EFFICIENT ESTIMATION OF SEMIPARAMETRIC MULTIVARIATE COPULA MODELS

by

Xiaohong Chen, Yanqin Fan, and Victor Tsyrennikov

Working Paper No. 04-W20

September 2004

DEPARTMENT OF ECONOMICS
VANDERBILT UNIVERSITY
NASHVILLE, TN 37235

www.vanderbilt.edu/econ

Efficient Estimation of Semiparametric Multivariate Copula Models*

Xiaohong Chen[†] Yanqin Fan[‡] Victor Tsyrennikov[§]

First Version: May 2002; This Version: September 2004

Abstract

We propose a sieve maximum likelihood (ML) estimation procedure for a broad class of semiparametric multivariate distribution models. A joint distribution in this class is characterized by a parametric copula function evaluated at nonparametric marginal distributions. This class of models 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 ML estimates 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 ML 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 ML estimates 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 O. Linton and participants at 2003 North American Econometric Society Summer Meetings in Chicago, 2004 DeMoSTAFI conference in Quebec City for helpful comments. 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.

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

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

§Victor Tsyrennikov, Department of Economics, New York University, 269 Mercer Street, 7th Floor, New York, NY 10003, USA; E-mail: vt23@nyu.edu.

1 Introduction

Suppose we observe an i.i.d. sample $\{Z_i \equiv (X_{1i}, ..., X_{mi})'\}_{i=1}^n$ 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 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.

A class of semiparametric multivariate distribution models has gained popularity in diverse fields in recent years 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. To introduce this class, let F_{oj} denote the true unknown marginal cdf of X_j, $j = 1, ..., m$. Assume that F_{oj}, $j = 1, ..., m$, are continuous. By the Sklar's (1959) theorem, there exists a unique copula function C_o such that $H_o(X_1, ..., X_m) \equiv C_o(F_{o1}(X_1), ..., F_{om}(X_m))$. Let f_{oj}, $j = 1, ..., m$, and $c_o(u_1, ..., u_m)$ denote the probability densities associated with F_{oj}, $j = 1, ..., m$, and C_o respectively. 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 $C(u_1, ..., u_m; \theta)$ is a class of parametric copula functions. Then for any $(x_1, ..., x_m) \in \mathcal{X}_1 \times ... \times \mathcal{X}_m$, the pdf h_o has the following representation:

$$h_o(x_1, ..., x_m) \equiv c(F_{o1}(x_1), ..., F_{om}(x_m); \theta_o) \prod_{j=1}^{m} f_{oj}(x_j), \tag{1}$$

where $c(u_1, \ldots, u_m; \theta_o)$ is the density of the copula $C(u_1, \ldots, u_m; \theta_o)$ and the functional forms of f_{oj}, $j = 1, ..., m$, are unknown. We refer to the class of multivariate distributions with density functions of the form (1) as the class of copula-based semiparametric multivariate distributions. It 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 $X_1, \ldots, X_m$. It is obvious that the copula measure of dependence is invariant to any increasing transformation of the univariate random variables X_j, $j = 1, \ldots, m$.

The class of semiparametric multivariate copula models has been used extensively in applied work, where modeling and estimating the dependence structure between several random variables are of interest. Specific applications include those in finance 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 (1982, 1983), Heckman and Honore (1989), Granger, et al. (2003) and Patton (2004)), to name only a few.

Because of its special role in a semiparametric multivariate copula model, estimation of the copula parameter θ_o has attracted much attention from researchers including Clayton (1978), Clayton and Cuzick (1985), Oakes (1982, 1986, 1994), Genest (1987) and Genest, et al. (1995). One of the most commonly used estimators of θ_o in recent applied work is the two-step estimator proposed by Oakes (1994) and Genest, et al. (1995):

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

where $\widetilde{F}_{nj}(x_j) = \frac{1}{n+1}\sum_{i=1}^{n} 1\{X_{ji} \le x_j\}$ 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 the two-step estimator $\widetilde{\theta}_n$. Shih and Louis (1995) independently propose the two-step estimator and establish its asymptotic properties for i.i.d. data with random censoring. The large sample results of Genest, et al. (1995) have also been extended to time series setting in Chen and Fan (2002, 2003).

Despite its popularity, the two-step estimator of the copula dependence parameter θ_o is generally asymptotically inefficient except for a few special cases; see Genest and Werker (2002). Klaassen and Wellner (1997) show that it is efficient for the Gaussian copula models and Genest, et al. (1995) show that it is efficient for the independence copula model. Bickel, et al. (1993, chapter 4.7) provide some efficiency score characterization for θ_o in general bivariate semiparametric copula models, but provide no efficient estimator for it. For semiparametric bivariate survival Clayton copula models, Maguluri (1993) also provides some efficiency score calculations for θ_o and conjectures that the estimator proposed in his paper might be efficient. To the best of our knowledge (see Genest and Werker (2002)), there does not exist any published results on efficient estimation procedure of θ_o for general bivariate (multivariate) semiparametric copula models. Moreover, 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 copula parameter θ_o and the marginal distributions F_{oj}, $j = 1, \ldots, m$. Except in models with the independence copula, it is clear that the univariate (rescaled) empirical distributions are generally inefficient estimates of the marginal distributions. Intuitively one could obtain more efficient estimates of F_{oj}, $j = 1, \ldots, m$ by utilizing the dependence information contained in the parametric copula. Unfortunately even for semiparametric models with the Gaussian copula, there is currently no efficient estimates of univariate marginal distributions, see Klaassen and Wellner (1997). For the special case of a bivariate copula model with one known marginal distribution and one unknown marginal distribution, Bickel, et al. (1993, chapter 6.7) provide some efficiency score calculations for the unknown margin, but they again do not present any efficient estimators.

In this paper, we propose a general sieve maximum likelihood (ML) estimation procedure for all the unknown parameters in a semiparametric multivariate copula model (1). Intuitively, we approximate the infinite-dimensional unknown marginal densities f_{oj}, $j = 1, ..., m$ by combinations of finite-dimensional known basis functions with increasing complexity (sieves), and then maximize the joint likelihood with respect to the copula dependence parameter and the sieve parameters of the approximation of the marginal densities. By applying the general theory of Shen (1997) on asymptotic efficiency of the sieve ML estimates, we show that our plug-in sieve ML estimates of all smooth functionals, including the copula parameter and the unknown marginal distributions, are semiparametrically efficient. Although the asymptotic variances of these smooth functionals cannot be derived in closed-form, they can be estimated easily and consistently. As our sieve ML procedure involves approximating and estimating one-dimensional unknown functions (marginal densities) only, it does not suffer from 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 distributions to produce more efficient estimates. Examples of such restrictions include equal but unknown marginal distributions, known parametric forms of some marginal distributions, to name only a few. Simulation studies show clearly the efficiency gains of our sieve ML estimates over the two-step estimator of θ_o and the empirical distribution estimates of the marginal distributions, especially so when prior restrictions are incorporated. We find that the sieve ML estimate of θ_o in models with some nonparametric and some parametric margins perform almost as well as the infeasible ML estimates of θ_o obtained as if all the marginal distributions are known.

The rest of this paper is organized as follows. In Section 2, we introduce our sieve ML estimators of the copula parameter and the unknown marginal distributions in models with or without restrictions on the marginal distributions. In Section 3, we show that for semiparametric multivariate copula models with unknown marginal distributions, the plug-in sieve ML estimates of all smooth functionals are root-n normal and semiparametrically efficient. These results are then applied to deliver the root-n asymptotic normality and efficiency of the sieve ML estimates of the copula parameter and the marginal distributions. We also provide simple consistent estimators of the asymptotic variances of these estimators. In Section 4, we extend the efficiency results in Section 3 to models with equal but unknown margins and models with some parametric margins. Section 5 provides results from a simulation study. All the proofs are gathered into Appendix A.

2 The Sieve ML Estimators

In this section, we will introduce sieve ML estimation of parameters in a semiparametric multivariate copula model in various cases including i) the marginal distributions are completely unspecified; ii) the marginal distributions are the same, but unspecified otherwise; iii) some of the marginal distributions are parameterized, but the others are unspecified.

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 $j = 1, \ldots, m$. 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), Robinson (1988), Newey (1997) and Horowitz (1998). 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., if 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. Define a Hölder ball with smoothness $r = J + \gamma$ as

$$\Lambda_K^r(\mathcal{X}_j) = \{h \in C^J(\mathcal{X}_j) : \sup_{x,y \in \mathcal{X}_j, x \neq y} \frac{|D^J h(x) - D^J h(y)|}{|x - y|^\gamma} \leq K\}.$$

2.1.1 Bounded support

It is known that functions in $\Lambda^r(\mathcal{X}_j)$ with $r > 1/2$ and $\mathcal{X}_j = [0,1]$ can be well approximated by many sieve bases such as the polynomial sieve $\text{Pol}(K_n)$, the trigonometric sieve $\text{TriPol}(K_n)$ and the cosine series $\text{CosPol}(K_n)$:

$$\text{Pol}(K_n) = \left\{ \sum_{k=0}^{K_n} a_k x^k,\ x \in [0,1] :\ a_k \in \mathcal{R} \right\};$$

$$\text{TriPol}(K_n) = \left\{ a_0 + \sum_{k=1}^{K_n} [a_k \cos(k\pi x) + b_k \sin(k\pi x)],\ x \in [0,1] :\ a_k, b_k \in \mathcal{R} \right\};$$

$$\text{CosPol}(K_n) = \left\{ a_0 + \sum_{k=1}^{K_n} a_k \cos(k\pi x),\ x \in [0,1] :\ a_k \in \mathcal{R} \right\}.$$

They can also be well approximated by the spline sieve $\text{Spl}(\gamma, K_n)$, which is a linear space of dimension $(K_n + \gamma + 1)$ consisting of spline functions of degree γ 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 $0 = t_0 < t_1 < \cdots < t_{K_n} < t_{K_n+1} = 1$. Partition $[0,1]$ into $K_n + 1$ subintervals $I_k = [t_k, t_{k+1})$, $k = 0, \ldots, K_n - 1$, and $I_{K_n} = [t_{K_n}, t_{K_n+1}]$. We assume that the knots $t_1, \ldots, t_{K_n}$ have bounded mesh ratio:

$$\frac{\max_{0 \le k \le K_n}(t_{k+1} - t_k)}{\min_{0 \le k \le K_n}(t_{k+1} - t_k)} \le const.$$

A function on $[0,1]$ is a spline of degree γ with knots $t_1, \ldots, t_{K_n}$ if it is: (i) a polynomial of degree γ or less on each interval I_k, $k = 0, \ldots, K_n$; and (ii) $(\gamma - 1)$-times continuously differentiable on $[0,1]$. See Schumaker (1981) for details on univariate splines.

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

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

2.1.2 Unbounded support

There are also many sieves that can approximate densities with support $\mathcal{X}_j = \mathcal{R}$. Here we present two examples: (i) if density f_{oj} has close to exponential thin tails over $\mathcal{X}_j = \mathcal{R}$, we can use the Hermite polynomial sieve to approximate f_{oj}:

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

where $K_{nj} \to \infty$, $K_{nj}/n \to 0$ as in Gallant and Nychka (1987); (ii) if density f_{oj} has polynomial fat tails over $\mathcal{X}_j = \mathcal{R}$, 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:

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

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

2.2 Sieve MLEs

To avoid introducing too many notations, we use the same notation $\hat{\alpha}_n$ to denote the sieve ML estimates for all cases considered with or without prior restriction on the marginal distributions. That is, it changes from case to case.

2.2.1 Unknown margins

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 $\widehat{\alpha}_n = (\widehat{\theta}'_n, \widehat{f}_{n1}, ..., \widehat{f}_{nm})' \in \Theta \times \prod_{j=1}^m \mathcal{F}_{nj} = \mathcal{A}_n$ denote the sieve ML estimator:

$$\begin{aligned} \widehat{\alpha}_n &= \operatorname{argmax}_{\theta\in\Theta, f_j\in\mathcal{F}_{nj}} \sum_{i=1}^{n} \log\left\{ c(U_{1i}, ..., U_{mi}; \theta) \prod_{j=1}^{m} f_j(X_{ji}) \right\} \quad (8) \\ \text{with } U_{ji} &\equiv F_j(X_{ji}) = \int_{\mathcal{X}_j} 1(x \le X_{ji}) f_j(x)dx, \quad j = 1, ..., m, \end{aligned}$$

where $f_j \in \mathcal{F}_{nj}$ for $j = 1, ..., m$, and the sieve space $\mathcal{F}_{nj}$ is (4) if $\mathcal{X}_j$ is a bounded interval, and $\mathcal{F}_{nj}$ 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 $\hat{F}_{nj}(x_j) = \int 1(y \le x_j)\hat{f}_{nj}(y)dy$, $j = 1, ..., m$.

Remark 1: We note that the sieve MLE optimization problem can be rewritten as an unconstrained optimization problem

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

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 the second order derivative of $f_j(\cdot; a_{jn})$, and λ_{jn}'s are penalization factors.

Noting 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 shall illustrate this in the next two subsections.

2.2.2 Equal but unknown margins

Now suppose the marginal distributions are all equal but unknown, i.e., $F_{oj} = F_o$ ($f_{oj} = f_o$) 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. The sieve MLE $\widehat{\alpha}_n = (\widehat{\theta}'_n, \widehat{f}_n)' \in \Theta \times \mathcal{F}_{n1} = \mathcal{A}_n$ is now given by:

$$\begin{aligned} \widehat{\alpha}_n &= \text{argmax}_{\theta\in\Theta, f\in\mathcal{F}_{n1}} \sum_{i=1}^{n} \log\left\{ c(U_{1i}, ..., U_{mi}; \theta) \prod_{j=1}^{m} f(X_{ji}) \right\} \qquad (9) \\ \text{with } U_{ji} &\equiv F(X_{ji}) = \int_{\mathcal{X}} 1(x \le X_{ji}) f(x) dx, \quad j = 1, ..., m. \end{aligned}$$

This procedure can be easily extended to the case where some but not all marginal distributions are equal.

2.2.3 Some parametric margins

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 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 $\widehat{\alpha}_n = (\widehat{\theta}'_n, \widehat{\beta}'_n, \widehat{f}_{n2}, ..., \widehat{f}_{nm})' \in \Theta \times \mathcal{B} \times \prod_{j=2}^{m} \mathcal{F}_{nj} = \mathcal{A}_n$ denote the sieve ML estimator:

$$\begin{aligned} \widehat{\alpha}_n &= \text{argmax}_{\substack{\theta\in\Theta,\beta\in\mathcal{B}, \\ f_j\in\mathcal{F}_{nj}, j=2,...,m}} \sum_{i=1}^{n} \log\left\{ c(U_{1i}, ..., U_{mi}; \theta) f_{o1}(X_{1i}, \beta) \prod_{j=2}^{m} f_j(X_{ji}) \right\} \qquad (10) \\ \text{with } U_{1i} &\equiv F_{o1}(X_{ji}, \beta),\ U_{ji} \equiv F_j(X_{ji}) = \int_{\mathcal{X}_j} 1(x \le X_{ji}) f_j(x) dx,\ j = 2, ..., m. \end{aligned}$$

When $F_{o1}(\cdot)$ is completely known, we simply take $\mathcal{B} = \{\beta_o\}$ and $\hat{\beta}_n = \beta = \beta_o$ in the above optimization problem (10).

3 Asymptotic Normality and Efficiency of Smooth Functionals

Let $\rho : \mathcal{A} \to \mathcal{R}$ be a functional of interest and $\rho(\widehat{\alpha}_n)$ be the plug-in sieve ML estimate 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 our sieve MLE estimator $\rho(\widehat{\alpha}_n)$ for smooth functionals ρ of $\alpha_o = (\theta'_o, f_{o1}, ..., f_{om})'$.

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

Let $\ell(\alpha, Z_i) = \log\{c(F_1(X_{1i}), ..., F_m(X_{mi}); \theta) \prod_{j=1}^{m} f_j(X_{ji})\}$ and $E_o(\cdot)$ be the expectation under 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: $\frac{\partial^2 \log c(u,\theta)}{\partial \theta^2}$, $\frac{\partial^2 \log c(u,\theta)}{\partial u_j \partial \theta}$, $\frac{\partial^2 \log c(u,\theta)}{\partial u_j \partial u_i}$ for $j, i = 1, ..., m$.

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

$$\langle v, \widetilde{v} \rangle \equiv E_o \left[\left(\frac{\partial \ell(\alpha_o, Z)}{\partial \alpha'}[v] \right) \left(\frac{\partial \ell(\alpha_o, Z)}{\partial \alpha'}[\widetilde{v}] \right) \right], \tag{11}$$

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

$$\overline{\mathbf{V}}_j = \left\{ v_j : \mathcal{X}_j \to \mathcal{R} : E_o \left(\frac{v_j(X_j)}{f_{oj}(X_j)} \right) = 0, E_o \left(\frac{v_j(X_j)}{f_{oj}(X_j)} \right)^2 < \infty \right\}. \tag{12}$$

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 $||v|| \le \varepsilon$, we have

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

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

$$\langle v^*, v \rangle = \frac{\partial \rho(\alpha_o)}{\partial \alpha'}[v] \quad \text{for all } v \in \mathbf{V} \tag{13}$$

and

$$||v^*||^2 = \|\frac{\partial \rho(\alpha_o)}{\partial \alpha'}\|^2 = \sup_{v \in \mathbf{V}:||v||>0} \frac{\left|\frac{\partial \rho(\alpha_o)}{\partial \alpha'}[v]\right|^2}{||v||^2} < \infty \tag{14}$$

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 the Appendix 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 ML estimates of unknown marginal densities $\widehat{f}_{nj}$, $j = 1, ..., m$. There exist many results on convergence rates of general sieve estimates of an univariate 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 variable, 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 $\upsilon^* = (\upsilon_\theta^{*\prime}, \upsilon_1^*, ..., \upsilon_m^*)' \in \overline{\mathbf{V}}$ is uniquely determined by $b^* = (\upsilon_\theta^{*\prime}, b_1^*, ..., b_m^*)' \in \overline{\mathbf{B}}$ (and vise versa) via the relation:

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

Then Assumption 4(2) can be replaced by

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

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

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 $\upsilon^* \in \overline{\mathbf{V}}$ satisfying (15) and (16):

$$\lambda'(\theta - \theta_o) = \langle \alpha - \alpha_o, \upsilon^*\rangle \quad \text{for any } \alpha - \alpha_o \in \overline{\mathbf{V}} \tag{15}$$

and

$$||\frac{\partial\rho(\alpha_o)}{\partial\alpha'}||^2 = ||\upsilon^*||^2 = \langle \upsilon^*, \upsilon^*\rangle = \sup_{v\neq 0, v\in\overline{\mathbf{V}}} \frac{|\lambda' v_\theta|^2}{||v||^2} < \infty. \tag{16}$$

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

$$\mathcal{S}_{\theta_o}' = \frac{\partial\log c(U_o,\theta_o)}{\partial\theta'} - \sum_{j=1}^m\left[\frac{\partial \log c(U_o,\theta_o)}{\partial u_j}\int_0^{U_{oj}} g_j^*(u)du + g_j^*(U_{oj})\right], \tag{17}$$

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 $k = 1, ..., d_\theta$,

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

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

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

Hence (16) 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.

Although the asymptotic variance $\mathcal{I}_*(\theta_o)^{-1}$ of $\hat{\theta}_n$ has no closed form expression, it can be consistently estimated by the following simple procedure. 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:

$$\mathbf{A}_n = \{(e_1, ..., e_{d_\theta}) : e_j(\cdot) \in \mathbf{B}_n, \ j = 1, ..., d_\theta\}, \tag{18}$$

$$\mathbf{B}_n = \{e(u) = \sum_{k=1}^{K_{n\theta}} a_k\sqrt{2}\cos(k\pi u), \ u \in [0,1], \sum_{k=1}^{K_{n\theta}} a_k^2 < \infty\}, \tag{19}$$

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

3.3 Sieve ML Estimates 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 ML estimate: $\rho(\widehat{\alpha}) = \widehat{F}_{nj}(x_j) = \int 1(y \leq x_j)\widehat{f}_{nj}(y)dy$, 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 $v^* \in \overline{\mathbf{V}}$ should satisfy (20) and (21):

$$\langle v^*, v\rangle = \frac{\partial \rho(\alpha_o)}{\partial \alpha'}[v] = E_o\left(1(X_j \leq x_j)\frac{v_j(X_j)}{f_{oj}(X_j)}\right) \quad \text{for all } v \in \mathbf{V} \tag{20}$$

$$\|\frac{\partial \rho(\alpha_o)}{\partial \alpha'}\|^2 = ||v^*||^2 = ||b^*||^2 = \sup_{b \in \overline{\mathbf{B}}: ||b||>0} \frac{\left|E_o\left(1(U_{oj} \leq F_{oj}(x_j)) b_j(U_{oj})\right)\right|^2}{||b||^2}. \tag{21}$$

Proposition 3. Let $v^* \in \overline{\mathbf{V}}$ solve (20) and (21). Suppose that Assumptions 1 - 2 and 4 - 6 hold. Then for any fixed $x_j \in \mathcal{X}_j$ and for $j = 1, ..., m$, $\sqrt{n}(\widehat{F}_{nj}(x_j) - F_{oj}(x_j)) \Rightarrow \mathcal{N}\left(0, ||v^*||^2\right)$. Moreover, $\widehat{F}_{nj}$ is semiparametrically efficient.

For general copulas including the Gaussian copula, there does not seem to be a closed-form solution to (20) and (21) for the representer $v^* \in \overline{\mathbf{V}}$ and the asymptotic variance $||v^*||^2$. Nevertheless, the asymptotic variance $||v^*||^2$ can again be consistently estimated. 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 (19).

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

Remark 2: In the special case of the independence copula ($c(u_1, ..., u_m, \theta) = 1$), we could solve (20) and (21) 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 (20) and (21) are satisfied with $v^*_j(X_j) = \{1(X_j \leq x_j) - E_o[1(X_j \leq x_j)]\} f_{oj}(X_j)$ and $v^*_k = 0$ for all $k \neq j$. Hence

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

Thus for models with the independence copula, the plug-in sieve ML estimate 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 $\widetilde{F}_{nj}(x_j) = \frac{1}{n}\sum_{i=1}^n 1\{X_{ji} \leq x_j\}$ of F_{oj}. For models with parametric copula functions that are not independent, we have $||v^*||^2 \leq F_{oj}(x_j)\{1 - F_{oj}(x_j)\}$.

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 subsections 2.2.2 and 2.2.3.

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 (12).

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, one immediately concludes 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 for 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 (18), and $\mathbf{B}_n$ is the sieve space (19).

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 (12).

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 $\sqrt{n}(\widehat{F}_{nj}(x) - F_{oj}(x)) \Rightarrow \mathcal{N}\left(0, ||\upsilon^*||^2\right)$ where $||\upsilon^*||^2 = ||b^*||^2 =$

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

Proposition 8. Under conditions for Proposition 7, we have:
(i) $\widehat{\sigma}_\theta^2 = \mathcal{I}_*(\theta_o) + o_p(1)$, where $\widehat{\sigma}_\theta^2 =$

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

(ii) $\widehat{\sigma}_{F_j}^2(x_j) = ||\upsilon^*||^2 + o_p(1)$, 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 $\hat{\alpha}_n = (\hat{\theta}_n, \hat{f}_{n2}, \ldots, \hat{f}_{nm})$ be defined as in (10) except that $\beta = \beta_o$ is treated as known. 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 $\{\widehat{\sigma}_\theta^2\}^{-1}$, with

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

and the asymptotic variance of $\sqrt{n}(\widehat{F}_{nj}(x) - F_{oj}(x))$ can be consistently estimated by $\{\widehat{\sigma}^2_{F_j}(x_j)\}^{-1}$, 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 (18), and $\mathbf{B}_n$ is the sieve space (19).

5 A Simulation Study

This section presents results from a small Monte Carlo study to assess the finite sample performance of the sieve ML estimates. We first introduce the simulation design and the estimators studied in this section and then present the Monte Carlo results.

5.1 Simulation Design and the Methods of Estimation

The data $\{(X_{1i}, X_{2i})\}_{i=1}^n$ are generated from a semiparametric bivariate copula-based model with the Clayton copula: $C(F_{o1}(x_1), F_{o2}(x_2); \theta_o)$, where the Clayton copula density $c(u_1, u_2; \theta)$ is given by

$$c(u_1, u_2; \theta) = (1+\theta) u_1^{-(\theta+1)} u_2^{-(\theta+1)} [u_1^{-\theta} + u_2^{-\theta} - 1]^{-(\theta^{-1}+2)}, \quad \text{where } \theta > 0.$$

We have used the algorithm of Genest and MacKay (1986) to simulate data from the Clayton copula and then transformed them to have marginals F_{o1} and F_{o2} respectively. Two classes of DGPs denoted by $(\theta_o, F_{o1}, F_{o2})$ are considered:

DGP I. The two marginals are different: $(\theta_o, F_{o1}, F_{o2}) = (\theta_o, t_{[5]}, t_{[25]})$ with $\theta_o = 5, 10, 15$.

DGP II. The two marginals are the same: $(\theta_o, F_{o1}, F_{o2}) = (\theta_o, t_{[5]}, t_{[5]})$ with $\theta_o = 5, 10, 15$.

In terms of estimation, we considered estimators that take into account prior information in the following cases:

Case I. the two marginals are different and are completely unknown;

Case II. the two marginals are the same, but otherwise completely unknown;

Case III. the first marginal is of a parametric form and the second one is completely unknown;

Case IV. the first marginal is completely known and the second one is completely unknown.

For each case, we consider the methods of sieve ML estimation, two-step estimation, and an infeasible ML estimation where both margins are assumed to be known. From Case I to Case IV, there is more and more information about the marginal distributions, our theoretical results suggest that the sieve MLE by taking into account the prior information should become more efficient. They also suggest that for a given case, the sieve MLE should be more efficient than the two-step estimator.

The sieve MLE of θ_o for each of the four cases was presented in Section 2. For clarity, we denote the sieve MLE in the four cases as $\hat{\theta}_I, \hat{\theta}_{II}, \hat{\theta}_{III}, \hat{\theta}_{IV}$ respectively. The infeasible MLE $\bar{\theta}_n$ is the same for all four cases and is defined as

$$\bar{\theta}_n = \arg\max_{\theta} \sum_{i=1}^{n} \log\{c(F_{o1}(X_{1i}), F_{o2}(X_{2i}); \theta)\}. \tag{22}$$

The two-step estimator for Case I was defined in (2), We denote it as $\tilde{\theta}_I$. For Case II, the two-step estimator under prior restrictions on marginal distributions $\widetilde{\theta}_{II}$ is defined as

$$\begin{aligned} \widetilde{\theta}_{II} &= \arg\max_{\theta} \sum_{i=1}^{n} \log\{c(\widetilde{F}(X_{1i}), \widetilde{F}(X_{2i}), \theta)\}, \\ \text{with} \quad \widetilde{F}(x) &= \frac{1}{2}\sum_{j=1}^{2} \widetilde{F}_{nj}(x), \quad \widetilde{F}_{nj}(x) = \frac{1}{n+1}\sum_{i=1}^{n} 1\{X_{ji} \le x\}. \end{aligned}$$

For Case III, the two-step estimator of θ_o under a parametric marginal $F_{o1}(x) = F_{o1}(x, \beta_o)$ is defined as

$$\begin{aligned} \widetilde{\theta}_{III} &= \arg\max_{\theta} \sum_{i=1}^{n} \log\{c(F_{o1}(X_{1i}, \widetilde{\beta}), \widetilde{F}_{n2}(X_{2i}), \theta)\}, \\ \text{with} \quad \widetilde{\beta} &= \arg\max_{\beta} \sum_{i=1}^{n} \log f_{o1}(X_{1i}, \beta), \end{aligned} \tag{23}$$

and for Case IV, the two-step estimator $\tilde{\theta}_{IV}$ is obtained from (23) by using $F_{o1}(X_{1i}, \beta_o)$ instead of $F_{o1}(X_{1i}, \tilde{\beta})$.

For marginal distributions, we used the plug-in sieve MLE $\widehat{F}_{nj}$ obtained in each case and the (rescaled) empirical distribution function $\widetilde{F}_{nj}$ and the modified estimator $\widetilde{F}(x)$ for DGP II in Case II.

The sieve MLEs were implemented by using the B-spline basis as follows. Let $\{B_\gamma(x-j)\}_{j=1}^{K}$ be the $\gamma-$th order B-spline basis. Then the marginal density functions f_{o1} and f_{o2} can be approximated by

$$f_k(x; a_k) = \frac{\left(\sum_{j=1}^{K} a_{jk} B_\gamma(x-j)\right)^2}{\int \left(\sum_{j=1}^{K} a_{jk} B_\gamma(x-j)\right)^2 dx},$$

where $k = 1, 2$. In the Monte Carlo experiment, we used the 3rd order B-splines, *i.e.*, $\gamma = 3$. We approximated 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 $b_1 = \max(z \le \min(X_{ji}) - s_{X_j} : z \text{ is integer})$, and $b_2 = \min(z \ge \max(X_{ji}) + s_{X_j} : z$ is integer). Then for B-splines of order γ, we need $K_n = b_2 - b_1 + 1 - \gamma$ sieve coefficients to 'cover' the interval $[b_1, b_2]$. To evaluate the integral that appears in the denominator we used a grid of equidistant points on $[b_1, b_2]$. The results reported in this paper correspond to grid size 0.005, but we also tried value 0.01, which gives very similar results. In each case, the sieve MLE is computed via penalization. We tried penalization factors of values 0.01, 0.001, and 0.0001 and found that the results are similar. The results reported use 0.001 as the penalization factor.

5.2 Monte Carlo Results

Results reported in this section are based on 100 simulations. For each estimator of θ_o, we computed its sample mean and sample mean squared error (MSE), as well as the sample mean of a consistent estimator of its asymptotic variance (Est.avar). The consistent estimators of the asymptotic variances for the sieve MLE are computed according to those described in Sections 3 and 4, with 8 number of cosine series terms. The consistent estimator of the asymptotic variance for the two-step estimator when all margins are unknown can be found in Genest, et al. (1995). In Appendix B, we provide some consistent estimators of the asymptotic variances of the modified two-step estimators under prior restrictions on marginal distributions; we also present a simple consistent estimator of the asymptotic variance of the infeasible MLE.

For each estimator of the marginal distributions, we computed its sample mean and sample mean squared error (MSE), as well as the sample mean of a consistent estimator of its asymptotic variance (Est.avar) at the 33th percentile and 66th percentile of the true distribution. In addition, we also computed the sample mean of the integrated MSE (IMSE) of each estimator of the marginal distributions.

Throughout the experiment, we considered two sample sizes $n = 400$ and $n = 800$. To save space, we will not report results for all cases corresponding to both sample sizes. Table 1 reports results for the estimation of the copula parameter θ_o for DGP I.

Table 1. Estimation of θ_o for DGP I (Case I)

	$\theta_o = 5, n = 400$			$\theta_o = 10, n = 400$		
Estimator	$\hat{\theta}_I$	$\bar{\theta}_n$	$\tilde{\theta}_I$	$\hat{\theta}_I$	$\bar{\theta}_n$	$\tilde{\theta}_I$
Mean	4.949	5.013	4.855	9.960	10.006	9.622
MSE	0.158	0.063	0.162	0.579	0.230	0.620
Est.avar	0.139	0.069	0.169	0.487	0.223	0.661
$\theta_o = 15$		$n = 400$			$n = 800$	
Mean	14.637	15.034	14.187	14.909	14.963	14.57
MSE	1.584	0.469	1.893	0.536	0.206	0.630
Est.avar	0.967	0.463	1.618	0.505	0.232	0.811

Results in Table 1 confirm the better performance of the sieve MLE over the two-step approach, although the MSE and the estimated asymptotic variance of the sieve MLE are closer to those of the two-step than the infeasible MLE, consistent with the theoretical finding that the sieve MLE is asymptotically efficient but not adaptive. As expected, both estimators perform better as the sample size n increases.

Table 2 reports results for the estimation of the marginal distributions for $\theta_o = 15$ and $n = 400, 800$. The sieve ML estimator of $F_{o1} = t_{[5]}$ at the 33th percentile of $t_{[5]}$ is 44% more efficient than the rescaled empirical cdf counterpart, and the sieve ML estimator of $F_{o2} = t_{[25]}$ is 38% more efficient; corresponding to 66th percentile efficiency gains are 65% and 86% for F_{o1} and F_{o2} respectively. In terms of the IMSE, the relative efficiency gain (computed as the ratio of the IMSE of the two-step estimator to that of the sieve MLE less 1) is 9 percent for $F_{o2} = t_{[25]}$ and 17 percent for $F_{o1} = t_{[5]}$.

Table 2. Point Estimates of Marginal Distributions for DGP I (Case I, $\theta_o = 15$)

x	$t_{[5],.33}$	$t_{[25],.33}$	$t_{[5],.66}$	$t_{[25],.66}$
$n = 400$				
Empirical Distribution $(\tilde{F}_{n1}, \tilde{F}_{n2})$				
Mean	0.3274	0.3347	0.6653	0.6573
MSE$\times 10^3$	0.6874	0.6499	0.7345	0.8055
Est. avar$\times 10^3$	0.5838	0.5798	0.6894	0.6750
Sieve ML estimates $(\hat{F}_{n1}, \hat{F}_{n2})$				
Mean	0.3292	0.3286	0.6608	0.6618
MSE$\times 10^3$	0.4773	0.4705	0.4440	0.4337
Est. avar$\times 10^3$	0.4064	0.4063	0.3974	0.3805
$n = 800$				
Empirical distribution $(\tilde{F}_{n1}, \tilde{F}_{n2})$				
Mean	0.3333	0.3406	0.6676	0.6606
MSE$\times 10^3$	0.2874	0.3489	0.2195	0.2726
Est. avar$\times 10^3$	0.2778	0.2807	0.2775	0.2803
Sieve ML estimates $(\hat{F}_{n1}, \hat{F}_{n2})$				
Mean	0.3361	0.3350	0.6656	0.6671
MSE$\times 10^3$	0.2287	0.2247	0.1766	0.1772
Est. avar$\times 10^3$	0.2057	0.2059	0.1875	0.1860
IMSE$\times 10^3$				
	$n = 400$		$n = 800$	
	F_{o1}	F_{o2}	F_{o1}	F_{o2}
Empirical distribution	1.9375	1.6424	0.7784	0.6695
Sieve ML distribution	1.7734	1.4078	0.6368	0.5481

To examine the further efficiency gain of sieve MLE from using prior information on the marginal distributions, we report in Tables 3 and 4 results for DGP II with $\theta_o = 15$, $F_{o1} = F_{o2} = t_{[5]}$, and $n = 400, 800$. For comparison purposes, we estimated $(\theta_o, F_{o1}, F_{o2})$ with and without using the prior information.

Table 3. Estimation of $\theta_o = 15$ for DGP II (Case I, Case II)

	$n = 400$			$n = 800$		
Estimator	$\hat{\theta}_I$	$\bar{\theta}_n$	$\tilde{\theta}_I$	$\hat{\theta}_I$	$\bar{\theta}_n$	$\tilde{\theta}_I$
Mean	15.271	15.058	14.391	15.189	15.018	14.575
MSE	1.087	0.392	1.300	0.570	0.147	0.707
Est.avar	1.116	0.475	1.637	0.532	0.232	0.776
Estimator	$\hat{\theta}_{II}$	$\bar{\theta}_n$	$\tilde{\theta}_{II}$	$\hat{\theta}_{II}$	$\bar{\theta}_n$	$\tilde{\theta}_{II}$
Mean	14.964	15.058	13.605	14.976	15.018	14.139
MSE	0.926	0.392	2.628	0.538	0.147	1.183
Est.avar	1.068	0.475	1.520	0.518	0.232	0.746

Comparing the results for $\hat{\theta}_I$ and $\hat{\theta}_{II}$, Table 3 reveals better performance of $\hat{\theta}_{II}$ than $\hat{\theta}_I$ in terms of all three measures. Surprisingly, the performance of the modified two-step $\tilde{\theta}_{II}$ is worse than that of the unmodified two-step $\tilde{\theta}_I$. The improved performance of sieve MLE of the marginal distribution over the empirical distribution is also evident from Table 4 below.

Table 4. Pointwise Estimates of the Marginal Distribution for DGP II (Case II, $\theta_o = 15, n = 400$)

Estimator	Empirical distribution ($\tilde{F}$)		Sieve ML estimates ($\hat{F}$)	
x	$t_{[5],.33}$	$t_{[5],.67}$	$t_{[5],.33}$	$t_{[5],.67}$
Mean	0.3295	0.6648	0.3320	0.6646
MSE$\times 10^3$	0.6323	0.5950	0.4371	0.3166
Est. avar$\times 10^3$	0.5526	0.5566	0.3901	0.3589
IMSE$\times 10^3$	1.4830		1.1605	

The last two tables report estimation results for DGP II, but under Case III and Case IV respectively.

Table 5. Estimation of θ_o for DGP II (Case III, Case IV)

Estimator	$\hat{\theta}_{III}$	$\bar{\theta}_n$	$\tilde{\theta}_{III}$	$\tilde{\theta}_I$	$\hat{\theta}_{IV}$	$\tilde{\theta}_{IV}$
$\theta_o = 5, n = 400$						
Mean	5.019	4.999	4.665	4.917	5.046	4.630
MSE	0.088	0.065	0.232	0.175	0.071	0.248
Est.avar	0.074	0.070	0.106	0.173	0.074	0.105
$\theta_o = 10, n = 400$						
Mean	9.956	9.989	8.481	9.689	10.058	8.373
MSE	0.310	0.206	2.982	0.626	0.238	3.418
Est.avar	0.232	0.224	0.402	0.672	0.235	0.396
$\theta_o = 15, n = 400$						
Mean	14.986	15.034	11.237	14.315	15.121	11.139
MSE	0.675	0.469	16.938	1.683	0.496	18.131
Est.avar	0.477	0.464	0.905	1.611	0.481	1.034
$\theta_o = 15, n = 800$						
Mean	14.961	15.017	12.536	14.672	15.073	12.297
MSE	0.275	0.234	7.280	0.709	0.239	9.037
Est.avar	0.239	0.235	0.551	0.780	0.242	0.465

Several interesting observations emerge from Table 5: i) the sieve MLE $\hat{\theta}_{III}$ under the parametric assumption on F_{o1} performs very similarly to the sieve MLE $\hat{\theta}_{IV}$ under the assumption that F_{o1} is completely known; ii) the performance of the sieve MLE $\hat{\theta}_{IV}$ ($\hat{\theta}_{III}$) is very close to that of the infeasible MLE $\bar{\theta}_n$; iii) both modified two-step estimators $\tilde{\theta}_{III}$ and $\tilde{\theta}_{IV}$ are much worse than the unmodified two-step estimator $\tilde{\theta}_I$ which we found puzzling. We also computed the values of the semiparametric efficiency bound for θ_o derived in Bickel, et al. (1993) for the case with one completely known marginal (Case IV). They are 0.069, 0.222, 0.463, 0.231 corresponding to (θ_o, n) =(5, 400), (10, 400), (15, 400), (15, 800) respectively. They are clearly very close to the estimated asymptotic variances of $\hat{\theta}_{IV}$ and $\hat{\theta}_{III}$, reconfirming the efficiency of the proposed sieve MLE procedure and its relevance in finite samples.

Table 6 below reveals a similar performance of the sieve MLE of the unknown marginal distribution F_{o2} to that of θ_o.

Table 6. IMSE ($\times 10^3$) of Estimators of F_{o2} for DGP II (Case III, Case IV)

	$n = 400$					
	Case III			Case IV		
θ_o	5	10	15	5	10	15
Empirical distribution	1.7673	1.7512	1.8490	1.7673	1.7512	1.8490
Sieve ML distribution	0.6469	0.5076	0.3983	0.6176	0.4701	0.3441

To summarize, we find: i) regardless of the prior information on marginal cdfs, the sieve MLE of θ_o has very small bias in finite samples; ii) when all the marginal cdfs are different and unknown, the relative improvement of sieve MLE $\widehat{\theta}_I$ over that of the two-step estimator $\tilde{\theta}_I$ is not very big; iii) incorporating prior information on the marginal distributions improves the performance of sieve MLE $\widehat{\theta}_j$ (j=II, III, IV) in terms of both finite sample MSE and the asymptotic variance estimate. Moreover, when one marginal cdf is known or of a parametric form, the sieve MLE $\widehat{\theta}_{III}$ or $\hat{\theta}_{IV}$ performs very well, almost as well as the infeasible MLE $\bar{\theta}_n$ and is much better than the corresponding two-step estimators; iv) incorporating prior information on marginal distributions seems to worsen the finite sample performance of the corresponding two-step estimator; v) as the amount of dependence increases, all three estimators of θ_o get slightly worse in terms of the finite sample MSEs and asymptotic variance estimates.

For the estimation of the marginal distributions, we find: i) incorporating prior information improves the finite sample performance of the sieve MLE; ii) as the amount of dependence increases, the efficiency gain of the sieve MLE over the rescaled empirical cdf estimate increases.

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 $j = 1, ..., m$,

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

and **(3)**

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

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

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

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:

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

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 Chebychev inequality, i.i.d. data, and $||\Pi_n \upsilon^* - \upsilon^*|| = o(1)$ which is satisfied given Assumption 4(2). For **(A1.2)** we notice

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

for some $\widetilde{\alpha} \in \mathcal{A}_n$ in between α, α_o. It is easy to check that for any $\upsilon = (\upsilon_\theta, \upsilon_1, ..., \upsilon_m)' \in \mathbf{V}$, and $\widetilde{\alpha} \in \mathcal{A}_n$ with $||\widetilde{\alpha} - \alpha_o|| = O(\delta_n)$ we have

$$\begin{aligned}
& E_o\left(\frac{\partial^2 \ell(\widetilde{\alpha}, Z)}{\partial\alpha\partial\alpha'}[\upsilon, \upsilon] - \frac{\partial^2 \ell(\alpha_o, Z)}{\partial\alpha\partial\alpha'}[\upsilon, \upsilon]\right) \\
= \ & \upsilon_\theta' E_o\left(\frac{\partial^2 \log c(\widetilde{\alpha})}{\partial\theta\partial\theta'} - \frac{\partial^2 \log c(\alpha_o)}{\partial\theta\partial\theta'}\right)\upsilon_\theta \\
& + 2\upsilon_\theta' \sum_{j=1}^m E_o\left(\left\{\frac{\partial^2 \log c(\widetilde{\alpha})}{\partial\theta\partial u_j} - \frac{\partial^2 \log c(\alpha_o)}{\partial\theta\partial u_j}\right\}\int^{X_j} \upsilon_j(x)dx\right) \\
& + \sum_{i=1}^m \sum_{j=1}^m E_o\left(\left\{\frac{\partial^2 \log c(\widetilde{\alpha})}{\partial u_i\partial u_j} - \frac{\partial^2 \log c(\alpha_o)}{\partial u_i\partial u_j}\right\}\int^{X_j} \upsilon_j(x)dx \int^{X_i} \upsilon_i(x)dx\right) \\
& - \sum_{j=1}^m E_o\left(\left[\frac{\upsilon_j(X_j)}{\widetilde{f}_j(X_j)}\right]^2 - \left[\frac{\upsilon_j(X_j)}{f_{oj}(X_j)}\right]^2\right).
\end{aligned}$$

Under Assumption 5, we have

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

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

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

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

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

(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 (16) 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 for the Infeasible MLE and the Restricted two-step Estimators

The **infeasible MLE** $\bar{\theta}_n$ given in (22) satisfies $\sqrt{n}(\bar{\theta}_n - \theta_o) \to \mathcal{N}\left(0, [\mathcal{I}(\theta_o)]^{-1}\right)$ where $\mathcal{I}(\theta_o) = E[-\frac{\partial^2}{\partial\theta^2}\log\{c(F_{o1}(X_{1i}), ..., F_{om}(X_{mi}), \theta_o)\}]$. Hence the asymptotic variance of $\bar{\theta}_n$ can be consistently estimated by

$$\widehat{avar(\bar{\theta}_n)} = \frac{1}{n\widehat{\mathcal{I}}(\bar{\theta}_n)} = \left[-\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\log\{c(F_{o1}(X_{1i}), ..., F_{om}(X_{mi}), \bar{\theta}_n)\}\right]^{-1}.$$

Two-step estimator with equal but unknown margins: When $m = 2$ and $F_{o1} = F_{o2} = F_o$, the modified two-step estimator $\widetilde{\theta}_{II}$ of θ_o satisfies

$$\sqrt{n}(\widetilde{\theta}_{II} - \theta_o) \to_d \mathcal{N}\left(0, \frac{1}{\mathcal{I}(\theta_o)} + \frac{var\{W_1(X_1) + W_2(X_2)\}}{[\mathcal{I}(\theta_o)]^2}\right)$$

where

$$\mathcal{I}(\theta_o) = E\left(-\frac{\partial^2}{\partial\theta^2}\log(c(F_o(X_{1i}), F_o(X_{2i}), \theta_o))\right),$$

and for $k = 1, 2$,

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

Using sample data we can estimate $\mathcal{I}(\theta_o)$ by

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

and $W_k(X_{ki})$ by

$$\widetilde{W}_k(X_{ki}) = \frac{-1}{n}\sum_{j:\widetilde{F}(X_{kj})\geqslant\widetilde{F}(X_{ki})}\frac{d\log(c(\widetilde{F}(X_{1j}), \widetilde{F}(X_{2j}), \widetilde{\theta}_{II}))}{d\theta}\frac{d\log(c(\widetilde{F}(X_{1j}), \widetilde{F}(X_{2j}), \widetilde{\theta}_{II}))}{du_k}.$$

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

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

Two-step estimator with a parametric margin: When $m = 2$ and $F_{o1}(\cdot) = F_{o1}(\cdot, \beta_o)$ is known up to unknown parameter $\beta_o \in int()$, the modified two-step estimator $\widetilde{\theta}_{III}$ of θ_o satisfies

$$\sqrt{n}(\widetilde{\theta}_{III} - \theta_o) \rightarrow_d \mathcal{N}\left(0, \frac{1}{\mathcal{I}(\theta_o)} + \frac{var(W_1(X_{1i}, \beta_o) + W_2(X_{2i}))}{[\mathcal{I}(\theta_o)]^2}\right)$$

where

$$\mathcal{I}(\theta_o) = E\left(-\frac{\partial^2}{\partial\theta^2}\log(c(F_{o1}(X_{1i}, \beta_o), F_{o2}(X_{2i}), \theta_o))\right),$$

$$W_2(X_{2i}) = -\int I(F_{o2}(X_{2i}) \leqslant u_2)\frac{d\log(c(u_1, u_2, \theta_o))}{d\theta}\frac{d\log(c(u_1, u_2, \theta_o))}{du_2}c(u_1, u_2, \theta_o)du_1du_2,$$

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

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

$$\widetilde{\sigma}^2 = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\log(c(\widetilde{F}_{o1}(X_{1i}), \widetilde{F}_{n2}(X_{2i}), \widetilde{\theta}_{III})), \tag{24}$$

and $\widetilde{W}_2(X_{2i}) =$

$$\frac{-1}{n}\sum_{j:\widetilde{F}_{n2}(X_{2j})\geqslant\widetilde{F}_{n2}(X_{2i})}\frac{d\log c(\widetilde{F}_{o1}(X_{1j}), \widetilde{F}_{n2}(X_{2j}), \widetilde{\theta}_{III})}{d\theta}\frac{d\log c(\widetilde{F}_{o1}(X_{1j}), \widetilde{F}_{n2}(X_{2j}), \widetilde{\theta}_{III})}{du_2}, \tag{25}$$

and $\widetilde{W}_{o1}(X_{1i}) =$

$$\left[\frac{-1}{n}\sum_{j=1}^{n}\frac{d\log c(\widetilde{F}_{o1}(X_{1j}), \widetilde{F}_{n2}(X_{2j}), \widetilde{\theta}_{III})}{d\theta}\frac{d\log c(\widetilde{F}_{o1}(X_{1j}), \widetilde{F}_{n2}(X_{2j}), \widetilde{\theta}_{III})}{du_1}\frac{dF_{o1}(X_{1j}, \widetilde{\beta})}{d\beta}\right]$$

$$\times\left(\frac{-1}{n}\sum_{j=1}^{n}\frac{\partial^2\log f_{o1}(X_{1j}, \widetilde{\beta})}{\partial\beta^2}\right)^{-1}\frac{d\log f_{o1}(X_{1i}, \widetilde{\beta})}{d\beta}.$$

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

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

Two-step estimator with a known margin: When $m = 2$ and $F_{o1}(\cdot) = F_{o1}(\cdot, \beta_o)$ is known with known β_o, the modified two-step estimator $\widetilde{\theta}_{IV}$ of θ_o satisfies

$$\sqrt{n}(\widetilde{\theta}_{IV} - \theta_o) \rightarrow_d \mathcal{N}\left(0, \frac{1}{\mathcal{I}(\theta_o)} + \frac{var(W_2(X_2))}{[\mathcal{I}(\theta_o)]^2}\right),$$

and a consistent estimator of the asymptotic variance of $\widetilde{\theta}_{IV}$ is given by

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

where $\widetilde{\sigma}^2$ and $\widetilde{W}_2(X_{2i})$ are given in (24) and (25) except we replace $F_{o1}(\cdot, \widetilde{\beta})$ by $F_{o1}(\cdot, \beta_o)$.

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 (2002). "Estimation of copula-based semiparametric time series models." Working paper, New York University and Vanderbilt University.

[5] Chen, X. and Y. Fan (2003). "Specification and estimation of a new class of semiparametric copula-based multivariate dynamic models." Manuscript, New York University and Vanderbilt University.

[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] Horowitz, J. (1998). *Semiparametric Methods in Econometrics*, New York Springer-Verlag.

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

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

[25] Lee, L. (1982). "Some approaches to the correction of selectivity bias." *Review of Economic Studies* 49, 355-372.

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

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

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

[29] Newey, W.K. (1997). "Convergence rates and asymptotic normality for series estimators." *Journal of Econometrics* 79, 147-168.

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

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

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

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

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

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

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

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

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

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

[40] Sklar, A. (1959). "Functions de r'epartition 'a n dimensionset leurs marges." *Publ. Inst. Statis. Univ. Paris* 8, 229-231.

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

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

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

[44] 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.