Efficient Estimation of Semiparametric Multivariate Copula Models
Author(s): Xiaohong Chen, Yanqin Fan and Viktor Tsyrennikov
Source: *Journal of the American Statistical Association*, Vol. 101, No. 475 (Sep., 2006), pp. 1228-1240
Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association
Stable URL: http://www.jstor.org/stable/27590797
Accessed: 07-04-2015 04:46 UTC

Efficient Estimation of Semiparametric Multivariate Copula Models

Xiaohong CHEN, Yanqin FAN, and Viktor TSYRENNIKOV

We propose a sieve maximum likelihood estimation 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 its flexibility in separately modeling the dependence structure and the marginal behaviors of a multivariate random variable, and its circumvention of the "curse of dimensionality" associated with purely nonparametric multivariate distributions. We show that the plug-in sieve maximum likelihood estimators (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 maximum likelihood estimation procedure to achieve further efficiency gains. Two such cases are studied: (a) the marginal distributions are equal but otherwise unspecified, and (b) some but not all marginal distributions are parametric. Monte Carlo studies indicate that the sieve MLEs perform well in finite samples, especially when prior information on the marginal distributions is incorporated.

KEY WORDS: Copula dependence parameter; Efficiency bound; Marginal distribution; Prior information; Sieve maximum likelihood.

1. INTRODUCTION

Let $\{Z_i \equiv (X_{1i}, \dots, X_{mi})'\}_{i=1}^n$ be a random sample from the distribution $H_o(x_1, \dots, x_m)$ of $Z \equiv (X_1, \dots, X_m)'$ in $\mathcal{X}_1 \times \cdots \times \mathcal{X}_m \subseteq \mathcal{R}^m$, $m \geq 2$. Assume that H_o is absolutely continuous with respect to the Lebesgue measure on $\mathcal{R}^m$ and let $h_o(x_1, \dots, x_m)$ be the probability density function (pdf) of Z. Clearly, estimation of H_o or h_o is one of the most important statistical problems. Because of 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 because of 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 cumulative distribution function (cdf) of X_j, $j = 1, \dots, m$. The characterization theorem of Sklar (1959) implies that there exists a unique copula function, $C_o(\cdot, \dots, \cdot): [0, 1]^m \to [0, 1]$, such that $H_o(x_1, \dots, x_m) \equiv C_o(F_{o1}(x_1), \dots, F_{om}(x_m))$ for all $(x_1, \dots, x_m) \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_m$. Suppose that the functional form of the copula $C_o(u_1, \dots, u_m)$ is known apart from a finite-dimensional parameter θ_o; that is, for any $(u_1, \dots, u_m) \in [0, 1]^m$, we have $C_o(u_1, \dots, u_m) = C(u_1, \dots, u_m; \theta_o)$, where $\{C(u_1, \dots, u_m; \theta): \theta \in \Theta\}$ is a parametric family of copula functions. Then the multivariate distribution H_o is of a semiparametric form,

$$H_o(x_1, \dots, x_m) = C\big(F_{o1}(x_1), \dots, F_{om}(x_m); \theta_o\big), \qquad (1)$$

with unknown finite-dimensional parameter θ_o and infinite-dimensional parameters F_{oj}, $j = 1, \dots, m$. Commonly used parametric copulas include the Gaussian copula, the Student t copula, and Clayton, Frank, and Gumbel copulas (see Joe 1997 and Nelsen 1999 for properties of many existing parametric copulas). Let f_{oj}, $j = 1, \dots, m$, and $c(u_1, \dots, u_m; \theta_o)$ denote the pdf's associated with F_{oj}, $j = 1, \dots, m$, and $C(u_1, \dots, u_m; \theta_o)$. Then for any $(x_1, \dots, x_m) \in \mathcal{X}_1 \times \cdots \times \mathcal{X}_m$, the pdf h_o of H_o given by (1) has the representation $h_o(x_1, \dots, x_m) = c(F_{o1}(x_1), \dots, F_{om}(x_m); \theta_o) \prod_{j=1}^m f_{oj}(x_j)$. 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; for any m, the joint density $h_o(x_1, \dots, x_m)$ 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}, $j = 1, \dots, m$, capture the marginal behavior of the univariate random variables X_j, $j = 1, \dots, m$, and the finite-dimensional parameter θ_o or, equivalently, the parametric copula $C(u_1, \dots, u_m, \theta_o)$ characterizes the dependence structure between $X_1, \dots, X_m$ that is invariant to any increasing transformations of the univariate random variables X_j, $j = 1, \dots, m$. Given the existence of a large number of parametric copulas and univariate distributions, this class of semiparametric multivariate distributions is able to jointly model any type of dependence with any types of marginal behaviors and has proven useful in diverse fields. Specific applications include those in finance and insurance (e.g., Frees and Valdez 1998; Embrechts, McNeil, and Straumann 2002), survival analysis (e.g., Joe 1997; Nelsen 1999; Oakes 1989), and econometrics (e.g., Lee 1983; Heckman and Honoré 1989; Granger, Teräsvirta, and Patton 2006; Patton 2004).

To estimate the multivariate distribution $H_o(x_1, \dots, x_m) \equiv C(F_{o1}(x_1), \dots, F_{om}(x_m); \theta_o)$, one must estimate both the copula parameter θ_o and the marginal cdf's F_{oj}, $j = 1, \dots, m$. Currently, the most popular estimator of F_{oj} is the empirical cdf, $F_{nj}(x_j) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{X_{ji} \leq x_j\}$ for $j = 1, \dots, m$. The most

Xiaohong Chen is Professor, Department of Economics, New York University, New York, NY 10003 (E-mail: *xiaohong.chen@nyu.edu*). Yanqin Fan is Professor, Department of Economics, Vanderbilt University, Nashville, TN 37235 (E-mail: *yanqin.fan@vanderbilt.edu*). Viktor Tsyrennikov is a Graduate Student, Department of Economics, New York University, New York, NY 10003 (E-mail: *viktor.tsyrennikov@nyu.edu*). The authors thank the coeditor, the associate editor, and three anonymous referees for very useful comments that greatly improved the article. The authors also thank Demian Pouzo for excellent assistance in simulation studies, and Oliver Linton, Wolfgang Haerdle and seminar participants at London School of Economics, and Humbolt University; and participants at the 2003 North American Econometric Society Summer Meetings in Chicago, the 2004 DeMoSTAFI conference in Quebec City, and the 2004 Semiparametrics in Rio for helpful comments. Chen acknowledges financial supports from the ESRC, the National Science Foundation (NSF), the CASE Center at Humboldt University, and the C. V. Starr Center at New York University. Fan acknowledges financial support from the NSF.

Journal of the American Statistical Association
September 2006, Vol. 101, No. 475, Theory and Methods
DOI 10.1198/016214506000000311

widely used estimator of θ_o is the two-step estimator $\widetilde{\theta}_n$ proposed by Oakes (1994) and Genest, Ghoudi, and Rivest (1995),

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

where $\widetilde{F}_{nj}(x_j) = \frac{1}{n+1}\sum_{i=1}^{n}\mathbb{1}\{X_{ji}\le x_j\}$ is the rescaled empirical cdf estimator of F_{oj}, $j=1,\ldots,m$. Genest et al. (1995) established the root-n consistency and asymptotic normality of $\widetilde{\theta}_n$. Shih and Louis (1995) independently proposed the two-step estimator for iid data with random censoring. The two-step estimator and its large-sample properties have been extended to time series setting by Chen and Fan (2005, 2006). Many previous articles have proposed 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; Genest 1987).

In many applications, efficient estimation of the entire multivariate distribution $H_o(x_1,\ldots,x_m)$ is desirable, which requires efficient estimation of both the marginal cdf's 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 cdf's F_{nj}, $j=1,\ldots,m$, are generally inefficient. Intuitively, one could obtain more efficient estimates of F_{oj}, $j=1,\ldots,m$, by using 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_o 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 is efficient for the Gaussian copula parameter when marginal cdf's are unknown (Klaassen and Wellner 1997). Unfortunately, even for the bivariate Gaussian copula model with unknown margins, currently there are no efficient estimates of univariate marginal cdf's (see Klaassen and Wellner 1997). For semiparametric bivariate survival Clayton copula models, Maguluri (1993) provided some efficiency score calculation for θ_o and conjectured that her proposed estimator might be efficient. For general bivariate semiparametric copula models, Bickel, Klaassen, Ritov, and Wellner (1993, chap. 4.7) presented 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, chap. 6.7) provided some efficiency bound calculations for the unknown margin, but again no efficient estimators. To the best of our knowledge (see Klaassen and Wellner 1997; Genest and Werker 2002), there are no published works providing efficient estimates of θ_o and F_{oj}, $j=1,\ldots,m$, for general multivariate semiparametric copula models.

In this article we propose a general sieve maximum likelihood (ML) estimation procedure for all of the unknown parameters in a semiparametric multivariate copula model (1). This procedure approximates the infinite-dimensional unknown marginal densities f_{oj}, $j=1,\ldots,m$, 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 maximum likelihood estimators (MLEs) of the marginal cdf's use all of the parametric dependence information, and our sieve MLE of the copula parameter effectively solves an approximate efficient score equation for θ_o (where the approximation error becomes negligible as sample size grows sufficiently large), intuition suggests that these estimators should be efficient. By applying the general theory of Shen (1997), we show that our plug-in sieve MLEs of all smooth functionals, including the unknown marginal cdf's and the copula parameter, are indeed semiparametrically efficient. Because 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 cdf's to produce more efficient estimates. Examples of such restrictions include equal but unknown marginal cdf's and known parametric forms of some (but not all) marginal cdf's. Results from an extensive simulation study for several copula families and marginal cdf's in both bivariate and trivariate models confirm the efficacy of the sieve MLE.

Although we establish that the sieve MLEs of the copula parameter and marginal cdf's achieve their efficiency bounds, there are no closed-form expressions for the efficiency bounds of copula parameter and marginal cdf's 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 possible only for a few special copula models. Nevertheless, for general semiparametric multivariate copula models with or without prior information on marginal cdf's, we can provide simple consistent estimates of the asymptotic variances of the sieve MLEs of the copula parameter and of the unknown marginal cdf's. 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 through 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 article is organized as follows. In Section 2 we introduce the sieve MLEs of the copula parameter and the unknown marginal cdf's in models with or without restrictions on the marginal cdf's. In Section 3 we show that for semiparametric multivariate copula models with unknown marginal cdf's, the plug-in sieve MLEs of all smooth functionals are root-n normal and semiparametrically efficient. We then use these results to deliver the root-n asymptotic normality and efficiency of the sieve MLEs of the copula parameter and the marginal cdf's. We also provide simple consistent estimators of the asymptotic variances of these sieve MLEs. In Section 4 we extend our results in Section 3 to models with equal but unknown margins and models with some parametric margins. In Section 5 we summarize 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. We also describe 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 cdf's to the empirical cdf's. The Appendix contains proofs.

2. THE SIEVE MAXIMUM LIKELIHOOD ESTIMATORS

We first introduce suitable sieve spaces for approximating an unknown univariate density function of certain smoothness. Based on this, we 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; Geman and Hwang 1982).

There exist many sieves for approximating a univariate pdf. In this article we focus on using linear sieves to directly approximate a square root density,

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

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). Suppose that 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 Jth 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 $\mathcal{X}_j$ that 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 $\mathrm{Spl}(s, K_n)$ with $s > [r]$ (the largest integer part of r). The spline sieve $\mathrm{Spl}(s, K_n)$ is a linear space of dimension $(K_n + s + 1)$ consisting of spline functions of degree s with almost equally spaced knots $t_1, \dots, t_{K_n}$ on $[0, 1]$. Let $t_0, t_1, \dots, t_{K_n}, t_{K_n+1}$ be real numbers with $0 = t_0 < t_1 < \cdots < t_{K_n} < t_{K_n+1} = 1$ and $\max_{0 \leq k \leq K_n}(t_{k+1} - t_k) \leq \text{const} \min_{0 \leq k \leq K_n}(t_{k+1} - t_k)$. Partition $[0, 1]$ into $K_n + 1$ subintervals $I_k = [t_k, t_{k+1})$, $k = 0, \dots, K_n - 1$, and $I_{K_n} = [t_{K_n}, t_{K_n+1}]$. A function on $[0, 1]$ is a spline of degree s with knots $t_1, \dots, t_{K_n}$ if it is (a) a polynomial of degree s or less on each interval I_k, $k = 0, \dots, K_n$, and (b) $(s - 1)$-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 $\mathrm{Pol}(K_n) = \{\sum_{k=0}^{K_n} a_k x^k, x \in [0, 1] : a_k \in \mathcal{R}\}$, the trigonometric sieve $\mathrm{TriPol}(K_n) = \{a_0 + \sum_{k=1}^{K_n}[a_k \cos(k\pi x) + b_k \sin(k\pi x)], x \in [0, 1] : a_k, b_k \in \mathcal{R}\}$, and the cosine series $\mathrm{CosPol}(K_n) = \{a_0 + \sum_{k=1}^{K_n} a_k \cos(k\pi x), x \in [0, 1] : a_k \in \mathcal{R}\}$.

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

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

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

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

as was done by Gallant and Nychka (1987), or if the density f_{oj} has polynomial fat tails, then we can use the spline wavelet sieve to approximate it,

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

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

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

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

2.2 Sieve Maximum Likelihood Estimators

To simplify presentation, we let $\ell(\alpha, Z_i)$ denote the contribution of the ith observation to the log-likelihood function and let $\hat{\alpha}_n$ denote the sieve MLE for all the cases (with or without prior information on marginal distributions) considered in this article. We first consider the case without any prior information on the marginal distributions. Let $\alpha = (\theta', f_1, \dots, f_m)'$ and let $\alpha_o = (\theta_o', f_{o1}, \dots, f_{om})' \in \Theta \times \prod_{j=1}^m \mathcal{F}_j = \mathcal{A}$ denote the true but unknown parameter value. Let

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

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

$$\hat{\alpha}_n = \arg\max_{\alpha \in \mathcal{A}_n} \sum_{i=1}^n \ell(\alpha, Z_i), \quad (8)$$

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 $\hat{F}_{nj}(x_j) = \int \mathbb{1}(y \le x_j)\hat{f}_{nj}(y)\,dy$, $j = 1, \ldots, m$.

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

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

where for $j = 1, \ldots, m$, $f_j(X_{ji}; a_{jn})$ 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, $\operatorname{Pen}(a_{jn})$, typically corresponds to the L_2-norm of either the first derivative or the second derivative of $f_j^{1/2}(\cdot; a_{jn})$, and the λ_{jn}'s are penalization factors. In our simulation study, we chose the penalization factors through cross-validation. In principle, we could use any model selection methodology, including cross-validation (see, e.g., Coppejans and Gallant 2002), covariance penalty (see, e.g., Shen and Ye 2002), and 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.

In the first case, 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 $j = 1, \ldots, m$, and 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 $\ell(\alpha, Z_i) = \log\{c(F(X_{1i}), \ldots, F(X_{mi}); \theta)\prod_{j=1}^{m} f(X_{ji})\}$, in which $F(X_{ji}) = \int_{\mathcal{X}} \mathbb{1}(x \le X_{ji}) f(x)\,dx$, $j = 1, \ldots, m$. Then the sieve MLE, $\hat{\alpha}_n = (\hat{\theta}_n', \hat{f}_n)' \in \Theta \times \mathcal{F}_{n1} = \mathcal{A}_n$, 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) considered 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 that 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, that is, $F_{o1}(x_1) = F_{o1}(x_1, \beta_o)$ for some $\beta_o \in \mathcal{B}$. The marginal distributions $F_{o2}, \ldots, F_{om}$ are unspecified. Let $\alpha = (\theta', \beta', f_2, \ldots, f_m)'$ and let $\alpha_o = (\theta_o', \beta_o', f_{o2}, \ldots, f_{om})' \in \Theta \times \mathcal{B} \times \prod_{j=2}^{m} \mathcal{F}_j = \mathcal{A}$ denote the true but unknown parameter value. Let $\ell(\alpha, Z_i) = \log\{c(F_{o1}(X_{1i}, \beta), \ldots, F_{om}(X_{mi}); \theta) \times f_{o1}(X_{1i}, \beta)\prod_{j=2}^{m} f_j(X_{ji})\}$, in which $F_j(X_{ji}) = \int_{\mathcal{X}_j} \mathbb{1}(x \le X_{ji}) f_j(x)\,dx$, $j = 2, \ldots, m$. Then the sieve MLE, denoted as $\hat{\alpha}_n = (\hat{\theta}_n', \hat{\beta}_n', \hat{f}_{n2}, \ldots, \hat{f}_{nm})' \in \Theta \times \mathcal{B} \times \prod_{j=2}^{m} \mathcal{F}_{nj} = \mathcal{A}_n$, 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 let $\rho(\hat{\alpha}_n)$ 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 $\rho(\hat{\alpha}_n)$ 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 $U_o \equiv (U_{o1}, \ldots, U_{om})' \equiv (F_{o1}(X_1), \ldots, F_{om}(X_m))'$ and let $u = (u_1, \ldots, u_m)'$ be an arbitrary value in $[0, 1]^m$. In addition, let $c(F_{o1}(X_1), \ldots, F_{om}(X_m); \theta_o) = c(U_o, \theta_o) = c(\alpha_o)$.

Assumption 1. (a) $\theta_o \in \operatorname{int}(\Theta)$, Θ a compact subset of $\mathcal{R}^{d_\theta}$; (b) for $j = 1, \ldots, m$, $\sqrt{f_{oj}} \in \Lambda^{r_j}(\mathcal{X}_j)$, $r_j > 1/2$; and (c) $\alpha_o = (\theta_o', f_{o1}, \ldots, f_{om})'$ 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: $\partial^2 \log c(u, \theta)/\partial\theta^2$, $\partial^2 \log c(u, \theta)/\partial u_j\,\partial\theta$, and $\partial^2 \log c(u, \theta)/\partial u_j\,\partial u_k$ for $j, k = 1, \ldots, m$.

Let $\mathbf{V}$ denote the linear span of $\mathcal{A} - \{\alpha_o\}$. Under Assumption 2, for any $v = (v_\theta', v_1, \ldots, v_m)' \in \mathbf{V}$, 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

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

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 $\overline{\mathbf{V}} = \{v = (v_\theta', v_1, \ldots, v_m)' \in \mathcal{R}^{d_\theta} \times \prod_{j=1}^{m} \overline{\mathbf{V}}_j : \|v\| < \infty\}$ 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 write

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

whenever the right-side limit is well defined and assume the following.

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

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

(b) 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

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

and

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

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

Assumption 4. (a) $\|\hat{\alpha}_n - \alpha_o\| = O_P(\delta_n)$ for a decreasing sequence δ_n satisfying $(\delta_n)^{\omega} = o(n^{-1/2})$, and (b) 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 and 6 stated in the Appendix hold. Then $\sqrt{n}(\rho(\hat{\alpha}_n) - \rho(\alpha_o)) \Rightarrow \mathcal{N}(0, \|\frac{\partial \rho(\alpha_o)}{\partial \alpha'}\|^2)$, and $\rho(\hat{\alpha}_n)$ is semiparametrically efficient.

Discussion of Assumptions. Assumptions 1 and 2 are standard assumptions. Assumption 3 is essentially the definition of a smooth functional. Assumption 4(a) is a requirement on the convergence rate of the sieve MLEs of unknown marginal densities $\hat{f}_{nj}$, $j = 1, \dots, m$. There are many results on convergence rates of general sieve estimates of a marginal density (see, e.g., Shen and Wong 1994; Wong and Shen 1995; 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(b) requires that the Riesz representer have some smoothness. Although Assumptions 3 and 4(b) are stated in terms of data $Z_i = (X_{1i}, \dots, X_{mi})'$ 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 a 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 $b_j(u_j) = v_j(F_{oj}^{-1}(u_j))/f_{oj}(F_{oj}^{-1}(u_j))$, and vice versa. Therefore, we can always rewrite $\frac{\partial \ell(\alpha_o, Z)}{\partial \alpha'}[v]$ as

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

and

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

Define

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

Then there is a one-to-one 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 $v^* = (v_\theta^{*\prime}, v_1^*, \dots, v_m^*)' \in \overline{\mathbf{V}}$ is uniquely determined by $b^* = (v_\theta^{*\prime}, b_1^*, \dots, b_m^*)' \in \overline{\mathbf{B}}$ (and vise versa) through the relation

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

Then Assumption 4(b) can be replaced by the following.

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

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

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

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

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

3.2 $\sqrt{n}$-Normality and Efficiency of $\hat{\theta}_n$

We take $\rho(\alpha) = \lambda'\theta$ for any arbitrarily fixed $\lambda \in \mathcal{R}^{d_\theta}$ with $0 < |\lambda| < \infty$. This satisfies Assumption 3(b) with $\frac{\partial\rho(\alpha_o)}{\partial\alpha'}[v] = \lambda' v_\theta$ and $\omega = \infty$. Assumption 3(a) 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

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

Note that

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

where

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

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

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

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

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

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

Assumption 3′. (a) $\partial \log c(U_o, \theta_o)/\partial\theta$ and $\partial \log c(U_o, \theta_o)/\partial u_j$, $j = 1, \ldots, m$, have finite second moments; (b) $\mathcal{I}(\theta_o) \equiv E_o[\frac{\partial \log c(U_o,\theta_o)}{\partial\theta} \frac{\partial \log c(U_o,\theta_o)}{\partial\theta'}]$ is finite and positive definite; (c) $\int \frac{\partial c(u,\theta_o)}{\partial u_j}\,du_{-j} = \frac{\partial}{\partial u_j}\int c(u,\theta_o)\,du_{-j} = 0$ for $(j,-j) = (1,\ldots,m)$ with $j \neq -j$; (d) $\int \frac{\partial^2 c(u,\theta_o)}{\partial u_j\,\partial\theta}\,du_{-j} = \frac{\partial^2}{\partial u_j\,\partial\theta} \times \int c(u,\theta_o)\,du_{-j} = 0$ for $(j,-j) = (1,\ldots,m)$ with $j \neq -j$; and (e) there exists a constant K such that

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

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

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

To make inferences on θ_o using the sieve MLE $\hat{\theta}_n$, 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 $\mathcal{I}_*(\hat{\theta}_n)$. Recently, Klaassen and Wellner (1997) derived a closed-form expression of $\mathcal{I}_*(\theta_o)$ for the bivariate Gaussian copula model with unknown margins. In general, however, there is no closed-form solution of $\mathcal{I}_*(\theta_o)$ for multivariate copula models with unknown margins. Hence direct plug-in estimation of $\mathcal{I}_*(\theta_o)$ is difficult. Instead, we propose a sieve estimator of $\mathcal{I}_*(\theta_o)$ based on its characterization in (15). Let $\hat{U}_i = (\hat{U}_{1i}, \ldots, \hat{U}_{mi})' = (\hat{F}_{n1}(X_{1i}), \ldots, \hat{F}_{nm}(X_{mi}))'$ for $i = 1, \ldots, n$. Let $\mathbf{A}_n$ be some sieve space, such as

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

and

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

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

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

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

Remark 2. Although there is no closed-form expression for $\mathcal{I}_*(\theta_o)$ in general, we could compute a simulated theoretical asymptotic variance expression, denoted by $\mathcal{I}_s(\theta_o)$,

$$\mathcal{I}_s(\theta_o) = \min_{\substack{g_j \in \mathbf{A}_S,\\ j=1,\ldots,m}} \frac{1}{S}\sum_{i=1}^{S}\Bigg\{\Bigg(\frac{\partial \log c(U_i,\theta_o)}{\partial\theta'} - \sum_{j=1}^{m}\bigg[\frac{\partial \log c(U_i,\theta_o)}{\partial u_j}\int_0^{U_{ji}} g_j(v)\,dv + g_j(U_{ji})\bigg]\Bigg)' \times \Bigg(\frac{\partial \log c(U_i,\theta_o)}{\partial\theta'} - \sum_{j=1}^{m}\bigg[\frac{\partial \log c(U_i,\theta_o)}{\partial u_j}\int_0^{U_{ji}} g_j(v)\,dv + g_j(U_{ji})\bigg]\Bigg)\Bigg\},$$

using a large (say, $S = 100{,}000$) simulated random sample, $\{U_i\}_{i=1}^S$, drawn from the true copula function $C(u,\theta_o)$. We use this in our simulation study where we know the value θ_o in a Monte Carlo design. Of course, in practice the true value θ_o is unknown, and instead we could compute a simulated estimator of the asymptotic variance, denoted by $\mathcal{I}_s(\hat\theta_n)$,

$$\mathcal{I}_s(\hat\theta_n) = \min_{\substack{g_j \in \mathbf{A}_S,\\ j=1,\ldots,m}} \frac{1}{S}\sum_{i=1}^{S}\Bigg\{\Bigg(\frac{\partial \log c(U_i^s,\hat\theta_n)}{\partial\theta'} - \sum_{j=1}^{m}\bigg[\frac{\partial \log c(U_i^s,\hat\theta_n)}{\partial u_j}\int_0^{U_{ji}^s} g_j(v)\,dv + g_j(U_{ji}^s)\bigg]\Bigg)' \times \Bigg(\frac{\partial \log c(U_i^s,\hat\theta_n)}{\partial\theta'} - \sum_{j=1}^{m}\bigg[\frac{\partial \log c(U_i^s,\hat\theta_n)}{\partial u_j}\int_0^{U_{ji}^s} g_j(v)\,dv + g_j(U_{ji}^s)\bigg]\Bigg)\Bigg\},$$

using a large (say, $S = 100{,}000$) simulated random sample, $\{U_i^s\}_{i=1}^S$, drawn from the copula function $C(u,\hat\theta_n)$, where $\hat\theta_n$ is the sieve MLE of θ_o. In our simulation study we computed both $\hat\sigma_\theta^2$ and $\mathcal{I}_s(\hat\theta_n)$ as consistent estimators of $\mathcal{I}_*(\theta_o)$, and their values are close to one another.

3.3 Sieve MLE of F_{oj}

For $j = 1,\ldots,m$, 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, $\rho(\hat\alpha) = \hat F_{nj}(x_j) = \int \mathbb{1}(y \le x_j)\hat f_{nj}(y)\,dy$, where $\hat f_{nj}$ is the sieve MLE from (8). Clearly, $\frac{\partial\rho(\alpha_o)}{\partial\alpha'}[v] = \int_{\mathcal{X}_j}\mathbb{1}(y\le x_j)v_j(y)\,dy$ for any $v = (v_\theta', v_1,\ldots,v_m)' \in \mathbf{V}$. It is easy to see that $\omega = \infty$ in Assumptions 3 and 4, and that

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

Hence the representer $v^* \in \overline{\mathbf{V}}$ should satisfy

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

and

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

Proposition 3. Let $v^* \in \overline{\mathbf{V}}$ 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 $j = 1,\ldots,m$, $\sqrt{n}(\hat F_{nj}(x_j) - F_{oj}(x_j)) \Rightarrow \mathcal{N}(0, \|v^*\|^2)$. Moreover, $\hat F_{nj}$ 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 $\|v^*\|^2$. Nevertheless, we can again consistently estimate $\|v^*\|^2$ by the sieve method. Let

$$\hat\sigma_{F_j}^2(x_j) = \max_{\substack{v_\theta\ne 0, b_k\in\mathbf{B}_n,\\ k=1,\ldots,m}}\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{\hat U_{ji}\le\hat F_{nj}(x_j)\}b_j(\hat U_{ji})\right|^2 \times\Bigg(\frac{1}{n}\sum_{i=1}^{n}\bigg[\frac{\partial\log c(\hat U_i,\hat\theta)}{\partial\theta'}v_\theta + \sum_{k=1}^{m}\bigg[\frac{\partial\log c(\hat U_i,\hat\theta)}{\partial u_k}\int_0^{\hat U_{ki}} b_k(u)\,du + b_k(\hat U_{ki})\bigg]\bigg]^2\Bigg)^{-1},$$

where $\hat U_i = (\hat F_{n1}(X_{1i}),\ldots,\hat F_{nm}(X_{mi}))'$ and $\mathbf{B}_n$ is given in (17).

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

Remark 3. (a) Although there is generally no closed-form expression for $\|v^*\|^2$, in simulation studies where the true value $(\theta_o, F_{o1},\ldots,F_{om})$ is known, we could compute a simulated theoretical asymptotic variance, denoted by $\|v_s\|^2$,

$$\|v_s\|^2 = \max_{\substack{v_\theta\ne 0, b_k\in\mathbf{B}_S,\\ k=1,\ldots,m}}\left|\frac{1}{S}\sum_{i=1}^{S}\mathbb{1}\{U_{ji}\le F_{oj}(x_j)\}b_j(U_{ji})\right|^2 \times\Bigg(\frac{1}{S}\sum_{i=1}^{S}\bigg[\frac{\partial\log c(U_i,\theta_o)}{\partial\theta'}v_\theta + \sum_{k=1}^{m}\bigg[\frac{\partial\log c(U_i,\theta_o)}{\partial u_k}\int_0^{U_{ki}} b_k(u)\,du + b_k(U_{ki})\bigg]\bigg]^2\Bigg)^{-1},$$

using a large (say, $S = 100{,}000$) simulated random sample, $\{U_i\}_{i=1}^S$, drawn from the true copula function $C(u,\theta_o)$. Of course, in practice, the true value $(\theta_o, F_{o1},\ldots,F_{om})$ is unknown, and instead we could compute a simulated estimator of the asymptotic variance, denoted by $\|\hat{\upsilon}_s\|^2$,

$$\|\hat{\upsilon}_s\|^2 = \max_{\substack{\upsilon_\theta \neq 0, b_k \in \mathbf{B}_S, \\ k=1,\dots,m}} \left| \frac{1}{S} \sum_{i=1}^{S} \mathbb{1}\{U_{ji}^s \leq \hat{F}_{nj}(x_j)\} b_j(U_{ji}^s) \right|^2 \times \left(\frac{1}{S} \sum_{i=1}^{S} \left[\frac{\partial \log c(U_i^s, \hat{\theta}_n)}{\partial \theta'} \upsilon_\theta + \sum_{k=1}^{m} \left[\frac{\partial \log c(U_i^s, \hat{\theta}_n)}{\partial u_k} \int_0^{U_{ki}^s} b_k(u)\, du + b_k(U_{ki}^s) \right] \right]^2 \right)^{-1},$$

using a large (say, $S = 100{,}000$) simulated random sample, $\{U_i^s\}_{i=1}^S$, drawn from the copula function $C(u, \hat{\theta}_n)$, where $(\hat{\theta}_n, \hat{F}_{n1}, \dots, \hat{F}_{nm})$ is the sieve MLE of $(\theta_o, F_{o1}, \dots, F_{om})$.

(b) In the special case of the independence copula $(c(u_1, \dots, u_m, \theta) = 1)$, we could solve (18) and (19) explicitly. Note that for the independence copula,

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

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

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

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

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

where its asymptotic variance coincides with that of the standard empirical cdf estimate $F_{nj}(x_j) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}\{X_{ji} \leq x_j\}$ 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 $\|\upsilon\|^2 = E_o\{\frac{\partial \ell(\alpha_o, Z)}{\partial \alpha'}[\upsilon]\}^2$ with

$$\frac{\partial \ell(\alpha_o, Z)}{\partial \alpha'}[\upsilon] = \frac{\partial \log c(U_o, \theta_o)}{\partial \theta'} \upsilon_\theta + \sum_{j=1}^{m} \left\{ \frac{\partial \log c(U_o, \theta_o)}{\partial u_j} \int^{X_j} \upsilon_1(x)\, dx + \frac{\upsilon_1(X_j)}{f_o(X_j)} \right\},$$

$U_o = (F_o(X_1), \dots, F_o(X_m))'$, and $\upsilon \in \overline{\mathbf{V}} = \{\upsilon = (\upsilon_\theta', \upsilon_1)' \in \mathcal{R}^{d_\theta} \times \overline{\mathbf{V}}_1 : \|\upsilon\| < \infty\}$, with $\overline{\mathbf{V}}_1$ as given in (10).

Proposition 5. Suppose that Assumptions 1, 2, 3′, and 4–6 hold and that $f_{oj} = f_o$ for $j = 1, \dots, m$. Then the following hold:

(a) $\hat{\theta}_n$ is semiparametrically efficient and $\sqrt{n}(\hat{\theta}_n - \theta_o) \Rightarrow \mathcal{N}(0, \mathcal{I}_*(\theta_o)^{-1})$, where

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

(b) For any fixed $x \in \mathcal{X}$, $\hat{F}_n(x) = \int \mathbb{1}(y \leq x) \hat{f}_n(y)\, dy$ is semiparametrically efficient and $\sqrt{n}(\hat{F}_n(x) - F_o(x)) \Rightarrow \mathcal{N}(0, \|\upsilon^*\|^2)$, where

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

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

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

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

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

(b) $\hat{\sigma}_F^2(x) = \|\upsilon^*\|^2 + o_p(1)$, where

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

in which $\hat{U}_i = (\hat{F}_n(X_{1i}), \dots, \hat{F}_n(X_{mi}))'$, $\mathbf{A}_n$ 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 $\|v\|^2 = E_o\{\partial \ell(\alpha_o, Z)/\partial \alpha'[v]\}^2$ with

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

$$\frac{\partial \ell(\alpha_o, Z)}{\partial \beta'} v_\beta = \left[\frac{\partial \log c(U_o, \theta_o)}{\partial u_1} \int^{X_1} \frac{\partial f_{o1}(x, \beta_o)}{\partial \beta'}\, dx + \frac{1}{f_{o1}(X_1, \beta_o)} \frac{\partial f_{o1}(X_1, \beta_o)}{\partial \beta'} \right] v_\beta,$$

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

Proposition 7. Suppose that Assumptions 1, 2, 3′, and 4–6 hold, $F_{o1}(\cdot) = F_{o1}(\cdot, \beta_o)$ for unknown $\beta_o \in \text{int}(\mathcal{B})$ and $E[\frac{\partial \log f_{o1}(X_1, \beta_o)}{\partial \beta} \frac{\partial \log f_{o1}(X_1, \beta_o)}{\partial \beta'}]$ is positive definite. Then the following hold:

(a) $\hat{\theta}_n$ is semiparametrically efficient and $\sqrt{n}(\hat{\theta}_n - \theta_o) \Rightarrow \mathcal{N}(0, \mathcal{I}_*(\theta_o)^{-1})$, where $\mathcal{I}_*(\theta_o) = E_o[\mathcal{S}_{\theta_o} \mathcal{S}_{\theta_o}']$ with $\mathcal{S}_{\theta_o}' = (\mathcal{S}_{\theta_{o1}}, \dots, \mathcal{S}_{\theta_{od_\theta}})$ and for $k = 1, \dots, d_\theta$,

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

solves the following optimization problem:

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

(b) For any fixed $x \in \mathcal{X}$ and for $j = 2, \dots, m$, $\hat{F}_{nj}(x) = \int \mathbb{1}(y \le x) \hat{f}_{nj}(y)\, dy$ is semiparametrically efficient and $\sqrt{n}(\hat{F}_{nj}(x) - F_{oj}(x)) \Rightarrow \mathcal{N}(0, \|\upsilon^*\|^2)$, where

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

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

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

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

(b) $\hat{\sigma}_{F_j}^2(x_j) = \|\upsilon^*\|^2 + o_p(1)$, where

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

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

Remark 4. 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 $\alpha_o = (\theta_o, \beta_o, f_{o2}, \ldots, f_{om})$. 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 $\sqrt{n}(\hat{\theta}_n - \theta_o)$ can be consistently estimated by $\{\hat{\sigma}_\theta^2\}^{-1}$, with

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

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

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

where $\hat{U}_i = (F_{o1}(X_{1i}), \hat{F}_{n2}(X_{2i}), \ldots, \hat{F}_{nm}(X_{mi}))'$, $\mathbf{A}_n$ 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 sieve MLEs of the copula parameter and the marginals, we conduct an extensive simulation study using models that exhibit a wide range of copula dependence structures and marginal behaviors. We consider three types of information on the marginals: (a) one margin is known or parametric; (b) all margins are equal but unknown; and (c) all margins are unequal and unknown. For comparison purposes, we include the ideal (or infeasible) MLE of the copula parameter in which all of the margins are assumed to be known, the two-step estimator of the copula parameter, and the empirical cdf's of the marginals. As a fair comparison, we also include the modified two-step estimator of the copula parameter in which the marginals are estimated using the prior information on marginals. All the simulation results are based on 500 Monte Carlo replications of estimates using random samples $\{Z_i \equiv (X_{1i}, \ldots, X_{mi})'\}_{i=1}^n$ with $n = 400$. Due to the lack of space, we provide only a brief summary of the Monte Carlo designs and the simulation findings here, and refer interested readers to the working paper version (Chen, Fan, and Tsyrennikov 2005) for details. Computer programs are available from Chen or Tsyrennikov upon request.

Monte Carlo Designs. We consider numerous combinations of copulas and marginals in bivariate and trivariate semiparametric copula models. The marginals used include Student's t, normal, and mixture of normals. The bivariate copulas used include the Clayton copula $[C_C(u_1, u_2; \theta) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}, \theta \ge 0]$, the Gumbel copula $[C_G(u_1, u_2; \theta) = \exp(-[(-\ln u_1)^\theta + (-\ln u_2)^\theta]^{1/\theta}), \theta \ge 1]$, the Gaussian copula $[\Phi_\theta(\Phi^{-1}(u_1), \Phi^{-1}(u_2))$ with $\Phi(\cdot)$ being the standard univariate normal cdf and $\Phi_\theta(u_1, u_2)$ the standard bivariate normal cdf with correlation coefficient θ, $|\theta| \le 1]$, and a mixture of Gaussian and Clayton copula $[.9\Phi_\theta(\Phi^{-1}(u_1), \Phi^{-1}(u_2)) + .1C_C(u_1, u_2; .5), |\theta| \le 1]$. These four copulas characterize four different dependence structures. The Clayton copula exhibits positive dependence (i.e., its Kendall's τ is always positive) and lower tail dependence, but no upper tail dependence; as θ increases, both its overall dependence (in terms of Kendall's τ) and its lower tail dependence increase. The Gumbel copula has positive dependence and upper tail dependence, but no lower tail dependence; as θ increases, both its overall dependence and its upper tail dependence increase. The bivariate Gaussian copula has symmetric positive and negative dependence but no tail dependence; as $|\theta|$ increases, its overall dependence increases. The mixture of Gaussian and Clayton copula exhibits asymmetric positive and negative dependence; as $|\theta|$ increases, its overall dependence increases, but its tail dependence remains unchanged. In addition, we consider two families of trivariate copulas: the mixture of Clayton and Gumbel copula, and the Student's t copula. The mixture of Clayton and Gumbel copula has positive and asymmetric tail dependence, whereas the Student's t-copula has positive and negative pairwise dependence and symmetric tail dependence.

Implementation of Sieve MLEs. We tried polynomial spline sieve to approximate a log-marginal density, B-spline sieve, and Hermite polynomial sieve to approximate the square root of a marginal density and obtained similar results. To save space, we describe only the use of a third-order B-spline sieve $\{B_3(x-k)\}_{k=1}^{K_{nj}}$ to approximate $\sqrt{f_{oj}}$ for $j = 1, \ldots, m$. In this case the marginal density f_{oj} is approximated by $f_j(x; a_j) = \{\sum_{k=1}^{K_{nj}} a_{jk} B_3(x-k)\}^2 / \int \{\sum_{k=1}^{K_{nj}} a_{jk} B_3(y-k)\}^2\, dy$. 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. Denote $b_{1j} = \max(z \le \min(X_{ji}) - s_{X_j} : z \text{ is integer})$ and $b_{2j} = \min(z \ge \max(X_{ji}) + s_{X_j} : z \text{ is integer})$. Then we need $K_{nj} = b_{2j} - b_{1j} - 2$ 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 $[b_{1j}, b_{2j}]$; we tried grid size of .01 and .005, and the simulation results are very similar. For each Monte Carlo design, the sieve MLE is computed as that discussed in Remark 1.

A Brief Summary of Simulation Findings. Through the simulation study, we aim to shed light on (1) the sensitivity of the performance of sieve MLE with respect to the choice of sieve bases and the choice of smoothing parameters, (2) the performance of the consistent estimators of the asymptotic variances for the sieve MLEs of the copula parameter and the marginals, (3) the accuracy of the simulated theoretical asymptotic variances of the sieve MLEs, (4) the relative performance of sieve MLE of the copula parameter to the two-step estimator, and (5) the relative performance of the sieve MLE of the marginal cdf's to the empirical cdf's. Regarding (1), we find the sieve MLEs are not sensitive to the choice of sieve bases and our crude way of choosing smoothing parameters work well in finite samples. Regarding (2) and (3), we find that our consistent estimators of the asymptotic variances are close to the simulated theoretical asymptotic variances. For general semiparametric copula models, there are no analytical closed-form expressions of asymptotic efficient variance bounds for copula parameters and marginal cdfs', hence we could not say how accurate our consistent asymptotic variance estimators and our simulated theoretical asymptotic variances are in general. However, for three specific models—(a) the bivariate Clayton copula with one known margin, (b) the bivariate Gaussian copula with one known margin, and (c) the bivariate Gaussian copula with unknown margins—the true efficiency bounds for the copula parameter and/or the marginals have been established by Bickel et al. (1993) for cases (a) and (b) and by Klaassen and Wellner (1997) for case (c). For these models, we find that both our estimators of asymptotic variances and our simulated theoretical asymptotic variances of the sieve MLEs are very close to the true efficiency bounds. Regarding (4), we find that for weak dependence (in terms of small absolute values of Kendall's τ) and/or little tail dependence, the sieve MLE and the two-step estimator of the copula parameter perform comparably in finite samples, but for strong dependence (in terms of large absolute values of Kendall's τ), the sieve MLE of copula parameter performs much better than the two-step estimator in terms of finite-sample Monte Carlo variance and mean squared error (MSE), as well as asymptotic variances, except for the bivariate Gaussian copula model with unknown margins, where both estimators (sieve MLE and the two-step estimator) of the Gaussian copula correlation parameter are asymptotically efficient. Regarding (5), we find that as long as there is some dependence, the sieve MLEs of marginal distributions are always better than the empirical distributions in terms of finite-sample MSEs and asymptotic variances; this is true even for the bivariate Gaussian copula model with unknown and unequal margins. Finally, the sieve MLEs perform better when prior information on the marginals is incorporated.

APPENDIX: MATHEMATICAL PROOFS

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

(a) $$\left| E_o\left(\frac{\partial^2 \log c(\widetilde{\alpha})}{\partial\theta\,\partial\theta'} - \frac{\partial^2 \log c(\alpha_o)}{\partial\theta\,\partial\theta'}\right)\right| \le K\|\widetilde{\alpha} - \alpha_o\|^{1-\epsilon_2},$$

(b) $$\left| E_o\left(\left\{\frac{\partial^2 \log c(\widetilde{\alpha})}{\partial\theta\,\partial u_j} - \frac{\partial^2 \log c(\alpha_o)}{\partial\theta\,\partial u_j}\right\}\int^{X_j} v_j(x)\,dx\right)\right| \le K\|v\|^{1-\epsilon_1}\|\widetilde{\alpha} - \alpha_o\|^{1-\epsilon_2} \quad \text{for all } j = 1, \dots, m,$$

(c) $$\left| E_o\left(\left\{\frac{\partial^2 \log c(\widetilde{\alpha})}{\partial u_i\,\partial u_j} - \frac{\partial^2 \log c(\alpha_o)}{\partial u_i\,\partial u_j}\right\} \times \int^{X_j} v_j(x)\,dx \int^{X_i} v_i(x)\,dx\right)\right| \le K\|v\|^{2(1-\epsilon_1)}\|\widetilde{\alpha} - \alpha_o\|^{1-\epsilon_2} \quad \text{for all } j, i = 1, \dots, m,$$

(d) $$\left| E_o\left(\left[\frac{v_j(X_j)}{\widetilde{f}_j(X_j)}\right]^2 - \left[\frac{v_j(X_j)}{f_{oj}(X_j)}\right]^2\right)\right| \le K\|v\|^{2(1-\epsilon_1)}\|\widetilde{\alpha} - \alpha_o\|^{1-\epsilon_2} \quad \text{for all } j = 1, \dots, m.$$

In the following assumption, 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. (a)

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

(b) for all $j = 1, \dots, m$,

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

(c)

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

Assumptions 5 and 6 are sufficient conditions for controlling 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, and Hölder classes) and are bounded away from 0. When unknown marginal densities are smooth but approach 0 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 of Theorem 1

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

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

In what follows, we show that

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \ell(\alpha_o, Z_i)}{\partial \alpha'}[\Pi_n \upsilon^* - \upsilon^*] = o_P(n^{-1/2}), \tag{A.1}$$

$$E_o(r[\hat{\alpha}, \alpha_o, Z_i] - r[\hat{\alpha} \pm \varepsilon_n \Pi_n \upsilon^*, \alpha_o, Z_i]) = \pm\varepsilon_n\langle \hat{\alpha} - \alpha_o, \upsilon^*\rangle + \varepsilon_n o_P(n^{-1/2}), \tag{A.2}$$

and

$$\mu_n(r[\hat{\alpha}, \alpha_o, Z_i] - r[\hat{\alpha} \pm \varepsilon_n \Pi_n \upsilon^*, \alpha_o, Z_i]) = \varepsilon_n o_P(n^{-1/2}). \tag{A.3}$$

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

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

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

To complete the proof, it remains to establish (A.1)–(A.3). Note that (A.1) is implied by Chebyshev inequality, iid data, and $\|\Pi_n \upsilon^* - \upsilon^*\| = o(1)$, which is satisfied given Assumption 4(b). For (A.2), we note that

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

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

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

Under Assumption 5, we have

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

where the last equality holds because Assumptions 4(a) and 4(b) imply that

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

Hence (A.2) is satisfied. For (A.3), we note that

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

where $\widetilde{\alpha} \in \mathcal{A}_n$ is between $\hat{\alpha}$ and $\hat{\alpha} \pm \varepsilon_n \Pi_n \upsilon^*$. Because

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

(A.3) is implied by Assumption 6.

The semiparametric efficiency is a direct application of theorem 4 of Shen (1997).

Proof of Proposition 1

Recall that the semiparametric efficiency bound for θ_o is $\mathcal{I}_*(\theta_o) = E_o\{\mathcal{S}_{\theta_o}\mathcal{S}_{\theta_o}'\}$, 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 of the score functions for the nuisance parameters f_{oj}, $j = 1, \ldots, m$. Moreover, θ_o is *$\sqrt{n}$-efficiently estimable* if and only if $E_o\{\mathcal{S}_{\theta_o}\mathcal{S}_{\theta_o}'\}$ is *nonsingular* (see, e.g., Bickel et al. 1993). Hence (14) is clearly a necessary condition for $\sqrt{n}$-normality and efficiency of $\hat{\theta}$ for θ_o.

Under Assumptions 2 and 3′, propositions 4.7.4 and 4.7.6 of Bickel et al. (1993, pp. 165–168) for bivariate copula models can be directly extended to the multivariate case (see also Klaassen and Wellner 1997, sec. 4). Therefore, with $\mathcal{S}_{\theta_o}$ as defined in (17), we have that $\mathcal{I}_*(\theta_o) = E_o\{\mathcal{S}_{\theta_o}\mathcal{S}_{\theta_o}'\}$ 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 that for any $\lambda \in \mathcal{R}^{d_\theta}$, $\lambda \ne 0$, we have $\sqrt{n}(\lambda'\hat{\theta} - \lambda'\theta_o) \Rightarrow \mathcal{N}(0, \lambda'\mathcal{I}_*(\theta_o)^{-1}\lambda)$. This implies Proposition 1.

Proofs of Propositions 2, 4, 6 and 8

The consistency of these asymptotic variances can be established by applying theorem 5.1 of Ai and Chen (2003).

[Received October 2004. Revised January 2006.]

REFERENCES

Ai, C., and Chen, X. (2003), "Efficient Estimation of Models With Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795–1843.

Barron, A., and Sheu, C. (1991), "Approximation of Density Functions by Sequences of Exponential Families," *The Annals of Statistics*, 19, 1347–1369.

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

Chen, X., and Fan, Y. (2005), "Estimation and Model Selection of Semiparametric Copula-Based Multivariate Dynamic Models Under Copula Misspecification," *Journal of Econometrics*, available online 6 Sept. 2005.

——— (2006), "Estimation of Copula-Based Semiparametric Time Series Models," *Journal of Econometrics*, 130, 307–335.

Chen, X., Fan, Y., and Tsyrennikov, V. (2005), "Efficient Estimation of Semiparametric Multivariate Copula Models," working paper, New York University.

Chen, X., and White, H. (1999), "Improved Rates and Asymptotic Normality for Nonparametric Neural Network Estimators," *IEEE Transactions on Information Theory*, 45, 682–691.

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

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.

Clayton, D., and Cuzick, J. (1985), "Multivariate Generalizations of the Proportional Hazards Model" (with discussion), *Journal of the Royal Statistical Society*, Ser. A, 148, 82–117.

Coppejans, M., and Gallant, A. R. (2002), "Cross-Validated SNP Density Estimates," *Journal of Econometrics*, 110, 27–65.

Embrechts, P., McNeil, A., and Straumann, D. (2002), "Correlation and Dependence Properties in Risk Management: Properties and Pitfalls," in *Risk Management: Value at Risk and Beyond*, ed. M. Dempster, Cambridge, U.K.: Cambridge University Press, pp. 176–223.

Frees, E. W., and Valdez, E. A. (1998), "Understanding Relationships Using Copulas," *North American Actuarial Journal*, 2, 1–25.

Gallant, A. R., and Nychka, D. (1987), "Semi-Non-Parametric Maximum Likelihood Estimation," *Econometrica*, 55, 363–390.

Geman, S., and Hwang, C. (1982), "Nonparametric Maximum Likelihood Estimation by the Method of Sieves," *The Annals of Statistics*, 10, 401–414.

Genest, C. (1987), "Frank's Family of Bivariate Distributions," *Biometrika*, 74, 549–555.

Genest, C., Ghoudi, K., and Rivest, L.-P. (1995), "A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions," *Biometrika*, 82, 534–552.

Genest, C., and Werker, B. (2002), "Conditions for the Asymptotic Semiparametric Efficiency of an Omnibus Estimator of Dependence Parameters in Copula Models," in *Proceedings of the Conference on Distributions With Given Marginals and Statistical Modelling*, eds. C. M. Cuadras and J. A. Rodríguez Lallena, Dordrecht: Kluwer Academic Publishers, pp. 103–112.

Granger, C. W. J., Teräsvirta, T., and Patton, A. (2006), "Common Factors in Conditional Distributions," *Journal of Econometrics*, 132, 43–57.

Grenander, U. (1981), *Abstract Inference*, New York: Wiley.

Heckman, J. J., and Honoré, B. E. (1989), "The Identifiability of the Competing Risks Model," *Biometrika*, 76, 325–330.

Joe, H. (1997), *Multivariate Models and Dependence Concepts*, Boca Raton, FL: Chapman & Hall/CRC.

Klaassen, C., and Wellner, J. (1997), "Efficient Estimation in the Bivariate Normal Copula Model: Normal Margins Are Least Favorable," *Bernoulli*, 3, 55–77.

Lee, L. (1983), "Generalized Econometric Models With Selectivity," *Econometrica*, 51, 507–512.

Maguluri, G. (1993), "Semiparametric Estimation of Association in a Bivariate Survival Function," *The Annals of Statistics*, 21, 1648–1662.

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

Oakes, D. (1982), "A Model for Association in Bivariate Survival Data," *Journal of the Royal Statistical Society*, Ser. B, 44, 414–422.

——— (1986), "Semiparametric Inference in a Model for Association in Bivariate Survival Data," *Biometrika*, 73, 353–361.

——— (1989), "Bivariate Survival Models Induced by Frailties," *Journal of the American Statistical Association*, 84, 487–493.

——— (1994), "Multivariate Survival Distributions," *Journal of Nonparametric Statistics*, 3, 343–354.

Patton, A. J. (2004), "On the Out-of-Sample Importance of Skewness and Asymmetric Dependence for Asset Allocation," *Journal of Financial Econometrics*, 2, 130–168.

Robinson, P. (1988), "Root-n-Consistent Semiparametric Regression," *Econometrica*, 56, 931–954.

Schumaker, L. (1981), *Spline Functions: Basic Theory*, New York: Wiley.

Shen, X. (1997), "On Methods of Sieves and Penalization," *The Annals of Statistics*, 25, 2555–2591.

Shen, X., and Wong, W. (1994), "Convergence Rate of Sieve Estimates," *The Annals of Statistics*, 22, 580–615.

Shen, X., and Ye, J. (2002), "Adaptive Model Selection," *Journal of the American Statistical Association*, 97, 210–221.

Shih, J. H., and Louis, T. A. (1995), "Inferences on the Association Parameter in Copula Models for Bivariate Survival Data," *Biometrics*, 51, 1384–1399.

Sklar, A. (1959), "Fonctions de répartition à n dimensionset leurs marges," *Publications de l'Institut de Statistique de L'Univerité de Paris*, 8, 229–231.

Stone, C. (1982), "Optimal Global Rates of Convergence for Nonparametric Regression," *The Annals of Statistics*, 10, 1040–1053.

——— (1990), "Large-Sample Inference for Log-Spline Models," *The Annals of Statistics*, 18, 717–741.

Van de Geer, S. (2000), *Empirical Processes in M-Estimation*, Cambridge, U.K.: Cambridge University Press.

Wong, W., and Shen, X. (1995), "Probability Inequalities for Likelihood Ratios and Convergence Rates for Sieve MLEs," *The Annals of Statistics*, 23, 339–362.