Center for Economic Institutions

Working Paper Series

No. 2010-10

"Semiparametric Estimation in Models of First-Price, Sealed-Bid Auctions with Affiliation"

Timothy P. Hubbard, Tong Li
and Harry J. Paarsch

January 2011

Institute of Economic Research
Hitotsubashi University
2-1 Naka, Kunitachi, Tokyo, 186-8603 JAPAN
http://cei.ier.hit-u.ac.jp/English/index.html
Tel:+81-42-580-8405/Fax:+81-42-580-8333

Semiparametric Estimation in Models of First-Price, Sealed-Bid Auctions with Affiliation

Timothy P. Hubbard[a], Tong Li[b], Harry J. Paarsch[c,*]

[a]*Department of Economics, Texas Tech University, USA*
[b]*Department of Economics, Vanderbilt University, USA*
[c]*Department of Economics, University of Melbourne, Australia*

Abstract

Within the affiliated private-values paradigm, we develop a tractable empirical model of equilibrium behaviour at first-price, sealed-bid auctions. The model is non-parametrically identified, but the rate of convergence in estimation is slow when the number of bidders is even moderately large, so we develop a semiparametric estimation strategy, focusing on the Archimedean family of copulae and implementing this framework using particular members—the Clayton, Frank, and Gumbel copulae. We apply our framework to data from low-price, sealed-bid auctions used by the Michigan Department of Transportation to procure road-resurfacing services, rejecting the hypothesis of independence and finding significant (and high) affiliation in cost signals.

Key words: first-price, sealed-bid auctions; copulae; affiliation.
JEL classification: C20, D44, L1.

1. Motivation and Introduction

During the past half century, economists have made remarkable progress in understanding the theoretical structure of equilibrium strategic behaviour under market mechanisms, such as auctions, when the number of potential participants is relatively small; see Krishna [18] for a comprehensive presentation and evaluation of progress.

One analytic device commonly used to describe bidder motivation at single-object auctions is a continuous random variable that represents individual-specific heterogeneity in valuations. The conceptual experiment involves each potential bidder's receiving an independent draw from a distribution of valuations. Conditional on his draw, a bidder is then assumed to act purposefully, maximizing either the expected profit or the expected utility of profit from winning the auction. Another, frequently-made assumption is that the independent valuation draws of bidders are from the same distribution of valuations; this is often referred to as the *symmetric independent private-values paradigm* (symmetric IPVP). Under this assumption, the researcher can then focus on a representative agent's decision rule when describing equilibrium behaviour.

*Corresponding author. Level 5, Economics & Commerce Building, Parkville, VIC, 3010. Telephone: +61 3 8344 4655; Facsimile: +61 3 8344 6899.

Email addresses: `Timothy.Hubbard@ttu.edu` (Timothy P. Hubbard), `Tong.Li@Vanderbilt.edu` (Tong Li), `HPaarsch@UniMelb.edu.au` (Harry J. Paarsch)

This version: *5 October 2010*

At many real-world auctions, the latent valuations of potential bidders are probably dependent in some way. In auction theory, it has been typically assumed that this dependence satisfies *affiliation*, a term coined by Milgrom and Weber [24]. Affiliation is a condition concerning the joint distribution of signals. In the case of continuous random variables, following Karlin [17], some refer to affiliation as *multivariate total positivity of order two* (or MTP_2, for short). Essentially, under affiliation for continuous random variables, the off-diagonal elements of the Hessian of the logarithm of the joint probability density of signals are all non-negative; i.e., the joint probability density function is log-supermodular. Under joint normality of signals, affiliation requires that the pair-wise covariances be weakly positive. Krishna [18] as well as de Castro [5] have also noted that affiliation implies positive correlation; i.e., affiliation is a much stronger condition than positive correlation.

Investigating equilibrium behaviour at auctions, empirically, when latent valuations are affiliated has challenged researchers for some time. Laffont and Vuong [20] have noted that identification is impossible to establish in many models when affiliation is present. In fact, one result of Laffont and Vuong is that any model within the affiliated-values paradigm (AVP) is observationally equivalent to a model within the affiliated private-values paradigm (APVP). For this reason, when admitting dependence, nearly all empirical workers have considered models within the APVP, the notable exception being Brendstrup and Paarsch [1].

Most structural econometric research devoted to investigating equilibrium behaviour at auctions has involved single-object auctions within the symmetric IPVP. Examples include Paarsch [28, 29]; Donald and Paarsch [6, 7, 8]; Laffont, Ossard, and Vuong [19]; Guerre, Perrigne, and Vuong [13] (hereafter GPV, for short); Haile and Tamer [14]; and Li [21]. Paarsch and Hong [30] have summarized some of the important empirical work in this area.

Only a few researchers have dealt explicitly with models within the APVP. In particular, Li, Perrigne, and Vuong [22] have demonstrated non-parametric identification within the conditional IPVP, a special case of the APVP, while Li, Perrigne, and Vuong [23] (hereafter LPV, for short) have demonstrated non-parametric identification within the APVP. When implementing their approach, LPV faced two related problems: first, the LPV non-parametric estimator suffers from the curse of dimensionality as the number of bidders gets large; second, their estimator is plagued by the curse of dimensionality as the number of covariates gets large. Consequently, the LPV estimator can be slow to converge. Most importantly, however, LPV do not impose affiliation in their estimation strategy, so the first-order condition used in estimation need not constitute an equilibrium.

In this paper, we investigate the advantages of using a semiparametric estimation strategy as a dimension-reducing device to speed-up convergence. Specifically, we address the curse of dimensionality stemming from the number of bidders at auction, not the number of covariates. Under our approach, we also use the properties of a particular family of copulae to impose affiliation, thus ensuring an equilibrium is satisfied by the measurement equation. We focus our efforts on affiliation within models of first-price, sealed-bid auctions, the most important auction format used in practice, at least in terms of the value of goods and services either sold or procured.

Our paper has seven remaining parts. In the next section, we briefly define affiliation and explain why it is used in theoretical models of auctions. Because the copula is central to our analysis, in section 3, we present a brief review of the theory concerning copulae. Subsequently, in section 4, we introduce a simple model of bidding at first-price, sealed-bid auctions in which affiliation is imposed on the copula to guarantee a unique, monotone pure-strategy equilibrium. In section 5, we propose a semiparametric estimator, demonstrating that it is consistent and deriving its asymptotic distribution; we also demonstrate that the proposed estimator attains the optimal rate of convergence. In section 6, we investigate the small-sample properties of our estimator using Monte Carlo methods, while in section 7, we apply our methods in an empirical investigation of low-price, sealed-bid, procurement-contract auctions held by the Department of Transportation in the State of Michigan. We summarize and conclude in section 8, the final section of the paper. In an appendix, we collect several lemmata (and their proofs) that are too cumbersome and detailed to include in the text of the paper; we also document the creation of the data set used.

2. Definition and Use of Affiliation

Suppose valuations V_1, V_2, …, V_n have joint probability density function $f_{\boldsymbol{V}}(\boldsymbol{v})$ where $\boldsymbol{v}$ collects $(v_1, v_2, \ldots, v_n)$, with lower-case letters denoting realizations of upper-case random variables. Consider $\boldsymbol{v}'$ and $\boldsymbol{v}''$. The random variables $\boldsymbol{V}$ are said to be affiliated if

$$f_{\boldsymbol{V}}(\boldsymbol{v}' \vee \boldsymbol{v}'')\, f_{\boldsymbol{V}}(\boldsymbol{v}' \wedge \boldsymbol{v}'') \geq f_{\boldsymbol{V}}(\boldsymbol{v}')\, f_{\boldsymbol{V}}(\boldsymbol{v}'') \tag{1}$$

where

$$(\boldsymbol{v}' \vee \boldsymbol{v}'') = [\max(v_1', v_1''), \max(v_2', v_2''), \ldots, \max(v_n', v_n'')]$$

denotes the component-wise maxima of $\boldsymbol{v}'$ and $\boldsymbol{v}''$, sometimes referred to as the *join*, while

$$(\boldsymbol{v}' \wedge \boldsymbol{v}'') = [\min(v_1', v_1''), \min(v_2', v_2''), \ldots, \min(v_n', v_n'')]$$

denotes the component-wise minima, sometimes referred to as the *meet*. Affiliation is sufficient to guarantee conditions important in delivering a unique, monotone, pure-strategy equilibrium (MPSE). de Castro [5] has also noted that affiliation is a stronger condition than is necessary to guarantee a unique MPSE.

3. Some Results concerning Copulae

The main analytic device we use to organize our analysis is the *copula*. Nelsen [26] has provided a detailed introduction to the theory of copulae. Here, we simply repeat some basic facts that are relevant to our later work as well as establish a notation. In what follows, for expositional reasons, for the most part, we restrict our discussion to bivariate copulae, but the results generalize to the case of n variables easily. Given two variables, U_1 and U_2, a bivariate copula $C(u_1, u_2)$ is a continuous function having the following properties:

1. Domain$(C) = [0, 1]^2$;
2. $C(u_1, 0) = 0 = C(0, u_2)$;
3. $C(u_1, 1) = u_1$ and $C(1, u_2) = u_2$;
4. C is a twice-increasing function, so

$$C(u_1^1, u_2^1) - C(u_1^0, u_2^1) - C(u_1^1, u_2^0) + C(u_1^0, u_2^0) \geq 0$$

for any u_1^0, u_2^0, u_1^1, $u_2^1 \in [0, 1]^2$, such that $u_1^0 \leq u_1^1$ and $u_2^0 \leq u_2^1$.

Because U_1 and U_2 are both defined on the unit interval, they can be viewed as uniform random variables with $C(u_1, u_2)$ being their joint distribution function. Alternatively, U_1 and U_2 can be viewed as the cumulative distribution functions of two random variables V_1 and V_2 which are collected in the vector $\boldsymbol{V}$. In this case, their marginal distribution functions $F_1(v_1)$ and $F_2(v_2)$ are linked to their joint distribution $F_{\boldsymbol{V}}(v_1, v_2)$ by

$$F_{\boldsymbol{V}}(v_1, v_2) = C[F_1(v_1), F_2(v_2)].$$

One attractive feature of copulae is that the marginal cumulative distribution functions do not depend on the choice of the dependence function for the two random variables in question. When one is interested in the association between random variables, copulae are a useful device to use because the dependence structure can be separated from the marginal cumulative distribution functions.

We know, too, from Sklar's Theorem, that the copula C always exists, and is a unique function linking $F_{\boldsymbol{V}}(v_1, v_2)$ with $F_1(v_1)$ and $F_2(v_2)$. Of course, when V_1 and V_2 are independent, the copula is a trivial function as

$$F_{\boldsymbol{V}}(v_1, v_2) = F_1(v_1) \times F_2(v_2).$$

Also, if we introduce the copulae

$$\mathcal{U}(u_1, u_2) = \min(u_1, u_2)$$

and

$$\mathcal{B}(u_1, u_2) = \max(0, u_1 + u_2 - 1),$$

then the following inequalities hold:

$$\mathcal{B}(u_1, u_2) = \max(0, u_1 + u_2 - 1) \leq C(u_1, u_2) \leq \min(u_1, u_2) = \mathcal{U}(u_1, u_2),$$

which are known as the *Fréchet–Hoeffding bounds*.

Consider now S_1 and S_2 which are, respectively, two strictly increasing functions of V_1 and V_2, denoted $\sigma_1(V_1)$ and $\sigma_2(V_2)$. Denote the cumulative distribution functions of S_1 and S_2 by $G_1(s_1)$ and $G_2(s_2)$, respectively. It is important to note that

$$C\left[G_1(s_1), G_2(s_2)\right] = C\left(F_1\left[\sigma_1^{-1}(s_1)\right], F_2\left[\sigma_2^{-1}(s_2)\right]\right) = C\left[F_1(v_1), F_2(v_2)\right].$$

To wit, the copulae of two random variables and two strictly increasing functions of those two random variables are identical. This result is Theorem 2.4.3 of Nelsen [26].

Different families of copulae exist. A simple, and commonly-used, family of copulae that admits non-linear dependence is the *Archimedean* family, which is uniquely characterized by its generator function $\zeta(\cdot)$ where

$$C_\zeta(u_1, u_2) = \zeta^{-1}[\zeta(u_1) + \zeta(u_2)]. \tag{2}$$

Here, $\zeta(\cdot)$ is a convex, decreasing function. Note, too, that $\zeta(1)$ must equal zero and $\zeta^{-1}(u)$ must be zero for any u exceeding $\zeta(0)$. These conditions are both necessary and sufficient for C_ζ to be a distribution function. Copulae within the Archimedean family have the following bivariate joint density function:

$$\boldsymbol{c}_\zeta(u_1, u_2) = -\frac{\zeta''(F_{\boldsymbol{V}})\zeta'(u_1)\zeta'(u_2)}{[\zeta'(F_{\boldsymbol{V}})]^3},$$

Table 1: Commonly-Used Archimedean Copulae

Member	Copula	$\zeta(u;\theta)$	Domain
Clayton	$(u_1^{-\theta}+u_2^{-\theta}-1)^{-\frac{1}{\theta}}$	$\frac{1}{\theta}(u^{-\theta}-1)$	$\theta \in [-1,\infty)\setminus\{0\}$
Frank	$-\frac{1}{\theta}\log\left\{1+\frac{[\exp(-\theta u_1)-1][\exp(-\theta u_2)-1]}{[\exp(-\theta)-1]}\right\}$	$-\log\left[\frac{\exp(-\theta u)-1}{\exp(-\theta)-1}\right]$	$\theta \in (-\infty,\infty)\setminus\{0\}$
Gumbel	$\exp\left\{-\left[(-\log u_1)^{\theta}+(-\log u_2)^{\theta}\right]^{\frac{1}{\theta}}\right\}$	$(-\log u)^{\theta}$	$\theta \in [1,\infty)$

which generalizes naturally to n-variates.

Three commonly-used members of the Archimedean family of copulae are the *Clayton*, *Frank*, and *Gumbel* copulae. In table 1, we present the copula and generator functions of each as a function of a dependence parameter θ. What interpretation can be given to the dependence parameter θ? Consider the Frank copula: in the bivariate case, the larger is a positive value of θ, the greater the concordance, positive dependence. On the other hand, a very negative value of θ indicates negative dependence. Independence obtains when θ approaches zero. Note that, for the Frank copula, when n exceeds two, θ is restricted to be positive because a negative θ would imply a non-monotonic inverse-generator function; see example 4.22 in Nelsen [26]. Similar interpretations exist for the Clayton and Gumbel copulae and can be also found in Nelsen [26].

Müller and Scarsini [25] have characterized various notions of positive dependence, such as MTP_2 and *conditionally increasingness in sequence* (CIS) for Archimedean copulae.[1] They have also presented a general condition that the generator of an arbitrary Archimedean copula must satisfy in order to guarantee that MTP_2 holds (cf. Theorem 2.11 in their paper).

For the Frank copula, Genest [11] has demonstrated that the relevant condition for *total positivity of order two* (TP_2) coincides with the condition that guarantees a monotonic inverse-generator function when n exceeds two; viz., θ must be positive. Although Genest's condition for TP_2 only concerned bivariate copulae, it can be applied to multivariate Frank copulae, too. For it is well-known that a function is MTP_2 if and only if it is TP_2 in all pairs of its arguments. For the Clayton copula, the parameter θ must be positive if that copula is to satisfy TP_2, while the parameter θ of the Gumbel copula must be (weakly) greater than one if that copula is to satisfy TP_2.

4. First-Price, Sealed-Bid Auction Model with Affiliation

We consider a model in which each of $n(\geq 2)$ potential bidders draws a valuation V from the joint distribution $F_{\boldsymbol{V}}(\boldsymbol{v})$ where $F_i(v_i)$ denotes the marginal cumulative distribution function of bidder i. Now, by Sklar's theorem, there exists a unique copula function C such that

$$F_{\boldsymbol{V}}(\boldsymbol{v}) = C[F_1(v_1), F_2(v_2), \ldots, F_n(v_n)].$$

For notational simplicity, however, we consider the case where $F_i(\cdot)$ is the same for all bidders and equals $F_0(\cdot)$; this is the *symmetric* APVP. The extension to the asymmetric APVP is straightforward, but incredibly cumbersome and tedious, notationally speaking. Under the symmetric APVP,

$$F_{\boldsymbol{V}}(\boldsymbol{v}) = C[F_0(v_1), F_0(v_2), \ldots, F_0(v_n)].$$

[1] We are grateful to Professor Alfred Müller for sharing with us his extensive knowledge and insights concerning MTP_2 and the Archimedean family of copulae.

Let $\boldsymbol{V}_{-1}$ denote $(V_2, \ldots, V_n)$, so $\boldsymbol{V}$ without V_1. The following lemma is key to our derivation of the equilibrium bid function:

Lemma 1. *Assume that $F_{\boldsymbol{V}}(v_1, \ldots, v_n)$ is a symmetric distribution that can be expressed as*

$$F_{\boldsymbol{V}}(v_1, \ldots, v_n) = C[F_0(v_1), \ldots, F_0(v_n)]$$

where C is a copula and $F_0(\cdot)$ is the marginal distribution of v_i. The conditional distribution $F_{\boldsymbol{V}_{-1}|V_1}(v_2, ...v_n|v_1)$ can be expressed as

$$F_{\boldsymbol{V}_{-1}|V_1}(v_2, ...v_n|v_1) = C_1[F_0(v_1), F_0(v_2), \ldots, F_0(v_n)]$$

where C_1 is the partial derivative of C with respect to the first component.

Proof.

Let $f_{\boldsymbol{V}}(v_1, \ldots, v_n)$ denote the joint density function related to $F_{\boldsymbol{V}}(v_1, \ldots, v_n)$. Now,

$$\begin{aligned} F_{\boldsymbol{V}}(v_1, \ldots, v_n) &= C[F_0(v_1), \ldots, F_0(v_n)] \\ &= \int^{v_1} \cdots \int^{v_n} f_{\boldsymbol{V}}(u_1, \ldots, u_n)\, \mathrm{d}u_1 \cdots \mathrm{d}u_n. \end{aligned}$$

Differentiating both sides of the last equality with respect to v_1 yields

$$C_1[F_0(v_1), \ldots, F_0(v_n)]\, f_0(v_1) = \int^{v_2} \cdots \int^{v_n} f_{\boldsymbol{V}}(v_1, u_2, \ldots, u_n)\, \mathrm{d}u_2 \cdots \mathrm{d}u_n.$$

Thus,

$$\frac{\int^{v_2} \cdots \int^{v_n} f_{\boldsymbol{V}}(v_1, u_2, \ldots, u_n)\, \mathrm{d}u_2 \cdots \mathrm{d}u_n}{f_0(v_1)} = C_1[F_0(v_1), \ldots, F_0(v_n)]. \tag{3}$$

But the left-hand side of equation (3) is just $F_{\boldsymbol{V}_{-1}|V_1}(v_2, ...v_n|v_1)$, so the desired result obtains.

Note that the assumption of symmetry is unnecessary to our proof; we use it to simplify notation as we shall only investigate the symmetric APVP in our empirical work below. Note, too, that at this stage no structure has been assumed of $C(\cdot)$. Specifically, we have not imposed affiliation. Affiliation is a sufficient condition assumed by Milgrom and Weber [24] to guarantee a unique MPSE in the game specified below.

Within the symmetric APVP, we can consider the behaviour of any bidder so, without loss of generality, we focus on bidder 1 who has value v_1 and is assumed to maximize his expected profit

$$\begin{aligned} \mathcal{E}[\pi(s_1, v_1)] &= (v_1 - s_1) \Pr[V_2 \le \sigma^{-1}(s_1), \ldots, V_n \le \sigma^{-1}(s_1)|v_1] \\ &= (v_1 - s_1) C_1 \left(F_0(v_1), F_0[\sigma^{-1}(s_1)], \ldots, F_0[\sigma^{-1}(s_1)]\right) \end{aligned}$$

by choice of bidding strategy s_1, where $\sigma(\cdot)$ denotes the strictly increasing bidding strategy and where the second equality follows from Lemma 1. Under symmetry, and after some algebra, the first-order condition yields

$$\sigma'(v_1) = [v_1 - \sigma(v_1)](n-1) f_0(v_1) \frac{C_{12}[F_0(v_1), \ldots, F_0(v_1)]}{C_1[F_0(v_1), \ldots, F_0(v_1)]}. \tag{4}$$

A sufficient condition for equation (4) to characterize a unique MPSE is that the copula $C(\cdot)$ satisfy MTP_2, which we now assume it does.

Introducing $G_0(\cdot)$ to denote the marginal distribution of equilibrium bids and $g_0(\cdot)$ to denote its corresponding probability density function, we can apply the GPV approach which involves noting that

$$G_0(s_1) = F_0(v_1)$$

and

$$g_0(s_1) = \frac{f_0(v_1)}{\sigma'(v_1)}.$$

Thus, re-arranging terms of the first-order condition in equation (4) yields

$$v_1 = s_1 + \frac{C_1[G_0(s_1), \ldots, G_0(s_1)]}{(n-1)g_0(s_1)C_{12}[G_0(s_1), \ldots, G_0(s_1)]}, \tag{5}$$

where we use the fact that the copula $C(\cdot)$ is invariant under strictly-increasing transformations of its arguments.

5. A Semiparametric Estimator

We frame the intuition behind our estimation strategy in terms of the previous literature and then, subsequently, demonstrate parameter consistency and asymptotic normality of our estimator later in this section. Consider a sample of T auctions at which no reserve price exists, so issues of participation can be safely ignored. In this case, each of the n participants has tendered a bid at the T auctions, so given the following data:

$$\{\{S_{it}\}_{i=1}^n\}_{t=1}^T,$$

one can non-parametrically estimate $G_0(s)$ and $g_0(s)$ using the methods proposed by GPV. Denote these estimates by $\widetilde{G}_0(s)$ and $\widetilde{g}_0(s)$. From these, again using the sample data, one can then estimate $C[F_0(v_1), \ldots, F_0(v_n)]$ non-parametrically using standard methods for copulae; see, for example, Nelsen [26] as well as Brendstrup and Paarsch [1]. Based on these, one can then form the pseudo-values according to

$$\widetilde{V}_{it} = S_{it} + \frac{\widetilde{C}_1[\widetilde{G}_0(S_{it}), \ldots, \widetilde{G}_0(S_{it})]}{(n-1)\widetilde{g}_0(S_{it})\widetilde{C}_{12}[\widetilde{G}_0(S_{it}), \ldots, \widetilde{G}_0(S_{it})]}.$$

Note, however, that standard kernel-smoothing techniques typically do not guarantee that $\widetilde{C}(\cdot)$ satisfies MTP_2. However, if the true copula $C^0(\cdot)$ satisfies MTP_2, then $\widetilde{C}(\cdot)$ will converge in probability to $C^0(\cdot)$.

In addition, if n is even moderately large, then a non-parametric estimator of the copula $C(\cdot)$ may be slow to converge. Thus, we advocate using semiparametric methods. In particular, one can still estimate both $G_0(s)$ and $g_0(s)$ non-parametrically, but now put some structure on the copula $C(\cdot)$. For example, suppose that $C(\cdot)$ is a member of the Archimedean family which is uniquely characterized by the generator function $\zeta(\cdot)$.

One flexible way of estimating the generating function $\zeta(\cdot)$ would involve introducing a family of shape-preserving polynomials. One would then estimate the coefficients of these polynomials. The estimation strategy would involve allowing the number of terms in the polynomials to increase as the sample size increased, but at a rate which is slower than that of the sample size. Sieve estimation is an example of such a method; see, for example, Chen and Shen [3]. A drawback of this approach is that the polynomial approximations of $\zeta(\cdot)$ need not satisfy MTP_2, so the first-order condition used to define the pseudo-valuations need not correspond to an equilibrium.

Alternatively, consider a family of copulae that is known up to some finite p-dimensional parameter vector $\boldsymbol{\theta}$, so $C(\cdot;\boldsymbol{\theta})$. A number of members of the Archimedean family have simple parametric representations which can be easily constrained to respect affiliation, MTP_2. When the true copula $C^0(\cdot)$ belongs to a parametric family

$$\mathcal{C} = \{C_{\boldsymbol{\theta}},\ \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$$

defined by the vector $\boldsymbol{\theta}$, then a parameter-consistent and asymptotically-normal estimator of $\boldsymbol{\theta}^0$, the true value, can be obtained by applying the method of maximum likelihood.

Now, under the hypothesis of equilibrium, from Theorem 2.4.3 of Nelsen [26], we know that the parameter vector $\boldsymbol{\theta}^0$ which characterizes the degree of affiliation in the joint distribution of valuations is the same parameter vector which characterizes the degree of affiliation in the joint distribution of equilibrium bids. Thus, we can focus on observed bids when estimating the dependence in unobserved valuations.

Thus, we estimate $G_0^0(s)$, the marginal cumulative distribution function of S using

$$\widetilde{G}_0(s) = \frac{1}{nT+1}\sum_{t=1}^{T}\sum_{i=1}^{n}\mathbf{1}(S_{it} \le s).$$

Here, $[1/(nT+1)]$ is used to scale the cumulative sum in the definition of the empirical distribution function to avoid boundary problems encountered when implementing the copulae. Subsequently, we insert $\widetilde{G}_0(\cdot)$ into the logarithm of the likelihood function and maximize with respect to the parameter vector $\boldsymbol{\theta}$.[2] Because $\widetilde{G}_0(\cdot)$ is different from the true population cumulative distribution function $G_0^0(\cdot)$, we refer to this method as *pseudo maximum-likelihood* (PML) estimation.

Of course, if C^0 depends on a parameter, then so too does the support of $G_0^0(\cdot)$, so one of the regularity conditions imposed by maximum-likelihood estimation is violated. To see this, consider the following fully parametric model where the marginal probability density function depends on a vector of unknown parameters $\boldsymbol{\gamma}$, so $f_0(v;\boldsymbol{\gamma})$. Under this assumption, not only does the marginal probability density function of S depend on $\boldsymbol{\gamma}$, it also depends on the copula parameter vector $\boldsymbol{\theta}$, so $g_0(s;\boldsymbol{\gamma},\boldsymbol{\theta})$. In addition, the upper bound of support of S depends on $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$, so

$$\underline{v} \le s \le \bar{s}(\boldsymbol{\gamma},\boldsymbol{\theta}) = \sigma(\bar{v})$$

where $\underline{v}$ and $\bar{v}$ are the lower and upper bounds of support of V, respectively.

What to do? Well, it turns out that, under our proposed approach, one can ignore this support problem. In fact, this is one of the most compelling features of our approach. To see how this works, consider the simple case where n is three. The joint density function of bids, can be

[2] Fermanian and Scaillet [10] have shown that the bias is smaller under a two-step approach than when the model is estimated at one go—i.e., when the parameters of the marginal distribution and those of the copula are estimated simultaneously. Furthermore, they have argued that there is “little to lose but lots to gain from shifting towards a semiparametric approach” in terms of efficiency.

written in terms of the copula and the marginal cumulative distribution and probability density functions of S at auction t as

$$C_{123}[G_0(s_{1t};\boldsymbol{\gamma},\boldsymbol{\theta}), G_0(s_{2t};\boldsymbol{\gamma},\boldsymbol{\theta}), G_0(s_{3t};\boldsymbol{\gamma},\boldsymbol{\theta}); \boldsymbol{\theta}]g_0(s_{1t};\boldsymbol{\gamma},\boldsymbol{\theta})g_0(s_{2t};\boldsymbol{\gamma},\boldsymbol{\theta})g_0(s_{3t};\boldsymbol{\gamma},\boldsymbol{\theta}),$$

the logarithm of which (when aggregated over the $t = 1, \ldots, T$ auctions) is then

$$\sum_{t=1}^{T} \quad \log\left(C_{123}[G_0(s_{1t};\boldsymbol{\gamma},\boldsymbol{\theta}), G_0(s_{2t};\boldsymbol{\gamma},\boldsymbol{\theta}), G_0(s_{3t};\boldsymbol{\gamma},\boldsymbol{\theta}); \boldsymbol{\theta}]\right) + \\ \log[g_0(s_{1t};\boldsymbol{\gamma},\boldsymbol{\theta})] + \log[g_0(s_{2t};\boldsymbol{\gamma},\boldsymbol{\theta})] + \log[g_0(s_{3t};\boldsymbol{\gamma},\boldsymbol{\theta})].$$

As mentioned above, one of the regularity conditions used to prove the parameter-consistency of the maximum-likelihood estimator is violated because $\bar{s}(\boldsymbol{\gamma},\boldsymbol{\theta})$, the upper bound of support of the strategy, depends on the parameters in the logarithm of the likelihood function.

Consider now our semiparametric, PML estimator. A major difference between our approach and the fully parametric approach is that we do not specify parametric functional forms for $F_0(\cdot)$ and, consequently, $G_0(\cdot)$. Instead, we use the observed bids to estimate $G_0(\cdot)$, which can then be used to estimate $g_0(\cdot)$. Using $\widetilde{G}_0(\cdot)$ and $\widetilde{g}_0(\cdot)$, we then maximimize, with respect to $\boldsymbol{\theta}$,

$$\sum_{t=1}^{T} \log\left(C_{123}[\widetilde{G}_0(s_{1t}), \widetilde{G}_0(s_{2t}), \widetilde{G}_0(s_{3t}); \boldsymbol{\theta}]\right)$$

instead of

$$\sum_{t=1}^{T} \log\left(C_{123}[\widetilde{G}_0(s_{1t}), \widetilde{G}_0(s_{2t}), \widetilde{G}_0(s_{3t}); \boldsymbol{\theta}]\right) + \log[\widetilde{g}_0(s_{1t})] + \log[\widetilde{g}_0(s_{2t})] + \log[\widetilde{g}_0(s_{3t})].$$

Clearly, the upper bound of support $\bar{s}(\boldsymbol{\gamma},\boldsymbol{\theta})$ still depends on $\boldsymbol{\theta}$, but this does not cause us any technical problems because, when we search over $\boldsymbol{\theta}$, $\widetilde{g}_0(s_{it})$ is always positive as it does not depend on $\boldsymbol{\theta}$, unlike in the fully parametric case, so it can effectively be ignored. In the fully parametric approach, values of $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$ can result in $g_0(s_{it};\boldsymbol{\gamma},\boldsymbol{\theta})$'s being zero, so its logarithm is then undefined.

5.1. Asymptotic Properties of the Estimator

As mentioned, given an initial estimate $\widetilde{G}_0(\cdot)$, we propose to estimate the copula dependence parameter by maximizing the following *pseudo log-likelihood* function:

$$\widetilde{\mathcal{L}}(\theta) = \frac{1}{T}\sum_{t=1}^{T} \log C_{1,\ldots,n}[\widetilde{G}_0(s_{1t}), \ldots, \widetilde{G}_0(s_{nt}); \boldsymbol{\theta}]$$

with respect to the unknown parameter vector $\boldsymbol{\theta}$, having constrained $\boldsymbol{\theta}$ to respect affiliation, MTP_2. Thus, the PML estimator $\hat{\boldsymbol{\theta}}$ is

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\ \widetilde{\mathcal{L}}(\boldsymbol{\theta}).$$

As noted by Chen and Fan [2], the main difficulty in establishing the asymptotic properties of $\hat{\boldsymbol{\theta}}$ is that the score function and its derivatives can approach infinity (become undefined) at the boundaries of the space under alternative choices of copula functions. To circumvent this problem, Chen and Fan [2] introduced a weighting function. Their approach is as follows: first, they established convergence of the non-parametric estimator for the marginal distribution in a weighted metric and, then, they established the asymptotic properties of the copula dependence parameter(s). Specifically, the weighting function is constructed so that it equals one when the score function and its derivatives are defined at the boundary; otherwise, the weighting function is defined to be some smooth function that becomes zero at the boundaries—zero and one. Although Chen and Fan [2] investigated the estimation of copula-based semiparametric time-series models using the copula to model the joint distribution of the time series Y_t and Y_{t-1}, leaving the marginal distribution unspecified, their weighting-function approach can be applied in our case. We make the following assumptions that apply to our auction application:

A1. $\{(S_{1t}, \ldots, S_{nt})\}_{t=1}^{T}$ is a random sample from the joint distribution modelled by the copula $C[G_0^0(\cdot), \ldots, G_0^0(\cdot); \boldsymbol{\theta}^0]$ where $G_0^0(\cdot)$ is absolutely continuous with respect to Lebesgue measure on the real line and $C(\cdot, \ldots, \cdot; \boldsymbol{\theta}^0)$ is the true parametric copula for $(S_{1t}, \ldots, S_{nt})$, which is absolutely continuous with respect to Lebesgue measure on $[0, 1]^n$, and does not attain either the lower or the upper bounds of Fréchet–Hoeffding described in section 3.

A2. $\boldsymbol{\theta}^0 \in \boldsymbol{\Theta}$, a compact set in $\mathbb{R}^p$, and $\mathcal{E}[\ell_{\boldsymbol{\theta}}(U_1, \ldots, U_n; \boldsymbol{\theta}^0)]$ equals zero, if and only if $\boldsymbol{\theta}$ equals $\boldsymbol{\theta}^0$, where

$$\ell_{\boldsymbol{\theta}}(u_1, \ldots, u_n; \boldsymbol{\theta}) \equiv \nabla_{\boldsymbol{\theta}} \log C_{1,\ldots,n}(u_1, \ldots, u_n; \boldsymbol{\theta}).$$

A3. $\ell_{\boldsymbol{\theta}}(u_1, \ldots, u_n; \boldsymbol{\theta})$ is defined for $(u_1, \ldots, u_n) \in [0, 1]^n \times \boldsymbol{\Theta}$ and for all $\boldsymbol{\theta}^0 \in \boldsymbol{\Theta}$; $\ell_{\boldsymbol{\theta}}(u_1, \ldots, u_n; \boldsymbol{\theta})$ is Lipschitz continuous at $\boldsymbol{\theta}$ almost surely; $\ell_{\boldsymbol{\theta},i}(u_1, \ldots, u_n; \boldsymbol{\theta})$ for $i = 1, \ldots, n$ are defined and continuous in $(u_1, \ldots, u_n; \boldsymbol{\theta}) \in [0, 1]^n \times \boldsymbol{\Theta}$.

A4. $\mathcal{E}\{\sup_{\theta \in \Theta} \|\ell_{\boldsymbol{\theta}}(U_1, \ldots, U_n; \boldsymbol{\theta})\| \log[1 + \|\ell_{\boldsymbol{\theta}}(U_1, \ldots, U_n; \boldsymbol{\theta})\|]\} < \infty$.

A5. $\mathcal{E}\{\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}, G \in \mathcal{G}_\delta} \|\ell_{\boldsymbol{\theta},i}[G(S_{1t}), \ldots, G(S_{nt}); \boldsymbol{\theta}]\| w(U_i)\} < \infty$ where $\mathcal{G}_\delta$ equals $\{G \in \mathcal{G} : \|G - G_0^0\|_{\mathcal{G}} \leq \delta\}$ with $\mathcal{G}$ being the space of continuous probability distributions over the support of S_{it}, and $w(v)$ denotes $[v(1 - v)]^{1-\xi}$ for $v \in (0, 1)$ and $\xi \in (0, 1)$.

Note that these assumptions are adapted from those in Chen and Fan [2]. While Chen and Fan considered time-series data with two variables Y_t and Y_{t-1}, we consider data from a cross-section of n variables $(S_{1t}, \ldots, S_{nt})$. The following proposition states the consistency result of $\hat{\boldsymbol{\theta}}$.

Proposition 1. *Under A1–A5,*

$$\hat{\boldsymbol{\theta}} \xrightarrow{\mathrm{P}} \boldsymbol{\theta}^0.$$

The proof of Proposition 1, which follows closely the proof of Proposition 4.2 in Chen and Fan [2], is omitted as it is straightforward. For the asymptotic distribution, following Chen and Fan [2], we make the following assumptions:

A6. (i) A2 is satisfied with $\boldsymbol{\theta}^0$ in the interior of $\boldsymbol{\Theta}$; (ii) $\mathbf{B} \equiv -\mathcal{E}[\ell_{\boldsymbol{\theta},\boldsymbol{\theta}}(U_1, \ldots, U_n; \theta^0)]$ is positive definite; (iii) $\boldsymbol{\Sigma} \equiv \lim_{T\to\infty} \mathrm{Var}(\sqrt{T}\mathbf{A}_T)$ is positive definite where

$$\mathbf{A}_T \equiv \frac{1}{T}\sum_{t=1}^{T}\left[\ell_{\boldsymbol{\theta}}(U_1, \ldots, U_n; \boldsymbol{\theta}^0) + \sum_{i=1}^{n} \mathcal{W}_i(U_i)\right]$$

with

$$\mathcal{W}_i(U_i) = \int_0^1 \cdots \int_0^1 [\mathbf{1}(U_i \le v_i) - v_i]\ell_{\boldsymbol{\theta},i}(v_1, \ldots, v_n; \boldsymbol{\theta}^0)C_{1,\ldots,n}(v_1, \ldots, v_n; \boldsymbol{\theta}^0)\,\mathrm{d}v_1 \cdots \mathrm{d}v_n;$$

(iv) $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^0 + o_P(1)$ and $\sup_s |[\widetilde{G}_0(s) - G_0^0(s)]/w[G_0^0(s)]| = O_P(T^{-1/2})$ where $w(\cdot)$ is defined in A5.

A7. $\ell_{\boldsymbol{\theta},\boldsymbol{\theta}}(u_1, \ldots, u_n; \boldsymbol{\theta})$ and $\ell_{\boldsymbol{\theta},i}(u_1, \ldots, u_n; \boldsymbol{\theta})$ for $i = 1, \ldots, n$ are all defined, and continuous in $(u_1, \ldots, u_n, \boldsymbol{\theta}) \in [0, 1]^n \times \mathrm{Int}(\boldsymbol{\Theta})$.

A8. It is valid to interchange the order of differentiation and integration of $\ell_{\boldsymbol{\theta}}[G_\eta(s_{1t}), \ldots, G_\eta(s_{nt}); \boldsymbol{\theta}_\eta]$ with respect to $\eta \in (0, 1)$ where $\{(\boldsymbol{\theta}_\eta, G_\eta) : \eta \in [0, 1]\} \subset \mathcal{F}_\delta \equiv \{(\boldsymbol{\theta}, G) \in \boldsymbol{\Theta} \times \mathcal{G}_\delta : \|\boldsymbol{\theta} - \boldsymbol{\theta}^0\| \le \delta\}$ for a small δ, is a one-dimensional smooth path in $\mathcal{F}_\delta$.

A9. Both $\mathcal{E}\{\sup_{(\boldsymbol{\theta},G)\in\mathcal{F}_\delta} \|\ell_{\boldsymbol{\theta}}[G(S_{1t}), \ldots, G(S_{nt}); \boldsymbol{\theta}]\|\}^2 < \infty$ and $\mathcal{E}\{\| \sum_{i=1}^n \mathcal{W}_i(U_i)\|^2\} < \infty$.

A10. $\mathcal{E}\{\sup_{(\boldsymbol{\theta},G)\in\mathcal{F}_\delta} \|\ell_{\boldsymbol{\theta},\boldsymbol{\theta}}[G(S_{1t}), \ldots, G(S_{nt}); \boldsymbol{\theta}]\|\}^2 < \infty$.

A11. $\mathcal{E}\{\sup_{(\boldsymbol{\theta},G)\in\mathcal{F}_\delta} \|\ell_{\boldsymbol{\theta},i}[G(S_{1t}), \ldots, G(S_{nt}); \boldsymbol{\theta}]\|w(U_i)\}^2 < \infty$ for $i = 1, \ldots, n$.

These assumptions are modified versions of those in Chen and Fan [2] because we consider data from a cross-section case, while they consider data from a time-series. We can now state the following result concerning the $\sqrt{T}$ asymptotic normality of $\hat{\boldsymbol{\theta}}$.

Proposition 2. *Under A6–A11,*

$$\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \xrightarrow{\mathrm{d}} \mathcal{N}(\mathbf{0}, \mathbf{B}^{-1}\boldsymbol{\Sigma}\mathbf{B}^{-1}).$$

Were $G_0^0(\cdot)$ known, then $\sum_{i=1}^n \mathcal{W}_i(U_i)$ would disappear from $\mathbf{A}_T$ and, thus, from $\boldsymbol{\Sigma}$. In other words, the term $\sum_{i=1}^n \mathcal{W}_i(U_i)$ is introduced because $G_0^0(\cdot)$ is unknown and must be estimated.

For ease of exposition, the preceding asymptotic results were established for the case where auctioned objects are homogenous. The results can, however, be readily extended to the case where the auctioned objects are heterogenous and are conditioned on a d-dimensional vector of covariates $\boldsymbol{X}$.

5.2. Optimal Uniform Convergence Rate: Estimator of f_V

Having obtained the PML estimator $\hat{\boldsymbol{\theta}}$, we can now calculate the pseudo private-values. To this end, in addition to the non-parametric estimator of the marginal cumulative distribution, we define the non-parametric estimator of the probability density function as follows:

$$\tilde{g}_0(s) = \frac{1}{nTh_g} \sum_{t=1}^{T} \sum_{i=1}^{n} \kappa_g \left(\frac{s - S_{it}}{h_g} \right)$$

where h_g is a bandwidth and $\kappa_g(\cdot)$ is a kernel with a compact support whose length is ρ_g. Let $S_{\min}$ and $S_{\max}$ denote the minimum and maximum of the nT observed bids, and define the pseudo private-value corresponding to S_{it} as

$$\widetilde{V}_{it} = \begin{cases} S_{it} + \frac{\widetilde{C}_1[\widetilde{G}_0(S_{it}),\ldots,\widetilde{G}_0(S_{it})]}{(n-1)\tilde{g}_0(S_{it})\widetilde{C}_{12}[\widetilde{G}_0(S_{it}),\ldots,\widetilde{G}_0(S_{it})]} & \text{if } S_{\min} + \rho_g h_g/2 \le S_{it} \le S_{\max} - \rho_g h_g/2 \\ \infty \text{ otherwise.} \end{cases}$$

Note that, here, the trimming, which follows GPV, is necessary because $\tilde{g}_0(\cdot)$ is biased near the boundaries. The pseudo private-values $\widetilde{V}_{it}$ can then be used to estimate the marginal probability density function of private values via

$$\tilde{f}_0(v) = \frac{1}{nTh_f} \sum_{t=1}^{T} \sum_{i=1}^{n} \kappa_f \left(\frac{v - \widetilde{V}_{it}}{h_f} \right)$$

where h_f is a bandwidth and $\kappa_f(\cdot)$ is a kernel with compact support.

Since we can also estimate the marginal distribution of private values by

$$\widetilde{F}_0(v) = \frac{1}{nT+1} \sum_{t=1}^{T} \sum_{i=1}^{n} \mathbf{1}(\widetilde{V}_{it} \le v),$$

we can now estimate the joint density of n private values by

$$\begin{aligned} \tilde{f}_V(v_1,\ldots,v_n) &= C_{1,\ldots,n}[\widetilde{F}_0(v_1),\ldots,\widetilde{F}_0(v_n);\hat{\boldsymbol{\theta}}] \prod_{i=1}^{n} \tilde{f}_0(v_i) \\ &\equiv \widetilde{C}_{1,\ldots,n}[\widetilde{F}_0(v_1),\ldots,\widetilde{F}_0(v_n)] \prod_{i=1}^{n} \tilde{f}_0(v_i). \end{aligned}$$

GPV studied the optimal rate of uniform convergence of the non-parametric density estimator $\tilde{f}_0(\cdot)$ when the n private values are independent. In this case, the private value can be expressed as

$$V_{it} = S_{it} + \frac{G_0(S_{it})}{(n-1)g_0(S_{it})}$$

and the pseudo private-values can be defined accordingly. It turns out that under suitable regularity conditions, we can establish the optimal rate of uniform convergence for $\tilde{f}_0(\cdot)$, which is given in the next proposition.

Proposition 3. *When (i) $f_0(\cdot)$ has R bounded, continuous derivatives inside its support; (ii)*

$$H(u,\boldsymbol{\theta}) \equiv \frac{C_1(u,\ldots,u;\boldsymbol{\theta})}{C_{12}(u,\ldots,u;\boldsymbol{\theta})}$$

is continuously differentiable in both u and $\boldsymbol{\theta}$ in $(0,1)\times\mathrm{Int}(\boldsymbol{\Theta})$ *with all derivatives being bounded in absolute value; and (iii)* $h_g = c_g(\log T/T)^{1/(2R+3)}$ *with* $h_f = c_f(\log T/T)^{1/(2R+3)}$*, then*

$$\sup_{v\in\zeta(V)} |\tilde{f}_0(v) - f_0(v)| = O((\log T/T)^{R/(2R+3)}) \text{ almost surely}$$

for any closed inner subset $\zeta(V)$ *of the support of* $f_0(\cdot)$ *where* $(\log T/T)^{R/(2R+3)}$ *is indeed the optimal uniform convergence rate for two-step non-parametric estimators of the density* $f_0(\cdot)$ *from the observed bids.*

Proposition 3 follows directly from Theorems 2 and 3 in GPV because of the differentiability of $H(u,\boldsymbol{\theta})$ assumed in (ii) above. A comparison of our estimator $\tilde{f}_0(\cdot)$ to the one in GPV indicates that our estimator behaves (asymptotically) in the same way as the one in GPV provided that (ii) is met: once this condition is satisfied, $\{\widetilde{C}_1[\widetilde{G}_0(S_{it}),\ldots,\widetilde{G}_0(S_{it})]/\widetilde{C}_{12}[\widetilde{G}_0(S_{it}),\ldots,\widetilde{G}_0(S_{it})]\}$, which is now $H[\widetilde{G}_0(S_{it}),\hat{\boldsymbol{\theta}}]$, behaves like $\widetilde{G}_0(S_{it})$, at least asymptotically, so

$$S_{it} + \frac{1}{(n-1)\tilde{g}_0(S_{it})}\frac{\widetilde{C}_1[\widetilde{G}_0(S_{it}),\ldots,\widetilde{G}_0(S_{it})]}{\widetilde{C}_{12}[\widetilde{G}_0(S_{it}),\ldots,\widetilde{G}_0(S_{it})]}$$

behaves like

$$S_{it} + \frac{\widetilde{G}_0(S_{it})}{(n-1)\tilde{g}_0(S_{it})},$$

at least asymptotically. In particular, as in Lemma B2 in GPV, we can show that the uniform rate of convergence of $H[\widetilde{G}_0(S_{it}),\hat{\boldsymbol{\theta}}]$ over an expanding subset is the same as that of $\widetilde{G}_0(S_{it})$, which is $(\log T/T)^{(R+1)/(2R+3)}$. Furthermore, because of this proposition, and the way in which we estimate the joint density of private values, we can obtain the same optimal uniform rate of convergence for the joint density estimator of private values, which is stated in the following corollary.

Corollary 1. *Under the assumptions of Proposition 3, the optimal uniform convergence rate of* $\tilde{f}_V(\cdot,\ldots,\cdot)$ *is also* $(\log T/T)^{R/(2R+3)}$.

The result given in Corollary 1 is worth some discussion. It implies that the joint density of private values can be estimated at the same rate as the marginal density of private values. The semiparametric nature of our approach—imposing a parametric copula specification for the joint distribution of private values (and, hence, bids) while leaving the marginal distribution of bids unspecified—delivers sufficient structure to guarantee this rate of convergence. Thus, the two-step nature of our estimation strategy parallels that considered by GPV. Consequently, the convergence rate is faster than the rate of the fully non-parametric LPV estimator.

Following Theorems 2 and 3 of GPV, Proposition 3 and Corollary 1 can be readily extended to the case of heterogenous objects; i.e., the probability density and cumulative distribution functions can be conditioned on a d-dimensional vector of covariates, viz., $\boldsymbol{X}$. Then, the optimal uniform convergence rates for both $\tilde{f}_0(\cdot|\cdot)$ and $\tilde{f}_V(\cdot,\ldots,\cdot|\cdot)$ are $(\log T/T)^{R/(2R+d+3)}$. Of course, in this case, the non-parametric estimators of $\widetilde{G}_0(s|\boldsymbol{x})$, $\tilde{g}_0(s|\boldsymbol{x})$, and $\tilde{f}_0(v|\boldsymbol{x})$ need to be accordingly modified following GPV.

As mentioned above, the LPV estimator suffers from the curse of dimensionality in two ways. The first is the dimension n—viz., the number of bidders at the auction, which determines the dimension of joint density of private values; the second is d, the dimension of the covariate vector. We have demonstrated that, under our approach, the optimal rate of convergence obtains when a parametric copula is used because this reduces the n dimension to one; the d dimension still remains. Our method is especially useful when the number of bidders is large. Our estimator, however, still does not address the problem caused by a large number of covariates. In this regard, it would be useful to consider possible extensions of our semiparametric estimator. One such extension would involve a single index model, which can be used to reduce the curse of dimensionality introduced by d. Of course, the price of the single-index assumption is a reduction of flexibility and generality.

Note, too, that when estimating the underlying distribution of private values using the observed bids, our semiparametric approach specifies a parametric copula with a dependence parameter, while nonparametrically estimating the marginal distribution of the private values. Therefore, our semiparametric estimator is different from most of the semiparametric estimators derived within a regression framework; for example those studied by Newey and McFadden [27]. The objective in that reseearch is to estimate the parametric part of the model, while treating the nonparametric part as a nuisance parameter (albeit of an infinite dimension). Despite this difference, our estimator of the dependence parameter attains $\sqrt{T}$ rate of convergence (see Proposition 2), which is the common rate for the class of semiparametric estimators studied by Newey and McFadden [27].

6. Some Monte Carlo Results

Below, we describe a Monte Carlo experiment designed to shed light on the small-sample properties of our estimation strategy. In the tradition of the theoretical literature concerning auctions, our model of bidding in section 4 was developed in terms of valuations for an object to be sold at auction under the first-price, sealed-bid format. Sealed-bid tenders are often used in procurement—i.e., low-price, sealed-bid auctions at which a buyer (often a government agency) seeks to find the lowest-cost producer of some good or service. Because our empirical example in section 7 involves investigating procurement of road resurfacing by a government agency, our simulation study is couched in terms of a procurement auction. For the case of low-price, sealed-bid procurement auctions, we collect several lemmata and their proofs in an appendix to this paper.

In all of the experiments, the simulated data involved a truncated Pareto random variable C (for costs) having the following marginal cumulative distribution function:

$$F_0(c) = \frac{F_C(c) - F_C(\underline{c})}{[F_C(\overline{c}) - F_C(\underline{c})]} \tag{6}$$

where

$$F_C(c) = 1 - \left(\frac{\gamma_0}{c}\right)^{\gamma_1} \quad 0 < \gamma_0 \leq c,\ 1 < \gamma_1.$$

Our simulation study involved samples of size T equal fifty, one hundred, and two hundred with n of three bidders; each sample was replicated $1,000$ times. The lower bound of support, the lowest cost $\underline{c}$, was one, while the upper bound of support, the highest cost $\overline{c}$, was three. The parameters of the Pareto distribution were γ_0 equal one and γ_1 equal two.

To compare estimators, we considered the strategies proposed by GPV and LPV as well as our copula approach for three members of the Archimedean family—the Clayton, Frank, and Gumbel copulae, which are the most frequently used copulae in empirical applications. The Archimedean family was discussed in section 3 and copula-specific equations needed for estimation, such as the survival representation, are presented in an appendix to this paper.

6.1. Performance using Independent Data

While the focus of our research concerns auctions in the presence of affiliation, it is useful to compare the performance of the estimation strategies when signals are independent. Such an analysis provides a benchmark to contrast the performance of a given strategy when dependence is introduced. In addition, researchers may not know *ex ante* whether a given data set contains dependence, or not, so there is value in investigating the performance of our semiparametric estimator, which nests independence as a special case.

In the case of independence, we first generated uniform draws from $U(0,1)$. To convert the i^{th} uniform draw u_i into a draw from the truncated Pareto distribution c_i, we simply inverted the cumulative distribution according to the following formula:

$$c_i = \frac{\gamma_0}{\left(1 - u_i\left[F_C(\overline{c}) - F_C(\underline{c})\right] - F_C(\underline{c})\right)^{1/\gamma_1}}.$$

Simulated costs c_i were then mapped into simulated bids b_i using the following symmetric equilibrium bid function:

$$b_i = \beta(c_i) = c_i + \frac{\int_{c_i}^{\overline{c}} [1 - F_0(u)]^{n-1}\, du}{[1 - F_0(c_i)]^{n-1}}$$

where $F_0(\cdot)$ is the truncated Pareto distribution specified in equation (6) and the integral was computed numerically using quadrature.

The estimation procedures proposed by GPV and LPV are discussed in detail in their respective papers. When implementing their estimators, we employed the kernel function and adopted the optimal bandwidths they suggested. To implement our copula approach, we first modified the approach of GPV to estimate $G_0(b)$ and $g_0(b)$ non-parametrically using the following estimators:

$$\widetilde{G}_0(b) = \frac{1}{nT+1} \sum_{t=1}^{T} \sum_{i=1}^{n} \mathbf{1}\left(b_{it} \le b\right) \tag{7}$$

and

$$\tilde{g}_0(b) = \frac{1}{(nT+1)h} \sum_{t=1}^{T} \sum_{i=1}^{n} \kappa\left(\frac{b - b_{it}}{h}\right). \tag{8}$$

Again, the division by $(nT+1)$ rather than nT is a rescaling to avoid numerical complications arising at the boundary of the copula; e.g., recall (from section 3 above) that $\mathcal{C}(1, u_2)$ equals u_2 and $\mathcal{C}(u_1, 1)$ equals u_1. We employed the triweight kernel

$$\kappa(u) = \frac{35}{32}\left(1 - u^2\right)^3 \mathbf{1}(|u| \le 1)$$

with bandwidth

$$h = d\left(\frac{4}{3}\right)^{1/5} \hat{\sigma}(nT+1)^{-1/5}$$

where d equal 2.978 is the bandwidth transformation constant from Härdle [15] and $\hat{\sigma}$ was the standard deviation of $\boldsymbol{b}$ which is the vector collecting the data $\{\{b_{it}\}_{i=1}^{n}\}_{t=1}^{T}$. We then estimated θ by the method of PML using the following pseudo log-likelihood function:

$$\widetilde{\mathcal{L}}(\theta; \boldsymbol{b}) = \sum_{t=1}^{T} \log\left(\boldsymbol{c}_\zeta\left[\widetilde{G}_0(b_{1t}), \widetilde{G}_0(b_{2t}), \widetilde{G}_0(b_{3t}); \theta\right]\right)$$

Table 2: Performance of Methods using IID Data, MSEP

Method	$T = 50$	$T = 100$	$T = 200$
GPV	0.00188	0.00105	0.00058
LPV	n/a	3.15053	0.02362
LPV (log transform)	1.13816	0.00945	0.00603
Clayton	0.00176	0.00100	0.00055
Frank	0.00185	0.00103	0.00057
Gumbel	0.00194	0.00111	0.00065

where

$$\boldsymbol{c}_\zeta(u_1, u_2, u_3) = \frac{\partial^3 C_\zeta(u_1, u_2, u_3)}{\partial u_1 \partial u_2 \partial u_3}.$$

Thus, the PML estimator $\hat{\theta}$ is defined by

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\ \widetilde{\mathcal{L}}(\theta; \boldsymbol{b}).$$

Using $\hat{\theta}$ as well as $\widetilde{G}_0(b)$ and $\tilde{g}_0(b)$ from equations (7) and (8), we then computed estimates of the partial derivatives of each respective survival copula according to

$$\widetilde{\mathcal{S}}_1(b_i) = \mathcal{S}_1 \left[1 - \widetilde{G}_0(b_i), 1 - \widetilde{G}_0(b_i), 1 - \widetilde{G}_0(b_i); \hat{\theta}\right]$$

and

$$\widetilde{\mathcal{S}}_{12}(b_i) = \mathcal{S}_{12}[1 - \widetilde{G}_0(b_i), 1 - \widetilde{G}_0(b_i), 1 - \widetilde{G}_0(b_i); \hat{\theta}]$$

when we were interested in recovering the pseudo-cost $\tilde{c}_i$ associated with bid b_i. We then used our $\tilde{g}_0(b)$ from equation (8) in conjunction with $\widetilde{\mathcal{S}}_1(b)$ and $\widetilde{\mathcal{S}}_{12}(b)$ to compute the pseudo-cost associated with any bid b

$$\tilde{c} = b - \frac{\widetilde{\mathcal{S}}_1(b)}{(n-1)\tilde{g}_0(b)\widetilde{\mathcal{S}}_{12}(b)}. \tag{9}$$

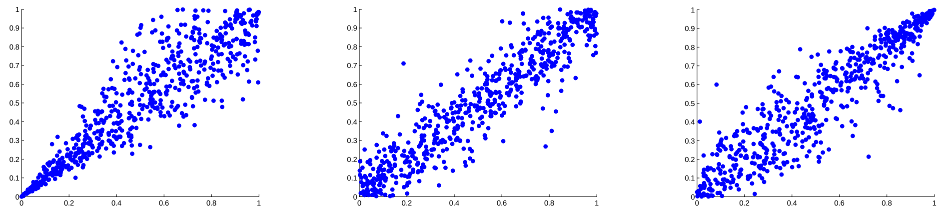
To account for biases, near the boundaries, that obtain because we kernel-smoothed the density, we only used the bid function in the prior step to recover the cost if the observed bid was in the range $[\underline{b} + h, \overline{b} - h]$ where $\underline{b}$ was $\min\{b_{it}\}$ and $\overline{b}$ was $\max\{b_{it}\}$.

By construction, in the Monte Carlo study, we knew the true costs associated with the simulated data $\boldsymbol{b}$. Each estimation strategy was used to recover a predicted pseudo-cost associated with each bid. This allowed us to evaluate the error associated with each of the predicted values; i.e., the error for the i^{th} cost is $(c_i - \tilde{c}_i)$. To compare the performance of the estimation strategies, we computed the mean squared error of the prediction (MSEP) using

$$\text{MSEP} = \sum_{i=1}^{M} \frac{(c_i - \tilde{c}_i)^2}{M}$$

where M is the number of costs that survived the trimming described above for the particular estimation strategy, in a given simulation.

In table 2, we present the MSEPs that obtained using the estimation strategies with independently- and identically-distributed (IID) data, where the Monte Carlo samples were held fixed across all



Clayton Frank Gumbel

Figure 1: Types of Dependence in Archimedean Copulae

methods. As expected, the MSEPs for all estimation strategies decrease as the sample size increases, which can be seen by comparing the three right-most columns of the table along a given row. Our results indicate that the copula estimator performed as well as the GPV estimator which was designed for IID data. The MSEPs associated with the copula approaches, regardless of the copula chosen, are comparable to the MSEPs obtained using the GPV approach. We considered the LPV approach using the raw bids as well as the logarithm of bids, which is what the authors suggested researchers consider when the distribution of bids is skewed. We found that the LPV approach had difficulty with IID data when T was small (for example, 50) because the kernel density took values close to zero and resulted in exceedingly high pseudo-costs. However, even for the largest sample size, which involved 200 auctions, the MSEP obtained using the LPV approach was forty times that of GPV or our copula approach. Performance of the LPV approach improved substantially when the logarithmic transformation was used and the pseudo costs were recovered using an appropriate modification of equation (9).

6.2. Performance using Affiliated Data

In our simulation study, with affiliated data, we considered data generated under different "levels" of dependence as well as different "types" of dependence. We generated dependent data from the Clayton, Frank, and Gumbel copulae to evaluate the performance of our method using data that exhibited different types of dependence. For example, in figure 1, we depict bivariate draws generated from Clayton, Frank, and Gumbel copulae. The figures illustrate the type of dependence inherent in the data varies depending on the copula used.[3] In the scatterplots, and in our simulation study, we fixed the level of dependence according to Kendall's τ—a non-parametric measure of association which is defined in terms of concordance. As was demonstrated by Genest and MacKay [12], a direct relationship exists between Kendall's τ and the parameter of an Archimedean copula. In table 3, we list the correspondence between Kendall's τ and the parameter of each Archimedean copula we considered, here denoted θ. We exploited these relationships to provide structure in our simulation study: they allowed us to fix the level of dependence across generating copulae by computing the copula parameters associated with given choices of Kendall's τ. For example, in figure 1, the parameters of the Clayton, Frank, and Gumbel copulae were chosen to correspond with a value for Kendall's τ of 0.75.

The discussion above centered around bivariate copulae, while in our simulation study we considered auctions with three bidders. However, the relationship between Kendall's τ and

[3]The type of dependence is easier to see using bivariate draws, which can be plotted in two dimensions. Of course, the relationships apply to multivariate draws, as will become clear in the discussion that follows.

Table 3: Kendall's τ and Archimedean Copula Parameters

Copula	Relationship
Clayton	$\tau = \frac{\theta}{\theta+2}$
Gumbel	$\tau = 1 - \frac{1}{\theta}$
Frank*	$\tau = 1 - \frac{4}{\theta}[D_1(\theta) - 1]$

* $D_1(\theta) = \frac{1}{\theta}\int_0^{\theta} \frac{x}{\exp(x)-1}\,dx$

Archimedean copulae derived by Genest and MacKay [12] can be applied to the multivariate copulae we considered. Because we are concerned with the symmetric affiliated private-cost paradigm (APCP), the marginal distribution functions of each random variable are identically distributed. Note, too, that a copula is considered *exchangeable* if the marginal distribution functions are the same and the arguments are treated equivalently by the copula; i.e., in the bivariate case,

$$C(u_1, u_2) = C(u_2, u_1)$$

given U_1 and U_2 are identically distributed. Thus, exchangeability is a form of symmetry. Jouini and Clemen [16] have demonstrated that, in the bivariate case, Kendall's τ is sufficient to construct an n-dimensional, exchangeable copula. This implies the unique relationships presented in table 3 can be used to relate Kendall's τ to the parameter of the multivariate Archimedean copulae of interest, all of which are exchangeable.

To simulate data from the Clayton, Frank, or Gumbel copulae with dependence, we followed the approach described by Cherubini, Luciano, and Vecchiato [4]. Theirs involves *conditional sampling* where, initially, w_1, a $U(0,1)$ random draw, is taken and then u_1 is set equal to it. The next (dependent) draw u_2 is found by solving

$$w_2 = C_2(u_2|u_1),$$

and so on, so that u_n is found by solving

$$w_n = C_n(u_n|u_1, \ldots, u_{n-1})$$

where all the w_is are independent $U(0,1)$ draws. Because the inverse function does not have a closed-form for some copulae, the procedure is computationally intensive because the root of each copula, conditional on the previous draws, must be solved numerically.

After converting the dependent draws into costs from the truncated Pareto distribution $F_0(\cdot)$, the simulated costs c_i were then mapped into simulated bids b_i using the following first-order condition for profit maximization:

$$\beta'(c_i) = [\beta(c_i) - c_i]\,\frac{(n-1)f_0(c_i)\mathcal{S}_{12}\,[1 - F_0(c_i), 1 - F_0(c_i), 1 - F_0(c_i)]}{\mathcal{S}_1\,[1 - F_0(c_i), 1 - F_0(c_i), 1 - F_0(c_i)]}$$

where $\beta(\cdot)$ is the equilibrium bid function. This first-order condition is a standard ordinary differential equation which, subject to the boundary condition

$$\beta(\overline{c}) = \overline{c},$$

yields the symmetric equilibrium bid function under affiliation

$$b_i = \beta(c_i) = c_i + \int_{c_i}^{\bar{c}} \exp\left(-\int_{c_i}^{u} \frac{(n-1) f_0(w) \mathcal{S}_{12} \left[1 - F_0(w), 1 - F_0(w), 1 - F_0(w)\right]}{\mathcal{S}_1 \left[1 - F_0(w), 1 - F_0(w), 1 - F_0(w)\right]} \, \mathrm{d}w\right) \mathrm{d}u.$$

The results from the discussion above allowed us to conduct a Monte Carlo study using dependent data in which we varied the number of auctions, the copula used to generate the dependent data, and the amount of dependence inherent in the generating copula. Kendall's τ provided a convenient way of assigning the same level of dependence to simulations generated under different copulae. Again, we compared the performance of the GPV and the LPV estimation strategies as well as our copula approach in which the Clayton, Frank, and Gumbel copulae were chosen for estimation, regardless of the generating copula. This exercise also provides insight into the question of which copula should be chosen in empirical applications and helps to determine the importance of copula choice in these auction models; cf. Fermanian [9] for goodness-of-fit tests for copulae.

In the next three tables, we present the MSEPs obtained using each of the estimation strategies with dependent data generated from a specific Archimedean copula. In particular, in table 4, we present results when data were generated from the Clayton copula. Likewise, in tables 5 and 6, we present the MSEPs when data were generated from the Frank and Gumbel copulae, respectively. In each table, the first column indicates the level of dependence, as measured by Kendall's τ. We generated data from each copula using the copula-specific parameter corresponding to Kendall's τ equal to 0.25, 0.50, and 0.75. The second column in each table describes the estimation strategy used, while the last three columns present the MSEPs for different samples sizes: T equal to fifty, one hundred, and two hundred auctions each having n equal to three bidders.

Each of the tables reveals that the GPV estimation strategy is an inappropriate one when signals are affiliated: as the dependence increases, the MSEPs increase. In addition, for a given level of dependence, increasing the sample size often yielded an higher MSEP under GPV. This is clearly unfortunate, but perhaps not surprising: under GPV, in the presence of affiliation, the measurement equation is mis-specified.

In contrast, the simulation results for the LPV non-parametric estimation strategy, which was designed to account for affiliation, show promise. Regardless of the copula used to generate the data or the level of dependence in the data, the performance of the LPV method improved as the sample size increased. Furthermore, the MSEPs obtained using the LPV method on dependent data were substantially lower than those found for that method when independent signals were used.

Interestingly, this does not hold for the LPV approach when a logarithmic transformation of the bids was used. While the transformation was attractive for independent data, the use of the raw bids in the LPV approach often outperformed the log-transformed bids, at least in terms of MSEP, for dependent data. In fact, the MSEP for the LPV approach when no transformation was used is always lower than that of the logarithmic transformation for τ equal 0.75, regardless of the copula used to generate the data. The logarithmic transformation was introduced so that fewer data would be trimmed for skewed bid distributions. With highly-dependent data, this rationale may no longer hold as one high (low) bid means other bids are more likely to be high (low). Consequently, the logarithmic transformation may result in an increase in the number of bids that are trimmed, which is what happened in the Monte Carlo study.

Perhaps not surprisingly, for any given simulation, the lowest MSEPs of all the estimation strategies obtained for our copula approach when it was used to estimate the model that generated

Table 4: Performance of Methods using Data Generated from Clayton Copula, MSEP

Dependence	Method	$T = 50$	$T = 100$	$T = 200$
	GPV	0.00289	0.00189	0.00143
	LPV	0.01079	0.00558	0.00341
$\tau = 0.25$	LPV (log transform)	0.00675	0.00352	0.00204
	Clayton	0.00185	0.00104	0.00059
	Frank	0.00195	0.00113	0.00070
	Gumbel	0.00215	0.00154	0.00114
	GPV	0.00386	0.00374	0.00487
	LPV	0.00682	0.00372	0.00228
$\tau = 0.50$	LPV (log transform)	0.00705	0.00487	0.00357
	Clayton	0.00187	0.00104	0.00057
	Frank	0.00229	0.00172	0.00136
	Gumbel	0.00333	0.00246	0.00199
	GPV	0.01033	0.01271	0.01613
	LPV	0.01364	0.01222	0.00980
$\tau = 0.75$	LPV (log transform)	0.02240	0.02343	0.01953
	Clayton	0.00267	0.00131	0.00064
	Frank	0.00481	0.00433	0.00382
	Gumbel	0.00641	0.00583	0.00545

the data. Note, too, that our copula approach always performed well when either the Frank or the Gumbel copulae were used in estimation, regardless of the copula that generated the data. The type of dependence inherent in the Clayton copula is quite specific, as illustrated in figure 1. Because of this, the simulation results show that when our copula approach used the Clayton copula in the estimation, but the data were not generated from a Clayton copula, the MSEPs were closer to those obtained using the LPV method.

7. An Empirical Application

To illustrate the feasibility of our estimation strategy, we have chosen to implement it using data from low-price, sealed-bid, procurement auctions held by the Department of Transportation in the State of Michigan. At these auctions, qualified firms are invited to bid on jobs that involve resurfacing roads in Michigan. We have chosen this type of auction because the task at hand is quite well-understood. In addition, there are reasons to believe that firm-specific characteristics make the private-cost paradigm a reasonable assumption; e.g., managerial ability at specific firms can differ considerably. On the other hand, other factors suggest that the cost signals of individual bidders could be dependent, even affiliated; e.g., these firms hire labour services in the same market and face similar costs for inputs, such as energy as well as paving inputs. Thus, the APCP seems reasonable. We have eschewed investigating issues relating to asymmetries across bidders (i.e., introducing F_is that vary across bidders) because we simply have insufficient data to identify such models. Instead, we focus on the symmetric APCP outlined above and developed further in the first section of the appendix to this paper. As no reserve price exists at these auctions, we treat the number of participants as if it were the number of potential bidders and focus on auctions at which three bidders participated. In short, we are ignoring the potential

Table 5: Performance of Methods using Data Generated from Frank Copula, MSEP

Dependence	Method	$T = 50$	$T = 100$	$T = 200$
	GPV	0.00468	0.00323	0.00238
	LPV	0.01064	0.00553	0.00329
$\tau = 0.25$	LPV (log transform)	0.00808	0.00419	0.00249
	Clayton	0.00290	0.00187	0.00135
	Frank	0.00217	0.00115	0.00064
	Gumbel	0.00236	0.00153	0.00106
	GPV	0.00413	0.00353	0.00380
	LPV	0.00972	0.00672	0.00457
$\tau = 0.50$	LPV (log transform)	0.01064	0.00920	0.00681
	Clayton	0.00484	0.00451	0.00476
	Frank	0.00178	0.00096	0.00055
	Gumbel	0.00292	0.00199	0.00151
	GPV	0.01658	0.01826	0.02052
	LPV	0.02348	0.02192	0.01780
$\tau = 0.75$	LPV (log transform)	0.03078	0.03175	0.02665
	Clayton	0.02398	0.02430	0.02428
	Frank	0.00160	0.00089	0.00048
	Gumbel	0.00333	0.00293	0.00298

importance of participation costs which others, including Li [21], have investigated elsewhere.

In table 7, we present the summary descriptive statistics concerning our sample of 834 observations from 278 auctions. Our bid variable is the price per mile. Notice that both the winning bids as well as all tendered bids vary considerably, from a low of \$41,760.32 per mile to an high of \$5,693,872.81 per mile. What explains this variation? Well, presumably heterogeneity in the tasks that need to be performed. One way to control for this heterogeneity would be to retrieve each and every contract and then to construct covariates from those contracts. Unfortunately, the State of Michigan cannot provide us with this information, at least not any time soon.

How can we deal with this heterogeneity? Well, in our case, we have an engineer's estimate x of the cost per mile of performing the project. Thus, we condition on this exogenous covariate when estimating $G_0^0(b|x)$ and $g_0^0(b|x)$. In table 8, we present the PML estimates as well as standard errors of the copula parameters. We used the bootstrap to calculate the standard errors: specifically, we drew 278 triplets of bids along with the engineer's estimate for that auction at random, with replacement, from the sample distribution to form bootstrap estimates $\widetilde{G}_0^*(b|x)$ and $\widetilde{g}_0^*(b|x)$. We then estimated the dependence parameter θ for each chosen copula, replicating this for $1,000$ bootstrap samples.

Of course, we cannot compare directly the parameter estimates across copulae because the values of the parameters are specific to a given copula and imply different levels of dependence. Thus, alongside these parameter estimates, we present the corresponding estimated values of Kendall's τ, with standard errors, again obtained via the bootstrap.

The estimates, and their standard errors, indicate considerable (and significantly) positive dependence—affiliation—within the Archimedean family of copulae. In figure 2, we depict $\widetilde{f}_0(c|x)$, a non-parametric estimate of $f_0^0(c|x)$ admitting dependence, evaluated at the sample median of x for each copula used in estimating the pseudo-costs. The PML estimates of $f_0^0(c|x)$ are

Table 6: Performance of Methods using Data Generated from Gumbel Copula, MSEP

Dependence	Method	$T = 50$	$T = 100$	$T = 200$
$\tau = 0.25$	GPV	0.00452	0.00295	0.00219
	LPV	0.01408	0.00747	0.00463
	LPV (log transform)	0.00967	0.00514	0.00323
	Clayton	0.00354	0.00228	0.00180
	Frank	0.00305	0.00178	0.00123
	Gumbel	0.00216	0.00114	0.00065
$\tau = 0.50$	GPV	0.00383	0.00323	0.00331
	LPV	0.01240	0.00984	0.00688
	LPV (log transform)	0.01110	0.00915	0.00697
	Clayton	0.00787	0.00694	0.00650
	Frank	0.00352	0.00291	0.00227
	Gumbel	0.00150	0.00084	0.00046
$\tau = 0.75$	GPV	0.02047	0.02107	0.02173
	LPV	0.03073	0.02827	0.02389
	LPV (log transform)	0.03266	0.03167	0.02728
	Clayton	0.02906	0.02760	0.02514
	Frank	0.00542	0.00530	0.00474
	Gumbel	0.00090	0.00047	0.00024

Table 7: Sample Descriptive Statistics—Dollars/Mile, $n = 3$, $T = 278$

Variable	Mean	St. Dev.	Median	Minimum	Maximum
Engineer's Estimate	475,544.54	491,006.52	307,331.26	54,574.41	3,694,272.59
Winning Bid	466,468.63	507,025.81	286,102.57	41,760.32	3,882,524.81
All Tendered Bids	507,332.42	564,842.58	317,814.77	41,760.32	5,693,872.81

all very close, regardless of the copula used in estimation.

In figure 3, we present the bid function (pseudo-costs) predicted by our approach. Notice how our estimates are very close to the 45°-line. In figure 4, we plot our pseudo-cost estimates versus those estimated using the methods of GPV. One of the pseudo-costs estimated by the method of GPV is negative, while no costs are estimated to be negative using our approach. The main point, however, is that our PML estimated pseudo-costs are systematically above the GPV estimates. In the presence of affiliation, an higher cost is implied for a given bid than under independence because players (typically) bid closer to their costs. The average absolute relative difference between the GPV estimates and the PML estimates obtained using the Clayton, Frank, and Gumbel copulae are 13.27 percent, 20.75 percent, and 19.24 percent, respectively.

Why is affiliation potentially interesting to an economist? In figure 5, we illustrate that, in the presence of affiliation, low-cost bidders (those who are likely to win the auction) behave more competitively than would be the case under independence. The figure was constructed using the analytic bid function for the Frank copula with parameters corresponding to the values for Kendall's τ denoted in the figure. Here, τ^* equals the value estimated using a Frank copula in the application given in table 8. In fact, even with just three bidders, the winners (typically those with low costs) bid very close to their costs. Affiliation disciplines bidders: when a bidder contemplates his having the lowest cost, and thus winning the auction, he must also recognize

Table 8: Copula Parameter and Kendall's τ Estimates

	Copula Parameter		Kendall's τ	
Copula	Estimate	St. Err.	Estimate	St. Err.
Clayton	3.8028	0.3177	0.6553	0.0211
Frank	11.2898	0.6645	0.6973	0.0167
Gumbel	3.0187	0.1319	0.6687	0.0176

that, under affiliation, his opponents will probably have costs close to his, and this forces him to bid more aggressively than he would under independence.

8. Summary and Conclusions

Within the affiliated private-values paradigm, we have developed a tractable empirical model of equilibrium behaviour at first-price, sealed-bid auctions. While the model is non-parametrically identified, the rate of convergence is slow when the number of bidders is even moderately large. Also, the affiliation sufficient for the measurement equation in an empirical specification to constitute an equilibrium is difficult to preserve in the course of non-parametric estimation. Thus, we have developed a semiparametric estimation strategy, which respects affiliation as well as avoids the curse of dimensionality that relates to the number of bidders, focusing our attention on the Archimedean family of copulae and implementing this framework using particular members—the Clayton, Frank, and Gumbel copulae. We have applied our framework to data from low-price, sealed-bid auctions used by the Michigan Department of Transportation to procure road-resurfacing services, rejecting the hypothesis of independence and finding significant (and high) affiliation in cost signals. Ignoring this potential affiliation has important implications concerning estimates of the procurement-cost distribution, which can be important to a policy-maker seeking to design the optimal procurement mechanism.

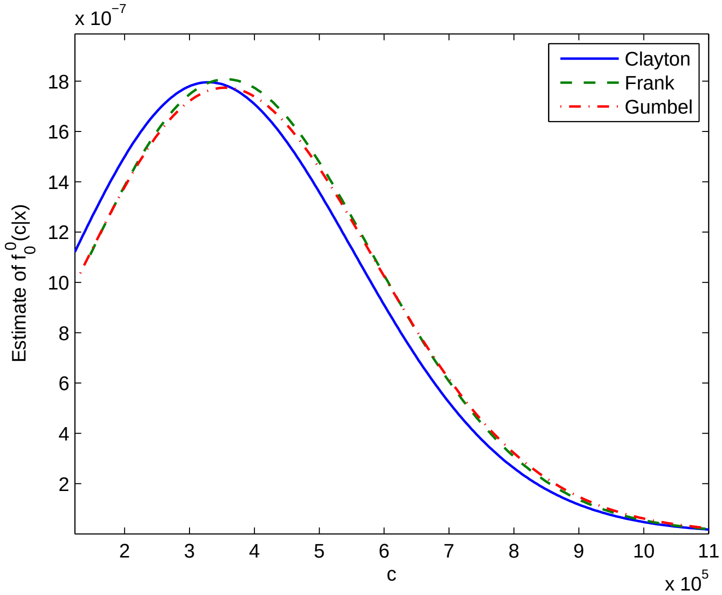


Figure 2: PML Estimates of $f_0^0(c|x)$ Evaluated at Sample Median of Engineer's Estimate x

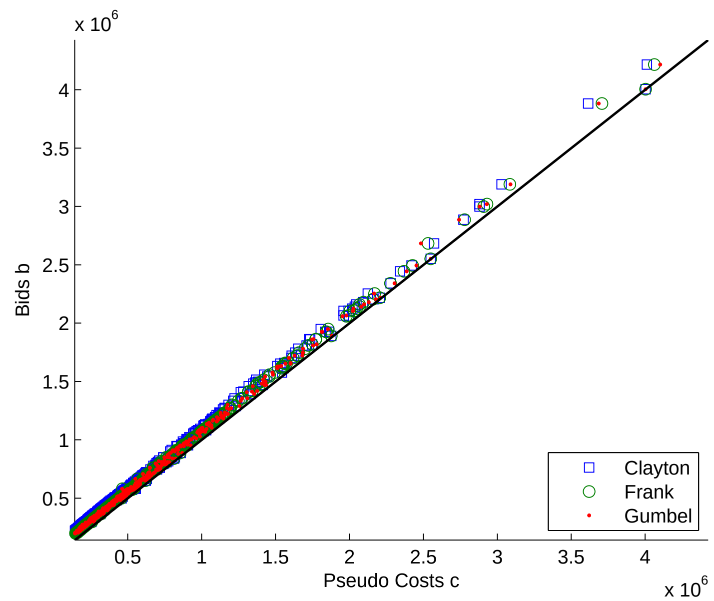


Figure 3: Bids versus PML Estimated Dependent Pseudo-Costs

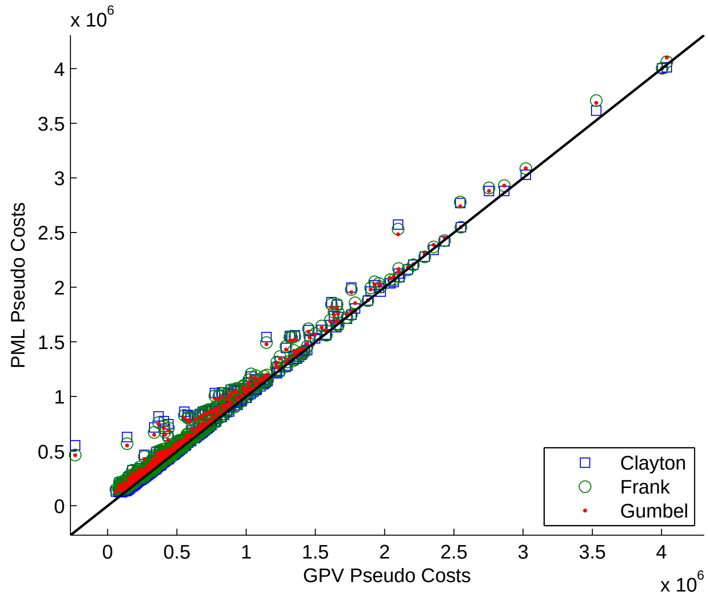


Figure 4: GPV versus PML Estimated Dependent Pseudo-Costs

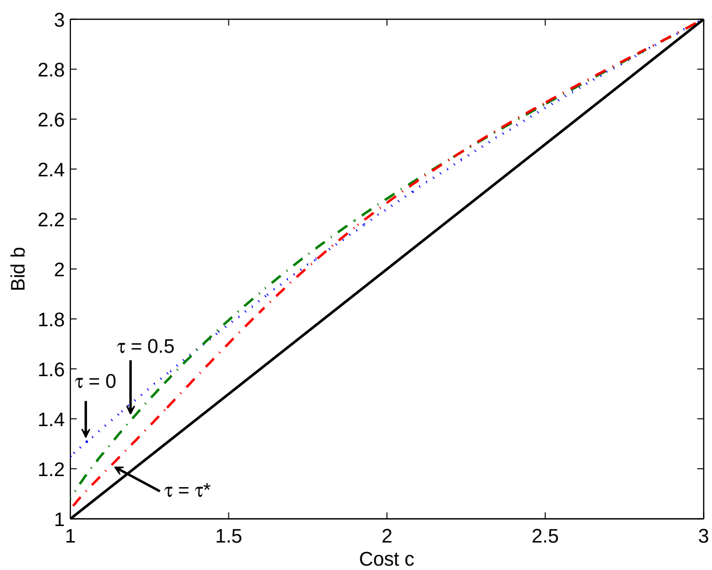


Figure 5: Bid Functions: Independence as well as Affiliation

Acknowledgements

The authors would like to thank Loren Lemmen of the Michigan Department of Transportation for his help in providing the data. They are also grateful to two anonymous referees as well as Bjarne Brendstrup, Martin W. Cripps, Luciano de Castro, Yanqin Fan, Marcelo Fernandes, Emmanuel Guerre, Han Hong, Joel L. Horowitz, Alfred Müller, Denis Nekipelov, Jean-Marc Robin, and Olivier Scaillet for useful comments and helpful suggestions. Li gratefully acknowledges the financial support of the National Science Foundation under grant number SES-0922109, while Paarsch gratefully acknowledges that the final revisions to this paper were completed while he was a visiting research scholar at the Center for Economic Institutions in the Institute of Economic Research at Hitotsubashi University in Kunitachi, Japan.

Appendix

In this appendix, we present calculations too cumbersome for inclusion in the text of the paper as well as describe the creation of the data set used.

Low-Price, Sealed-Bid, Procurement-Auction Model

In this section of the appendix, we present the calculations necessary to implement the low-price, sealed-bid, procurement-auction model investigated in section 6 and implemented in section 7.

We first extend the result of Nelsen [26] concerning the copula representation of the bivariate survival copula to the case of n of three or greater, and then use this result, in conjunction with the GPV transformation, to isolate the pseudo-costs in a symmetric affiliated private-cost model of a procurement auction.

Lemma 2:

$$\begin{aligned}\Pr(C_1 \geq c_1, \ldots, C_n \geq c_n) = 1 - \sum_{i=1}^{n} \Pr(C_i < c_i) + \\ \sum_{1 \leq i \leq j \leq n} \Pr(C_i < c_i, C_j < c_j) - \ldots + \\ (-1)^n \Pr(C_1 < c_1, \ldots, C_n < c_n).\end{aligned}$$

Proof.

The proof is by induction. We begin by demonstrating the result for the case of n equal two, and then demonstrate it for n greater than two. Note that, when n is two, we have

$$\begin{aligned}\Pr(C_1 \geq c_1, C_2 \geq c_2) &= \Pr(C_1 \geq c_1) - \Pr(C_1 \geq c_1, C_2 < c_2) \\ &= 1 - \Pr(C_1 < c_1) - [\Pr(C_1 \geq c_1, C_2 < c_2) - \Pr(C_1 < c_1, C_2 < c_2)] \\ &= 1 - \Pr(C_1 < c_1) - \Pr(C_2 < c_2) + \Pr(C_1 < c_1, C_2 < c_2).\end{aligned}$$

Now, suppose the result holds for the case of $(n-1)$ where n is three or greater, then

$$\begin{aligned}\Pr(C_1 \geq c_1, \ldots, C_n \geq c_n) = \Pr(C_1 \geq c_1, \ldots, C_{n-1} \geq c_{n-1}) - \\ \Pr(C_1 \geq c_1, \ldots, C_{n-1} \geq c_{n-1}, C_n < c_n)\end{aligned}$$

$$
\begin{aligned}
&= 1 - \sum_{i=1}^{n-1} \Pr(C_i < c_i) + \\
&\quad \sum_{1 \le i \le j \le n-1} \Pr(C_i < c_i, C_j < c_j) - \ldots + \\
&\quad (-1)^{n-1} \Pr(C_1 < c_1, \ldots, C_{n-1} < c_{n-1}) - \\
&\quad \Pr(C_1 \ge c_1, \ldots, C_{n-1} \ge c_{n-1} | C_n < c_n) \Pr(C_n < c_n) \\
&= 1 - \sum_{i=1}^{n-1} \Pr(C_i < c_i) + \\
&\quad \sum_{1 \le i \le j \le n-1} \Pr(C_i < c_i, C_j < c_j) - \ldots + \\
&\quad (-1)^{n-1} \Pr(C_1 < c_1, \ldots, C_{n-1} < c_{n-1}) - \\
&\quad [1 - \sum_{i=1}^{n-1} \Pr(C_i < c_i | C_n < c_n) + \\
&\quad \sum_{1 \le i \le j \le n-1} \Pr(C_i < c_i, C_j < c_j | C_n < c_n) - \ldots + \\
&\quad (-1)^{n-1} \Pr(C_1 < c_1, \ldots, C_{n-1} < c_{n-1} | C_n < c_n)] \\
&\quad \Pr(C_n < c_n)] \\
&= 1 - \sum_{i=1}^{n} \Pr(C_i < c_i) + \\
&\quad \sum_{1 \le i \le j \le n-1} \Pr(C_i < c_i, C_j < c_j) - \ldots + \\
&\quad (-1)^n \Pr(C_1 < c_1, \ldots, C_n < c_n)
\end{aligned}
$$

Note that this lemma gives a copula representation of the survival copula, which is useful in the characterization of the first-order condition of the equilibrium bid at a procurement auction. The lemma is a generalization of the case for n of two given by Nelsen [26] on page 28. We introduce the notation $\mathcal{S}$ to denote the survival copula and define a survival copula as

$$
\begin{aligned}
\mathcal{S}[1 - F_1(c_1), \ldots, 1 - F_n(c_n)] \equiv 1 &- \sum_{i=1}^{n} \Pr(C_i < c_i) + \\
&\sum_{1 \le i \le j \le n-1} \Pr(C_i < c_i, C_j < c_j) - \ldots + \\
&(-1)^n \Pr(C_1 < c_1, \ldots, C_n < c_n)
\end{aligned}
$$

where $F_i(\cdot)$ denotes the cumulative distribution function of variable C_i.

Lemma 3:

$$\Pr(C_2 \ge c_2, \ldots, C_n \ge c_n | c_1) = \mathcal{S}_1[1 - F_0(c_1), \ldots, 1 - F_0(c_n)]$$

where $\mathcal{S}_1$ denotes the partial derivative of $\mathcal{S}$ with respect to the first component.

Proof.

Let $f_{\boldsymbol{C}}(c_1, \ldots, c_n)$ denote the probability density function corresponding to $F_{\boldsymbol{C}}(c_1, \ldots, c_n)$. Now,

$$\begin{aligned}\Pr(C_1 \geq c_1, \ldots, C_n \geq c_n) &= \mathcal{S}_1[1 - F_0(c_1), \ldots, 1 - F_0(c_n)] \\ &= \int_{c_1} \cdots \int_{c_n} f_{\boldsymbol{C}}(u_1, \ldots, u_n)\, \mathrm{d}u_1, \ldots \mathrm{d}u_n.\end{aligned}$$

Differentiating both sides of the last equality with respect to c_1 yields

$$-\mathcal{S}_1[1 - F_0(c_1), \ldots, 1 - F_0(c_n)]\, f_0(c_1) = -\int_{c_2} \cdots \int_{c_n} f_{\boldsymbol{C}}(c_1, u_2, \ldots, u_n)\, \mathrm{d}u_2, \ldots \mathrm{d}u_n.$$

Thus, we have

$$\frac{\int_{c_2} \cdots \int_{c_n} f_{\boldsymbol{C}}(c_1, u_2, \ldots, u_n)\, \mathrm{d}u_2, \ldots \mathrm{d}u_n}{f_0(c_1)} = \mathcal{S}_1[1 - F_0(c_1), \ldots, 1 - F_0(c_n)]. \tag{10}$$

But the left-hand side of equation (10) is just $\Pr(C_2 \geq c_2, \ldots, C_n \geq c_n | c_1)$. The desired result then follows.

Within the APCP, for any bidder (focus, say, on bidder 1), we assume that he maximizes expected profit

$$\mathcal{E}[\pi(b_1, c_1)] = (b_1 - c_1)\mathcal{S}_1 \left(1 - F_0(c_1), 1 - F_0[\beta^{-1}(b_1)], \ldots, 1 - F_0[\beta^{-1}(b_1)]\right)$$

by choice of bid b_1 where c_1 is bidder 1's private cost and where $\beta(\cdot)$ is a strictly monotonically increasing function. The first-order condition for expected-profit maximization implies

$$\begin{aligned}\frac{\partial \mathcal{E}[\pi(b_1, c_1)]}{\partial b_1} &= 0 \\ &= \mathcal{S}_1 \left(1 - F_0(c_1), 1 - F_0[\beta^{-1}(b_1)], \ldots, 1 - F_0[\beta^{-1}(b_1)]\right) - \\ &\quad (b_1 - c_1) \sum_{j=2}^{n} f_0[\beta^{-1}(b_1)] \frac{d\beta^{-1}(b_1)}{db_1} \\ &\quad \times \mathcal{S}_{1j} \left(1 - F_0(c_1), 1 - F_0[\beta^{-1}(b_1)], \ldots, 1 - F_0[\beta^{-1}(b_1)]\right)\end{aligned}$$

which can be re-written as

$$\begin{aligned}&\mathcal{S}_1 \left(1 - F_0(c_1), 1 - F_0[\beta^{-1}(b_1)], \ldots, 1 - F_0[\beta^{-1}(b_1)]\right) \\ &\quad = (b_1 - c_1)(n - 1) f_0[\beta^{-1}(b_1)] \frac{d\beta^{-1}(b_1)}{db_1} \\ &\quad\quad \times \mathcal{S}_{12} \left(1 - F_0(c_1), 1 - F_0[\beta^{-1}(b_1)], \ldots, 1 - F_0[\beta^{-1}(b_1)]\right)\end{aligned}$$

because

$$\begin{aligned}&\mathcal{S}_{1j} \left(1 - F_0(c_1), 1 - F_0[\beta^{-1}(b_1)], \ldots, 1 - F_0[\beta^{-1}(b_1)]\right) = \\ &\quad \mathcal{S}_{1k} \left(1 - F_0(c_1), 1 - F_0[\beta^{-1}(b_1)], \ldots, 1 - F_0[\beta^{-1}(b_1)]\right) \ \forall j \neq k.\end{aligned}$$

Given this last equality, we can use $\mathcal{S}_{12}(\cdot)$ without loss of generality. Applying the GPV approach, we note that

$$G_0(b_1) = F_0(c_1)$$

and

$$g_0(b_1) = \frac{f_0(c_1)}{\beta'(c_1)},$$

so we can re-arrange terms to get

$$c_1 = b_1 - \frac{\mathcal{S}_1[1 - G_0(b_1), 1 - G_0(b_1), \ldots, 1 - G_0(b_1)]}{(n-1)g_0(b_1)\mathcal{S}_{12}[1 - G_0(b_1), 1 - G_0(b_1), \ldots, 1 - G_0(b_1)]},$$

which expresses bidder 1's cost solely as a function of the bid.

Data

The data for the empirical part of this paper, which concern procurement contracts for road resurfacing, were provided in raw form by the Department of Transportation of the State of Michigan. To create the final data set, we first extracted contract length and determined the number of bidders at each auction. Next, we looked for observations having missing data, scanning each file to ensure that all contract lengths were properly extracted. Finally, we checked to ensure that all data were in the same units; e.g., we converted contracts that were measured in kilometres into miles. At this point, we had $4,200$ observations concerning $1,041$ contracts. Subsequently, we focused on auctions having n of three; this constrained us to 278 auctions.

References

[1] Brendstrup, B., Paarsch, H., 2007. Semiparametric identification and estimation multi-object, English auctions. Journal of Econometrics 141, 84–108.

[2] Chen, X. and Fan, Y., 2006. Estimation of copula-based semiparametric time series models. Journal of Econometrics 130, 307–335.

[3] Chen, X. and Shen, X., 1998. Sieve extremum estimation for weakly dependent data. Econometrica 66, 289–314.

[4] Cherubini, U., Luciano, E., Vecchiato, W., 2004. Copula Methods in Finance. John Wiley & Sons, New York.

[5] de Castro, L., 2007. Affiliation, equilibrium existence and the revenue ranking of auctions. Typescript, Department of Economics, Carlos III University.

[6] Donald, S., Paarsch, H., 1993. Piecewise pseudo-maximum likelihood estimation in empirical models of auctions. International Economic Review 34, 121–148.

[7] Donald, S., Paarsch, H., 1996. Identification, estimation, and testing in parametric empirical models of auctions within the independent private values paradigm. Econometric Theory 12, 517–567.

[8] Donald, S., Paarsch, H., 2002. Superconsistent estimation and inference in structural econometric models using extreme order statistics. Journal of Econometrics 109, 305–340.

[9] Fermanian J.-D., 2005. Goodness-of-fit tests for copulas. Journal of Multivariate Analysis 95, 119-152.

[10] Fermanian J.-D., Scaillet, O., 2005. Some statistical pitfalls in copula modeling for financial applications. In: Klein, E. (Ed.), Capital Formation, Governance and Banking. Nova Science Publishers, New York, 57–72.

[11] Genest, C., 1987. Frank's family of bivariate distributions. Biometrika 74, 549–555.

[12] Genest, C., MacKay, J., 1986. The joy of copulas: bivariate distributions with uniform marginals. The American Statistician 40, 280–283.

[13] Guerre, E., Perrigne, I., Vuong, Q., 2000. Optimal nonparametric estimation of first-price auctions. Econometrica 68, 525–574.

[14] Haile, P., Tamer, E., 2003. Inference with an incomplete model of English auctions. Journal of Political Economy 111, 1–51.

[15] Härdle, W., 1991. Smoothing Techniques with Implementation in S. Springer–Verlag, New York.

[16] Jouini, M., Clemen, R., 1996. Copula models for aggregating expert opinions. Operations Research 44, 444–457.

[17] Karlin, S., 1968. Total Positivity, Volume 1. Stanford University Press, Stanford.

[18] Krishna, V., 2002. Auction Theory. Academic Press, San Diego.
[19] Laffont, J.-J., Ossard, H., Vuong, Q., 1995. Econometrics of first-price auctions. Econometrica 63, 953–980.
[20] Laffont, J.-J., Vuong Q., 1996. Structural analysis of auction data. American Economic Review 86, 414–420.
[21] Li, T., 2005. Econometrics of first-price auctions with entry and binding reservation prices. Journal of Econometrics 126, 173–200.
[22] Li, T., Perrigne, I., Vuong, Q., 2000. Conditionally independent private information in OCS wildcat auctions. Journal of Econometrics 98, 129–161.
[23] Li, T., Perrigne, I., Vuong, Q., 2002. Structural estimation of the affiliated private value model. RAND Journal of Economics 33, 171–193.
[24] Milgrom, P., Weber, R., 1982. A theory of auctions and competitive bidding. Econometrica 50, 1089–1122.
[25] Müller, A., Scarsini, M., 2005. Archimedean copulae and positive dependence. Journal of Multivariate Analysis 93, 434–445.
[26] Nelsen, R., 1999. An Introduction to Copulas. Springer, New York.
[27] Newey, W., McFadden D., 1994. Large sample estimation and hypothesis testing. In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics, Volume 4. Elsevier, New York.
[28] Paarsch, H., 1992. Deciding between the common and private value paradigms in empirical models of auctions. Journal of Econometrics 51, 191–215.
[29] Paarsch, H., 1997. Deriving an estimate of the optimal reserve price: an application to British Columbian timber sales. Journal of Econometrics 78, 333–357.
[30] Paarsch, H., Hong, H., 2006. An Introduction to the Structural Econometrics of Auction Data. MIT Press, Cambridge, Massachusetts.