloaded from the Grant Street Group's website, http://www.grantstreet.com/auctions/results.

Aegon NV have large increments. However, these are essentially non-serious bids as they would have little chance of winning in the prevailing market conditions. The majority of *serious* bids are in fact submitted in the smallest increments. This particular auction illustrates a general phenomenon observed in GIC Internet auctions: the wide majority of these auctions conform closely to the button model.

Figure 3.2 exhibits a histogram of a measure of jump bidding, *the money left on the table*. This is simply the difference between the winning bid and the second-highest. There is a pronounced spike at (approximately) 0, and as well pronounced clustering around 0. More than 80% of all auctions have the money left on the table amounting to less than 5 percentage points.

The total raw number of participants is equal to 43. However, these raw bidders were aggregated since several bidder groups in fact belonged to a single corporate entity. As a result of this aggregation, the final list of bidders, reported in Table 3.2, contained 30 bidders. Table 3.2 exhibits the identities of the bidders, along with the number of bids submitted. The average bid rate is 3.87 with a standard deviation of 1.42. The maximum bid is 6.55. The minimum number of bidders in an auction is 2, and the maximum is 13, with the average being 7 bids.

3.4 De-censoring Approach (Independent Case)

3.4.1 Empirical Results

In order to implement our collusion test assuming independence, we need to know the identity of at least one competitive bidder. In order to increase the precision of our estimates, it is in fact desirable to have several competitive bidders, so that the highest bid among them reveals the valuation of the losing cartel leader relatively often. In Table 3.2, we identify for each bidder whether or not it was on the defendant list in (i) CAC, (ii) SCAC and (iii) the Los Angeles complaint. As can be seen, the Los Angeles complaint provides the most extensive list, overlapping to some extent with the list on CAC complaint. The SCAC list, on the other hand, is much smaller subset of CAC.

For the purposes of our collusion test, we decided to use the list of firms in any of the complaints mentioned in Table 3.2 as our collusive superset $\mathcal{N}_{col}$. In order to remove the effect of auction heterogeneity, both observed and unobserved, we follow Bajari, Hong, and Ryan (2010) and Bajari, Houghton, and Tadelis (2014) and estimate a fixed-effects regression with auction-level fixed effects. We then apply our procedure to the residuals of this regression.

To begin illustrating our procedure, we have picked Rabobank as the alleged conspirator, as the bank that has submitted most bids among all alleged conspirators. Figure 3.3 shows the estimated CDFs for Rabobank. The blue curve is the empirical CDF of Rabobank's estimated residuals in the fixed-effect regression. The red curve is the predicted CDF assuming Rabobank is competitive, estimated by following our de-censoring approach. The figure shows that the CDFs are actually quite close to each other, and cross several times. There is no visual evidence of stochastic dominance as would be if Rabobank colluded. Our test of collusion has a p-value of 0.26, which implies that the hypothesis of the competitive behavior for Rabobank cannot be rejected at the customary levels of confidence.

3.4.2 Holm-Bonferroni Test

Next, we have implemented the Holm-Bonferroni test. Numerically, we have found our estimator to be unreliable for banks that have submitted fewer than 40 bids, so only 9 banks with the number of bids above this threshold were included. The test results are shown in Table 3.3. There is one participant (XL Capital with a p-value $= 3\%$) for whom the p-value is individually significant at the 5% level. However, it does not pass the rejection cutoff of the Holm-Bonferroni procedure. At customary significance levels, the test does not reject competition.[10]

3.5 Copula-Based Approach (Dependent Case)

As outlined in the second chapter, we propose a copula-based approach to establish identification and estimation of model primitives within English auctions under the absence of independence, which is implemented in this section. Copula-based methodology is implemented here to recover and estimate linear and non-linear association between the variables of interest due to its ability it can parametrize the link between the variables in a joint distribution. Finally, we apply a semi-parametric strategy, based

[10] The Holm-Bonferroni adjusted p-value is 0.27.

on Archimedean copulae, to identify and estimate the model primitives and take into account the correlation between bids. One advantage that this approach has is that it allows us to separate the estimation of the marginal distribution from the estimation of the joint distribution of underlying bidder values. Finally, we will allow us to uncover the unobservable latent value distributions using our sample data.

3.5.1 Empirical Results

In the application of copula theory, as formulated in the second chapter, our initial goal is to select a copula among the Archimedean class that is flexible enough, allowing any level of positive dependence, while it will accurately capture the dependency structure for each pair of variables. Therefore, we have used the Clayton copula in our empirical test for convenience, however we can easily generalize this to any other copula type in the Archimedean class without loss of generality. The Clayton copula also allows for independence, treated as a special case. As in the independence case, we need to have at least one competitive bidder in the data. We will use the list of allegedly colluding firms outlined in Table 3 as our collusive superset. Initially, we need to estimate the copula parameter, θ, in order to replace its unknown counterpart in the generator function. Using our main estimating equation for θ and replacing the unknown distributions with their empirical counterparts from the sample, we obtain bidder-level estimator $\hat{\theta}_i$. After estimating the individual parameters for each bidder, we obtain the bidder-invariant copula parameter of 0.55 for our test, by estimating a weighted average across bidders. Subsequently, we can use the estimator obtained in the previous step $\hat{\theta}$ to predict the counterfactual distributions - in other words, how would the collusive firms behave if they were competitive. If the bidder is competitive, then the counterfactual and actual distributions will coincide. However, if the bidder is collusive, we show that the counterfactual competitive bid distribution stochastically dominates the actual collusive one. Hence, we apply this statistical test of the null hypothesis of competitive bidding against the alternative of collusive bidding on each individual bidder.

As we have discussed it in the independent case, we take into account the fact that it is probable to observe auctions where non-homogenous objects are auctioned off. Once again, we control for observed and unobserved heterogeneity by incorporating the approach used Bajari et al. (2010) and Bajari et al. (2014), and estimate a fixed-effects regression with auction-level fixed effects. This regression is estimated parametrically, and our estimators are applied to the homogenized bids resulting from this regression.

We have picked Salomon as the alleged cartel member for illustration purposes. Figure 3.4 illustrates the estimated cumulative distribution functions for Salomon, with the blue curve depicting the empirical CDF, and the red curve depicting the counterfactual or predicted CDF assuming the exhibition of competitive behaviour of the bidder in question. We can see that the two curves overlap for the most part, without evidence of stochastic dominance or indicators of collusive behaviour. In sum, once we account for dependence in our model, we obtain a much closer match between the actual empirical distribution function of values and the predicted one obtained in our model. As we see in 3.4, the estimated curves are strikingly close to each other, in line in what we found for the independent case. Once again, there is no visual evidence of stochastic dominance implying collusion.

3.5.2 Holm-Bonferroni Test

We used the Holm-Bonferroni procedure to test for testing sequential testing procedure that we now describe. Let α denote the overall significance level. As in the de-censoring approach, we are only including banks that have submitted more than 40 bids, hence only 9 banks with the number of bids above this threshold were included. We conduct the procedure on the bids from the original sample, and adjusted bids controlling for heterogeneity. The test results are shown in Table 3.4 and 3.5. At customary significance levels, using original bid data, the test does not reject competition. Once tested with adjusted bids, there are two bidders (Salomon with a p-value $= 0.5\%$, and Morgan Stanley with a p-value $= 0.25\%$) for whom the p-value is individually significant at the 5% level, hence we do not reject collusive behaviour for these two bidders. However, since total number of bids submitted by both is not significant overall, it is unlikely that this pair of bidders would form a cartel. Therefore, we reject collusive behaviour for the adjusted sample as well.

3.6 Concluding remarks

Based on this empirical study of the municipal GIC auctions, we conclude that despite the fact that there have been allegations of collusion in this market, our test does not detect deviations from competition. Both methodologies that we have implemented in this empirical study have produced consistent results with the aforementioned conclusion. A plausible explanation of this finding is that the Internet auction design involves very limited information disclosure.

All in all, whether or not the open Internet GIC auctions have been successful in achieving the goal of combatting collusion is an interesting empirical question investigated in our paper. We take advantage of the fact that the set of alleged conspirators in GIC auctions can be determined from court case filings for non-Internet auctions. Our test finds no evidence of collusive behavior.

Finally, the research in this paper can be extended in a number of directions. Below, we discuss three important but challenging extensions.

First, we restrict attention to English auctions. Can our approach be extended to another popular format, first-price auctions (FPA)? In English auctions, bidders stay in the auction up to their valuations. As we have shown, this crucial feature allows one to identify the the distribution of valuations of a given bidder *regardless* of whether other bidders are colluding and who participates in the cartel. In FPAs, bidders bid less than their values, and the competitive bids depend on whether there is a cartel, and on the cartel composition. A combination of our approach with the identification and estimation methodology for first-price auctions proposed in Guerre, Perrigne, and Vuong (2000) is clearly desirable.

The second extension concerns relaxation of the efficient cartel hypothesis. While many papers in the empirical auction literature assume efficient collusion, this is obviously a limitation. This is not always the case when bidders are asymmetric. Asker (2010) has recently estimated a structural model of a knockout auction for a stamp dealer cartel and found evidence of inefficient allocation. Incorporating richer leader selection rules supported is desirable and left for future work. As Asker (2010) has demonstrated for a postal stamp cartel, a cartel large enough to exercise market power may include bidders that are quite different, and may adopt a knockout auction that leads to inefficient allocation. If the form of the knockout auction is known to the researcher, one could use this information to extend our approach. More specifically, given the distributions of valuations, any cartel mechanism determines the leader selection probabilities. And vise versa, given the leader selection probabilities, the value distributions could be identified using our approach. This fixed-point reasoning opens up a way to identify the distributions of valuations of the members of the (potentially inefficient) cartel. However, this may be challenging not least because knockout auctions with asymmetric bidders may be difficult to solve even numerically. This extension is left for future research.

Third, our approach relies on the button model of the English auction, which as we have argued, is applicable to recent Internet auction designs with minimal information disclosure, where bidders only see the status of their bid (winning or losing). The auction format provides incentives for bidders to bid up to their true values. Indeed, it is easy to see that that the closed-exit format matches *exactly* the button-, or thermometer-auction paradigm first proposed in Vickrey (1961), where bidding own valuation is a weakly-dominant strategy. Moreover, the closed-exit rule ensures that this equilibrium is unique. In particular, the model is suitable for our empirical application. In this model, it is a dominant strategy for a bidder to drop out at its valuation. Haile and Tamer (2003) argue that this assumption is unrealistic in traditional English auctions and develop sharp nonparametric bounds on the distributions of valuations when it does not hold. Whether or not their bounding approach could be extended to collusion is an open question also left for future research.

3.7 Tables

Table 3.1: Timeline

November, 2006	FBI allegedly rides the offices and seizes documents of financial broker firms Rubin/Chambers, Dunhill Insurance Services (CDR), Investment Management Advisory Group (IMAGE), and Sound Capital Management Inc.
December, 2006	DOJ Antitrust brought their case to the Southern District Court of New York (S.D.C.N.Y.).
January, 2007	One of the defendants, Bank of America, enters into the DOJ leniency program. Subsequently, several municipalities filed complaints to various courts.
August, 2008	Consolidated Class Action Complaint (CAC) filed against more than 40 corporate defendants. However, the defendants almost immediately filed a motion to dismiss.
April, 2009	The S.D.C.N.Y. granted the defendants their motion, citing lack of factual evidence.
June, 2009	The CAC plaintiffs filed Second Class Action Complaint (SCAC) against a smaller list of defendants. The defendants immediately respondent with a motion to dismiss.
September, 2009	The City of Los Angeles files a first amended complaint against a number of corporate defendants, including both GIC providers and brokers.
March, 2010	The S.D.C.N.Y. denied the SCAC defendants' motion.
December, 2010	Bank of America settles for $137 million.
May, 2011	One of the defendants in SCAC, UBS AG, agrees to settle and pay $160 million for its anticompetitive conduct in the municipal derivative market.
July, 2011	Defendant JP Morgan Chase Inc., agrees to settle and pay $228 million for its anticompetitive conduct in the municipal derivative market.
December, 2011	Defendant GE Funding Capital Market Services Inc. agrees to settle and pay $70 million for its anticompetitive conduct in the municipal derivative market.
January, 2012	An executive and former executive of CDR pleaded guilty for participating in bid rigging.

Table 3.2: Internet Auction Participants

		Complaints:		
Bidder	Number of bids	CAC	SCAC	Los Angeles
ABN AMRO	4			
AEGON	144			
AMBAC Capital Funding	14			
American Internation Group, Inc.	140	X		X
Bank of America	8	X	X	X
Bayerische Landesbank	103			X
Bear Stearns Inc.	11	X	X	
Citigroup	2			X
Credit Agricole	42			
DEPFA Bank	82			
Financial Guaranty Insurance Co. LLC	22	X		X
Financial Security Assurance Ltd.	49	X		X
First Union National Bank	8			
GE Funding Capital Market Services, Inc.	20	X		
HSBC Bank	11			
Hypo Real Estate Bank	63			
ING Bank	9			
JP Morgan Chase	13	X	X	X
Lehman Brothers	4			
MBIA Inc.	70			X
Merrill Lynch Inc.	10	X		X
Morgan Stanley	30	X	X	X
Natixis S.A.	48	X	X	X
Rabobank	138			X
Royal Bank of Canada	8			
Societe Generale SA	49	X	X	X
UBS AG	1	X	X	X
Wells Fargo	7			X
Westdeutsche Landesbank	11			
XL Capital	42	X		X

Table 3.3: Test Results (Independent Case)

Bidder name	p-value	Holm-Bonferroni cutoff
XL Capital	0.03	0.006
Rabobank	0.24	0.006
American International Group Inc.	0.33	0.007
Natixis	0.35	0.008
FSA	0.38	0.010
Bayerische Landesbank	0.52	0.013
Salomon	0.71	0.017
MBIA	1	0.025
Morgan Stanley	1	0.050

Table 3.4: Test Results - Raw Bids (Dependent Case)

Bidder name	p-value	Holm-Bonferroni cutoff
Salomon (Citigroup)	0.01	0.006
Morgan Stanley	0.0175	0.006
American International Group Inc.	0.0475	0.007
Natixis	0.06	0.008
Bayerische Landesbank	0.0875	0.010
MBIA	0.11	0.013
XL Capital	0.1275	0.017
FSA	0.185	0.025
Rabobank	0.925	0.050

Table 3.5: Test Results - Adjusted Bids (Dependent Case)

Bidder name	p-value	Holm-Bonferroni cutoff
Morgan Stanley	0.0025	0.00556
Salomon (Citigroup)	0.005	0.00625
XL Capital	0.0075	0.007
Natixis	0.0275	0.008
American International Group Inc.	0.0475	0.010
MBIA	0.0775	0.013
Bayerische Landesbank	0.185	0.017
FSA	0.225	0.025
Rabobank	0.36	0.050

3.8 Figures

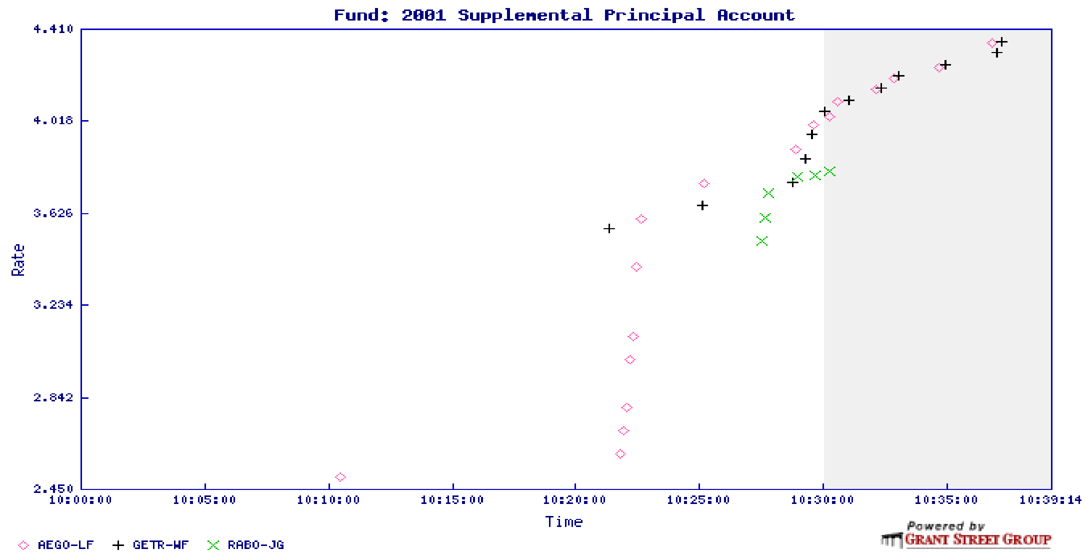

Figure 3.1: A tight GIC bidding race.

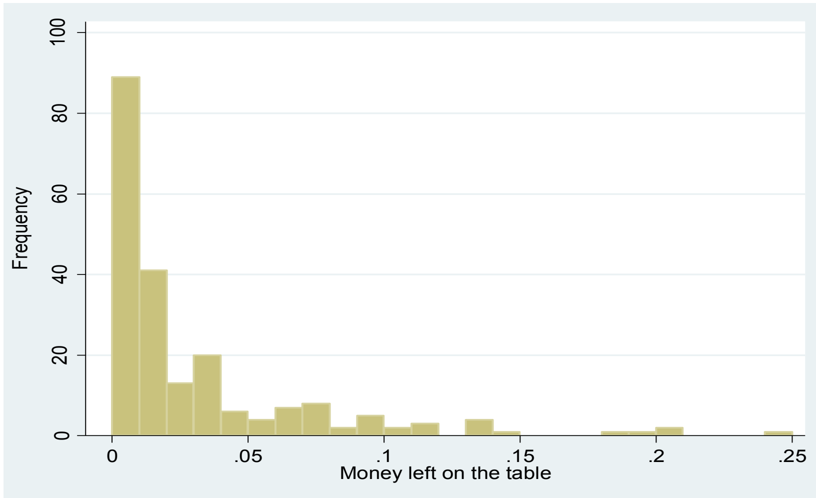

Figure 3.2: Money left on the table.

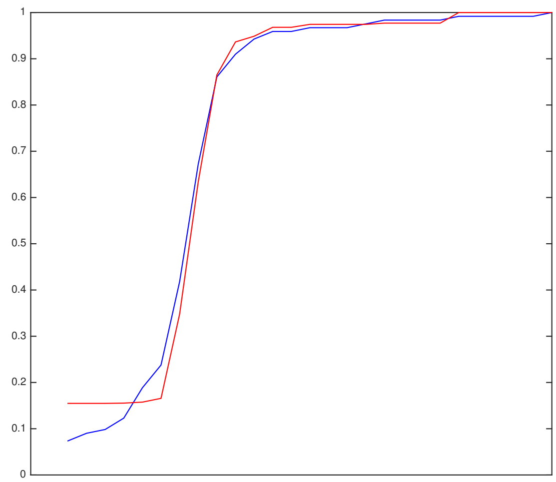

Figure 3.3: Rabobank: Predicted competitive (red) vs. actual (blue) CDFs of bids.

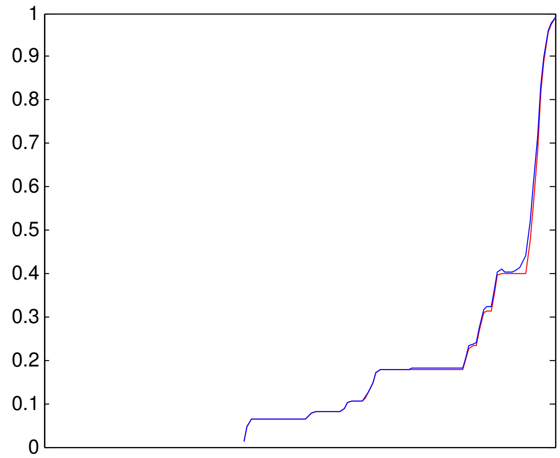

Figure 3.4: Salomon: Predicted competitive (red) vs. actual (blue) CDFs of bids.

References

Aalen, O. (1978). Nonparametric inference for a family of counting processes. *Annals of Statistics*, *6*(4), 701–726.

Aradillas-López, A., Gandhi, A., & Quint, D. (2011). *Identification and inference in ascending auctions with correlated private values.* (Working Paper. University of Wisconsin, Madison)

Aryal, G., & Gabrielli, M. F. (2013). Testing for collusion in asymmetric first-price auctions. *International Journal of Industrial Organization*, *31*(1), 26–35.

Asker, J. (2010). A study of the internal organization of a bidding cartel. *American Economic Review*, *100*(3), 724–762.

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

Bajari, P., Hong, H., & Ryan, S. P. (2010). Identification and estimation of a discrete game of complete information. *Econometrica*, *78*(5), 1529–1568.

Bajari, P., Houghton, S., & Tadelis, S. (2014). Bidding for incomplete contracts: An empirical analysis of adaptation costs. *American Economic Review*, *104*(4), 1288–1319.

Bajari, P., & Ye, L. (2003). Deciding between competition and collusion. *Review of Economics and Statistics*, *85*(4), 971–989.

Baldwin, L., Marshall, R., & Richard, J.-F. (1997). Bidder collusion at forest service timber sales. *Journal of Political Economy*, *105*(4), 657-699.

Braekers, R., & Veraverbeke, N. (2005). A copula-graphic estimator for the conditional survival function under dependent censoring. *The Canadian Journal of Statistics*, *33*, 429–447.

Brendstrup, B., & Paarsch, H. (n.d.). Semiparametric identification and estimation in multi-object, english auctions.

Chen, K., & Lo, S. (1997). On a mapping approach to investigating the bootstrap accuracy. *Probability Theory and Related Fields*, *107*(2), 197–217.

Coey, D., Larsen, B., & Sweeney, K. (2014). *The bidder exclusion effect.* (NBER Working Paper 20523)

Embrechts, P. (2009). Copulas: a personal view. *Journal of Risk and Insurance*, *76*, 639–650.

Fan, Y., & Liu, X. (2013). Identification and inference in censored parametric quantile regression: Sensitivity to independent censoring.

Fermanian, J.-D., & Scaillet, O. (2003). Nonparametric estimation of copulas for time series. *Journal of Risk*, *5*, 25–54.

Genest, C., & MacKay, R. (1986). Copules archimediennes et familles de lois bidimensionnelles dont les marges sont données. *La Revue Canadienne de Statistique*, *14*, 145–159.

Genest, C., & Rivest, L. (1993). Statistical inference procedures for bivariate archimedean copula. *Journal of the American Statistical Association*, *8*, 1034–1043.

Graham, D. A., & Marshall, R. C. (1987). Collusive bidder behavior at single-object second-price and english auctions. *Journal of Political Economy*, *95*(6), 1217–1239.

Guerre, E., Perrigne, I., & Vuong, Q. (2000). Optimal nonparametric estimation of first-price auctions. *Econometrica*, *68*(3), 525–74.

Haile, P., Hong, H., & Shum, M. (2003). *Nonparametric tests for common values at first-price sealed-bid auctions.* (NBER Working Paper 10105)

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

Hanoch, G., & Levy, H. (1969). The efficiency analysis of choices involving risk. *Review of Economic Studies*, *36*(3), 335-346.

Harrington, J. E. (2008). Detecting cartels. In P. Buccirossi (Ed.), *Handbook of antitrust economics* (pp. 213–258). Cambridge, MA: MIT Press.

Hendricks, K., & Porter, R. (1989). Collusion in auctions. *Annales d'Economie et de Statistique*, *15–16*, 217–230.

Hendricks, K., Porter, R., & Tan, G. (2008). Bidding rings and the winner's curse. *RAND Journal of Economics*, *39*(4), 1018–1041.

Hofert, M., & Scherer, M. (2011). Cdo pricing with nested archimedean copulas. *Quantitative Finance*, *11(5)*, 775–787.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.

Hubbard, T., Li, T., & Paarsch, H. (2012). Semiparametric estimation in models of first-price, sealed-bid auctions with affiliation. *Journal of Econometrics*, *168*, 4–16.

Joe, H. (1997). *Multivariate models and dependence concepts*. London: Chapman and Hall.

Klemperer, P. (2002). What really matters in auction design. *Journal of Economic Perspectives*, *16*(1), 169–189.

Komarova, T. (2013). Nonparametric identification in asymmetric second-price auctions: a new approach. *Quantitative Economics*, *4*(2), 269–328.

Mailath, G., & Zemsky, P. (1991). Collusion in second price auctions with heterogeneous bidders. *Games and Economic Behavior*, *3*(4), 467–486.

Marshall, R. C., & Marx, L. M. (2009). The vulnerability of auctions to bidder collusion. *Quarterly Journal of Economics*, *124*(2), 883–910.

Meilijson, I. (1981). Estimation of the lifetime distribution of the parts from the autopsy statistics of the machine. *Journal of Applied Probability*, *18*(4), pp. 829-838.

Milgrom, P. R. (2004). *Putting auction theory to work*. Cambridge, UK: Cambridge University Press.

Nelsen, R. (2006). *An introduction to copulas*. New York: Springer.

Nelson, W. (1969). Hazard plotting for incomplete failure data. *Journal of Quality Technology*, *1*(1), 27–52.

Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics*, *14*(4), 945–966.

Oakes, D. (1989). Bivariate survival models induced by frailties. *J. Am. Statistical Association*, *84*, 487–493.

Oh, D.-H., & Patton, A. (2012). Modelling dependence in high dimensions with factor copulas. (Working paper, Duke University)

Pesendorfer, M. (2000). A study of collusion in first-price auctions. *Review of Economic Studies*, *67*(3), 381–411.

Porter, R., & Zona, D. (1993). Detection of bid rigging in procurement auctions. *Journal of Political Economy*, *101*(3), 518–538.

Porter, R., & Zona, J. (1999). Ohio school milk markets: An analysis of bidding. *RAND Journal of Economics*, *30*(2), 263–288.

Revuz, D., & Yor, M. (1999). *Continuous martingales and brownian motion*. Berlin: Springer.

Sancetta, A., & Satchell, S. (2004). The bernstein copula and its applications to modeling and approximations of multivariate distributions. *Econometric Theory*, *20*, 535–562.

Schurter, K. (2017). *Identification and inference in first-price auctions with collusion.* (Working Paper, University of Chicago)

Shneyerov, A. (2006). An empirical study of auction revenue rankings: The case of municipal bonds. *RAND Journal of Economics*, *37*(4), 1005–1022.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de lInstitut de Statistique de lUniversité de Paris*, *68*, 229–231.

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

Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences of the United States of America*, *72(1)*, 20–22.

van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge: Cambridge University Press.

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