

A MODEL SELECTION TEST FOR BIVARIATE FAILURE-TIME DATA

by

Xiaohong Chen and Yanqin Fan



Working Paper No. 04-W21

October 2004

DEPARTMENT OF ECONOMICS
VANDERBILT UNIVERSITY
NASHVILLE, TN 37235

www.vanderbilt.edu/econ

A Model Selection Test for Bivariate Failure-Time Data*

Xiaohong Chen[†] Yanqin Fan[‡]

August 2004

Abstract

In this paper, we address two important issues in 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 model selection, 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 applications provide strong support to our theoretical findings.

KEY WORDS: Archimedean copula; Bivariate survival function; Data reuse; Minimum-distance estimation; Model selection.

*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 the Fortran code for computing the bivariate Kaplan-Meier estimator of Dabrowska (1988).

[†]Xiaohong Chen, Department of Economics, New York University, 269 Mercer Street, 7th Floor, New York, NY 10003, USA; E-mail: xiaohong.chen@nyu.edu; Tel: 212-998-8970; Fax: 212-995-4186.

[‡]Yanqin Fan, Department of Economics, Vanderbilt University, VU Station B #351819, 2301 Vanderbilt Place, Nashville, TN 37235, USA. E-mail: yanqin.fan@vanderbilt.edu; Tel: 615-322-3796; Fax: 615-343-8495.

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 $F_j^o(\cdot)$ ($j = 1, 2$). Assume that F_1^o and F_2^o are continuous. By the Sklar's (1959) theorem, there exists a unique copula function C^o such that $F^o(x, y) \equiv C^o(F_1^o(x), F_2^o(y))$. This decomposition of the joint survival function leads naturally to the class of semiparametric survival functions in which the marginal survival functions are unspecified, but the copula function is parameterized: $C^o(u, v) = C^o(u, v; \alpha^o)$ for some parametric copula function $C^o(u, v; \alpha)$ and some $\alpha^o \in \mathcal{A}$. This class of semiparametric survival functions has been used widely in survival analysis, where modeling and estimating the dependence structure between lifetime variables is of interest, see Joe (1997), Nelsen (1999), Clayton (1978) and Oakes (1982) for examples of applications.

One important issue that an applied researcher faces in applying this class of semiparametric survival functions 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, et al. (2003), Fermanian (2003), and Genest, et al. (2003), 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 copula, extending existing results for complete data to censored 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 square difference between the true copula and a parametric copula for different parametric copula models; the one with the smallest value of the integrated square 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 square difference, we need to estimate the parametric copulas. In the above application, Wang and Wells (2000) estimate the parametric copulas by inverting Kendall's τ instead of minimizing the integrated square difference over all possible values

of the parameter. 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 square 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 copulas.

The purpose of this paper is two-fold: First, we demonstrate that for model selection, estimators of the parametric copulas based on minimizing the selection criterion function may be preferred to other estimators, since 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 Pena 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 minimizes a distance between itself and the underlying data-generating process.” Second, we provide a formal statistical test for copula model selection to address the issue of data reuse as raised by Pena in his Comment on Wang and Wells (2000). Our test draws on the reality check for data snooping in White (2000).

The remainder of this paper is organized as follows. Section 2 provides a rigorous treatment of the estimation of the criterion function. In Section 3, we put copula model selection in the context of hypothesis testing and develop a test for model selection. Section 4 presents results based on a simulation experiment and two real data sets. Section 5 concludes. All the proofs are gathered to Appendix B.

2 Model Selection

We build on the recent work of Wang and Wells (2000). Our notation will be similar. Let (C_1, C_2) be the nuisance censoring variables. With right-censored data, one observes $(\tilde{X}, \tilde{Y}) = (X \wedge C_1, Y \wedge C_2)$ and a pair of indicators, $(\delta_1, \delta_2) = \{I(X \leq C_1), I(Y \leq C_2)\}$, where $X \wedge C_1 = \min(X, C_1)$ and $I(\cdot)$ is the indicator function. Suppose n i.i.d. observations $\{(\tilde{X}_i, \tilde{Y}_i, \delta_{1i}, \delta_{2i})\}_{i=1}^n$ are available, where $(\tilde{X}_i, \tilde{Y}_i) = (X_i \wedge C_{1i}, Y_i \wedge C_{2i})$ and $(\delta_{1i}, \delta_{2i}) = \{I(X_i \leq C_{1i}), I(Y_i \leq C_{2i})\}$.

For each $i = 1, 2, \dots, M$, let $\mathcal{M}_i = \{C_i(u_1, u_2; \alpha_i) : \alpha_i \in \mathcal{A}_i \subset \mathcal{R}^{a_i}\}$ be a class of

parametric Archimedean copulas. Wang and Wells (2000) propose a copula model selection procedure given a finite number of models $\{\mathcal{M}_i\}_{i=1}^M$. In particular, they build their model selection procedure on the integrated square difference between $\{K_i(v; \alpha_i) : \alpha_i \in \mathcal{A}_i\}$ and $K^o(v)$, where $K^o(v)$ is defined as $K^o(v) \equiv \Pr(F^o(X, Y) \leq v)$, the distribution function of the random variable $F^o(X, Y)$, and $K_i(v; \alpha_i)$ is the analogue of $K^o(v)$ associated with model class \mathcal{M}_i . The justification for using the functions $\{K_i(v; \alpha_i)\}_{i=1}^M$ and $K^o(v)$ is that they uniquely determine the corresponding copulas in the Archimedean family, see Genest and Rivest (1993).

Following Wang and Wells (2000), we define

$$S_i(\alpha_i) = \int_{\varsigma}^1 [K^o(v) - K_i(v; \alpha_i)]^2 dv \text{ for } i = 1, \dots, M,$$

where ς is introduced in Wang and Wells (2000), see Lemma A.1 in Appendix A. Let $\alpha_i^* = \arg \min_{\alpha_i} S_i(\alpha_i)$. Then $K_i(v; \alpha_i^*)$ corresponds to the copula in $\{C_i(u_1, u_2; \alpha_i) : \alpha_i \in \mathcal{A}_i\}$ closest to the true copula according to the integrated square difference criterion and the distance from the i -th model class to the true copula is $S_i(\alpha_i^*)$ ($i = 1, \dots, M$).

For $i = 1, \dots, M$, the value α_i^* is referred to as the pseudo-true value, as it leads to the copula function in the i -th model class that best approximates the true copula. Obviously if the copula $C_i(u_1, u_2; \alpha_i)$ correctly specifies the true copula in the sense that there exists α_{io} such that $K^o(v) = K_i(v; \alpha_{io})$, then $\alpha_i^* = \alpha_{io}$ and $S_i(\alpha_i^*) = 0$. In general, however, all the parametric copulas being entertained may misspecify the true copula and the M best copulas $K_i(v; \alpha_i^*)$, one from each class \mathcal{M}_i , $i = 1, \dots, M$, 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 $j = 1, \dots, M$ and $j \neq i$. In practice, neither $K^o(v)$ nor α_i^* is known. Any model selection procedure will then be based on estimates of $K^o(v)$ and α_i^* or equivalently of $S_i(\alpha_i^*)$, $i = 1, \dots, M$.

2.1 Estimation of $K^o(v)$ and α_i^*

When complete observations on (X, Y) , $\{(X_i, Y_i)\}_{i=1}^n$, are available, Genest and Rivest (1993) propose to estimate $K^o(v)$ by $K_n(v) = \frac{1}{n} \sum_{i=1}^n I(F_n(X_i, Y_i) \leq v)$, 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, see also Ghoudi and Remillard (1998, 2003).

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

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

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

Lemma A.1 in Appendix A restates asymptotic properties of the estimator $\tilde{K}(v)$ established in Wang and Wells (2000). Conditions under which the results hold are satisfied by existing estimators $\hat{F}(x, y)$ under different censoring schemes, see Wang and Wells (2000) for references. 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 choice of an appropriate estimator of α_i^* is a delicate issue and deserves a rigorous treatment. Several consistent estimators of α_i^* , when \mathcal{M}_i correctly specifies the true model, are available in the literature, including the two-step estimator¹ of Shih and Louis (1995), the minimum distance estimator proposed in Wang and Wells (2000), and the estimator based on Kendall's τ . In this subsection, we study asymptotic properties of the minimum distance estimator in the general case that \mathcal{M}_i misspecifies the true model and discuss the pros and cons of basing model selection on other estimators in the next subsection.

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

$$\hat{\alpha}_i = \arg \min_{\alpha_i} \int_{\varsigma}^1 [\tilde{K}(v) - K_i(v; \alpha_i)]^2 dv \equiv \arg \min_{\alpha_i} \tilde{S}_i(\alpha_i).$$

Wang and Wells (2000) establish the asymptotic properties of $\hat{\alpha}_i$ when the copula class \mathcal{M}_i correctly specifies the true copula of (X, Y) . In this case, $\alpha_i^* = \alpha_{io}$. In general, since $\alpha_i^* = \arg \min_{\alpha_i} \int_{\varsigma}^1 [K^o(v) - K_i(v; \alpha_i)]^2 dv$ and $\tilde{K}(v)$ is a consistent estimator of $K^o(v)$, we expect $\hat{\alpha}_i$ to be a consistent estimator of α_i^* regardless of whether \mathcal{M}_i correctly specifies the true copula.

¹This is the same as the two-step estimator in Genest, et al. (1995) for complete data.

Proposition 2.1 *Suppose \mathcal{A}_i is a compact subset of \mathcal{R}^{a_i} with non-empty interior and α_i^* is the unique minimizer of $S_i(\alpha_i)$ over \mathcal{A}_i . Moreover, $K_i(v; \alpha_i)$ is twice differentiable with respect to α_i with bounded derivatives.*

(i) $\hat{\alpha}_i$ is consistent for α_i^* ;

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

$$Q_i = \left\{ \int_{\zeta}^1 [D_{\alpha}K_i(v; \alpha_i^*)D'_{\alpha}K_i(v; \alpha_i^*)] dv - \int_{\zeta}^1 [K^{\circ}(v) - K_i(v; \alpha_i^*)]D_{\alpha}^2K_i(v; \alpha_i^*)dv \right\}^{-1} \times \int_{\zeta}^1 X(v)D_{\alpha}K_i(v; \alpha_i^*)dv,$$

in which $X(v)$ is defined in Lemma A.1 in Appendix A;

(iii) $\tilde{S}_i(\hat{\alpha}_i)$ is a consistent estimator of $S_i(\alpha_i^*)$;

(iv) If \mathcal{M}_i misspecifies the true copula, then

$$\sqrt{n}[\tilde{S}_i(\hat{\alpha}_i) - S_i(\alpha_i^*)] \rightarrow \int_{\zeta}^1 X(v)[K^{\circ}(v) - K_i(v; \alpha_i^*)]dv \text{ in distribution.}$$

On the other hand, if \mathcal{M}_i correctly specifies the true copula, then $n\tilde{S}_i(\hat{\alpha}_i) \rightarrow \int_{\zeta}^1 \hat{X}_i^2(v)dv$, where $\hat{X}_i(v) = X(v) - D_{\alpha}K_i(v; \alpha_i^*)Q_i$.

Proposition 2.1 shows that in general $\hat{\alpha}_i$ consistently estimates the pseudo-true value α_i^* , extending the consistency result of Wang and Wells (2000) for α_{i0} when the i -th model class correctly specifies the true model. Moreover $\tilde{S}_i(\hat{\alpha}_i)$ consistently estimates $S_i(\alpha_i^*)$, justifying the use of $\tilde{S}_i(\hat{\alpha}_i)$ for model selection based on minimizing $S_i(\alpha_i^*)$ over $i = 1, \dots, M$. Proposition 2.1 (iv) also implies that the limiting distribution of $\sqrt{n}[\tilde{S}_i(\hat{\alpha}_i) - S_i(\alpha_i^*)]$ is the same as that of $\sqrt{n}[\tilde{S}_i(\alpha_i^*) - S_i(\alpha_i^*)]$ irrespective of whether \mathcal{M}_i correctly specifies the true model. The only difference is that when \mathcal{M}_i correctly specifies the true model, $\sqrt{n}[\tilde{S}_i(\alpha_i^*) - S_i(\alpha_i^*)]$ has a degenerate limiting distribution.

2.2 Which Parametric Estimator to Use?

Many estimators, including the two-step estimator of Shih and Louis (1995), share the consistency property of $\hat{\alpha}_i$ for α_{i0} when the i -th model correctly specifies the true model, but

may not share the consistency property for the pseudo-true value α_i^* in general. The reason is that unlike the true value α_{i0} when the i -th copula class correctly specifies the true copula, the pseudo-true value α_i^* when the i -th copula class misspecifies the true copula may change if a different distance measure is used.

To be more specific, let $\tilde{\alpha}_i$ be a parametric estimator other than $\hat{\alpha}_i$. 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, $\tilde{L}_i(\alpha_i)$. For example, for an Archimedean copula, the estimator based on inverting Kendall's tau can be obtained by minimizing $\tilde{L}_i(\alpha_i) = |\int_0^1 [\tilde{K}(v) - K_i(v; \alpha_i)] dv|$, since $\tau = 4 \int_0^1 [v - K(v)] dv + 1$. The two-step estimator of Shih and Louis (1995) minimizes the pseudo-likelihood function. Under regularity conditions, $\tilde{\alpha}_i = \alpha_{iL}^* + o_p(1)$, where $\alpha_{iL}^* = \operatorname{argmin}_{\alpha_i} [\operatorname{plim} \tilde{L}_i(\alpha_i)]$. For example, for the estimator based on inverting Kendall's τ , $\alpha_{iL}^* = \operatorname{argmin}_{\alpha_i} |\int_0^1 [K^o(v) - K_i(v; \alpha_i)] dv|$, while for the two-step estimator in Shih and Louis (1995), α_{iL}^* minimizes the Kullback-Leibler Information Criterion. In general, α_{iL}^* differs from α_i^* unless \mathcal{M}_i correctly specifies the true copula in which case both α_{iL}^* and α_i^* equal the true value of the parameter α_{i0} .

Proposition 2.2 *Suppose \mathcal{A}_i is a compact subset of \mathcal{R}^{α_i} with non-empty interior and $K_i(v; \alpha_i)$ is twice differentiable with respect to α_i with bounded derivatives. In addition, assume the estimator $\tilde{\alpha}_i$ satisfies: $\tilde{\alpha}_i = \alpha_{iL}^* + o_p(1)$ and $\sqrt{n}[\tilde{\alpha}_i - \alpha_{iL}^*] \rightarrow Q_{iL}$ in distribution for some $\alpha_{iL}^* \in \mathcal{A}_i$ and a non-degenerate distribution Q_{iL} .*

(i) $\tilde{S}_i(\tilde{\alpha}_i)$ is a consistent estimator of $S_i(\alpha_{iL}^*)$;

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

$$\sqrt{n}[\tilde{S}_i(\tilde{\alpha}_i) - S_i(\alpha_{iL}^*)] \rightarrow \int_{\zeta}^1 \hat{X}_{iL}(v) [K^o(v) - K_i(v; \alpha_{iL}^*)] dv \text{ in distribution,}$$

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

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

Given that in general $\alpha_{iL}^* \neq \alpha_i^*$ unless copula class \mathcal{M}_i correctly specifies the true copula, Proposition 2.2 shows that $\tilde{S}_i(\tilde{\alpha}_i)$ is in general an inconsistent estimator of $S_i(\alpha_i^*)$. Any model selection procedure based on $\tilde{S}_i(\tilde{\alpha}_i)$ ($i = 1, \dots, M$) may thus lead to incorrect inferences

unless the ranking of the M models based on $S_i(\alpha_{iL}^*)$ is the same as that based on $S_i(\alpha_i^*)$, or at the very least, the model that minimizes $S_i(\alpha_{iL}^*)$ over $i = 1, \dots, M$ also minimizes $S_i(\alpha_i^*)$ over $i = 1, \dots, M$. Another difference between using $\tilde{\alpha}_i$ and $\hat{\alpha}_i$ in model selection is that the limiting distribution of $\sqrt{n}[\tilde{S}_i(\tilde{\alpha}_i) - S_i(\alpha_{iL}^*)]$ is not the same as that of $\sqrt{n}[\tilde{S}_i(\alpha_{iL}^*) - S_i(\alpha_{iL}^*)]$ unless \mathcal{M}_i 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 Pena 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 $\tilde{S}_i(\hat{\alpha}_i)$ or $\tilde{S}_i(\tilde{\alpha}_i)$ when \mathcal{M}_i misspecifies the true model are most relevant to model selection.

3 A Model Selection Test

Results in the previous section suggest the use of $\tilde{S}_i(\hat{\alpha}_i)$ in model selection. Two drawbacks of basing model selection on point estimates $\tilde{S}_i(\hat{\alpha}_i)$ only are: First, the limiting distribution of $\sqrt{n}[\tilde{S}_i(\hat{\alpha}_i) - S_i(\alpha_i^*)]$ depends on model i under misspecification, even though it is the same as that of $\sqrt{n}[\tilde{S}_i(\alpha_i^*) - S_i(\alpha_i^*)]$; Second, the same data is used twice in obtaining $\tilde{S}_i(\hat{\alpha}_i)$, see Pena for more discussion on 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:

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

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 $\{C_i(u_1, u_2; \alpha_i)\}_{i=2}^M$ be the candidate copulas. The null and alternative hypotheses of interest are expressed as

$$H_0 : \max_{i=2, \dots, M} [S_1(\alpha_1^*) - S_i(\alpha_i^*)] \leq 0 \text{ and } H_1 : \max_{i=2, \dots, M} [S_1(\alpha_1^*) - S_i(\alpha_i^*)] > 0.$$

Under H_0 , \mathcal{M}_1 will be selected and under H_1 , the model minimizing $\tilde{S}_i(\hat{\alpha}_i)$ for $i = 2, \dots, M$ will be selected.

In contrast to goodness-of-fit tests developed in Genest, et al. (2003) 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 square difference.

Let $T_{in} = \tilde{S}_1(\hat{\alpha}_1) - \tilde{S}_i(\hat{\alpha}_i)$, where recall $\tilde{S}_i(\hat{\alpha}_i) = \int_{\zeta}^1 [\tilde{K}(v) - K_i(v; \hat{\alpha}_i)]^2 dv$ for $i = 1, \dots, M$. The following proposition provides the basis for our test.

Proposition 3.1 *Suppose that for $i = 1, 2, \dots, M$, the copula model i satisfies the conditions of Proposition 2.1. Then jointly $n^{1/2} \{T_{in} - [S_1(\alpha_1^*) - S_i(\alpha_i^*)]\}_{i=2, \dots, M} \rightarrow (Z_2, \dots, Z_M)'$ in distribution, 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_{ii} denote the variance of Z_i . Then $\sigma_{ii} = 0$ if and only if model i and the benchmark model 1 are generalized nested in the sense that $K_i(v; \alpha_i^*) = K_1(v; \alpha_1^*)$ for almost all $v \in [0, 1]$.

Proposition 3.1 and the continuous mapping theorem imply

$$\max_{i=2, \dots, M} n^{1/2} \{T_{in} - [S_1(\alpha_1^*) - S_i(\alpha_i^*)]\} \rightarrow \max_{i=2, \dots, M} Z_i \text{ in distribution.}$$

Hence under the Least Favorable Configuration (i.e., $S_i(\alpha_i^*) = S_1(\alpha_1^*)$ for $i = 2, \dots, M$), $\max_{i=2, \dots, M} [\sqrt{n}T_{in}] \rightarrow \max_{i=2, \dots, M} Z_i$ in distribution. This allows us to construct a test for H_0 .

Let $T_n = \max_{i=2, \dots, M} [\sqrt{n}T_{in}]$. Suppose $\sigma_{ii} > 0$ for at least one $i = 2, \dots, M$, then we will reject H_0 if $T_n > Z_{\alpha}$, where Z_{α} is the upper α -percentile of the distribution of $\max_{i=2, \dots, M} Z_i$. In general, the distribution of $\max_{i=2, \dots, M} Z_i$ is unknown, depending on the true distribution function $F^o(x, y)$ and the pseudo-true values α_i^* . To circumvent this difficulty, we propose to use 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, since 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 $K_i(v; \hat{\alpha}_i)$. 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 $\{(\tilde{X}_i^*, \tilde{Y}_i^*, \delta_{1i}^*, \delta_{2i}^*)\}_{i=1}^n$ be a random sample with replacement from the original data $\{(\tilde{X}_i, \tilde{Y}_i, \delta_{1i}, \delta_{2i})\}_{i=1}^n$ and let $\hat{F}^*(x, y)$, $\tilde{K}^*(v)$, and $\hat{\alpha}_i^*$ be the bootstrapped counterparts of $\hat{F}(x, y)$, $\tilde{K}(v)$, and $\hat{\alpha}_i$.

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

$$T_{inC}^* = T_{in}^* - T_{in}, \quad i = 2, \dots, M.$$

Step 3. Repeat Steps 1-2 for a large number of times and define the bootstrap value of T_n as $T_n^* = \max_{i=2, \dots, M} [n^{1/2} T_{inC}^*]$.

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

We show in Appendix B that the above bootstrap works for T_n .

THEOREM 3.2 *Under the conditions of Proposition 3.1,*

$$\sup_{z \in \mathcal{R}} |P^*(T_n^* \leq z) - P(\max_{i=2, \dots, M} [n^{1/2} T_{in} - \{S_1(\alpha_1^*) - S_i(\alpha_i^*)\}] \leq z)| = o_p(1),$$

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

4 Numerical Studies

Following Wang and Wells (2000), we use five copulas from the Archimedean family: Clayton, Frank, Gumbel, Log-copula and I-Gaussian. The expressions for $K(v)$ are given in Table 1.

Clayton	$K(v) = v(1 + \frac{1-v^\alpha}{\alpha})$	$\alpha > 0$
Frank	$K(v) = v + \left(\log \left[\frac{1-\exp(-\alpha)}{1-\exp(-\alpha v)} \right] \frac{1-\exp(-\alpha v)}{\alpha \exp(-\alpha v)} \right)$	$\alpha \in \mathcal{R}$
Gumbel	$K(v) = v(1 - \frac{\log(v)}{\alpha+1})$	$\alpha \geq 0$
Log-copula	$K(v) = v(1 - \frac{1-(1-\log(v))^{\alpha+1}}{(\alpha+1)(1-\log(v))^\alpha})$	$\alpha > 0$
I-Gaussian	$K(v) = v(1 - \frac{\log(v)(\log(v)-2\alpha)}{2(\log(v)-\alpha)})$	$\alpha > 0$

Table 1: Expressions for $K(v)$

4.1 Simulation Results

The simulation design is the same as that of Wang and Wells (2000) except that we included the I-Gaussian copula in the experiment. In all cases, (C_1, C_2) were generated from Clayton copula with $\tau = 0.3$, and the censoring rates in both components were controlled to be between 10 and 20%. The sample size was $n = 250$. The number of simulation repetitions was 200 and the number of bootstrap repetitions was 100. All the computations were done in Fortran.²

4.1.1 Results Based on $\tilde{S}_i(\hat{\alpha}_i)$

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

We 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 $\tilde{S}_i(\hat{\alpha}_i)$ ($\tilde{S}_i(\tilde{\alpha}_i)$) 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 applications. As the results for Clayton and Frank copulas are very close, we report results for Clayton copula only in Tables 2 and 3.

In addition to the percentage of times that each copula is selected, we also reported the mean and standard error of $\tilde{S}_i(\hat{\alpha}_i)$ ($\tilde{S}_i(\tilde{\alpha}_i)$). To fully understand the results in Tables 2 and 3, we provide the ranking of the remaining 4 copulas in terms of their distance to the true Clayton copula $S_i(\alpha_i^*)$ and their ranking based on $S_i(\alpha_{iL}^*)$ in Table 4, where α_{iL}^* is the probability limit of $\tilde{\alpha}_i$, i.e., $\alpha_{iL}^* = \operatorname{argmin}_{\alpha_i} |\int_0^1 K^o(v)dv - \int_0^1 K_i(v; \alpha_i)|$ in which $K^o(v)$ is the K function for Clayton copula and $\{K_i(v; \alpha_i)\}_{i=1}^4$ 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 4 copulas is the same regardless of whether $S_i(\alpha_i^*)$ is used or $S_i(\alpha_{iL}^*)$ is used. Moreover, in most of the cases, the respective values of $S_i(\alpha_i^*)$ and $S_i(\alpha_{iL}^*)$ are very close, although they are clearly different. Our theoretical results in Section 2 predict that the selection results based on $\tilde{S}_i(\hat{\alpha}_i)$ and $\tilde{S}_i(\tilde{\alpha}_i)$ should be similar. Indeed, Tables 2 and 3 provide an overwhelming amount of evidence supporting this prediction.

²Fortran code is available from the authors upon request.

	Minimum Distance $\hat{\alpha}_i$			Kendall's Tau $\tilde{\alpha}_i$		
	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$
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
I-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
I-Gaussian	5.30	17.63	53.75	5.78	18.47	53.75
Std. 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
I-Gaussian	3.23	6.99	12.38	3.42	7.18	12.38

Table 2: The true copula is Clayton, included in the selection

	Minimum Distance $\hat{\alpha}_i$			Kendall's Tau $\tilde{\alpha}_i$		
	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$
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
I-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
I-Gaussian	5.06	17.86	52.99	5.50	18.71	52.99
Std. 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
I-Gaussian	2.66	6.88	12.21	2.99	7.11	12.21

Table 3: The true copula is Clayton, excluded from the selection

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

Table 4: Rankings from the least distant to most distant to Clayton copula

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 $\hat{\alpha}_i$ or $\tilde{\alpha}_i$ 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, the mean and standard error of $\tilde{S}_i(\hat{\alpha}_i)$ are smaller than the mean and standard error of $\tilde{S}_i(\tilde{\alpha}_i)$. What is more important and reassuring is the fact that based on $\tilde{S}_i(\hat{\alpha}_i)$, 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 4. Only when $\tau = 0.3$, the probability that the Log-copula is being selected decreases slightly and the I-Gaussian copula being the second closest to the true copula gains a small probability of being chosen. The corresponding results based on $\tilde{\alpha}_i$ reveal similar patterns qualitatively, although for $\tau = 0.3$, the probability of selecting the Log-copula using $\tilde{\alpha}_i$ is smaller than using $\hat{\alpha}_i$. In all cases, the mean and standard error of $\tilde{S}_i(\hat{\alpha}_i)$ are smaller than the mean and standard error of $\tilde{S}_i(\tilde{\alpha}_i)$.

4.1.2 Results Based on the Model Selection Test

In this section, we applied the model selection test proposed in Section 3 to the same data sets generated in the previous subsection. 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 Tables 5 and 6.

	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$
Benchmark	Clayton (H_0 holds)		
5%	0	0	0
10%	0	0	0
Benchmark	Log-copula (H_1 holds)		
5%	0	0	0
10%	0.02	0.02	0
Benchmark	I-Gaussian (H_1 holds)		
5%	0.03	0.63	0.95
10%	0.10	0.81	0.99
Benchmark	Frank (H_1 holds)		
5%	0.41	0.61	0.04
10%	0.55	0.74	0.13
Benchmark	Gumbel (H_1 holds)		
5%	0.67	0.92	0.46
10%	0.82	0.99	0.58

Table 5: Rejection Rates: Clayton included

Both tables indicate that the test is under-sized regardless of whether the true copula is included in the selection and the true value of τ . In the first case, when the copula other than the Clayton is used as the benchmark, the rejection rates start to increase. Since 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 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 Tables 5 and 6 with the distance of each copula to the true copula reported in Table 4, we conclude that the rejection rate of the test mainly depends on the distance of the benchmark to the true copula instead of the value of τ . When the true copula is excluded from the selection, the same general conclusion carries over, see Table 6. 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 I-Gaussian, Frank, or Gumbel is used as the benchmark, the null hypothesis fails and the power of the test increases as the distance of

	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$
Benchmark	Log-copula (H_0 holds)		
5%	0	0	0
10%	0	0	0
Benchmark	I-Gaussian (H_1 holds)		
5%	0.005	0.61	0.94
10%	0.015	0.85	0.98
Benchmark	Frank (H_1 holds)		
5%	0.49	0.73	0.02
10%	0.61	0.79	0.06
Benchmark	Gumbel (H_1 holds)		
5%	0.84	0.96	0.29
10%	0.91	0.99	0.53

Table 6: Rejection Rates: Clayton excluded

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 $\tilde{S}(\hat{\alpha}_i)$.

4.2 Real Data Examples

In this section, we applied the proposed test to the two data sets studied in Wang and Wells (2000); the data set in Danahy, et al. (1977) and the data set in McGilchrist and Aisbett (1991). We suggest the readers to consult the original papers or Wang and Wells (2000) for backgrounds on these data sets. Based on the value of $\tilde{S}_i(\tilde{\alpha}_i)$, 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 7, we report p-values of the model selection test, values of $\tilde{S}_i(\tilde{\alpha}_i)$ and $\tilde{S}_i(\hat{\alpha}_i)$.

	Danahy, et al. (1977) ($\tau = 0.41$)			McGilchrist and Aisbett (1991) ($\tau = 0.19$)		
Benchmark	p-value	$\tilde{S}_i(\hat{\alpha}_i) \times 10^3$	$\tilde{S}_i(\tilde{\alpha}_i) \times 10^3$	p-value	$\tilde{S}_i(\hat{\alpha}_i) \times 10^3$	$\tilde{S}_i(\tilde{\alpha}_i) \times 10^3$
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

Table 7: p-values of the test

Based on the p-values of the test, Clayton copula is clearly rejected for both data sets and

none of the other three copulas is rejected. This is consistent with the values of $\tilde{S}_i(\hat{\alpha}_i)$; they are small for Log-copula, Frank, and Gumbel compared with the value for Clayton copula. Based on the minimum value criterion, the Gumbel copula is selected for both data sets.³ In addition to using the values of $\tilde{S}_i(\hat{\alpha}_i)$ for copula model selection, Table 7 indicates that p-values of the test provide an alternative model selection criterion. For both data sets, all three criteria $\tilde{S}_i(\hat{\alpha}_i)$, $\tilde{S}_i(\tilde{\alpha}_i)$, and the p-value select the same copula class.

5 Conclusions

This paper makes two contributions to copula model selection for bivariate failure-time data. First, we provide a rigorous discussion on the appropriate choice of a parametric estimator of the copula parameter in model selection. In fact, this applies to not only copula model selection, but model selection in general. Second, we address the issue of data reuse in copula model selection pointed out by Pena in his Comment on Wang and Wells (2000) by establishing a formal statistical test for copula model selection for bivariate failure-time data. This draws on the literature on the selection of forecasting models, see White (2000).

Extension of our model selection test to multivariate failure-time data is straightforward. As the test is based on comparing $K^o(v)$ with $K_i(v; \alpha_i^*)$, it is consistent for Archimedean copulas. For selection of general parametric copulas for censored data, one may develop pseudo-likelihood ratio procedures using the two-step estimator of Shih and Louis (1995). For complete data, goodness-of-fit statistics such as those in Chen, et al. (2003) and Fermanian (2003) may also be adopted.

³The values of $\tilde{S}_i(\tilde{\alpha}_i)$ 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 $\tilde{S}_i(\hat{\alpha}_i)$ for both data sets.

Appendix A: Asymptotic Properties of $\tilde{K}(v)$

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

- A1.** The distribution function $K^o(v)$ admits a continuous density $k^o(v)$.
- A2.** Let \mathcal{T} denote the support of (\tilde{X}, \tilde{Y}) . Given $F^o(x, y) = v$, there exists a version of the conditional distribution of (X, Y) and a countable family \mathcal{P} of partition \mathcal{C} on \mathcal{T} into a finite number of Borel sets satisfying $\inf_{C \in \mathcal{P}} \max_{C \in \mathcal{C}} \text{diam}(C) = 0$, such that for all $C \in \mathcal{C}$, the mapping $v \rightarrow \mu_v(C) = k^o(v) \Pr\{(X, Y) \in C | F^o(X, Y) = v\}$, 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 $\mathcal{D}(\mathcal{T})$.

The following lemma summarizes the results of Theorem 1 in Wang and Wells (2000).

Lemma A.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 $\mathcal{D}[\varsigma, 1]$,*

$$n^{1/2}\{\tilde{K}(v) - K^o(v)\} \Rightarrow X(v) = - \int \int I(F^o(x, y) > v) W(dx, dy) - \int \int W(x, y) \mu_v(dx, dy).$$

Appendix B: Technical Proofs

Proof of Proposition 2.1. 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 2.1, 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 inequality below:

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

(ii) By Theorem 6.2 in White (1994), it suffices to show that

- a) $n^{1/2} D_{\alpha} \tilde{S}_i(\alpha_i^*) \rightarrow -2 \int_{\varsigma}^1 X(v) D_{\alpha} K_i(v; \alpha_i^*) dv$ 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

$$D_\alpha \tilde{S}_i(\alpha_i^*) = -2 \int_{\zeta}^1 [\tilde{K}(v) - K^o(v)] D_\alpha K_i(v; \alpha_i^*) dv. \quad (\text{B.1})$$

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

$$D_\alpha^2 \tilde{S}_i(\alpha_i) = 2 \int_{\zeta}^1 [D_\alpha K_i(v; \alpha_i) D'_\alpha K_i(v; \alpha_i)] dv - \int_{\zeta}^1 [\tilde{K}(v) - K_i(v; \alpha_i)] D_\alpha^2 K_i(v; \alpha_i) dv. \quad (\text{B.2})$$

(iii) It is straightforward and omitted.

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

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

Lemma A.1 and Proposition 2.1 (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 $\hat{\alpha}_i$, the second equality follows from the fact that $D_\alpha S_i(\alpha_i^*) = 0$ and the last from Proposition 2.1 (i) (ii). Consequently, we have

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

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 A.1, the continuous mapping theorem, and (B.3). For the correct specification case, see Theorem 3 in Wang and Wells (2000).

□

Proof of Proposition 2.2. (i) is straightforward and omitted.

(ii) A similar proof to that of Proposition 2.1 (iv) suggests that

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

The result follows from Lemma A.1, the assumption on $\tilde{\alpha}_i$, the above equation, and the continuous mapping theorem.

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

□

Proof of Proposition 3.1. Equation (B.3) implies that

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

The conclusion follows immediately from Lemma A.1 and the continuous mapping theorem.

□

To prove Theorem 3.2, 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 Goncalves 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 B.1 *Under the conditions of Proposition 2.1,*

- 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.

Proof of Theorem 3.2. It suffices to show that

$$\sup_{z \in \mathcal{R}^{M-1}} |P^*\left(\begin{pmatrix} n^{1/2}T_{2nC}^* \\ \vdots \\ n^{1/2}T_{MnC}^* \end{pmatrix} \leq z\right) - P\left(\begin{pmatrix} n^{1/2}(T_{2n} - [S_1(\alpha_1^*) - S_2(\alpha_2^*)]) \\ \vdots \\ n^{1/2}(T_{Mn} - [S_1(\alpha_1^*) - S_M(\alpha_M^*)]) \end{pmatrix} \leq z\right)| = o_p(1). \quad (\text{B.5})$$

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

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

By Lemma B.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,

$$\begin{aligned} B_3^* &= \int_{\zeta}^1 [\tilde{K}(v) - K_i(v; \hat{\alpha}_i)] D_{\alpha} K_i(v; \hat{\alpha}_i) dv (\hat{\alpha}_i - \hat{\alpha}_i^*) \\ &\quad + (\hat{\alpha}_i - \hat{\alpha}_i^*)' \int_{\zeta}^1 [\tilde{K}(v) - K_i(v; \hat{\alpha}_i)] D_{\alpha}^2 K_i(v; \hat{\alpha}_i) dv (\hat{\alpha}_i - \hat{\alpha}_i^*) \\ &= (\hat{\alpha}_i - \hat{\alpha}_i^*)' \int_{\zeta}^1 [\tilde{K}(v) - K_i(v; \hat{\alpha}_i)] D_{\alpha}^2 K_i(v; \hat{\alpha}_i) dv (\hat{\alpha}_i - \hat{\alpha}_i^*) \\ &= O_{p^*}(n^{-1}) \text{ pr-P} \end{aligned}$$

where $\bar{\alpha}_i^*$ lies between $\hat{\alpha}_i^*$ and $\hat{\alpha}_i$, the second equality follows from the fact that $D_{\alpha} \tilde{S}_i(\hat{\alpha}_i) = 0$ and the last from Lemma B.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 A.1, B.1, and Proposition 2.1.

□

Proof of Lemma B.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 $\bar{K}(F)(\cdot): D[0, \tau_1] \times [0, \tau_2] \rightarrow l^\infty([0, 1])$ is Hadamard differentiable at F tangentially to $C[0, \tau_1] \times [0, \tau_2]$, where $\bar{K}(F)(\cdot) = E[I(F(X, Y) > \cdot)]$. In fact for $H \in C[0, \tau_1] \times [0, \tau_2]$, its Hadamard derivative at F is given by

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

Let $t_n \rightarrow 0$ and $H_n \rightarrow H \in C[0, \tau_1] \times [0, \tau_2]$. We need to show that

$$\frac{\bar{K}(F + t_n H_n) - \bar{K}(F)}{t_n} \rightarrow \bar{K}_F(H). \quad (\text{B.6})$$

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

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

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 (B.6) by replacing weak convergence and tightness in their proof with uniform 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 Goncalves 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, since

$$\sup_{\alpha_i} |\tilde{S}_i^*(\alpha_i) - \tilde{S}_i(\alpha_i)| \tag{B.8}$$

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

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

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

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

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

$$\begin{aligned} \sup_{\alpha_i} |D_{\alpha}^2 \tilde{S}_i^*(\alpha_i) - D_{\alpha}^2 \tilde{S}_i(\alpha_i)| &\leq 2 \int_{\zeta}^1 |\tilde{K}^*(v) - \tilde{K}(v)| \sup_{\alpha_i} |D_{\alpha}^2 K_i(v; \alpha_i)| dv \\ &= o_{p^*}(1) \text{pr-P} \end{aligned}$$

by part i) of this Lemma, Lemma A.1, and boundedness of $D_{\alpha}^2 K_i(v; \alpha_i)$.

□

References

- [1] Barbe, P., A. Genest, K. Ghoudi, and R. Bruno (1996), "On Kendall's Process," *Journal of Multivariate Analysis* 58, 197-229.
- [2] Chen, X., Y. Fan and 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.
- [3] 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.

- [4] Dabrowska, D. M. (1988), “Kaplan-Meier Estimates on the Plane,” *The Annals of Statistics* 16, 1475-1489.
- [5] Dabrowska, D. M. (1989), “Kaplan-Meier Estimate on the Plane: Weak Convergence, LIL, and the Bootstrap,” *Journal of Multivariate Analysis* 29, 308-325.
- [6] Danahy, D. J. , D. T. Burwell, W. S. Aranow, and R. Prakash (1977), “Sustained Hemodynamic and Antianginal Effect of High-Dose Oral Isosorbide Dinitrate,” *Circulation* 55, 381-387.
- [7] Fermanian, J.-D. (2003), “Goodness of Fit Tests for Copulas,” Mimeo, CREST.
- [8] Genest, C. , K. Ghoudi, and L.-P. Rivest (1995), “A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions,” *Biometrika* 82 (3), 543-552.
- [9] Genest, C. , J-F. Quessy, and B. Remillars (2003), “Goodness-of-fit Procedures for Copula Models Based on the Integral Probability Transformation,” Working Paper, *Les Cahiers du GERAD* G-2003-54.
- [10] Genest, C. and L.-P. Rivest (1993), “Statistical Inference Procedures for Bivariate Archimedean Copulas,” *Journal of the American Statistical Association* 88, 1034-1043.
- [11] Ghoudi, K. and B. Remillard (1998), “Empirical Processes Based on Pseudo-Observations,” *Asymptotic Methods in Probability and Statistics: a Volume in Honour of Miklos Csorgo.* (ed. B. Szyskowitz), 171-197, Elsevier, Amsterdam.
- [12] Ghoudi, K. and B. Remillard (2003), “Empirical Processes Based on Pseudo-Observations II: The Multivariate Case,” *Asymptotic Methods in Stochastics.* (eds. L. Horvath and B. Szyskowitz), in Press, The Fields Institute Communications Series, AMS, Providence, RI.
- [13] Gonçalves, S. and H. White (2004), “Maximum Likelihood and the Bootstrap for Non-linear Dynamic Models,” forthcoming in *Journal of Econometrics*.
- [14] Joe, H. (1997), *Multivariate Models and Dependence Concepts*, Chapman & Hall/CRC.
- [15] McGilchrist, C. A. and C. W. Aisbett (1991), “Regression with Frailty in Survival Analysis,” *Biometrics* 47, 461-466.

- [16] Nelsen, R. B. (1999), *An Introduction to Copulas*, Springer.
- [17] Oaks, D. (1982), "A Concordance Test for Independence in the Presence of Bivariate Censoring," *Biometrics* 38, 451-455.
- [18] Shih, J. and T. Louis (1995), "Inferences on the Association Parameter in Copula Models for Bivariate Survival Data," *Biometrics* 51, 1384-1399.
- [19] Sklar, A. (1959), "Fonctions de r'epartition 'a n dimensions et leurs marges," *Publ. Inst. Statis. Univ. Paris* 8, 229-231.
- [20] Van der Vaart, A. and J. Wellner (1996), *Weak Convergence and Empirical Processes: with Applications to Statistics*. New York: Springer-Verlag.
- [21] Vuong, Q. H. (1989), "Likelihood Ratio Test for Model Selection and Non-nested Hypotheses," *Econometrica* 57, 307-333.
- [22] Wang, W. and M. Wells (2000), "Model Selection and Semiparametric Inference for Bivariate Failure-Time Data," *Journal of the American Statistical Association* 95, 62-76.
- [23] White, H. (1994), *Estimation, Inference and Specification Analysis*, Cambridge University Press, Cambridge, UK.
- [24] White, H. (2000), "A Reality Check for Data Snooping," *Econometrica* 68, 1097-1126.