Econometric Theory, **23**, 2007, 414–439. Printed in the United States of America.
DOI: 10.1017/S0266466607070181

A MODEL SELECTION TEST FOR BIVARIATE FAILURE-TIME DATA

XIAOHONG CHEN
New York University

YANQIN FAN
Vanderbilt University

In this paper, we address two important issues in semiparametric survival model selection for censored data generated by the Archimedean copula family: method of estimating the parametric copulas and data reuse. We demonstrate that for selection among candidate copula models that might all be misspecified, estimators of the parametric copulas based on minimizing the selection criterion function may be preferred to other estimators. To handle the issue of data reuse, we put model selection in the context of hypothesis testing and propose a simple test for model selection from a finite number of parametric copulas. Results from a simulation study and two empirical illustrations confirm our theoretical findings.

1. INTRODUCTION

Let (X,Y) be the lifetime variables of interest with joint survival function $F^o(x,y) = \Pr(X > x, Y > y)$ and marginal survival functions $F_j^o(\cdot)$, $j = 1,2$. Let $F(x,y)$ denote the corresponding joint cumulative distribution function (c.d.f.) with marginal distributions $F_j(\cdot)$, $j = 1,2$. Assume that F_1 and F_2 are continuous. By the Sklar (1959) theorem, there exists a unique copula function C_d such that

$$F(x,y) \equiv C_d(F_1(x), F_2(y)), \tag{1}$$

where the copula function $C_d(\cdot,\cdot)$ is itself a bivariate c.d.f. on $[0,1]^2$, which captures all the scale-invariant dependence between X and Y. It follows from (1) that the representation

We thank the editor Peter Phillips, three anonymous referees, and Hal White for their comments, which greatly improved the paper. An earlier version of this paper was presented at the 2005 World Congress Meetings of the Econometric Society, the 2005 Joint Statistical Meetings, the University of Waterloo, and the University of Western Ontario. Chen acknowledges support from the National Science Foundation and the C.V. Starr Center at NYU. Fan acknowledges support from the National Science Foundation. We thank Demian Pouzo for excellent research assistance on the numerical work in this paper and Weijing Wang for providing us with the Fortran code for computing the bivariate Kaplan–Meier estimator of Dabrowska (1988). Address correspondence to Xiaohong Chen, Department of Economics, New York University, 269 Mercer Street, 7th Floor, New York, NY 10003, USA; e-mail: xiaohong.chen@nyu.edu.

 © 2007 Cambridge University Press 0266-4666/07 $15.00

https://doi.org/10.1017/S0266466607070181 Published online by Cambridge University Press

$$F^o(x,y) \equiv C^o(F_1^o(x), F_2^o(y)) \tag{2}$$

holds where $C^o(u,v) \equiv u + v - 1 + C_d(1-u, 1-v)$ is itself a copula function,[1] that is, a bivariate c.d.f. on $[0,1]^2$.

This decomposition of the joint distribution (survival) function leads naturally to a class of semiparametric bivariate distribution (survival) functions in which the marginal distribution (survival) functions are unspecified, but the copula function is parameterized as $C^o(u,v) = C(u,v;\alpha_o)$ for some parametric copula function $C(u,v;\alpha)$ and some unknown parameter α_o. Commonly used parametric copulas include the Gaussian copula, Student's t copula, and the Clayton, Frank, and Gumbel copulas; see Joe (1997) and Nelsen (1999) for properties of many parametric copulas.

Given the existence of a large number of parametric copulas, this class of semiparametric bivariate distribution (survival) functions is able to model many different types of dependence, allowing for any type of marginal behaviors. It has proved useful in diverse fields. Specific applications include those in finance and insurance (e.g., Frees and Valdez, 1998; Embrechts, McNeil, and Straumann, 2002); in survival analysis (e.g., Clayton, 1978; Joe, 1997; Nelsen, 1999; Oakes, 1982, 1989); in econometrics (e.g., Lee, 1983; Heckman and Honoré, 1989; Granger, Teräsvirta, and Patton, 2006; Patton, 2004, 2006; Hu, 2006), to name only a few.

Because the parametric copula function $C(u,v;\alpha_o)$ captures all the dependence between X and Y, it is important to estimate the true but unknown copula parameter α_o in any application of this class of semiparametric bivariate distribution (survival) functions. The most widely used estimator of the copula parameter α_o is the two-step estimator proposed by Oakes (1994) and Genest, Quessy, and Rémillard (2006) for complete data and by Shih and Louis (1995) for censored data. In this approach, one first estimates the unknown marginal distribution (survival) functions by the rescaled empirical marginal c.d.f.s (the Kaplan–Meier estimators) and then estimates α_o by maximizing the estimated log-likelihood function. Genest, Ghoudi, and Rivest (1995) and Shih and Louis (1995) establish the root-n consistency and asymptotic normality of the two-step estimator for a random sample and censored data of size n, respectively. The two-step estimator and its large-sample properties have been extended to time series settings in Chen and Fan (2006a, 2006b). Earlier papers that propose specific estimators of the copula parameter α_o for specific parametric copula models include Clayton (1978), Clayton and Cuzick (1985), Oakes (1982, 1986), and Genest (1987). Except for the Gaussian copula and the independence copula, the two-step estimator of the copula parameter α_o is generally inefficient; see Genest and Werker (2002). Recently, Chen, Fan, and Tsyrennikov (2006) have proposed a sieve maximum likelihood estimation (MLE) procedure for all the unknown parameters in a general semiparametric multivariate copula model. For independent and identically distributed (i.i.d.) data without censoring, they show that the sieve MLEs of α_o, F_1, and F_2 are all asymptotically efficient.

https://doi.org/10.1017/S0266466607070181 Published online by Cambridge University Press

One important issue in applying this class of semiparametric distribution (survival) function models to a given data set is how to choose an appropriate parametric copula. Effort has been made to resolve this issue. For complete data, Chen, Fan, and Patton (2003), Fermanian (2005), and Genest et al. (2006), among others, develop goodness-of-fit tests for the correct specification of a parametric copula. For censored data, Wang and Wells (2000) propose a goodness-of-fit test for the correct specification of a parametric Archimedean copula, extending existing results of Genest and Rivest (1993) and Barbe, Genest, Ghoudi, and Bruno (1996) for complete data. One drawback of these tests is that if the null hypothesis of correct parametric specification is rejected, they provide no guidance as to which copula model to use.

In addition to the goodness-of-fit test, Wang and Wells (2000) also propose a model selection procedure based on comparing estimates of the integrated squared difference between the true Archimedean copula and a parametric copula for different parametric Archimedean copula models; the one with the smallest value of the integrated squared difference is chosen over the rest of the models. For example, Wang and Wells (2000) fit the Clayton, Gumbel, Frank, and log-copula models to the data set in McGilchrist and Aisbett (1991) from a study of the recurrence time of infection in kidney patients using a portable dialysis machine. They find that the corresponding values ($\times 10^{-4}$) of the integrated square difference are 21.72, 11.23, 12.44, and 16.14 and conclude that the Gumbel model provides the best fit to the data. Note that to compute the values of the integrated squared difference one needs to estimate the copula parameter associated with each of the parametric Archimedean copula models. In the preceding application, instead of estimating the copula parameter by minimizing the integrated squared difference over all possible values of the parameter, Wang and Wells (2000) estimate the copula parameter of each Archimedean copula model by inverting Kendall's tau (τ). As such, the corresponding estimates of the parametric copulas may not estimate the closest copulas from each parametric family to the true data generating process, and the values of the integrated squared difference being compared may not estimate the distance between each parametric copula family to the true copula. Moreover, because of data reuse, it is not clear if the Gumbel model provides a significantly better fit to the data than the rest of the copula models.

The purpose of this paper is twofold. First, we demonstrate that for selection among candidate copula models that might all be misspecified, estimators of the copula parameters based on minimizing the selection criterion function may be preferred to other estimators, because the former consistently estimate the copulas in each parametric family least distant to the true copula. This ensures that the copulas being compared in the selection process are indeed the closest members from each parametric family to the true copula, consistent with the statement that Wang and Wells made in their Rejoinder to the Comment of Peña on their original article: "Our specific focus on an information-based criterion leads directly to a sensible definition of a best model as the model that mini-

https://doi.org/10.1017/S0266466607070181 Published online by Cambridge University Press

mizes a distance between itself and the underlying data-generating process." Second, we provide a formal statistical test for Archimedean copula model selection to address the issue of data reuse as raised by Peña in his comment on Wang and Wells (2000). Our test extends the reality check for data snooping of White (2000) to the semiparametric Archimedean copula model selection for censored data.

The remainder of this paper is organized as follows. Section 2 provides a brief review of dependence properties of commonly used parametric Archimedean copulas and the nonparametric method for estimating an Archimedean copula based on censored data. In Section 3, we introduce the model selection criterion used in this paper and provide a rigorous treatment of the estimation of the criterion function. In Section 4, we put copula model selection in the context of hypothesis testing and develop a test for model selection. Section 5 presents results based on a simulation experiment and two real data sets. Some concluding remarks are provided in Section 6. All the proofs are gathered in the Appendix. In what follows, we use $\Rightarrow$ to denote weak convergence.

2. ARCHIMEDEAN COPULAS AND THEIR NONPARAMETRIC ESTIMATION

2.1. A Brief Review of Archimedean Copulas

A (bivariate) copula is said to be Archimedean if it can be expressed in the form

$$C_\phi(u,v) = \phi^{-1}[\phi(u) + \phi(v)], \qquad 0 < u,v < 1 \tag{3}$$

for some convex, strictly decreasing function $\phi(u)$ from $[0,1]$ to $[0,\infty]$ satisfying $\phi(1) = 0$. By convention, $\phi^{-1}(v)$ is taken equal to zero whenever $v \geq \phi(0)$. The function $\phi(\cdot)$ is often referred to as a generator of the corresponding Archimedean copula. For a parametric Archimedean copula $C(u,v;\alpha)$, we use $\phi_\alpha(u)$ to denote its generator. For an Archimedean copula C_ϕ with generator ϕ, its Kendall's tau (τ) is given by

$$\tau = 1 + 4\int_0^1 \frac{\phi(t)}{\phi'(t)}\,dt, \tag{4}$$

and its lower and upper tail dependence parameters are (see Nelsen, 1997):

$$\lambda_L \equiv \lim_{u\to 0^+} \frac{C_\phi(u,u)}{u} = \lim_{t\to\infty} \frac{\phi^{-1}(2t)}{\phi^{-1}(t)}, \tag{5}$$

$$\lambda_U \equiv \lim_{u\to 1^-} \frac{1 - 2u + C_\phi(u,u)}{1-u} = 2 - \lim_{t\to 0^+} \frac{1-\phi^{-1}(2t)}{1-\phi^{-1}(t)}. \tag{6}$$

https://doi.org/10.1017/S0266466607070181 Published online by Cambridge University Press

Commonly used parametric Archimedean copulas include Frank, Clayton, and Gumbel copulas. The expressions of their generators are listed in Table 1. These copulas have different dependence properties. For example, the Clayton copula exhibits positive dependence (i.e., its Kendall's τ is always positive) and lower tail dependence ($\lambda_L \in (0,1]$) but no upper tail dependence ($\lambda_U = 0$); as α 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 α increases, both its overall dependence and its upper tail dependence increase. The Frank copula has symmetric positive and negative dependence but no tail dependence; as $|\alpha|$ increases, its overall dependence increases.

The class of semiparametric survival distributions with Archimedean copulas includes the commonly used frailty models in empirical studies of survival data as a subclass. This partially explains their popularity; see Joe (1997), Nelsen (1999), Clayton (1978), and Oakes (1982) for examples of applications. A bivariate frailty model for survival or lifetime variables X, Y assumes that X and Y are conditionally independent given a positive random variable W called frailty and that they satisfy

$$\Pr(X > x|W = w) = \{B_X(x)\}^w, \qquad \Pr(Y > y|W = w) = \{B_Y(y)\}^w, \tag{7}$$

where $B_X(\cdot)$ and $B_Y(\cdot)$ are baseline survival functions. In the numerical studies reported in Section 5 of this paper, we include two Archimedean copulas that are generated from frailty models: log-copulas corresponding to Hougaard's power variance functions (see Hougaard, 1986) and inverse-Gaussian copulas

Table 1. Examples of families of parametric Archimedean copulas

Family	ϕ	K	Parameter(s)
Clayton	$\frac{(v^{-\alpha}-1)}{\alpha}$	$K(v) = v\left(1 + \frac{1-v^{\alpha}}{\alpha}\right)$	$\alpha > 0$
Frank	$\log\left(\frac{1-\exp(-\alpha)}{1-\exp(-\alpha v)}\right)$	$K(v) = v + \left(\log\left[\frac{1-\exp(-\alpha)}{1-\exp(-\alpha v)}\right] \times \frac{1-\exp(-\alpha v)}{\alpha \exp(-\alpha v)}\right)$	$\alpha \in \mathcal{R}$
Gumbel	$(-\log(v))^{\alpha+1}$	$K(v) = v\left(1 - \frac{\log(v)}{\alpha+1}\right)$	$\alpha \geq 0$
Log-copula	$\left(1 - \frac{\log(v)}{\alpha\gamma}\right)^{\alpha+1} - 1$	$K(v) = v\left(1 - \frac{1-(1-\log(v))^{\alpha+1}}{(\alpha+1)(1-\log(v))^{\alpha}}\right)$	$\alpha > 0, \gamma > 0$
Inverse-Gaussian	$\log(v)\left(\frac{\log(v)-2\alpha}{2\alpha^2}\right)$	$K(v) = v\left(1 - \frac{\log(v)(\log(v)-2\alpha)}{2(\log(v)-\alpha)}\right)$	$\alpha > 0$

https://doi.org/10.1017/S0266466607070181 Published online by Cambridge University Press

corresponding to the inverse-Gaussian frailties (see Oakes, 1989). Their generators are also listed in Table 1. As shown in Genest and Rivest (1993), the log-copulas include Gumbel and Clayton copulas as limiting cases. Namely, the log-copula with parameters α, γ converges to Gumbel copula with parameter α as $\gamma \to 0$ and to the Clayton copula with parameter $1/\gamma$ as $\alpha \to \infty$. Using definitions (5) and (6), one can show that inverse-Gaussian copulas have neither upper nor lower tail dependence, and for $\alpha, \gamma \in (0,\infty)$, log-copulas have no tail dependence either.

Many properties of Archimedean copulas can be found in Joe (1997) and Nelsen (1999). We end this brief review by stating an important property of Archimedean copulas that is the basis for our model selection test. Let (U,V) be uniform random variables with an Archimedean copula function $C_\phi(u,v)$. For $v \in [0,1]$, define $K(v) = \Pr(C_\phi(U,V) \le v)$, the distribution function of $C_\phi(U,V)$. Genest and Rivest (1993) show that there is a one-to-one mapping between the generator ϕ and K: $\phi(v) = \exp[\int_{v_0}^{v}(t - K(t))^{-1}\,dt]$ for an arbitrary constant $0 < v_0 < 1$ and hence a one-to-one mapping between the Archimedean copula and K. For convenience, the expressions for K corresponding to the parametric copulas used in the numerical studies in Section 5 are listed in Table 1.

2.2. Nonparametric Estimation

Let (W_1, W_2) be the nuisance censoring variables. With right-censored data, one observes $(\widetilde{X}, \widetilde{Y}) = (X \wedge W_1, Y \wedge W_2)$ and a pair of indicators, $(\delta_1, \delta_2) = \{I(X \le W_1), I(Y \le W_2)\}$, where $X \wedge W_1 = \min(X, W_1)$ and $I(\cdot)$ is the indicator function. Suppose n i.i.d. observations $\{(\widetilde{X}_i, \widetilde{Y}_i, \delta_{1i}, \delta_{2i})\}_{i=1}^n$ are available, where $(\widetilde{X}_i, \widetilde{Y}_i) = (X_i \wedge W_{1i}, Y_i \wedge W_{2i})$ and $(\delta_{1i}, \delta_{2i}) = \{I(X_i \le W_{1i}), I(Y_i \le W_{2i})\}$. Throughout the paper we assume that the true but unknown joint survival function of (X,Y) takes an Archimedean form: $F^o(x,y) \equiv C^o(F_1^o(x), F_2^o(y)) \equiv \phi_o^{-1}[\phi_o(F_1^o(x)) + \phi_o(F_2^o(y))]$, where $\phi_o(\cdot)$ is a generator function of the copula $C^o(\cdot,\cdot)$.

Let $K^o(v) \equiv \Pr(F^o(X,Y) \le v)$, the distribution function of the random variable $F^o(X,Y)$. Because there is a one-to-one mapping from the Archimedean copula C^o and K^o, we will focus on the nonparametric estimation of K^o.

When complete observations on $(X,Y), \{(X_i,Y_i)\}_{i=1}^n$, are available, Genest and Rivest (1993) propose estimating $K^o(v)$ by $K_n(v) = (1/n)\sum_{i=1}^n I(F_n(X_i, Y_i) \le v)$, where $F_n(x,y)$ is the empirical distribution function of $\{(X_i,Y_i)\}_{i=1}^n$. Barbe et al. (1996) establish the weak convergence of the process $\{K_n(\cdot)\}$ to a Gaussian process.

For censored data, Wang and Wells (2000) propose a consistent estimator of $K^o(v)$:

$$\widetilde{K}(v) = 1 - \sum_{i=1}^{n}\sum_{j=1}^{n} I(\hat{F}(\widetilde{X}_{(i)}, \widetilde{Y}_{(j)}) > v)\hat{F}(\Delta\widetilde{X}_{(i)}, \Delta\widetilde{Y}_{(j)}),$$

https://doi.org/10.1017/S0266466607070181 Published online by Cambridge University Press

where $\widetilde{X}_{(1)} \leq \widetilde{X}_{(2)} \leq \cdots \leq \widetilde{X}_{(n)}$ and $\widetilde{Y}_{(1)} \leq \widetilde{Y}_{(2)} \leq \cdots \leq \widetilde{Y}_{(n)}$ are ordered observations of $\{(\widetilde{X}_i, \widetilde{Y}_i)\}_{i=1}^n$ and $\hat{F}(\Delta\widetilde{X}_{(i)}, \Delta\widetilde{Y}_{(j)}) = \hat{F}(\widetilde{X}_{(i)}, \widetilde{Y}_{(j)}) - \hat{F}(\widetilde{X}_{(i-1)}, \widetilde{Y}_{(j)}) - \hat{F}(\widetilde{X}_{(i)}, \widetilde{Y}_{(j-1)}) + \hat{F}(\widetilde{X}_{(i-1)}, \widetilde{Y}_{(j-1)})$ is the estimated mass at $(\widetilde{X}_{(i)}, \widetilde{Y}_{(j)})$ in which $\hat{F}(x, y)$ is a nonparametric estimate of $F(x, y)$.

For the paper to be self-contained, we restate the conditions and asymptotic properties of $\tilde{K}(v)$ established in Wang and Wells (2000).

A1. The distribution function $K^o(v)$ admits a continuous density $k^o(v)$.

A2. Let $\mathcal{T}$ denote the support of $(\widetilde{X}, \widetilde{Y})$. Given $F^o(x, y) = v$, there exist a version of the conditional distribution of (X, Y) and a countable family $\mathcal{P}$ of partitions $\mathcal{C}$ on $\mathcal{T}$ into a finite number of Borel sets satisfying $\inf_{\mathcal{C}\in\mathcal{P}} \max_{A\in\mathcal{C}} \operatorname{diam}(A) = 0$, such that for all $A \in \mathcal{C}$, the mapping $v \to \mu_v(A) = k^o(v)\Pr\{(X, Y) \in A | F^o(X, Y) = v\}$ is continuous.

B1. $\hat{F}(x, y)$ is a uniformly and strongly consistent estimator of $F^o(x, y)$ for $(x, y) \in \mathcal{T}$.

B2. $n^{1/2}\{\hat{F}(x, y) - F^o(x, y)\}$ converges weakly to $W(x, y)$, where $W(x, y)$ is a continuous mean-zero Gaussian process on $\mathcal{D}(\mathcal{T})$, where $\mathcal{D}(\mathcal{T})$ denotes Skorohod space.

For independent censoring, the bivariate Kaplan–Meier estimator introduced in Dabrowska (1988) satisfies B1 and B2; see Dabrowska (1989). In addition to weak convergence of the bivariate Kaplan–Meier estimator, Dabrowska (1989) also establishes the law of iterated logarithm and the validity of the bootstrap version of the bivariate Kaplan–Meier estimator. This estimator will be used in the numerical studies in this paper. The following lemma summarizes the results of Theorem 1 in Wang and Wells (2000).

LEMMA 2.1. *Under B1, for* $0 < \varsigma = F^o(\tau_1, \tau_2) \leq v \leq 1$ *with* $(\tau_1, \tau_2) \in \mathcal{T}$, $\tilde{K}(v) - K^o(v) = o_p(1)$. *Furthermore, if A1, A2, and B2 hold, then on* $\mathcal{D}[\varsigma, 1]$,

$$n^{1/2}\{\tilde{K}(v) - K^o(v)\} \Rightarrow X(v)$$

$$\equiv -\iint I(F^o(x, y) > v) W(dx, dy)$$

$$-\iint W(x, y)\mu_v(dx, dy).$$

3. MODEL SELECTION AND PARAMETRIC ESTIMATION

For each $j = 1, 2, \ldots, M$, let $\mathcal{M}_j = \{C_j(u_1, u_2; \alpha_j) : \alpha_j \in \mathcal{A}_j \subset \mathcal{R}^{a_j}\}$ be a class of parametric Archimedean copulas. Let $K_j(v; \alpha_j) \equiv \Pr(C_j(F_1^o(X), F_2^o(Y); \alpha_j) \leq v)$. Then, by the result of Genest and Rivest (1993), model selection of semiparametric Archimedean survival models can be based on the difference between

https://doi.org/10.1017/S0266466607070181 Published online by Cambridge University Press

$\{K_j(v;\alpha_j):\alpha_j \in \mathcal{A}_j\}_{j=1}^M$ and $K^o(v)$. In particular, Wang and Wells (2000) propose a copula model selection procedure using the integrated squared difference between $\{K_j(v;\alpha_j):\alpha_j \in \mathcal{A}_j\}$ and $K^o(v)$.

Define the integrated squared difference as

$$S_j(\alpha_j) = \int_{\varsigma}^{1} [K^o(v) - K_j(v;\alpha_j)]^2\, dv \quad \text{for } j = 1,\ldots,M,$$

where ς is the smallest value of $F^o(x,y)$ such that the estimator $\tilde{K}(v)$ consistently estimates $K^o(v)$ for $v \in [\varsigma,1]$; see Lemma 2.1 in Section 2.2. Define

$$\alpha_j^* = \arg\min_{\alpha_j \in \mathcal{A}_j} S_j(\alpha_j) \quad \text{for } j = 1,\ldots,M.$$

Then $K_j(v;\alpha_j^*)$ corresponds to the copula in $\{C_j(u_1,u_2;\alpha_j):\alpha_j \in \mathcal{A}_j\}$ closest to the true copula according to the integrated squared difference criterion, and the distance from the jth model class to the true copula is $S_j(\alpha_j^*)$ $(j = 1,\ldots,M)$.

For $j = 1,\ldots,M$, the value α_j^* is referred to as the pseudo-true value, as it leads to the copula function in the jth model class that best approximates the true copula. Obviously if the copula $C_j(u_1,u_2;\alpha_j)$ correctly specifies the true copula in the sense that there exists α_{jo} such that $K^o(v) = K_j(v;\alpha_{jo})$, then $\alpha_j^* = \alpha_{jo}$ and $S_j(\alpha_j^*) = 0$. In general, however, all the parametric copulas being entertained may misspecify the true copula, and the M best copulas $K_j(v;\alpha_j^*)$, one from each class $\mathcal{M}_j$, $j = 1,\ldots,M$, are compared with each other in the selection process. The class $\mathcal{M}_i$ is selected if $S_i(\alpha_i^*) < S_j(\alpha_j^*)$ for all $j = 1,\ldots,M$ and $j \neq i$. In practice, neither $K^o(v)$ nor α_j^* is known. Any model selection procedure will then be based on estimates of $K^o(v)$ and α_j^* or equivalently of $S_j(\alpha_j^*)$, $j = 1,\ldots,M$.

3.1. Estimation of α_j^*

The choice of an appropriate estimator of α_j^* is a delicate issue and deserves a rigorous treatment. When $\mathcal{M}_j$ correctly specifies the true copula model, several consistent estimators of α_j^* are available in the literature, including the two-step estimator of Shih and Louis (1995), the minimum distance estimator proposed in Wang and Wells (2000), and the estimator based on Kendall's τ. In Section 3.1, we study asymptotic properties of the minimum distance estimator in the general case that $\mathcal{M}_j$ misspecifies the true model, and we discuss the pros and cons of basing model selection on different parametric estimators in Section 3.2.

Let $\tilde{S}_j(\alpha_j) = \int_{\varsigma}^{1}[\tilde{K}(v) - K_j(v;\alpha_j)]^2\, dv$. The minimum distance estimator is defined as

$$\hat{\alpha}_j = \arg\min_{\alpha_j} \int_{\varsigma}^{1} [\tilde{K}(v) - K_j(v;\alpha_j)]^2\, dv \equiv \arg\min_{\alpha_j} \tilde{S}_j(\alpha_j).$$

https://doi.org/10.1017/S0266466607070181 Published online by Cambridge University Press

Wang and Wells (2000) establish the asymptotic properties of $\hat{\alpha}_j$ when the copula class $\mathcal{M}_j$ correctly specifies the true copula of (X,Y). In this case, $\alpha_j^* = \alpha_{jo}$. In general, because $\alpha_j^* = \arg\min_{\alpha_j} \int_\varsigma^1 [K^o(v) - K_j(v;\alpha_j)]^2\,dv$ and $\tilde{K}(v)$ is a consistent estimator of $K^o(v)$ for $v \in [\varsigma,1]$, we expect $\hat{\alpha}_j$ to be a consistent estimator of α_j^* regardless of whether $\mathcal{M}_j$ correctly specifies the true copula. The formal results are stated in the following proposition.

In what follows, D_α stands for the operator of differentiation with respect to α: $D_\alpha f(\cdot,\alpha) = \partial f(\cdot,\alpha)/\partial\alpha$ for a function $f(\cdot,\alpha)$ depending on α.

PROPOSITION 3.1. *Suppose $\mathcal{A}_j$ is a compact subset of $\mathcal{R}^{a_j}$ with nonempty interior and α_j^* is the unique minimizer of $S_j(\alpha_j)$ over $\mathcal{A}_j$. Moreover, assume that $K_j(v;\alpha_j)$ is twice differentiable with respect to α_j with bounded derivatives. Then the following results hold.*

(i) $\hat{\alpha}_j$ *is consistent for* α_j^*.

(ii) Let α_j^* *be an interior point of* $\mathcal{A}_j$. *Under A1, A2, B1, and B2,* $\sqrt{n}(\hat{\alpha}_j - \alpha_j^*)$ *converges in distribution to* Q_j, *where*

$$Q_j = 2\left\{2\int_\varsigma^1 [D_\alpha K_j(v;\alpha_j^*) D'_\alpha K_j(v;\alpha_j^*)]\,dv - \int_\varsigma^1 [K^o(v) - K_j(v;\alpha_j^*)] D_\alpha^2 K_j(v;\alpha_j^*)\,dv\right\}^{-1} \times \int_\varsigma^1 X(v) D_\alpha K_j(v;\alpha_j^*)\,dv$$

and $X(v)$ *is defined in Lemma 2.1.*

(iii) $\tilde{S}_j(\hat{\alpha}_j)$ *is a consistent estimator of* $S_j(\alpha_j^*)$.

(iv) If $\mathcal{M}_j$ *misspecifies the true copula, then*

$$\sqrt{n}[\tilde{S}_j(\hat{\alpha}_j) - S_j(\alpha_j^*)] \Rightarrow 2\int_\zeta^1 X(v)[K^o(v) - K_j(v;\alpha_j^*)]\,dv.$$

On the other hand, if $\mathcal{M}_j$ *correctly specifies the true copula, then* $n\tilde{S}_j(\hat{\alpha}_j) \Rightarrow \int_\zeta^1 \hat{X}_j^2(v)\,dv$, *where* $\hat{X}_j(v) = X(v) - D_\alpha K_j(v;\alpha_j^*) Q_j$.

Proposition 3.1 shows that in general $\hat{\alpha}_j$ consistently estimates the pseudo-true value α_j^*, extending the consistency result of Wang and Wells (2000) for α_{jo} when the jth model class correctly specifies the true model. Moreover $\tilde{S}_j(\hat{\alpha}_j)$ consistently estimates $S_j(\alpha_j^*)$, justifying the use of $\tilde{S}_j(\hat{\alpha}_j)$ for model selection based on minimizing $S_j(\alpha_j^*)$ over $j = 1,\ldots,M$. Proposition 3.1(iv) also implies that the limiting distribution of $\sqrt{n}[\tilde{S}_j(\hat{\alpha}_j) - S_j(\alpha_j^*)]$ is the same as that of $\sqrt{n}[\tilde{S}_j(\alpha_j^*) - S_j(\alpha_j^*)]$ irrespective of whether $\mathcal{M}_j$ correctly specifies the true

https://doi.org/10.1017/S0266466607070181 Published online by Cambridge University Press

copula model. The only difference is that when $\mathcal{M}_j$ correctly specifies the true model, $\sqrt{n}[\tilde{S}_j(\alpha_j^*) - S_j(\alpha_j^*)]$ has a degenerate limiting distribution.

3.2. Which Parametric Estimator to Use?

Many estimators, including the two-step estimator of Shih and Louis (1995), share the consistency property of $\hat{\alpha}_j$ for α_{jo} when the jth model correctly specifies the true model, but may not share the consistency property for the pseudo-true value α_j^* in general. The reason is that unlike the true value α_{jo} when the jth copula class correctly specifies the true copula, the pseudo-true value α_j^* when the jth copula class misspecifies the true copula may change if a different distance measure is used.

To be more specific, let $\tilde{\alpha}_j$ be a parametric estimator other than $\hat{\alpha}_j$. Examples include the estimator based on inverting Kendall's τ used in the numerical studies in Wang and Wells (2000) and the two-step estimator in Shih and Louis (1995). Most of these estimators are optimization estimators in the sense that they can be obtained by minimizing some criterion function, say, $\tilde{L}_j(\alpha_j)$. For example, for an Archimedean copula, the estimator based on inverting Kendall's τ can be obtained by minimizing $\tilde{L}_j(\alpha_j) = |\int_\varsigma^1 [\tilde{K}(v) - K_j(v;\alpha_j)]\,dv|$, because $\tau = 4\int_0^1 [v - K(v)]\,dv + 1$. The two-step estimator of Shih and Louis (1995) maximizes the pseudo-likelihood function. Under regularity conditions, $\tilde{\alpha}_j = \alpha_{jL}^* + o_p(1)$, where $\alpha_{jL}^* = \operatorname{argmin}_{\alpha_j}[\operatorname{plim} \tilde{L}_j(\alpha_j)]$. For example, for the estimator based on inverting Kendall's τ, $\alpha_{jL}^* = \operatorname{argmin}_{\alpha_j} |\int_\varsigma^1 [K^o(v) - K_j(v;\alpha_j)]\,dv|$, whereas for the two-step estimator in Shih and Louis (1995), α_{jL}^* minimizes the Kullback–Leibler information criterion. In general, α_{jL}^* differs from α_j^* unless $\mathcal{M}_j$ correctly specifies the true copula, in which case both α_{jL}^* and α_j^* equal the true value of the parameter α_{jo}.

PROPOSITION 3.2. *Suppose $\mathcal{A}_j$ is a compact subset of $\mathcal{R}^{a_j}$ with nonempty interior and $K_j(v;\alpha_j)$ is twice differentiable with respect to α_j with bounded derivatives. In addition, assume that the estimator $\tilde{\alpha}_j$ satisfies $\tilde{\alpha}_j = \alpha_{jL}^* + o_p(1)$ and $\sqrt{n}[\tilde{\alpha}_j - \alpha_{jL}^*] \Rightarrow Q_{jL}$ for some $\alpha_{jL}^* \in \mathcal{A}_j$ and a nondegenerate distribution Q_{jL}.*

(i) $\tilde{S}_j(\tilde{\alpha}_j)$ is a consistent estimator of $S_j(\alpha_{jL}^*)$.

(ii) If $\mathcal{M}_j$ misspecifies the true copula, then

$$\sqrt{n}[\tilde{S}_j(\tilde{\alpha}_j) - S_j(\alpha_{jL}^*)] \Rightarrow 2\int_\zeta^1 \hat{X}_{jL}(v)[K^o(v) - K_j(v;\alpha_{jL}^*)]\,dv,$$

where $\hat{X}_{jL}(v) = X(v) - D_\alpha K_j(v;\alpha_{jL}^*)Q_{jL}$.

(iii) If $\mathcal{M}_j$ correctly specifies the true copula, then $n\tilde{S}_j(\tilde{\alpha}_j) \Rightarrow \int_\zeta^1 \hat{X}_{jL}^2(v)\,dv$.

https://doi.org/10.1017/S0266466607070181 Published online by Cambridge University Press

Given that in general $\alpha_{jL}^* \neq \alpha_j^*$ unless copula class $\mathcal{M}_j$ correctly specifies the true copula, Proposition 3.2 shows that $\tilde{S}_j(\tilde{\alpha}_j)$ is in general an inconsistent estimator of $S_j(\alpha_j^*)$. Any model selection procedure based on $\tilde{S}_j(\tilde{\alpha}_j)$ ($j = 1, \ldots, M$) may thus lead to incorrect inference unless the ranking of the M models based on $S_j(\alpha_{jL}^*)$ is the same as that based on $S_j(\alpha_j^*)$, or at the very least, the model that minimizes $S_j(\alpha_{jL}^*)$ over $j = 1, \ldots, M$ also minimizes $S_j(\alpha_j^*)$ over $j = 1, \ldots, M$. Another difference between using $\tilde{\alpha}_j$ and $\hat{\alpha}_j$ in model selection is that the limiting distribution of $\sqrt{n}[\tilde{S}_j(\tilde{\alpha}_j) - S_i(\alpha_{jL}^*)]$ is not the same as that of $\sqrt{n}[\tilde{S}_j(\alpha_{jL}^*) - S_j(\alpha_{jL}^*)]$ unless $\mathcal{M}_j$ correctly specifies the true model, in which case the limiting distribution is degenerate. As Wang and Wells argue in their rejoinder to the comment of Peña on their paper: "in practice, only best approximating models exist," it is an exception rather than the rule that any simple statistical model is able to characterize the true data generating process correctly. As a result, properties of $\tilde{S}_j(\hat{\alpha}_j)$ or $\tilde{S}_j(\tilde{\alpha}_j)$ when $\mathcal{M}_j$ misspecifies the true model are most relevant to model selection.

4. A MODEL SELECTION TEST

Results in the previous section suggest the use of $\tilde{S}_j(\hat{\alpha}_j)$ in model selection. Two drawbacks of basing model selection on point estimates $\tilde{S}_j(\hat{\alpha}_j)$ only are as follows. First, the limiting distribution of $\sqrt{n}[\tilde{S}_j(\hat{\alpha}_j) - S_j(\alpha_j^*)]$ depends on model j under misspecification, even though it is the same as that of $\sqrt{n}[\tilde{S}_j(\alpha_j^*) - S_j(\alpha_j^*)]$. Second, the same data are used twice in obtaining $\tilde{S}_j(\hat{\alpha}_j)$; see Peña's discussion of both issues in his Comment on Wang and Wells (2000). A procedure for assessing the significance of the selection result is called for. In this section, we provide a formal statistical procedure to address this issue by putting copula model selection in the context of hypothesis testing along the lines of Vuong (1989) and White (2000).

To clarify the underlying idea, consider the case where only two models are being selected ($M = 2$). The relevant null and alternative hypotheses are

$$H_0 : S_1(\alpha_1^*) \leq S_2(\alpha_2^*) \quad \text{and} \quad H_1 : S_1(\alpha_1^*) > S_2(\alpha_2^*).$$

So under H_0, $\mathcal{M}_1$ will be selected and under H_1, $\mathcal{M}_2$ will be selected. We will construct a test for H_0 based on $[\tilde{S}_1(\hat{\alpha}_1) - \tilde{S}_2(\hat{\alpha}_2)]$.

For more than two models, we will formulate the testing problem in the same way as White (2000). Let $C_1(u_1, u_2; \alpha_1)$ be the benchmark copula and $\{C_j(u_1, u_2; \alpha_j)\}_{j=2}^M$ be the candidate copulas. The null and alternative hypotheses of interest are expressed as

$$H_0 : \max_{j=2,\ldots,M} [S_1(\alpha_1^*) - S_j(\alpha_j^*)] \leq 0 \quad \text{and} \quad H_1 : \max_{j=2,\ldots,M} [S_1(\alpha_1^*) - S_j(\alpha_j^*)] > 0.$$

Under H_0, $\mathcal{M}_1$ will be selected, and under H_1, the model minimizing $S_j(\alpha_j^*)$ over $j = 2, \ldots, M$ will be selected.

https://doi.org/10.1017/S0266466607070181 Published online by Cambridge University Press

In contrast to goodness-of-fit tests developed in Genest and Rivest (1993) and Wang and Wells (2000), our model selection test does not require any of the parametric copulas to correctly specify the true copula under H_0. In practical applications this is most likely to be the case, and the best one can do is to select the model that best approximates the true model according to some criterion such as the integrated squared difference.

Let $T_{jn} = \tilde{S}_1(\hat{\alpha}_1) - \tilde{S}_j(\hat{\alpha}_j)$, where recall that $\tilde{S}_j(\hat{\alpha}_j) = \int_\zeta^1 [\tilde{K}(v) - K_j(v;\hat{\alpha}_j)]^2\,dv$ for $j = 1,\ldots,M$. The following proposition provides the basis for our test.

PROPOSITION 4.1. *Suppose that for $j = 1,2,\ldots,M$, the copula model j satisfies the conditions of Proposition 3.1. Then jointly $n^{1/2}\{T_{jn} - [S_1(\alpha_1^*) - S_j(\alpha_j^*)]\}_{j=2,\ldots,M} \Rightarrow (Z_2,\ldots,Z_M)'$, where*

$$\begin{pmatrix} Z_2 \\ \vdots \\ Z_M \end{pmatrix} \sim 2 \begin{pmatrix} \int_\zeta^1 X(v)[K_2(v;\alpha_2^*) - K_1(v;\alpha_1^*)]\,dv \\ \vdots \\ \int_\zeta^1 X(v)[K_M(v;\alpha_M^*) - K_1(v;\alpha_1^*)]\,dv \end{pmatrix}.$$

Moreover, let σ_{jj} denote the variance of Z_j. Then $\sigma_{jj} = 0$ if and only if model j and the benchmark model 1 are generalized nested in the sense that $K_j(v;\alpha_j^*) = K_1(v;\alpha_1^*)$ for almost all $v \in [0,1]$.

Proposition 4.1 and the continuous mapping theorem imply

$$\max_{j=2,\ldots,M} n^{1/2}\{T_{jn} - [S_1(\alpha_1^*) - S_j(\alpha_j^*)]\} \Rightarrow \max_{j=2,\ldots,M} Z_j.$$

Hence under the least favorable configuration (i.e., $S_j(\alpha_j^*) = S_1(\alpha_1^*)$ for $j = 2,\ldots,M$), $\max_{j=2,\ldots,M}[\sqrt{n}T_{jn}] \Rightarrow \max_{j=2,\ldots,M} Z_j$. This allows us to construct a test for H_0.

Let $T_n = \max_{j=2,\ldots,M}[\sqrt{n}T_{jn}]$. Suppose $\sigma_{jj} > 0$ for at least one $j = 2,\ldots,M$. Then we will reject H_0 if $T_n > Z_\alpha$, where Z_α is the upper α-percentile of the distribution of $\max_{j=2,\ldots,M} Z_j$. The asymptotic power properties of this test against fixed alternatives and Pitman local alternatives follow immediately from Proposition 4.1 and are summarized in the following theorem.

THEOREM 4.2. *The test based on T_n is consistent against fixed alternatives of the form H_1 and has nontrivial power against local alternatives satisfying*

$$\max_{j=2,\ldots,M} \lim_{n\to\infty} \{\sqrt{n}[S_1(\alpha_1^*) - S_j(\alpha_j^*)]\} > 0.$$

https://doi.org/10.1017/S0266466607070181 Published online by Cambridge University Press

In general, the distribution of $\max_{j=2,\ldots,M} Z_j$ is unknown, depending on the true survival function $F^o(x,y)$ and the pseudo-true values α_j^*. To circumvent this difficulty, we propose using the naive bootstrap to approximate the null distribution of T_n. Bootstrap is also used in Wang and Wells (2000) to approximate the null limiting distribution of their goodness-of-fit test. However, because in Wang and Wells (2000) the parametric copula model being tested correctly specifies the true copula under the null hypothesis, they suggest the use of a parametric bootstrap procedure based on the estimated null model $K_j(v;\hat{\alpha}_j)$. This approach will not work in our context, as all the parametric copula models may misspecify the true copula. The naive bootstrap, on the other hand, provides a valid procedure.

Step 1. Let $\{(\widetilde{X}_i^*, \widetilde{Y}_i^*, \delta_{1i}^*, \delta_{2i}^*)\}_{i=1}^n$ be a random sample with replacement from the original data $\{(\widetilde{X}_i, \widetilde{Y}_i, \delta_{1i}, \delta_{2i})\}_{i=1}^n$ and let $\hat{F}^*(x,y), \hat{K}^*(v)$, and $\hat{\alpha}_j^*$ be the bootstrapped counterparts of $\hat{F}(x,y), \hat{K}(v)$, and $\hat{\alpha}_j$.

Step 2. Let T_{jn}^* be the bootstrapped value of T_{jn} and define its recentered value as

$$T_{jnC}^* = T_{jn}^* - T_{jn}, \qquad j = 2,\ldots,M.$$

Step 3. Repeat steps 1 and 2 for a large number of times and define the bootstrap value of T_n as $T_n^* = \max_{j=2,\ldots,M}[n^{1/2}T_{jnC}^*]$.

Step 4. Use the empirical distribution function of the resulting values T_n^* to approximate the null distribution of T_n.

We show in the Appendix that the preceding bootstrap works for T_n.

THEOREM 4.3. *Under the conditions of Proposition 4.1,*

$$\sup_{z\in\mathcal{R}}\left| P^*(T_n^* \le z) - P\left(\max_{j=2,\ldots,M}[n^{1/2}T_{jn} - \{S_1(\alpha_1^*) - S_j(\alpha_j^*)\}] \le z\right)\right| = o_p(1),$$

where $P^*(T_n^* \le z)$ is the conditional distribution of T_n^* given the original sample.

5. NUMERICAL STUDIES

Following Wang and Wells (2000), we consider five Archimedean copula models: Clayton, Frank, Gumbel, log-copula, and inverse-Gaussian. Their generators and expressions for $K(v)$ are given in Table 1 in Section 2. For the log-copula, we assume $\alpha\gamma = 1$ as in Wang and Wells (2000). Hence, the families of Clayton and Gumbel copulas are not nested in the family of log-copulas considered in the numerical studies.

https://doi.org/10.1017/S0266466607070181 Published online by Cambridge University Press

5.1. Simulation Results

The simulation design is the same as that of Wang and Wells (2000) except that we include the inverse-Gaussian copula in the experiment. In all cases, (W_1, W_2) are generated from Clayton copula with Kendall's $\tau = 0.3$, and the censoring rates in both components are controlled to be between 10 and 20%. The sample size is $n = 250$.[2] The number of simulation repetitions is 200, and the number of bootstrap repetitions is 100. All the computations are done in Fortran.[3]

5.1.1. Results Based on $\tilde{S}_j(\hat{\alpha}_j)$. In this section we examine performance of the model selection procedure based on the minimum value of $\tilde{S}_j(\hat{\alpha}_j)$. For comparison purposes, we also report results based on $\tilde{S}_j(\tilde{\alpha}_j)$, where $\tilde{\alpha}_j$ is the estimator based on Kendall's τ used in Wang and Wells (2000).

In the simulation study we have used Clayton copula and Frank copula as the true copula, respectively. For each true copula, we calculated the percentage of times that each copula is selected based on having the minimum value $\tilde{S}_j(\hat{\alpha}_j)$ $(\tilde{S}_j(\tilde{\alpha}_j))$ in two cases: (i) the true copula is included in the family of copula models to be selected $(M = 5)$; (ii) the true copula is not included in the family of copula models to be selected $(M = 4)$. As we argued earlier, case (ii) is most likely to characterize real empirical applications. As the results for Clayton and Frank copulas are very close, we report results for Clayton copula only in Table 2.

In addition to the percentage of times that each copula is selected, we also reported the mean and standard error of $\tilde{S}_j(\hat{\alpha}_j)$ $(\tilde{S}_j(\tilde{\alpha}_j))$. To fully understand the results in Table 2, we provide the ranking of the remaining four copulas in terms of their distance to the true Clayton copula $S_j(\alpha_j^*)$ and their ranking based on $S_j(\alpha_{jL}^*)$ in Table 3, where α_{jL}^* is the probability limit of $\tilde{\alpha}_j$, that is, $\alpha_{jL}^* = \text{argmin}_{\alpha_j} \left| \int_0^1 K^o(v)\, dv - \int_0^1 K_j(v;\alpha_j) \right|$ in which $K^o(v)$ is the K function for Clayton copula and $\{K_j(v;\alpha_j)\}_{j=1}^4$ are associated with the remaining copulas in Table 1. In all cases considered $(\tau = 0.3, 0.5, 0.7)$, the ranking of the remaining four copulas is the same regardless of whether $S_j(\alpha_j^*)$ is used or $S_j(\alpha_{jL}^*)$ is used. Moreover, in most of the cases, the respective values of $S_j(\alpha_j^*)$ and $S_j(\alpha_{jL}^*)$ are very close, although they are clearly different. Our theoretical results in Section 2 predict that the selection results based on $\tilde{S}_j(\hat{\alpha}_j)$ and $\tilde{S}_j(\tilde{\alpha}_j)$ should be similar. Indeed, Table 2 provides an overwhelming amount of evidence supporting this prediction.

Table 2 reveals that when the true Clayton copula is included, it always has the highest probability of being selected based on the minimum value criterion regardless of whether $\hat{\alpha}_j$ or $\tilde{\alpha}_j$ is being used and the value of τ. The copula with the second highest probability of being selected is the log-copula, the closest to the true copula, although its probability of being selected is much smaller than that of the Clayton copula. In all cases except for the inverse-Gaussian with $\tau = 0.7$, the mean and standard error of $\tilde{S}_j(\hat{\alpha}_j)$ are smaller than the mean and standard error of $\tilde{S}_j(\tilde{\alpha}_j)$. What is more important and reassuring is the fact that,

https://doi.org/10.1017/S0266466607070181 Published online by Cambridge University Press

TABLE 2. The true copula is Clayton

	Minimum distance $\hat{\alpha}_j$			Kendall's tau $\tilde{\alpha}_j$		
	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$
Clayton included						
Percentage						
Clayton	0.72	0.79	0.80	0.76	0.80	0.78
Frank	0	0	0	0.01	0	0
Gumbel	0	0	0	0	0	0
Log-copula	0.23	0.21	0.20	0.17	0.20	0.22
Inverse-Gaussian	0.05	0	0	0.07	0	0
Mean $\times 10^4$						
Clayton	2.70	2.58	2.24	3.30	3.01	2.74
Frank	8.51	12.30	10.58	15.41	16.29	12.13
Gumbel	13.37	20.02	17.05	70.14	36.40	18.54
Log-copula	3.90	4.30	3.71	4.81	4.82	4.15
Inverse-Gaussian	5.30	17.63	53.75	5.78	18.47	53.75
Standard error $\times 10^4$						
Clayton	1.59	1.34	1.06	2.09	1.75	1.46
Frank	4.25	4.84	3.87	8.60	20.18	16.87
Gumbel	5.62	6.45	5.19	15.48	8.70	5.43
Log-copula	2.37	2.40	1.90	2.89	2.81	2.14
Inverse-Gaussian	3.23	6.99	12.38	3.42	7.18	12.38
Clayton excluded						
Percentage						
Frank	0.01	0	0	0.01	0	0
Gumbel	0	0	0	0	0	0
Log-copula	0.87	1	1	0.76	1	1
Inverse-Gaussian	0.12	0	0	0.23	0	0
Mean $\times 10^4$						
Frank	8.12	12.74	10.65	14.50	20.65	12.36
Gumbel	12.98	20.61	17.14	72.25	37.44	18.70
Log-copula	3.74	4.49	3.77	4.61	5.02	4.08
Inverse-Gaussian	5.06	17.86	52.99	5.50	18.71	52.99
Standard error $\times 10^4$						
Frank	3.68	4.94	4.27	9.25	28.59	21.10
Gumbel	5.10	6.29	5.40	16.44	9.05	5.42
Log-copula	1.89	2.44	2.14	2.63	2.80	2.35
Inverse-Gaussian	2.66	6.88	12.21	2.99	7.11	12.21

based on $\tilde{S}_j(\hat{\alpha}_j)$, when the true copula is not included, which is typically the case in empirical applications, the log-copula, being the closest to the true Clayton copula, has the highest probability of being selected and as the value of τ exceeds 0.3, the probability of selecting the log-copula reaches 1; see Table 2.

https://doi.org/10.1017/S0266466607070181 Published online by Cambridge University Press

TABLE 3. Rankings from the least distant to most distant to Clayton copula

Ranking based on $S_j(\alpha_j^*)$				
$\tau = 0.3$	Log-copula	Inverse-Gaussian	Frank	Gumbel
$S_j(\alpha_j^*) \times 10^4$	1.6765	3.1612	6.7928	12.5798
$\tau = 0.5$	Log-copula	Frank	Inverse-Gaussian	Gumbel
$S_j(\alpha_j^*) \times 10^4$	2.3559	11.2060	16.9010	19.8090
$\tau = 0.7$	Log-copula	Frank	Gumbel	Inverse-Gaussian
$S_j(\alpha_j^*) \times 10^4$	1.7295	8.8827	15.6128	50.0980
Ranking based on $S_j(\alpha_{jL}^*)$				
$\tau = 0.3$	Log-copula	Inverse-Gaussian	Frank	Gumbel
$S_j(\alpha_{jL}^*) \times 10^4$	1.6766	3.2540	6.8211	12.5798
$\tau = 0.5$	Log-copula	Frank	Inverse-Gaussian	Gumbel
$S_j(\alpha_{jL}^*) \times 10^4$	2.3560	11.2540	20.0106	20.0106
$\tau = 0.7$	Log-copula	Frank	Gumbel	Inverse-Gaussian
$S_j(\alpha_{jL}^*) \times 10^4$	1.7429	8.9097	16.2396	54.0411

Only when $\tau = 0.3$, the probability that the log-copula is being selected decreases slightly and the inverse-Gaussian copula being the second closest to the true copula gains a small probability of being chosen. The corresponding results based on $\tilde{\alpha}_j$ reveal similar patterns qualitatively, although for $\tau = 0.3$, the probability of selecting the log-copula using $\tilde{\alpha}_j$ is smaller than that using $\hat{\alpha}_j$. In all cases except for the inverse-Gaussian with $\tau = 0.7$, the mean and standard error of $\tilde{S}_j(\hat{\alpha}_j)$ are smaller than the mean and standard error of $\tilde{S}_j(\tilde{\alpha}_j)$.

5.1.2. Results Based on the Model Selection Test. In this section, we applied the model selection test proposed in Section 4 to the same data sets generated in Section 5.1.1. When Clayton copula is included, H_0 holds with Clayton being the benchmark, and H_1 holds when any other copula is used as the benchmark. When Clayton copula is excluded, H_0 holds with log-copula being the benchmark. The rejection rates are reported in Table 4.

Table 4 indicates that the test is undersized regardless of whether the true copula is included in the selection and the true value of τ.[4] In the first case, when a copula other than the Clayton is used as the benchmark, the rejection rates start to increase. Because for all values of τ, the log-copula is the closest to the true copula and the distance between the log-copula and the true copula is small, the rejection rate for the null hypothesis with log-copula as the benchmark is very low. But as the benchmark copula moves away from the true copula, the rejection rate of the test increases quickly. For example, when $\tau = 0.5$ and the Gumbel is used as the benchmark, the rejection rate exceeds 0.9. By comparing the rejection rates in Table 4 with the distance of each copula to the true copula reported in Table 3, we conclude that the rejection rate of the test mainly depends on the distance of the benchmark to the true copula instead of

https://doi.org/10.1017/S0266466607070181 Published online by Cambridge University Press

TABLE 4. Rejection rates

	Clayton included			Clayton excluded		
	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$
Benchmark	Clayton (H_0 holds)			Log-copula (H_0 holds)		
5%	0	0	0	0	0	0
10%	0	0	0	0	0	0
Benchmark	Log-copula (H_1 holds)			Inverse-Gaussian (H_1 holds)		
5%	0	0	0	0.005	0.61	0.94
10%	0.02	0.02	0	0.015	0.85	0.98
Benchmark	Inverse-Gaussian (H_1 holds			Frank (H_1 holds))		
5%	0.03	0.63	0.95	0.49	0.73	0.02
10%	0.10	0.81	0.99	0.61	0.79	0.06
Benchmark	Frank (H_1 holds)			Gumbel (H_1 holds)		
5%	0.41	0.61	0.04	0.84	0.96	0.29
10%	0.55	0.74	0.13	0.91	0.99	0.53
Benchmark	Gumbel (H_1 holds)					
5%	0.67	0.92	0.46			
10%	0.82	0.99	0.58			

the value of τ. When the true copula is excluded from the selection, the same general conclusion carries over; see Table 4. In this case, all the copulas misspecify the true copula. But when the log-copula is used as the benchmark, the null hypothesis holds. When inverse-Gaussian, Frank, or Gumbel is used as the benchmark, the null hypothesis fails, and the power of the test increases as the distance of the benchmark copula to the true copula increases. In summary, besides Clayton copula, the only copula that the model selection test fails to reject as the benchmark model for all values of τ considered is the log-copula, supporting the selection result based on values of $\tilde{S}_j(\hat{\alpha}_j)$.

5.2. Real Data Examples

In this section, we apply our proposed model selection test to the two data sets studied in Wang and Wells (2000): the data set in Danahy, Burwell, Aranow, and Prakash (1977) and the data set in McGilchrist and Aisbett (1991). We suggest that readers consult the original papers or Wang and Wells (2000) for backgrounds on these data sets. Based on the value of $\tilde{S}_j(\tilde{\alpha}_j)$, Wang and Wells (2000) recommend the choice of Clayton or log-copula for the first data set and Gumbel for the second data set. In Table 5, we report bootstrap p-values of our model selection test, values of $\tilde{S}_j(\tilde{\alpha}_j)$ and $\tilde{S}_j(\hat{\alpha}_j)$.

Based on the p-values of the test, Clayton copula is rejected for both data sets at 10% significance level and none of the other three copulas is rejected.

https://doi.org/10.1017/S0266466607070181 Published online by Cambridge University Press

TABLE 5. *p*-values of the test

	Danahy et al. (1977) ($\tau = 0.41$)			McGilchrist and Aisbett (1991) ($\tau = 0.19$)		
Benchmark	*p*-value	$\tilde{S}_j(\hat{\alpha}_j) \times 10^3$	$\tilde{S}_j(\tilde{\alpha}_j) \times 10^3$	*p*-value	$\tilde{S}_j(\hat{\alpha}_j) \times 10^3$	$\tilde{S}_j(\tilde{\alpha}_j) \times 10^3$
Log-copula	0.32	1.99	2.20	0.42	1.81	1.81
Clayton	0.08	2.95	3.38	0.09	2.41	2.49
Frank	0.72	1.34	1.60	0.56	1.36	1.42
Gumbel	0.91	1.08	1.17	0.60	1.24	1.28

This is consistent with the values of $\tilde{S}_j(\hat{\alpha}_j)$; they are small for log-copula, Frank, and Gumbel compared with the value for Clayton copula. Because Clayton copula is the only one among these four copulas that has lower tail dependence, this suggests that neither data set is likely to have lower tail dependence. Based on the minimum value criterion, the Gumbel copula is selected for both data sets,[5] suggesting the possibility of upper tail dependence in both data sets. In addition to using the values of $\tilde{S}_j(\hat{\alpha}_j)$ for copula model selection, Table 5 indicates that *p*-values of the test provide an alternative model selection criterion. For both data sets, all three criteria, $\tilde{S}_j(\hat{\alpha}_j)$, $\tilde{S}_j(\tilde{\alpha}_j)$, and the *p*-value, select the same copula class.

6. CONCLUSIONS

This paper makes two contributions to semiparametric survival model selection for bivariate failure-time data. First, we provide a rigorous treatment of the appropriate choice of the parametric estimator of the copula parameter in model selection. This differs from model specification testing as done in Wang and Wells (2000) because in model specification testing, under the null hypothesis, the parametric model correctly specifies the true model and hence any root-*n* consistent parametric estimator under the null hypothesis can be used to construct model specification tests. In model selection tests, on the other hand, there is no guarantee that any of the models being entertained correctly specifies the true model even under the null hypothesis. This makes the choice of the parametric estimator critical in model selection tests, and we recommend estimating the pseudo-true parameter by minimizing the selection criterion function. Although in this paper we address this issue in the specific context of model selection for bivariate failure-time data using the integrated squared difference criterion, it is clear that the whole idea applies to model selection tests in other contexts and to any selection criterion function. Second, we address the issue of data reuse in copula model selection pointed out by Peña in his

https://doi.org/10.1017/S0266466607070181 Published online by Cambridge University Press

Comment on Wang and Wells (2000) by establishing a formal statistical test for copula model selection for bivariate failure-time data.

Extension of our model selection test to multivariate failure-time data is straightforward. As the test is based on comparing $K^o(v)$ with $K_j(v;\alpha_j^*)$, it is consistent for Archimedean copulas. For selection of multiple multivariate parametric copulas for complete data, Chen and Fan (2005) proposed pseudo-likelihood ratio tests that are consistent for any parametric copulas, including Archimedean and non-Archimedean families. Using the two-step estimator of Shih and Louis (1995) and that of Lin and Ying (1993) for various censored data, we can develop pseudo–likelihood ratio tests for general copula model selection for censored data. We are currently working on this.

NOTES

1. As suggested by a referee, we emphasize here that the function C^o in (2) is not a survival function and, likewise, the function C_α in representation (1) of Wang and Wells (2000) is a joint c.d.f., but not a survival function.

2. We choose a small sample size $n = 250$ to match that in Wang and Wells (2000). Nevertheless, we have also tried larger sample size $n = 500$. For the Monte Carlo studies with larger sample size, the power of our model selection test is much improved.

3. Fortran code is available from the authors upon request.

4. As discussed in Romano and Wolf (2005), White's reality check test is asymptotically undersized unless the least favorable curve (LFC) holds. In our case, because $S_j(\alpha_j^*) \neq S_1(\alpha_1^*)$, we do not have LFC and hence expect our test to be undersized. As demonstrated by Romano and Wolf (2005), Studentization may improve the size and power of the reality check test. This will be explored elsewhere.

5. The values of $\tilde{S}_j(\tilde{\alpha}_j)$ we obtained for the first data set lead to a different ranking of the copulas from that of Wang and Wells (2000), although they lead to the same ranking as that in Wang and Wells (2000) for the second data set. They also lead to the same ranking of the copulas as $\tilde{S}_j(\hat{\alpha}_j)$ for both data sets.

REFERENCES

Barbe, P., A. Genest, K. Ghoudi, & R. Bruno (1996) On Kendall's process. *Journal of Multivariate Analysis* 58, 197–229.

Chen, X. & Y. Fan (2005) Pseudo-likelihood ratio tests for model selection in semiparametric multivariate copula models. *Canadian Journal of Statistics* 33, 389–414.

Chen, X. & Y. Fan (2006a) Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *Journal of Econometrics* 135, 125–154.

Chen, X. & Y. Fan (2006b) Estimation of copula-based semiparametric time series models. *Journal of Econometrics* 130, 307–335.

Chen, X., Y. Fan, & A. Patton (2003) Simple Tests for Models of Dependence between Multiple Financial Time Series: With Applications to U.S. Equity Returns and Exchange Rates. Manuscript, New York University, Vanderbilt University, and London School of Economics.

Chen, X., Y. Fan, & V. Tsyrennikov (2006) Efficient estimation of semiparametric multivariate copula models. *Journal of the American Statistical Association* 101, 1228–1240.

Clayton, D.G. (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. & J. Cuzick (1985) Multivariate generalizations of the proportional hazards model (with discussion). *Journal of the Royal Statistical Society, Series A* 148, 82–117.

Dabrowska, D.M. (1988) Kaplan-Meier estimates on the plane. *Annals of Statistics* 16, 1475–1489.

https://doi.org/10.1017/S0266466607070181 Published online by Cambridge University Press

Dabrowska, D.M. (1989) Kaplan-Meier estimate on the plane: Weak convergence, LIL, and the bootstrap. *Journal of Multivariate Analysis* 29, 308–325.

Danahy, D.J., D.T. Burwell, W.S. Aranow, & R. Prakash (1977) Sustained henodynamic and anti-anginal effect of high-dose oral isosorbide dinitrate. *Circulation* 55, 381–387.

Embrechts, P., A. McNeil, & D. Straumann (2002) Correlation and dependence properties in risk management: Properties and pitfalls. In M. Dempster (ed.), *Risk Management: Value at Risk and Beyond*, pp. 176–223. Cambridge University Press.

Fermanian, J.-D. (2005) Goodness of fit tests for copulas. *Journal of Multivariate Analysis* 95, 119–152.

Frees, Edward W. & E.A. Valdez (1998) Understanding relationships using copulas. *North American Actuarial Journal* 2, 1–25.

Genest, C. (1987) Frank's family of bivariate distributions. *Biometrika* 74, 549–555.

Genest, C., K. Ghoudi, & L.-P. Rivest (1995) A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* 82, 543–552.

Genest, C., J.-F. Quessy, & B. Rémillard (2006) Goodness-of-fit procedures for copula models based on the integral probability transformation. *Scandinavian Journal of Statistics* 33, 337–366.

Genest, C. & L.-P. Rivest (1993) Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association* 88, 1034–1043.

Genest, C. & B. Werker (2002) Conditions for the asymptotic semiparametric efficiency of an omnibus estimator of dependence parameters in copula models. In C.M. Cuadras, J. Fortiana, & J.A. Rodríguez Lallena (eds.), *Distributions with Given Marginals and Statistical Modelling*, pp. 103–112. Kluwer.

Gonçalves, S. & H. White (2004) Maximum likelihood and the bootstrap for nonlinear dynamic models. *Journal of Econometrics* 119, 199–219.

Granger, C.W.J., T. Teräsvirta, & A. Patton (2006) Common factors in conditional distributions for bivariate time series. *Journal of Econometrics* 132, 43–57.

Heckman, J.J. & B.E. Honoré (1989) The identifiability of the competing risks model. *Biometrika* 76, 325–330.

Hougaard, P. (1986) Survival models for heterogeneous populations derived from stable distributions. *Biometrika* 73, 387–396.

Hu, L. (2006) Dependence patterns across financial markets: A mixed copula approach. *Applied Financial Economics* 16, 717–729.

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

Lee, L. (1983) Generalized econometric models with selectivity. *Econometrica* 51, 507–512.

Lin, D. & Z. Ying (1993) A simple nonparametric estimator of the bivariate survival function under univariate censoring. *Biometrika* 80, 573–581.

McGilchrist, C.A. & C.W. Aisbett (1991) Regression with frailty in survival analysis. *Biometrics* 47, 461–466.

Nelsen, R.B. (1997) Dependence and order in families of Archimedean copulas. *Journal of Multivariate Analysis* 60, 111–122.

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

Oakes, D. (1982) A concordance test for independence in the presence of bivariate censoring. *Biometrics* 38, 451–455.

Oakes, D. (1986) Semiparametric inference in a model for association in bivariate survival data. *Biometrika* 73, 353–361.

Oakes, D. (1989) Bivariate survival models induced by frailties. *Journal of the American Statistical Association* 84, 487–493.

Oakes, D. (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.

Patton, A.J. (2006) Modeling asymmetric exchange rate dependence. *International Economic Review* 47, 527–556.

https://doi.org/10.1017/S0266466607070181 Published online by Cambridge University Press

Romano, J.P. & M. Wolf (2005) Stepwise multiple testing as formalized data snooping. *Econometrica* 73, 1237–1282.

Shih, J. & T. Louis (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* dimensions et leurs marges. *Publications de l'Institut de statistique de l'Université de Paris* 8, 229–231.

Van der Vaart, A. & J. Wellner (1996) *Weak Convergence and Empirical Processes: With Applications to Statistics.* Springer-Verlag.

Vuong, Q.H. (1989) Likelihood ratio test for model selection and non-nested hypotheses. *Econometrica* 57, 307–333.

Wang, W. & M. Wells (2000) Model selection and semiparametric inference for bivariate failure-time data (with Comment by V. Peña and Rejoinder by the authors). *Journal of the American Statistical Association* 95, 62–76.

White, H. (1994) *Estimation, Inference and Specification Analysis*. Cambridge University Press.

White, H. (2000) A reality check for data snooping. *Econometrica* 68, 1097–1126.

APPENDIX: Technical Proofs

Proof of Proposition 3.1. Recall that $\hat{\alpha}_i = \text{argmin}_{\alpha_i} \tilde{S}_i(\alpha_i)$, where $\tilde{S}_i(\alpha_i) = \int_\varsigma^1 [\tilde{K}(v) - K_i(v;\alpha_i)]^2\, dv$. One can apply Theorems 3.4 and 6.2 in White (1994) to prove (i) and (ii).

(i) Given the conditions of Proposition 3.1, we only need to show the uniform convergence of $\tilde{S}_i(\alpha_i)$ to $S_i(\alpha_i)$ over $\mathcal{A}_i$ in probability; see Theorem 3.4 in White (1994). This follows from consistency of $\tilde{K}(v)$, the fact that $K^o(v)$ and $K_i(v;\alpha_i)$ are bounded, and the next inequality:

$$|\tilde{S}_i(\alpha_i) - S_i(\alpha_i)| \leq \int_\varsigma^1 |\tilde{K}(v) - K^o(v)|\,|\tilde{K}(v) + K^o(v) - 2K_i(v;\alpha_i)|\, dv.$$

(ii) By Theorem 6.2 in White (1994), it suffices to show that
(a) $n^{1/2} D_\alpha \tilde{S}_i(\alpha_i^*) \to -2\int_\varsigma^1 X(v) D_\alpha K_i(v;\alpha_i^*)\, dv$ in distribution, and
(b) $D_\alpha^2 \tilde{S}_i(\alpha_i) = 2\int_\varsigma^1 [D_\alpha K_i(v;\alpha_i) D_\alpha' K_i(v;\alpha_i)]\, dv - \int_\varsigma^1 [K^o(v) - K_i(v;\alpha_i)] D_\alpha^2 K_i(v;\alpha_i)\, dv + o_p(1)$ uniformly in $\alpha_i \in \mathcal{A}_i$.

Using the fact that $D_\alpha S_i(\alpha_i^*) = 0$, one can easily show that

$$D_\alpha \tilde{S}_i(\alpha_i^*) = -2\int_\varsigma^1 [\tilde{K}(v) - K^o(v)] D_\alpha K_i(v;\alpha_i^*)\, dv. \tag{A.1}$$

(a) follows from (A.1) and Lemma 2.1, whereas (b) follows from Lemma 2.1 and the following equation:

$$D_\alpha^2 \tilde{S}_i(\alpha_i) = 2\int_\varsigma^1 [D_\alpha K_i(v;\alpha_i) D_\alpha' K_i(v;\alpha_i)]\, dv - \int_\varsigma^1 [\tilde{K}(v) - K_i(v;\alpha_i)] D_\alpha^2 K_i(v;\alpha_i)\, dv. \tag{A.2}$$

https://doi.org/10.1017/S0266466607070181 Published online by Cambridge University Press

(iii) It is straightforward and omitted.

(iv) As in Wang and Wells (2000), we decompose $\tilde{S}_i(\hat{\alpha}_i)$ as follows:

$$\begin{aligned}
\tilde{S}_i(\hat{\alpha}_i) &= \int_\zeta^1 [\tilde{K}(v) - K_i(v;\hat{\alpha}_i)]^2\, dv \\
&= \int_\zeta^1 [\tilde{K}(v) - K^o(v)]^2\, dv + \int_\zeta^1 [K^o(v) - K_i(v;\alpha_i^*)]^2\, dv \\
&\quad + \int_\zeta^1 [K_i(v;\alpha_i^*) - K_i(v;\hat{\alpha}_i)]^2\, dv \\
&\quad + 2\int_\zeta^1 [\tilde{K}(v) - K^o(v)][K^o(v) - K_i(v;\alpha_i^*)]\, dv \\
&\quad + 2\int_\zeta^1 [\tilde{K}(v) - K^o(v)][K_i(v;\alpha_i^*) - K_i(v;\hat{\alpha}_i)]\, dv \\
&\quad + 2\int_\zeta^1 [K^o(v) - K_i(v;\alpha_i^*)][K_i(v;\alpha_i^*) - K_i(v;\hat{\alpha}_i)]\, dv \\
&= A_1 + A_2 + A_3 + 2B_1 + 2B_2 + 2B_3.
\end{aligned}$$

Lemma 2.1 and Proposition 3.1(ii) imply $A_1 = O_p(n^{-1})$, $A_3 = O_p(n^{-1})$, and $B_2 = O_p(n^{-1})$. By Taylor series expansion,

$$\begin{aligned}
B_3 &= \int_\zeta^1 [K^o(v) - K_i(v;\alpha_i^*)] D_\alpha K_i(v;\alpha_i^*)\, dv(\alpha_i^* - \hat{\alpha}_i) \\
&\quad + (\alpha_i^* - \hat{\alpha}_i)' \int_\zeta^1 [K^o(v) - K_i(v;\alpha_i^*)] D_\alpha^2 K_i(v;\bar{\alpha}_i^*)\, dv(\alpha_i^* - \hat{\alpha}_i) \\
&= (\alpha_i^* - \hat{\alpha}_i)' \int_\zeta^1 [K^o(v) - K_i(v;\alpha_i^*)] D_\alpha^2 K_i(v;\bar{\alpha}_i^*)\, dv(\alpha_i^* - \hat{\alpha}_i) \\
&= O_p(n^{-1}),
\end{aligned}$$

where $\bar{\alpha}_i^*$ lies between α_i^* and $\hat{\alpha}_i$ and the second equality follows from the fact that $D_\alpha S_i(\alpha_i^*) = 0$ and the last from Proposition 3.1(i) and (ii). Consequently, we have

$$\tilde{S}_i(\hat{\alpha}_i) - S_i(\alpha_i^*) = 2B_1 + O_p(n^{-1}). \tag{A.3}$$

The limiting distribution of $\sqrt{n}[\tilde{S}_i(\hat{\alpha}_i) - S_i(\alpha_i^*)]$ for the case where $\mathcal{M}_i$ misspecifies the true copula follows from Lemma 2.1, the continuous mapping theorem, and (A.3). For the correct specification case, see Theorem 3 in Wang and Wells (2000). ■

https://doi.org/10.1017/S0266466607070181 Published online by Cambridge University Press

Proof of Proposition 3.2.

(i) is straightforward and omitted.

(ii) A similar proof to that of Proposition 3.1(iv) suggests that

$$\begin{aligned}
&\tilde{S}_i(\tilde{\alpha}_i) - S_i(\alpha_{iL}^*) \\
&\quad = 2\int_\zeta^1 [\tilde{K}(v) - K^o(v)][K^o(v) - K_i(v;\alpha_{iL}^*)]\,dv \\
&\qquad + 2\int_\zeta^1 [K^o(v) - K_i(v;\alpha_{iL}^*)][K_i(v;\alpha_{iL}^*) - K_i(v;\tilde{\alpha}_i)]\,dv + O_p(n^{-1}) \\
&\quad = 2\int_\zeta^1 \{[\tilde{K}(v) - K^o(v)] - D_\alpha K_i(v;\alpha_{iL}^*)(\tilde{\alpha}_i - \alpha_{iL}^*)\}[K^o(v) - K_i(v;\alpha_{iL}^*)]\,dv \\
&\qquad + O_p(n^{-1}). \qquad \textbf{(A.4)}
\end{aligned}$$

The result follows from Lemma 2.1, the assumption on $\tilde{\alpha}_i$, the preceding equation, and the continuous mapping theorem.

(iii) Theorem 3 in Wang and Wells (2000). ■

Proof of Proposition 4.1. Equation (A.3) implies that

$$\begin{aligned}
\sqrt{n}\{T_{in} - [S_1(\alpha_1^*) - S_i(\alpha_i^*)]\} &= 2\int_\zeta^1 \{\sqrt{n}[\tilde{K}(v) - K^o(v)]\} \\
&\quad \times [K_i(v;\alpha_i^*) - K_1(v;\alpha_1^*)]\,dv + O_p(n^{-1/2}).
\end{aligned}$$

The conclusion follows immediately from Lemma 2.1 and the continuous mapping theorem. ■

To prove Theorem 4.3, we first verify the consistency of the proposed bootstrap procedure for $n^{1/2}[\tilde{K}(v) - K^o(v)]$ and $n^{1/2}[\hat{\alpha}_i - \alpha_i^*]$. The results are stated in the following lemma. For compactness, we borrow the notation $o_{p^*}(1)$ pr-P and $O_{p^*}(1)$ pr-P from Gonçalves and White (2004): $W_n^* = o_{p^*}(1)$ pr-P if W_n^* approaches zero in P^* probability conditional on the original data and for all samples except a set of measure approaching zero, where $W_n^* = O_{p^*}(1)$ pr-P if W_n^* is bounded in P^* probability conditional on the original data and for all samples except a set of measure approaching zero.

LEMMA A.1. *Under the conditions of Proposition 3.1,*

(i) the limit process of $n^{1/2}[\tilde{K}^*(\cdot) - \tilde{K}(\cdot)]$ conditional on the original data converges to the limit process of $n^{1/2}[\tilde{K}(\cdot) - K^o(\cdot)]$ in probability;

(ii) the conditional distribution of $n^{1/2}[\hat{\alpha}^* - \hat{\alpha}]$ given the original data converges to the distribution of $n^{1/2}[\hat{\alpha} - \alpha^*]$ in probability.

https://doi.org/10.1017/S0266466607070181 Published online by Cambridge University Press

Proof of Theorem 4.3. It suffices to show that

$$\sup_{z\in\mathcal{R}^{M-1}}\left|P^*\left(\begin{pmatrix} n^{1/2}T^*_{2nC}\\ \vdots\\ n^{1/2}T^*_{MnC}\end{pmatrix}\le z\right)\right.$$
$$\left.-P\left(\begin{pmatrix} n^{1/2}(T_{2n}-[S_1(\alpha^*_1)-S_2(\alpha^*_2)])\\ \vdots\\ n^{1/2}(T_{Mn}-[S_1(\alpha^*_1)-S_M(\alpha^*_M)])\end{pmatrix}\le z\right)\right|=o_p(1). \tag{A.5}$$

Similar to the decomposition of $\tilde{S}_i(\hat{\alpha}_i)$, we get

$$\begin{aligned}\tilde{S}^*_i(\hat{\alpha}^*_i)&=\int_\zeta^1[\tilde{K}^*(v)-K_i(v;\hat{\alpha}^*_i)]^2\,dv\\ &=\int_\zeta^1[\tilde{K}^*(v)-\tilde{K}(v)]^2\,dv+\int_\zeta^1[\tilde{K}(v)-K_i(v;\hat{\alpha}_i)]^2\,dv\\ &\quad+\int_\zeta^1[K_i(v;\hat{\alpha}_i)-K_i(v;\hat{\alpha}^*_i)]^2\,dv\\ &\quad+2\int_\zeta^1[\tilde{K}^*(v)-\tilde{K}(v)][\tilde{K}(v)-K_i(v;\hat{\alpha}_i)]\,dv\\ &\quad+2\int_\zeta^1[\tilde{K}^*(v)-\tilde{K}(v)][K_i(v;\hat{\alpha}_i)-K_i(v;\hat{\alpha}^*_i)]\,dv\\ &\quad+2\int_\zeta^1[\tilde{K}(v)-K_i(v;\hat{\alpha}_i)][K_i(v;\hat{\alpha}_i)-K_i(v;\hat{\alpha}^*_i)]\,dv\\ &=A^*_1+\tilde{S}_i(\hat{\alpha}_i)+A^*_3+2B^*_1+2B^*_2+2B^*_3.\end{aligned}$$

By Lemma A.1, we get $A^*_1=O_{p^*}(n^{-1})$ pr-P, $A^*_3=O_{p^*}(n^{-1})$ pr-P, and $B^*_2=O_{p^*}(n^{-1})$ pr-P. By Taylor series expansion,

$$\begin{aligned}B^*_3&=\int_\zeta^1[\tilde{K}(v)-K_i(v;\hat{\alpha}_i)]D_\alpha K_i(v;\hat{\alpha}_i)\,dv(\hat{\alpha}_i-\hat{\alpha}^*_i)\\ &\quad+(\hat{\alpha}_i-\hat{\alpha}^*_i)'\int_\zeta^1[\tilde{K}(v)-K_i(v;\hat{\alpha}_i)]D^2_\alpha K_i(v;\bar{\alpha}^*_i)\,dv(\hat{\alpha}_i-\hat{\alpha}^*_i)\\ &=(\hat{\alpha}_i-\hat{\alpha}^*_i)'\int_\zeta^1[\tilde{K}(v)-K_i(v;\hat{\alpha}_i)]D^2_\alpha K_i(v;\bar{\alpha}^*_i)\,dv(\hat{\alpha}_i-\hat{\alpha}^*_i)\\ &=O_{p^*}(n^{-1})\ \text{pr}-\text{P},\end{aligned}$$

https://doi.org/10.1017/S0266466607070181 Published online by Cambridge University Press

where $\bar{\alpha}_i^*$ lies between $\hat{\alpha}_i^*$ and $\hat{\alpha}_i$ and the second equality follows from the fact that $D_\alpha \tilde{S}_i(\hat{\alpha}_i) = 0$ and the last from Lemma A.1. Consequently, we have

$$\sqrt{n}\{T_{in}^* - T_{in}\} = 2\int_\zeta^1 \{\sqrt{n}[\tilde{K}^*(v) - \tilde{K}(v)]\}[K_i(v;\hat{\alpha}_i) - K_1(v;\hat{\alpha}_1)]\,dv + o_{p^*}(1).$$

The conclusion follows from Lemmas 2.1 and A.1 and Proposition 3.1. ■

Proof of Lemma A.1.

(i) By Theorem 3.9.11 in Van der Vaart and Wellner (1996) and Corollary 2.2 in Dabrowska (1989), it suffices to show that the map $\bar{K}(F)(\cdot)$: $\mathcal{D}([0,\tau_1] \times [0,\tau_2]) \to l^\infty([0,1])$ is Hadamard differentiable at F tangentially to $\mathcal{C}([0,\tau_1] \times [0,\tau_2])$, where $\bar{K}(F)(\cdot) = E[I(F(X,Y) > \cdot)]$. In fact for $H \in \mathcal{C}([0,\tau_1] \times [0,\tau_2])$, its Hadamard derivative at F is given by

$$\bar{K}_F(H)(v) = \iint H(x,y)\mu_v(dx,dy) + \iint I(F(x,y) > v)H(dx,dy).$$

Let $t_n \to 0$ and $H_n \to H \in \mathcal{C}([0,\tau_1] \times [0,\tau_2])$. We need to show that

$$\frac{\bar{K}(F + t_n H_n) - \bar{K}(F)}{t_n} \to \bar{K}_F(H). \tag{A.6}$$

The left-hand side of (A.6) can be decomposed as follows:

$$\begin{aligned}
&\frac{\bar{K}(F + t_n H_n) - \bar{K}(F)}{t_n}(v) \\
&= t_n^{-1}\left\{\iint I(F(x,y) + t_n H_n(x,y) > v)(F + t_n H_n)(dx,dy)\right. \\
&\qquad \left. - \iint I(F(x,y) > v)F(dx,dy)\right\} \\
&= t_n^{-1}\iint [I(F(x,y) + t_n H_n(x,y) > v) - I(F(x,y) > v)]F(dx,dy) \\
&\quad + \iint I(F(x,y) > v)H_n(dx,dy) \\
&\quad + \iint [I(F(x,y) + t_n H_n(x,y) > v) - I(F(x,y) > v)]H_n(dx,dy) \\
&\equiv A_n(v) + B_n(v) + C_n(v).
\end{aligned} \tag{A.7}$$

Noting the similarity between $A_n(v)$, $B_n(v)$, and $C_n(v)$ and $\hat{\alpha}(v)$, $\hat{\beta}(v)$, and $\hat{\gamma}(v)$, respectively, in the proof of Theorem 1 in Wang and Wells (2000), one can follow their proof to verify (A.6) by replacing weak convergence and tightness in their proof with uniform

https://doi.org/10.1017/S0266466607070181 Published online by Cambridge University Press

convergence and equicontinuity for the relevant functions in our case. For example, H_n plays the role of $\sqrt{n}[\hat{F} - F]$ and t_n that of $n^{-1/2}$ in Wang and Wells (2000).

(ii) First, we show that $[\hat{\alpha}_i^* - \hat{\alpha}_i] = o_{p^*}(1)$ pr-P. Using Lemma A.2 in Gonçalves and White (2004), it suffices to show that $\sup_{\alpha_i} |\tilde{S}_i^*(\alpha_i) - \tilde{S}_i(\alpha_i)| = o_{p^*}(1)$ pr-P. It follows from (i) of this lemma, because

$$\sup_{\alpha_i} |\tilde{S}_i^*(\alpha_i) - \tilde{S}_i(\alpha_i)|$$

$$= \sup_{\alpha_i} \left| \int_\zeta^1 \{[\tilde{K}^*(v) - K_i(v;\alpha_i)]^2 - [\tilde{K}(v) - K_i(v;\alpha_i)]^2\}\, dv \right| \tag{A.8}$$

$$\leq \int_\zeta^1 |\tilde{K}^*(v) - \tilde{K}(v)| \left[\tilde{K}^*(v) + \tilde{K}(v) + 2 \sup_{\alpha_i} K_i(v;\alpha_i) \right] dv. \tag{A.9}$$

Now we can apply Lemma A.3 in Gonçalves and White (2004) to show (ii). It suffices to verify conditions (b3) and (b4) of Lemma A.3 in Gonçalves and White (2004). They are

(b3) The conditional distribution of $n^{1/2} D_\alpha \tilde{S}_i^*(\hat{\alpha}_i)$ converges to $-2 \int_\varsigma^1 X(v)\, D_\alpha K_i(v;\alpha_i^*)\, dv$ in probability;

(b4) $\sup_{\alpha_i} |D_\alpha^2 \tilde{S}_i^*(\alpha_i) - D_\alpha^2 \tilde{S}_i(\alpha_i)| = o_{p^*}(1)$ pr-P.

Using the fact that $D_\alpha \tilde{S}_i(\hat{\alpha}_i) = 0$, one can show that

$$n^{1/2} D_\alpha \tilde{S}_i^*(\hat{\alpha}_i) = -2 \int_\varsigma^1 \{n^{1/2}[\tilde{K}^*(v) - \tilde{K}(v)]\} D_\alpha K_i(v;\hat{\alpha}_i)\, dv. \tag{A.10}$$

(b3) follows from (A.10) and (i) of this lemma. Similar to (A.2), one can show that

$$\sup_{\alpha_i} |D_\alpha^2 \tilde{S}_i^*(\alpha_i) - D_\alpha^2 \tilde{S}_i(\alpha_i)| \leq 2 \int_\varsigma^1 |\tilde{K}^*(v) - \tilde{K}(v)| \sup_{\alpha_i} |D_\alpha^2 K_i(v;\alpha_i)|\, dv$$

$$= o_{p^*}(1) \text{pr-P}$$

by part (i) of this lemma, Lemma 2.1, and boundedness of $D_\alpha^2 K_i(v;\alpha_i)$. ■

https://doi.org/10.1017/S0266466607070181 Published online by Cambridge University Press