Journal of Multivariate Analysis 123 (2014) 330–344

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Efficient estimation of semiparametric copula models for bivariate survival data

Guang Cheng [a,*], Lan Zhou [b], Xiaohong Chen [c], Jianhua Z. Huang [b]

[a] *Department of Statistics, Purdue University, West Lafayette, IN, 47907, United States*

[b] *Department of Statistics, Texas A&M University, College Station, TX, 77843, United States*

[c] *Department of Economics, Yale University, New Haven, CT, 06520, United States*

ARTICLE INFO

Article history:
Received 11 November 2011
Available online 29 October 2013

AMS 2000 subject classifications:
primary 62F40
secondary 62G20

Keywords:
Semiparametric copula model
Bivariate survival data
B-spline
Efficiency
Consistent covariance estimation

ABSTRACT

A semiparametric copula model for bivariate survival data is characterized by a parametric copula model of dependence and nonparametric models of two marginal survival functions. Efficient estimation for the semiparametric copula model has been recently studied for the complete data case. When the survival data are censored, semiparametric efficient estimation has only been considered for some specific copula models such as the Gaussian copulas. In this paper, we obtain the semiparametric efficiency bound and efficient estimation for general semiparametric copula models for possibly censored data. We construct an approximate maximum likelihood estimator by approximating the log baseline hazard functions with spline functions. We show that our estimates of the copula dependence parameter and the survival functions are asymptotically normal and efficient. Simple consistent covariance estimators are also provided. Numerical results are used to illustrate the finite sample performance of the proposed estimators.

1. Introduction

Economic, financial and medical multivariate survival data are typically non-normally distributed and exhibit nonlinear dependence among their component variables. Some well-known examples include the Danish Twin Study [30], the diabetic retinopathy study [9], the dual infection kidney dialysis study [27], and the reproductive health study of the association of age at a marker event and age at menopause [16]. In all these studies, the assessment of survival functions and the dependence among component variables, e.g. twins, are of major interests. Another example in financial area is the insurance company data on losses and allocated loss adjustment expenses (ALAEs) analyzed in [6].

The distribution of bivariate survival data can be characterized by a bivariate survival function. Nonparametric estimation of the bivariate survival function for right-censored data under independent censorship has been thoroughly studied; see [29] for a review. However, we focus on semiparametric models in this paper. By Sklar's [24] theorem, any bivariate survival function with continuous marginal survival functions can be uniquely represented by its (survival) copula function evaluated at its marginal survival functions, where the copula function captures all the dependence among the component variables. We consider a large class of semiparametric models for bivariate survival data—the semiparametric copula models, in which a bivariate survival function is modeled as a parametric (survival) copula function evaluated at nonparametric marginal survival functions. Note that our methodology and theory in this paper can be easily extended to accommodate multivariate survival data.

* Corresponding author.
E-mail addresses: guang.cheng.stat@gmail.com, chengg@purdue.edu, chengg@stat.purdue.edu (G. Cheng), lzhou@stat.tamu.edu (L. Zhou), xiaohong.chen@yale.edu (X. Chen), jianhua@stat.tamu.edu (J.Z. Huang).

0047-259X/$ – see front matter
http://dx.doi.org/10.1016/j.jmva.2013.10.008

For complete data, i.e., data without censoring or truncation, Oakes [18] and Genest et al. [7] proposed a semiparametric two-step estimation procedure: In the first step the unknown marginal distribution functions are estimated by rescaled empirical distribution functions, and in the second step the copula dependence parameter is estimated by maximizing the estimated log-likelihood function where the marginal distributions are fixed at the estimated ones from the first step. The two-step estimator of the copula parameter is inefficient in general, see [8], except for the independence case [7] and Gaussian Copula case [12]. This is because the two-step estimator is generally not the solution of the semiparametric efficient score equation for the copula parameter. Moreover, the two-step procedure does not produce an efficient estimator of marginal distribution (survival) function except for the trivial independence case. Intuitively, one could obtain more efficient estimates of the copula parameter and the marginal survival functions by utilizing the dependence information contained in the copula. Chen, Fan and Tsyrennikov [4] proposed a sieve maximum likelihood estimation of a general semiparametric copula model, and established that their estimators of both the copula parameter and the marginal distributions are semiparametrically efficient. In addition, their simulation studies demonstrate that the two-step procedure underestimates the dependence in copula models with strong tail dependence such as Clayton copula, Gumbel copula and others.

On the other hand, the efficient estimation problem is not generally solved for the multivariate survival data when there is censoring. For right censored data, Shih and Louis [23] and Chen et al. [3] proposed the same two-step procedure as Oakes [18] and Genest et al. [7] except that the Kaplan–Meier estimators of marginal survival functions are used in the first step, and developed the corresponding asymptotic theory for inference. Wang and Ding [28] proposed a two-step procedure for current status data using nonparametric MLE in the first step. This two-step procedure is computationally convenient but is generally inefficient as discussed above for the complete data. Recently, for similar right censored data, Li, Prentice and Lin [14] developed the semiparametric efficient maximum likelihood estimation for the normal transformation model, which corresponds to the Gaussian copula. However, their estimation approach and the relevant theoretical proofs are only tailored for the special class of Gaussian copula, and thus cannot be easily generalized to other semiparametric copula models. To the best of our knowledge, there is no general theory on efficient estimation of semiparametric copula model for censored data yet.

The purpose of this paper is to develop an efficient estimation procedure of the semiparametric copula model for the right censored survival data under the general copula framework. Our procedure is based on spline approximation of the log marginal hazard functions. We derive the semiparametric efficiency bound for the general semiparametric copula model and show that, under regularity conditions, our proposed estimators are asymptotic normally distributed and achieve the semiparametric efficiency bound. Although the asymptotic covariance matrix of our estimators have no closed-form expressions for general semiparametric copula models, we provide simple consistent estimates of the asymptotic covariance matrix of our estimators.

The rest of this paper is organized as follows. Section 2 defines the semiparametric copula models for censored data. Section 3 presents our proposed estimator and its asymptotic properties. In particular, Section 3.1 gives the exact likelihood function and the approximate likelihood function under spline approximation of the log marginal hazard functions. Sections 3.2 and 3.3 focus on the copula dependence parameter. While Section 3.2 contains the semiparametric efficiency bound calculation and establishes the asymptotic normality and efficiency of our copula parameter estimator, Section 3.3 gives a simple consistent estimator of its asymptotic covariance matrix. Section 3.4 discusses the efficient estimation of the marginal cumulative hazard functions and consistent estimators of their asymptotic variances. Section 4 presents some simulation results to illustrate the finite sample properties of the proposed estimator.

2. Semiparametric survival copula models

Consider a pair of possibly correlated survival times (T_1, T_2) with joint survival function $S(t_1, t_2) = P(T_1 > t_1, T_2 > t_2)$ and marginal survival functions $S_j(t_j) = P(T_j > t_j)$ for $j = 1, 2$. Sklar's [24] theorem implies that, for continuous S_j's, there exists a unique survival copula function C such that $S(t_1, t_2) = C(S_1(t_1), S_2(t_2))$, where $C(\cdot)$ captures the dependence structure of (T_1, T_2). This decomposition of the joint survival function leads naturally to the class of semiparametric survival copula models in which the marginal survival functions are unspecified, but the survival copula function is parameterized as $C(u, v) = C(u, v; \theta_0)$ for some Euclidean parameter θ_0. This class of semiparametric models represents the joint survival function as

$$S(t_1, t_2) = C(S_1(t_1), S_2(t_2); \theta_0). \tag{1}$$

We refer to Nelsen [17] and Joe [11] for properties of different parametric (survival) copulas. The semiparametric copula model (1) is very general and includes many existing models as special cases. For example, if the Gaussian copula function is used in (1), the model is equivalent to the normal transformation model considered by Li, Prentice and Lin [14]. If the Archimedean copula function is used, the model can be transformed into the frailty model for the multivariate survival data; see [15].

In the following, we will consider the right censored data, and shall use (D_1, D_2) to denote the censoring times. We observe $(Y_1, Y_2) \equiv (T_1 \wedge D_1,\ T_2 \wedge D_2)$ and $(\delta_1, \delta_2) \equiv (I\{T_1 \le D_1\}, I\{T_2 \le D_2\})$. The support of the marginal distribution of Y_1 and Y_2 is assumed to be a compact interval and is further assumed to be [0, 1] for notational simplicity. We also assume that (T_1, T_2) is independent of (D_1, D_2). This independence assumption allows various censoring mechanisms, e.g., random censoring, fixed censoring and no censoring. The above general censoring scheme applies to the loss–ALAE data set [6] where ALAE is not censored and loss is censored by a constant differing from each individual to another.

Remark 2.1. The survival copula $C(\cdot,\cdot)$ should not be confused with the commonly used copula $\bar{C}(\cdot,\cdot)$ unless the latter is radially symmetric, e.g., Gaussian copula. However, the survival copula C is related to $\bar{C}$ in the following manner:

$$C(u,v) = u + v - 1 + \bar{C}(1-u, 1-v). \tag{2}$$

3. Semiparametric maximum likelihood estimation

3.1. Maximum likelihood

Suppose n i.i.d. observations $\{X_i = (Y_{1i}, Y_{2i}, \delta_{1i}, \delta_{2i})\}_{i=1}^n$ are available. Denote the joint distribution and density for (T_1, T_2) as $F(\cdot,\cdot)$ and $f(\cdot.\cdot)$, the marginal distribution and density as $F_j(\cdot)$ and $f_j(\cdot)$ for $j = 1, 2$. For notational simplicity, we denote $s_1 = S_1(y_1)$, $s_2 = S_2(y_2)$, $C_j(s_1, s_2; \theta) = (\partial/\partial s_j)C(s_1, s_2; \theta)$, and $C_{12}(s_1, s_2; \theta) = (\partial^2/\partial s_1 \partial s_2)C(s_1, s_2; \theta)$. The log-likelihood for (θ, S_1, S_2) is written as

$$\begin{aligned}\ell(\theta, S_1, S_2) = {} & \delta_1\delta_2 \log[C_{12}(s_1, s_2; \theta) f_1(y_1) f_2(y_2)] + \delta_1(1-\delta_2)\log[C_1(s_1, s_2; \theta) f_1(y_1)] \\ & + \delta_2(1-\delta_1)\log[C_2(s_1, s_2; \theta) f_2(y_2)] + (1-\delta_1)(1-\delta_2)\log C(s_1, s_2; \theta).\end{aligned} \tag{3}$$

Direct maximization of the above exact log-likelihood over an infinite dimensional space containing continuous survival functions is not feasible. To achieve this, we may need to rewrite (3) in the form of empirical likelihood considering monotonicity constraints on $S_j(\cdot)$'s. For example, Li, Prentice and Lin [14] assume that the cumulative hazard function $\Lambda_j(\cdot) \equiv -\log S_j(\cdot)$ to be cadlag and piecewise constant in the case of Gaussian copula. Unfortunately, this strategy requires a case-by-case analysis for different copula functions. In this paper, we propose an approximate log-likelihood approach, which applies to a general class of copula functions and yields smooth estimate for $S_j(\cdot)$ and $\Lambda_j(\cdot)$. Our approach can also be generalized to consider the cluster survival data of varying sizes.

Denoted the log baseline hazard function as $h_j(\cdot) \equiv \log \dot{\Lambda}_j(\cdot) \equiv \log \lambda_j(\cdot)$. Our procedure approximates $h_j(\cdot)$ by a linear combination of a finite number of smooth basis functions, i.e., B-splines, and maximizes the likelihood with respect to the copula parameter and the B-spline coefficient parameters. Specifically, we approximate $h_j(\cdot)$ as follows

$$h_j(t) \approx \sum_{k=1}^{K} \gamma_{jk} B_{jk}(t) = \gamma_j' \mathbf{B}_j(t), \tag{4}$$

where $\mathbf{B}_j(\cdot)$ is a vector of B-splines. When h_j is a smooth function, basis functions can be chosen such that the above approximation is very accurate. In our theoretical analysis, we assume the parameter space for h_j is $\mathcal{F}_j \equiv \{h_j(\cdot) : h_j \in \mathrm{H}_{c_j'}^{r_j}[0,1]$ with $\|h_j\|_\infty \le c_j\}$, where $\mathrm{H}_c^r(\mathcal{Y})$ is a Hölder ball containing widely used smooth functions in the nonparametric estimation. Specifically, $h \in \mathrm{H}_c^r(\mathcal{Y})$ if and only if it is $J < r$ times continuously differentiable on $\mathcal{Y}$ and its J-th derivative satisfies a Hölder condition with exponent $\kappa \equiv r - J \in (0, 1]$, i.e.,

$$\sup_{x,y\in\mathcal{Y}, x\neq y} \frac{|h^{(J)}(x) - h^{(J)}(y)|}{|x-y|^\kappa} \le c.$$

We assume that $r_j > 1$. Given a system of basis functions $\mathbf{B}_j(t)$, we define the approximate parameter space as

$$\mathcal{F}_{jn} = \left\{ h_j(\cdot) : h_j(t) = \sum_{k=1}^{K_j} \gamma_{jk} B_{jk}(t) = \gamma_j' \mathbf{B}_j(t) \text{ and } \|h_j\|_\infty \le c_j \right\},$$

where $K_j \to \infty$ and $K_j/n \to 0$. Note that, for any $h \in \mathrm{H}_c^r(\mathcal{Y})$, there exists a spline function $\gamma'\mathbf{B}$ with degree $d \ge (r-1)$ such that, as $K_j \to \infty$,

$$\|h - \gamma'\mathbf{B}\|_\infty \asymp K_j^{-r}, \tag{5}$$

where the notation "$\asymp$" means that the ratio of both sides is bounded away from zero and infinity [20]. We assume that Θ is a compact subset of $\mathbb{R}^p$ in which the true value θ_0 is an interior point. Let $\mathcal{A} = \Theta \times \mathcal{F}_1 \times \mathcal{F}_2$ and $\mathcal{A}_n = \Theta \times \mathcal{F}_{1n} \times \mathcal{F}_{2n}$.

Denote the unknown parameters collectively as $\alpha = (\theta', h_1, h_2)'$ with $\alpha_0 = (\theta_0', h_{10}, h_{20})'$ being the true value. The log-likelihood (3) can be written as a function of these parameters as

$$\begin{aligned}\ell(\alpha) = {} & \delta_1\delta_2 \log[C_{12}(s_1, s_2; \theta) s_1 s_2 \exp\{h_1(y_1)\}\exp\{h_2(y_2)\}] + \delta_1(1-\delta_2)\log[C_1(s_1, s_2; \theta) s_1 \exp\{h_1(y_1)\}] \\ & + \delta_2(1-\delta_1)\log[C_2(s_1, s_2; \theta) s_2 \exp\{h_2(y_2)\}] + (1-\delta_1)(1-\delta_2)\log\{C(s_1, s_2; \theta)\},\end{aligned} \tag{6}$$

where $s_j = \exp[-\int_0^{y_j} \exp\{h_j(x)\}dx]$ for $j = 1, 2$. We denote the log-likelihood for the observation i as $\ell_i(\alpha)$. The B-spline estimate of α is defined as

$$\widehat{\alpha} = (\widehat{\theta}', \widehat{h}_1, \widehat{h}_2)' = (\widehat{\theta}', \widehat{\gamma}_1'\mathbf{B}_1, \widehat{\gamma}_2'\mathbf{B}_2)' = \arg\max_{\alpha} \sum_{i=1}^{n} \ell_i(\alpha). \tag{7}$$

Consequently, the maximum likelihood estimates of the marginal cumulative hazard function and survival function are $\widehat{\Lambda}_j(y) = \int_0^y \exp(\widehat{h}_j(s))ds$ and $\widehat{S}_j(y) = \exp\{-\widehat{\Lambda}_j(y)\}$, respectively. In the case of no censoring, i.e., $D_j = +\infty$ for $j = 1, 2$, our estimate $\widehat{\alpha}$ is the same as the general sieve estimate proposed in [4]. Note that by modeling the log hazard function h_j directly, our approach naturally leads to the positivity and monotonicity of the resulting cumulative hazard function $\Lambda_j(t) = \int_0^t \lambda_j(s)ds$, and avoids the constrained optimization required in the implementation of monotone spline estimation, e.g., [31].

Remark 3.1. In the theoretical proofs and numerical calculations, the exact values of c_j's required in $\mathcal{F}_j$ and $\mathcal{F}_{jn}$ are not necessary. Instead, only the boundedness condition, equivalently the compactness of parameter spaces and spline spaces, is needed. Here we assume this boundedness condition, which can be relaxed by invoking the chaining arguments (see [19]), only for simplifying our theoretical derivations

Remark 3.2. Our B-spline estimation framework is very flexible so that the prior restrictions on the marginal cumulative hazard function or survival function can be easily taken into account. For example, when the two marginal survivals are assumed to be equal, we can set $\gamma_1 = \gamma_2 = \gamma$ and $\mathbf{B}_1(\cdot) = \mathbf{B}_2(\cdot) = \mathbf{B}(\cdot)$ to obtain $(\widehat{\theta}, \widehat{\gamma})$. Another example is that one marginal survival is of a particular parametric form, e.g., $S_1(\cdot, z; \beta)$ follows the Cox regression $\exp(-e^{\beta' z}\Lambda_1(\cdot))$, but another is left unspecified. In this case, we can replace s_1 and $\exp\{h_1(y_1)\}$ with $S_1(y_1, z; \beta)$ and $-[(\partial/\partial y_1)S_1(y_1, z; \beta)]/S_1(y_1, z; \beta)$ in (6) to obtain $(\widehat{\theta}, \widehat{\beta}, \widehat{\gamma}_1, \widehat{\gamma}_2)$.

3.2. Asymptotic properties of the copula estimator

In this section, we show that $\widehat{\theta}$ defined in (7) is asymptotically normal and semiparametric efficient in the sense that it achieves the semiparametric efficiency bound, i.e., the minimal possible asymptotic covariance matrix over all the regular estimators.

We first derive the semiparametric efficiency bound using the concept of the hardest one-dimensional submodel as in the general formulation of Bickel et al. [1]. Consider a class of parametric submodels perturbing around α_0 along the directions $v = (v_\theta', v_1, v_2)'$, i.e., $\{t \mapsto \ell(\alpha_0 + tv) : v \in \mathbf{V}\}$. Here $\mathbf{V}$ represents some perturbation space and is defined as the closure of the linear span of $\{\alpha - \alpha_0 : \alpha \in \mathcal{A}\}$. For each fixed $v \in \mathbf{V}$, the corresponding parametric submodel has the score function

$$\begin{aligned}\dot{\ell}(x; \alpha_0)[v] &= (d/dt)|_{t=0}\ell(\alpha_0 + tv) \\ &= v_\theta'\dot{\ell}_\theta(x; \alpha_0) + \dot{\ell}_{h_1}(x; \alpha_0)[v_1] + \dot{\ell}_{h_2}(x; \alpha_0)[v_2],\end{aligned} \tag{8}$$

where

$$\dot{\ell}_\theta(x; \alpha) = \frac{\partial H(s_1, s_2; \theta)}{\partial \theta}, \tag{9}$$

$$\dot{\ell}_{h_j}(x; \alpha)[v_j] = \left\{v_j(y_j) - G_j[v_j](y_j)\right\}\delta_j - G_j[v_j](y_j)\frac{\partial H(s_1, s_2; \theta)}{\partial s_j}s_j, \tag{10}$$

and

$$H(s_1, s_2; \theta) = \delta_1\delta_2 \log C_{12} + \delta_1(1 - \delta_2)\log C_1 + \delta_2(1 - \delta_1)\log C_2 + (1 - \delta_1)(1 - \delta_2)\log C,$$

$$G_j[v_j](y_j) = \int_0^{y_j} \exp(h_j(x))v_j(x)dx.$$

Note that the definition of the score function and Fisher information requires some regularity conditions, i.e. Condition R1, which we defer to later sections when we present the asymptotic results. The Fisher information of the parametric model is thus calculated as $I_0(v) = E\{\dot{\ell}(\alpha_0)[v]\}^2$. Since $I_0(v)$ is a quadratic form of v, we refer to it as the Fisher norm and denote it as $\|v\|^2$. We assume that the information $I_0(v)$ is bounded away from zero in the following sense: there exists a constant $c_0 > 0$ such that $I_0(v) \geq c_0(|v_\theta| + \|v_1\|_2^2 + \|v_2\|_2^2)$ for all $v = (v_\theta', v_1, v_2)'$. This positive information assumption rules out the irregular cases for which there exist parametric submodels with zero information.

To obtain the semiparametric efficiency bound of θ, we first focus our attention on the transformed variable $\rho_\lambda(\alpha_0) = \lambda'\theta_0$ for any fixed $\lambda \in \mathbb{R}^p$. Estimation of $\rho_\lambda(\alpha_0)$ in the semiparametric model $\ell(\alpha)$ corresponds to the estimation of $\psi_\lambda(t) = \lambda'(\theta_0 + tv_\theta)$ at $t = 0$ in the above perturbed submodel, which has the Cramér–Rao lower bound

$$CR_\lambda(v) = \frac{[\dot{\psi}_\lambda(0)]^2}{I_0(v)} = \frac{(\lambda' v_\theta)^2}{\|v\|^2}.$$

The largest Cramér–Rao lower bound in the above class of parametric submodels, i.e.,

$$CR_\lambda^* = \sup_{v\in\mathbf{V}:\|v\|>0} CR_\lambda(v) = \sup_{v_\theta}\left\{\frac{\lambda' v_\theta v_\theta'\lambda}{\inf_{v_1,v_2}\|v\|^2}\right\}, \tag{11}$$

corresponds to the semiparametric efficiency bound for estimating $\lambda'\theta_0$; see [1]. A more transparent form of CR_λ^* is obtained as follows.

Let $\mathcal{T}$ denote the closure in $L_2(P)$, where P is the probability measure of the underlying data generating process, of the linear span of $\{\sum_{j=1}^2 \dot\ell_{h_j}(\alpha_0)[v_j]\}$. Let (w_{1k}^*, w_{2k}^*) be the k-th coordinate projection of $\dot\ell_\theta(\alpha_0)$ onto $\mathcal{T}$ in the sense that $(w_{1k}, w_{2k}) \mapsto E\{[\dot\ell_\theta(\alpha_0)]_k + \sum_{j=1}^2 \dot\ell_{h_j}(\alpha_0)[w_{jk}]\}^2$ achieves its minimal at (w_{1k}^*, w_{2k}^*). Define

$$\widetilde{\ell}_0 = \dot\ell_\theta(\alpha_0) + \sum_{j=1}^2 \dot\ell_{h_j}(\alpha_0)[\bar w_j^*] \quad\text{and}\quad \widetilde{I}_0 = E\widetilde{\ell}_0\widetilde{\ell}_0', \tag{12}$$

where $\bar w_1^* = (w_{11}^*, \ldots, w_{1p}^*)'$ and $\bar w_2^* = (w_{21}^*, \ldots, w_{2p}^*)'$. Considering the fact that the projection of $v_\theta'\dot\ell_\theta(\alpha_0)$ onto $\mathcal{T}$ is $\sum_{j=1}^2 \dot\ell_{h_j}(\alpha_0)[v_\theta'\bar w_j^*]$, we can decompose $\|v\|^2 \equiv E\{\dot\ell(\alpha_0)[v]\}^2$ as

$$v_\theta'\widetilde{I}_0 v_\theta + E\left(\sum_{j=1}^2 \dot\ell_{h_j}(\alpha_0)[v_j^{res}]\right)^2,$$

where $v_j^{res} \equiv v_j - v_\theta'\bar w_j^*$ for $j = 1, 2$. Thus, by setting $v_j^{res} = 0$, i.e., $v_j = v_\theta'\bar w_j^*$, we obtain $\inf_{v_1,v_2}\|v\|^2$ in (11) as $v_\theta'\widetilde{I}_0 v_\theta$, which further reduces the form of (11) to

$$CR_\lambda^* = \sup_{v_\theta}\frac{\lambda' v_\theta v_\theta'\lambda}{v_\theta'\widetilde{I}_0 v_\theta} = \lambda'\widetilde{I}_0^{-1}\lambda, \tag{13}$$

where the optimizer is $v_\theta^* = \widetilde{I}_0^{-1}\lambda$ by the Cauchy–Schwarz inequality. According to the above analysis, we know that the largest Cramér–Rao lower bound of estimating $\lambda'\theta_0$ is achieved in the parametric submodel $t \mapsto \ell(\alpha_0 + tv^*)$, where

$$v^* = \begin{pmatrix} v_\theta^* \\ v_1^* \\ v_2^* \end{pmatrix} = \begin{pmatrix} I_p \\ (\bar w_1^*)' \\ (\bar w_2^*)' \end{pmatrix} \times \widetilde{I}_0^{-1}\lambda \quad\text{and}\quad I_p \text{ is the } p\times p \text{ identity matrix.} \tag{14}$$

By some algebra, we can derive that $\|v^*\|^2 = \lambda'\widetilde{I}_0^{-1}\lambda$ based on (12).

Since $\widehat\theta$ is asymptotically efficient if and only if $\lambda'\widehat\theta$ is asymptotically efficient for any $\lambda \in \mathbb{R}^p$, it follows from (13) that the largest Cramér–Rao lower bound of estimating θ_0 in a p-dimensional parametric submodel is $\widetilde{I}_0^{-1}$. This lower bound is achieved in the p-dimensional submodel $\mathbf{s} \mapsto \ell(\alpha_0 + u^*\mathbf{s})$, where

$$\mathbf{s} = \begin{pmatrix} s_1 \\ \vdots \\ s_p \end{pmatrix} \quad\text{and}\quad u^* = \begin{pmatrix} I_p \\ (\bar w_1^*)' \\ (\bar w_2^*)' \end{pmatrix}.$$

The above parametric submodel is called the least favorable submodel (LFS) and its corresponding Cramér–Rao bound is the semiparametric efficiency bound for estimating θ_0. The score function, i.e., $\widetilde{\ell}_0$, and information matrix, i.e., $\widetilde{I}_0$, of the LFS are called the efficient score function and efficient information matrix in the semiparametric literature. Given the above LFS, we can obtain an asymptotic linear expansion of the efficient estimate, which involves u^*. This expansion suggests a plug-in efficient estimate if u^* can be estimated consistently. However, this plug-in estimation approach is in general not feasible since there is no closed-form expression of u^*.

Our estimator $\widehat\theta$ defined in the previous subsection does not require the knowledge of u^*. We will show that $\widehat\theta$ is asymptotically normal and semiparametric efficient by applying a general theory in Appendix A.1, i.e., Lemma A.1, which works for any sufficiently smooth functional $\rho(\alpha)$. In particular, we will assume $\rho_\lambda(\alpha) = \lambda'\theta$ to show Theorem 3.1; see Appendix A.3 for details.

Theorem 3.1. *Suppose that Conditions* M1 *and* M2 *in the Appendix hold and* $\widetilde{I}_0$ *is nonsingular, we have*

$$\sqrt{n}(\widehat\theta - \theta_0) \xrightarrow{d} N\left(0, \widetilde{I}_0^{-1}\right). \tag{15}$$

3.3. Consistent estimator for the asymptotic covariance matrix of $\hat\theta$

The asymptotic covariance matrix of $\widehat\theta$, which is related to the infinite dimensional optimization problem (11), usually has no closed-form except in some special case, e.g., bivariate Gaussian copula model for the complete data in [12]. However, we can use the approximate likelihood to construct an explicit estimator of $\widetilde{I}_0$ by treating it as the parametric likelihood with increasing dimension.

We define the B-spline version of score functions:

$$\dot{\ell}_{\gamma_1}(X;\alpha) = \left(\dot{\ell}_{h_1}(X;\alpha)[B_{11}], \ldots, \dot{\ell}_{h_1}(X;\alpha)[B_{1K_1}]\right)',$$
$$\dot{\ell}_{\gamma_2}(X;\alpha) = \left(\dot{\ell}_{h_2}(X;\alpha)[B_{21}], \ldots, \dot{\ell}_{h_2}(X;\alpha)[B_{2K_2}]\right)'.$$

Hence, the observed information for $(\theta', \gamma_1', \gamma_2')'$ is

$$\widehat{J} = \begin{pmatrix} \widehat{I}_{\theta\theta} & \widehat{I}_{\theta\gamma_1} & \widehat{I}_{\theta\gamma_2} \\ \widehat{I}_{\gamma_1\theta} & \widehat{I}_{\gamma_1\gamma_1} & \widehat{I}_{\gamma_1\gamma_2} \\ \widehat{I}_{\gamma_2\theta} & \widehat{I}_{\gamma_2\gamma_1} & \widehat{I}_{\gamma_2\gamma_2} \end{pmatrix}_{(p+K_1+K_2)\times(p+K_1+K_2)},$$

where $\widehat{I}_{jk} = \sum_{i=1}^n \dot{\ell}_j(X_i;\widehat{\alpha})\dot{\ell}_k'(X_i;\widehat{\alpha})/n$, for $j, k = \theta, \gamma_1, \gamma_2$. The observed information can be used to empirically check the positive information assumption stated in Section 3.2: existence of zero or extremely small eigenvalues of $\widehat{J}$ would suggest violation of the assumption.

The theory of parametric inference implies that the information for θ is of the form

$$\widehat{I} = \widehat{I}_{\theta\theta} - \widehat{I}_{\theta\eta}\widehat{I}_{\eta\eta}^{-1}\widehat{I}_{\eta\theta}, \tag{16}$$

where $\eta = (\gamma_1', \gamma_2')'$, $\widehat{I}_{\theta\eta} = (\widehat{I}_{\theta\gamma_1}, \widehat{I}_{\theta\gamma_2})$, $\widehat{I}_{\eta\theta} = \widehat{I}_{\theta\eta}'$, and

$$\widehat{I}_{\eta\eta} = \begin{pmatrix} \widehat{I}_{\gamma_1\gamma_1} & \widehat{I}_{\gamma_1\gamma_2} \\ \widehat{I}_{\gamma_2\gamma_1} & \widehat{I}_{\gamma_2\gamma_2} \end{pmatrix}.$$

We use (16) as our estimator for $\widetilde{I}_0$. The next result shows that this estimator is consistent.

Theorem 3.2. *Under Conditions* M1–M6 *in the Appendix, we have* $\widehat{I} \xrightarrow{P} \widetilde{I}_0$.

We call the following Condition R1 as the *regularity conditions* for the semiparametric copula models in consideration:

R1. $C(s_1, s_2;\theta)$, $C_1(s_1, s_2;\theta)$, $C_2(s_1, s_2;\theta)$ and $C_{12}(s_1, s_2;\theta)$ are twice continuously differentiable w.r.t. (s_1, s_2, θ) when θ is around θ_0.

Note that the true baseline hazard functions λ_j are bounded away from zero since $\lambda_j = \exp(h_j)$ and h_j belongs to some Hölder ball with bounded supremum norm. Under R1 and some other mild conditions, we can show that Conditions M1–M6 are satisfied; see Appendices A.2 and A.4 for more details. In fact, we verify R1 in seven commonly used copula functions, including Gaussian, Gumbel and Clayton copulas, listed in Appendix B of Chen et al. [3] which gives explicit expressions of those copulas and their derivatives.

The rigorous proof of Theorem 3.2, which heavily relies on the empirical processes theory, will be given in Appendix A.5. Here, we give some theoretical insight why the result holds without resorting to the empirical processes tool. First we observe from (11) and (13) that

$$\lambda'\widetilde{I}_0^{-1}\lambda = \sup_{v\in\mathbf{V}:\|v\|>0} \frac{|\lambda' v_\theta|^2}{\|v\|^2}.$$

Thus, we expect $\widehat{I}$ to be consistent if it can be characterized as

$$\lambda'\widehat{I}^{-1}\lambda = \sup_{(v_\theta,\gamma_1,\gamma_2)\in\mathbb{R}^p\times\mathbb{R}^{K_1}\times\mathbb{R}^{K_2}} \left\{\frac{|\lambda' v_\theta|^2}{\|v\|_n^2}\right\}, \tag{17}$$

where $\|\cdot\|_n$ is the estimated Fisher norm, i.e.,

$$\|v\|_n^2 = \frac{1}{n}\sum_{i=1}^n \{v_\theta'\dot{\ell}_\theta(X_i;\widehat{\alpha}) + \gamma_1'\dot{\ell}_{\gamma_1}(X_i;\widehat{\alpha}) + \gamma_2'\dot{\ell}_{\gamma_2}(X_i;\widehat{\alpha})\}^2. \tag{18}$$

Note that (18) can be further written as

$$v_\theta'\widehat{I}_{\theta\theta}v_\theta + 2v_\theta'\widehat{I}_{\theta\eta}\begin{pmatrix}\gamma_1\\ \gamma_2\end{pmatrix} + (\gamma_1', \gamma_2')\widehat{I}_{\eta\eta}\begin{pmatrix}\gamma_1\\ \gamma_2\end{pmatrix}. \tag{19}$$

To verify that $\widehat{I}$ indeed satisfies (17), we rewrite the right hand side of (17) as

$$\sup_{v_\theta}\left\{\frac{|\lambda' v_\theta|^2}{\inf_{(\gamma_1,\gamma_2)}\|v\|_n^2}\right\}. \tag{20}$$

According to (19), for fixed v_θ, $\|v\|_n^2$ is minimized when $(\gamma_1', \gamma_2')' = -\widehat{I}_{\eta\eta}^{-1}\widehat{I}_{\eta\theta}v_\theta$ with the minimum value being $v_\theta'\widehat{I}v_\theta$. Thus (20) becomes $\sup_{v_\theta}\{|\lambda' v_\theta|/v_\theta'\widehat{I}v_\theta\}$. An application of the Cauchy–Schwarz inequality similar to (13) confirms (17).

We next provide an intuitive explanation on the consistency of $\widehat{I}$. Define a B-spline estimate $(\bar{\gamma}_j^*)'\mathbf{B}_j$ of $\bar{w}_j^*$ (i.e., the characteristics of the LFS), where $[\bar{\gamma}_j^*]_{K_j\times p} = (\gamma_{j1}^*, \ldots, \gamma_{jp}^*)$, $((\gamma_{1k}^*)', (\gamma_{2k}^*)')' = -\widehat{I}_{\eta\eta}^{-1}\widehat{I}_{\eta\theta}1_k$ and 1_k represents the p-vector with its k-th element as one and others as zeros, for $j = 1, 2$ and $k = 1, \ldots, p$. We can rewrite (16) as the following explicit form

$$\widehat{I} = \frac{1}{n}\sum_{i=1}^{n}\left[\dot{\ell}_\theta(X_i; \widehat{\alpha}) + \sum_{j=1}^{2}\dot{\ell}_{h_j}(X_i; \widehat{\alpha})[(\bar{\gamma}_j^*)'\mathbf{B}_j]\right]^{\otimes 2}. \tag{21}$$

Thus the above $\widehat{I}$ can be viewed as a plug-in estimate of the efficient information $\widetilde{I}_0$ defined in (12). We expect $\widehat{I}$ to be consistent if $\widehat{\alpha}$ is consistent and $(\bar{\gamma}_j^*)'\mathbf{B}_j$ is a consistent estimate of $\bar{w}_j^*$ which is implicitly required to be smooth. The latter consistency could be verified under our Conditions M1–M6.

3.4. *Asymptotic properties of marginal cumulative hazard estimator*

Our asymptotic analysis for the copula estimator in Section 3.2 can be generalized to the other estimators which can be expressed as smooth functional of the B-spline estimator, i.e., $\rho(\widehat{\alpha})$, according to Lemma A.1 in the Appendix. This is the case for the cumulative hazard estimator $\widehat{\Lambda}_j(y_j)$ in this section, which can be written as a smooth transformation of $\widehat{\alpha}$, say $\rho_{\Lambda_j}(\widehat{\alpha}) = \int_0^{y_j}\exp(\widehat{h}_j(x))dx = \widehat{\Lambda}_j(y_j)$. Following similar discussions in Section 3.2, we expect the B-spline estimate $\widehat{\Lambda}_j(y_j)$ to be asymptotically normal and semiparametric efficient with the asymptotic variance

$$V_j = \sup_{v\in\mathbf{V}:\|v\|>0}\frac{\{G_{j0}[v_j](y_j)\}^2}{\|v\|^2},$$

where $G_{j0}[v_j](y_j) = \int_0^{y_j}\exp(h_{j0}(x))v_j(x)dx$. The asymptotic normality and semiparametric efficiency of $\widehat{S}_j(y_j)$ trivially follows since it is a smooth transformation of $\widehat{\Lambda}_j(y_j)$.

Theorem 3.3. *Suppose that Conditions* M1–M2 *in the* Appendix *hold and* $V_j < \infty$. *Then for any fixed* $y_j \in (0, 1)$, *we have*

$$\sqrt{n}(\widehat{\Lambda}_j(y_j) - \Lambda_{j0}(y_j)) \xrightarrow{d} N(0, V_j),$$
$$\sqrt{n}(\widehat{S}_j(y_j) - S_{j0}(y_j)) \xrightarrow{d} N\left(0, S_{j0}^2(y_j)V_j\right).$$

Moreover, $\widehat{\Lambda}_j(y_j)$ *and* $\widehat{S}_j(y_j)$ *are semiparametric efficient.*

In general, there is no closed-form for V_j even for the Gaussian copula and independence copula. However, we are able to provide an explicit B-spline estimate for V_j given below. For simplicity, we assume $j = 1$ for now. Following similar derivations for $\widehat{I}$ in Section 3.3, we know that a consistent information estimator of γ_1 is

$$\widehat{I}_{\gamma_1} = \widehat{I}_{\gamma_1\gamma_1} - (\widehat{I}_{\gamma_1\theta}, \widehat{I}_{\gamma_1\gamma_2})\begin{pmatrix}\widehat{I}_{\theta\theta} & \widehat{I}_{\theta\gamma_2}\\ \widehat{I}_{\gamma_2\theta} & \widehat{I}_{\gamma_2\gamma_2}\end{pmatrix}^{-1}\begin{pmatrix}\widehat{I}_{\theta\gamma_1}\\ \widehat{I}_{\gamma_2\gamma_1}\end{pmatrix}.$$

Note that $\widehat{\Lambda}_1(y_1)$ is proven to be efficient and is also a function of $\widehat{\gamma}_1$, i.e., $\widehat{\Lambda}_1(y_1) = \int_0^{y_1}\exp(\widehat{\gamma}_1'\mathbf{B}_1(x))dx$. This implies the asymptotic variance estimator, i.e., the Cramér–Rao lower bound estimator, for $\widehat{\Lambda}_j(y_j)$ is of the following form:

$$\widehat{V}_1 = \widehat{G}_{1n}'[\mathbf{B}_1]\widehat{I}_{\gamma_1}^{-1}\widehat{G}_{1n}[\mathbf{B}_1],$$

where $\widehat{G}_{1n}[\mathbf{B}_1] = (\widehat{G}_{1n}[B_{11}](y_1), \ldots, \widehat{G}_{1n}[B_{1K_1}](y_1))'$ and $\widehat{G}_{1n}[B_{1k}](y_1) = \int_0^{y_1}\exp(\widehat{h}_1(s))B_{1k}(s)ds$. We expect $\widehat{V}_1$ to be consistent if $\widehat{\alpha}$ is consistent. The asymptotic variance estimator for $\widehat{S}_j(y_j)$ is simply $\exp[-2\int_0^{y_j}\exp(\widehat{\gamma}_j'\mathbf{B}_j(s))ds]\widehat{V}_j$.

4. Simulation studies

We performed a series of simulation studies to examine the finite sample properties of the proposed semiparametric efficient estimator (and its asymptotic covariance matrix estimate (16)) and to compare it with the two-step estimator of Chen et al. [3]. In all simulation setups, the marginal distributions of T_1 and T_2 were specified as exponential distribution with unit mean, and the censoring distribution was chosen to be exponential with mean 2. Two censoring mechanisms were used for the simulation setups: (1) both variables were censored at the same censoring time; (2) the variables were censored independently from the same censoring distribution. For each simulation run, we randomly simulated 100 subjects. The experiment for each simulation setup was repeated 400 times.

Now we give the details of the copula model specification for the simulation setups. We use the following convention: when we refer to a named copula function, we mean that the named copula function is applied to the marginal distribution

Table 1
Summary of simulation results comparing the two-step procedure (2-step) and the efficient joint estimation (joint). Reported are the mean squared errors (MSEs) of estimating the copula parameter (θ), Spearman's ρ, Kendall's τ, and the mean integrated squared errors (MISEs) of the hazard functions. The MSEs for θ marked with * and the MSEs for ρ and τ are the actual numbers multiplied by 100.

Setup	Method	Euclidean parameter			Hazard functions	
		θ	ρ	τ	Func. # 1	Func. #2
1	2-step	0.214*	0.231	0.231	11.010	10.044
	Joint	0.178*	0.193	0.205	5.019	4.594
2	2-step	0.236*	0.254	0.254	11.010	10.351
	Joint	0.195*	0.211	0.220	4.162	3.840
3	2-step	0.163	0.199	0.242	10.089	9.585
	Joint	0.163	0.133	0.180	4.774	4.646
4	2-step	0.168	0.175	0.223	10.089	10.023
	Joint	0.170	0.132	0.182	4.299	4.117
5	2-step	0.154	0.145	0.190	10.420	8.713
	Joint	0.159	0.107	0.155	5.153	4.719
6	2-step	0.169	0.159	0.208	10.420	9.938
	Joint	0.180	0.119	0.172	5.024	4.817
7	2-step	0.374	0.200	0.216	10.592	9.875
	Joint	0.298	0.135	0.154	5.295	5.187
8	2-step	0.397	0.213	0.229	10.592	10.281
	Joint	0.321	0.142	0.163	5.127	5.323

functions to define the bivariate joint distribution function; when we use the prefix "survival" on a named copula function, we mean that the named copula function is applied to the marginal survival functions to define the bivariate joint survival function.

- Setup 1. Gaussian copula, same censoring time. The copula function has the form

  $$C(u_1, u_2; \theta) = \Psi(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \theta),$$

  where Ψ is the distribution function of bivariate standard normal with correlation coefficient θ and $\Phi(\cdot)$ is the distribution function of the standard normal distribution. The copula parameter used is $\theta = 0.8$. The corresponding values of Spearman's ρ and Kendall's τ are 0.786 and 0.590.
- Setup 2. Gaussian copula, independent censoring. This setup is the same as Setup 1, except that independent censoring is used.
- Setup 3. Gumbel copula, same censoring time. The copula function has the form

  $$C(u_1, u_2; \theta) = \exp[-\{(-\log u_1)^\theta + (-\log u_2)^\theta\}^{1/\theta}].$$

  The copula parameter used is $\theta = 3$. The corresponding values of Spearman's ρ and Kendall's τ are 0.849 and 0.667.
- Setup 4. Gumbel copula, independent censoring. This setup is the same as Setup 3, except that independent censoring is used.
- Setup 5. Survival Gumbel copula, same censoring time. This setup is the same as Setup 3, except that the Gumbel copula is applied on the marginal survival functions to define the joint survival function.
- Setup 6. Survival Gumbel copula, independent censoring. This setup is the same as Setup 5, except that independent censoring is used.
- Setup 7. Survival Clayton copula, same censoring time. The copula function has the form

  $$C(u_1, u_2; \theta) = \{(u_1^{-\theta} + u_2^{-\theta}) - 1\}^{-\theta^{-1}}.$$

  The copula parameter used is $\theta = 3$. The corresponding values of Spearman's ρ and Kendall's τ are 0.786 and 0.600.
- Setup 8. Survival Clayton copula, independent censoring. This setup is the same as Setup 7, except that independent censoring is used.

Both the two-step procedure and our efficient joint estimation procedure were applied to each simulated data set. Table 1 reports the mean squared errors (MSEs) of estimation of the copula parameter θ, Spearman's ρ, Kendall's τ, and the mean integrated squared errors of estimation of the marginal hazard functions. For estimating ρ and τ, the joint estimation always yields smaller MSEs. For estimating θ, the MSEs of the joint estimation are not always smaller. We think the asymptotic for estimating θ has not kicked in for the sample size in consideration; when we increased the sample size, we observed that the MSEs of the joint estimation become smaller than the 2-step procedure. On the other hand, the advantage of the joint estimation is profound in the estimation of the marginal hazard functions. The reduction of mean integrated squared errors is more than 50% in most setups. Table 2 shows that the average of estimated variance are reasonably close to the Monte Carlo variance of the copula parameter estimates. The estimated variance is biased towards giving conservative inference, but this bias gradually goes away when the sample increases.

Table 2
Comparison of estimated variances (est var) and Monte Carlo variances (var) of the copula parameter estimates. The results for setups 1 and 2 are multiplied by 100.

	Setup 1	Setup 2	Setup 3	Setup 4	Setup 5	Setup 6	Setup 7	Setup 8
Var	0.178	0.195	0.155	0.164	0.139	0.156	0.252	0.272
Est var	0.226	0.250	0.183	0.200	0.155	0.167	0.290	0.310

Acknowledgments

Chen's research was partially sponsored by NSF (SES-0838161). Cheng's research was sponsored by NSF (DMS-0906497 and CAREER Award DMS-1151692). Huang's research was partly sponsored by NSF (DMS-0907170, DMS-1007618), and Award Number KUS-CI-016-04, made by King Abdullah University of Science and Technology (KAUST). Zhou's research was partially sponsored by NSF (DMS-0907170). The authors thank the editor, the associate editor, and one referee for insightful comments that led to important improvements in the paper.

Appendix

A.1. Asymptotic normality and efficiency of smooth functionals

We present a general theory on the asymptotic normality and semiparametric efficiency of $\rho(\widehat{\alpha})$, where $\rho : \mathcal{A} \mapsto \mathbb{R}^1$ is a sufficiently smooth functional.

We first introduce some necessary notations. Under the assumption that $\ell(\alpha)$ is second order differentiable around α_0, we define the second order directional derivative $\ddot{\ell}(\alpha)[v_a, v_b]$ as $(d/dt)|_{t=0}\dot{\ell}(\alpha_0 + tv_b)[v_a]$. The Fisher inner product and Fisher norm on the space $\mathbf{V}$ are defined as

$$\langle v, \widetilde{v} \rangle = E\left[\dot{\ell}(\alpha_0)[v]\dot{\ell}(\alpha_0)[\widetilde{v}]\right] \quad \text{and} \quad \|v\|^2 = \langle v, v \rangle , \tag{A.1}$$

respectively. Of course, we know that $\|v\|^2$ also equals $-E\ddot{\ell}(\alpha_0)[v, v]$. We form a Hilbert space $(\overline{\mathbf{V}}, \|\cdot\|)$ by defining $\overline{\mathbf{V}}$ to be the completion of $\mathbf{V}$ under the Fisher norm. It is easy to show that $\overline{\mathbf{V}} = \{v = (v_\theta', v_1, v_2)' \in \mathbb{R}^p \times L_2([0, 1]) \times L_2([0, 1]) : \|v\| < \infty\}$, where $L_2([0, 1]) = \{v(\cdot) : \int_0^1 v^2(s)ds < \infty\}$.

For any $v \in \mathbf{V}$ and any smooth functional $\rho(\cdot)$, we define

$$\dot{\rho}(\alpha_0)[v] = \lim_{t\to 0} \frac{\rho(\alpha_0 + tv) - \rho(\alpha_0)}{t} \quad \text{and} \quad \|\dot{\rho}(\alpha_0)\| = \sup_{v\in\mathbf{V}:\|v\|>0} \frac{|\dot{\rho}(\alpha_0)[v]|}{\|v\|}. \tag{A.2}$$

We assume that $v \mapsto \dot{\rho}(\alpha_0)[v]$ is a linear functional, and also assume the following smoothness condition S on $\rho(\cdot)$.

S. (a) For any $v \in \mathbf{V}$, $\rho(\alpha_0 + tv)$ is continuously differentiable for small t and $\|\dot{\rho}(\alpha_0)\| < \infty$.
(b) There exist constants $\omega > 0$ and $\varepsilon > 0$ such that, for any $v \in \mathbf{V}$ with $\|v\| \le \varepsilon$,

$$|\rho(\alpha_0 + v) - \rho(\alpha_0) - \dot{\rho}(\alpha_0)[v]| = O(\|v\|^{\omega}),$$
$$\|\widehat{\alpha} - \alpha_0\|^{\omega} = o_P(n^{-1/2}).$$

The Riesz representation theorem implies that there exists a $v^* = ((v_\theta^*)', v_1^*, v_2^*)' \in \overline{\mathbf{V}}$ such that

$$\langle v^*, v \rangle = \dot{\rho}(\alpha_0)[v] \quad \text{for all } v \in \mathbf{V} \quad \text{and} \quad \|v^*\| = \|\dot{\rho}(\alpha_0)\|. \tag{A.3}$$

The concrete form of v^* depends on the form of $\rho(\cdot)$.

We assume the following convergence rate condition M1 and semiparametric copula model condition M2 in Lemma A.1. Define $\Pi_n v^*$ and $\Pi_n \alpha_0$ respectively as the projection of v^* and α_0 onto $\mathcal{A}_n$ in terms of the Fisher inner product defined above. Let $\mathbb{G}_n f = (1/\sqrt{n})\sum_{i=1}^n (f(X_i) - Ef)$.

M1. We assume that $\|\Pi_n v^* - v^*\| = o(1)$ and $\|\widehat{\alpha} - \alpha_0\| = o_P(\delta_n)$ for some $\delta_n = o(n^{-1/3})$ satisfying $\delta_n \times \|\Pi_n v^* - v^*\| = o(n^{-1/2})$.

M2. Let $\mathcal{A}_n(\delta_n) = \{\alpha_n \in \mathcal{A}_n : \|\alpha_n - \alpha_0\| \le \delta_n\}$, where δ_n is defined in M1. We assume that

$$\sup_{\alpha_n\in\mathcal{A}_n(\delta_n)} \left|\mathbb{G}_n\left(\dot{\ell}(\alpha_n)[\Pi_n v^*] - \dot{\ell}(\alpha_0)[\Pi_n v^*]\right)\right| = o_P(1), \tag{A.4}$$

and

$$E\{\ddot{\ell}(\alpha_n)[v_n, v_n] - \ddot{\ell}(\alpha_0)[v_n, v_n]\} = O(\delta_n^3), \tag{A.5}$$

for all $\alpha_n \in \mathcal{A}_n(\delta_n)$ and $v_n \in \mathbf{V}$ satisfying $\|v_n\| = O(\delta_n)$.

Lemma A.1. *Suppose that Conditions* S *and* M1–M2 *hold. Then, we have*

$$\sqrt{n}(\rho(\widehat{\alpha}) - \rho(\alpha_0)) \xrightarrow{d} N\left(0, \|\dot{\rho}(\alpha_0)\|^2\right),$$

and $\rho(\widehat{\alpha})$ is semiparametrically efficient.

The proof of Lemma A.1 is completely analogous to that of Theorem 1 in [4], which applies the general theory of Shen [21] to semiparametric copula models, and is thus skipped.

A.2. Verifications of conditions M1–M2

To verify the convergence rate of $\widehat{\alpha}$ in M1, we can apply Theorems 1–2 in [22]. In fact, the conditions C1–C2 of their Theorem 1 are easily satisfied with $\alpha = \beta = 1$ since we use the convenient Fisher norm and $\ell(\alpha)$ is bounded in the present setting. We also know that the ϵ-entropy number of the class of function $\{\ell(\alpha) - \ell(\Pi_n\alpha_0) : \alpha \in \mathcal{A}_n\}$ is $O((K_1 \vee K_2)\log(1 + (c_1 \vee c_2)/\epsilon))$ (in terms of L_∞-norm) according to Lemma 2.5 of van de Geer [25]. This yields that $\delta_n = O((n^{-1/2}K_1^{1/2} \vee K_1^{-r_1}) \vee (n^{-1/2}K_2^{1/2} \vee K_2^{-r_2}))$ in view of (5). Therefore, by choosing

$$K_j \asymp n^{1/(1+2r_j)} \quad \text{for } j = 1, 2, \tag{A.6}$$

we have $\delta_n = O(n^{-r/(2r+1)})$, where $r = r_1 \wedge r_2$. The range that $r_j > 1$ implies $1/3 < r/(2r+1) < 1/2$.

To derive the rate of $\|\Pi_n v^* - v^*\|$ needed in Condition M1, we usually need to show that the Riesz representer v^* has some smoothness. As discussed in Appendix A.1, the form (and also the property) of v^* depends on that of $\rho(\cdot)$. For illustrative purpose, we focus on the case that $\rho = \rho_\lambda(\alpha) = \lambda'\theta$, which corresponds to Section 3.2 on the copula parameter; see Proof of Theorem 3.1 below. Specifically, we shall show that w_{jk}^* solves a system of integral equations through which its smoothness is implied by reasonable assumptions on the underlying density of X, the form of the copula function and the smoothness of h_{j0}'s. Recall that (w_{1k}^*, w_{2k}^*) is defined as the minimizer of

$$(w_{1k}, w_{2k}) \mapsto E\left\{[\dot{\ell}_\theta(\alpha_0)]_k + \sum_{j=1}^{2} \dot{\ell}_{h_j}(\alpha_0)[w_{jk}]\right\}^2.$$

Let e_k denote a p-vector whose k-th element is 1 and other elements are zero. The above minimization problem can be restated in terms of the Fisher norm as follows:

$$(w_{1k}^*, w_{2k}^*) = \underset{(w_{1k}, w_{2k})}{\operatorname{argmin}} \|(e_k, 0, 0) - (0, w_{1k}, w_{2k})\|^2.$$

By the Hilbert projection theorem, (w_{1k}^*, w_{2k}^*) exists and is unique in the topology induced by the Fisher norm. According to the positive information assumption stated in Section 3.2, the Fisher norm is stronger than the L_2 norm and thus (w_{1k}^*, w_{2k}^*) exists and is unique in L_2.

Clearly, (w_{1k}^*, w_{2k}^*) solves the following stationary equation:

$$E\left\{\left([\dot{\ell}_\theta(\alpha_0)]_k + \sum_{j'=1}^{2} \dot{\ell}_{h_{j'}}(\alpha_0)[w_{j'k}^*]\right) \sum_{j=1}^{2} \dot{\ell}_{h_j}(\alpha_0)[w_{jk}]\right\} = 0 \tag{A.7}$$

for any $w_{1k}, w_{2k} \in L_2([0, 1])$ and $k = 1, 2, \ldots, p$. In general, (A.7) corresponds to the system of integral equations that involves (w_{1k}^*, w_{2k}^*). For notational simplicity, we write

$$\begin{aligned}\dot{\ell}_{h_j}(\alpha)[w_{jk}] &= w_{jk}(y_j)\delta_j - \int_0^{y_j}\left[\delta_j + \frac{\partial H(s_1, s_2; \theta)}{\partial s_j}s_j\right]\exp(h_j(x))w_{jk}(x)dx \\ &\equiv w_{jk}(y_j)\delta_j - \int_0^{y_j} A_j(x)w_{jk}(x)dx.\end{aligned}$$

Denote $[\dot{\ell}_\theta(\alpha_0)]_k$ as L_k. We next analyze each term of the expansion in (A.7) as follows:

$$E\{L_k w_{jk}\delta_j\} = E\{w_{jk}E(L_k\delta_j|Y_j)\} = \int_0^1 w_{jk}(y_j)E(L_k\delta_j|y_j)f_{Y_j}(y_j)dy_j \equiv \int_0^1 w_{jk}(x)B_j(x)dx,$$

$$\begin{aligned}E\left\{L_k\int_0^{Y_j} A_j(x)w_{jk}(x)dx\right\} &= E\left\{\int_0^{Y_j} A_j(x)w_{jk}(x)dxE(L_k|Y_j)\right\} \\ &= \int_0^1\left(\int_0^1 E(L_k|y_j)1\{y_j \geq x\}f_{Y_j}(y_j)dy_j\right)A_j(x)w_{jk}(x)dx \\ &\equiv \int_0^1 w_{jk}(x)C_j(x)dx,\end{aligned}$$

$$E\{w_{1k}^*\delta_1 w_{1k}\delta_1\} = \int_0^1 f_{Y_1}(x)E(\delta_1|Y_1 = x)w_{1k}^*(x)w_{1k}(x)dx \equiv \int_0^1 D_{11}(x)w_{1k}^*(x)w_{1k}(x)dx,$$

$$\begin{aligned} E\{w_{1k}^*\delta_1 w_{2k}\delta_2\} &= \int_0^1\int_0^1 w_{1k}^*(x)w_{2k}(\bar{x})E(\delta_1\delta_2|Y_1 = x, Y_2 = \bar{x})f_{Y_1,Y_2}(x,\bar{x})d\bar{x}dx \\ &\equiv \int_0^1\int_0^1 w_{1k}^*(x)w_{2k}(\bar{x})D_{12}(x,\bar{x})d\bar{x}dx, \end{aligned}$$

$$\begin{aligned} E\left\{w_{1k}^*\delta_1\int_0^{Y_1} A_1(x)w_{1k}(x)dx\right\} &= \int_0^1\int_0^1 w_{1k}^*(y_1)E(\delta_1|y_1)f_{Y_1}(y_1)1\{x \le y_1\}A_1(x)w_{1k}(x)dxdy_1 \\ &\equiv \int_0^1\int_0^1 w_{1k}^*(x)w_{1k}(\bar{x})E_{11}(x,\bar{x})d\bar{x}dx, \end{aligned}$$

$$\begin{aligned} E\left\{w_{1k}^*\delta_1\int_0^{Y_2} A_2(x)w_{2k}(x)dx\right\} &= \int_0^1\int_0^1 w_{1k}^*(y_1)w_{2k}(x)A_2(x) \\ &\quad\times\left(\int_0^1 f_{Y_1,Y_2}(y_1,y_2)E(\delta_1|Y_1 = y_1, Y_2 = y_2)1\{x \le y_2\}dy_2\right)dxdy_1 \\ &\equiv \int_0^1\int_0^1 w_{1k}^*(x)w_{2k}(\bar{x})E_{12}(x,\bar{x})d\bar{x}dx, \end{aligned}$$

$$\begin{aligned} E\left\{\int_0^{Y_1} A_1(x)w_{1k}^*(x)dx\int_0^{Y_1} A_1(x)w_{1k}(x)dx\right\} &= \int_0^1\int_0^1 w_{1k}^*(x)w_{1k}(\bar{x})P(Y_1 \ge x \vee \bar{x})A_1(x)A_1(\bar{x})d\bar{x}dx \\ &= \int_0^1\int_0^1 w_{1k}^*(x)w_{1k}(\bar{x})F_{11}(x,\bar{x})d\bar{x}dx, \end{aligned}$$

$$\begin{aligned} E\left\{\int_0^{Y_1} A_1(x)w_{1k}^*(x)dx\int_0^{Y_2} A_2(x)w_{2k}(x)dx\right\} &= \int_0^1\int_0^1 w_{1k}^*(x)w_{2k}(\bar{x})P(Y_1 \ge x, Y_2 \ge \bar{x})A_1(x)A_2(\bar{x})d\bar{x}dx \\ &= \int_0^1\int_0^1 w_{1k}^*(x)w_{2k}(\bar{x})F_{12}(x,\bar{x})d\bar{x}dx, \end{aligned}$$

where f_{Y_j} is the density function for Y_j and $f_{Y_1,Y_2}(y_1, y_2)$ is the joint density for (Y_1, Y_2). We can also define D_{22}, E_{21}, E_{22} and F_{22} for other cross terms in similar fashion. In summary, (A.7) is expanded as

$$\begin{aligned} &\int_0^1[w_{1k}(B_1 - C_1) + w_{2k}(B_2 - C_2)]dx + \int_0^1[w_{1k}^*w_{1k}D_{11} + w_{2k}^*w_{2k}D_{22}]dx \\ &\quad+\int_0^1\int_0^1[w_{1k}^*w_{2k} + w_{1k}w_{2k}^*]D_{12}d\bar{x}dx - \int_0^1\int_0^1[(w_{1k}w_{2k}^*E_{21}) + (w_{1k}^*w_{2k}E_{12})]d\bar{x}dx \\ &\quad-\int_0^1\int_0^1[w_{1k}^*w_{1k}E_{11} + w_{2k}^*w_{2k}E_{22}]d\bar{x}dx - \int_0^1\int_0^1[w_{1k}w_{1k}^*E_{11} + w_{1k}w_{2k}^*E_{12}]d\bar{x}dx \\ &\quad+\int_0^1\int_0^1[w_{1k}^*w_{1k}F_{11} + w_{1k}w_{2k}^*F_{12}]d\bar{x}dx - \int_0^1\int_0^1[w_{2k}w_{1k}^*E_{21} + w_{2k}w_{2k}^*E_{22}]d\bar{x}dx \\ &\quad+\int_0^1\int_0^1[w_{1k}^*w_{2k}F_{12} + w_{2k}^*w_{2k}F_{22}]d\bar{x}dx = 0, \end{aligned} \tag{A.8}$$

where we have used the shortened notations, e.g., $\int_0^1\int_0^1 w_{1k}^*(x)w_{2k}(\bar{x})D_{12}(x,\bar{x})d\bar{x}dx$.

Since (A.8) holds for any $w_{1k}, w_{2k} \in L([0, 1])$, we can show that

$$(B_1 - C_1) + w_{1k}^*D_{11} = \int_0^1 w_{2k}^*(E_{21} + E_{12} - D_{12} - F_{12})d\bar{x} + \int_0^1 w_{1k}^*(E_{11} + \bar{E}_{11} - \bar{F}_{11})d\bar{x}, \tag{A.9}$$

where $\bar{E}_{11}(x,\bar{x}) = E_{11}(\bar{x}, x)$, $\bar{F}_{11}(x,\bar{x}) = F_{11}(\bar{x}, x)$. Hence, the smoothness of w_{1k}^* is determined by that of h_{10}, h_{20}, the density for X, $(\partial/\partial\theta)|_{\alpha=\alpha_0}H(s_1, s_2; \theta)$, $(\partial/\partial s_1)|_{\alpha=\alpha_0}H(s_1, s_2; \theta)$ and $(\partial/\partial s_2)|_{\alpha=\alpha_0}H(s_1, s_2; \theta)$. In the end, by applying Lemma 5 in [5], we can figure out more primitive conditions on the above quantities in order to obtain the desired smoothness of w_{1k}^*, i.e., belongs to some Hölder ball. Note that the smoothness of w_{jk}^* is not required in the application of the above lemma (we need the fact $w_{jk}^* \in L_2([0, 1])$, though). Similar smoothness analysis also applies to w_{2k}^*.

To verify the asymptotic equicontinuity Condition (A.4) in M2, we can use Lemma 3.4.2 of van der Vaart and Wellner [26]. This boils down to calculate the bracketing entropy $H_B(\epsilon, \mathcal{G}_n, L_2(P_X))$, where $\mathcal{G}_n \equiv \{\dot{\ell}(\alpha_n)[v_n] - \dot{\ell}(\alpha_0)[v_n] : \alpha_n \in$

$\mathcal{A}_n(\delta_n), v_n \in \mathcal{A}_n\}$. Under regularity conditions R1, we have $H_B(\epsilon, \mathcal{G}_n, L_2(P_X)) = O((K_1 \vee K_2)\log(1+\delta_n/\epsilon))$. In addition, we have $Ef^2 \le \delta_n^2$, where $\delta_n = O(n^{-r/(2r+1)})$, and $\|f\|_\infty \le M < \infty$ for any $f \in \mathcal{G}_n$.

The verification of (A.5) in M2 is more tedious, and thus a set of sufficient conditions is provided here. We first write down the explicit form of $\ddot{\ell}(\alpha)[v, v]$ based on (6):

$$\ddot{\ell}(\alpha)[v,v] = v_\theta' \frac{\partial^2 H}{\partial \theta^2} v_\theta - v_\theta' \sum_{l=1}^2 \frac{\partial^2 H}{\partial\theta\partial s_l} G_l[v_l] + \sum_{i,j=1}^2 \frac{\partial^2 H}{\partial s_i \partial s_j} s_i s_j G_i[v_i] G_j[v_j] - \sum_{l=1}^2 \left(\frac{\partial H}{\partial s_l} s_l N_l[v_l] + G_l[v_l^2]\delta_l \right), \tag{A.10}$$

where $H = H(s_1, s_2; \theta)$ and $N_l[v_l] = G_l[v_l^2] - G_l^2[v_l]$. We write $\alpha_n = (\theta_n', h_{1n}, h_{2n})'$ and $v_n = (v_{\theta n}', v_{1n}, v_{2n})'$. We define $s_{l0} = \exp[-\int_0^{y_l} \exp(h_{l0}(x))dx]$ and $s_{ln} = \exp[-\int_0^{y_l} \exp(h_{ln}(x))dx]$. The same notation rule applies to $G_l[v_l]$ and $N_l[v_l]$. Based on (A.10), we have the following sufficient conditions for (A.5):

$$
\begin{aligned}
&E\left(\frac{\partial^2 K(\alpha_n)}{\partial\theta^2} - \frac{\partial^2 K(\alpha_0)}{\partial\theta^2} \right) \lesssim \|\alpha_n - \alpha_0\|,\\
&E\left(\frac{\partial^2 K(\alpha_n)}{\partial\theta\partial s_l} G_{ln}[v_{ln}] - \frac{\partial^2 K(\alpha_0)}{\partial\theta\partial s_l} G_{l0}[v_{ln}] \right) \lesssim \|v_n\| \, \|\alpha_n - \alpha_0\|,\\
&E\left(\frac{\partial^2 K(\alpha_n)}{\partial s_i \partial s_j} s_{in}s_{jn} G_{in}[v_{in}]G_{jn}[v_{jn}] - \frac{\partial^2 K(\alpha_0)}{\partial s_i\partial s_j} s_{i0}s_{j0}G_{i0}[v_{in}]G_{j0}[v_{jn}] \right) \lesssim \|v_n\|^2 \|\alpha_n - \alpha_0\|,\\
&E\left(\frac{\partial K(\alpha_{ln})}{\partial s_l} s_{ln} N_{ln}[v_{ln}] - \frac{\partial K(\alpha_0)}{\partial s_l} s_{l0} N_{l0}[v_{ln}] \right) \lesssim \|v_n\|^2 \|\alpha_n - \alpha_0\|,\\
&E\left(G_{ln}[v_{ln}^2] - G_{l0}[v_{ln}^2] \right) \lesssim \|v_n\|^2 \|\alpha_n - \alpha_0\|,
\end{aligned}
$$

where $K(\cdot) = \log C(\cdot)$, $\log C_1(\cdot)$, $\log C_2(\cdot)$ or $\log C_{12}(\cdot)$ and $\lesssim$ denotes smaller than, up to an universal constant.

A.3. Proof of Theorem 3.1

We apply Lemma A.1 to prove Theorem 3.1 by taking $\rho_\lambda(\alpha) = \lambda'\theta$ for any fixed $\lambda \in \mathbb{R}^p$ with $0 < \|\lambda\| < \infty$. Therefore, it suffices to show that Condition S holds when $\widetilde{I}_0$ is nonsingular. Condition S(b) is trivially satisfied with $\omega = \infty$ since $\rho_\lambda(\alpha)$ is a linear functional. It remains to show that $\|\dot{\rho}_\lambda(\alpha_0)\| < \infty$. This follows from the nonsingularity of $\widetilde{I}_0$ since $\|\dot{\rho}_\lambda(\alpha_0)\|^2 = \lambda' \widetilde{I}_0^{-1}\lambda$ as shown in (11) and (13) according to (A.2). □

A.4. Conditions M3–M6 and verifications

Define $\|\alpha - \alpha_0\|_s = \|\theta - \theta_0\| + \sum_{j=1}^2 \|h_j - h_{j0}\|_\infty$. Let

$$
\begin{aligned}
S_k(X; \alpha, w_k) &= [\dot{\ell}_\theta(X;\alpha)]_k + \sum_{j=1}^2 \dot{\ell}_{h_j}(X;\alpha)[w_{jk}],\\
\mathcal{N}_0 &= \{\alpha \in \mathcal{A}_n : \|\alpha - \alpha_0\|_s \le \epsilon_n\},\\
\rho_{jn} &= \max_{1\le k\le p} \inf_{f\in\mathcal{F}_{jn}} \|f - w_{jk}^*\|_\infty.
\end{aligned}
$$

M3. We assume that

$$E \sup_{w_k \in \mathcal{F}_{1n}\times\mathcal{F}_{2n}} |S_k(X;\alpha,w_k) - S_k(X;\alpha_0,w_k)|^2 \lesssim \|\alpha - \alpha_0\|_s^2 \tag{A.11}$$

for all $\alpha \in \mathcal{N}_0$,

$$\sum_j \|w_{jk}^* - w_{jk}\|_2^2 \lesssim E\{S_k(X;\alpha_0,w_k^*) - S_k(X;\alpha_0,w_k)\}^2, \tag{A.12}$$

and

$$\|\dot{\ell}_{h_j}(\alpha_0)[w_{jk}^*] - \dot{\ell}_{h_j}(\alpha_0)[w_{jk}]\|_2 \lesssim \|w_{jk}^* - w_{jk}\|_\infty, \tag{A.13}$$

where $\|\cdot\|_2$ represents the L_2 norm, for all $w_{jk} \in \mathcal{F}_{jn}$.

M4. We assume that $\|w_{jk}^*\|_\infty < c_j$ for $j = 1, 2$ and $k = 1, \ldots, p$, and $\rho_{jn} \to 0$ for $j = 1, 2$.

M5. We assume that $\epsilon_n \to 0$ in $\mathcal{N}_0$ and $E[\sup_{\alpha\in\mathcal{N}_0, w_k\in\mathcal{F}_{1n}\times\mathcal{F}_{2n}} |S_k(X;\alpha,w_k)|^2] \le$ const. $< \infty$ for any $k = 1, \ldots, p$.

M6. We assume that $\|\widehat{\alpha} - \alpha_0\|_s = o_P(1)$.

Based on the form of $S_k(X;\alpha,w_k)$ and the definition of $\|\cdot\|_s$, we can easily verify (A.11) when the regularity conditions R1 are satisfied. The assumption (A.12) is essentially the positive information assumption stated in Section 3.2. As for (A.13), it follows from the explicit form of $\dot{\ell}[\alpha](\cdot)$ and the fact that $G_j[\cdot]$ is a bounded linear operator. The verification of the convergence rates in assumption M4 follows from the smoothness result of w^*_{jk} derived in Appendix A.2. The smoothness of w^*_{jk} ensures the boundedness of w^*_{jk} but not necessarily implies $\|w^*_{jk}\|_\infty < c_j$ for the c_j used in the definition of $\mathcal{F}_j$, although we expect this holds when c_j is chosen sufficiently large. Thus the first part of assumption M4 is necessary. In practice, we can get indication of the validity of $\|w^*_{jk}\|_\infty < c_j$ by checking the spline estimate of w^*_{jk}'s, defined as in (A.16) but with the boundedness requirement removed from $\mathcal{F}_{1n}$ and $\mathcal{F}_{2n}$. The triangular inequality together with the regularity conditions R1 implies M5. We can apply Theorem 3.1 of Chen [2] to show the consistency of $\widehat{\alpha}$ in M6.

A.5. Proof of Theorem 3.2

For simplicity, we write $S_k(X;\alpha_0,w_k)$ and $S_k(X;\widehat{\alpha},w_k)$ as $S^0_k[w_k](X)$ and $\widehat{S}_k[w_k](X)$, respectively. Based on the definitions of $\widetilde{I}_0$ and $\widehat{I}$, i.e., (12) and (21), we know their (k,k')-th entries can be written as

$$\widetilde{I}_0(k,k') = ES^0_k[w^*_k]S^0_{k'}[w^*_{k'}], \tag{A.14}$$

$$\widehat{I}(k,k') = \frac{1}{n}\sum_{i=1}^n \widehat{S}_k[\widehat{w}^*_k](X_i)\widehat{S}_{k'}[\widehat{w}^*_{k'}](X_i), \tag{A.15}$$

where $\widehat{w}^*_k = ((\gamma^*_{1k})'\mathbf{B}_1, (\gamma^*_{2k})'\mathbf{B}_2)$. Note that

$$\widehat{w}^*_k = \arg\min_{w\in\mathcal{F}_{1n}\times\mathcal{F}_{2n}} \mathbb{P}_n\widehat{S}^2_k[w]. \tag{A.16}$$

Assumption M5 implies that $\{S_k(x;\alpha,w):\alpha\in\mathcal{N}_0, w\in\mathcal{F}_{1n}\times\mathcal{F}_{2n}\}$ is P-Glivenko–Cantelli. It follows from Corollary 9.27 of Kosorok [13] that, uniformly over $w_k, w_{k'}\in\mathcal{F}_{1n}\times\mathcal{F}_{2n}$,

$$\frac{1}{n}\sum_{i=1}^n \widehat{S}_k[w_k](X_i)\widehat{S}_{k'}[w_{k'}](X_i) = E\widehat{S}_k[w_k]\widehat{S}_{k'}[w_{k'}] + o_P(1). \tag{A.17}$$

Thus, uniformly over $w_k, w_{k'}\in\mathcal{F}_{1n}\times\mathcal{F}_{2n}$,

$$\begin{aligned}
&\left|E\widehat{S}_k[w_k]\widehat{S}_{k'}[w_{k'}] - ES^0_k[w_k]S^0_{k'}[w_{k'}]\right| \le E\left|\widehat{S}_k[w_k](\widehat{S}_{k'}[w_{k'}] - S^0_{k'}[w_{k'}])\right| + E\left|S^0_{k'}[w_{k'}](\widehat{S}_k[w_k] - S^0_k[w_k])\right| \\
&\quad\le \sqrt{E\widehat{S}^2_k[w_k]E\left(\widehat{S}_{k'}[w_{k'}] - S^0_{k'}[w_{k'}]\right)^2} + \sqrt{E\left(S^0_{k'}[w_{k'}]\right)^2E\left(\widehat{S}_k[w_k] - S^0_k[w_k]\right)^2} \\
&\quad\le o_P(1),
\end{aligned} \tag{A.18}$$

where the last inequality follows from assumptions (A.11) (together with the consistency of $\widehat{\alpha}$) and M5. Combining (A.17) and (A.18), we have obtained that

$$\sup_{w_k,w_{k'}\in\mathcal{F}_{1n}\times\mathcal{F}_{2n}}\left|\frac{1}{n}\sum_{i=1}^n \widehat{S}_k[w_k](X_i)\widehat{S}_{k'}[w_{k'}](X_i) - ES^0_k[w_k]S^0_{k'}[w_{k'}]\right| = o_P(1), \tag{A.19}$$

which implies that

$$\widehat{I}(k,k') = ES^0_k[\widehat{w}^*_k]S^0_{k'}[\widehat{w}^*_{k'}] + o_P(1). \tag{A.20}$$

To finish the proof, we need to introduce $\widetilde{w}^*_k \equiv \arg\min_{w_k\in\mathcal{F}_{1n}\times\mathcal{F}_{2n}} E\{S^0_k[w_k]\}^2$ as a bridge. It suffices to show that

$$ES^0_k[\widehat{w}^*_k]S^0_{k'}[\widehat{w}^*_{k'}] - ES^0_k[\widetilde{w}^*_k]S^0_{k'}[\widetilde{w}^*_{k'}] = o_P(1), \tag{A.21}$$

$$ES^0_k[\widetilde{w}^*_k]S^0_{k'}[\widetilde{w}^*_{k'}] - \widetilde{I}_0(k,k') = o_P(1). \tag{A.22}$$

By an argument similar to (A.18), we know that (A.21) holds under assumption M5 if $E\{S^0_k[\widetilde{w}^*_k] - S^0_k[\widehat{w}^*_k]\}^2 = o_P(1)$ for $k=1,\ldots,p$. We will show later that

$$E(S^0_k[\widetilde{w}^*_k] - S^0_k[\widehat{w}^*_k])^2 = E(S^0_k[\widehat{w}^*_k])^2 - E(S^0_k[\widetilde{w}^*_k])^2. \tag{A.23}$$

It follows from (A.19) and $\widetilde{w}^*_k\in\mathcal{F}_{1n}\times\mathcal{F}_{2n}$ that

$$\begin{aligned}
E(S^0_k[\widehat{w}^*_k])^2 - E(S^0_k[\widetilde{w}^*_k])^2 &= \frac{1}{n}\sum_{i=1}^n \widehat{S}^2_k[\widehat{w}^*_k](X_i) - E(S^0_k[\widetilde{w}^*_k])^2 + o_p(1), \\
&= M_n(\widehat{w}^*_k) - M(\widetilde{w}^*_k) + o_P(1).
\end{aligned}$$

Note that

$$M_n(\widehat{w}_k^*) - M(\widehat{w}_k^*) \le M_n(\widehat{w}_k^*) - M(\widetilde{w}_k^*) \le M_n(\widetilde{w}_k^*) - M(\widetilde{w}_k^*)$$

by the definitions of $\widehat{w}_k^*$ and $\widetilde{w}_k^*$. Therefore, by (A.19), we obtain $M_n(\widehat{w}_k^*) - M(\widetilde{w}_k^*) = o_P(1)$, which further implies (A.21). Similarly, to show (A.22) holds, we need only show that $E(S_k^0[\widetilde{w}_k^*] - S_k^0[w_k^*])^2 = o_P(1)$. By the definitions of $\widetilde{w}_k^*$ and w_k^*, we have

$$\begin{aligned}
E(S_k^0[\widetilde{w}_k^*] - S_k^0[w_k^*])^2 &= \inf_{g_1 \in \mathcal{F}_{1n}, g_2 \in \mathcal{F}_{2n}} E\left(\sum_{j=1}^{2} \dot{\ell}_{h_j}(X;\alpha_0)[w_{jk}^*] - \sum_{j=1}^{2} \dot{\ell}_{h_j}(X;\alpha_0)[g_j]\right)^2 \\
&\lesssim \inf_{g_1 \in \mathcal{F}_{1n}, g_2 \in \mathcal{F}_{2n}} \left\{\sum_{j=1}^{2} E\left(\dot{\ell}_{h_j}(X;\alpha_0)[w_{jk}^*] - \dot{\ell}_{h_j}(X;\alpha_0)[g_j]\right)^2\right\} \\
&\lesssim \sum_{j=1}^{2} \left\{\inf_{g_j \in \mathcal{F}_{jn}} E\left(\dot{\ell}_{h_j}(X;\alpha_0)[w_{jk}^*] - \dot{\ell}_{h_j}(X;\alpha_0)[g_j]\right)^2\right\} \\
&\lesssim \sum_{j=1}^{2} \left\{\inf_{g_j \in \mathcal{F}_{jn}} \|w_{jk}^* - g_j\|_\infty^2\right\} \\
&\lesssim O(\rho_{1n}^2 \vee \rho_{2n}^2) \to 0,
\end{aligned}$$

based on assumptions (A.13) and M4.

It remains to show that (A.23) holds. The argument is complicated by the fact that $\mathcal{F}_{jn}$ is not a linear space and so the Pythagorean theorem does not directly apply. Let $\mathcal{F}_{jn}^\dagger$ be similarly defined as $\mathcal{F}_{jn}$ except that the boundedness restriction is removed. Let $w_{jk}^\dagger$ be defined similarly as $\widetilde{w}_{jk}^*$ except that the projection is onto $\mathcal{F}_{jn}^\dagger$. By the Pythagorean theorem, (A.23) holds if $\widetilde{w}_{jk}^*$ is replaced by $w_{jk}^\dagger$. Thus we need only show that $\|w_{jk}^\dagger\|_\infty \le c_j$, since it implies that $w_{jk}^\dagger = \widetilde{w}_{jk}^*$. The smoothness of w_{jk}^* implies that it is bounded. We make the assumption that $\|w_{jk}^*\|_\infty < c_j$; this assumption is satisfied for large enough c_j. Then by the approximation theory, there exists $\widetilde{w}_{jk} \in \mathcal{F}_{jn}^\dagger$ such that $\|\widetilde{w}_{jk}\|_\infty < c_j$ and $\|\widetilde{w}_{jk} - w_{jk}^*\|_\infty = O(\rho_{jn})$. The same argument as at the end of the previous paragraph yields

$$\sum_j E(S_k^0[w_{jk}^\dagger] - S_k^0[w_{jk}^*])^2 = O(\rho_{1n}^2 \vee \rho_{2n}^2).$$

This together with Condition (A.12) implies that

$$\sum_j \|w_{jk}^\dagger - w_{jk}^*\|_2^2 = O(\rho_{1n}^2 \vee \rho_{2n}^2).$$

Therefore,

$$\sum_j \|w_{jk}^\dagger - \tilde{w}_{jk}\|_2^2 = O(\rho_{1n}^2 \vee \rho_{2n}^2).$$

Using relationship between the L_∞ and L_2 norm on a spline space (see [10]), we obtain that

$$\|w_{jk}^\dagger - \tilde{w}_{jk}\|_\infty \le \sqrt{K_j}\|w_{jk}^\dagger - \tilde{w}_{jk}\|_2 = \sqrt{K_j}O(\rho_{1n} \vee \rho_{2n}) = o(1).$$

The last equality uses $\rho_{jn} \asymp K_j^{-\alpha}$ for some $\alpha > 1/2$ [20]. Hence $\|w_{jk}^\dagger\|_\infty \le \|\tilde{w}_{jk}\|_\infty + o(1) \le c_j$, as desired. The proof is complete. □

A.6. Proof of Theorem 3.3

We need to verify Condition S in order to apply Lemma A.1. It follows from the definition of $\rho_{\Lambda_j}(\alpha) = \int_0^{y_j} \exp(h_j(x))dx$ that $\dot{\rho}_{\Lambda_j}(\alpha_0)[v] = G_{j0}[v_j](y_j) = \int_0^{y_j} \exp(h_{j0}(x))v_j(x)dx$. We can verify that $w = 2$ in Condition S(b) by the Taylor expansion. Thus, we have $\|\widehat{\alpha} - \alpha_0\|^2 = o_P(n^{-1/2})$ since we have shown that $\|\widehat{\alpha} - \alpha_0\| = O_P(n^{-r/(2r+1)})$, where $1/3 < r/(2r+1) < 1/2$. We also assume that V_j, which is exactly $\|\dot{\rho}_{\Lambda_j}(\alpha_0)\|^2$, is finite. This competes the proof. □

References

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

[2] X. Chen, Large sample sieve estimation of semi-nonparametric models, in: J.J. Heckman, E.E. Leamer. (Eds.), Handbook of Econometrics, Vol. 6B, North-Holland, Amsterdam, 2007, Chapter 76.

[3] X. Chen, Y. Fan, D. Pouzo, Z. Ying, Estimation and model selection of semiparametric multivariate survival functions under general censorship, Journal of Econometrics 157 (2010) 129–142.
[4] X. Chen, Y. Fan, V. Tsyrennikov, Efficient estimation of semiparametric multivariate copula models, Journal of the American Statistical Association 475 (2006) 1228–1240.
[5] G. Cheng, X. Wang, Semiparametric additive transformation model under current status data, Electronic Journal of Statistics 5 (2011) 1735–1764.
[6] Edward W. Frees, E.A. Valdez, Understanding relationships using copulas, North American Actuarial Journal 2 (1998) 1–25.
[7] C. Genest, K. Ghoudi, L.-P. Rivest, A semiparametric estimation procedure of dependence parameters in multivariate families of distributions, Biometrika 82 (1995) 534–552.
[8] C. Genest, B. Werker, Conditions for the asymptotic semiparametric efficiency of an omnibus estimator of dependence parameters in Copula models, in: C.M. Cuadras and J.A. Rodríguez Lallena (Eds.), Proceedings of the Conference on Distributions with Given Marginals and Statistical Modelling, 2002.
[9] P. Hougaard, Analysis of Multivariate Survival Data, Springer-Verlag, New York, 2000.
[10] J.Z. Huang, Projection estimation in multiple regression with application to functional ANOVA models, Annals of Statistics 26 (1998) 242–272.
[11] H. Joe, Multivariate Models and Dependence Concepts, Chapman & Hall/CRC, 1997.
[12] C. Klaassen, J. Wellner, Efficient estimation in the bivariate normal Copula model: normal margins are least-favorable, Bernoulli 3 (1997) 55–77.
[13] M.R. Kosorok, Introduction to Empirical Processes and Semiparametric Inference, Springer, New York, 2008.
[14] Y. Li, R.L. Prentice, X. Lin, Semiparametric maximum likelihood estimation in normal transformation models for bivariate survival data, Biometrika 95 (2008) 947–960.
[15] A.W. Marshall, I. Olkin, Families of multivariate distributions, Journal of the American Statistical Association 83 (1988) 834–841.
[16] B. Nan, X. Lin, L.D. Lisabeth, S.D. Harlow, Piecewise constant cross-ration estimation for association of age at a marker event and age at menopause, Journal of the American Statistical Association 101 (2006) 65–77.
[17] R.B. Nelsen, An Introduction to Copulas, Springer-Verlag, New York, 1999.
[18] D. Oakes, Multivariate survival distributions, Journal of Nonparametric Statistics 3 (1994) 343–354.
[19] D. Pollard, Bracketing methods in statistics and econometrics, in: W.A. Barnett, J. Powell, G.E. Tauchen (Eds.), Nonparametric and Semiparametric Methods in Econometrics and Statistics, Cambridge Uniersity Press, 1989, pp. 337–355.
[20] L. Schumaker, Spline Functions: Basic Theory, John Wiley & Sons, New York, 1981.
[21] X. Shen, On methods of sieves and penalization, The Annals of Statistics 25 (1997) 2555–2591.
[22] X. Shen, W. Wong, Convergence rate of sieve estimates, The Annals of Statistics 22 (1994) 580–615.
[23] J.H. Shih, T.A. Louis, Inferences on the association parameter in Copula models for bivariate survival data, Biometrics 51 (1995) 1384–1399.
[24] A. Sklar, Fonctions de répartition à n dimensionset leurs marges, Publications de l'Institut de Statistique de l'Université de Paris 8 (1959) 229–231.
[25] S. van de Geer, Empirical Processes in M-Estimation, Cambridge University Press, 2000.
[26] A.W. van der Vaart, J.A. Wellner, Weak Convergence and Empirical Processes: With Applications to Statistics, Springer, New York, 1996.
[27] I. van Keilegom, T.P. Hettmansperger, Inference on multivariate M-estimators based on bivariate censored data, Journal of the American Statistical Association 97 (2002) 328–336.
[28] W. Wang, A.A. Ding, On assessing the association for bivariate current status data, Biometrika 87 (2000) 879–893.
[29] W. Wang, M.T. Wells, Nonparametric estimators of the bivariate survival function under simplified structures, Biometrika 84 (1997) 863–880.
[30] A. Wienke, P. Lichtenstein, A. Yashin, A bivariate frailty model with a cure fraction for modeling familial correlations in diseases, Biometrics 59 (2003) 1178–1183.
[31] Y. Zhang, L. Hua, J. Huang, A spline-based semiparametric maximum likelihood estimation method for the Cox model with interval-censored data, Scandinavian Journal of Statistics 37 (2010) 338–354.