



Model Selection and Semiparametric Inference for Bivariate Failure-Time Data

Author(s): Weijing Wang and Martin T. Wells

Source: *Journal of the American Statistical Association*, Vol. 95, No. 449 (Mar., 2000), pp. 62-72

Published by: [American Statistical Association](#)

Stable URL: <http://www.jstor.org/stable/2669523>

Accessed: 28/04/2014 04:17

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



American Statistical Association is collaborating with JSTOR to digitize, preserve and extend access to *Journal of the American Statistical Association*.

<http://www.jstor.org>

Model Selection and Semiparametric Inference for Bivariate Failure-Time Data

Weijing WANG and Martin T. WELLS

We propose model selection procedures for bivariate survival models for censored data generated by the Archimedean copula family. In route to constructing the selection methodology, we develop estimates of some time-dependent association measures, including estimates of the local and global Kendall's tau, local odds ratio, and other measures defined throughout the literature. We propose a goodness-of-fit-based model selection methodology as well as a graphical approach. We show that the proposed methods have desirable asymptotic properties and perform well in finite samples.

KEY WORDS: Archimedean copula; Bivariate survival function; Frailty distribution; Kendall's tau; Model selection; Odds ratio estimation; Time-dependent association.

1. INTRODUCTION

In recent years substantial research effort has been devoted to developing methodology for multivariate failure-time data. Applications of multivariate survival analysis arise in various fields. Examples in biomedical applications include lifetime analysis in match-paired case control studies, studies of time to occurrence of a disease to paired organs, and the examination of duration times of critical stages of a multistage disease process. Specifically, in Section 4 we consider an assessment of the effect of a medical intervention on angina pectoris. Danahy, Burwell, Aranow, and Prakash (1977) collected data on 21 cardiac disease pectoris and recorded exercise time until angina pectoris and the exercise time until angina pectoris 3 hours after taking oral isosorbide dinitrate. One needs to account for the censoring induced by patient fatigue. It is clearly important to assess the bivariate relationship between the control and treatment times while accounting for within-subject dependence, and a marginal analysis would not yield the important treatment effect information. As for other applications, in demographic studies of the dynamics of mortality, multivariate models incorporate an exchangeable dependence structure by the inclusion of a cluster-specific random effect (see Vaupel, Manton, and Stallard 1979). In engineering applications, modeling the multivariate nature of mechanical or electronic components in a parallel or a serial system has become increasingly important (see Marshall and Olkin 1988). The canonical problem of interest is to study the dependence relationship among several lifetime random variables. In many applications, it is often believed that the level of association varies across time, and it is of particular interest to investigate the time-dependent association. In this article we focus on the bivariate case, although many of the ideas could be extended to multivariate problems.

Let (X, Y) be the lifetime variables of interest with joint survival function $F(x, y) = \Pr(X > x, Y > y)$ and marginal survival functions $F_i(\cdot)$ ($i = 1, 2$). If the components of (X, Y) are locally independent at a point (x, y) , then $F(x, y) = F_1(x)F_2(y)$. Hence the most simplistic method of assessing local dependence is by checking whether $F(x, y)/\{F_1(x)F_2(y)\} = 1$. There exist many other time-dependent association measures constructed for different purposes of analysis; these include the local odds ratio function, the local Kendall's tau function (Oakes 1989), and the covariance function of the marginal martingale components (Prentice and Cai 1992). Anderson, Louis, Holm, and Harvald (1992) have given a nice general discussion on these measures.

A more modern approach to investigating local dependence is through model fitting. Although more assumptions may be required, modeling provides a systematic way to summarize joint relationships. The past decade has seen a substantial research effort toward deriving a unified approach to studying models that are generated by a system of random effects. For example, Hougaard (1986) and Oakes (1989) discussed a family of correlated bivariate distributions induced by a latent frailty variable. Lindeboom and Van Den Berg (1994) and Marshall and Olkin (1988) studied a class of distributions generated by bivariate mixtures. Applications of frailty models were discussed by Bandeene-Roche and Liang (1996), Clayton and Cuzick (1985), Murphy (1994), Nielsen, Gill, Andersen, and Sorenson (1992), and Vaupel et al. (1979). Genest and MacKay (1986), Genest and Rivest (1993), and Joe (1993) studied the mathematical properties of the copula and Archimedean copula (AC) classes. The copula class separates the dependence structure from the marginal effects. Following development of the general modeling techniques, there has been growing research interest in developing methodology for selecting a particular model from a given class. For instance, Oakes (1989) used the local odds ratio function to identify the underlying frailty distribution. Genest and Rivest (1993) derived a measure based on a decomposition of Kendall's tau statistic to identify a particular AC model. However, in the

Weijing Wang is Associate Professor, Institute of Statistics, National Chiao-Tung University, Hsin-chu, Taiwan (E-mail: wjwang@stat.nctu.edu.tw). Martin T. Wells is Associate Professor, Department of Social Statistics, Cornell University, Ithaca, NY 14851. The support of The National Science Council (grant 87-2118) and National Science Foundation (grants DMS 9625440 and 9971586) is gratefully acknowledged, as are the kind and helpful comments of the editor, associate editor, and referees. The authors thank Phillip Hougaard for providing the references for the datasets.

© 2000 American Statistical Association
Journal of the American Statistical Association
March 2000, Vol. 95, No. 449, Theory and Methods

important case where censoring is present, there have been no results to date.

Censoring is common in the analysis of lifetime data. In bivariate survival analysis (X, Y) may both be subject to censoring, this complicates the construction of statistical inference procedures. Specifically, 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, $(\delta_1, \delta_2) = \{\mathbb{1}(X \leq C_1), \mathbb{1}(Y \leq C_2)\}$, where $a \wedge b = \min(a, b)$ and $\mathbb{1}(\cdot)$ is the indicator function. Recently, tremendous effort has been spent on the derivation of nonparametric estimators of $F(x, y)$ for bivariate censored data. Estimators of $F(x, y)$ have been proposed by Campbell and Földes (1982), Dabrowska (1988), Lin and Ying (1993), Prentice and Cai (1992), Tsai et al. (1986), van der Laan (1996), and Wang and Wells (1997), to name just a few. In this article we show how previous results can be used to derive inferential methods for parameter estimation and model selection. The proposed methods, which utilize the von Mises functional technique of Gill (1989), provide a unified inferential approach for estimating quantities that can be expressed as statistical functionals of F . Because the censoring issue is handled in the stage of estimating F , the proposed approach is sufficiently flexible to deal with various censoring mechanisms. Extra information about the marginal distributions or covariates can easily be incorporated into the analysis through the estimation of F .

In Section 2 we develop the theory for a new goodness-of-fit procedure and propose a graphical method to select a particular AC model for bivariate censored data. In Section 3 we derive an estimator of the local odds ratio function for models in the AC class. We present two real data analyses and simulation results on model selection in Section 4, and give some concluding remarks and point out some direction for further study in Section 5. We provide the proofs of the results in an Appendix.

2. MODEL SELECTION METHODS

2.1 Definitions

Many well-known bivariate lifetime distributions with continuous marginals, such as those proposed by Clayton (1978), Frank (1979), Gumbel (1960, 1961), and Hougaard (1986), are of the form

$$F(x, y) = C_\alpha\{F_1(x), F_2(y)\}, \quad (1)$$

where $F(\cdot, \cdot)$ denotes the joint survival function of (X, Y) , $F_i(\cdot)$ ($i = 1, 2$) are the marginal survival functions of X and Y , and $\alpha \in \mathbb{R}^k$ denotes an unknown association parameter. Note that the copula function, $C_\alpha(\cdot, \cdot)$, is itself a survival function on $[0, 1]^2$. A special feature of the copula class is that the dependence structure is separated from the marginal effects, so the dependence relationship can be studied without specifying the marginal distributions. The parameter α can be viewed as a global association parameter related to Kendall's tau,

specifically

$$\begin{aligned} \tau &= 4 \int \int F(x, y) F(dx, dy) - 1 \\ &= 4 \int_0^1 \int_0^1 C_\alpha(s, t) C_\alpha(ds, dt) - 1. \end{aligned}$$

Given the same level of overall association measured by α or τ , $C_\alpha(s, t)$ determines the degree of local dependence at $(s, t) \in [0, 1]^2$, where (s, t) indicates the joint survival status. Note that all of the models reduce to the same form when the overall association approaches to the extreme levels, under independence ($\tau = 0$), $C(s, t) = st$ and under positive maximal dependence ($\tau = 1$), $C(s, t) = s \wedge t$, the upper Fréchet bound (see Marshall and Olkin 1988).

Recent research has focused on a subclass of (1), the AC class, which indexes $C_\alpha(\cdot, \cdot)$ by a univariate function and thus has more tractable analytical properties. The survival functions in the AC class are of the form

$$F(x, y) = \phi_\alpha^{-1}[\phi_\alpha\{F_1(x)\} + \phi_\alpha\{F_2(y)\}], \quad (2)$$

where $\phi_\alpha(\cdot)$ is a convex function defined on $[0, 1]$ satisfying $\phi_\alpha(1) = 0$. This class also contains many useful models, including the bivariate frailty family when $\phi_\alpha^{-1}(\cdot)$ is the Laplace transform of the underlying frailty distribution (Oakes 1989). Genest and Rivest (1993) showed that the function $\phi_\alpha(\cdot)$ in (2) can be recovered by the estimable univariate function $K(v) \equiv \Pr\{F(X, Y) \leq v\}$. Specifically, $K(v)$ is related to $\phi_\alpha(\cdot)$ through the differential equation

$$\lambda(v) = v - K(v) = \frac{\phi_\alpha(v)}{\phi'_\alpha(v)}, \quad (3)$$

where $\phi'_\alpha(v) = \partial\phi_\alpha(v)/\partial v$. The foregoing expression yields the inversion formula

$$\phi_\alpha(v) = \exp \left[\int_{v_0}^v \frac{1}{t - K(t)} dt \right] = \exp \left[\int_{v_0}^v \frac{1}{\lambda(t)} dt \right], \quad (4)$$

where $0 < v_0 < 1$ is an arbitrary constant. Thus $K(v)$, or, equivalently $\lambda(v)$, plays a key role in the identification of $\phi_\alpha(\cdot)$, which in turn determines the underlying dependence structure for the AC class. The function $K(v)$ has a general geometric interpretation related to contour analysis. Define the contour curve of $F(\cdot, \cdot)$ at level v for $v \in [0, 1]$ by $\psi(v) = \{(x, y) : F(x, y) = v, (x, y) \in \mathbb{R}_+^2\}$. Because $F(\cdot, \cdot)$ is monotone, $K(v)$ measures the mass between the contour curves $\psi(0)$ and $\psi(v)$. Equation (4) implies that members in the AC class are classified according to the distribution of the mass within the contour curves.

2.2 Estimation of $K(v)$ Under Bivariate Censoring

When the data are complete, Genest and Rivest (1993) proposed estimating $K(v)$ by constructing "pseudo observations" of $V_i = F(X_i, Y_i)$ by $\hat{V}_i = \sum_{j=1}^n \mathbb{1}(x_j > X_i, y_j > Y_i) / (n - 1)$ ($i = 1, \dots, n$) and then estimating $K(v)$ by the empirical distribution function of the \hat{V}_i ($i = 1, \dots, n$). Specifically, Genest and Rivest's estimate of $K(v)$ is given by $\bar{K}(v) = \sum_{i=1}^n \mathbb{1}(\hat{V}_i \leq v) / n$. This approach is not viable if some of $\{(X_i, Y_i), i = 1, \dots, n\}$ cannot be directly observed due to censoring. Here we propose an estimator

of $K(v)$ for bivariate censored data, $\{(\tilde{X}_i, \tilde{Y}_i, \delta_{1i}, \delta_{2i}), i = 1, \dots, n\}$. Consider the following expressions of $K(v)$:

$$\begin{aligned} K(v) &= E[\mathbb{1}\{F(X, Y) \leq v\}] \\ &= \int_0^\infty \int_0^\infty \mathbb{1}[F(x, y) \leq v] F(dx, dy), \end{aligned} \quad (5)$$

which can be estimated nonparametrically by plugging in an estimator of F in the foregoing integral form. Specifically, let $\tilde{x}_{(1)} \leq \dots \leq \tilde{x}_{(n)}$ and $\tilde{y}_{(1)} \leq \dots \leq \tilde{y}_{(n)}$ be ordered observations of $\{(\tilde{X}_i, \tilde{Y}_i), i = 1, \dots, n\}$. The first proposed estimator of $K(v)$ is given by

$$\begin{aligned} \hat{K}(v) &= \int \int \mathbb{1}[\hat{F}(x, y) \leq v] \hat{F}(dx, dy) \\ &= \sum_{i=1}^n \sum_{j=1}^n \mathbb{1}\{\hat{F}(\tilde{x}_{(i)}, \tilde{y}_{(j)}) \leq v\} \hat{F}(\Delta\tilde{x}_{(i)}, \Delta\tilde{y}_{(j)}), \end{aligned} \quad (6)$$

where \hat{F} is a nonparametric estimator of F 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 $(\tilde{x}_{(i)}, \tilde{y}_{(j)})$. When there is no censoring, $\hat{F}(x_i, y_i) = \sum_{j=1}^n \mathbb{1}(x_j > x_i, y_j > y_i)/n$, $(i = 1, \dots, n)$, is the empirical survival estimate at (x_i, y_i) , $\hat{F}(\Delta x_i, \Delta y_i) = 1/n$ ($i = 1, \dots, n$) and $\hat{F}(\Delta x_i, \Delta y_j) = 0$ if $i \neq j$. It is easy to see that $\hat{K}(v) - \bar{K}(v) = 1/n \sum_{i=1}^n \mathbb{1}[v < \hat{V}_i \leq (n/n - 1)v] \rightarrow^p \Pr(F(X, Y) = v)$, which equals 0 under (2) and the assumed continuity (Genest and MacKay 1986). Hence \bar{K} and \hat{K} are asymptotically equivalent in the case of no censoring.

The nonparametric estimators of F mentioned previously were all derived under the assumption that (X, Y) are independent of (C_1, C_2) . Sometimes the independent censoring assumption is not plausible. Lin, Sun, and Ying (1999), Visser (1996), and Wang and Wells (1998) discussed a dependent censoring situation when (X, Y) are the duration times for successive events and proposed nonparametric estimators to adjust for dependent censoring. Note that because the issue of censoring is handled in the stage of estimating F , the proposed idea for estimating $K(\cdot)$ is flexible under various censoring structures as long as an appropriate estimator of F is used.

Properties of $\hat{K}(v)$ depend on properties of the underlying estimator \hat{F} . Denote the support of (\tilde{X}, \tilde{Y}) by $\mathcal{T} = \{(x, y) : \Pr(\tilde{X} > x, \tilde{Y} > y) > 0\}$, and let \mathcal{V} be the image of \mathcal{T} under F . Under right censoring, \mathcal{T} is contained in the support of (X, Y) . Asymptotic results for most nonparametric estimators of F are valid only for points in \mathcal{T} , the restricted support. Because $F(x, y)$ cannot capture the mass outside \mathcal{T} , $\hat{K}(v)$ must be modified. The modified estimator of $K(v)$ is based on the equivalent expression, $K(v) = 1 - \Pr(F(X, Y) > v)$, and is given by

$$\begin{aligned} \tilde{K}(v) &= 1 - \sum_{i=1}^n \sum_{j=1}^n \mathbb{1}[\hat{F}(\tilde{x}_{(i)}, \tilde{y}_{(j)}) > v] \\ &\quad \times \hat{F}(\Delta\tilde{x}_{(i)}, \Delta\tilde{y}_{(j)}). \end{aligned} \quad (7)$$

The following theorem shows that if the underlying estimator of F is consistent and converges weakly to a Gauss-

ian process on $\mathcal{D}[\mathcal{T}]$, then $\tilde{K}(v)$ will inherit some nice properties on $\mathcal{D}[\mathcal{V}]$. The weak convergence (\Rightarrow) result for $\sqrt{n}\{\tilde{K}(v) - K(v)\}$ has been established by Barbe, Genest, Choudi, and Rémillard (1996) for complete data. To deduce the weak convergence result for $\sqrt{n}\{\tilde{K}(v) - K(v)\}$ we need the following hypotheses:

H1. The distribution function $K(v)$ of $V = F(X, Y)$ admits a continuous density $k(v)$.

H2. Given $F(x, y) = v$, 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 $v \rightarrow \mu_v(C) = k(v) \Pr\{(X, Y) \in C | F(X, Y) = v\}$, is continuous.

Theorem 1. If $\hat{F}(x, y)$ is a uniformly and strongly consistent estimator of $F(x, y)$ for $(x, y) \in \mathcal{T}$, then for $0 < \xi \equiv F(\tau_1, \tau_2) \leq v \leq 1$ with $(\tau_1, \tau_2) \in \mathcal{T}$, $\tilde{K}(v) \xrightarrow{p} K(v)$ on $[\xi, 1]$. Furthermore, if $n^{1/2}\{\hat{F}(x, y) - F(x, y)\} \Rightarrow W(x, y)$, where $W(x, y)$ is a continuous mean-zero Gaussian process on $\mathcal{D}[\mathcal{T}]$, then under H1 and H2, it follows that on $\mathcal{D}[\xi, 1]$,

$$\begin{aligned} n^{1/2}\{\tilde{K}(v) - K(v)\} &\Rightarrow X(v) \equiv - \int \int \mathbb{1}(F(x, y) > v) \\ &\quad \times W(dx, dy) - \int \int W(x, y) \mu_v(dx, dy). \end{aligned}$$

The weak convergence of various estimators of F was demonstrated by Gill, van der Laan, and Wellner (1993) using functional delta-method theorems to establish functional central limit theorems. (For an extensive weak convergence theory, see Gill 1989 and van der Vaart and Wellner 1993.)

The asymptotic variance of $\tilde{K}(v)$ depends on the asymptotic variance of \hat{F} . However, in general it is too complex to give a closed-form expression of the asymptotic variance for most estimators of F . Therefore, the bootstrap method becomes a practical alternative for obtaining the variance estimate (Dabrowska 1989). Specifically, let $\{(\tilde{X}_j^*, \tilde{Y}_j^*, \delta_{1j}^*, \delta_{2j}^*), j = 1, \dots, m\}$ be a random sample with replacement from the original data $\{(\tilde{X}_i, \tilde{Y}_i, \delta_{1i}, \delta_{2i}), i = 1, \dots, n\}$ and let $\hat{F}^*(x, y)$ and $\tilde{K}^*(v)$ be the bootstrapped counterparts of $\hat{F}(x, y)$ and $\tilde{K}(v)$. Using a functional delta-method theorem of Gill (1989), it can be shown that as $m \wedge n \rightarrow \infty$, the bootstrap process $m^{1/2}\{\hat{F}^*(x, y) - \hat{F}(x, y)\}$ converges to the same limiting process as $n^{1/2}\{\hat{F}(x, y) - F(x, y)\}$. A similar argument may be applied to show that the bootstrap version, $m^{1/2}\{\tilde{K}^*(v) - \tilde{K}(v)\}$, also converges to the same limit of $n^{1/2}\{\tilde{K}(v) - K(v)\}$.

Once an estimator of $K(v)$ is obtained, $\phi(\cdot)$ can be estimated nonparametrically by using the inversion formula in (4); that is,

$$\hat{\phi}(v) = \exp \left[\int_{v_0}^v \frac{1}{t - \tilde{K}(t)} dt \right]. \quad (8)$$

Because $\tilde{K}(v)$ is a step function, to evaluate (8), one must smooth $\tilde{K}(v)$ and then perform numerical integration. However, in general $\hat{\phi}(v)$ does not have a tractable form. For

inferential purposes, it is more appealing to select a parametric family of $\phi_\alpha(\cdot)$ that best describes the data. The following section introduces a goodness-of-fit statistic for testing whether the data are drawn from a hypothesized model.

2.3 Goodness-of-Fit Statistics

A number of metrics could be used as goodness-of-fit statistics to measure the discrepancy between a hypothesized model and the empirical model. A natural choice is the L^2 -norm distance,

$$S(\alpha) = \int_{\xi}^1 \{\tilde{K}(v) - K_\alpha(v)\}^2 dv. \quad (9)$$

Note that to evaluate $K_\alpha(\cdot)$ for the hypothesized model, one usually needs to estimate α . A preliminary estimator of α may be obtained via an estimator of Kendall's tau based on the relationship

$$\tau = 4E[F(X, Y)] - 1 = 4 \int_0^1 \lambda_\alpha(v) dv + 1 \equiv \Lambda(\alpha). \quad (10)$$

If $\Lambda(\alpha) = \tau$ is a one-to-one function, then α can be estimated by $\hat{\alpha} = \Lambda^{-1}(\hat{\tau})$, where $\hat{\tau}$ is an estimate of τ . If τ is not a one-to-one function of α , then some artificial constraints may be imposed. For example, the log-copula model is indexed by two parameters, α and γ (see Table 1), and for convenience, one may assume that $\alpha\gamma = 1$. Nonparametric estimation of Kendall's tau under censoring is a complex problem. It turns out the estimators proposed by Brown, Hollander, and Korwar (1974) and Oakes (1982) are not consistent if $\tau \neq 0$. Wang and Wells (1999) showed that if the largest observations, $\tilde{x}_{(n)}$ and $\tilde{y}_{(n)}$, are both uncensored, then

$$\hat{\tau}_0 = 4 \sum_{i=1}^n \sum_{j=1}^n \hat{F}(\tilde{x}_{(i)}, \tilde{y}_{(j)}) \hat{F}(\Delta \tilde{x}_{(i)}, \Delta \tilde{y}_{(j)}) - 1 \quad (11)$$

is a consistent estimate of τ . However, in general $\hat{\tau}_0 \rightarrow 4 \int \int_{\mathcal{T}} F(x, y) F(dx, dy) - 1 \leq \tau$. Alternatively, α can be estimated by using a minimum distance-type estimate $\tilde{\alpha} = \arg\min_{\alpha} \int \{\tilde{K}(v) - K_\alpha(v)\}^2 dv$. Similar types of minimum Cramer-von Mises estimates were discussed by Shorack and Wellner (1984, p. 254). The following theorem summarizes the asymptotic properties of $\tilde{\alpha}$.

Theorem 2. If $K_\alpha(v)$ is twice differentiable with respect to α with bounded derivatives and $\sqrt{n}\{\tilde{K}(v) - K_\alpha(v)\} \Rightarrow X(v)$, where $X(v)$ is defined in Theorem 1, then $n^{1/2}(\tilde{\alpha} - \alpha)$ converges in distribution to $[\int \{[\partial K_\alpha(v)]/\partial \alpha\}^2 dv]^{-1} \int (\partial K_\alpha(v)/\partial \alpha) X(v) dv$.

The proof of Theorem 2 is given in the Appendix. Note that due to the complexity of $X(v)$, it is also difficult to derive a closed-form expression of the asymptotic variance of $\tilde{\alpha}$. The bootstrap method discussed earlier can be used to obtain the variance estimate. Other semiparametric estimators of α for copula models have been proposed by Genest, Ghoudi, and Rivest (1995), Hsu and Prentice (1996) and

Shih and Louis (1995). These estimators have the desirable properties, and the variance of their estimators can be estimated analytically.

The following two results supply the needed asymptotic theory for construction of the formal test procedure. The first is a simple consequence of the continuous mapping theorem (Shorack and Wellner 1986, chap. 5) for integrals of squared Gaussian processes.

Proposition 1. If $\sqrt{n}\{\tilde{K}(v) - K_\alpha(v)\} \Rightarrow X(v)$ on $\mathcal{D}[\xi, 1]$, then $nS(\alpha) \Rightarrow \int_{\xi}^1 X^2(v) dv$.

With α estimated by $\hat{\alpha}$, the ideas from the area of goodness-of-fit testing for a composite hypothesis may be applied (see Shorack and Wellner 1996, chap. 5).

Theorem 3. If $K_\alpha(\cdot)$ is twice differentiable with respect to α with bounded derivatives and $\sqrt{n}(\hat{\alpha} - \alpha) \Rightarrow Q$, where $\hat{\alpha}$ is an estimate of α , then $\sqrt{n}\{\tilde{K}(v) - K_{\hat{\alpha}}(v)\} \Rightarrow \hat{X}(v) \equiv X(v) - \{[\partial K_\alpha(v)]/\partial \alpha\}Q$. Furthermore, $nS(\hat{\alpha}) \Rightarrow \int_{\xi}^1 \hat{X}^2(v) dv$.

The asymptotic variance of $S(\hat{\alpha})$, denoted by σ_S^2 , is also difficult to estimate analytically. However, the naive bootstrap approach discussed earlier by resampling bootstrap replicates from the original sample, $\{(\tilde{X}_i, \tilde{Y}_i, \delta_{1i}, \delta_{2i}), i = 1, \dots, n\}$, is not valid for obtaining a reasonable estimate of σ_S^2 . We give a heuristic explanation in the next section. Romano (1988) discussed the validity of using the bootstrap to approximate the asymptotic distribution of some general distance statistics. According to the general theory, in our case one generates a bootstrap sample, $\{V_1^*, \dots, V_n^*\}$, from the distribution $V \sim K_{\hat{\alpha}}(v)$. Using the fact that $\tau = 4E(V) - 1$, where $V = F(X, Y)$, with the bootstrap sample one can compute $\tilde{K}^*(v) = \sum_{i=1}^n \mathbb{1}(V_i^* \leq v)/n$, $\hat{\tau}^* = 4 \sum_{i=1}^n V_i^*/n - 1$, $\hat{\alpha}^* = \Lambda^{-1}(\hat{\tau}^*)$, $K_{\hat{\alpha}^*}(v)$, and then $S^*(\hat{\alpha}^*) = \int_{\xi}^1 \{\tilde{K}^*(v) - K_{\hat{\alpha}^*}(v)\}^2 dv$. By repeating the bootstrap procedure many times, one can estimate σ_S^2 by the sample variance of bootstrap estimates of $S^*(\hat{\alpha}^*)$. Note that it does not matter whether originally $\hat{\alpha}$ is obtained by inverting an estimate τ or not, because the difference of $nS(\hat{\alpha})$ by using different consistent estimates of α is of smaller order and hence negligible.

2.4 A Graphical Model Selection Method

Practitioners often need to select the best-fitting model among some competing model alternatives for the data at hand. Genest and Rivest (1993) proposed a graphical model selection procedure by plotting empirical estimates of $\lambda(v) = v - K(v)$ with theoretical curves of $\lambda_\alpha(v)$ for models under consideration. The best-fitting model is the one whose theoretical curve is closest to the empirical estimates. The function $\lambda(v)$ can be estimated by $\tilde{\lambda}(v) = v - \tilde{K}(v)$. Plotting $\tilde{\lambda}(v)$ instead of $\tilde{K}(v)$ gives a better visual comparison between $\tilde{\lambda}(v)$ and $\lambda(v)$ than the difference between $\tilde{K}(v)$ and $K(v)$. Note that because $|\tilde{K}(v) - K_{\hat{\alpha}}(v)| = |\tilde{\lambda}(v) - \lambda_{\hat{\alpha}}(v)|$, $K(v)$ and $\lambda(v)$ yield the same information measures.

The relative magnitude of the test statistic, $S(\hat{\alpha})$, under different model assumptions can be used to rank the model

proposals. We need not include a model complexity penalty if the models under consideration all have roughly the same number of parameters. Note that because α is estimated, σ_S^2 depends on the form of $K_\alpha(v)$ and thus is different for different model hypotheses. It is not clear whether the standardized statistic $T(\hat{\alpha}) \equiv S(\hat{\alpha})/\sigma_S$, which accounts for the variation of $S(\hat{\alpha})$, is a better measure for model selection than $S(\hat{\alpha})$. We find that the value of σ_S is much smaller under the true model. To simplify the discussion, suppose that parameters of the model alternatives all have one-to-one correspondence with Kendall's tau. Denote $K_\tau(\cdot)$ as the distribution function of $F(X, Y)$ under the true model with the true τ and $K_\tau(\cdot)$ as the function for any hypothesized model with the true level of τ . It follows that

$$\begin{aligned} nS(\hat{\tau}) = & \int n\{\tilde{K}(v) - K_\tau^\circ(v)\}^2 dv \\ & + \int n\{K_\tau(v) - K_{\hat{\tau}}(v)\}^2 dv \\ & + \int n\{K_\tau^\circ(v) - K_\tau(v)\}^2 dv \\ & + 2 \int n\{\tilde{K}(v) - K_\tau^\circ(v)\}\{K_\tau(v) - K_{\hat{\tau}}(v)\} dv \\ & + 2 \int n\{K_\tau(v) - K_{\hat{\tau}}(v)\}\{K_\tau^\circ(v) - K_\tau(v)\} dv \\ & + 2 \int n\{\tilde{K}(v) - K_\tau^\circ(v)\}\{K_\tau^\circ(v) - K_\tau(v)\} dv. \end{aligned}$$

When the hypothesized model is the true one, $K_\alpha^\circ(\cdot) = K_\alpha(\cdot)$ and $nS(\hat{\alpha}) \Rightarrow \int \hat{X}^2(v) dv$, as stated in Theorem 3. However, if the hypothesized model is misspecified (i.e., $K_\alpha^\circ(v) - K_\alpha(v) = c(v) \neq 0$, where $c(v)$ is a constant varying with v), then $nS(\hat{\tau})$ will diverge as $n \rightarrow \infty$ and produce large variation in finite samples. Because $T(\hat{\alpha})$ would impose an unnecessary penalty on the correct model, we suggest ranking the models based on $S(\hat{\alpha})$ and not on its standardized version.

The foregoing decomposition of $nS(\hat{\tau})$ can also be used to illustrate why the naive bootstrap procedure is not valid for estimating the asymptotic variance of $nS(\hat{\alpha})$. Although the naive bootstrap procedure provides a good approximation of $n^{1/2}\{\tilde{K}(v) - K_\tau^\circ(v)\}$ and $n^{1/2}\{K_\tau(v) - K_{\hat{\tau}}(v)\}$, it cannot mimic the behavior of $n^{1/2}\{K_\tau^\circ(v) - K_\tau(v)\}$. In simulations, not presented here, we found that the variance estimate using the naive bootstrap method is much larger than its theoretical value, especially when the hypothesized model is the true model.

Table 1. Examples of $\theta_\alpha(v)$

Family	$\phi_\alpha(v)$	Range of α and γ	$\theta_\alpha(v)$
Clayton	$(v^{-\alpha} - 1)/\alpha$	$(0, \infty)$	$\alpha + 1$
Frank	$\log\left(\frac{1 - \exp(-\alpha)}{1 - \exp(-\alpha v)}\right)$	$(-\infty, \infty)$	$\frac{v\alpha}{1 - \exp(-\alpha v)}$
Gumbel	$\{-\log(v)\}^{\alpha+1}$	$[0, \infty)$	$1 - \frac{\alpha}{\log(v)}$
Log-copula	$\{1 - \log(v)/\alpha\}^{\alpha+1} - 1$	$(0, \infty)$	$1 + \frac{\alpha\gamma - \log v}{\alpha}$
I-Gaussian	$\{\log v\}\{\log v - 2\alpha\}/(2\alpha^2)$	$(0, \infty)$	$1 + \frac{1}{\alpha - \log v}$

To simplify numerical analysis, $S(\alpha)$ can be replaced by its Riemann sum approximate; that is, $\hat{S}(\alpha) = \sum_{i=1}^n \{\tilde{K}(v_{(i)}) - K_\alpha(v_{(i)})\}^2 (v_{(i)} - v_{(i-1)})$, where $v_0 = 0$ and $v_{(1)} \leq \dots \leq v_{(n)}$ are the ordered values of $\{v_j = \hat{F}(\tilde{x}_j, \tilde{y}_j), j = 1, \dots, n\}$. The proposed graphical model selection procedure that can handle bivariate censored data is described by the following steps: Select an appropriate plug-in estimator of F ; compute $\tilde{K}(v)$ and then $\tilde{\lambda}(v)$; plot $\tilde{\lambda}(v) = v - \tilde{K}(v)$ for $v = \hat{F}(\tilde{x}_i, \tilde{y}_i)$ ($i = 1, \dots, n$); estimate α_j , the parameter for the j th model, by $\hat{\alpha}_j$; compare the empirical plot of $\tilde{\lambda}(v)$ with theoretical plots of $\lambda_{\hat{\alpha}_j}(v)$ for models under consideration; and choose a model that provides the closest fit to the estimated curve—that is, select the model with smallest $\hat{S}(\hat{\alpha}_j)$. Note that iterating the last two steps provides a visual way to obtain $\tilde{\alpha}$, the estimator that minimizes the L^2 distance between $\tilde{K}(v)$ and $K_\alpha(v)$.

3. ESTIMATING THE ODDS RATIO FUNCTION FOR THE ARCHIMEDEAN COPULA CLASS

The local odds ratio function first proposed by Oakes (1989) provides an intuitive way to describe local association, which is defined as

$$\theta^*(x, y) \equiv \frac{F(x, y)D_1D_2F(x, y)}{\{D_1F(x, y)\}\{D_2F(x, y)\}}, \quad (12)$$

where D_i ($i = 1, 2$) denote $-\partial/\partial x$ and $-\partial/\partial y$. In general, $\theta^*(x, y)$ is a bivariate function measuring local dependence and equals 1 if (X, Y) are independent at (x, y) . For (X, Y) in the AC class, Oakes showed that $\theta^*(x, y)$ depends on (x, y) through some univariate function $\theta\{F(x, y)\}$ such that $\theta_\alpha(v) = -v\{[\phi''_\alpha(v)]/[\phi'_\alpha(v)]\}$, where $\phi_\alpha(v) = \int_v^1 \exp\{\int_z^{1-c} [\theta_\alpha(u)/u] du\} dz$, where $c > 0$ is a constant. The form of $\theta_\alpha(v)$ for several models in the AC class are listed in Table 1. (For more detailed properties of the odds ratio function, see Anderson et al. 1991 and Oakes 1989.) Oakes also derived another local dependence measure, called local Kendall's tau and denoted by $\tau(v)$, which is given by $\tau_\alpha(v) = E[\text{sign}\{(X_1 - X_2)(Y_1 - Y_2)\} | X_1 \wedge X_2 = x, Y_1 \wedge Y_2 = y] = \{[\theta_\alpha(v) - 1]/[\theta_\alpha(v) + 1]\}$, where $v = F(x, y)$. Note that $-1 \leq \tau_\alpha(v) \leq 1$. Figure 1 plots $\theta_\alpha(v)$ and $\tau_\alpha(v)$ for several AC models. The patterns of $\theta_\alpha(v)$ and $\tau_\alpha(v)$ are the same at two levels of τ : $\tau = .3$ and $\tau = .7$. Larger values of $\theta(v)$ and $\tau(v)$ indicate higher dependence. Because v represents the joint survival probability, as time passes, v changes from 1 to 0. It can be seen that for Gumbel's type II model of extreme values (which has a positive stable frailty), $\theta(v)$ decreases exponentially as v decreases from 1 to 0, whereas for Frank's model and the log-copula model, $\theta(v)$ decreases fairly linearly. For Clayton's model, $\theta(v)$ and $\tau(v)$ both stay at the same level. In fact, $\theta^*(x, y)$ is a constant for all (x, y) for the Clayton model. When $\tau = .3$, the inverse Gaussian frailty model behaves like Frank's model, whereas it will approach Gumbel's model as $\tau \rightarrow .5$ (not shown in Fig. 1).

Oakes (1989) proposed a nonparametric estimator of $\theta^*(x, y)$ that applies only to discrete or grouped data. Based on (12), estimating $\theta^*(x, y)$ for continuous distributions in-

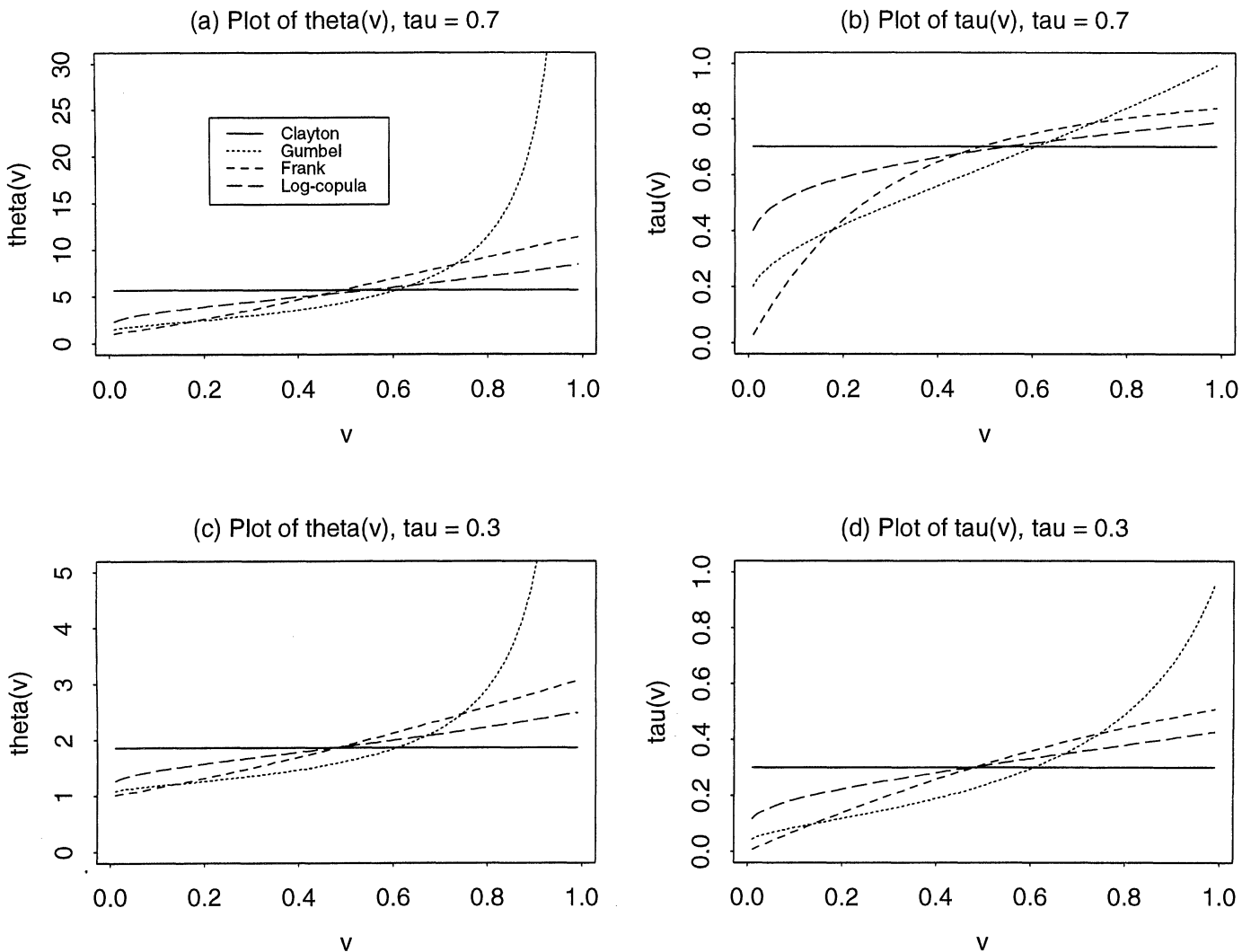


Figure 1. Plots of $\theta_\alpha(v)$ and $\tau_\alpha(v)$. — Clayton, ... Gumbel, --- Frank, -.- Log-copula. (a) $\theta(v)$, $\tau = .7$; (b) $\tau(v)$, $\tau = .7$; (c) $\theta(v)$, $\tau = .3$; (d) $\tau(v)$, $\tau = .3$.

volves estimating the derivatives of $F(x, y)$. Let $f(x, y) = D_1 D_2 F(x, y)$ be the joint density function of (X, Y) . A simple kernel density estimator of $f(x, y)$ for bivariate censored data is given by $\hat{f}_b(x, y) = 1/b^2 \sum_{i=1}^n \sum_{j=1}^n \Phi(\{[x - \tilde{x}_{(i)}]/b\}, \{[y - \tilde{y}_{(j)}]/b\}) \hat{F}(\Delta \tilde{x}_{(i)}, \Delta \tilde{y}_{(j)})$, where $\Phi(\cdot, \cdot)$ is a bivariate kernel function and b is a bandwidth parameter. Wells and Yeo (1996) discussed kernel density estimation for bivariate censored data. Estimates of $D_1 F(x, y)$ and $D_2 F(x, y)$ can be obtained by integrating $\hat{f}(x, y)$. Therefore, a nonparametric estimator of $\theta^*(x, y)$ for continuous (X, Y) is given by

$$\hat{\theta}^*(x, y) = \frac{\left\{ \int_x^\infty \int_y^\infty \hat{f}_{b_3}(s, t) ds dt \right\} \hat{f}_b(x, y)}{\left\{ \int_x^\infty \hat{f}_{b_1}(s, y) ds \right\} \left\{ \int_y^\infty \hat{f}_{b_2}(x, t) dt \right\}}, \quad (13)$$

where b, b_1, b_2 , and b_3 denote the bandwidth parameters for each component. Different bandwidths are used, because it is known that the optimal rate of convergence for the bandwidth of a kernel density estimator and a kernel estimator of an integral of a density are different. Thus $\hat{\theta}^*(x, y)$ re-

quires choosing different bandwidths for b, b_1, b_2 , and b_3 . In a simulation study not shown here, we found that $\hat{\theta}^*(x, y)$ is very sensitive to the choice of bandwidths and the normalizing constants.

For models in the AC class, a simpler estimator of the odds ratio function can be derived. Let $k(v)$ be the density of $V = F(X, Y)$. It follows that $k(v) = \{\phi(v)/[\phi'(v)]\} \{\phi''(v)/[\phi'(v)]\} = (K(v) - v)[\theta(v)/v]$ and hence $\theta(v) = \{[vk(v)]/[K(v) - v]\}$. Note that the univariate density function $k(v)$ can be estimated by the kernel estimator, $k(v) = 1/h \sum_{i=1}^n \sum_{j=1}^n \Psi([v - \hat{F}(\tilde{x}_{(i)}, \tilde{y}_{(j)})]/h) \hat{F}(\Delta \tilde{x}_{(i)}, \Delta \tilde{y}_{(j)})$, where $\Psi(\cdot)$ is a univariate kernel function satisfying the usual regularity conditions and h is a positive bandwidth parameter sequence. Hence $\theta(v)$ can be estimated by $\hat{\theta}(v) = \{[vk(v)]/[\hat{K}(v) - v]\}$. Note that when v is close to 0 or 1, the performance of $\hat{\theta}(v)$ will be less stable, because its denominator is close to 0. Under fairly weak conditions, $\hat{k}(v)$ is a pointwise consistent estimator of $k(v)$, and hence $\hat{\theta}(v)$ is a pointwise consistent estimator of $\theta(v)$. The proof of the next result is given in the Appendix. Theorem 4 can easily be shown by applying Theorem 1 and Proposition 2.

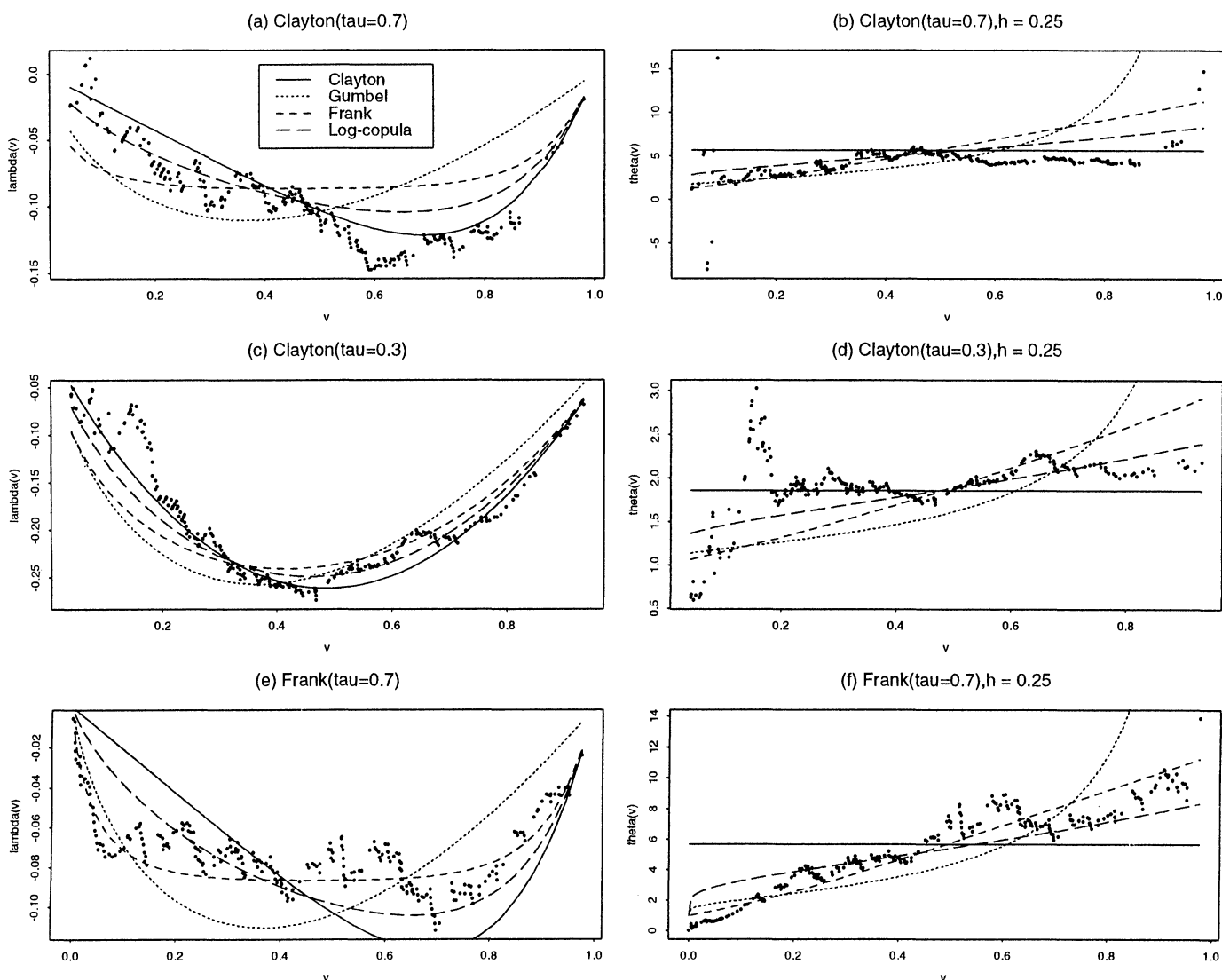


Figure 2. Plots of Empirical and Hypothesized $\lambda_\alpha(v)$. — Clayton, ... Gumbel, - - - Frank, - · - Log-copula. (a) Clayton ($\tau = .7$); (b) Clayton ($\tau = .7$, $h = .25$); (c) Clayton ($\tau = .3$); (d) Clayton ($\tau = .3$, $h = .25$); (e) Frank ($\tau = .7$); (f) Frank ($\tau = .7$, $h = .25$).

Proposition 2. Suppose that the kernel function $\Psi(\cdot)$ is bounded and satisfies $\int |\Psi(t)| dt < \infty$, $\int \Psi(t) dt = 1$, and $|t\Psi(t)| \rightarrow 0$ as $|t| \rightarrow \infty$. Suppose that the bandwidth sequence satisfies $h \rightarrow 0$ and $nh \rightarrow \infty$ as $n \rightarrow \infty$. If $\tilde{K}(v) \xrightarrow{P} K(v)$ on $v \in [\xi, 1]$, then for $v \in (\xi, 1]$, $\tilde{k}(v) \xrightarrow{P} k(v)$.

Theorem 4. Suppose that $\tilde{K}(v) \xrightarrow{P} K(v)$ and $\tilde{k}(v) \xrightarrow{P} k(v)$ for $v \in (\xi, 1]$, then $\tilde{\theta}(v) = \{[v\tilde{k}(v)]/[\tilde{K}(v)-v]\} \xrightarrow{P} \theta(v)$, for each $v \in (\xi, 1]$.

4. NUMERICAL STUDIES

4.1 Real Data Examples

Two datasets were analyzed for illustrative purposes. Dabrowska's estimator was used as the plug-in estimator of F in computing $\tilde{K}(v)$ and $\tilde{\theta}(v)$. The first dataset (Danahy et al. 1977), introduced in Section 1, was from a study on the length of exercise time required to induce angina pectoris for 21 heart disease patients. Here T_1 is the exercise time to angina pectoris at time 0 and T_2 is the exercise time to angina pectoris 3 hours after taking oral isosorbide dini-

trate. Only four observations of T_2 were censored due to patient fatigue. The estimated value of τ is .39 from Wang and Wells (1999) and .48 from Brown et al. (1974). It turns out that theoretical curves of $\lambda_{\hat{\alpha}}(v)$ using $\hat{\alpha}$ inverted from $\hat{\tau} = .48$ are farther away from the nonparametric estimates of $\lambda(v)$ than those from using $\hat{\tau} = .39$, which, however, still may not be the best choice. After trying several values of $\hat{\tau}$, we found that the theoretical curves of $\lambda_{\hat{\alpha}}(v)$ with $\hat{\alpha}$ inverted based on $\hat{\tau} = .37$ are closest to the empirical $\tilde{\lambda}(v)$. By inverting from $\hat{\tau} = .37$, we obtained estimates of α for the models under comparison; the values ($\times 10^{-4}$) of $\hat{S}(\hat{\alpha})$ for the Clayton, Gumbel, Frank, and log-copula models are 12.47, 23.56, 17.29, and 13.27. This implies that the data more likely come from the Clayton model or the log-copula model than from the other two models.

The second dataset (McGilchrist and Aisbett 1991) was from a study of the recurrence time of infection in kidney patients using a portable dialysis machine. Two successive recurrence times, measured from insertion until the next infection, were recorded. The catheter must be removed if an infection occurs. After the infection clears up, the catheter

Table 2. Large-Sample Simulation Results for Frank's Model (Based on 1,000 Replications)

v	.3	.35	.4	.45	.5	.55	.6	.65	.7
$\tau = .3$ ($\alpha = .292$)									
$\tilde{\lambda}(v)$.04 (2.50)	-.61 (2.51)	-.83 (2.43)	-1.47 (2.30)	-1.22 (2.19)	-1.0 (2.11)	-.34 (2.01)	-.43 (1.86)	-.45 (1.66)
$\tilde{\theta}(v)$	7.85 (1.90)	5.50 (2.09)	3.89 (2.25)	2.10 (2.34)	1.12 (2.53)	-.22 (2.72)	-1.32 (2.91)	-3.41 (3.09)	-5.90 (3.26)
$\tau = .5$ ($\alpha = 5.8$)									
$\tilde{\lambda}(v)$	-4.22 (2.27)	-2.99 (2.14)	-2.55 (2.09)	-2.57 (2.11)	-2.79 (2.13)	-2.11 (1.97)	-1.56 (1.86)	-1.36 (1.78)	-1.30 (1.68)
$\tilde{\theta}(v)$	5.83 (3.69)	3.72 (3.69)	1.79 (3.96)	-.41 (4.34)	-2.73 (4.89)	-4.34 (5.15)	-6.67 (5.41)	-11.2 (5.72)	-18.5 (5.97)
$\tau = .7$ ($\alpha = 11.5$)									
$\tilde{\lambda}(v)$	-1.20 (1.71)	-2.14 (1.73)	-1.07 (1.77)	-.85 (1.72)	-1.29 (1.78)	-.77 (1.85)	-.69 (1.74)	-.71 (1.81)	-2.21 (1.75)
$\tilde{\theta}(v)$	258 (20.1)	189 (19.4)	143 (19.8)	85.6 (20.0)	29.6 (21.1)	-19.9 (20.8)	-76.3 (21.8)	-128 (21.7)	-189 (20.8)

NOTE: In each cell, the first row ($\times 10^3$) is the average bias of $\tilde{\lambda}(v)$ and the second row ($\times 10^2$) is the standard deviation of $\tilde{\lambda}(v)$. The third row ($\times 10^2$) is the average bias of $\tilde{\theta}(v)$ and the fourth row ($\times 10^1$) is the standard deviation of $\tilde{\theta}(v)$.

is then reinserted. Censoring may be due to removal for other reasons or the end-of-study effect (for the second infection). Let T_1 be the time to the first infection, and let T_2 be the time to the second infection. There were 38 observations. Here 6 observations of T_1 were censored, 12 observations of T_2 were censored, and 3 observations were doubly censored. The estimate of τ by Wang and Wells (1999) and by Brown et al. (1974) are both close to .21. Using $\hat{\alpha} = \Lambda^{-1}(\hat{\tau})$ with $\hat{\tau} = .21$, the value ($\times 10^{-4}$) of $\hat{S}_n(\hat{\alpha})$ for the Clayton, Gumbel, Frank and log-copula models are 21.72, 11.23, 12.44, and 16.14. The result implies that the Gumbel model provides the best fit to the data.

4.2 Simulation Results

A series of simulation studies were carried out to investigate the performance of the proposed methods. In all of the simulations, we used an algorithm of Genest (1987) to simulate the Frank model and an algorithm of Prentice and Cai (1992) to generate data from the Clayton model. In all of

the cases, (C_1, C_2) were generated from Clayton's model with $\tau = .3$, and the censoring rates in both components were controlled to be between 10 and 20%. The sample size was $n = 250$. The Dabrowska estimator of $F(x, y)$ was used as the plug-in estimator to compute $\tilde{K}(v)$, $\tilde{k}(v)$, $\tilde{\lambda}(v)$, and $\tilde{\theta}(v)$, and $\hat{\tau}$ and α were estimated by $\hat{\alpha} = \Lambda^{-1}(\hat{\tau})$. In computing $\tilde{k}(v)$, we used the Epanechnikov kernel (i.e., $\Psi(v) = \frac{3}{4}[1 - v^2] \mathbb{I}[|v| \leq 1]$), and set the bandwidth as $h = .25$.

Figure 2 shows the diagnostic plots of $\tilde{\lambda}(v)$ and $\tilde{\theta}(v)$ based on one sample replication when the true model is generated from the Clayton and Frank models. The estimated curves in general can capture the shape of the true theoretical curves. It should be noted that for Clayton's model, $\tilde{\theta}(v)$ is very unstable for $v < .2$. Table 2 presents summary statistics (based on 1,000 replications) of $\tilde{\lambda}(v)$ and $\tilde{\theta}(v)$ for the Frank model. In general, $\tilde{\lambda}(v)$ has smaller bias and smaller variation than $\tilde{\theta}(v)$, which is reasonable because $\tilde{\theta}(v)$ involves estimating a density term that has a slower convergence rate. Note that as τ increases, the standard deviation of $\tilde{\theta}(v)$ also increases. The bias of $\tilde{\theta}(v)$ for Frank's model also increases drastically when $\tau = .7$.

The proposed model selection approach has been evaluated; Table 3 presents the results. The "best" fitted model that gives the smallest value of $\hat{S}(\hat{\alpha})$ is selected, where α for the j th model is estimated by $\hat{\alpha}_j = \Lambda_j^{-1}(\hat{\tau})$. It can be seen that when the fitted model is the true one, $\hat{S}(\hat{\alpha})$ has the smallest mean and standard deviation. Table 3 shows that when $\tau \geq .5$, the probability of selecting the true model exceeds 80%. In the case when $\tau = .3$, the probability of selecting the true model decreases to 70%. The log-copula model has the second-highest probability of being chosen when the true model is Frank's or Clayton's model. When the data are drawn from Clayton's model, it seems very unlikely to choose a Gumbel model. As we have seen from the univariate descriptive plots in Figure 1, when τ is small, the log-copula model is between Frank's model and Clayton's model, and the Gumbel model is very different from Clayton's but more similar to Frank's model. The results imply that the proposed model selection method can protect from

Table 3. Large-Sample Simulation Results on \hat{S} (Based on 200 Replications)

		Frank	Clayton	Gumbel	Log-copula
Frank $\tau = .3$	a	68.5%	1%	11.5%	19%
	b	2.6	9.6	5.0	4.5
	c	(1.5)	(5.3)	(2.7)	(3.0)
Frank $\tau = .5$	a	84.0%	0%	7.5%	8.5%
	b	2.5	14.1	6.0	6.1
	c	(1.3)	(5.4)	(2.6)	(3.0)
Frank $\tau = .7$	a	87.5%	0.5%	5.5%	6.5%
	b	2.6	11.5	5.6	5.5
	c	(1.5)	(4.5)	(2.4)	(2.6)
Clayton $\tau = .3$	a	1%	74%	0.5%	24.5%
	b	9.7	3.4	15.5	4.7
	c	(5.7)	(1.8)	(7.0)	(3.4)
Clayton $\tau = .5$	a	0.5%	84%	0%	15.5%
	b	14.7	3.2	23.5	5.8
	c	(6.42)	(1.6)	(7.6)	(3.5)
Clayton $\tau = .7$	a	0%	80.5%	0%	19.5%
	b	11.2	3.3	18.3	4.2
	c	(4.8)	(1.8)	(5.6)	(2.8)

NOTE: In each cell, the first row is the percentage of the column model being selected. The second row ($\times 10^4$) is the mean of \hat{S} , and the third row ($\times 10^4$) is the standard deviation of \hat{S} .

fitting a bad model. When τ is small, the models become more alike, so that the mistake of selecting a wrong model becomes less serious.

The bootstrap method discussed in Section 2.3 for estimating $\sigma_{\hat{S}}^2$ was also evaluated. For a sample $\{(\tilde{X}_i, \tilde{Y}_i, \delta_{1i}, \delta_{2i}) : i = 1, \dots, n\}$, we estimated α by $\hat{\alpha}$ and then simulated (V_1^*, \dots, V_n^*) from $V \sim K_{\hat{\alpha}}(v)$, which we then used to compute $\hat{S}^*(\hat{\alpha}^*)$. The bootstrap process was repeated 100 times. The sample standard deviation of $\hat{S}^*(\hat{\alpha}^*)$, denoted by $\hat{\sigma}_{\hat{S}}$, became an estimate of $\sigma_{\hat{S}}$. When the true model was generated from the Clayton model, the sample means of $\hat{\sigma}_{\hat{S}}$ ($\times 10^4$) based on 200 simulation replications were 1.99, 1.69, and 1.92 under $\tau = .3, .5$, and $.7$. When the true model was from Frank's model, the sample means of $\hat{\sigma}_{\hat{S}}$ ($\times 10^4$) were 1.7, 1.56, and 1.66 under $\tau = .3, .5$, and $.7$. Compared with the true value of $\sigma_{\hat{S}}$ listed in the third row of each entry in Table 3, we find that $\hat{\sigma}_{\hat{S}}$ tends to slightly overestimate $\sigma_{\hat{S}}$, but the bias is small.

5. CONCLUSIONS

In this article we have studied several model selection strategies for the class of Archimedean copula models with bivariate censored data. We expressed the important function $K(v)$ as a statistical functional of the joint survival function so that it can be estimated using the von Mises functional techniques discussed by Gill (1989). The proposed approach is quite flexible for adjusting various data-generating mechanisms. As mentioned previously, the plugged-in estimator \hat{F} should account for the underlying censoring structure. Auxiliary information can also be incorporated in the stage of estimating F . For example, if the marginal distributions are specified and can be estimated by $\hat{F}_i^*(\cdot)$ ($i = 1, 2$), the plug-in estimator may be estimated by $\hat{F}(x, y) = \hat{F}_1^*(x)\hat{F}_2^*(y)\hat{\varphi}(x, y)$, where $\hat{\varphi}(x, y)$ can be obtained nonparametrically based on a bivariate product limit expression of Dabrowska (1988) or an integral equation related to the martingale covariance function of Prentice and Cai (1992). When covariate information is available, \hat{F} may be estimated using the idea described by Prentice and Cai (1992).

APPENDIX: PROOFS

A.1 Proof of Theorem 1: Weak Convergence of $\sqrt{n}\{\hat{K}(v) - K(v)\}$

Assume that $\hat{W}(x, y) = \sqrt{n}\{\hat{F}(x, y) - F(x, y)\} \Rightarrow W(x, y)$, where $W(x, y)$ is a continuous Gaussian process on the space $\mathcal{D}[0, \tau_1] \times [0, \tau_2]$ with the Skorohod topology and $F(\tau_1, \tau_2) \equiv \xi > 0$. One can write $-\sqrt{n}\{\hat{K}(v) - K(v)\} = \hat{\alpha}(v) + \hat{\beta}(v) + \hat{\gamma}(v)$, where

$$\hat{\alpha}(v) \equiv \int \int \mathbb{1}(F(x, y) > v) \sqrt{n}\{\hat{F}(dx, dy) - F(dx, dy)\},$$

$$\hat{\beta}(v) \equiv \int \int \sqrt{n}\{\mathbb{1}(\hat{F}(x, y) > v) - \mathbb{1}(F(x, y) > v)\} F(dx, dy),$$

and

$$\hat{\gamma}(v) \equiv \int \int \{\mathbb{1}(\hat{F}(x, y) > v) - \mathbb{1}(F(x, y) > v)\} \times \sqrt{n}\{\mathbb{1}(\hat{F}(dx, dy) > v) - \mathbb{1}(F(dx, dy) > v)\}.$$

The proof contains three steps. Specifically, we show that $\hat{\alpha}(v) \Rightarrow \int \int \mathbb{1}(F(x, y) > v) W(dx, dy) \equiv \alpha(v)$, $\hat{\beta}(v) \Rightarrow \int \int W(x, y) \mu_v(dx, dy) \equiv \beta(v)$, and $\hat{\gamma}(v) = o_p(1)$. Let $\cup_{j=1}^J C_j$ be a partition of $[0, \tau_1] \times [0, \tau_2]$. Define $\omega(f, r) = \sup_{d((x_1, y_1), (x_2, y_2)) \leq r} |f(x_1, y_1) - f(x_2, y_2)|$ to be the modulus of continuity of f .

A.1.1 Part I: Proof of Weak Convergence of $\hat{\alpha}(v)$. First, we show that $\hat{\alpha}(v)$ converges in distribution to $\alpha(v)$. The set $\{(x, y) : F(x, y) > v\}$ can be approximated by a union of disjoint rectangles. For any rectangular set, $B = \{(x, y) : a \leq x \leq b, c \leq y \leq d\}$, the weak convergence of $\hat{W}(x, y)$ to $W(x, y)$ implies that $\int_a^b \int_c^d \hat{W}(dx, dy) = \hat{W}(b, d) - \hat{W}(b, c) - \hat{W}(a, d) + \hat{W}(a, c) \Rightarrow \int_a^b \int_c^d W(dx, dy)$. Letting the mesh of rectangular grids become finer, one can show that for each $v \in [\xi, 1]$, $\hat{\alpha}(v) \Rightarrow \alpha(v)$. Furthermore, one can show that for any finite sequence (v_1, \dots, v_k) in $[\xi, 1]$, the distribution of $\{\hat{\alpha}(v_1), \dots, \hat{\alpha}(v_k)\}$ converges to that of $\{\alpha(v_1), \dots, \alpha(v_k)\}$. It remains to show that the process $\hat{\alpha}(v)$ is tight, which can be proved by showing that for any ε and $\eta > 0$, there exists $\delta > 0$ such that as n large for $v \in [\xi, 1]$,

$$\Pr\left(\sup_{v \leq u \leq v+\delta} |\hat{\alpha}(v) - \hat{\alpha}(u)| > \varepsilon\right) \leq \delta\eta. \quad (\text{A.1})$$

For $v \leq u$, it is easy to see that $\mathbb{1}(F(x, y) > v) - \mathbb{1}(F(x, y) > u) = \mathbb{1}(v < F(x, y) \leq u)$. Let $C_j(v, \delta) = C_j \cap \{(x, y) : v < F(x, y) \leq v + \delta\}$ and let $\cup_{j=1}^{J^*} C_j(v, \delta)$ be a collection of nonempty sets partitioning $\{(x, y) : v < F(x, y) \leq v + \delta\}$. Note that $J^* \ll J$. Without loss of generality, assume that $C_j(v, \delta)$ is a rectangle for each $j = 1, \dots, J^*$; otherwise, it can be approximated by a union of rectangles. For a rectangle set B , it is easy to see that $|\int_B \hat{W}(dx, dy)| \leq 2 \sup_B |\hat{W}(x_1, y_1) - \hat{W}(x_2, y_2)|$, where \sup_B means $\sup_{(x_1, y_1), (x_2, y_2) \in B}$. Given $\varepsilon > 0$, one can find ε_2 such that

$$\begin{aligned} & \Pr\left(\sup_{v < u \leq v+\delta} |\hat{\alpha}(v) - \hat{\alpha}(u)| > \varepsilon\right) \\ & \leq \sum_{j=1}^{J^*} \Pr\left(\sup_{C_j(v, \delta)} |\hat{W}(x_1, y_1) - \hat{W}(x_2, y_2)| > \varepsilon_2\right) \\ & \leq J^* \times \Pr\{\omega(\hat{W}, \max_{1 \leq j \leq J^*} \text{diam}(C_j(v, \delta))) > \varepsilon_2\}. \end{aligned}$$

By tightness of $\hat{W}(x, y)$, for any ε_2 , $\eta > 0$, it is possible that with appropriate choice of $\delta > 0$ and the partition, we have $\Pr\{\omega(\hat{W}, \max_{1 \leq j \leq J^*} \text{diam}(C_j(v, \delta))) > \varepsilon_2\} \leq \delta\eta/J^*$ as $n \rightarrow \infty$.

A.1.2 Part II: Proof of Weak Convergence of $\hat{\beta}(v)$. Let \hat{W}^- and \hat{W}^+ be the negative and positive parts of \hat{W} , which can be shown to converge weakly to $W^-(x, y)$ and $W^+(x, y)$. One can write $\hat{\beta}(v) = \hat{\delta}(v) - \hat{\kappa}(v)$, where

$$\hat{\delta}(v) = \sqrt{n} \int \int \mathbb{1}[v - (\hat{F} - F)^+(x, y) < F(x, y) \leq v] F(dx, dy)$$

and

$$\hat{\kappa}(v) = \sqrt{n} \int \int \mathbb{1}[v < F(x, y) \leq v + (\hat{F} - F)^-(x, y)] F(dx, dy).$$

We first show that

$$\sup_{\xi \leq v \leq 1} \left| \hat{\delta}(v) - \int \int \sqrt{n}(\hat{F} - F)^+(x, y) \mu_v(dx, dy) \right| \rightarrow 0, \quad (\text{A.2})$$

$$\sup_{\xi \leq v \leq 1} \left| \hat{\kappa}(v) - \int \int \sqrt{n}(\hat{F} - F)^-(x, y) \mu_v(dx, dy) \right| \rightarrow 0, \quad (\text{A.3})$$

and

$$\sup_{\xi \leq v \leq 1} \left| \hat{\beta}(v) - \int \int \sqrt{n}(\hat{F} - F)(x, y) \mu_v(dx, dy) \right| \rightarrow 0. \quad (\text{A.4})$$

Now we prove (A.2). Let $S_{n,j} = \sup_{(x,y) \in C_j} \sqrt{n} \{\hat{F} - F\}^+(x, y)$ and $I_{n,j} = \inf_{(x,y) \in C_j} \sqrt{n} \{\hat{F} - F\}^+(x, y)$. To establish (A.2), we can write $\hat{\delta}(v) = \sum_{j=1}^J \hat{\delta}_j(v)$, where

$$\hat{\delta}_j(v) \equiv \sqrt{n} \int \int_{C_j} \mathbb{1}[v - (\hat{F} - F)^+(x, y) < F(x, y) \leq v] F(dx, dy).$$

It follows that

$$\begin{aligned} \hat{\delta}_j(v) &\leq \sqrt{n} \int \int_{C_j} \mathbb{1}[v - S_{n,j}/\sqrt{n} < F(x, y) \leq v] F(dx, dy) \\ &= \sqrt{n} \{\Pr(F(X, Y) \leq v, (X, Y) \in C_j) - \Pr(F(X, Y) \leq v - S_{n,j}/\sqrt{n}, (X, Y) \in C_j)\} \\ &= \sqrt{n} \int_{v - S_{n,j}/\sqrt{n}}^v \mu_s(C_j) ds \quad (\text{under H1 and H2}) \\ &= \sqrt{n} \int_{v - S_{n,j}/\sqrt{n}}^v \{\mu_s(C_j) - \mu_v(C_j)\} ds \\ &\quad + \int \int_{C_j} \sqrt{n} \{\hat{F} - F\}^+(x, y) \mu_v(dx, dy) \\ &\quad + S_{n,j} \mu_v(C_j) - \int \int_{C_j} \sqrt{n} \{\hat{F} - F\}^+(x, y) \mu_v(dx, dy) \\ &\leq \sqrt{n} \int_{v - S_{n,j}/\sqrt{n}}^v \{\mu_s(C_j) - \mu_v(C_j)\} ds \\ &\quad + \int \int_{C_j} \sqrt{n} \{\hat{F} - F\}^+(x, y) \mu_v(dx, dy) \\ &\quad + \{S_{n,j} - I_{n,j}\} \mu_v(C_j). \end{aligned}$$

Similarly, one can show that

$$\begin{aligned} \hat{\delta}_j(v) &\geq \sqrt{n} \int_{v - I_{n,j}/\sqrt{n}}^v \{\mu_s(C_j) - \mu_v(C_j)\} ds \\ &\quad + \int \int_{C_j} \sqrt{n} \{\hat{F} - F\}^+(x, y) \mu_v(dx, dy) \\ &\quad - \{S_{n,j} - I_{n,j}\} \mu_v(C_j). \end{aligned}$$

It is easy to see that $S_{n,j}/\sqrt{n} \rightarrow^p 0$ and $I_{n,j}/\sqrt{n} \rightarrow^p 0$. Under H2, $\mu_s(C_j)$ is continuous at $s \in [\xi, 1]$, and it then follows that as $\Delta \rightarrow 0$,

$$\sqrt{n} \sum_{j=1}^J \sup_{\xi \leq v \leq 1} \left| \int_v^{v+\Delta} \{\mu_s(C_j) - \mu_v(C_j)\} ds \right| \rightarrow^p 0.$$

Note that

$$\begin{aligned} &\sum_{j=1}^J \mu_v(C_j) (S_{n,j} - I_{n,j}) \\ &\leq \sum_{j=1}^J \mu_v(C_j) \max_{1 \leq j \leq J} \sup_{C_j} \sqrt{n} \{\hat{F} - F\}^+(x_1, y_1) \\ &\quad - \{\hat{F} - F\}^+(x_2, y_2) \\ &\leq k(v) \left\{ \sum_{j=1}^J \Pr(X, Y) \in C_j | F(X, Y) = v \right\} \\ &\quad \times \max_{1 \leq j \leq J} \sup_{C_j} \sqrt{n} \{\hat{F} - F\}^+(x_1, y_1) \\ &\quad - \{\hat{F} - F\}^+(x_2, y_2) \\ &\leq k(v) \omega \{\sqrt{n}(\hat{F} - F)^+, \max_{1 \leq j \leq J} \text{diam}(C_j)\}. \end{aligned}$$

By tightness of $n^{1/2} \{\hat{F}(x, y) - F(x, y)\}$, it is again possible to choose an appropriate mesh to get $\omega \{\sqrt{n}(\hat{F} - F)^+, \max_{1 \leq j \leq J} \text{diam}(C_j)\} \rightarrow 0$. Thus (A.2) is proved. Similar arguments hold to show that (A.3) and (A.4) is a consequence of (A.2) and (A.3).

Note that the mapping $v \mapsto \int_0^{\tau_1} \int_0^{\tau_2} f(x, y) \mu_v(dx, dy)$ is a continuous function of $v \in [\xi, 1]$ and also the mapping $f \mapsto \int_0^{\tau_1} \int_0^{\tau_2} f(x, y) \mu_v(dx, dy)$ is a bounded linear (and hence continuous) function of f . Then by theorem 2.3.5 of Shorack and Wellner (1986, p. 48),

$$\int \int \sqrt{n} \{\hat{F} - F\}(x, y) \mu_v(dx, dy) \Rightarrow \int \int W(x, y) \mu_v(dx, dy),$$

and hence for the equivalent process (with the sample path almost surely identical),

$$\hat{\beta}(v) \Rightarrow \int \int W(x, y) \mu_v(dx, dy).$$

A.1.3 Part III: Proof of $\hat{\gamma}(v) = o_p(1)$. Using similar arguments as in Part II, one can write

$$\begin{aligned} \hat{\gamma}(v) &= \int \int \mathbb{1}[v - (\hat{F} - F)^+(x, y) < F(x, y) \leq v] \hat{W}(dx, dy) \\ &\quad - \int \int \mathbb{1}[v < F(x, y) \leq v + (\hat{F} - F)^-(x, y)] \hat{W}(dx, dy). \end{aligned}$$

Then it follows that

$$\begin{aligned} |\hat{\gamma}(v)| &\leq \sum_{j=1}^J \left| \int \int_{C_j} \mathbb{1}[v - S_{n,j}/\sqrt{n} < F(x, y) \leq v] \hat{W}(dx, dy) \right| \\ &\quad + \sum_{j=1}^J \left| \int \int_{C_j} \mathbb{1}[v < F(x, y) \leq v + S_{n,j}^*/\sqrt{n}] \hat{W}(dx, dy) \right|, \end{aligned}$$

where $S_{n,j}^* = \sup_{(x,y) \in C_j} |(\hat{F} - F)^-(x, y)|$. Using similar techniques as in Part I, one can prove that for any $\varepsilon > 0$, $\Pr(|\hat{\gamma}(v)| > \varepsilon) \rightarrow 0$. Note that uniform consistency of $\hat{K}(v)$ can be proved by showing that $\sup_{v \in \mathcal{V}} |\hat{\alpha}(v)/\sqrt{n}| + \sup_{v \in \mathcal{V}} |\hat{\beta}(v)/\sqrt{n}| + \sup_{v \in \mathcal{V}} |\hat{\gamma}(v)/\sqrt{n}| \rightarrow^p 0$. The details are omitted.

A.2 Proof of Theorem 2: Limiting Distribution of $n^{1/2} \{\tilde{\alpha} - \alpha\}$

The proof of the result is like the classical proof of the limiting normality of an M estimator (see Shorack and Wellner 1986, p. 254). $\tilde{\alpha}$ solves $M(\tilde{\alpha}) = 0$, where $M(\alpha) = \partial/\partial\alpha \int \{\tilde{K}(v) - K_\alpha(v)\}^2 dv = 0$. If $K_\alpha(v)$ is twice differentiable with respect to α and the derivatives are bounded, then it follows that $n^{1/2}(\tilde{\alpha} - \alpha) = -([\partial M(\alpha)]/\partial\alpha)^{-1} n^{1/2} M(\alpha) + o_p(1)$, where $-n^{1/2} M(\alpha) \Rightarrow \int \{[\partial K_\alpha(v)]/\partial\alpha\} X(v) dv$ and $[\partial M(\alpha)]/\partial\alpha = 2 \int \{[\partial K_\alpha(v)]/\partial\alpha\}^2 dv - 2 \int \{[\partial^2 K_\alpha(v)]/\partial\alpha^2\} \{\tilde{K}(v) - K_\alpha(v)\} dv = 2 \int \{[\partial K_\alpha(v)]/\partial\alpha\}^2 dv + o_p(1)$.

A.3 Proof of Proposition 2: Consistency of $\tilde{k}(v)$

Because $\tilde{K}(v)$ only jumps at $v = \hat{F}(\tilde{x}_{(i)}, \tilde{y}_{(j)})$ with mass $\hat{F}(\Delta\tilde{x}_{(i)}, \Delta\tilde{y}_{(j)})$ for $v \in [\xi, 1]$, one can write $\tilde{k}(v) = 1/h \int_\xi^1 \Psi[(v - u)/h] d\tilde{K}(u)$. By writing $\tilde{k}(v) - k(v) = \tilde{k}(v) - \hat{k}(v) + \hat{k}(v) - k(v)$, where $\hat{k}(v) = 1/h \int_\xi^1 \Psi[(v - u)/h] dK(u)$, we want to show that $\tilde{k}(v) - \hat{k}(v) = o_p(1)$ and $\hat{k}(v) - k(v) = o_p(1)$. By change of variables and integration by parts, one can write $\hat{k}(v) - k(v) = 1/h \int_\xi^1 \Psi[(v - u)/h] \{d\tilde{K}(u) - dK(u)\} = -1/h \Psi(t) [\tilde{K}(v - ht) - K(v - ht)]|_c^d + 1/h \int_c^d \{\tilde{K}(v - ht) - K(v - ht)\} d\Psi(t)$, where

$[c, d] = [(v-1)/h, (v-\xi)/h]$. By the consistency of \tilde{K} , it can be shown that $\tilde{K}(v-ht) \xrightarrow{P} K(v-ht)$ and by boundedness of $\Psi(\cdot)$, thus $\tilde{k}(v) \xrightarrow{P} k(v)$. Convergence of $\hat{k}(v)$ to $k(v)$ can be shown using techniques similar to those of Parzen (1962).

[Received September 1996. Revised April 1999.]

REFERENCES

- Anderson, J. E., Louis, T. A., Holm, N. V., and Harvald, B. (1992), "Time-Dependent Association Measures for Bivariate Survival Distributions," *Journal of the American Statistical Association*, 87, 641-650.
- Bandein-Roche, K. J., and Liang, K. Y. (1996), "Modeling Failure-Time Associations in Data With Multiple Levels of Clustering," *Biometrika*, 83, 29-40.
- Barbe, P., Genest, C., Ghoudi, K., and Bruno, R. (1996), "On Kendall's Process," *Journal of Multivariate Analysis*, 58, 197-229.
- Brown, B. W. M., Hollander, M., and Korwar, R. M. (1974), "Nonparametric Tests of Independence for Censored Data With Applications to Heart Transplant Studies," in *Reliability and Biometry: Statistical Analysis of Life Length*, Serfling, R. J. (Ed.) pp. 327-354.
- Campbell, G., and Földes, A. (1982), "Large-Sample Properties of Nonparametric Bivariate Estimators with Censored Data," in *Nonparametric Statistical Inference, Colloquia Mathematica-Societatis, János Bolyai*, eds. B. V. Gnedenko, M. L. Puri, and I. Vincze, Amsterdam: North-Holland, pp. 103-122.
- 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. G., and Cuzick, J. (1985), "Multivariate Generalizations of the Proportional Hazards Model," *Journal of the Royal Statistical Society, Ser. A*, 148, 82-117.
- Dabrowska, D. M. (1988), "Kaplan-Meier Estimates on the Plane," *The Annals of Statistics*, 16, 1475-1489.
- (1989), "Kaplan-Meier Estimate on the Plane: Weak Convergence, LIL, and the Bootstrap," *Journal of Multivariate Analysis*, 29, 308-325.
- Danahy, D. J., Burwell, D. T., Aranow, W. S., and Prakash, R. (1977), "Sustained Hemodynamic and Antianginal Effect of High-Dose Oral Isosorbide Dinitrate," *Circulation*, 55, 381-387.
- Frank, M. J. (1979), "On the Simultaneous Associativity of $F(x, y)$ and $1 - x + y - F(x, y)$," *Aequationes Mathematicae*, 19, 194-226.
- Genest, C. (1987), "Frank's Family of Bivariate Distributions," *Biometrika*, 74, 549-555.
- Genest, C., Ghoudi, K., and Rivest, L. P. (1995), "A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions," *Biometrika*, 82, 543-552.
- Genest, C., and MacKay, J. (1986), "The Joy of Copulas; Bivariate Distributions With Uniform Marginals," *The American Statistician*, 40, 280-283.
- Genest, C., and Rivest, L. P. (1993), "Statistical Inference Procedures for Bivariate Archimedean Copulas," *Journal of the American Statistical Association*, 88, 1034-1043.
- Gill, R. D. (1989), "Non- and Semi-Parametric Maximum Likelihood Estimators and the von Mises Method (Part 1)," *Scandinavian Journal of Statistics*, 16, 97-128.
- Gill, R. D., van der Laan, M. J. and Wellner, J. A. (1993), "Inefficient Estimators for Three Multivariate Models," Preprint 769, Department of Mathematics, University of Utrecht.
- Gumbel, E. J. (1960), "Bivariate Exponential Distributions," *Journal of the American Statistical Association*, 55, 698-707.
- (1961), "Bivariate Logistic Distributions," *Journal of the American Statistical Association*, 56, 335-349.
- Hougaard, P. (1986), "A Class of Multivariate Failure Time Distributions," *Biometrika*, 73, 671-678.
- Hsu, L., and Prentice, R. L. (1996), "On Assessing the Strength of Dependence Between Failure Time Variates," *Biometrika*, 83, 491-506.
- Joe, H. (1993), "Parametric Families of Multivariate Distributions With Given Margins," *Journal of Multivariate Analysis*, 46, 262-282.
- Lin, D. Y., and Ying, Z. (1993), "A Simple Nonparametric Estimator of the Bivariate Survival Function under Univariate Censoring," *Biometrika*, 80, 573-581.
- Lin, D. Y., Sun, W., and Ying, Z. (1999), "Nonparametric Estimation of the Gap Time Distributions for Serial Events with Censored Data," *Biometrika*, 86, 59-70.
- Lindeboom, M., Van den Berg, G. J. and Gerald, J. (1994), "Heterogeneity in Models for Bivariate Survival: The Importance of the Mixing Distribution," *Journal of the Royal Statistical Society, Ser. B*, 56, 49-60.
- McGilchrist, C. A., and Aisbett, C. W. (1991), "Regression with Frailty in Survival Analysis," *Biometrics*, 47, 461-466.
- Marshall, A. W., and Olkin, I. (1988), "Families of Multivariate Distributions," *Journal of the American Statistical Association*, 83, 834-841.
- Murphy, S. (1994), "Consistency in a Proportional Hazards Model Incorporating a Random Effect," *The Annals of Statistics*, 22, 712-731.
- Nielsen, G. G., Gill, R. D., Andersen, P. K., and Sorensen, T. I. (1992), "A Counting Process Approach to Maximum Likelihood Estimation in Frailty Models," *Scandinavian Journal of Statistics*, 19, 25-44.
- Oakes, D. (1982), "A Concordance Test for Independence in the Presence of Bivariate Censoring," *Biometrics*, 38, 451-455.
- (1989), "Bivariate Survival Models Induced by Frailties," *Journal of the American Statistical Association*, 84, 487-493.
- Parzen, E. (1962), "On Estimation of a Probability Density Function and Mode," *The Annals of Mathematical Statistics*, 33, 1065-1076.
- Prentice, R. L., and Cai, J. (1992), "Covariance and Survivor Function Estimation Using Censored Multivariate Failure Time Data," *Biometrika*, 79, 495-512.
- Romano, J. P. (1988), "A Bootstrap Revival of Some Nonparametric Distance Tests," *Journal of the American Statistical Association*, 83, 698-708.
- Shih, J. H., and Louis, T. A. (1995), "Inferences on the Association Parameter in Copula Models for Bivariate Survival Data," *Biometrics*, 51, 1384-1399.
- Shorack, G., and Wellner, J. A. (1986), *Empirical Processes With Applications to Statistics*, New York: Wiley.
- Tsai, W. Y., Leurgans, S., and Crowley, J. J. (1986), "Nonparametric Estimation of a Bivariate Survival Function in Presence of Censoring," *The Annals of Statistics*, 14, 1351-1365.
- van der Laan, M. J. (1992), "Efficient Estimator of the Bivariate Survival Function for Right-Censored Data," Technical Report 337, University of California, Berkeley, Dept. of Statistics.
- (1996), "Efficient Estimation in the Bivariate Censoring Model and Repairing MLE," *The Annals of Statistics*, 24, 596-627.
- van der Vaart, A. W., and Wellner, J. A. (1993), *Weak Convergence and Empirical Processes*, New York: Springer-Verlag.
- Vaupel, J. W., Manton, K. G., and Stallard, E. (1979), "The Impact of Heterogeneity in Individual Frailty and the Dynamics of Mortality," *Demography*, 16, 439-454.
- Visser, M. (1996), "Nonparametric Estimation of the Bivariate Survival Function with an Application to Vertically Transmitted AIDS," *Biometrika*, 83, 507-518.
- Wang, W., and Wells, M. T. (1997), "Nonparametric Estimators of the Bivariate Survival Function under Simplified Censoring Conditions," *Biometrika*, 84, 863-880.
- Wang, W., and Wells, M. T. (1998), "Nonparametric Estimation of Successive Duration Times under Dependent Censoring," *Biometrika*, 85, 561-572.
- Wang, W., and Wells, M. T. (1999), "Estimation of Kendall's Tau Under Censoring," technical report, Institute of Statistical Science, Academia Sinica, Taiwan.
- Wells, M. T., and Yeo, K. P. (1996), "Density Estimation With Bivariate Censored Data," *Journal of the American Statistical Association*, 91, 1566-1574.