Efficient Estimation of Semiparametric Multivariate Copula Models*

Xiaohong Chen† Yanqin Fan‡ Victor Tsyrennikov§

First version: May 2002; Second revision: November 2005

Abstract

We propose a sieve maximum likelihood estimation (MLE) procedure for a broad class of semiparametric multivariate distributions. A joint distribution in this class is characterized by a parametric copula function evaluated at nonparametric marginal distributions. This class of distributions has gained popularity in diverse fields due to a) its flexibility in separately modeling the dependence structure and the marginal behaviors of a multivariate random variable, and b) its circumvention of the "curse of dimensionality" associated with purely nonparametric multivariate distributions. We show that the plug-in sieve MLEs of all smooth functionals, including the finite dimensional copula parameters and the unknown marginal distributions, are semiparametrically efficient; and that their asymptotic variances can be estimated consistently. Moreover, prior restrictions on the marginal distributions can be easily incorporated into the sieve MLE procedure to achieve further efficiency gains. Two such cases are studied in the paper: (i) the marginal distributions are equal but otherwise unspecified, and (ii) some but not all marginal distributions are parametric. Monte Carlo studies indicate that the sieve MLEs perform well in finite samples, especially so when prior information on the marginal distributions is incorporated.

KEY WORDS: Multivariate copula; Sieve maximum likelihood; Semiparametric efficiency

---

*We thank the coeditor, the associate editor and three anonymous referees for very useful comments that greatly improved the paper. We also thank Oliver Linton and participants at 2003 North American Econometric Society Summer Meetings in Chicago, 2004 DeMoSTAFI conference in Quebec City and 2004 Semiparametrics in Rio for helpful comments. We also thank Demian Pouzo for excellent assistance in simulation studies. Chen acknowledges financial supports from the ESRC/UK, the NSF/USA and the C.V. Starr Center at NYU. Fan acknowledges financial support from the NSF/USA.

†The corresponding author. Department of Economics, New York University, 269 Mercer Street, 7th Floor, New York, NY 10003, USA; E-mail: xiaohong.chen@nyu.edu.

‡Department of Economics, Vanderbilt University, VU Station B #351819, 2301 Vanderbilt Place, Nashville, TN 37235, USA. E-mail: yanqin.fan@vanderbilt.edu.

§Department of Economics, New York University, 269 Mercer Street, 7th Floor, New York, NY 10003, USA; E-mail: victor.tsyrennikov@nyu.edu.

1 Introduction

Let $\{Z_i \equiv (X_{1i}, ..., X_{mi})'\}_{i=1}^n$ be a random sample from the distribution $H_o(x_1, \ldots, x_m)$ of $Z \equiv (X_1, ..., X_m)'$ in $\mathcal{X}_1 \times ... \times \mathcal{X}_m \subseteq \mathcal{R}^m$, $m \geq 2$. Assume that H_o is absolutely continuous with respect to the Lebesgue measure on $\mathcal{R}^m$ and let $h_o(x_1, ..., x_m)$ be the probability density function (pdf) of Z. Clearly estimation of H_o or h_o is one of the most important statistical problems. Due to the well-known "curse of dimensionality," it is undesirable to estimate H_o or h_o fully nonparametrically in high dimensions. This motivates the development of many semiparametric models for H_o.

One class of semiparametric multivariate distributions has gained popularity in diverse fields due to its flexibility in separately modeling the dependence structure and the marginal behaviors of a multivariate random variable. To introduce this class, let F_{oj} denote the true unknown marginal cdf of X_j, $j = 1, ..., m$. The characterization theorem of Sklar (1959) implies that there exists a unique copula function $C_o()$ such that $H_o(X_1, ..., X_m) \equiv C_o(F_{o1}(X_1), ..., F_{om}(X_m))$. Suppose that the functional form of the copula $C_o(u_1, ..., u_m)$ is known apart from a finite dimensional parameter θ_o, i.e., for any $(u_1, \ldots, u_m) \in [0, 1]^m$, we have $C_o(u_1, ..., u_m) = C(u_1, ..., u_m; \theta_o)$, where $\{C(u_1, ..., u_m; \theta) : \theta \in \Theta\}$ is a parametric family of copula functions.[1] Then the multivariate distribution H_o is of a semiparametric form:

$$H_o(X_1, ..., X_m) = C(F_{o1}(X_1), ..., F_{om}(X_m); \theta_o) \tag{1}$$

with unknown finite dimensional parameter θ_o and infinite dimensional parameters F_{oj}, $j = 1, \ldots, m$. Let f_{oj}, $j = 1, ..., m$, and $c(u_1, ..., u_m; \theta_o)$ denote the pdfs associated with F_{oj}, $j = 1, ..., m$, and $C(u_1, ..., u_m; \theta_o)$ respectively. Then for any $(x_1, ..., x_m) \in \mathcal{X}_1 \times ... \times \mathcal{X}_m$, the pdf h_o of H_o given by (1) has the representation: $h_o(x_1, ..., x_m) = c(F_{o1}(x_1), ..., F_{om}(x_m); \theta_o) \prod_{j=1}^m f_{oj}(x_j)$. We refer to the class of multivariate distributions of the form (1) as the class of copula-based semiparametric multivariate distributions.

This class of copula-based models achieves the aim of dimension reduction; as for any m, the joint density $h_o(x_1, \ldots, x_m)$ depends on nonparametric functions of only one dimension. In addition, the parameters in models of this class are easy to interpret: the marginal distributions F_{oj}, $j = 1, \ldots, m$, capture the marginal behavior of the univariate random variables X_j, $j = 1, \ldots, m$; and the finite dimensional parameter θ_o, or equivalently the parametric copula $C(u_1, \ldots, u_m, \theta_o)$, characterizes the dependence structure between $X_1, \ldots, X_m$ that is invariant to any increasing transformations of the univariate random variables X_j, $j = 1, \ldots, m$. Given the existence of a large number of parametric copulas and univariate distributions, this class of semiparametric multivariate distributions is very flexible to model jointly any type of dependence with any types of marginal behaviors, and has proven to be useful in diverse fields. Specific applications include those in finance

[1] Commonly used parametric copulas include the Gaussian copula, Student's t copula, Clayton, Frank and Gumbel copulas; see Joe (1997) and Nelsen (1999) for properties of many existing parametric copulas.

and insurance (e.g., Frees and Valdez (1998) and Embrechts, et al. (2002)); in survival analysis (e.g. Joe (1997), Nelsen (1999) and Oakes (1989)); in econometrics (e.g. Lee (1983), Heckman and Honore (1989), Granger, et al. (2003) and Patton (2004)), to name only a few.

To estimate the multivariate distribution $H_o(x_1, \ldots, x_m) \equiv C(F_{o1}(x_1), \ldots, F_{om}(x_m); \theta_o)$, one has to estimate both the copula parameter θ_o and the marginal cdfs F_{oj}, $j = 1, \ldots, m$. In current literature, the most popular estimator of F_{oj} is the empirical cdf $F_{nj}(x_j) = \frac{1}{n} \sum_{i=1}^{n} 1\{X_{ji} \leq x_j\}$ for $j = 1, \ldots, m$. And the most widely used estimator of θ_o is the two-step estimator $\widetilde{\theta}_n$ proposed by Oakes (1994) and Genest, et al. (1995):

$$\widetilde{\theta}_n = \arg\max_{\theta \in \Theta} \left[\sum_{i=1}^{n} \log c(\tilde{F}_{n1}(X_{1i}), ..., \tilde{F}_{nm}(X_{mi}); \theta) \right], \tag{2}$$

where $\widetilde{F}_{nj}(x_j) = \frac{1}{n+1} \sum_{i=1}^{n} 1\{X_{ji} \leq x_j\}$ is the rescaled empirical cdf estimator of $F_{oj}, j = 1, ..., m$. Genest, et al. (1995) establish the root-n consistency and asymptotic normality of $\widetilde{\theta}_n$.[2]

In many applications, efficient estimation of the entire multivariate distribution $H_o(x_1, \ldots, x_m) \equiv C(F_{o1}(x_1), \ldots, F_{om}(x_m); \theta_o)$ is desirable, which requires efficient estimation of both the marginal cdfs F_{oj}, $j = 1, \ldots, m$ and the copula dependence parameter θ_o. Except when $X_1, \ldots, X_m$ are independent, it is clear that the empirical cdfs F_{nj}, $j = 1, \ldots, m$ are generally inefficient. Intuitively one could obtain more efficient estimates of F_{oj}, $j = 1, \ldots, m$ by utilizing the dependence information contained in the parametric copula. Except for a few special cases, the two-step estimator of the copula parameter θ_o is inefficient in general (see Genest and Werker, 2002). This is because the two-step estimator $\widetilde{\theta}_n$ does not solve the efficient score equation for θ in general. Currently there are only two known special cases where the two-step estimator is asymptotically efficient; it is efficient at independence (Genest, et al., 1995), and it is efficient for the Gaussian copula parameter when marginal cdfs are unknown (Klaassen and Wellner, 1997). Unfortunately even for the bivariate Gaussian copula model with unknown margins, there is presently no efficient estimates of univariate marginal cdfs; see Klaassen and Wellner (1997). For semiparametric bivariate survival Clayton copula models, Maguluri (1993) provides some efficiency score calculation for θ_o and conjectures that his proposed estimator might be efficient. For general bivariate semiparametric copula models, Bickel, et al. (1993, chapter 4.7) present some efficiency bound characterizations for θ_o, but no efficient estimators. For a bivariate copula model with one known marginal cdf and one unknown marginal cdf, Bickel, et al. (1993, chapter 6.7) provide some efficiency bound calculations for the unknown margin, but again no efficient estimators. To the best of our knowledge (see Klaassen and Wellner (1997), and Genest and Werker (2002)), there does not exist any published

[2] Shih and Louis (1995) independently propose the two-step estimator for i.i.d. data with random censoring. The two-step estimator and its large sample properties have been extended to time series setting in Chen and Fan (2005a, b). There are many earlier papers that propose specific estimators of the copula parameter θ_o for specific parametric copula models; see e.g., Clayton (1978), Clayton and Cuzick (1985), Oakes (1982, 1986) and Genest (1987).

work on efficient estimation of θ_o and F_{oj}, $j = 1, \ldots, m$ for general multivariate semiparametric copula models.

In this paper, we propose a general sieve maximum likelihood estimation (MLE) procedure for all the unknown parameters in a semiparametric multivariate copula model (1). This procedure approximates the infinite-dimensional unknown marginal densities f_{oj}, $j = 1, ..., m$ by linear combinations of finite-dimensional known basis functions with increasing complexity (sieves), and then maximizes the joint likelihood with respect to the copula parameter and the sieve parameters of the approximating marginal densities. Because our sieve MLEs of the marginal cdfs utilize all the parametric dependence information, and our sieve MLE of the copula parameter effectively solves an approximate efficient score equation for θ (where the approximation error becomes negligible as sample size grows large enough), intuition suggests that these estimators should be efficient. By applying the general theory of Shen (1997) we can show that our plug-in sieve MLEs of all smooth functionals, including the unknown marginal cdfs and the copula parameter, are indeed semiparametrically efficient. As our sieve MLE procedure involves approximating and estimating one-dimensional unknown functions (marginal densities) only, it avoids the "curse of dimensionality" and is simple to compute. In addition, it can be easily adapted to estimating semiparametric multivariate copula models with prior restrictions on the marginal cdfs to produce more efficient estimates. Examples of such restrictions include equal but unknown marginal cdfs, known parametric forms of some (but not all) marginal cdfs. Results from an extensive simulation study for several copula families and marginal cdfs in both bivariate and tri-variate models confirm the efficacy of the sieve MLE.

Although we establish that the sieve MLEs of copula parameter and marginal cdfs achieve their efficiency bounds, there is no closed-form expressions for the efficiency bounds of copula parameter and marginal cdfs in general semiparametric copula models (except for a few special bivariate copula models such as the bivariate Clayton copula model with one known margin). As a result, direct estimation of the asymptotic variances of sieve MLEs using the analytic expressions of the efficiency bounds is only possible for a few special copula models. Nevertheless, for general semiparametric multivariate copula models with or without prior information on marginal cdfs, we are able to provide simple consistent estimates of the asymptotic variances of the sieve MLEs of the copula parameter and of the unknown marginal cdfs. This greatly broadens the applicability of our sieve MLEs. Using the closed-form expressions in the special model of bivariate Clayton copula with one known margin, we demonstrate via simulation that our consistent estimators of the asymptotic variances of the sieve MLEs for both the copula parameter and the unknown marginals perform extremely well.

The rest of this paper is organized as follows. Section 2 introduces the sieve MLEs of the copula parameter and the unknown marginal cdfs in models with or without restrictions on the marginal cdfs. In Section 3, we show that for semiparametric multivariate copula models with unknown marginal cdfs, the plug-in sieve MLEs of all smooth functionals are root-n normal and semiparametrically efficient. These results are then employed to deliver the root-n asymptotic normality and efficiency of the sieve MLEs of the copula parameter and the marginal cdfs. We also provide simple consistent estimators of the asymptotic variances of these sieve MLEs. Section 4 extends results in Section 3 to models with equal but unknown margins and models with some parametric margins. Section 5 provides simulation results on finite sample performance of the sieve MLEs for various models of different combinations of marginals and copulas that exhibit a wide range of dependence structures. It also reveals some important features of the relative behaviors of the sieve MLE of the copula parameter to the two-step estimator, and of the sieve MLEs of the marginal cdfs to the empirical cdfs. Appendix A contains the proofs. Appendix B presents asymptotic variances of the modified two-step estimator of θ_o under restrictions on the marginals.

2 The Sieve ML Estimators

We first introduce suitable sieve spaces for approximating an unknown univariate density function of certain smoothness, based on which we will then present our sieve MLEs.

2.1 Sieve Spaces for Approximating a Univariate Density

Let the true density function f_{oj} belong to $\mathcal{F}_j$ for $j = 1, \dots, m$. Recall that a space $\mathcal{F}_{nj}$ is called a sieve space for $\mathcal{F}_j$ if for any $g_j \in \mathcal{F}_j$, there exists an element $\Pi_n g_j \in \mathcal{F}_{nj}$ such that $d(g_j, \Pi_n g_j) \to 0$ as $n \to \infty$ where d is a metric on $\mathcal{F}_j$; see e.g. Grenander (1981) and Geman and Hwang (1982).

There exist many sieves for approximating a univariate probability density function. In this paper, we will focus on using linear sieves to directly approximate a square root density:

$$\mathcal{F}_{nj} = \left\{ f_{K_{nj}}(x) = \left[\sum_{k=1}^{K_{nj}} a_k A_k(x) \right]^2, \quad \int f_{K_{nj}}(x)dx = 1 \right\}, \quad K_{nj} \to \infty, \frac{K_{nj}}{n} \to 0, \tag{3}$$

where $\{A_k(\cdot) : k \geq 1\}$ consists of known basis functions, and $\{a_k : k \geq 1\}$ is the collection of unknown sieve coefficients.

Before presenting some concrete examples of known sieve basis functions $\{A_k(\cdot) : k \geq 1\}$, we first recall a popular smoothness function class used in the nonparametric estimation literature; see, e.g. Stone (1982) and Robinson (1988). Suppose the support $\mathcal{X}_j$ (of the true f_{oj}) is either a compact interval (say $[0, 1]$) or the whole real line $\mathcal{R}$. A real-valued function h on $\mathcal{X}_j$ is said to be *r-smooth* if it is J times continuously differentiable on $\mathcal{X}_j$ and its J-th derivative satisfies a Hölder condition with exponent $\gamma \equiv r - J \in (0, 1]$ (i.e., there is a positive number K such that $|D^J h(x) - D^J h(y)| \leq K|x - y|^\gamma$ for all $x, y \in \mathcal{X}_j$). We denote $\Lambda^r(\mathcal{X}_j)$ as the class of all real-valued functions on $\mathcal{X}_j$ which are *r-smooth*; it is called a Hölder space.

The appropriate sieve bases for approximating functions in $\Lambda^r(\mathcal{X}_j)$ depend on the support $\mathcal{X}_j$. If the support is bounded such as $\mathcal{X}_j = [0,1]$, then functions in $\Lambda^r(\mathcal{X}_j)$ with $r > 1/2$ can be well approximated by the spline sieve $\text{Spl}(s, K_n)$ with $s > [r]$ (the largest integer part of r). The spline sieve $\text{Spl}(s, K_n)$ is a linear space of dimension $(K_n + s + 1)$ consisting of spline functions of degree s with almost equally spaced knots $t_1, \ldots, t_{K_n}$ on $[0,1]$. Let $t_0, t_1, \ldots, t_{K_n}, t_{K_n+1}$ be real numbers with $0 = t_0 < t_1 < \cdots < t_{K_n} < t_{K_n+1} = 1$ and $\max_{0\le k\le K_n}(t_{k+1}-t_k) \le const. \min_{0\le k\le K_n}(t_{k+1}-t_k)$. Partition $[0,1]$ into K_n+1 subintervals $I_k = [t_k, t_{k+1})$, $k = 0, \ldots, K_n - 1$, and $I_{K_n} = [t_{K_n}, t_{K_n+1}]$. A function on $[0,1]$ is a spline of degree s with knots $t_1, \ldots, t_{K_n}$ if it is: (i) a polynomial of degree s or less on each interval I_k, $k = 0, \ldots, K_n$; and (ii) $(s-1)$-times continuously differentiable on $[0,1]$. See Schumaker (1981) for details on univariate splines. Other sieve spaces for approximating functions in $\Lambda^r(\mathcal{X}_j)$ with $r > 1/2$ and $\mathcal{X}_j = [0,1]$ include the polynomial sieve $\text{Pol}(K_n) = \{\sum_{k=0}^{K_n} a_k x^k,\ x \in [0,1] :\ a_k \in \mathcal{R}\}$, the trigonometric sieve $\text{TriPol}(K_n) = \{a_0 + \sum_{k=1}^{K_n}[a_k \cos(k\pi x) + b_k \sin(k\pi x)],\ x \in [0,1] :\ a_k, b_k \in \mathcal{R}\}$ and the cosine series $\text{CosPol}(K_n) = \{a_0 + \sum_{k=1}^{K_n} a_k \cos(k\pi x),\ x \in [0,1] :\ a_k \in \mathcal{R}\}$.

If the true unknown marginal densities are such that $\sqrt{f_{oj}} \in \Lambda^{r_j}(\mathcal{X}_j)$, $\mathcal{X}_j$ bounded interval, then we can let $\mathcal{F}_{nj}$ in (3) be

$$\mathcal{F}_{nj} = \left\{ \begin{array}{c} f(x) = [g(x)]^2 :\ \int [g(x)]^2\, dx = 1, \\ g \in \text{Spl}([r_j]+1, K_n) \text{ or } \text{Pol}(K_n) \text{ or } \text{TriPol}(K_n) \text{ or } \text{CosPol}(K_n) \end{array} \right\}. \tag{4}$$

There are also sieve bases that can be used to approximate densities with unbounded support: $\mathcal{X}_j = \mathcal{R}$. For example, (i) if the density f_{oj} has close to exponential thin tails, we can use the Hermite polynomial sieve to approximate it:

$$\mathcal{F}_{nj} = \left\{ \begin{array}{c} f_{K_{nj}}(x) = \frac{\epsilon_0 + \{\sum_{k=1}^{K_{nj}} a_k (\frac{x-\varsigma_0}{\sigma})^k\}^2}{\sigma} \exp\{-\frac{(x-\varsigma_0)^2}{2\sigma^2}\} : \\ \epsilon_0 > 0, \sigma > 0, \varsigma_0, a_k \in \mathcal{R}, \quad \int f_{K_{nj}}(x)dx = 1 \end{array} \right\} \tag{5}$$

as in Gallant and Nychka (1987); (ii) if the density f_{oj} has polynomial fat tails, we can use the spline wavelet sieve to approximate it:

$$\mathcal{F}_{nj} = \left\{ f_{K_{nj}}(x) = \left[\sum_{k=0}^{K_{nj}} \sum_{l \in \mathcal{K}_n} a_{kl} 2^{k/2} B_\gamma(2^k x - l) \right]^2, \quad \int f_{K_{nj}}(x) dx = 1 \right\} \tag{6}$$

where $B_\gamma(\cdot)$ denotes the cardinal B-spline of order γ:

$$B_\gamma(y) = \frac{1}{(\gamma-1)!} \sum_{i=0}^{\gamma} (-1)^i \begin{pmatrix} \gamma \\ i \end{pmatrix} [\max(0, y-i)]^{\gamma-1}. \tag{7}$$

See Chui (1992, Chapter 4) for the approximation property of this sieve.

2.2 Sieve MLEs

To simplify presentation, regardless there is any prior information on marginal distributions, we let $\ell(\alpha, Z_i)$ denote the contribution of the i-th observation to the log-likelihood function and $\widehat{\alpha}_n$ denote the sieve MLE for all the cases being considered in the paper.

First we consider the completely unrestricted case. Let $\alpha = (\theta', f_1, ..., f_m)'$ and denote $\alpha_o = (\theta_o', f_{o1}, ..., f_{om})' \in \Theta \times \prod_{j=1}^m \mathcal{F}_j = \mathcal{A}$ as the true but unknown parameter value. Let

$$\ell(\alpha, Z_i) = \log\{c(F_1(X_{1i}), ..., F_m(X_{mi}); \theta) \prod_{j=1}^{m} f_j(X_{ji})\}$$

in which $F_j(X_{ji}) = \int_{\mathcal{X}_j} 1(x \leq X_{ji}) f_j(x) dx$, $j = 1, ..., m$, and $\widehat{\alpha}_n = (\widehat{\theta}_n', \widehat{f}_{n1}, ..., \widehat{f}_{nm})' \in \Theta \times \prod_{j=1}^m \mathcal{F}_{nj} = \mathcal{A}_n$ denote the sieve MLE:

$$\widehat{\alpha}_n = \text{argmax}_{\alpha \in \mathcal{A}_n} \sum_{i=1}^{n} \ell(\alpha, Z_i) \tag{8}$$

where the sieve space $\mathcal{F}_{nj}$ could be (4) if $\mathcal{X}_j$ is a bounded interval, and could be (5) or (6) if $\mathcal{X}_j = \mathcal{R}$. The plug-in sieve MLE of the marginal distribution $F_{oj}(\cdot)$ is given by $\hat{F}_{nj}(x_j) = \int 1(y \leq x_j) \hat{f}_{nj}(y) dy$, $j = 1, ..., m$.

Remark 1: The sieve MLE optimization problem can be rewritten as an unconstrained optimization problem:

$$\max_{\theta, a_{1n}, \ldots, a_{mn}} \sum_{i=1}^{n} \{\log c(F_1(X_{1i}; a_{1n}), ..., F_m(X_{mi}; a_{mn}); \theta) + \sum_{j=1}^{m} [\log f_j(X_{ji}; a_{jn}) + \lambda_{jn} Pen(a_{jn})]\},$$

where for $j = 1, ..., m$, $f_j(X_{ji}; a_{jn})$ is a known (up to unknown sieve coefficients a_{jn}) sieve approximation to the unknown true f_{oj}, and $F_j(X_{ji}; a_{jn})$ is the corresponding sieve approximation to the unknown true F_{oj}. The smoothness penalization term $Pen(a_{jn})$ typically corresponds to the L_2-norm of either the first derivative or the second order derivative of $f_j^{1/2}(\cdot; a_{jn})$, and λ_{jn}'s are penalization factors. In our simulation study, we chose the penalization factors via cross-validation. In principle, we could use any model selection methodology such as cross-validation, covariance penalty (see, e.g., Shen and Ye (2002), Shen, et al. (2004)), among many others, to choose the number of terms K_{nj} in the sieve approximation.

Note that once the unknown marginal density functions are approximated by the appropriate sieves, the sieve MLEs are obtained by maximization over a finite dimensional parameter space. The properties of the resulting sieve MLEs depend on the approximation properties of the sieves. Prior restrictions on the marginal distributions can be easily taken into account in the choice of the sieves, leading to further efficiency gain in the resulting sieve MLEs. We now illustrate this for two cases.

The first is the case where the marginal distributions are the same, but unspecified otherwise. Let $F_{oj} = F_o$ ($f_{oj} = f_o$) and $\mathcal{X}_j = \mathcal{X}$ for all $j = 1, ..., m$. Let $\alpha = (\theta', f)'$ and let $\alpha_o = (\theta_o', f_o)' \in \Theta \times \mathcal{F}_1 = \mathcal{A}$ be the true but unknown parameter value. Let $\ell(\alpha, Z_i) = \log\{c(F(X_{1i}), ..., F(X_{mi}); \theta) \prod_{j=1}^m f(X_{ji})\}$ in which $F(X_{ji}) = \int_{\mathcal{X}} 1(x \leq X_{ji}) f(x) dx$, $j = 1, ..., m$. Then the sieve MLE $\widehat{\alpha}_n = (\widehat{\theta}_n', \widehat{f}_n)' \in \Theta \times \mathcal{F}_{n1} = \mathcal{A}_n$ is given by (8). This procedure can be easily extended to the case where some but not all marginal distributions are equal.

Bickel, et al. (1993) consider a semiparametric bivariate copula model in which one marginal cdf is completely known and the other marginal is left unspecified. The sieve ML estimation procedure we just introduced can be easily modified to exploit this information. To be more specific, let the marginal distribution F_{o1} be of parametric form, i.e., $F_{o1}(x_1) = F_{o1}(x_1, \beta_o)$ for some $\beta_o \in \mathcal{B}$. The marginal distributions $F_{o2}, \ldots, F_{om}$ are unspecified. Let $\alpha = (\theta', \beta', f_2, ..., f_m)'$ and denote $\alpha_o = (\theta_o', \beta_o', f_{o2}, ..., f_{om})' \in \Theta \times \mathcal{B} \times \prod_{j=2}^m \mathcal{F}_j = \mathcal{A}$ as the true but unknown parameter value. Let $\ell(\alpha, Z_i) = \log\left\{c(F_{o1}(X_{1i}, \beta), ..., F_{om}(X_{mi}); \theta) f_{o1}(X_{1i}, \beta) \prod_{j=2}^m f_j(X_{ji})\right\}$ in which $F_j(X_{ji}) = \int_{\mathcal{X}_j} 1(x \le X_{ji}) f_j(x) dx$, $j = 2, ..., m$. Then the sieve MLE denoted as $\widehat{\alpha}_n = (\widehat{\theta}_n', \widehat{\beta}_n', \widehat{f}_{n2}, ..., \widehat{f}_{nm})' \in \Theta \times \mathcal{B} \times \prod_{j=2}^m \mathcal{F}_{nj} = \mathcal{A}_n$ is again given by (8). When $F_{o1}(\cdot)$ is completely known as in Bickel, et al. (1993), we simply take $\mathcal{B} = \{\beta_o\}$.

3 Asymptotic Normality and Efficiency of Smooth Functionals

Let $\rho : \mathcal{A} \to \mathcal{R}$ be a smooth functional and $\rho(\widehat{\alpha}_n)$ be the plug-in sieve MLE of $\rho(\alpha_o)$, where $\hat{\alpha}_n$ and α_o are defined in Section 2. In this section, we consider models with unrestricted marginals and apply the general theory of Shen (1997) to establish the asymptotic normality and semiparametric efficiency of the plug-in sieve MLE $\rho(\widehat{\alpha}_n)$ of $\rho(\alpha_o)$.

3.1 Asymptotic Normality and Efficiency of $\rho(\hat{\alpha}_n)$

Let $E_o(\cdot)$ denote the expectation under the true parameter α_o. Let $U_o \equiv (U_{o1}, ..., U_{om})' \equiv (F_{o1}(X_1), ..., F_{om}(X_m))'$ and $u = (u_1, ..., u_m)'$ be an arbitrary value in $[0,1]^m$. In addition, let $c(F_{o1}(X_1), ..., F_{om}(X_m); \theta_o) = c(U_o, \theta_o) = c(\alpha_o)$.

Assumption 1. (1) $\theta_o \in int(\Theta)$, Θ a compact subset of $\mathcal{R}^{d_\theta}$; (2) for $j = 1, ..., m$, $\sqrt{f_{oj}} \in \Lambda^{r_j}(\mathcal{X}_j)$, $r_j > 1/2$; (3) $\alpha_o = (\theta_o', f_{o1}, ..., f_{om})'$ is the unique maximizer of $E_o[\ell(\alpha, Z_i)]$ over $\mathcal{A} = \Theta \times \prod_{j=1}^m \mathcal{F}_j$ with $\mathcal{F}_j = \{f_j = g^2 : g \in \Lambda^{r_j}(\mathcal{X}_j), \int [g(x)]^2 dx = 1\}$.

Assumption 2. the following second order partial derivatives are all well-defined in the neighborhood of α_o: $\frac{\partial^2 \log c(u,\theta)}{\partial \theta^2}$, $\frac{\partial^2 \log c(u,\theta)}{\partial u_j \partial \theta}$, $\frac{\partial^2 \log c(u,\theta)}{\partial u_j \partial u_k}$ for $j, k = 1, ..., m$.

Denote $\mathbf{V}$ as the linear span of $\mathcal{A} - \{\alpha_o\}$. Under Assumption 2, for any $v = (v_\theta', v_1, ..., v_m)' \in \mathbf{V}$, we have that $\ell(\alpha_o + tv, Z)$ is continuously differentiable in small $t \in [0,1]$. Define the directional derivative of $\ell(\alpha, Z)$ at the direction $v \in \mathbf{V}$ (evaluated at α_o) as:

$$
\begin{aligned}
\frac{d\ell(\alpha_o + tv, Z)}{dt}|_{t=0} &\equiv \frac{\partial \ell(\alpha_o, Z)}{\partial \alpha'}[v] = \frac{\partial \ell(\alpha_o, Z)}{\partial \theta'}[v_\theta] + \sum_{j=1}^m \frac{\partial \ell(\alpha_o, Z)}{\partial f_j}[v_j] \\
&= \frac{\partial \log c(\alpha_o)}{\partial \theta'} v_\theta + \sum_{j=1}^m \left\{ \frac{\partial \log c(\alpha_o)}{\partial u_j} \int 1(x \le X_j) v_j(x) dx + \frac{v_j(X_j)}{f_{oj}(X_j)} \right\}.
\end{aligned}
$$

Define the Fisher inner product on the space $\mathbf{V}$ as

$$\langle v, \widetilde{v} \rangle \equiv E_o \left[\left(\frac{\partial \ell(\alpha_o, Z)}{\partial \alpha'}[v] \right) \left(\frac{\partial \ell(\alpha_o, Z)}{\partial \alpha'}[\widetilde{v}] \right) \right], \tag{9}$$

and the Fisher norm for $v \in \mathbf{V}$ as $||v||^2 = \langle v, v \rangle$. Let $\overline{\mathbf{V}}$ be the closed linear span of $\mathbf{V}$ under the Fisher norm. Then $(\overline{\mathbf{V}}, ||\cdot||)$ is a Hilbert space. It is easy to see that $\overline{\mathbf{V}} = \{v = (v_\theta', v_1, ..., v_m)' \in \mathcal{R}^{d_\theta} \times \prod_{j=1}^m \overline{\mathbf{V}}_j : ||v|| < \infty\}$ with

$$\overline{\mathbf{V}}_j = \left\{ v_j : \mathcal{X}_j \to \mathcal{R} : E_o \left(\frac{v_j(X_j)}{f_{oj}(X_j)} \right) = 0, E_o \left(\frac{v_j(X_j)}{f_{oj}(X_j)} \right)^2 < \infty \right\}. \tag{10}$$

It is known that the asymptotic properties of $\rho(\hat{\alpha}_n)$ depend on the smoothness of the functional ρ and the rate of convergence of $\hat{\alpha}_n$. For any $v \in \mathbf{V}$, we denote

$$\frac{\partial \rho(\alpha_o)}{\partial \alpha'}[v] \equiv \lim_{t \to 0}[(\rho(\alpha_o + tv) - \rho(\alpha_o))/t]$$

whenever the right hand-side limit is well defined and assume:

Assumption 3. (1) for any $v \in \mathbf{V}$, $\rho(\alpha_o + tv)$ is continuously differentiable in $t \in [0,1]$ near $t = 0$, and

$$\|\frac{\partial \rho(\alpha_o)}{\partial \alpha'}\| \equiv \sup_{v \in \mathbf{V}: ||v|| > 0} \frac{\left| \frac{\partial \rho(\alpha_o)}{\partial \alpha'}[v] \right|}{||v||} < \infty;$$

(2) there exist constants $c > 0, \omega > 0$, and a small $\varepsilon > 0$ such that for any $v \in \mathbf{V}$ with $||v|| \leq \varepsilon$, we have

$$\left| \rho(\alpha_o + v) - \rho(\alpha_o) - \frac{\partial \rho(\alpha_o)}{\partial \alpha'}[v] \right| \leq c||v||^\omega.$$

Under Assumption 3, by the Riesz representation theorem, there exists $v^* \in \overline{\mathbf{V}}$ such that

$$\langle v^*, v \rangle = \frac{\partial \rho(\alpha_o)}{\partial \alpha'}[v] \quad \text{for all } v \in \mathbf{V} \tag{11}$$

and

$$||v^*||^2 = \|\frac{\partial \rho(\alpha_o)}{\partial \alpha'}\|^2 = \sup_{v \in \mathbf{V}: ||v|| > 0} \frac{\left| \frac{\partial \rho(\alpha_o)}{\partial \alpha'}[v] \right|^2}{||v||^2} < \infty. \tag{12}$$

We make the following assumption on the rate of convergence of $\hat{\alpha}_n$:

Assumption 4. (1) $||\widehat{\alpha}_n - \alpha_o|| = O_P(\delta_n)$ for a decreasing sequence δ_n satisfying $(\delta_n)^\omega = o(n^{-1/2})$; (2) there exists $\Pi_n v^* \in \mathcal{A}_n - \{\alpha_o\}$ such that $\delta_n \times ||\Pi_n v^* - v^*|| = o(n^{-1/2})$.

Theorem 1. Suppose that Assumptions 1-4 and 5-6 stated in Appendix A hold. Then $\sqrt{n}(\rho(\widehat{\alpha}_n) - \rho(\alpha_o)) \Rightarrow \mathcal{N}\left(0, \|\frac{\partial \rho(\alpha_o)}{\partial \alpha'}\|^2\right)$ and $\rho(\widehat{\alpha}_n)$ is semiparametrically efficient.

Discussion of assumptions. Assumptions 1-2 are standard ones. Assumption 3 is essentially the definition of a smooth functional. Assumption 4(1) is a requirement on the convergence rate of the sieve MLEs of unknown marginal densities $\widehat{f}_{nj}$, $j = 1, ..., m$. There exist many results on convergence rates of general sieve estimates of a marginal density; see e.g., Shen and Wong (1994), Wong and Shen (1995), and Van der Geer (2000). There are also many results on particular sieve density estimates; see e.g. Stone (1990) for spline sieve, Barron and Sheu (1991) for polynomial, trigonometric and spline sieves, Chen and White (1999) for neural network sieve, Coppejans and Gallant (2002) for Hermite polynomial sieve. Assumption 4(2) requires that the Riesz representer has a little bit of smoothness. Although Assumptions 3 and 4(2) are stated in terms of data $Z_i = (X_{1i}, ..., X_{mi})'$, and the Fisher norm $||v||$ on the perturbation space $\overline{\mathbf{V}}$, it is often easier to verify these assumptions in terms of transformed variables. Let

$$\mathcal{L}_2^0([0,1]) \equiv \left\{ e : [0,1] \to \mathcal{R} : \int_0^1 e(v)dv = 0, \int_0^1 [e(v)]^2 dv < \infty \right\}.$$

By change of variables, for any $v_j \in \overline{\mathbf{V}}_j$ there is a unique function $b_j \in \mathcal{L}_2^0([0,1])$ with $b_j(u_j) = \frac{v_j(F_{oj}^{-1}(u_j))}{f_{oj}(F_{oj}^{-1}(u_j))}$, and vice versa. Therefore we can always rewrite $\frac{\partial \ell(\alpha_o, Z)}{\partial \alpha'}[v]$ as follows:

$$\begin{aligned} \frac{\partial \ell(\alpha_o, Z)}{\partial \alpha'}[v] &= \frac{\partial \ell(\alpha_o, U_o)}{\partial \alpha'}[(v_\theta', b_1, ..., b_m)'] \\ &= \frac{\partial \log c(\alpha_o)}{\partial \theta'} v_\theta + \sum_{j=1}^m \left\{ \frac{\partial \log c(\alpha_o)}{\partial u_j} \int_0^{U_{oj}} b_j(y)dy + b_j(U_{oj}) \right\} \end{aligned}$$

and

$$\begin{aligned} ||v||^2 &= E_o \left[\left(\frac{\partial \ell(\alpha_o, U_o)}{\partial \alpha'}[(v_\theta', b_1, ..., b_m)'] \right)^2 \right] \\ &= E_o \left[\left(\frac{\partial \log c(\alpha_o)}{\partial \theta'} v_\theta + \sum_{j=1}^m \left\{ \frac{\partial \log c(\alpha_o)}{\partial u_j} \int_0^{U_{oj}} b_j(y)dy + b_j(U_{oj}) \right\} \right)^2 \right] \end{aligned}$$

Define

$$\overline{\mathbf{B}} = \left\{ b = (v_\theta', b_1, ..., b_m)' \in \mathcal{R}^{d_\theta} \times \prod_{j=1}^m \mathcal{L}_2^0([0,1]) : ||b||^2 \equiv E_o \left[\left(\frac{\partial \ell(\alpha_o, U_o)}{\partial \alpha'}[b] \right)^2 \right] < \infty \right\}.$$

Then there is an one-to-one onto mapping between the two Hilbert spaces $(\overline{\mathbf{B}}, ||\cdot||)$ and $(\overline{\mathbf{V}}, ||\cdot||)$. Now it is easy to see that the Riesz representer $v^* = (v_\theta^{*\prime}, v_1^*, ..., v_m^*)' \in \overline{\mathbf{V}}$ is uniquely determined by $b^* = (v_\theta^{*\prime}, b_1^*, ..., b_m^*)' \in \overline{\mathbf{B}}$ (and vise versa) via the relation:

$$v_j^*(x_j) = b_j^*(F_{oj}(x_j)) f_{oj}(x_j) \quad \text{for all } x_j \in \mathcal{X}_j, \quad \text{for } j = 1, ..., m.$$

Then Assumption 4(2) can be replaced by

Assumption 4'(2): there exists $\Pi_n b^* = (v_\theta^{*\prime}, \Pi_{n1} b_1^*, ..., \Pi_{nm} b_m^*)' \in \mathcal{R}^{d_\theta} \times \prod_{j=1}^m \mathbf{B}_{nj}$ such that

$$||\Pi_n b^* - b^*||^2 = E_o \left(\sum_{j=1}^m \left\{ \frac{\partial \log c(\alpha_o)}{\partial u_j} \int_0^{U_{oj}} \{\Pi_n b_j^* - b_j^*\}(y)dy + \{\Pi_n b_j^* - b_j^*\}(U_{oj}) \right\} \right)^2 = o\left(\frac{1}{n\delta_n^2} \right)$$

where for $j = 1, ..., m$, $\mathbf{B}_{nj}$ is a sieve for $\mathcal{L}_2^0([0,1])$.

Although many sieves including Spl$(1, K_n)$, Pol(K_n) and TriPol(K_n) can be used as $\mathbf{B}_{nj}$ for the space $\mathcal{L}_2^0([0,1])$, due to its simple structure we recommend the following one:

$$\mathbf{B}_{nj} = \{e(u) = \sum_{k=1}^{K_{nj}} a_k \sqrt{2} \cos(k\pi u),\ u \in [0,1], \sum_{k=1}^{K_{nj}} a_k^2 < \infty\}.$$

3.2 $\sqrt{n}$−Normality and Efficiency of $\widehat{\theta}_n$

We take $\rho(\alpha) = \lambda'\theta$ for any arbitrarily fixed $\lambda \in \mathcal{R}^{d_\theta}$ with $0 < |\lambda| < \infty$. It satisfies Assumption 3(2) with $\frac{\partial \rho(\alpha_o)}{\partial \alpha'}[v] = \lambda' v_\theta$ and $\omega = \infty$. Assumption 3(1) is equivalent to finding a Riesz representer $v^* \in \overline{\mathbf{V}}$ satisfying (13) and (14):

$$\lambda'(\theta - \theta_o) = \langle \alpha - \alpha_o, v^* \rangle \quad \text{for any } \alpha - \alpha_o \in \overline{\mathbf{V}} \tag{13}$$

and

$$\|\frac{\partial \rho(\alpha_o)}{\partial \alpha'}\|^2 = ||v^*||^2 = \langle v^*, v^* \rangle = \sup_{v \neq 0, v \in \overline{\mathbf{V}}} \frac{|\lambda' v_\theta|^2}{||v||^2} < \infty. \tag{14}$$

Notice that

$$\begin{aligned} \sup_{v \neq 0, v \in \overline{\mathbf{V}}} \frac{|\lambda' v_\theta|^2}{||v||^2} &= \sup_{b \neq 0, b \in \overline{\mathbf{B}}} \left\{ \frac{|\lambda' v_\theta|^2}{E_o \left[\left(\frac{\partial \log c(\alpha_o)}{\partial \theta'} v_\theta + \sum_{j=1}^m \left\{ \frac{\partial \log c(\alpha_o)}{\partial u_j} \int_0^{U_{oj}} b_j(y) dy + b_j(U_{oj}) \right\} \right)^2 \right]} \right\} \\ &= \lambda' \mathcal{I}_*(\theta_o)^{-1} \lambda = \lambda' \left(E_o[\mathcal{S}_{\theta_o} \mathcal{S}'_{\theta_o}] \right)^{-1} \lambda \end{aligned}$$

where

$$\mathcal{S}'_{\theta_o} = \frac{\partial \log c(U_o, \theta_o)}{\partial \theta'} - \sum_{j=1}^m [\frac{\partial \log c(U_o, \theta_o)}{\partial u_j} \int_0^{U_{oj}} g_j^*(u) du + g_j^*(U_{oj})], \tag{15}$$

and $g_j^* = (g_{j,1}^*, ..., g_{j,d_\theta}^*) \in \prod_{k=1}^{d_\theta} \mathcal{L}_2^0([0,1]), j = 1, ..., m$ solves the following infinite-dimensional optimization problems for $k = 1, ..., d_\theta$,

$$\inf_{g_{1,k}, \dots, g_{m,k} \in \mathcal{L}_2^0([0,1])} E_o \left\{ \left(\frac{\partial \log c(U_o, \theta_o)}{\partial \theta_k} - \sum_{j=1}^m [\frac{\partial \log c(U_o, \theta_o)}{\partial u_j} \int_0^{U_{oj}} g_{j,k}(v) dv + g_{j,k}(U_{oj})] \right)^2 \right\}.$$

Therefore $b^* = (v_\theta^{*\prime}, b_1^*, ..., b_m^*)'$ with $v_\theta^* = \mathcal{I}_*(\theta_o)^{-1}\lambda$ and $b_j^*(u_j) = -g_j^*(u_j) \times v_\theta^*$, and

$$v^* = (I_{d_\theta}, -g_1^*(F_{o1}(x_1)) f_{o1}(x_1), ..., -g_m^*(F_{om}(x_m)) f_{om}(x_m)) \times \mathcal{I}_*(\theta_o)^{-1}\lambda.$$

Hence (14) is satisfied if and only if $\mathcal{I}_*(\theta_o) = E_o[\mathcal{S}_{\theta_o} \mathcal{S}'_{\theta_o}]$ is *non-singular,* which in turn is satisfied under the following assumption:

Assumption 3': (1) $\frac{\partial \log c(U_o, \theta_o)}{\partial \theta}$, $\frac{\partial \log c(U_o, \theta_o)}{\partial u_j}$, $j = 1, ..., m$ have finite second moments;

(2) $\mathcal{I}(\theta_o) \equiv E_o[\frac{\partial \log c(U_o,\theta_o)}{\partial \theta}\frac{\partial \log c(U_o,\theta_o)}{\partial \theta'}]$ is finite and positive definite;
(3) $\int \frac{\partial c(u,\theta_o)}{\partial u_j} du_{-j} = \frac{\partial}{\partial u_j}\int c(u,\theta_o)du_{-j} = 0$ for $(j,-j) = (1,...,m)$ with $j \neq -j$;
(4) $\int \frac{\partial^2 c(u,\theta_o)}{\partial u_j \partial \theta} du_{-j} = \frac{\partial^2}{\partial u_j \partial \theta}\int c(u,\theta_o)du_{-j} = 0$ for $(j,-j) = (1,...,m)$ with $j \neq -j$;
(5) there exists a constant K such that

$$\max_{j=1,...,m} \sup_{0<u_j<1} E\left[\left(u_j(1-u_j)\frac{\partial \log c(U_o,\theta_o)}{\partial u_j}\right)^2 \mid U_{oj} = u_j\right] \leq K.$$

We can now apply Theorem 1 to obtain the following result:

Proposition 1. Suppose that Assumptions 1 - 2, 3', 4 - 6 hold. Then $\sqrt{n}(\widehat{\theta}_n - \theta_o) \Rightarrow \mathcal{N}\left(0, \mathcal{I}_*(\theta_o)^{-1}\right)$ and $\widehat{\theta}_n$ is semiparametrically efficient.

To make inferences on θ_o using the sieve MLE $\widehat{\theta}_n$, we need to estimate its asymptotic variance or $\mathcal{I}_*(\theta_o)$. If there is a closed-form expression of $\mathcal{I}_*(\theta_o)$ then it can be consistently estimated by the direct plug-in estimator $\mathcal{I}_*(\widehat{\theta}_n)$. Unfortunately, only recently Klaassen and Wellner (1997) derive a closed-form expression of $\mathcal{I}_*(\theta_o)$ for the bivariate Gaussian copula model with unknown margins. In general there is no closed-form solutions of $\mathcal{I}_*(\theta_o)$ for multivariate copula models with unknown margins, hence direct plug-in estimation of $\mathcal{I}_*(\theta_o)$ is difficult. We propose a sieve estimator of $\mathcal{I}_*(\theta_o)$ based on its characterization in (15). Let $\widehat{U}_i = (\widehat{U}_{1i}, ..., \widehat{U}_{mi})' = (\widehat{F}_{n1}(X_{1i}), ..., \widehat{F}_{nm}(X_{mi}))'$ for $i = 1, ..., n$. Let $\mathbf{A}_n$ be some sieve space such as:

$$\mathbf{A}_n = \{(e_1, ..., e_{d_\theta}) : e_j(\cdot) \in \mathbf{B}_n,\, j = 1, ..., d_\theta\}, \tag{16}$$

$$\mathbf{B}_n = \{e(u) = \sum_{k=1}^{K_{n\theta}} a_k \sqrt{2}\cos(k\pi u),\, u \in [0,1], \sum_{k=1}^{K_{n\theta}} a_k^2 < \infty\}, \tag{17}$$

where $K_{n\theta} \to \infty$, $(K_{n\theta})^{d_\theta}/n \to 0$. We can now compute

$$\widehat{\sigma}_\theta^2 = \min_{\substack{g_j \in \mathbf{A}_n,\\ j=1,...,m}} \frac{1}{n}\sum_{i=1}^n \left\{ \begin{array}{c} \left(\frac{\partial \log c(\widehat{U}_i,\widehat{\theta}_n)}{\partial \theta'} - \sum_{j=1}^m [\frac{\partial \log c(\widehat{U}_i,\widehat{\theta}_n)}{\partial u_j}\int_0^{\widehat{U}_{ji}} g_j(v)dv + g_j(\widehat{U}_{ji})]\right)' \times \\ \left(\frac{\partial \log c(\widehat{U}_i,\widehat{\theta}_n)}{\partial \theta'} - \sum_{j=1}^m [\frac{\partial \log c(\widehat{U}_i,\widehat{\theta}_n)}{\partial u_j}\int_0^{\widehat{U}_{ji}} g_j(v)dv + g_j(\widehat{U}_{ji})]\right) \end{array} \right\}.$$

Proposition 2. Under the assumptions in Proposition 1, we have: $\widehat{\sigma}_\theta^2 = \mathcal{I}_*(\theta_o) + o_p(1)$.

3.3 Sieve MLE of F_{oj}

For $j = 1, ..., m$, we consider the estimation of $\rho(\alpha_o) = F_{oj}(x_j)$ for some fixed $x_j \in \mathcal{X}_j$ by the plug-in sieve MLE: $\rho(\widehat{\alpha}) = \widehat{F}_{nj}(x_j) = \int 1(y \leq x_j)\widehat{f}_{nj}(y)dy$, where $\widehat{f}_{nj}$ is the sieve MLE from (8). Clearly $\frac{\partial \rho(\alpha_o)}{\partial \alpha'}[v] = \int_{\mathcal{X}_j} 1(y \leq x_j)v_j(y)dy$ for any $v = (v_\theta', v_1, ..., v_m)' \in \mathbf{V}$. It is easy to see that $\omega = \infty$ in Assumptions 3 and 4, and

$$\|\frac{\partial \rho(\alpha_o)}{\partial \alpha'}\|^2 = \sup_{v \in \mathbf{V}: ||v|| > 0} \frac{\left|\int_{\mathcal{X}_j} 1(y \leq x_j)v_j(y)dy\right|^2}{||v||^2} < \infty.$$

Hence the representer $\upsilon^* \in \overline{\mathbf{V}}$ should satisfy (18) and (19):

$$\langle \upsilon^*, v\rangle = \frac{\partial \rho(\alpha_o)}{\partial \alpha'}[v] = E_o\left(1(X_j \leq x_j)\frac{v_j(X_j)}{f_{oj}(X_j)}\right) \quad \text{for all } v \in \mathbf{V} \tag{18}$$

$$\|\frac{\partial \rho(\alpha_o)}{\partial \alpha'}\|^2 = ||\upsilon^*||^2 = ||b^*||^2 = \sup_{b \in \overline{\mathbf{B}}: ||b||>0} \frac{\left|E_o\left(1(U_{oj} \leq F_{oj}(x_j))b_j(U_{oj})\right)\right|^2}{||b||^2}. \tag{19}$$

Proposition 3. Let $\upsilon^* \in \overline{\mathbf{V}}$ solve (18) and (19). Suppose that Assumptions 1 - 2 and 4 - 6 hold. Then for any fixed $x_j \in \mathcal{X}_j$ and for $j = 1, ..., m$, $\sqrt{n}(\widehat{F}_{nj}(x_j) - F_{oj}(x_j)) \Rightarrow \mathcal{N}\left(0, ||\upsilon^*||^2\right)$. Moreover, $\widehat{F}_{nj}$ is semiparametrically efficient.

Again for general semiparametric copula models including the Gaussian copula with unknown margins, there are currently no closed-form solutions for the asymptotic variance $||\upsilon^*||^2$. Nevertheless, we can again consistently estimate $||\upsilon^*||^2$ by the sieve method. Let

$$\widehat{\sigma}^2_{F_j}(x_j) = \max_{\substack{v_\theta \neq 0, b_k \in \mathbf{B}_n,\\ k=1,...,m}} \frac{\left|\frac{1}{n}\sum_{i=1}^n 1\{\widehat{U}_{ji} \leq \widehat{F}_{nj}(x_j)\} b_j(\widehat{U}_{ji})\right|^2}{\frac{1}{n}\sum_{i=1}^n \left[\frac{\partial \log c(\widehat{U}_i, \widehat{\theta})}{\partial \theta'} v_\theta + \sum_{k=1}^m \left[\frac{\partial \log c(\widehat{U}_i, \widehat{\theta})}{\partial u_k} \int_0^{\widehat{U}_{ki}} b_k(u)du + b_k(\widehat{U}_{ki})\right]\right]^2},$$

where $\widehat{U}_i = (\widehat{F}_{n1}(X_{1i}), ..., \widehat{F}_{nm}(X_{mi}))'$, and $\mathbf{B}_n$ is given in (17).

Proposition 4. Under assumptions in Proposition 3, we have for any fixed $x_j \in \mathcal{X}_j$ and $j = 1, ..., m$, $\widehat{\sigma}^2_{F_j}(x_j) = ||\upsilon^*||^2 + o_p(1)$.

Remark 2: In the special case of the independence copula ($c(u_1, ..., u_m, \theta) = 1$), we could solve (18) and (19) explicitly. We note that for the independence copula,

$$\langle \widetilde{v}, v\rangle = \sum_{k=1}^m E_o\left(\frac{\widetilde{v}_k(X_k)}{f_{ok}(X_k)}\frac{v_k(X_k)}{f_{ok}(X_k)}\right) \quad \text{for all } \widetilde{v}, v \in \mathbf{V}.$$

Thus (18) and (19) are satisfied with $\upsilon^*_j(X_j) = \{1(X_j \leq x_j) - E_o[1(X_j \leq x_j)]\} f_{oj}(X_j)$ and $\upsilon^*_k = 0$ for all $k \neq j$. Hence

$$||\upsilon^*||^2 = E_o\left(1(X_j \leq x_j)\left\{1(X_j \leq x_j) - E_o[1(X_j \leq x_j)]\right\}\right) = F_{oj}(x_j)\{1 - F_{oj}(x_j)\}.$$

Thus for models with the independence copula, the plug-in sieve MLE of F_{oj} satisfies

$$\sqrt{n}\left(\widehat{F}_{nj}(x_j) - F_{oj}(x_j)\right) \Rightarrow \mathcal{N}\left(0, F_{oj}(x_j)\{1 - F_{oj}(x_j)\}\right),$$

where its asymptotic variance coincides with that of the standard empirical cdf estimate $F_{nj}(x_j) = \frac{1}{n}\sum_{i=1}^n 1\{X_{ji} \leq x_j\}$ of F_{oj}. For models with parametric copula functions that are not independent, we have $||\upsilon^*||^2 \leq F_{oj}(x_j)\{1 - F_{oj}(x_j)\}$.

4 Sieve MLE with Restrictions on Marginals

In this section, we present the asymptotic normality and efficiency results for sieve MLEs of θ_o and F_{oj} under restrictions on marginal distributions considered in Section 2.

4.1 Equal but Unknown Margins

Now the Fisher norm becomes $||v||^2 = E_o\{\frac{\partial \ell(\alpha_o,Z)}{\partial \alpha'}[v]\}^2$ with

$$\frac{\partial \ell(\alpha_o, Z)}{\partial \alpha'}[v] = \frac{\partial \log c(U_o, \theta_o)}{\partial \theta'} v_\theta + \sum_{j=1}^{m} \left\{ \frac{\partial \log c(U_o, \theta_o)}{\partial u_j} \int^{X_j} v_1(x)dx + \frac{v_1(X_j)}{f_o(X_j)} \right\},$$

$U_o = (F_o(X_1), ..., F_o(X_m))'$ and $v \in \overline{\mathbf{V}} = \{v = (v_\theta', v_1)' \in \mathcal{R}^{d_\theta} \times \overline{\mathbf{V}}_1 : ||v|| < \infty\}$ with $\overline{\mathbf{V}}_1$ given in (10).

Proposition 5. Suppose Assumptions 1-2, 3', 4-6 hold and $f_{oj} = f_o$ for $j = 1, ..., m$. Then

(i) $\widehat{\theta}_n$ is semiparametrically efficient and $\sqrt{n}(\widehat{\theta}_n - \theta_o) \Rightarrow \mathcal{N}\left(0, \mathcal{I}_*(\theta_o)^{-1}\right)$ where $\mathcal{I}_*(\theta_o) =$

$$\inf_{g \in \prod_{k=1}^{d_\theta} \mathcal{L}_2^0([0,1])} E_o \left\{ \begin{array}{c} \left(\frac{\partial \log c(U_o,\theta_o)}{\partial \theta'} - \sum_{j=1}^m [\frac{\partial \log c(U_o,\theta_o)}{\partial u_j} \int_0^{U_{oj}} g(u)du + g(U_{oj})]\right)' \times \\ \left(\frac{\partial \log c(U_o,\theta_o)}{\partial \theta'} - \sum_{j=1}^m [\frac{\partial \log c(U_o,\theta_o)}{\partial u_j} \int_0^{U_{oj}} g(u)du + g(U_{oj})]\right) \end{array} \right\};$$

(ii) for any fixed $x \in \mathcal{X}$, $\widehat{F}_n(x) = \int 1(y \leq x)\widehat{f}_n(y)dy$ is semiparametrically efficient and $\sqrt{n}(\widehat{F}_n(x) - F_o(x)) \Rightarrow \mathcal{N}\left(0, ||v^*||^2\right)$ where $||v^*||^2 = ||b^*||^2 =$

$$\sup_{\substack{v_\theta \neq 0, \\ b \in \mathcal{L}_2^0([0,1])}} \frac{|E_o\{1(U_{o1} \leq F_o(x))b(U_{o1})\}|^2}{E_o\left[\left(\frac{\partial \log c(U_o,\theta_o)}{\partial \theta'} v_\theta + \sum_{k=1}^m \left\{\frac{\partial \log c(U_o,\theta_o)}{\partial u_k} \int_0^{U_{ok}} b(u)du + b(U_{ok})\right\}\right)^2\right]}.$$

Comparing the asymptotic variances of the estimators of θ_o and F_{oj} in Proposition 5 with those in Propositions 1 and 3, we see that exploiting the restriction of equal marginals in general leads to more efficient estimators of the copula parameter θ_o and the marginal distributions.

Proposition 6. Under conditions in Proposition 5, we have:

(i) $\widehat{\sigma}_\theta^2 = \mathcal{I}_*(\theta_o) + o_p(1)$, where

$$\widehat{\sigma}_\theta^2 = \min_{g \in \mathbf{A}_n} \frac{1}{n} \sum_{i=1}^n \left\{ \begin{array}{c} \left(\frac{\partial \log c(\widehat{U}_i,\widehat{\theta}_n)}{\partial \theta'} - \sum_{j=1}^m [\frac{\partial \log c(\widehat{U}_i,\widehat{\theta}_n)}{\partial u_j} \int_0^{\widehat{U}_{ji}} g(u)du + g(\widehat{U}_{ji})]\right)' \times \\ \left(\frac{\partial \log c(\widehat{U}_i,\widehat{\theta}_n)}{\partial \theta'} - \sum_{j=1}^m [\frac{\partial \log c(\widehat{U}_i,\widehat{\theta}_n)}{\partial u_j} \int_0^{\widehat{U}_{ji}} g(u)du + g(\widehat{U}_{ji})]\right) \end{array} \right\};$$

(ii) $\widehat{\sigma}_F^2(x) = ||v^*||^2 + o_p(1)$, where

$$\widehat{\sigma}_F^2(x) = \max_{v_\theta \neq 0, b \in \mathbf{B}_n} \frac{\left|\frac{1}{n}\sum_{i=1}^n 1\{\widehat{U}_{1i} \leq \widehat{F}_n(x)\} b(\widehat{U}_{1i})\right|^2}{\frac{1}{n}\sum_{i=1}^n \left[\frac{\partial \log c(\widehat{U}_i,\widehat{\theta}_n)}{\partial \theta'} v_\theta + \sum_{k=1}^m [\frac{\partial \log c(\widehat{U}_i,\widehat{\theta}_n)}{\partial u_k} \int_0^{\widehat{U}_{ki}} b(u)du + b(\widehat{U}_{ki})]\right]^2},$$

in which $\widehat{U}_i = (\widehat{F}_n(X_{1i}), ..., \widehat{F}_n(X_{mi}))'$, $\mathbf{A}_n$ is the sieve space (16), and $\mathbf{B}_n$ is the sieve space (17).

4.2 Models with a Parametric Margin

In this case, the Fisher norm becomes $||v||^2 = E_o\{\frac{\partial \ell(\alpha_o,Z)}{\partial \alpha'}[v]\}^2$ with

$$\begin{aligned}
\frac{\partial \ell(\alpha_o, Z)}{\partial \alpha'}[v] &= \frac{\partial \log c(U_o,\theta_o)}{\partial \theta'} v_\theta + \frac{\partial \ell(\alpha_o, Z)}{\partial \beta'} v_\beta + \sum_{j=2}^m \left\{ \frac{\partial \log c(U_o,\theta_o)}{\partial u_j} \int^{X_j} v_j(x)dx + \frac{v_j(X_j)}{f_{oj}(X_j)} \right\}, \\
\frac{\partial \ell(\alpha_o, Z)}{\partial \beta'} v_\beta &= \left[\frac{\partial \log c(U_o,\theta_o)}{\partial u_1} \int^{X_1} \frac{\partial f_{o1}(x,\beta_o)}{\partial \beta'} dx + \frac{1}{f_{o1}(X_1,\beta_o)} \frac{\partial f_{o1}(X_1,\beta_o)}{\partial \beta'} \right] v_\beta,
\end{aligned}$$

where $U_o = (F_{o1}(X_1,\beta_o), F_{o2}(X_2), ..., F_{om}(X_m))'$ and $v \in \overline{\mathbf{V}} = \{v = (v_\theta', v_\beta', v_2, \ldots, v_m)' \in \mathcal{R}^{d_\theta} \times \mathcal{R}^{d_\beta} \times \Pi_{j=2}^m \overline{\mathbf{V}}_j : ||v|| < \infty\}$ with $\overline{\mathbf{V}}_j$ given in (10).

Proposition 7. Suppose that Assumptions 1-2, 3', 4-6 hold, $F_{o1}(\cdot) = F_{o1}(\cdot,\beta_o)$ for unknown $\beta_o \in int(\mathcal{B})$ and $E\left[\frac{\partial \log f_{o1}(X_1,\beta_o)}{\partial \beta} \frac{\partial \log f_{o1}(X_1,\beta_o)}{\partial \beta'}\right]$ is positive definite. Then
(i) $\widehat{\theta}_n$ is semiparametrically efficient and $\sqrt{n}(\widehat{\theta}_n - \theta_o) \Rightarrow \mathcal{N}\left(0, \mathcal{I}_*(\theta_o)^{-1}\right)$ where $\mathcal{I}_*(\theta_o) = E_o[\mathcal{S}_{\theta_o}\mathcal{S}_{\theta_o}']$ with $\mathcal{S}_{\theta_o}' = \left(\mathcal{S}_{\theta_{o1}}, ..., \mathcal{S}_{\theta_{od_\theta}}\right)$ and for $k = 1, ..., d_\theta$,

$$\mathcal{S}_{\theta_{ok}} = \frac{\partial \log c(U_o,\theta_o)}{\partial \theta_k} - \frac{\partial \ell(\alpha_o, Z)}{\partial \beta'} a_k^* - \sum_{j=2}^m [\frac{\partial \log c(U_o,\theta_o)}{\partial u_j} \int_0^{U_{oj}} g_{j,k}^*(u)du + g_{j,k}^*(U_{oj})]$$

solves the following optimization problem:

$$\inf_{\substack{a_k \in \mathcal{R}^{d_\beta}, a_k \neq 0, \\ g_{j,k} \in \mathcal{L}_2^0([0,1])}} E_o\left\{ \left(\frac{\partial \log c(U_o,\theta_o)}{\partial \theta_k} - \frac{\partial \ell(\alpha_o, Z)}{\partial \beta'} a_k - \sum_{j=2}^m [\frac{\partial \log c(U_o,\theta_o)}{\partial u_j} \int_0^{U_{oj}} g_{j,k}(u)du + g_{j,k}(U_{oj})] \right)^2 \right\};$$

(ii) for any fixed $x \in \mathcal{X}$ and for $j = 2, \ldots, m$, $\widehat{F}_{nj}(x) = \int 1(y \leq x)\widehat{f}_{nj}(y)dy$ is semiparametrically efficient and $\sqrt{n}(\widehat{F}_{nj}(x) - F_{oj}(x)) \Rightarrow \mathcal{N}\left(0, ||v^*||^2\right)$ where $||v^*||^2 = ||b^*||^2 =$

$$\sup_{\substack{v_\theta \neq 0, v_\beta \neq 0, \\ b_k \in \mathcal{L}_2^0([0,1])}} \frac{|E_o\{1(U_{oj} \leq F_{oj}(x))b_j(U_{oj})\}|^2}{E_o\left[\left(\frac{\partial \log c(U_o,\theta_o)}{\partial \theta'} v_\theta + \frac{\partial \ell(\alpha_o,Z)}{\partial \beta'} v_\beta + \sum_{k=2}^m \left\{\frac{\partial \log c(U_o,\theta_o)}{\partial u_k} \int_0^{U_{ok}} b_k(u)du + b_k(U_{ok})\right\}\right)^2\right]}.$$

Proposition 8. Under conditions in Proposition 7, we have:

(i) $\widehat{\sigma}_\theta^2 = \mathcal{I}_*(\theta_o) + o_p(1)$, where $\widehat{\sigma}_\theta^2 =$

$$\min_{\substack{a \neq 0, \\ g_j \in \mathbf{A}_n}} \frac{1}{n} \sum_{i=1}^n \left\{ \begin{array}{l} \left(\frac{\partial \log c(\widehat{U}_i,\widehat{\theta}_n)}{\partial \theta'} - \frac{\partial \ell(\widehat{\alpha},Z_i)}{\partial \beta'} a - \sum_{j=2}^m [\frac{\partial \log c(\widehat{U}_i,\widehat{\theta}_n)}{\partial u_j} \int_0^{\widehat{U}_{ji}} g_j(v)dv + g_j(\widehat{U}_{ji})] \right)' \\ \left(\frac{\partial \log c(\widehat{U}_i,\widehat{\theta}_n)}{\partial \theta'} - \frac{\partial \ell(\widehat{\alpha},Z_i)}{\partial \beta'} a - \sum_{j=2}^m [\frac{\partial \log c(\widehat{U}_i,\widehat{\theta}_n)}{\partial u_j} \int_0^{\widehat{U}_{ji}} g_j(v)dv + g_j(\widehat{U}_{ji})] \right) \end{array} \right\};$$

(ii) $\widehat{\sigma}_{F_j}^2(x_j) = ||v^*||^2 + o_p(1)$, where $\widehat{\sigma}_{F_j}^2(x_j) =$

$$\max_{\substack{v_\theta \neq 0, v_\beta \neq 0, \\ b_k \in \mathbf{B}_n}} \frac{\frac{1}{n}\left|\sum_{i=1}^n 1\{\widehat{U}_{ji} \leq \widehat{F}_{nj}(x_j)\} b_j(\widehat{U}_{ji})\right|^2}{\sum_{i=1}^n \left[\frac{\partial \log c(\widehat{U}_i,\widehat{\theta})}{\partial \theta'} v_\theta + \frac{\partial \ell(\widehat{\alpha},Z_i)}{\partial \beta'} v_\beta + \sum_{k=2}^m [\frac{\partial \log c(\widehat{U}_i,\widehat{\theta})}{\partial u_k} \int_0^{\widehat{U}_{ki}} b_k(u)du + b_k(\widehat{U}_{ki})]\right]^2},$$

where $\widehat{U}_i = (F_{o1}(X_{1i};\widehat{\beta}), ..., \widehat{F}_{nm}(X_{mi}))'$.

Remark 3: Suppose further that the margin $F_{o1}(\cdot) = F_{o1}(\cdot, \beta_o)$ is completely known. Let $\hat{\alpha}_n = (\hat{\theta}_n, \beta_o, \hat{f}_{n2}, \ldots, \hat{f}_{nm})$ be the corresponding sieve MLE of $\alpha_o = (\theta_o, \beta_o, f_{o2}, ..., f_{om})$. Then the conclusions of Proposition 7 still hold after we drop the term "$\frac{\partial \ell(\alpha_o,Z)}{\partial \beta'} v_\beta$" from the definition of the Fisher norm and from the calculation of asymptotic variances. Moreover, the asymptotic variance of $\sqrt{n}(\widehat{\theta}_n - \theta_o)$ can be consistently estimated by $\{\widehat{\sigma}^2_\theta\}^{-1}$, with

$$\widehat{\sigma}^2_\theta = \min_{\substack{g_j \in \mathbf{A}_n, \\ j=2,...,m}} \frac{1}{n} \sum_{i=1}^{n} \left\{ \begin{array}{c} \left(\frac{\partial \log c(\widehat{U}_i, \widehat{\theta}_n)}{\partial \theta'} - \sum_{j=2}^{m} [\frac{\partial \log c(\widehat{U}_i, \widehat{\theta}_n)}{\partial u_j} \int_0^{\widehat{U}_{ji}} g_j(v) dv + g_j(\widehat{U}_{ji})] \right)' \times \\ \left(\frac{\partial \log c(\widehat{U}_i, \widehat{\theta}_n)}{\partial \theta'} - \sum_{j=2}^{m} [\frac{\partial \log c(\widehat{U}_i, \widehat{\theta}_n)}{\partial u_j} \int_0^{\widehat{U}_{ji}} g_j(v) dv + g_j(\widehat{U}_{ji})] \right) \end{array} \right\},$$

and the asymptotic variance of $\sqrt{n}(\widehat{F}_{nj}(x) - F_{oj}(x))$ can be consistently estimated by $\widehat{\sigma}^2_{F_j}(x_j)$, with

$$\widehat{\sigma}^2_{F_j}(x_j) = \max_{\substack{v_\theta \neq 0, b_k \in \mathbf{B}_n, \\ k=2,...,m}} \frac{\left| \frac{1}{n} \sum_{i=1}^{n} 1\{\widehat{U}_{ji} \leq \widehat{F}_{nj}(x_j)\} b_j(\widehat{U}_{ji}) \right|^2}{\frac{1}{n} \sum_{i=1}^{n} \left[\frac{\partial \log c(\widehat{U}_i, \widehat{\theta})}{\partial \theta'} v_\theta + \sum_{k=2}^{m} [\frac{\partial \log c(\widehat{U}_i, \widehat{\theta})}{\partial u_k} \int_0^{\widehat{U}_{ki}} b_k(u) du + b_k(\widehat{U}_{ki})] \right]^2},$$

where $\widehat{U}_i = (F_{o1}(X_{1i}), \widehat{F}_{n2}(X_{2i}), ..., \widehat{F}_{nm}(X_{mi}))'$, $\mathbf{A}_n$ is the sieve space (16), and $\mathbf{B}_n$ is the sieve space (17).

5 Simulation Study

To investigate the finite sample performance of the proposed sieve MLEs of the copula parameter and the marginals, we conduct an extensive simulation study that covers three families of marginals, four families of bivariate copulas and two families of tri-variate copulas. These copulas exhibit a wide range of dependence structures. For comparison purposes, we include the ideal MLE of the copula parameter in which all the margins are assumed to be known, the two-step estimator of the copula parameter, and the empirical cdfs of the marginals. We also investigate the relative performance of sieve MLEs of the copula parameter and the marginals under two kinds of prior information on marginals: (a) one margin is known or parametric, (b) all margins are equal but are unspecified. As a fair comparison, we include the modified two-step estimator of the copula parameter in which the first step marginals are estimated under prior information on marginals. All the results reported in this section are based on 500 Monte Carlo replications of estimates using a random sample $\{Z_i \equiv (X_{1i}, ..., X_{mi})'\}_{i=1}^n$ with $n = 400$.

5.1 Implementation of Sieve MLEs

We have tried both B-spline sieve and Hermite polynomial sieve to approximate square root of a marginal density in our simulation studies and have found out they perform similarly. To save space we only report the simulation results for the sieve MLEs using B-spline basis. Let $\{B_\gamma(x-k)\}_{k=1}^{K_{nj}}$

be the γ-th order B-spline basis. Then the marginal density function f_{oj} can be approximated by

$$f_j(x; a_j) = \frac{\left(\sum_{k=1}^{K_{nj}} a_{jk} B_\gamma(x-k)\right)^2}{\int \left(\sum_{k=1}^{K_{nj}} a_{jk} B_\gamma(x-k)\right)^2 dx},$$

where $j = 1, \ldots, m$. Throughout the simulation study, we use the 3rd order B-splines, *i.e.*, $\gamma = 3$. We approximate the density f_{oj} on the support $[\min(X_{ji}) - s_{X_j}, \max(X_{ji}) + s_{X_j}]$, where s_{X_j} is the sample standard deviation of $\{X_{ji}\}_{i=1}^n$. The number of sieve coefficients is dictated by the support of the density. Let $b_{1j} = \max(z \le \min(X_{ji}) - s_{X_j} : z \text{ is integer})$, and $b_{2j} = \min(z \ge \max(X_{ji}) + s_{X_j} : z \text{ is integer})$. Then for B-splines of order γ, we need $K_{nj} = b_{2j} - b_{1j} + 1 - \gamma$ sieve coefficients to 'cover' the interval $[b_{1j}, b_{2j}]$. To evaluate the integral that appears in the denominator we use a grid of equidistant points on $[b_{1j}, b_{2j}]$. The results reported in this paper correspond to grid size 0.01, but we also tried value 0.005, which gives very similar results. In each case, the sieve MLE is computed as that in Remark 1 using $Pen(f_j) = \|(\sqrt{f_j})'\|_2^2$ penalty.[3] We have tried penalization factors of values 0.01, 0.001, 0.0001 and have found the results are similar. The results reported in this section use 0.001 as the penalization factor for all the bivariate models and 0.01 for all the tri-variate models.

5.2 Bivariate Copula Models

We consider the estimation of semiparametric bivariate copula models for four different copulas: the Clayton copula, the Gumbel copula, the Gaussian copula, a mixture of Gaussian and Clayton copulas. The Clayton copula of dimension m is of the form:

$$C_C(u_1, \ldots, u_m; \theta) = (\sum_{j=1}^m u_j^{-\theta} - (m-1))^{-1/\theta}, \ \theta \ge 0.$$

The Gumbel copula of dimension m is given by

$$C_G(u_1, \ldots, u_m; \theta) = \exp(-[\sum_{j=1}^m (-\ln u_j)^\theta]^{1/\theta}), \ \theta \ge 1.$$

The bivariate Gaussian copula is simply $\Phi_\theta(\Phi^{-1}(u_1), \Phi^{-1}(u_2))$, where $\Phi(\cdot)$ is the standard univariate normal distribution, $\Phi_\theta(u_1, u_2)$ is the bivariate normal distribution with means zero, variances 1, and correlation coefficient θ ($|\theta| \le 1$). The mixture of Gaussian and Clayton copulas we use is

$$C_M(u_1, u_2; \theta) = 0.9\Phi_\theta(\Phi^{-1}(u_1), \Phi^{-1}(u_2)) + 0.1 C_C(u_1, u_2; 0.5), \ |\theta| \le 1, \tag{20}$$

These four copulas are chosen to represent four different dependence structures. The bivariate Clayton copula exhibits positive dependence (i.e., its Kendall's τ is always positive), lower tail

[3]For the bivariate cases, we also used the penalty $\|(f_j^{1/2})''\|_2^2$ and found the results are similar for both penalties.

dependence but no upper tail dependence. As θ increases, both the overall dependence as measured by Kendall's τ and the lower tail dependence of the Clayton copula model increase; it approaches perfect positive dependence when $\theta \to \infty$ and independence when $\theta \to 0$. The bivariate Gumbel copula has positive dependence, upper tail dependence but no lower tail dependence. As θ increases, both the overall dependence and the upper tail dependence of the Gumbel copula increase; it approaches perfect positive dependence when $\theta \to \infty$ and reduces to the independence copula when $\theta = 1$. The bivariate Gaussian copula has symmetric positive and negative dependence but no tail dependence. As $|\theta|$ increases its overall dependence increases; it reduces to the independence copula when $\theta = 0$. The mixture of Gaussian and Clayton copulas (20) exhibits asymmetric positive and negative dependence; it has lower tail dependence because of the Clayton copula. As $|\theta|$ increases, its overall dependence as measured by Kendall's τ increases, but its tail dependence remains unchanged.

To sum up, the bivariate semiparametric copula models examined in this section can be expressed as $C(F_{o1}(x_1), F_{o2}(x_2); \theta_o)$, where the copula $C(u_1, u_2; \theta)$ is either the Clayton, Gumbel, Gaussian, or the mixture of Gaussian and Clayton copulas described above. The first marginal F_{o1} is either $U[0,1]$, $N(0,1)$ or $t_{[5]}$, and the second marginal F_{o2} is either $t_{[3]}$, $t_{[5]}$ or a mixture of normals $mn \equiv 0.4N(-1,1) + 0.6N(2,1)$. For the unknown marginal distribution F_{o2} we estimate its value at the 1/3 percentile (q_1) and 2/3 percentile (q_2) of the true distribution. For the unknown copula parameter θ_o we estimate its value corresponding to Kendall's τ from small to big. For each estimator of θ_o and F_{o2}, we compute its Monte Carlo sample mean (Mean), sample variance (Var), sample mean squared error (MSE), the Monte Carlo sample mean of the consistent estimator of its asymptotic variance ($\widehat{AVar}$), and the simulated theoretical asymptotic variance (AVar). The theoretical asymptotic variance of the two-step estimator and its consistent estimator are computed using the expressions in Genest, et al. (1995). The theoretical asymptotic variance of the modified two-step estimator and its consistent estimator are provided in Appendix B. The consistent estimators of the asymptotic variances for the sieve MLEs are computed according to those described in Sections 3 and 4, with 12 cosine series terms when Kendall's $|\tau| \leq 0.5$, and 24 cosine series terms when Kendall's $|\tau| > 0.5$. The simulated theoretical asymptotic variances of the sieve MLEs are computed in the same way as those for the estimators of asymptotic variances, except that we use the true parameter values $(\theta_o, F_{o1}, F_{o2})$ and a huge simulated sample ($n = 100,000$). Throughout this section, the reported values for the Var, MSE, $\widehat{AVar}$ and AVar for the marginal distribution are all multiplied by 10^3.

Through this simulation study, we aim to shed light on i) the relative performance of sieve MLE of the copula parameter to the two-step estimator; ii) the relative performance of the sieve MLE of the marginal cdfs to the empirical cdfs; iii) the performance of the consistent estimators of the asymptotic variances for the sieve MLEs of the copula parameter and the marginals; and iv) the accuracy of the simulated theoretical asymptotic variances of the sieve MLEs. For the lack of space, we present more detailed results for the Clayton copula and the Gaussian copula, but only brief results for the Gumbel copula and the mixture copula.

5.2.1 Bivariate copula models with one known or parametric margin

Tables 1, 2, 4, 5 and 6 present simulation results for different bivariate copula models when the first margin F_{o1} is completely known and the second margin F_{o2} is unknown. Without loss of generality $F_{o1} = U[0,1]$, and the true unknown marginal distribution F_{o2} is either $mn \equiv 0.4N(-1,1) + 0.6N(2,1)$ or $t_{[3]}$, where the former is asymmetric and bimodal and the latter has very fat tails. The modified two-step estimator of θ_o under a known $F_{o1}(x)$ maximizes $\sum_{i=1}^{n} \log\{c(F_{o1}(X_{1i}), \widetilde{F}_{n2}(X_{2i}), \theta)\}$.

Tables 1 and 2 report results for the estimation of the Clayton copula parameter θ_o and the unknown margin F_{o2}, where the true unknown F_{o2} is either mn or $t_{[3]}$. We have also tried F_{o2} being Beta, $t_{[5]}$ and $t_{[10]}$, and the results are virtually the same as the ones reported in Tables 1 and 2 for unknown $F_{o2} = t_{[3]}$. These results illustrate that the estimation of the copula parameter and the unknown margins is not sensitive to the particular functional forms of the unknown margins. Because of these findings and the lack of space, in Tables 3 - 6 we only report results corresponding to the unknown margin F_{o2} being mn.

Table 3 reports simulation results for the bivariate Clayton copula when the first margin F_{o1} is parametric and taken to be normal $N(0, \beta_o)$ with true but unknown $\beta_o = 1$.[4] The modified two-step estimator of θ_o in this case maximizes $\sum_{i=1}^{n} \log\{c(F_{o1}(X_{1i}; \widetilde{\beta}), \widetilde{F}_{n2}(X_{2i}), \theta)\}$ where $\widetilde{\beta}$ is the parametric MLE for the parametric margin $F_{o1}(x, \beta_o)$. As expected, the sample MSE, the estimated asymptotic variance and the simulated theoretical asymptotic variance of all the estimators of θ_o are slightly larger than the corresponding values in Table 1 due to the additional parameter in the parametric margin $F_{o1} = N(0, \beta_o)$ that we have to estimate. However, we are happy to find out that the difference is small, and all the qualitative patterns in Tables 1 and 2 for one known margin case carry over to this one parametric margin case.

For the bivariate Clayton copula model with a known margin F_{o1}, Bickel, et al. (1993) provide closed-form expressions for the semiparametric efficient variance bounds for estimating the copula parameter θ_o and the unknown margin F_{o2}. These are respectively the true asymptotic variances of our sieve MLE of θ_o and of F_{o2}, which are reported in Tables 1 and 2 as AvarT. For the bivariate Gaussian copula model with a known margin F_{o1}, Bickel, et al. (1993) only derive the closed-form expression for the semiparametric efficient variance bound for estimating the copula parameter θ_o but not the unknown margin F_{o2}. Their result allows us to report the true asymptotic variance of our sieve MLE of θ_o (AVarT) in Table 5. Results in Tables 1, 2 and 5 indicate that, regardless

[4]We have also tried parametric margin $F_{o1} = t_{[\beta_o]}$ with true but unknown $\beta_o = 5$ design, and the simulation results are very close to the ones reported in Table 3.

Table 1: Clayton copula – Known F_{o1}: Estimation of copula parameter

		$F_{o2} = mn$				$F_{o2} = t_{[3]}$			
		Sieve	Ideal	M-2step	2step	Sieve	Ideal	M-2step	2step
$\theta = 0.22$ ($\tau = 0.1$)	Mean	0.2197	0.2174	0.2222	0.2282	0.2158	0.2174	0.2184	0.2236
	Var	0.0049	0.0045	0.0049	0.0053	0.0048	0.0046	0.0048	0.0051
	MSE	0.0049	0.0045	0.0049	0.0053	0.0049	0.0046	0.0048	0.0051
	$\widehat{AVar}$	0.0047	0.0045	0.0049	0.0053	0.0047	0.0046	0.0049	0.0053
	AVar	0.0044	0.0043	0.0046	0.0049	0.0044	0.0043	0.0046	0.0049
	AVarT	0.0044				0.0044			
$\theta = 0.5$ ($\tau = 0.2$)	Mean	0.5018	0.4968	0.5019	0.5139	0.4959	0.4963	0.4955	0.5034
	Var	0.0072	0.0062	0.0069	0.0084	0.0071	0.0063	0.0068	0.0080
	MSE	0.0072	0.0062	0.0069	0.0086	0.0071	0.0063	0.0068	0.0080
	$\widehat{AVar}$	0.0068	0.0062	0.0073	0.0089	0.0068	0.0063	0.0072	0.0086
	AVar	0.0065	0.0061	0.0069	0.0091	0.0065	0.0061	0.0069	0.0091
	AVarT	0.0061				0.0061			
$\theta = 2$ ($\tau = 0.5$)	Mean	2.0134	1.9992	1.9483	2.0039	2.0163	1.9993	1.9491	1.9794
	Var	0.0214	0.0193	0.0212	0.0348	0.0215	0.0194	0.0213	0.0339
	MSE	0.0216	0.0193	0.0239	0.0348	0.0218	0.0194	0.0239	0.0344
	$\widehat{AVar}$	0.0216	0.0194	0.0261	0.0403	0.0217	0.0195	0.0258	0.0392
	AVar	0.0203	0.0191	0.0244	0.0402	0.0203	0.0191	0.0244	0.0402
	AVarT	0.0196				0.0196			
$\theta = 8$ ($\tau = 0.8$)	Mean	8.0780	8.0083	6.9319	7.8419	8.0911	8.0084	7.0131	7.7800
	Var	0.1596	0.1539	0.4204	0.3141	0.1575	0.1537	0.3739	0.3054
	MSE	0.1657	0.1539	1.5612	0.3391	0.1658	0.1537	1.3479	0.3538
	$\widehat{AVar}$	0.1615	0.1520	0.2550	0.4206	0.1622	0.1523	0.2682	0.4190
	AVar	0.1520	0.1504	0.2231	0.3565	0.1520	0.1504	0.2231	0.3565
	AVarT	0.1548				0.1548			
$\theta = 18$ ($\tau = 0.9$)	Mean	18.1766	18.0088	12.4746	17.2412	18.1792	18.0082	12.7451	17.1128
	Var	0.6974	0.6812	4.5407	1.5142	0.6985	0.6816	4.3782	1.4710
	MSE	0.7285	0.6813	35.0708	2.0900	0.7306	0.6817	31.9927	2.2581
	$\widehat{AVar}$	0.6825	0.6550	1.2229	2.5173	0.6828	0.6561	1.3745	2.5706
	AVar	0.6553	0.6433	1.3285	1.5661	0.6553	0.6433	1.3285	1.5661
	AVarT	0.6559				0.6559			

$F_{o1} = U[0, 1]$.

the dependence structure and the degree of dependence as measured by Kendall's τ, our simulated theoretical asymptotic variance (AVar) and our estimated asymptotic variance ($\widehat{AVar}$) are close to each other, and the simulated asymptotic variance (AVar) is very close to the true asymptotic variance (AVarT). Because of these findings, in all the tables we use simulated theoretical asymptotic variance (AVar) as a good approximate to the true theoretical asymptotic variance for general semiparametric copula models.

Tables 1, 3, 4, 5 and 6 reveal several patterns on estimation of copula parameter θ_o under known or parametric margin F_{o1}: (1) For small values of Kendall's τ, the sample MSE and the estimated asymptotic variance of the sieve MLE, the ideal estimator, the two-step estimator and the modified two-step estimator are all very close; but as Kendall's τ increases, the MSE and the estimated asymptotic variance of the sieve MLE are significantly smaller than those of the two-step estimator. (2) The asymptotic relative efficiency (ARE) of sieve MLE over the two-step estimator increases as Kendall's τ increases in absolute value. (3) For Clayton and Gumbel copulas, both Kendall's τ and

Table 2: Clayton copula – Known F_{o1}: Estimation of marginal distribution

		Sieve, $F_{o2}=mn$		Empir., $F_{o2}=mn$		Sieve, $F_{o2}=t_{[3]}$		Empir., $F_{o2}=t_{[3]}$	
		$F_{o2}(q_1)$	$F_{o2}(q_2)$	$F_{o2}(q_1)$	$F_{o2}(q_2)$	$F_{o2}(q_1)$	$F_{o2}(q_2)$	$F_{o2}(q_1)$	$F_{o2}(q_2)$
	Mean	0.3348	0.6703	0.3337	0.6664	0.3395	0.6668	0.3335	0.6668
	Var	0.5128	0.4865	0.5501	0.5417	0.4459	0.4262	0.5530	0.5445
$\theta=0.22$	MSE	0.5151	0.4998	0.5503	0.5418	0.4845	0.4262	0.5530	0.5445
$(\tau=0.1)$	$\widehat{AVar}$	0.5175	0.5167	0.5545	0.5544	0.5121	0.5313	0.5554	0.5552
	AVar	0.5299	0.5418			0.5299	0.5418		
	AVarT	0.5387	0.5484	0.5556	0.5556	0.5387	0.5484	0.5556	0.5556
	Mean	0.3345	0.6705	0.3334	0.6675	0.3393	0.6667	0.3331	0.6678
	Var	0.4575	0.4562	0.5460	0.5619	0.4065	0.3991	0.5489	0.5661
$\theta=0.5$	MSE	0.4589	0.4710	0.5460	0.5626	0.4419	0.3991	0.5489	0.5675
$(\tau=0.2)$	$\widehat{AVar}$	0.4798	0.4978	0.5542	0.5535	0.4710	0.5185	0.5551	0.5542
	AVar	0.4824	0.5214			0.4824	0.5214		
	AVarT	0.4947	0.5287	0.5556	0.5556	0.4947	0.5287	0.5556	0.5556
	Mean	0.3340	0.6702	0.3334	0.6679	0.3383	0.6659	0.3330	0.6683
	Var	0.2540	0.3494	0.6163	0.5691	0.1953	0.2825	0.6189	0.5719
$\theta=2$	MSE	0.2544	0.3617	0.6163	0.5707	0.2196	0.2830	0.6190	0.5745
$(\tau=0.5)$	$\widehat{AVar}$	0.2697	0.3864	0.5541	0.5531	0.2655	0.4017	0.5549	0.5539
	AVar	0.2733	0.3971			0.2733	0.3971		
	AVarT	0.2836	0.4038	0.5556	0.5556	0.2836	0.4038	0.5556	0.5556
	Mean	0.3333	0.6684	0.3329	0.6665	0.3364	0.6645	0.3326	0.6668
	Var	0.0721	0.1323	0.5930	0.5813	0.0320	0.0895	0.5940	0.5801
$\theta=8$	MSE	0.0721	0.1355	0.5932	0.5813	0.0412	0.0943	0.5946	0.5801
$(\tau=0.8)$	$\widehat{AVar}$	0.0782	0.1584	0.5537	0.5543	0.0798	0.1596	0.5545	0.5551
	AVar	0.0828	0.1665			0.0828	0.1665		
	AVarT	0.0863	0.1692	0.5556	0.5556	0.0863	0.1692	0.5556	0.5556
	Mean	0.3336	0.6677	0.3337	0.6665	0.3356	0.6647	0.3334	0.6669
	Var	0.0201	0.0464	0.5692	0.5724	0.0086	0.0287	0.5691	0.5699
$\theta=18$	MSE	0.0202	0.0476	0.5693	0.5724	0.0138	0.0325	0.5691	0.5700
$(\tau=0.9)$	$\widehat{AVar}$	0.0336	0.0830	0.5545	0.5542	0.0335	0.0637	0.5553	0.5551
	AVar	0.0327	0.0723			0.0327	0.0723		
	AVarT	0.0393	0.0787	0.5556	0.5556	0.0393	0.0787	0.5556	0.5556

$F_{o1}=U[0,1]$; $F_{o2}(q_1)=0.33$, $F_{o2}(q_2)=0.67$.
Reported Var, MSE, $\widehat{AVar}$, AVar, AvarT are the true values multiplied by 10^3.

the tail dependence increase as θ increases; for these copulas, the MSE, the estimated asymptotic variance and the theoretical asymptotic variance of all the estimators increase as θ increases, see Tables 1, 3 and 4. For Gaussian and the mixture copulas, Kendall's τ increases in absolute value as $|\theta|$ increases, but tail dependence does not change; for these copulas, the MSE, the estimated asymptotic variance and the theoretical asymptotic variance of all the estimators decrease as $|\theta|$ increases, see Tables 5 and 6. (4) the asymptotic variance of the modified two-step estimator is always smaller than that of the two-step estimator, confirming asymptotic efficiency gains of making use of one known or parametric margin information. In addition, for most values of Kendall's $|\tau|$ except large ones, the sample MSE of the modified two-step estimator is also smaller than that of the two-step estimator. But for strong dependence as measured by large $|\tau|$, the finite sample bias and MSE of the modified two-step estimator is larger than those of the two-step estimator. The modified two-step estimator has big MSE whenever there is very strong tail dependence that are

Table 3: Clayton copula – Parametric F_{o1}

		Copula parameter				Sieve		Empirical	
		Sieve	Ideal	M-2Step	2Step	$F_{o2}(q_1)$	$F_{o2}(q_2)$	$F_{o2}(q_1)$	$F_{o2}(q_2)$
$\theta = 0.22$ ($\tau = 0.1$)	Mean	0.2225	0.2190	0.2230	0.2281	0.3329	0.6677	0.3308	0.6637
	Var	0.0051	0.0046	0.0051	0.0056	0.5216	0.5109	0.5806	0.5855
	MSE	0.0051	0.0046	0.0051	0.0057	0.5217	0.5120	0.5870	0.5944
	$\widehat{AVar}$	0.0049	0.0045	0.0049	0.0053	0.5215	0.5366	0.5520	0.5566
	AVar	0.0046	0.0043	0.0046	0.0049	0.5316	0.5428	0.5556	0.5556
$\theta = 0.5$ ($\tau = 0.2$)	Mean	0.5041	0.5019	0.5051	0.5160	0.3342	0.6681	0.3324	0.6647
	Var	0.0073	0.0059	0.0074	0.0088	0.4829	0.5063	0.5499	0.6081
	MSE	0.0073	0.0059	0.0074	0.0090	0.4836	0.5084	0.5507	0.6122
	$\widehat{AVar}$	0.0074	0.0063	0.0074	0.0088	0.4875	0.5141	0.5534	0.5557
	AVar	0.0070	0.0061	0.0073	0.0091	0.4860	0.5229	0.5556	0.5556
$\theta = 2$ ($\tau = 0.5$)	Mean	2.0085	2.0050	1.9649	2.0068	0.3319	0.6673	0.3302	0.6634
	Var	0.0276	0.0182	0.0280	0.0341	0.2779	0.3737	0.5467	0.5809
	MSE	0.0277	0.0182	0.0292	0.0341	0.2800	0.3742	0.5568	0.5915
	$\widehat{AVar}$	0.0281	0.0196	0.0306	0.0404	0.2975	0.3986	0.5515	0.5568
	AVar	0.0269	0.0191	0.0288	0.0402	0.2984	0.4012	0.5556	0.5556
$\theta = 8$ ($\tau = 0.8$)	Mean	8.0075	8.0175	7.1069	7.8285	0.3322	0.6677	0.3309	0.6631
	Var	0.2033	0.1353	0.3524	0.3113	0.0837	0.1497	0.5756	0.5459
	MSE	0.2034	0.1356	1.1501	0.3407	0.0850	0.1507	0.5813	0.5590
	$\widehat{AVar}$	0.2135	0.1533	0.2735	0.4252	0.0871	0.1707	0.5521	0.5572
	AVar	0.2005	0.1504	0.2401	0.3565	0.0903	0.1739	0.5556	0.5556
$\theta = 18$ ($\tau = 0.9$)	Mean	17.8516	18.0442	13.1491	17.2123	0.3318	0.6688	0.3313	0.6636
	Var	0.9031	0.5816	3.5128	1.4752	0.0401	0.0615	0.5631	0.5586
	MSE	0.9251	0.5836	27.0442	2.0956	0.0423	0.0660	0.5674	0.5682
	$\widehat{AVar}$	0.8173	0.6625	1.4862	2.5965	0.0408	0.0885	0.5524	0.5567
	AVar	0.7410	0.6433	1.3384	1.5661	0.0371	0.0742	0.5556	0.5556

$F_{o1} = N(0, \beta)$ with unknown $\beta = 1$; $F_{o2} = mn, F_{o2}(q_1) = 0.33$, $F_{o2}(q_2) = 0.67$.
Reported Var, MSE, $\widehat{AVar}$, AVar in last 4 columns are the true values multiplied by 10^3.

exhibited in the Clayton, Gumbel, survival-Clayton[5] copulas with large θ.

Tables 2, 3, 4, 5 and 6 reveal several patterns on estimation of unknown margin F_{o2} given known or parametric margin F_{o1}: (1) As Kendall's $|\tau|$ increases, the performance of the sieve MLE improves greatly in terms of its sample variance, MSE and estimated asymptotic variance. This is expected because the sieve MLE of F_{o2} makes use of the dependence information between the two samples $\{X_{1i}\}$ and $\{X_{2i}\}$. (2) For very small values of $|\tau|$, the performance of the empirical distribution and that of sieve MLE are comparable, but for large values of $|\tau|$, sieve MLE is much more efficient than the empirical distribution. (3) For Clayton copula model with lower tail dependence, the lower quantile is generally better estimated than the upper quantile in the sense of having smaller theoretical asymptotic variance. For Gumbel copula model with upper tail dependence, the lower quantile is generally more difficult to estimate than the upper quantile. For Gaussian copula model with symmetric dependence and no tail dependence, the lower and upper quantile have about the same sample MSE, estimated asymptotic variance and theoretical asymptotic variances.

[5]Simulation results on survival-Clayton are similar to those for Clayton and Gumbel copulas hence are skipped.

Table 4: Gumbel copula – Known F_{o1}

		Copula parameter				Sieve		Empirical	
		Sieve	Ideal	M-2Step	2Step	$F_{o2}(q_1)$	$F_{o2}(q_2)$	$F_{o2}(q_1)$	$F_{o2}(q_2)$
$\theta = 1.11$ $(\tau = 0.1)$	Mean	1.1121	1.1114	1.1137	1.1163	0.3356	0.6686	0.3342	0.6652
	Var	0.0014	0.0013	0.0014	0.0015	0.5327	0.4454	0.5818	0.5245
	MSE	0.0014	0.0013	0.0014	0.0015	0.5377	0.4491	0.5825	0.5266
	$\widehat{AVar}$	0.0014	0.0014	0.0015	0.0015	0.5295	0.5214	0.5548	0.5555
	AVar	0.0014	0.0013	0.0014	0.0015	0.5383	0.5275	0.5556	0.5556
$\theta = 1.25$ $(\tau = 0.2)$	Mean	1.2529	1.2514	1.2539	1.2588	0.3359	0.6685	0.3350	0.6658
	Var	0.0021	0.0019	0.0021	0.0024	0.5125	0.4223	0.5970	0.5706
	MSE	0.0022	0.0019	0.0021	0.0025	0.5191	0.4258	0.5998	0.5714
	$\widehat{AVar}$	0.0023	0.0021	0.0023	0.0025	0.5100	0.4909	0.5555	0.5549
	AVar	0.0022	0.0021	0.0022	0.0025	0.5171	0.4929	0.5556	0.5556
$\theta = 2$ $(\tau = 0.5)$	Mean	2.0098	2.0022	1.9936	2.0129	0.3351	0.6679	0.3353	0.6669
	Var	0.0069	0.0062	0.0070	0.0100	0.3701	0.2496	0.5876	0.5676
	MSE	0.0070	0.0062	0.0071	0.0101	0.3733	0.2510	0.5916	0.5676
	$\widehat{AVar}$	0.0075	0.0067	0.0079	0.0103	0.3706	0.3195	0.5557	0.5539
	AVar	0.0070	0.0066	0.0078	0.0102	0.3694	0.3110	0.5556	0.5556
$\theta = 5$ $(\tau = 0.8)$	Mean	5.0421	5.0066	4.7379	4.9731	0.3329	0.6675	0.3335	0.6678
	Var	0.0449	0.0407	0.0645	0.0804	0.1214	0.0700	0.5365	0.5453
	MSE	0.0467	0.0408	0.1332	0.0811	0.1216	0.0708	0.5365	0.5465
	$\widehat{AVar}$	0.0471	0.0440	0.0566	0.0855	0.1491	0.1260	0.5543	0.5533
	AVar	0.0445	0.0435	0.0638	0.0916	0.1491	0.1100	0.5556	0.5556
$\theta = 10$ $(\tau = 0.9)$	Mean	10.1075	10.0141	8.3804	9.7689	0.3329	0.6672	0.3336	0.6673
	Var	0.1771	0.1636	0.6066	0.3265	0.0522	0.0222	0.5495	0.5402
	MSE	0.1887	0.1638	3.2295	0.3799	0.0524	0.0225	0.5495	0.5406
	$\widehat{AVar}$	0.1874	0.1768	0.2575	0.4441	0.0681	0.0415	0.5544	0.5537
	AVar	0.1762	0.1747	0.3252	0.4330	0.0689	0.0301	0.5556	0.5556

$F_{o1} = U[0,1]$; $F_{o2} = mn, F_{o2}(q_1) = 0.33$, $F_{o2}(q_2) = 0.67$.
Reported Var, MSE, $\widehat{AVar}$, AVar in last 4 columns are the true values multiplied by 10^3.

5.2.2 Bivariate copula models with two unknown margins

Tables 7, 8 and 9 report simulation results for Clayton, Gumbel and Gaussian copulas with unknown but equal margins $F_{o1} = F_{o2} = t_{[5]}$. The modified two-step estimator taking into account the equal margin information is defined as the maximizer of $\sum_{i=1}^{n} \log\{c(\widetilde{F}(X_{1i}), \widetilde{F}(X_{2i}), \theta)\}$, where $\widetilde{F}(x) = \{\widetilde{F}_{n1}(x) + \widetilde{F}_{n2}(x)\}/2$ (denoted as M-empirical). Table 10 reports results for Gaussian copula with unknown and unequal margins $F_{o1} = t_{[5]}$, $F_{o2} = mn$.

For Gaussian copula with unknown margins, Klaassen and Wellner (1997) obtain the closed-form expression of semiparametric efficient variance bound for estimation of Gaussian correlation coefficient θ_o, and point out that the two-step estimator achieves this efficient variance bound. This result allows us to compute theoretical asymptotic variance for the sieve MLE of θ_o as well, which is reported as AVarT in Tables 9 and 10. Our simulation results in Tables 9 and 10 indicate that both sieve MLE and two-step estimator of θ_o perform comparably regardless the value of Kendall's τ. Moreover, our simulated theoretical asymptotic variance (AVar) is again very close to the true theoretical asymptotic variance (AVarT). Tables 9 and 10 also reveal that, although sieve MLE of θ_o has no efficiency gain over the two-step estimator, but the sieve MLEs of marginals do have

Table 5: Gaussian copula – Known F_{o1}

		Copula parameter				Sieve		Empirical	
		Sieve	Ideal	M-2Step	2Step	$F_{o2}(q_1)$	$F_{o2}(q_2)$	$F_{o2}(q_1)$	$F_{o2}(q_2)$
$\theta = 0.9511$ ($\tau = 0.80$)	Mean	0.9515	0.9509	0.9478	0.9502	0.3334	0.6669	0.3324	0.6646
	Var	0.0111	0.0106	0.0173	0.0242	0.1101	0.0910	0.5887	0.5466
	MSE	0.0113	0.0106	0.0280	0.0250	0.1101	0.0910	0.5896	0.5510
	$\widehat{AVar}$	0.0132	0.0124	0.0214	0.0276	0.1317	0.1404	0.5533	0.5559
	AVar	0.0126	0.0120	0.0176	0.0242	0.1268	0.1389		
	AVarT	0.0125						0.5556	0.5556
$\theta = 0.7071$ ($\tau = 0.5$)	Mean	0.7071	0.7062	0.7061	0.7090	0.3327	0.6675	0.3315	0.6646
	Var	0.4189	0.3718	0.4276	0.5826	0.3206	0.3412	0.5715	0.6008
	MSE	0.4189	0.3726	0.4287	0.5861	0.3209	0.3419	0.5748	0.6051
	$\widehat{AVar}$	0.4976	0.4330	0.5179	0.6341	0.3505	0.3588	0.5526	0.5558
	AVar	0.4688	0.4178	0.4938	0.6621	0.3520	0.3579		
	AVarT	0.4688						0.5556	0.5556
$\theta = 0.1564$ ($\tau = 0.1$)	Mean	0.1545	0.1542	0.1553	0.1570	0.3324	0.6667	0.3307	0.6635
	Var	2.0710	2.0661	2.0869	2.1716	0.5400	0.5349	0.5885	0.5946
	MSE	2.0747	2.0709	2.0881	2.1720	0.5409	0.5349	0.5955	0.6045
	$\widehat{AVar}$	2.4865	2.3894	2.5063	2.5300	0.5282	0.5387	0.5519	0.5567
	AVar	2.3611	2.3272	2.3782	2.4185	0.5465	0.5389		
	AVarT	2.3500						0.5556	0.5556
$\theta = -0.1564$ ($\tau = -0.1$)	Mean	-0.1589	-0.1580	-0.1600	-0.1616	0.3323	0.6671	0.3309	0.6633
	Var	2.0314	1.9856	2.0505	2.1571	0.5313	0.5147	0.5372	0.5249
	MSE	2.0374	1.9882	2.0632	2.1839	0.5324	0.5148	0.5430	0.5362
	$\widehat{AVar}$	2.4691	2.3749	2.4908	2.5152	0.5281	0.5379	0.5522	0.5570
	AVar	2.3595	2.3283	2.3623	2.4109	0.5466	0.5438		
	AVarT	2.3500						0.5556	0.5556
$\theta = -0.7071$ ($\tau = -0.5$)	Mean	-0.7089	-0.7072	-0.7078	-0.7105	0.3330	0.6657	0.3319	0.6628
	Var	0.3916	0.3508	0.4115	0.5835	0.3532	0.3248	0.5223	0.5352
	MSE	0.3948	0.3508	0.4119	0.5952	0.3533	0.3258	0.5243	0.5505
	$\widehat{AVar}$	0.4911	0.4288	0.5118	0.6290	0.3482	0.3611	0.5530	0.5574
	AVar	0.4684	0.4182	0.4804	0.6619	0.3598	0.3546		
	AVarT	0.4688						0.5556	0.5556
$\theta = -0.9511$ ($\tau = -0.8$)	Mean	-0.9516	-0.9510	-0.9479	-0.9502	0.3331	0.6662	0.3316	0.6641
	Var	0.0108	0.0103	0.0179	0.0255	0.1179	0.0966	0.4936	0.5702
	MSE	0.0112	0.0103	0.0281	0.0262	0.1179	0.0969	0.4966	0.5766
	$\widehat{AVar}$	0.0131	0.0124	0.0213	0.0280	0.1321	0.1422	0.5529	0.5562
	AVar	0.0126	0.0120	0.0156	0.0255	0.1356	0.1209		
	AVarT	0.0125						0.5556	0.5556

$F_{o1} = U[0,1]$; $F_{o2} = mn$, $F_{o2}(q_1) = 0.33, F_{o2}(q_2) = 0.67$.
Reported Var, MSE, $\widehat{AVar}$, AVar, AVarT are the true values multiplied by 10^3.

efficiency gains over the empirical distributions.

Tables 7 and 8 reveal several patterns on estimation of copula parameter θ_o under unknown but equal margins: First, the patterns (1), (2) and (3) on estimation of copula parameter θ_o under one known or parametric margin carry over here, except that the efficiency gain for large $|\tau|$ is not as big as in the one marginal known or parametric case. This is as expected, because intuitively the known or parametric marginal case represents stronger prior information than the equal, but unknown marginals case. Second, the asymptotic variance of the modified two-step estimator for θ_o is the same as that of the two-step estimator for θ_o, which means that the two estimators are asymptotically equivalent. This is also consistent with the simulation results that sample MSE,

Table 6: $0.9Gaussian(\theta) + 0.1Clayton(0.5)$ copula – Known F_{o1}

		Copula parameter				Sieve		Empirical	
		Sieve	Ideal	M-2Step	2Step	$F_{o2}(q_1)$	$F_{o2}(q_2)$	$F_{o2}(q_1)$	$F_{o2}(q_2)$
$\theta = 0.95$ $(\tau = 0.72)$	Mean	0.9507	0.9500	0.9468	0.9492	0.3330	0.6669	0.3314	0.6640
	Var	0.0192	0.0187	0.0290	0.0325	0.1354	0.1067	0.5585	0.5623
	MSE	0.0197	0.0187	0.0394	0.0331	0.1355	0.1068	0.5622	0.5696
	$\widehat{AVar}$	0.0208	0.0200	0.0287	0.0333	0.1509	0.1618	0.5526	0.5564
	AVar	0.0205	0.0197	0.0235	0.0311	0.1452	0.1608	0.5556	0.5556
$\theta = 0.90$ $(\tau = 0.67)$	Mean	0.9008	0.8998	0.8972	0.8998	0.3328	0.6669	0.3313	0.6633
	Var	0.0850	0.0799	0.0998	0.1269	0.2146	0.1739	0.5770	0.5696
	MSE	0.0856	0.0800	0.1077	0.1269	0.2149	0.1740	0.5812	0.5812
	$\widehat{AVar}$	0.0872	0.0816	0.1026	0.1250	0.2120	0.2204	0.5524	0.5569
	AVar	0.0855	0.0811	0.0945	0.1224	0.2070	0.2328	0.5556	0.5556
$\theta = 0.42$ $(\tau = 0.28)$	Mean	0.4184	0.4177	0.4197	0.4231	0.3325	0.6661	0.3310	0.6624
	Var	1.8876	1.7746	1.8789	2.1290	0.4756	0.4566	0.6132	0.5936
	MSE	1.8903	1.7799	1.8790	2.1388	0.4763	0.4570	0.6185	0.6116
	$\widehat{AVar}$	2.0847	1.9122	2.0798	2.2266	0.4687	0.4719	0.5521	0.5576
	AVar	1.9925	1.8797	2.0431	2.2318	0.4957	0.4960	0.5556	0.5556
$\theta = 0.155$ $(\tau = 0.11)$	Mean	0.1526	0.1524	0.1535	0.1553	0.3319	0.6660	0.3301	0.6627
	Var	2.7043	2.7098	2.7415	2.8631	0.5626	0.5049	0.6106	0.5718
	MSE	2.7099	2.7164	2.7437	2.8631	0.5645	0.5053	0.6211	0.5879
	$\widehat{AVar}$	3.0627	2.9442	3.0873	3.1269	0.5251	0.5384	0.5513	0.5574
	AVar	3.0121	2.9856	3.0260	3.0710	0.5258	0.5274	0.5556	0.5556
$\theta = -0.25$ $(\tau = -0.11)$	Mean	-0.2526	-0.2514	-0.2542	-0.2568	0.3316	0.6662	0.3301	0.6625
	Var	2.3795	2.3154	2.4230	2.5873	0.5223	0.4894	0.5409	0.5431
	MSE	2.3861	2.3174	2.4404	2.6333	0.5253	0.4896	0.5512	0.5603
	$\widehat{AVar}$	2.9226	2.7943	2.9333	2.9786	0.5128	0.5126	0.5515	0.5576
	AVar	2.8326	2.7606	2.8208	2.8821	0.5098	0.5133	0.5556	0.5556
$\theta = -0.50$ $(\tau = -0.28)$	Mean	-0.5024	-0.5007	-0.5034	-0.5071	0.3324	0.6659	0.3311	0.6627
	Var	1.4143	1.3295	1.4116	1.6505	0.4698	0.4376	0.5374	0.5365
	MSE	1.4199	1.3299	1.4229	1.7011	0.4706	0.4383	0.5423	0.5523
	$\widehat{AVar}$	1.8148	1.6700	1.8093	1.9356	0.4585	0.4528	0.5523	0.5575
	AVar	1.7924	1.6982	1.7571	1.9312	0.4763	0.4721	0.5556	0.5556
$\theta = -0.95$ $(\tau = -0.67)$	Mean	-0.9508	-0.9500	-0.9465	-0.9490	0.3330	0.6664	0.3309	0.6638
	Var	0.0183	0.0172	0.0283	0.0329	0.1452	0.1114	0.4975	0.5654
	MSE	0.0190	0.0172	0.0405	0.0339	0.1453	0.1114	0.5034	0.5734
	$\widehat{AVar}$	0.0196	0.0189	0.0292	0.0347	0.1460	0.1528	0.5523	0.5565
	AVar	0.0185	0.0179	0.0221	0.0316	0.1474	0.1605	0.5556	0.5556

Known margin is $F_{o1} = U[0,1]$; $F_{o2} = mn$, $F_{o2}(q_1) = 0.33$, $F_{o2}(q_2) = 0.67$.
Reported Var, MSE, $\widehat{AVar}$, AVar are the true values multiplied by 10^3.

estimated asymptotic variance of the two estimators are close to each other for all the different copulas and all different value of τ.

Tables 7, 8 and 9 reveal several patterns on estimation of unknown but equal margins $F_{o2} = F_{o1}$: (1) As Kendall's $\tau > 0$ increases, both the sieve MLE and the modified empirical become worse in terms of the Var, MSE, $\widehat{AVar}$ and AVar, but the modified empirical distribution deteriorates faster than the sieve MLE. However, for Kendall's $\tau < 0$, as $|\tau|$ increases, both the sieve MLE and the modified empirical become much better in terms of the Var, MSE, $\widehat{AVar}$ and AVar; see Table 9.[6] Intuitively, both sieve MLE and the modified empirical distribution make use of the

[6] We also find similar pattern for the mixture of Gaussian and Clayton copulas model with unknown but equal

Table 7: Clayton copula – Margins are unknown but equal

		Copula parameter				Sieve		M-Empirical	
		Sieve	Ideal	M-2Step	2Step	$F_{o2}(q_1)$	$F_{o2}(q_2)$	$F_{o2}(q_1)$	$F_{o2}(q_2)$
$\theta = 0.22$ $(\tau = 0.1)$	Mean	0.2132	0.2197	0.2169	0.2240	0.3367	0.6651	0.3315	0.6645
	Var	0.0049	0.0045	0.0049	0.0052	0.2386	0.2359	0.2957	0.2942
	MSE	0.0049	0.0045	0.0049	0.0052	0.2496	0.2384	0.2992	0.2989
	$\widehat{AVar}$	0.0048	0.0045	0.0052	0.0052	0.2800	0.2778	0.3090	0.2969
	AVar	0.0045	0.0043	0.0049	0.0049	0.2712	0.2860	0.3110	0.2965
$\theta = 0.5$ $(\tau = 0.2)$	Mean	0.4929	0.5028	0.4941	0.5071	0.3370	0.6659	0.3324	0.6652
	Var	0.0078	0.0058	0.0078	0.0080	0.3016	0.2729	0.3688	0.3391
	MSE	0.0079	0.0058	0.0078	0.0080	0.3153	0.2736	0.3698	0.3413
	$\widehat{AVar}$	0.0074	0.0063	0.0086	0.0087	0.3347	0.3244	0.3445	0.3184
	AVar	0.0070	0.0061	0.0091	0.0091	0.3305	0.3118	0.3448	0.3172
$\theta = 2$ $(\tau = 0.5)$	Mean	1.9937	2.0113	1.9518	1.9917	0.3368	0.6660	0.3327	0.6657
	Var	0.0327	0.0187	0.0324	0.0342	0.3799	0.3399	0.4897	0.4168
	MSE	0.0328	0.0189	0.0347	0.0343	0.3921	0.3403	0.4902	0.4177
	$\widehat{AVar}$	0.0286	0.0195	0.0392	0.0396	0.3938	0.3791	0.4417	0.3914
	AVar	0.0271	0.0191	0.0402	0.0402	0.3910	0.3639	0.4421	0.3904
$\theta = 8$ $(\tau = 0.8)$	Mean	7.9258	8.0175	7.4861	7.7532	0.3342	0.6647	0.3306	0.6635
	Var	0.2941	0.1353	0.2686	0.3104	0.3725	0.3352	0.5300	0.4721
	MSE	0.2996	0.1356	0.5328	0.3713	0.3733	0.3390	0.5375	0.4822
	$\widehat{AVar}$	0.2319	0.1533	0.4066	0.4232	0.4055	0.3787	0.5189	0.4913
	AVar	0.2206	0.1504	0.3565	0.3565	0.4118	0.3704	0.5210	0.4883

$F_{o1} = F_{o2} = t_{[5]}$; $F_{o2}(q_1) = 0.33$, $F_{o2}(q_2) = 0.67$.
Reported Var, MSE, $\widehat{AVar}$, AVar in last 4 columns are the true values multiplied by 10^3.

Table 8: Gumbel copula – Margins are unknown but equal

		Copula parameter				Sieve		M-Empirical	
		Sieve	Ideal	M-2Step	2Step	$F_{o2}(q_1)$	$F_{o2}(q_2)$	$F_{o2}(q_1)$	$F_{o2}(q_2)$
$\theta = 1.11$ $(\tau = 0.1)$	Mean	1.1083	1.1119	1.1145	1.1141	0.3379	0.6666	0.3329	0.6654
	Var	0.0014	0.0014	0.0015	0.0015	0.2588	0.2394	0.3143	0.2971
	MSE	0.0014	0.0014	0.0015	0.0015	0.2792	0.2394	0.3145	0.2988
	$\widehat{AVar}$	0.0014	0.0014	0.0015	0.0015	0.2981	0.3188	0.3967	0.3543
	AVar	0.0014	0.0013	0.0015	0.0015	0.2931	0.2994	0.3966	0.3529
$\theta = 1.25$ $(\tau = 0.2)$	Mean	1.2460	1.2521	1.2551	1.2547	0.3378	0.6668	0.3319	0.6657
	Var	0.0024	0.0021	0.0025	0.0025	0.2630	0.2541	0.3264	0.3090
	MSE	0.0024	0.0021	0.0025	0.0025	0.2833	0.2541	0.3285	0.3099
	$\widehat{AVar}$	0.0024	0.0021	0.0025	0.0025	0.3196	0.3462	0.4052	0.3609
	AVar	0.0024	0.0021	0.0025	0.0025	0.3219	0.3383	0.4056	0.3598
$\theta = 2$ $(\tau = 0.5)$	Mean	1.9871	2.0044	1.9999	2.0018	0.3378	0.6674	0.3315	0.6654
	Var	0.0095	0.0068	0.0098	0.0100	0.3071	0.2915	0.3672	0.3744
	MSE	0.0097	0.0069	0.0098	0.0100	0.3273	0.2921	0.3706	0.3759
	$\widehat{AVar}$	0.0096	0.0067	0.0101	0.0100	0.3688	0.3609	0.4413	0.3917
	AVar	0.0094	0.0066	0.0102	0.0102	0.3823	0.3545	0.4421	0.3904
$\theta = 5$ $(\tau = 0.8)$	Mean	4.9468	5.0119	4.8934	4.9411	0.3378	0.6665	0.3311	0.6642
	Var	0.0758	0.0443	0.0721	0.0775	0.3505	0.3216	0.4373	0.4233
	MSE	0.0786	0.0445	0.0835	0.0809	0.3687	0.3225	0.4424	0.4296
	$\widehat{AVar}$	0.0750	0.0441	0.0833	0.0828	0.4324	0.4352	0.5004	0.4602
	AVar	0.0726	0.0435	0.0916	0.0916	0.4391	0.4467	0.5018	0.4576

$F_{o1} = F_{o2} = t_{[5]}$; $F_{o2}(q_1) = 0.33$, $F_{o2}(q_2) = 0.67$.
Reported Var, MSE, $\widehat{AVar}$, AVar in last 4 columns are the true values multiplied by 10^3.

margins, but we do not report them due to the lack of space.

Table 9: Gaussian copula – Margins are unknown but equal

		Copula parameter				Sieve		M-Empirical	
		Sieve	Ideal	M-2Step	2Step	$F_{o2}(q_1)$	$F_{o2}(q_2)$	$F_{o2}(q_1)$	$F_{o2}(q_2)$
	Mean	0.9506	0.9510	0.9496	0.9502	0.3365	0.6650	0.3311	0.6647
	Var	0.0233	0.0115	0.0232	0.0232	0.4049	0.4106	0.5016	0.5599
$\theta = 0.9511$	MSE	0.0235	0.0115	0.0254	0.0240	0.4152	0.4133	0.5064	0.5640
$(\tau = 0.80)$	$\widehat{AVar}$	0.0240	0.0124	0.0277	0.0320	0.4066	0.3764	0.4961	0.4991
	AVar	0.0222	0.0120	0.0242	0.0242	0.4172	0.4168		
	AVarT	0.0228		0.0228	0.0228			0.4938	0.4998
	Mean	0.7017	0.7070	0.7047	0.7059	0.3373	0.6675	0.3335	0.6665
	Var	0.6976	0.4284	0.6729	0.6713	0.3639	0.3331	0.4110	0.3987
$\theta = 0.7071$	MSE	0.7273	0.4284	0.6789	0.6727	0.3792	0.3337	0.4110	0.3987
$(\tau = 0.50)$	$\widehat{AVar}$	0.6710	0.4281	0.6401	0.6640	0.3878	0.4005	0.4144	0.4145
	AVar	0.6136	0.4178	0.6621	0.6621	0.3915	0.4033		
	AVarT	0.6250		0.6250	0.6250			0.4154	0.4126
	Mean	0.1544	0.1564	0.1573	0.1587	0.3375	0.6660	0.3334	0.6659
	Var	2.5872	2.5490	2.6722	2.6616	0.2446	0.2449	0.2929	0.3198
$\theta = 0.1564$	MSE	2.5913	2.5490	2.6729	2.6667	0.2622	0.2453	0.2929	0.3204
$(\tau = 0.10)$	$\widehat{AVar}$	2.4966	2.3763	2.5099	2.5219	0.2952	0.3001	0.3038	0.3044
	AVar	2.3895	2.3272	2.4185	2.4185	0.2929	0.2871		
	AVarT	2.3791		2.3791	2.3791			0.3032	0.3057
	Mean	-0.1570	-0.1562	-0.1600	-0.1588	0.3369	0.6659	0.3327	0.6659
	Var	2.5645	2.5023	2.6675	2.6672	0.1882	0.1933	0.2453	0.2527
$\theta = -0.1564$	MSE	2.5648	2.5024	2.6801	2.6728	0.2007	0.1939	0.2457	0.2532
$(\tau = -0.10)$	$\widehat{AVar}$	2.4855	2.3743	2.5118	2.5143	0.2467	0.2534	0.2519	0.2526
	AVar	2.3870	2.3283	2.4109	2.4109	0.2431	0.2468		
	AVarT	2.3791		2.3791	2.3791			0.2514	0.2541
	Mean	-0.7031	-0.7070	-0.7067	-0.7057	0.3355	0.6659	0.3321	0.6662
	Var	0.6759	0.4261	0.6569	0.6601	0.1048	0.0988	0.1637	0.1647
$\theta = -0.7071$	MSE	0.6921	0.4262	0.6571	0.6621	0.1096	0.0994	0.1653	0.1649
$(\tau = -0.50)$	$\widehat{AVar}$	0.6570	0.4271	0.6424	0.6572	0.1500	0.1608	0.1655	0.1655
	AVar	0.6142	0.4182	0.6619	0.6619	0.1539	0.1434		
	AVarT	0.6250		0.6250	0.6250			0.1656	0.1645
	Mean	-0.9514	-0.9511	-0.9510	-0.9502	0.3360	0.6640	0.3316	0.6653
	Var	0.0242	0.0120	0.0245	0.0261	0.0563	0.0563	0.1442	0.1479
$\theta = -0.9511$	MSE	0.0243	0.0120	0.0245	0.0268	0.0636	0.0633	0.1471	0.1496
$(\tau = -0.80)$	$\widehat{AVar}$	0.0227	0.0124	0.0283	0.0297	0.0900	0.0951	0.1395	0.1382
	AVar	0.0227	0.0120	0.0241	0.0241	0.0914	0.0834		
	AVarT	0.0228		0.0228	0.0228			0.1390	0.1420

$F_{o1} = F_{o2} = t_{[5]}$; $F_{o2}(q_1) = 0.33$, $F_{o2}(q_2) = 0.67$.
Reported Var, MSE, $\widehat{AVar}$, AVar, AVarT are the true values multiplied by 10^3.

additional information in the observations from one distribution, say F_{o2} to estimate the other distribution F_{o1}. As the positive dependence between the two variables increases, there is less and less information to add to the observations from F_{o1}, but as negative dependence increases, there is more and more information to add. (2) For all values of $|\tau|$, both the sieve MLE and the modified empirical distribution perform much better than the empirical distribution.

5.3 Tri-variate Copula Models

We have also conducted Monte Carlo experiments for a number of other copulas and marginal distributions in three dimensions. To save space, we only report a few results for two tri-variate

Table 10: Gaussian copula – Margins are unknown and unequal

		Copula parameter			Sieve		Empirical	
		Sieve	Ideal	2Step	$F_{o2}(q_1)$	$F_{o2}(q_2)$	$F_{o2}(q_1)$	$F_{o2}(q_2)$
$\theta = 0.9511$ ($\tau = 0.80$)	Mean	0.9507	0.9511	0.9496	0.3352	0.6696	0.3327	0.6662
	Var	0.0239	0.0128	0.0250	0.4501	0.4847	0.5593	0.6246
	MSE	0.0240	0.0128	0.0270	0.4536	0.4930	0.5597	0.6248
	$\widehat{AVar}$	0.0240	0.0122	0.0282	0.4791	0.4867	0.5550	0.5558
	AVar	0.0235	0.0120	0.0242	0.4844	0.4610		
	AVarT	0.0228		0.0228			0.5556	0.5556
$\theta = 0.7071$ ($\tau = 0.50$)	Mean	0.7063	0.7082	0.7085	0.3342	0.6694	0.3327	0.6661
	Var	0.6541	0.4458	0.6255	0.5528	0.5265	0.6112	0.5982
	MSE	0.6547	0.4471	0.6274	0.5537	0.5339	0.6115	0.5985
	$\widehat{AVar}$	0.6638	0.4233	0.6329	0.5315	0.5392	0.5549	0.5559
	AVar	0.6443	0.4178	0.6621	0.5046	0.5244		
	AVarT	0.6250		0.6250			0.5556	0.5556
$\theta = 0.1564$ ($\tau = 0.10$)	Mean	0.1597	0.1609	0.1623	0.3344	0.6687	0.3327	0.6659
	Var	2.5043	2.3994	2.5726	0.5513	0.5293	0.5713	0.5851
	MSE	2.5148	2.4193	2.6073	0.5523	0.5335	0.5717	0.5857
	$\widehat{AVar}$	2.5545	2.3524	2.5056	0.5557	0.5614	0.5550	0.5561
	AVar	2.4193	2.3272	2.4185	0.5404	0.5488		
	AVarT	2.3791		2.3791			0.5556	0.5556
$\theta = -0.1564$ ($\tau = -0.10$)	Mean	-0.1505	-0.1521	-0.1533	0.3337	0.6687	0.3326	0.6661
	Var	2.4909	2.3676	2.5554	0.5400	0.5286	0.5575	0.5760
	MSE	2.5262	2.3865	2.5653	0.5401	0.5327	0.5580	0.5763
	$\widehat{AVar}$	2.5703	2.3778	2.5205	0.5563	0.5637	0.5549	0.5560
	AVar	2.4163	2.3283	2.4109	0.5287	0.5364		
	AVarT	2.3791		2.3791			0.5556	0.5556
$\theta = -0.7071$ ($\tau = -0.50$)	Mean	-0.7028	-0.7061	-0.7051	0.3337	0.6686	0.3332	0.6660
	Var	0.6593	0.4217	0.6340	0.4952	0.5536	0.5251	0.6189
	MSE	0.6778	0.4227	0.6381	0.4953	0.5575	0.5251	0.6193
	$\widehat{AVar}$	0.6760	0.4300	0.6432	0.5339	0.5318	0.5555	0.5560
	AVar	0.6419	0.4182	0.6619	0.4987	0.5081		
	AVarT	0.6250		0.6250			0.5556	0.5556
$\theta = -0.9511$ ($\tau = -0.80$)	Mean	-0.9505	-0.9510	-0.9493	0.3330	0.6679	0.3322	0.6654
	Var	0.0245	0.0124	0.0236	0.4687	0.4669	0.5713	0.5550
	MSE	0.0248	0.0124	0.0265	0.4688	0.4683	0.5725	0.5566
	$\widehat{AVar}$	0.0244	0.0123	0.0284	0.4822	0.4839	0.5546	0.5566
	AVar	0.0236	0.0120	0.0241	0.4883	0.4785		
	AVarT	0.0228		0.0228			0.5556	0.5556

$F_{o1} = t_{[5]}, F_{o2} = mn$; $F_{o2}(q_1) = 0.33$, $F_{o2}(q_2) = 0.67$.
Reported Var, MSE, $\widehat{AVar}$, AVar, AVarT are the true values multiplied by 10^3.

copula models: a mixture of Clayton and Gumbel copulas and the Student's t copula, while all the marginal distributions are assumed to be equal but otherwise unspecified.

The mixture of Clayton and Gumbel copulas is defined as

$$C(u_1, u_2, u_3; \theta) = \lambda C_C(u_1, u_2, u_3; \theta_1) + (1 - \lambda) C_G(u_1, u_2, u_3; \theta_2),$$

where the copula parameter is $\theta = (\theta_1, \theta_2, \lambda)$, $\theta_1 \geq 0$, $\theta_2 \geq 1$, $0 < \lambda < 1$. This mixture copula has positive pairwise dependence and asymmetric tail dependence. The tri-variate Student's t-copula is

$$C_t(u_1,u_2,u_3;\theta)=\int_{-\infty}^{u_1}\int_{-\infty}^{u_2}\int_{-\infty}^{u_3}\frac{\Gamma(\frac{\nu+3}{2})}{\Gamma(\frac{\nu}{2})\sqrt{(\pi\nu)^3|\Sigma|}}\left(1+\frac{(T_\nu^{-1}(s))'\Sigma^{-1}T_\nu^{-1}(s)}{\nu}\right)^{-\frac{\nu+3}{2}}ds_1ds_2ds_3,$$

where the copula parameter is $\theta=(\rho_{12},\rho_{13},\rho_{23},\nu)$, Σ is the correlation matrix with upper diagonal elements $\rho_{12},\rho_{13},\rho_{23}\in(-1,1)$ and ν is the degree of freedom. The Student's t-copula has positive and negative pairwise dependence and symmetric tail dependence.

Tables 11 and 12 report results for these two tri-variate copula models, where the true unknown margins are $F_{o1}=F_{o2}=F_{o3}=t_{[10]}$, the true unknown mixture copula parameter is $\theta_o=(15,10,0.25)$, and the true unknown Student's t copula parameter is $\theta_o=(-0.9,0.1,-0.5,5)$. The relative performance of different estimators is qualitatively the same as what we have observed in the bivariate case. Since the multivariate Student's t copula model has been widely used in financial risk management, it is very encouraging to know that our sieve ML estimates of its copula parameter and marginal distributions are not only asymptotically efficient but also perform well in finite samples.

Table 11: Trivariate copulas: Estimation of copula parameters

	Sieve	Ideal	2step
	Mixture of Clayton and Gumbel: $\theta=(\theta_1,\theta_2,\lambda)=(15,10,0.25)$		
Mean	(15.1411, 9.8246, 0.2566)	(15.0278,10.0164, 0.2495)	(15.4347, 9.3144, 0.2751)
Var	(3.0018, 0.3148, 0.0024)	(2.1472, 0.1561, 0.0014)	(3.0275, 0.2954, 0.0028)
MSE	(3.0217, 0.3455, 0.0025)	(2.1479, 0.1564, 0.0014)	(3.2165, 0.7654, 0.0034)
	Student's t copula: $\theta=(\rho_{12},\rho_{13},\rho_{23},\nu)=(-0.9,0.1,-0.5,5)$		
Mean	(-0.8996, 0.1008, -0.5008, 5.6443)	(-0.8999, 0.1013, -0.5014, 5.0572)	(-0.8927, 0.0944, -0.5015, 5.4235)
Var	(0.0185, 0.4884, 0.2890, 1.2455)	(0.0152, 0.4410, 0.2372, 0.1941)	(0.1369, 3.5522, 1.8233, 1.6428)
MSE	(0.0187, 0.4891, 0.2896, 1.6615)	(0.0152, 0.4427, 0.2391, 0.1974)	(0.1902, 3.5835, 1.8255, 1.8222)

$F_{o1}=F_{o2}=F_{o3}=t_{[10]}$.
Reported Var and MSE of $\{\rho_{ij}\}$ are the true values multiplied by 10^3.

Table 12: Trivariate copulas: Estimation of marginal distributions

	Empirical						M-Empirical		Sieve	
	$F_{o1}(q_1)$	$F_{o2}(q_1)$	$F_{o3}(q_1)$	$F_{o1}(q_2)$	$F_{o2}(q_2)$	$F_{o3}(q_2)$	$F_{o1}(q_1)$	$F_{o1}(q_2)$	$F_{o1}(q_1)$	$F_{o1}(q_2)$
	Mixture of Clayton and Gumbel: $\theta=(\theta_1,\theta_2,\lambda)=(15,10,0.25)$									
Mean	0.2497	0.2502	0.2506	0.7565	0.7564	0.7568	0.2498	0.7570	0.2501	0.7569
Var	0.5187	0.5045	0.5363	0.4196	0.4454	0.4223	0.4893	0.3948	0.3856	0.2749
MSE	0.5188	0.5045	0.5367	0.4618	0.4864	0.4685	0.4893	0.4438	0.3856	0.3225
	Student's t copula: $\theta=(\rho_{12},\rho_{13},\rho_{23},\nu)=(-0.9,0.1,-0.5,5)$									
Mean	0.2526	0.2435	0.2493	0.7564	0.7496	0.7512	0.2480	0.7528	0.2509	0.7550
Var	0.5045	0.4108	0.5019	0.4603	0.4927	0.5009	0.1169	0.1094	0.0618	0.0387
MSE	0.5113	0.4531	0.5024	0.5013	0.4929	0.5023	0.1209	0.1172	0.0626	0.0637

$F_{o1}=F_{o2}=F_{o3}=t_{[10]}$; $F_{o1}(q_1)=0.25$, and $F_{o1}(q_2)=0.75$.
Reported Var and MSE are the true values multiplied by 10^3.

Appendix A. Mathematical Proofs

Assumption 5. there exist constants $\epsilon_1>0,\epsilon_2>0$ with $2\epsilon_1+\epsilon_2<1$ such that $(\delta_n)^{3-(2\epsilon_1+\epsilon_2)}=o(n^{-1})$, and the followings (1)-(4) hold for all $\widetilde{\alpha}\in\mathcal{A}_n$ with $||\widetilde{\alpha}-\alpha_o||\leq\delta_n$ and all $v=(v_\theta,v_1,...,v_m)'\in\mathbf{V}$ with $||v||\leq\delta_n$:

(1) $\left|E_o\left(\frac{\partial^2 \log c(\widetilde{\alpha})}{\partial\theta\partial\theta'} - \frac{\partial^2 \log c(\alpha_o)}{\partial\theta\partial\theta'}\right)\right| \leq c||\widetilde{\alpha} - \alpha_o||^{1-\epsilon_2}$;
(2) $\left|E_o\left(\left\{\frac{\partial^2 \log c(\widetilde{\alpha})}{\partial\theta\partial u_j} - \frac{\partial^2 \log c(\alpha_o)}{\partial\theta\partial u_j}\right\}\int^{X_j} v_j(x)dx\right)\right| \leq c||v||^{1-\epsilon_1}||\widetilde{\alpha} - \alpha_o||^{1-\epsilon_2}$ for all $j = 1, ..., m$;
(3) $\left|E_o\left(\left\{\frac{\partial^2 \log c(\widetilde{\alpha})}{\partial u_i\partial u_j} - \frac{\partial^2 \log c(\alpha_o)}{\partial u_i\partial u_j}\right\}\int^{X_j} v_j(x)dx\int^{X_i} v_i(x)dx\right)\right| \leq c||v||^{2(1-\epsilon_1)}||\widetilde{\alpha}-\alpha_o||^{1-\epsilon_2}$ for all $j, i = 1, ..., m$;
(4) $\left|E_o\left(\left[\frac{v_j(X_j)}{\widetilde{f}_j(X_j)}\right]^2 - \left[\frac{v_j(X_j)}{f_{oj}(X_j)}\right]^2\right)\right| \leq c||v||^{2(1-\epsilon_1)}||\widetilde{\alpha} - \alpha_o||^{1-\epsilon_2}$ for all $j = 1, ..., m$.

In the following we denote $\mu_n(g) = \frac{1}{n}\sum_{i=1}^n[g(Z_i) - E_o(g(Z_i))]$ as the empirical process indexed by g.
Assumption 6. **(1)**

$$\sup_{\alpha\in\mathcal{A}_n:||\alpha-\alpha_o||=O(\delta_n)} \mu_n\left(\frac{\partial \log c(\alpha)}{\partial\theta'} - \frac{\partial \log c(\alpha_o)}{\partial\theta'}\right) = o_P(n^{-1/2});$$

(2) for all $j = 1, ..., m$,

$$\sup_{\alpha\in\mathcal{A}_n:||\alpha-\alpha_o||=O(\delta_n)} \mu_n\left(\left\{\frac{\partial \log c(\alpha)}{\partial u_j} - \frac{\partial \log c(\alpha_o)}{\partial u_j}\right\}\int 1(x \leq X_j)\Pi_n\upsilon_j^*(x)dx\right) = o_P(n^{-1/2});$$

and **(3)**

$$\sup_{\alpha\in\mathcal{A}_n:||\alpha-\alpha_o||=O(\delta_n)} \mu_n\left(\left\{\frac{1}{f_j(X_j)} - \frac{1}{f_{oj}(X_j)}\right\}\Pi_n\upsilon_j^*(X_j)\right) = o_P(n^{-1/2}).$$

Assumptions 5 and 6 are sufficient conditions to control the second order term in the expansion of the sample log-likelihood criterion function. They are easily satisfied when copula density is twice continuously differentiable around true α_o and the unknown marginal densities are in some smooth function classes (such as Sobolev, Besov, Hölder classes) and are bounded away from zero. When unknown marginal densities are smooth but approach zero at the tails, one might have to do some trimming or weighting to take care of the tails; see e.g. Wong and Shen (1995)

Proof. (**Theorem 1**): Let ε_n be any positive sequence satisfying $\varepsilon_n = o(\frac{1}{\sqrt{n}})$ and $(\delta_n)^{3-\epsilon} = \varepsilon_n \times o(n^{-1/2})$, [for instance we can take $\varepsilon_n = \frac{1}{\sqrt{n}\log n}$]. Also define $r[\alpha, \alpha_o, Z_i] \equiv \ell(\alpha, Z_i) - \ell(\alpha_o, Z_i) - \frac{\partial\ell(\alpha_o,Z_i)}{\partial\alpha'}[\alpha - \alpha_o]$. Then by definition of $\widehat{\alpha}$, we have

$$0 \leq \frac{1}{n}\sum_{i=1}^n[\ell(\widehat{\alpha}, Z_i) - \ell(\widehat{\alpha} \pm \varepsilon_n\Pi_n\upsilon^*, Z_i)] = \mu_n\left(\ell(\widehat{\alpha}, Z_i) - \ell(\widehat{\alpha} \pm \varepsilon_n\Pi_n\upsilon^*, Z_i)\right) + E_o\left(\ell(\widehat{\alpha}, Z_i) - \ell(\widehat{\alpha} \pm \varepsilon_n\Pi_n\upsilon^*, Z_i)\right)$$
$$= \mp\varepsilon_n \times \frac{1}{n}\sum_{i=1}^n \frac{\partial\ell(\alpha_o, Z_i)}{\partial\alpha'}[\Pi_n\upsilon^*] + \mu_n\left(r[\widehat{\alpha}, \alpha_o, Z_i] - r[\widehat{\alpha} \pm \varepsilon_n\Pi_n\upsilon^*, \alpha_o, Z_i]\right) + E_o\left(r[\widehat{\alpha}, \alpha_o, Z_i] - r[\widehat{\alpha} \pm \varepsilon_n\Pi_n\upsilon^*, \alpha_o, Z_i]\right)$$

In the following we will show that:

$$\textbf{(A1.1)}\quad \frac{1}{n}\sum_{i=1}^n \frac{\partial\ell(\alpha_o, Z_i)}{\partial\alpha'}[\Pi_n\upsilon^* - \upsilon^*] = o_P(n^{-1/2});$$
$$\textbf{(A1.2)}\quad E_o\left(r[\widehat{\alpha}, \alpha_o, Z_i] - r[\widehat{\alpha} \pm \varepsilon_n\Pi_n\upsilon^*, \alpha_o, Z_i]\right) = \pm\varepsilon_n \times \langle\widehat{\alpha} - \alpha_o, \upsilon^*\rangle + \varepsilon_n \times o_P(n^{-1/2});$$
$$\textbf{(A1.3)}\quad \mu_n\left(r[\widehat{\alpha}, \alpha_o, Z_i] - r[\widehat{\alpha} \pm \varepsilon_n\Pi_n\upsilon^*, \alpha_o, Z_i]\right) = \varepsilon_n \times o_P(n^{-1/2}).$$

Under (A1.1) - (A1.3), together with $E_o\left(\frac{\partial\ell(\alpha_o,Z_i)}{\partial\alpha'}[\upsilon^*]\right) = 0$, we have:

$$0 \leq \frac{1}{n}\sum_{i=1}^n[\ell(\widehat{\alpha}, Z_i) - \ell(\widehat{\alpha} \pm \varepsilon_n\Pi_n\upsilon^*, Z_i)] = \mp\varepsilon_n \times \mu_n\left(\frac{\partial\ell(\alpha_o, Z_i)}{\partial\alpha'}[\upsilon^*]\right) \pm \varepsilon_n \times \langle\widehat{\alpha} - \alpha_o, \upsilon^*\rangle + \varepsilon_n \times o_P(n^{-1/2}).$$

Hence $\sqrt{n}\langle\widehat{\alpha} - \alpha_o, \upsilon^*\rangle = \sqrt{n}\mu_n\left(\frac{\partial\ell(\alpha_o,Z_i)}{\partial\alpha'}[\upsilon^*]\right) + o_P(1) \Rightarrow \mathcal{N}\left(0, ||\upsilon^*||^2\right)$. This, Assumption 3 and Assumption 4(1) together imply $\sqrt{n}(\rho(\widehat{\alpha}) - \rho(\alpha_o)) = \sqrt{n}\langle\widehat{\alpha} - \alpha_o, \upsilon^*\rangle + o_P(1) \Rightarrow \mathcal{N}\left(0, ||\upsilon^*||^2\right)$.

To complete the proof, it remains to establish (A1.1) - (A1.3). Notice that **(A1.1)** is implied by Chebyshev inequality, i.i.d. data, and $||\Pi_n \upsilon^* - \upsilon^*|| = o(1)$ which is satisfied given Assumption 4(2). For **(A1.2)** we notice

$$E_o\left(r[\alpha,\alpha_o,Z_i]\right) = E_o\left(\ell(\alpha,Z_i) - \ell(\alpha_o,Z_i) - \frac{\partial \ell(\alpha_o,Z_i)}{\partial\alpha'}[\alpha-\alpha_o]\right)$$

$$= E_o\left(\frac{1}{2}\frac{\partial^2 \ell(\alpha_o,Z_i)}{\partial\alpha\partial\alpha'}[\alpha-\alpha_o,\alpha-\alpha_o]\right) + \frac{1}{2}E_o\left(\frac{\partial^2 \ell(\widetilde{\alpha},Z_i)}{\partial\alpha\partial\alpha'}[\alpha-\alpha_o,\alpha-\alpha_o] - \frac{\partial^2 \ell(\alpha_o,Z_i)}{\partial\alpha\partial\alpha'}[\alpha-\alpha_o,\alpha-\alpha_o]\right)$$

for some $\widetilde{\alpha} \in \mathcal{A}_n$ in between α, α_o. It is easy to check that for any $v = (v_\theta, v_1, ..., v_m)' \in \mathbf{V}$, and $\widetilde{\alpha} \in \mathcal{A}_n$ with $||\widetilde{\alpha} - \alpha_o|| = O(\delta_n)$ we have

$$E_o\left(\frac{\partial^2 \ell(\widetilde{\alpha},Z)}{\partial\alpha\partial\alpha'}[v,v] - \frac{\partial^2 \ell(\alpha_o,Z)}{\partial\alpha\partial\alpha'}[v,v]\right)$$

$$= v_\theta' E_o\left(\frac{\partial^2 \log c(\widetilde{\alpha})}{\partial\theta\partial\theta'} - \frac{\partial^2 \log c(\alpha_o)}{\partial\theta\partial\theta'}\right)v_\theta + 2v_\theta'\sum_{j=1}^m E_o\left(\left\{\frac{\partial^2 \log c(\widetilde{\alpha})}{\partial\theta\partial u_j} - \frac{\partial^2 \log c(\alpha_o)}{\partial\theta\partial u_j}\right\}\int^{X_j} v_j(x)dx\right)$$

$$+\sum_{i=1}^m\sum_{j=1}^m E_o\left(\left\{\frac{\partial^2 \log c(\widetilde{\alpha})}{\partial u_i\partial u_j} - \frac{\partial^2 \log c(\alpha_o)}{\partial u_i\partial u_j}\right\}\int^{X_j} v_j(x)dx\int^{X_i} v_i(x)dx\right) - \sum_{j=1}^m E_o\left(\left[\frac{v_j(X_j)}{\widetilde{f}_j(X_j)}\right]^2 - \left[\frac{v_j(X_j)}{f_{oj}(X_j)}\right]^2\right).$$

Under Assumption 5, we have

$$E_o\left(r[\widehat{\alpha},\alpha_o,Z_i] - r[\widehat{\alpha}\pm\varepsilon_n\Pi_n\upsilon^*,\alpha_o,Z_i]\right) = -\frac{||\widehat{\alpha}-\alpha_o||^2 - ||\widehat{\alpha}\pm\varepsilon_n\Pi_n\upsilon^* - \alpha_o||^2}{2} + o_P(\varepsilon_n n^{-1/2})$$

$$= \pm\varepsilon_n \times \langle\widehat{\alpha}-\alpha_o, \Pi_n\upsilon^*\rangle + \frac{||\varepsilon_n\Pi_n\upsilon^*||^2}{2} + o_P(\varepsilon_n n^{-1/2}) = \pm\varepsilon_n \times \langle\widehat{\alpha}-\alpha_o, \upsilon^*\rangle + o_P(\varepsilon_n n^{-1/2})$$

where the last equality holds since Assumption 4(1)(2) implies

$$\langle\widehat{\alpha}-\alpha_o, \Pi_n\upsilon^* - \upsilon^*\rangle = o_P(n^{-1/2}) \text{ and } ||\Pi_n\upsilon^*||^2 \to ||\upsilon^*||^2 < \infty.$$

Hence **(A1.2)** is satisfied. For **(A1.3),** we notice

$$\mu_n\left(r[\widehat{\alpha},\alpha_o,Z_i] - r[\widehat{\alpha}\pm\varepsilon_n\Pi_n\upsilon^*,\alpha_o,Z_i]\right)$$

$$= \mu_n\left(\ell(\widehat{\alpha},Z_i) - \ell(\widehat{\alpha}\pm\varepsilon_n\Pi_n\upsilon^*,Z_i) - \frac{\partial \ell(\alpha_o,Z_i)}{\partial\alpha'}[\mp\varepsilon_n\Pi_n\upsilon^*]\right) = \mp\varepsilon_n \times \mu_n\left(\frac{\partial \ell(\widetilde{\alpha},Z_i)}{\partial\alpha'}[\Pi_n\upsilon^*] - \frac{\partial \ell(\alpha_o,Z_i)}{\partial\alpha'}[\Pi_n\upsilon^*]\right)$$

where $\widetilde{\alpha} \in \mathcal{A}_n$ is in between $\widehat{\alpha}$, $\widehat{\alpha}\pm\varepsilon_n\Pi_n\upsilon^*$. Since

$$\frac{\partial \ell(\widetilde{\alpha},Z)}{\partial\alpha'}[\Pi_n\upsilon^*] = \frac{\partial \log c(\widetilde{\alpha})}{\partial\theta'}\upsilon_\theta^* + \sum_{j=1}^m\left\{\frac{\partial \log c(\widetilde{\alpha})}{\partial u_j}\int 1(x\le X_j)\Pi_n\upsilon_j^*(x)dx + \frac{\Pi_n\upsilon_j^*(X_j)}{\widetilde{f}_j(X_j)}\right\},$$

(A1.3) is implied by Assumption 6.

The semiparametric efficiency is a direct application of Theorem 4 in Shen (1997). □

Proof. (**Proposition 1**): Recall that the semiparametric efficiency bound for θ_o is $\mathcal{I}_*(\theta_o) = E_o\left\{\mathcal{S}_{\theta_o}\mathcal{S}_{\theta_o}'\right\}$, where $\mathcal{S}_{\theta_o}$ is the *efficient score* function for θ_o, which is defined as the ordinary score function for θ_o minus its population least squares orthogonal projection onto the closed linear span (clsp) of the score functions for the nuisance parameters f_{oj}, $j = 1, ..., m$. And θ_o is $\sqrt{n}$-*efficiently estimable* if and only if $E_o\left\{\mathcal{S}_{\theta_o}\mathcal{S}_{\theta_o}'\right\}$ is *non-singular*; see e.g. Bickel, et al. (1993). Hence (14) is clearly a necessary condition for $\sqrt{n}$-normality and efficiency of $\widehat{\theta}$ for θ_o.

Under Assumptions 2 and 3', Propositions 4.7.4 and 4.7.6 of Bickel, et al. (1993, pages 165 - 168) for bivariate copula models can be directly extended to the multivariate case; see also Klaassen and Wellner (1997, Section 4). Therefore with $\mathcal{S}_{\theta_o}$ defined in (17), we have that $\mathcal{I}_*(\theta_o) = E_o\left\{\mathcal{S}_{\theta_o}\mathcal{S}_{\theta_o}'\right\}$ is finite, positive-definite. This implies that Assumption 3 is satisfied with $\rho(\alpha) = \lambda'\theta$ and $\omega = \infty$ and $||\upsilon^*||^2 = \|\rho_{\alpha_o}'\|^2 = \lambda'\mathcal{I}_*(\theta_o)^{-1}\lambda < \infty$. Hence Theorem 1 implies, for any $\lambda \in \mathcal{R}^{d_\theta}, \lambda \neq 0$, we have $\sqrt{n}(\lambda'\widehat{\theta} - \lambda'\theta_o) \Rightarrow \mathcal{N}\left(0, \lambda'\mathcal{I}_*(\theta_o)^{-1}\lambda\right)$. This implies Proposition 1. □

Proof. (**Propositions 2, 4, 6, 8**): The consistency of these asymptotic variances can be established by applying Ai and Chen (2003). □

Appendix B. Asymptotic Variances of the Ideal MLE and Modified Two-step Estimators of θ_o

To simplify notation and to save space, we present results for $m = 2$ and scalar θ only. We use $\widehat{\theta}_I$ to denote the ideal MLE and $\tilde{\theta}_M$ the modified two-step estimator. The asymptotic variance of the **ideal MLE** is $[n\mathcal{I}(\theta_o)]^{-1}$, where

$$\mathcal{I}(\theta_o) \equiv E[-\frac{\partial^2}{\partial\theta^2}\log\{c(F_{o1}(X_{1i}), F_{o2}(X_{2i}), \theta_o)\}].$$

Hence the asymptotic variance of $\widehat{\theta}_I$ can be consistently estimated by

$$[n\widehat{\mathcal{I}}(\widehat{\theta}_I)]^{-1} = \left[-\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\log\{c(F_{o1}(X_{1i}), F_{o2}(X_{2i}); \widehat{\theta}_I)\}\right]^{-1}.$$

Two-step estimator with a parametric margin: When $F_{o1}(\cdot) = F_{o1}(\cdot, \beta_o)$ is known up to unknown parameter $\beta_o \in int(\mathcal{B})$, the asymptotic variance of the modified two-step estimator is given by $([\mathcal{I}(\theta_o)] + Var(W_1(X_{1i}, \beta_o) + W_2(X_{2i})))\,[\mathcal{I}(\theta_o)]^{-2}$, where

$$\begin{aligned} W_2(X_{2i}) &= -\int I(F_{o2}(X_{2i}) \leqslant u_2)\frac{d\log(c(u_1, u_2, \theta_o))}{d\theta}\frac{d\log(c(u_1, u_2, \theta_o))}{du_2}c(u_1, u_2, \theta_o)du_1du_2, \\ W_1(X_{1i}, \beta_o) &= -E\left[\frac{d\log(c(U_{o1}, U_{o2}, \theta_o))}{d\theta}\frac{d\log(c(U_{o1}, U_{o2}, \theta_o))}{du_1}\frac{dF_{o1}(X_1, \beta_o)}{d\beta}\right] \\ &\quad \times\left(E\{-\frac{\partial^2\log f_{o1}(X_1, \beta_o)}{\partial\beta^2}\}\right)^{-1}\frac{d\log f_{o1}(X_{1i}, \beta_o)}{d\beta}. \end{aligned}$$

Using sample data and let $\widetilde{F}_{o1}(\cdot) = F_{o1}(\cdot, \widetilde{\beta})$, we can estimate $\mathcal{I}(\theta_o)$, $W_2(X_{2i})$ and $W_1(X_{1i}, \beta_o)$ respectively by

$$\widetilde{\sigma}^2 = \frac{-1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\log(c(\widetilde{F}_{o1}(X_{1i}), \widetilde{F}_{n2}(X_{2i}), \widetilde{\theta}_M)), \tag{21}$$

$$\widetilde{W}_2(X_{2i}) = \frac{-1}{n}\sum_{j:\widetilde{F}_{n2}(X_{2j})\geqslant\widetilde{F}_{n2}(X_{2i})}\frac{d\log c(\widetilde{F}_{o1}(X_{1j}), \widetilde{F}_{n2}(X_{2j}), \widetilde{\theta}_M)}{d\theta}\frac{d\log c(\widetilde{F}_{o1}(X_{1j}), \widetilde{F}_{n2}(X_{2j}), \widetilde{\theta}_M)}{du_2} \tag{22}$$

$$\begin{aligned} \widetilde{W}_{o1}(X_{1i}) &= \left[\frac{-1}{n}\sum_{j=1}^{n}\frac{d\log c(\widetilde{F}_{o1}(X_{1j}), \widetilde{F}_{n2}(X_{2j}), \widetilde{\theta}_M)}{d\theta}\frac{d\log c(\widetilde{F}_{o1}(X_{1j}), \widetilde{F}_{n2}(X_{2j}), \widetilde{\theta}_M)}{du_1}\frac{dF_{o1}(X_{1j}, \widetilde{\beta})}{d\beta}\right] \\ &\quad \times\left(\frac{-1}{n}\sum_{j=1}^{n}\frac{\partial^2\log f_{o1}(X_{1j}, \widetilde{\beta})}{\partial\beta^2}\right)^{-1}\frac{d\log f_{o1}(X_{1i}, \widetilde{\beta})}{d\beta}. \end{aligned}$$

Hence a consistent estimator of the asymptotic variance of $\widetilde{\theta}_M$ is given by

$$\widehat{avar(\widetilde{\theta}_M)} = \frac{1}{n\widetilde{\sigma}^2}\left[1 + \widetilde{\sigma}^{-2}\frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{W}_{o1}(X_{1i}) + \widetilde{W}_2(X_{2i})\right)^2\right].$$

Two-step estimator with a known margin: When $F_{o1}(\cdot) = F_{o1}(\cdot, \beta_o)$ is known with known β_o, the asymptotic variance of the modified two-step estimator $\widetilde{\theta}_M$ of θ_o is given by $(\mathcal{I}(\theta_o) + var(W_2(X_2)))\,[\mathcal{I}(\theta_o)]^{-2}$, and a consistent estimator of the asymptotic variance of $\widetilde{\theta}_M$ is given by

$$\widehat{avar(\widetilde{\theta}_M)} = \frac{1}{n\widetilde{\sigma}^2}\left[1 + \widetilde{\sigma}^{-2}\frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{W}_2(X_{2i})\right)^2\right],$$

where $\widetilde{\sigma}^2$ and $\widetilde{W}_2(X_{2i})$ are given in (21) and (22) except we replace $F_{o1}(\cdot, \widetilde{\beta})$ by $F_{o1}(\cdot, \beta_o)$.
Two-step estimator with equal but unknown margins: When $F_{o1} = F_{o2} = F_o$, the asymptotic variance of the modified two-step estimator is $(\mathcal{I}(\theta_o) + var\{W_1(X_1) + W_2(X_2)\})\,[\mathcal{I}(\theta_o)]^{-2}$, where

$$W_k(X_k) = -\int I(F_o(X_k) \leqslant u_k)\frac{d\log(c(u_1,u_2,\theta_o))}{d\theta}\frac{d\log(c(u_1,u_2,\theta_o))}{du_k}c(u_1,u_2,\theta_o)du_1du_2,$$

for $k = 1, 2$. Note that when $F_{o1} = F_{o2}$, this asymptotic variance coincides with the asymptotic variance of the original two-step estimator proposed in Genest, et al. (1995).

Using the sample data we can estimate $\mathcal{I}(\theta_o)$ and $W_k(X_{ki})$ respectively by

$$\begin{aligned}
\widetilde{\sigma}^2 &= -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\log(c(\widetilde{F}(X_{1i}), \widetilde{F}(X_{2i}), \widetilde{\theta}_M)),\\
\widetilde{W}_k(X_{ki}) &= -\frac{1}{n}\sum_{j:\widetilde{F}(X_{kj})\geqslant\widetilde{F}(X_{ki})}\frac{d\log(c(\widetilde{F}(X_{1j}), \widetilde{F}(X_{2j}), \widetilde{\theta}_M))}{d\theta}\frac{d\log(c(\widetilde{F}(X_{1j}), \widetilde{F}(X_{2j}), \widetilde{\theta}_M))}{du_k}.
\end{aligned}$$

Hence a consistent estimator of the asymptotic variance of $\widetilde{\theta}_M$ is given by

$$\widehat{avar(\widetilde{\theta}_M)} = \frac{1}{n\widetilde{\sigma}^2}\left[1 + \widetilde{\sigma}^{-2}\frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{W}_1(X_{1i}) + \widetilde{W}_2(X_{2i})\right)^2\right].$$

References

[1] Ai, C. and X. Chen (2003). "Efficient estimation of models with conditional moment restrictions containing unknown functions." *Econometrica* 71, 1795-1843.

[2] Barron, A. and C. Sheu (1991). "Approximation of density functions by sequences of exponential families." *The Annals of Statistics* 19, 1347-1369.

[3] Bickel, P., C. Klaassen, Y. Ritov, and J. Wellner (1993). *Efficient and Adaptive Estimation for Semiparametric Models*, Baltimore, Maryland: John Hopkins University Press.

[4] Chen, X. and Y. Fan (2005a). "Estimation of copula-based semiparametric time series models." Forthcoming in *Journal of Econometrics.*

[5] Chen, X. and Y. Fan (2005b). "Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification." Forthcoming in *Journal of Econometrics.*

[6] Chen, X. and H. White (1999). "Improved rates and asymptotic normality for nonparametric neural network estimators." *IEEE Tran. Information Theory* 45, 682-691.

[7] Chui, C.K. (1992). *An Introduction to Wavelets*, San Diego, CA: Academic Press, Inc.

[8] Clayton, D. (1978). "A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence." *Biometrika* 65, 141-151.

[9] Clayton, D. and J. Cuzick (1985). "Multivariate generalizations of the proportional hazards model" (with discussion). *J. R. Statist. Soc.* A 148, 82-117.

[10] Coppejans, M. and A.R. Gallant (2002). "Cross-validated SNP density estimates." *Journal of Econometrics* 110, 27-65.

[11] Embrechts, P. , A. McNeil, and D. Straumann (2002). "Correlation and dependence properties in risk management: properties and pitfalls," in M. Dempster, ed., *Risk Management: Value at Risk and Beyond*, Cambridge University Press, 176-223.

[12] Frees, Edward W. and E. A. Valdez (1998). "Understanding relationships using copulas." *North American Actuarial Journal* 2, 1-25.

[13] Gallant, A.R. and D. Nychka (1987). "Semi-non-parametric maximum likelihood estimation." *Econometrica* 55, 363-390.

[14] Geman, S. and C. Hwang (1982). "Nonparametric maximum likelihood estimation by the method of sieves." *The Annals of Statistics* 10, 401-414.

[15] Genest, C. (1987). "Frank's family of bivariate distributions." *Biometrika* 74, 549-555.

[16] Genest, C. and J. MacKay (1986). "The joy Of copulas: Bivariate distributions with uniform marginals." *American Statistician* 40, 280-283.

[17] Genest, C., K. Ghoudi and L.-P. Rivest (1995). "A semiparametric estimation procedure of dependence parameters in multivariate families of distributions." *Biometrika* 82, 534-552.

[18] Genest, C. and B. Werker (2002). "Conditions for the asymptotic semiparametric efficiency of an omnibus estimator of dependence parameters in copula models." *Proceedings of the Conference on Distributions with Given Marginals and Statistical Modelling*, C. M. Cuadras and J. A. Rodríguez Lallena (eds).

[19] Granger, C.W.J., T. Teräsvirta and A. Patton (2003). "Common factors in conditional distributions." Forthcoming in *Journal of Econometrics*.

[20] Grenander, U. (1981). *Abstract Inference*, New York: Wiley Series.

[21] Heckman, J. J. and B. E. Honoré (1989). "The identifiability of the competing risks model." *Biometrika* 76(2), 325-330.

[22] Joe, H. (1997). *Multivariate Models and Dependence Concepts*, Chapman & Hall/CRC.

[23] Klaassen, C. and J. Wellner (1997). "Efficient estimation in the bivariate Normal copula model: Normal margins are least-favorable." *Bernoulli* 3, 55-77.

[24] Lee, L. (1983). "Generalized econometric models with selectivity." *Econometrica* 51, 507-512.

[25] Maguluri, G. (1993). "Semiparametric estimation of association in a bivariate survival function." *The Annals of Statistics* 21, 1648-1662.

[26] Nelsen, R. B. (1999). *An Introduction to Copulas*, Springer-Verlag, New York.

[27] Oakes, D. (1982). "A model for association in bivariate survival data." *J. R. Statist. Soc. B* 44, 414-422.

[28] Oakes, D. (1986). "Semiparametric inference in a model for association in bivariate survival data." *Biometrika* 73, 353-361.

[29] Oakes, D. (1989). "Bivariate survival models induced by frailties." *J. Amer. Statist. Assoc.* 84, 487-493.

[30] Oakes, D. (1994). "Multivariate survival distributions." *J. Nonparametric Statist.* 3, 343-354.

[31] Patton, A.J. (2004). "On the importance of skewness and asymmetric dependence in stock returns for asset allocation." Forthcoming, *Journal of Financial Econometrics*.

[32] Robinson, P. (1988). "Root-N-consistent semiparametric regression." *Econometrica* 56, 931-954.

[33] Schumaker, L. (1981). *Spline Functions: Basic Theory,* New York: John Wiley & Sons.

[34] Shen, X. (1997). "On methods of sieves and penalization." *The Annals of Statistics* 25, 2555-2591.

[35] Shen, X. , H.-C. Huang, and J. Ye (2004). "Inference after model selection." *Journal of the American Statistical Association* 99, 751-762.

[36] Shen, X. and W. Wong (1994). "Convergence rate of sieve estimates." *The Annals of Statistics* 22, 580-615.

[37] Shen, X. and J. Ye (2002). "Adaptive model selection." *Journal of the American Statistical Association* 97, 210-221.

[38] Shih, J.H. and T.A. Louis (1995). "Inferences on the association parameter in copula models for bivariate survival data." *Biometrics* 51, 1384-1399.

[39] Sklar, A. (1959). "Fonctions de répartition à n dimensionset leurs marges." *Publ. Inst. Statis. Univ. Paris* 8, 229-231.

[40] Stone, C. (1982). "Optimal global rates of convergence for nonparametric regression." *The Annals of Statistics* 10, 1040-1053.

[41] Stone, C. (1990). "Large-sample inference for log-spline models." *The Annals of Statistics* 18, 717-741.

[42] Van de Geer, S. (2000). *Empirical Processes in M-estimation*, Cambridge University Press.

[43] Wong, W. and X. Shen (1995). "Probability inequalities for likelihood ratios and convergence rates for sieve MLE's." *The Annals of Statistics* 23, 339-362.