The Canadian Journal of Statistics
Vol. 33, No 2, 2005, Pages ???–???
La revue canadienne de statistique

Pseudo-likelihood ratio tests for semiparametric multivariate copula model selection

Xiaohong CHEN and Yanqin FAN

Key words and phrases: Bootstrap; copula; model misspecification; non-nested hypothesis; pseudo-likelihood ratio test.

MSC 2000: Primary 62H20; secondary 62P10.

Abstract: The authors propose pseudo-likelihood ratio tests for selecting semiparametric multivariate copula models in which the marginal distributions are unspecified, but the copula function is parameterized and can be misspecified. For the comparison of two models, the tests differ depending on whether the two copulas are generalized non-nested or generalized nested. For more than two models, the procedure is built on the reality check test of White (2000). Unlike the latter, however, the test statistic is automatically standardized for generalized non-nested models (with the benchmark) and ignores generalized nested models asymptotically. The authors illustrate their approach with American insurance claim data.

Tests du rapport des pseudo-vraisemblances pour la sélection de modèles de copules multivariés semiparamétriques

Résumé : Les auteurs proposent l'emploi de tests du rapport des pseudo-vraisemblances pour la sélection de modèles de copules multivariés semiparamétriques dans lesquels les marges ne sont pas précisées et la copule paramétrique peut éventuellement être mal spécifiée. La forme du test permettant de comparer deux modèles varie selon que les copules sous-jacentes sont emboîtées ou non dans un sens large. La procédure permettant de comparer plusieurs modèles à la fois s'inspire du test de réalisme de White (2000). À la différence de ce dernier, cependant, la statistique du test est automatiquement standardisée (par rapport à un étalon) pour les modèles non-emboîtés et fait fi, asymptotiquement, des modèles emboîtés. Les auteurs illustrent leur approche à l'aide de données américaines de sinistres en assurance.

1. INTRODUCTION

Let $X_1 = (X_{11}, \ldots, X_{d1})^\top, \ldots, X_n = (X_{1n}, \ldots, X_{dn})^\top$ be a random sample from the unknown multivariate distribution function $F^0(x_1, \ldots, x_d)$ with continuous marginal distributions $F_1^0, \ldots, F_d^0$. The characterization theorem of Sklar (1959) implies that there exists a unique copula C^0 such that

$$F^0(x_1, \ldots, x_d) = C^0\{F_1^0(x_1), \ldots, F_d^0(x_d)\}$$

for all $x_1, \ldots, x_d \in \mathbb{R}$. Conversely, for any marginal distributions $F_1, \ldots, F_d$ and any copula function C, the function $C\{F_1(x_1), \ldots, F_d(x_d)\}$ is a multivariate distribution function with given marginal distributions $F_1, \ldots, F_d$. This theorem provides the theoretical foundation for the widespread use of the copula approach in generating multivariate distributions from univariate distributions. For reviews, see Joe (1997) and Nelsen (1999).

One important class of copula-based multivariate models is that of semiparametric multivariate copula models. Models in this class are based on parametric copulas but nonparametric marginal distributions. As a multivariate distribution in this class depends on nonparametric functions of only one dimension, it achieves dimension reduction and is still more flexible than purely parametric models. Due to the well-known problem of "curse of dimensionality" associated with the estimation of fully nonparametric multivariate distributions, this dimension reduction technique is particularly useful in high-dimensional modeling.

The commonly used approach to estimating semiparametric multivariate copula distributions is the two-step approach proposed in Oakes (1994), Genest, Ghoudi & Rivest (1995) and Shih & Louis (1995). Let $C(v_1, \ldots, v_d; \alpha)$ be a class of parametric copulas with unknown parameter α. These authors propose a two-step estimator of α, namely

$$\hat{\alpha} = \text{argmax}_\alpha \left[\frac{1}{n} \sum_{t=1}^{n} \log c \left\{ \widetilde{F}_1(X_{1t}), \ldots, \widetilde{F}_d(X_{dt}); \alpha \right\} \right],$$

where $c(v_1, \ldots, v_d; \alpha)$ is the density of the parametric copula $C(v_1, \ldots, v_d; \alpha)$ and $\widetilde{F}_j(x)$ is the rescaled empirical distribution function of $X_{j1}, \ldots, X_{jn}$, namely

$$\widetilde{F}_j(x) = \frac{1}{n+1} \sum_{t=1}^{n} 1(X_{jt} \leq x), \quad j = 1, \ldots, d.$$

Assuming that the parametric copula class $C(v_1, \ldots, v_d; \alpha)$ correctly specifies the true copula C^0, Genest, Ghoudi & Rivest (1995) establish the asymptotic normality of $\hat{\alpha}$ and provide a consistent estimator of its asymptotic variance. When the true copula is the Gaussian copula, $\hat{\alpha}$ is known to be semiparametrically efficient; see, e.g., Klaassen & Wellner (1997) and Genest & Werker (2002). For a class of copula-based semiparametric time series models, Chen & Fan (2002) establish the asymptotic properties of the two-step estimator.

As commonly used parametric copulas (such as the Gaussian, Clayton, Frank or Student's t-copula) lead to multivariate models that may have very different dependence properties, one important issue is the choice of an appropriate parametric copula. A number of existing papers have attempted to address this issue. Notable theoretical contributions were made by Oakes (1989), Genest & Rivest (1993), Wang & Wells (2000), Chen, Fan & Patton (2003), Fermanian (2003), and Genest, Quessy & Rémillard (2005). Practical applications include those of Frees & Valdez (1998), Klugman & Parsa (1999), Junker & May (2002), Breymann, Dias & Embrechts (2003), and Denuit, Purcaru & Van Keilegom (2004). While these represent important steps towards formal statistical model selection in the context of copula-based multivariate models, none of these papers addresses the issue of the statistical significance of the selection result. This paper attempts to bridge this gap.

We first consider model selection between two copula models and then develop tests for model selection when there are more than two copula models under consideration. In contrast to Breymann, Dias & Embrechts (2003), who select the model whose Akaike Information Criterion (AIC) is the smallest value, our tests take into account randomness of the AIC, which ensures that none of the remaining models under consideration performs significantly better than the model selected by our tests. In the case of two models, our tests extend the likelihood ratio tests for model selection of parametric models in Vuong (1989) to semiparametric multivariate copula models. They allow both competing copula models to be misspecified. In particular, unlike the test of Cox (1961) for non-nested hypotheses, our procedure does not require that one of the copula models be correctly specified under the null hypothesis. Although the idea of our testing approach follows that in Vuong (1989), we allow for infinite-dimensional nuisance parameters (marginal distributions) in our model selection criterion. Hence our tests are Pseudo- (or quasi-) Likelihood Ratio (PLR) tests, and the limiting distributions of our test statistics depend on the estimates of the unknown marginal distributions.

As noted earlier, in empirical applications of copulas, it is more common to use several copulas to fit the data and compare the results obtained from different models. To address the model selection issue in this case, we extend the PLR test developed for two competing models to more than two models along the lines of the reality check of White (2000). In this case, the candidate copula models are compared with a benchmark copula model. If no candidate model is closer to the true model (according to the KLIC distance) than the benchmark model, the latter is chosen; otherwise, the candidate model that is closest to the true model is selected.

White (2000) proposes the reality-check test for the superior predictive (forecasting) accuracy of at least one candidate parametric model over the benchmark parametric model when at least one of the candidate models is non-nested with the benchmark model. Hansen (2003) shows via simulation that the power of the reality check of White (2000) can be unduly influenced by certain candidate models. Hansen (2003) is then led to propose a standardized version of the test which relies implicitly on the assumption that all candidate models are non-nested with the benchmark model, and hence has limited applicability compared with the original test of White (2000).

In this paper, we develop a novel test that shares the advantages of both the reality check of White (2000) and the standardized test of Hansen (2003). Our test statistic not only automatically standardizes the PLR statistic associated with generalized non-nested candidate models (with the benchmark model), but also asymptotically removes the effect of generalized nested candidate models (with the benchmark model). Consequently, our test potentially has power gains over the original reality check of White (2000) and is not restricted to the class of non-nested candidate models like the standardized test of Hansen (2003). To illustrate the usefulness of our testing procedure, we applied it to copula model selection for the loss-ALAE data considered by Frees & Valdez (1998), Genest, Ghoudi & Rivest (1998), Klugman & Parsa (1999), and Denuit, Purcaru & Van Keilegom (2004). Compared to their procedures, our test does not restrict the parametric copulas to be Archimedean and takes into account the randomness of the criterion being used.

The remainder of this paper is organized as follows. In Section 2, we introduce the null hypothesis and the PLR statistic. Section 3 extends the asymptotic results of Genest, Ghoudi & Rivest (1995) to the case where the parametric copula misspecifies the true copula. Section 4 establishes the PLR tests and Section 5 presents a sequential test for model selection between two competing models. Section 6 addresses the selection issue for more than two models and provides an empirical application. Section 7 concludes. All the proofs are gathered in an Appendix.

2. THE NULL HYPOTHESIS AND AIC

For each $i = 1, 2$, let $\{C_i(u_1, \ldots, u_d; \alpha_i) : \alpha_i \in \mathcal{A}_i \subset \mathbb{R}^{p_i}\}$ be a class of parametric copulas. We are interested in selecting a parametric copula, $C_1(u_1, \ldots, u_d; \alpha_1)$, say, such that the resulting semi-parametric multivariate distribution $C_1\{F_1^0(x_1), \ldots, F_d^0(x_d); \alpha_1\}$ is closer to the true multivariate distribution

$$F^0(x_1, \ldots, x_d) \equiv C^0\{F_1^0(x_1), \ldots, F_d^0(x_d)\}$$

than $C_2\{F_1^0(x_1), \ldots, F_d^0(x_d); \alpha_2\}$ is.

Standard approaches for model comparisons are AIC and Bayesian Information Criterion (BIC). AIC can be extended to comparisons of semiparametric copula models in a straightforward way. The comments on AIC in the rest of this paper hold for BIC as well provided that the corresponding penalization term in AIC is replaced with that in BIC.

For $i = 1, 2$, let $\hat{\alpha}_i$ denote the two-step estimator defined as

$$\hat{\alpha}_i = \arg\max_{\alpha_i \in \mathcal{A}_i} \left[\frac{1}{n} \sum_{t=1}^{n} \log c_i \left\{ \widetilde{F}_1(X_{1t}), \ldots, \widetilde{F}_d(X_{dt}); \alpha_i \right\} \right], \tag{1}$$

where $c_i(\cdot; \alpha_i)$ is the density function of the copula $C_i(\cdot; \alpha_i)$.

In our semiparametric context, AIC can be carried out by comparing the values of AIC for the two copulas defined as

$$AIC_i = -\frac{2}{n}\sum_{t=1}^{n}\log c_i\left\{\widetilde{F}_1(X_{1t}),\ldots,\widetilde{F}_d(X_{dt});\hat{\alpha}_i\right\} + \frac{2}{n}\,p_i,$$

where p_i is the dimension of α_i for $i = 1, 2$. Hence, Copula Model 1 will be selected if $AIC_1 < AIC_2$ or equivalently if

$$LR_n\left(\widetilde{F}_1,\ldots,\widetilde{F}_d;\hat{\alpha}_2,\hat{\alpha}_1\right) + \frac{1}{n}\,(p_1 - p_2) < 0, \tag{2}$$

where

$$LR_n(\widetilde{F}_1,\ldots,\widetilde{F}_d;\hat{\alpha}_2,\hat{\alpha}_1) = \frac{1}{n}\sum_{t=1}^{n}\log\left[\frac{c_2\left\{\widetilde{F}_1(X_{1t}),\ldots,\widetilde{F}_d(X_{dt});\hat{\alpha}_2\right\}}{c_1\left\{\widetilde{F}_1(X_{1t}),\ldots,\widetilde{F}_d(X_{dt});\hat{\alpha}_1\right\}}\right]$$

is the PLR statistic.

Noting, however, that the PLR statistic or AIC_i is a random variable, the fact that inequality (2) holds for a random sample $X_1,\ldots,X_n$ may not imply that Copula Model 1 performs significantly better than Copula Model 2; it may occur by chance. This motivates us to develop tests for the null hypothesis that Copula Model 1 indeed performs significantly better than Copula Model 2 in the sense that the value of AIC for Copula Model 1 is significantly smaller than the value of AIC for Copula Model 2.

To introduce formally the null hypothesis for model selection between the two copula models, we let $U_{jt} = F_j^0(X_{jt})$ for $j = 1,\ldots,d$ and define

$$\ell_{t,i}(\alpha_i) = \sum_{j=1}^{d}\log\left\{f_j^0(X_{jt})\right\} + \sum_{j=1}^{d}\log\left\{c_i(U_{1t},\ldots,U_{dt};\alpha_i)\right\},$$

in which f_j^0 is the density function of the true marginal distribution function F_j^0 for $j = 1,\ldots,d$. Throughout this paper, we let E^0 and var^0 denote respectively the expectation and variance taken with respect to the true distribution $C^0(F_1^0,\ldots,F_d^0)$. Let

$$\alpha_i^* = \arg\max_{\alpha_i\in\mathcal{A}_i}\mathrm{E}^0\{\ell_{t,i}(\alpha_i)\} = \arg\max_{\alpha_i\in\mathcal{A}_i}\mathrm{E}^0\{\log c_i(U_{1t},\ldots,U_{dt};\alpha_i)\}. \tag{3}$$

be the pseudo true value associated with Copula Model i, $i = 1, 2$.

Following Vuong (1989), we measure the closeness of a semiparametric multivariate copula model to the true model by the minimum of the KLIC over the distributions in the copula model or equivalently by the maximum of $\mathrm{E}^0\{\ell_{t,i}(\alpha_i)\}$. Since only the second term in the expression of $\ell_{t,i}(\alpha_i)$ depends on the copula, an equivalent measure of the closeness of the ith copula model to the true copula model is $\mathrm{E}^0\log\{c_i(U_{1t},\ldots,U_{dt};\alpha_i^*)\}$, $i = 1, 2$; the larger $\mathrm{E}^0\log\{c_i(U_{1t},\ldots,U_{dt};\alpha_i^*)\}$, the closer is the model to the true copula model. This suggests that Copula Model 1 be selected if H_0 holds, where

$$H_0:\ \mathrm{E}^0\left[\log\left\{\frac{c_2(U_{1t},\ldots,U_{dt};\alpha_2^*)}{c_1(U_{1t},\ldots,U_{dt};\alpha_1^*)}\right\}\right] \le 0,$$

and Copula Model 2 be selected if H_1 holds, where

$$H_1:\ \mathrm{E}^0\left[\log\left\{\frac{c_2(U_{1t},\ldots,U_{dt};\alpha_2^*)}{c_1(U_{1t},\ldots,U_{dt};\alpha_1^*)}\right\}\right] > 0.$$

Obviously, when the two copulas have the same number of parameters, namely $p_1 = p_2$, Copula Model 1 is chosen according to AIC if $LR_n(\widetilde{F}_1,\ldots,\widetilde{F}_d;\hat{\alpha}_2,\hat{\alpha}_1) < 0$. To test H_0, however, we need to take into account the randomness of the PLR statistic. This will be accomplished in the rest of this paper.

3. ASYMPTOTIC PROPERTIES OF THE TWO-STEP ESTIMATOR UNDER MISSPECIFICATION

Almost all the existing work on the asymptotic properties of the two-step estimator $\hat{\alpha}$ introduced in Section 1 assumes that the parametric copula correctly specifies the true copula function. One exception is the recent paper by Cebrián, Denuit & Scaillet (2003), in which the asymptotic distribution of the two-step estimator under misspecification is provided. However, these authors do not give explicit regularity conditions for either the asymptotic normality or consistency of the estimator to the pseudo true value.

In this section, we establish asymptotic properties of the two-step estimator $\hat{\alpha}$ when the parametric copula misspecifies the true copula function. Although White (1982) has established asymptotic properties of the maximum likelihood estimator under misspecified parametric models, his results are not directly applicable here since the two-step estimation of a misspecified semiparametric copula model depends on the estimates of unknown marginal distributions. These results are not only useful in their own right but essential to the construction of our PLR tests.

Let $\mathcal{A} \subset \mathbb{R}^p$ be the parameter space. For α, $\alpha^* \in \mathcal{A}$, we use $||\alpha - \alpha^*||$ to denote the usual Euclidean metric. In addition, for $j = 1, \ldots, d$ let

$$\ell(v_1, \ldots, v_d; \alpha) = \log c(v_1, \ldots, v_d; \alpha), \quad \ell_\alpha(v_1, \ldots, v_d; \alpha) = \frac{\partial}{\partial \alpha} \ell(v_1, \ldots, v_d; \alpha),$$

$$\ell_j(v_1, \ldots, v_d; \alpha) = \frac{\partial}{\partial v_j} \ell(v_1, \ldots, v_d; \alpha), \quad \ell_{\alpha\alpha}(v_1, \ldots, v_d; \alpha) = \frac{\partial^2}{\partial \alpha \partial \alpha'} \ell(v_1, \ldots, v_d; \alpha)$$

and

$$\ell_{\alpha j}(v_1, \ldots, v_d; \alpha) = \frac{\partial^2}{\partial v_j \partial \alpha} \ell(v_1, \ldots, v_d; \alpha).$$

3.1. Convergence in probability.

Throughout the paper, we let $U_t = (U_{1t}, \ldots, U_{dt})^\top$ with $U_{jt} = F_j^0(X_{jt})$, $j = 1, \ldots, d$. The following conditions are sufficient to ensure the convergence of the two-step estimator $\hat{\alpha}$ to the pseudo true value α^* defined in C1.

- C1: $\mathrm{E}^0\{\ell(U_{1t}, \ldots, U_{dt}; \alpha)\}$ has a unique maximum α^* in $\mathcal{A}$, where $\mathcal{A}$ is a compact subset of $\mathbb{R}^p$.
- C2: $X_1 = (X_{11}, \ldots, X_{d1})^\top, \ldots, X_n = (X_{1n}, \ldots, X_{dn})^\top$ is an i.i.d. sample from the unknown distribution $F^0(x_1, \ldots, x_d)$ with continuous marginal distributions $F_1^0, \ldots, F_d^0$.
- C3: The true (unknown) copula function $C^0(u_1, \ldots, u_d)$ has continuous partial derivatives.
- C4: (i) For any $u \in (0,1)^d$, $\ell(u; \alpha)$ is a continuous function of α;
  (ii) $\mathrm{E}^0\{\sup_{\alpha \in \mathcal{A}} |\ell(U_{1t}, \ldots, U_{dt}; \alpha)|\} < \infty$.

Proposition 1. *Under conditions C1–C4, we have:*

(a) $||\widehat{\alpha} - \alpha^*|| = o_{a.s.}(1)$;

(b) $\displaystyle \frac{1}{n} \sum_{t=1}^{n} \ell\left\{\widetilde{F}_1(X_{1t}), \ldots, \widetilde{F}_d(X_{dt}); \widehat{\alpha}\right\} = \mathrm{E}^0\left\{\ell(F_1^0(X_{1t}), \ldots, F_d^0(X_{dt}); \alpha^*)\right\} + o_{a.s.}(1).$

Proposition 1 (a) states that the two-step estimator $\hat{\alpha}$ is a consistent estimator of the pseudo true value α^*. If the parametric copula correctly specifies the true copula in the sense that there exists $\alpha^0 \in \mathcal{A}$ such that $C(v_1, \ldots, v_d; \alpha^0) = C^0(v_1, \ldots, v_d)$ for almost all $(v_1, \ldots, v_d) \in (0,1)^d$, then $\alpha^* = \alpha^0$ and $\hat{\alpha}$ consistently estimates α^0 as shown in Genest, Ghoudi & Rivest (1995) for independent observations and in Chen & Fan (2002) for time series models.

3.2. Limiting distribution.

The following conditions are sufficient to ensure the $\sqrt{n}$–asymptotic normality of $\hat{\alpha}$.

A1: (i) C1 holds with $\alpha^* \in \text{int}(\mathcal{A})$;

(ii) $B \equiv -\mathrm{E}^0\{\ell_{\alpha\alpha}(U_{1t}, \ldots, U_{dt}; \alpha^*)\}$ is positive definite;

(iii) $\Sigma \equiv \text{var}^0\{\ell_\alpha(U_{1t}, \ldots, U_{dt}; \alpha^*) + \sum_{j=1}^d W_j(U_{jt}; \alpha^*)\}$ is finite, positive definite, where

$$W_j(U_{jt}; \alpha^*) \equiv \mathrm{E}^0[\ell_{\alpha j}(U_{1s}, \ldots, U_{ds}; \alpha^*)\{I(U_{jt} \le U_{js}) - U_{js}\}|U_{jt}].$$

A2: For $j = 1, \ldots, d$, $\ell_{\alpha j}(u_1, \ldots, u_d; \alpha^*)$ is well-defined and continuous in $(u_1, \ldots, u_d) \in (0,1)^d$.

A3: (i) $||\ell_\alpha(v_1, \ldots, v_d; \alpha^*)|| \le \text{constant} \times \prod_{j=1}^d \{v_j(1-v_j)\}^{-a_j}$ for some $a_j \ge 0$ such that

$$\mathrm{E}^0\left[\prod_{j=1}^d \{U_{jt}(1-U_{jt})\}^{-2a_j}\right] < \infty;$$

(ii) $||\ell_{\alpha k}(v_1, \ldots, v_d; \alpha^*)|| \le \text{constant} \times \{v_k(1-v_k)\}^{-b_k} \prod_{j=1, j\neq k}^d \{v_j(1-v_j)\}^{-a_j}$ for some $b_k > a_k$ such that

$$\mathrm{E}^0\left[\{U_{kt}(1-U_{kt})\}^{\xi_k - b_k} \prod_{j=1, j\neq k}^d \{U_{jt}(1-U_{jt})\}^{-a_j}\right] < \infty$$

for some $\xi_k \in (0, 1/2)$.

A4: (i) For every $(u_1, \ldots, u_d) \in (0,1)^d$, $\ell_{\alpha\alpha}(u_1, \ldots, u_d; \alpha)$ is a continuous function of α in a neighbourhood of α^*;

(ii) $\mathrm{E}^0\left\{\sup_{\alpha \in \mathcal{A}: ||\alpha - \alpha^*|| = o(1)} ||\ell_{\alpha\alpha}(U_{1t}, \ldots, U_{dt}; \alpha)||\right\} < \infty.$

Condition A3 allows the score function and its partial derivatives with respect to the first d arguments to blow up at the boundaries, which do occur for many popular copula functions such as Gaussian copula, t-copula and Clayton copula.

PROPOSITION 2. *Under conditions C2–C4, A1–A4, we have* $\sqrt{n}(\widehat{\alpha} - \alpha^*) \to \mathcal{N}(0, B^{-1}\Sigma B^{-1})$ *in distribution, where* B *and* Σ *are defined in A1.*

The additional term $\sum_{j=1}^d W_j(U_{jt}; \alpha^*)$ in the definition of Σ is introduced by the need to estimate the marginal distribution functions $F_1^0, \ldots, F_d^0$. If the latter are completely known, this term disappears.

Remark 1. Since the parametric copula may be misspecified, the information matrix equality

$$B = \mathrm{E}^0 \left\{ \ell_\alpha(U_t; \alpha^*) \ell_\alpha(U_t; \alpha^*)^\top \right\}$$

may not hold. Therefore, the asymptotic variance of $\widehat{\alpha}$ cannot be reduced to

$$B^{-1} + B^{-1} \mathrm{var}^0 \left\{ \sum_{j=1}^{d} W_j(U_{jt}; \alpha^*) \right\} B^{-1}$$

as in Genest, Ghoudi & Rivest (1995) for the correctly specified copula case.

Remark 2. The asymptotic variance of $\widehat{\alpha}$ can be consistently estimated by $\widehat{B}^{-}\widehat{\Sigma}\widehat{B}^{-}$, where $\widehat{B}^{-}$ is the generalized inverse of

$$\widehat{B} = -\frac{1}{n} \sum_{t=1}^{n} \ell_{\alpha\alpha}(\tilde{U}_t; \hat{\alpha})$$

in which $\widetilde{U}_t = (\widetilde{U}_{1t}, \ldots, \widetilde{U}_{dt})^\top$, $\widetilde{U}_{jt} = \widetilde{F}_j(X_{jt})$ for $j = 1, \ldots, d$, and

$$\widehat{\Sigma} = \frac{1}{n} \sum_{t=1}^{n} \left\{ \ell_\alpha(\tilde{U}_t; \hat{\alpha}) + \sum_{j=1}^{d} \widehat{W}_j(\widetilde{U}_{jt}; \hat{\alpha}) \right\} \left\{ \ell_\alpha(\tilde{U}_t; \hat{\alpha}) + \sum_{j=1}^{d} \widehat{W}_j(\widetilde{U}_{jt}; \hat{\alpha}) \right\}^\top,$$

with

$$\widehat{W}_j(U_{jt}; \hat{\alpha}) = \frac{1}{n} \sum_{s=1, s\neq t}^{n} \ell_{\alpha j}(U_s; \hat{\alpha}) \{ I(U_{jt} \leq U_{js}) - U_{js} \}.$$

Any inference drawn based on $\hat{\alpha}$ and the variance estimator $\widehat{B}^{-}\widehat{\Sigma}\widehat{B}^{-}$ would still be valid, except that it is about the pseudo true value α^* and the estimated parametric copula estimates the closest copula in the parametric family to the true copula in terms of minimizing the KLIC.

4. ASYMPTOTIC PROPERTIES OF THE PLR STATISTIC AND PLR TESTS

As will be shown later, the asymptotic distribution of the PLR statistic takes different form depending on whether the two closest parametric copulas to the true copula are equal. To distinguish between these two cases, we introduce the concept of generalized non-nested and of generalized nested copula models.

Definition 1.

(i) Two models are *generalized non-nested* if the set

$$\{(v_1, \ldots, v_d) : c_1(v_1, \ldots, v_d; \alpha_1^*) \neq c_2(v_1, \ldots, v_d; \alpha_2^*)\}$$

has positive Lebesgue measure.

(ii) Two models are *generalized nested* if $c_1(v_1, \ldots, v_d; \alpha_1^*) = c_2(v_1, \ldots, v_d; \alpha_2^*)$ for almost all $(v_1, \ldots, v_d) \in (0,1)^d$.

Given the definition of the pseudo true value α_i^* in (3), the closest $c_i(\cdot; \alpha_i^*)$ to the true copula c^0 (according to KLIC) in a parametric class of copulas $\{c_i(\cdot; \alpha_i) : \alpha_i \in \mathcal{A}_i\}$ depends on the true (but unknown) copula. Hence it is not obvious a priori whether two parametric classes of copulas are generalized non-nested or generalized nested.

Some commonly used parametric classes of copulas such as the Clayton copula, the Gumbel–Hougaard copula, the Frank copula and the Gaussian copula are generalized nested only when the closest member to the true copula in each class is the independence copula; otherwise they are generalized non-nested. Also, the Gaussian copula and the student t-copula are generalized nested if the closest member to the true copula in the student t-copula family is the one with degree of freedom being infinity (i.e., a Gaussian copula).

By applying Proposition 1 (b), we immediately obtain the probability limit of the PLR statistic.

Theorem 1. *Suppose that Copula Models* 1 *and* 2 *satisfy the conditions of Proposition 1. Then the PLR statistic* $LR_n(\tilde{F}_1, \ldots, \tilde{F}_d; \hat{\alpha}_2, \hat{\alpha}_1)$ *converges almost surely to*

$$\mathrm{E}^0 \left[\log \left\{ \frac{c_2(U_{1t}, \ldots, U_{dt}; \alpha_2^*)}{c_1(U_{1t}, \ldots, U_{dt}; \alpha_1^*)} \right\} \right].$$

In the following, we adopt the convention that all the notations involving the copula function $C(u_1, \ldots, u_d; \alpha)$ introduced in Section 3 are now indexed by a subscript i for $i = 1, 2$ to make explicit their dependence on the parametric Copula Model i. In addition, we define

$$\Sigma_{12} = \mathrm{cov}^0 \left\{ \ell_{1,\alpha}\left(U_s; \alpha_1^*\right) + \sum_{j=1}^{d} W_{1,j}\left(U_{js}; \alpha_1^*\right), \ell_{2,\alpha}(U_s; \alpha_2^*) + \sum_{j=1}^{d} W_{2,j}(U_{js}; \alpha_2^*) \right\},$$

and

$$\sigma^2 = \mathrm{var}^0 \left[\{\ell_2(U_s; \alpha_2^*) - \ell_1(U_s; \alpha_1^*)\} + \sum_{j=1}^{d} \{Q_{2,j}(U_{js}; \alpha_2^*) - Q_{1,j}(U_{js}; \alpha_1^*)\} \right]$$

where

$$Q_{i,j}(U_{js}, \alpha_i^*) \equiv \mathrm{E}^0 \left[\ell_{i,j}(U_t; \alpha_i^*)\{I(U_{js} \le U_{jt}) - U_{jt}\} | U_{js} \right], \quad 1 \le j \le d, \quad i = 1, 2. \tag{4}$$

Theorem 2. *Suppose that Copula Models* 1 *and* 2 *satisfy the conditions of Proposition 2. Then*

(a) *For the generalized non-nested case,*

$$\sqrt{n} \left[LR_n(\tilde{F}_1, \ldots, \tilde{F}_d; \hat{\alpha}_2, \hat{\alpha}_1) - \mathrm{E}^0 \left[\log \left\{ \frac{c_2(U_{1t}, \ldots, U_{dt}; \alpha_2^*)}{c_1(U_{1t}, \ldots, U_{dt}; \alpha_1^*)} \right\} \right] \right] \to \mathcal{N}(0, \sigma^2).$$

(b) *For the generalized nested case,*

$$\begin{aligned} 2nLR_n\left(\tilde{F}_1, \ldots, \tilde{F}_d; \hat{\alpha}_2, \hat{\alpha}_1\right) &= n(\alpha_2^* - \hat{\alpha}_2)^\top B_2(\alpha_2^* - \hat{\alpha}_2) - n(\alpha_1^* - \hat{\alpha}_1)^\top B_1(\alpha_1^* - \hat{\alpha}_1) + o_p(1) \\ &\to M_{p_1+p_2}(\cdot; \lambda^*), \end{aligned}$$

where $M_{p_1+p_2}(\cdot; \lambda^*)$ *is the distribution of a weighted sum of independent* $\chi^2_{(1)}$ *random variables in which the weight* $\lambda^* = (\lambda_1^*, \ldots, \lambda_{p_1+p_2}^*)^\top$ *is the* $(p_1 + p_2) \times 1$ *vector of eigenvalues of the matrix* W *defined as*

$$W = \begin{pmatrix} \Sigma_2 B_2^{-1} & -\Sigma_{12}^\top B_1^{-1} \\ \Sigma_{12} B_2^{-1} & -\Sigma_1 B_1^{-1} \end{pmatrix}.$$

Compared with Theorem 3.3 in Vuong (1989), the variance of the asymptotic distribution of the PLR statistic for generalized non-nested models has the additional term due to

$$\sum_{j=1}^{d}\{Q_{2,j}(U_{jt};\alpha_2^*) - Q_{1,j}(U_{jt};\alpha_1^*)\}.$$

This term is introduced by the first step estimation of the marginal distributions $F_1^0, \ldots, F_d^0$.

The following proposition shows that $\sigma^2 > 0$ if and only if the two copula models are generalized non-nested. Consequently, for generalized non-nested models, the null limiting distribution of $\sqrt{n}\, LR_n(\tilde{F}_1, \ldots, \tilde{F}_d; \hat{\alpha}_2, \hat{\alpha}_1)$ is a normal distribution with a positive variance. This is the basis for the PLR test for model selection in generalized non-nested case developed in this paper.

Proposition 3. *Let*

$$\sigma_a^2 = \mathrm{var}^0\left[\log\left\{\frac{c_2(U_{1t}, \ldots, U_{dt};\alpha_2^*)}{c_1(U_{1t}, \ldots, U_{dt};\alpha_1^*)}\right\}\right].$$

Under the conditions of Theorem 2, $\sigma^2 = 0$ if and only if $\sigma_a^2 = 0$, and $\sigma_a^2 = 0$ if and only if the two copula models under selection are generalized nested.

4.1. A PLR test for selection of generalized non-nested models.

Unlike Vuong (1989), for generalized non-nested models, the null hypothesis in our paper is a composite hypothesis. As a result, the asymptotic distribution of the PLR statistic under the null is not uniquely determined, see Theorem 2 (a). The usual approach to handling this problem is based on the Least Favourable Configuration (LFC), which is the point least favourable to the alternative. In our case, the LFC satisfies

$$\mathrm{E}^0\left[\log\left\{\frac{c_2(U_{1t}, \ldots, U_{dt};\alpha_2^*)}{c_1(U_{1t}, \ldots, U_{dt};\alpha_1^*)}\right\}\right] = 0.$$

Under the LFC, Theorem 2 (a) implies that $\sqrt{n}\, LR_n(\tilde{F}_1, \ldots, \tilde{F}_d; \hat{\alpha}_2, \hat{\alpha}_1) \to \mathcal{N}(0, \sigma^2)$. Moreover, $\sigma^2 > 0$ by Proposition 3. We now provide a consistent estimator of σ^2. First, we need to estimate

$$\sum_{j=1}^{d}\{Q_{2,j}(U_{js};\alpha_2^*) - Q_{1,j}(U_{js};\alpha_1^*)\}.$$

For a given s, define

$$\sum_{j=1}^{d}\left\{\widehat{Q}_{2,j}(U_{js};\hat{\alpha}_2) - \widehat{Q}_{1,j}(U_{js};\hat{\alpha}_1)\right\} = \frac{1}{n}\sum_{t=1, t\neq s}^{n}\hat{P}_n(U_t, U_s),$$

where

$$\hat{P}_n(U_t, U_s) = \sum_{j=1}^{d}\{\ell_{2,j}(U_t;\hat{\alpha}_2) - \ell_{1,j}(U_t;\hat{\alpha}_1)\}\{I(U_{js} \leq U_{jt}) - U_{jt}\}.$$

Thus a consistent estimator of σ^2 is given by

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}\left[\log\frac{c_2(\tilde{U}_t;\hat{\alpha}_2)}{c_1(\tilde{U}_t;\hat{\alpha}_1)} - \frac{1}{n}\sum_{s=1}^{n}\log\frac{c_2(\tilde{U}_s;\hat{\alpha}_2)}{c_1(\tilde{U}_s;\hat{\alpha}_1)} + \sum_{j=1}^{d}\left\{\widehat{Q}_{2,j}(\widetilde{U}_{jt};\hat{\alpha}_2) - \widehat{Q}_{1,j}(\widetilde{U}_{jt};\hat{\alpha}_1)\right\}\right]^2,$$

where $\tilde{U}_t = (\tilde{U}_{1t}, \ldots, \tilde{U}_{dt})^\top = (\tilde{F}_1(X_{1t}), \ldots, \tilde{F}_d(X_{dt}))^\top$.

Define the statistic for the PLR test for the selection of generalized non-nested models as

$$T_n^N = \frac{\sqrt{n}}{\hat{\sigma}} LR_n\left(\tilde{F}_1, \ldots, \tilde{F}_d; \hat{\alpha}_2, \hat{\alpha}_1\right),$$

where the superscript N is meant for non-nested models and normal limiting distributions.

Theorem 3. *Suppose the conditions of Theorem 2 hold and that the two models are generalized non-nested. Then under the LFC, $T_n^N \to \mathcal{N}(0,1)$.*

Theorems 1 and 3 suggest the following directional test for H_0: Given a significance level α, reject H_0 in favour of H_1 if $T_n^N > Z_\alpha$, where Z_α is the upper α–percentile of the standard normal distribution.

Remark 3. Like AIC, one can extend the above test to take into account the dimensions of the two copula models. Let

$$T_n^{NP} = \frac{\sqrt{n}}{\hat{\sigma}} \left\{ LR_n\left(\tilde{F}_1, \ldots, \tilde{F}_d; \hat{\alpha}_2, \hat{\alpha}_1\right) + \frac{1}{n}(p_1 - p_2) \right\}.$$

Since $(p_1 - p_2)/\sqrt{n} \to 0$, the comparison of T_n^{NP} and Z_α provides a valid test for H_0 versus H_1.

4.2. A PLR test for selection of generalized nested models.

We now consider the case where under H_0, the two models are generalized nested. In this case, the null hypothesis becomes a simple hypothesis, namely $C_1(v_1, \ldots, v_d; \alpha_1^*) = C_2(v_1, \ldots, v_d; \alpha_2^*)$ for almost all $(v_1, \ldots, v_d) \in [0,1]^d$.

Define the test statistic

$$T_n^Q = 2nLR_n\left(\tilde{F}_1, \ldots, \tilde{F}_d; \hat{\alpha}_2, \hat{\alpha}_1\right),$$

where the superscript Q in T_n^Q is meant for nested models and quadratic limiting statistics. Theorem 2 (b) implies that in this case the null limiting distribution of the PLR statistic T_n^Q is not distribution-free; it depends on the eigenvalues of the matrix W. One way to obtain the asymptotic critical values for the test based on T_n^Q is to estimate the matrix W and compute eigenvalues of the estimate of W.

Alternatively, one may use the method of bootstrap to approximate the critical values of the test. We observe from the proof of Theorem 4.2 in the Appendix that

$$LR_n\left(\tilde{F}_1, \ldots, \tilde{F}_d; \hat{\alpha}_2, \hat{\alpha}_1\right) = LR_n\left(\tilde{F}_1, \ldots, \tilde{F}_d; \alpha_2^*, \alpha_1^*\right) + D_n + o_p(1/n), \tag{5}$$

where

$$D_n = \frac{1}{2}(\hat{\alpha}_2 - \alpha_2^*)^\top B_2(\hat{\alpha}_2 - \alpha_2^*) - \frac{1}{2}(\hat{\alpha}_1 - \alpha_1^*)^\top B_1(\hat{\alpha}_1 - \alpha_1^*).$$

In the generalized nested case,

$$LR_n\left(\tilde{F}_1, \ldots, \tilde{F}_d; \alpha_2^*, \alpha_1^*\right) = 0$$

and hence $T_n^Q = 2nD_n + o_p(1)$, implying that the null limiting distribution of T_n^Q is given by the limiting distribution of $2nD_n$. Moreover the form of the limiting distribution of $2nD_n$ remains the same regardless of whether the null hypothesis holds or not. This suggests that we bootstrap the statistic $2nD_n$. The detailed procedure is summarized below.

Step 1: Draw a bootstrap random sample $X_1^b, \ldots, X_n^b$ with replacement from $X_1, \ldots, X_n$.

Step 2: Compute the two-step estimates $\hat{\alpha}_1^b$ and $\hat{\alpha}_2^b$ using the bootstrap sample $X_1^b, \ldots, X_n^b$.

Step 3: Compute

$$2D_n^b = (\hat{\alpha}_2^b - \hat{\alpha}_2)^\top \hat{B}_2 \left(\hat{\alpha}_2^b - \hat{\alpha}_2\right) - \left(\hat{\alpha}_1^b - \hat{\alpha}_1\right)^\top \hat{B}_1 \left(\hat{\alpha}_1^b - \hat{\alpha}_1\right),$$

where

$$\hat{B}_i = n^{-1} \sum_{t=1}^{n} \ell_{i,\alpha\alpha} \left\{\tilde{F}_1(X_{1t}), \ldots, \tilde{F}_d(X_{dt}); \hat{\alpha}_i\right\}, \quad i = 1, 2.$$

Step 4: Repeat Steps 1–3 a large number of times and use the empirical distribution of the resulting values of $2nD_n^b$ to approximate the null distribution of the test statistic T_n^Q.

In the sequel, we let $\mathrm{P}^*(\cdot|X_1, \ldots, X_n)$ denote the conditional distribution given the data $X_1, \ldots, X_n$. Theorem 4 below justifies the bootstrap test: reject H_0 if $T_n^Q > c^b$, where c^b is the critical value of the conditional distribution of $2nD_n^b$ given $X_1, \ldots, X_n$.

Theorem 4. *Under the conditions of Theorem* 2 (b), $\mathrm{P}^*(2nD_n^b \leq x|X_1, \ldots, X_n)$ *converges in probability to* $M_{p_1+p_2}(x; \lambda^*)$.

Remark 4. Let T_n^{Qb} denote the bootstrap version of the test statistic T_n^Q. The proposed bootstrap procedure is equivalent to using the distribution of the recentered T_n^{Qb} at T_n^Q, i.e., $(T_n^{Qb} - T_n^Q)$. To see why, consider the analog of (5) for the bootstrap sample:

$$LR_n\left(\tilde{F}_1^b, \ldots, \tilde{F}_d^b; \hat{\alpha}_2^b, \hat{\alpha}_1^b\right) = LR_n\left(\tilde{F}_1, \ldots, \tilde{F}_d; \hat{\alpha}_2, \hat{\alpha}_1\right) + D_n^b + o_{p^*}(1/n). \tag{6}$$

Unlike for the original sample, the first term on the right-hand side of (6) may not be zero and hence must be subtracted from the left hand side of (6) to form the bootstrap statistic which, when multiplied by $2n$, has the conditional distribution given the original sample converging in probability to the null limiting distribution of the test statistic T_n^Q.

Remark 5. The fact that regardless of whether H_0 holds, the conditional distribution of $2nD_n^b$ always converges in probability to $M_{p_1+p_2}(x; \lambda^*)$, the null limiting distribution of the test statistic T_n^Q, ensures good power properties of the bootstrap test. For example, since T_n^Q diverges to $+\infty$ if H_0 fails, the power of the bootstrap test will approach 1 as $n \to \infty$.

5. A SEQUENTIAL PLR TEST FOR MODEL SELECTION

The tests presented in the previous section are general in the sense that they are valid whether the parametric models are correct or not. However, one needs to know whether the two models are generalized non-nested or nested. As the pseudo true values α_1^* and α_2^* are unknown, it is generally unknown a priori if this is the case. We will follow Vuong (1989) by providing a sequential test in which one first tests if the two models are generalized non-nested and then determines which test to use based on the result of the pre-test.

It follows from Proposition 3 that the null hypothesis of generalized nested models can be tested by testing the null hypothesis $\sigma_a^2 = 0$. A consistent estimator of σ_a^2 is given by

$$\hat{\sigma}_a^2 = \frac{1}{n} \sum_{t=1}^{n} \left[\log\left\{\frac{c_2(\tilde{U}_t; \hat{\alpha}_2)}{c_1(\tilde{U}_t; \hat{\alpha}_1)}\right\} - \frac{1}{n} \sum_{s=1}^{n} \log\left\{\frac{c_2(\tilde{U}_s; \hat{\alpha}_2)}{c_1(\tilde{U}_s; \hat{\alpha}_1)}\right\}\right]^2.$$

In the following, we denote $\lambda^{*2} = (\lambda_1^{*2}, \ldots, \lambda_{p_1+p_2}^{*2})^\top$ as the vector of squares of $\lambda^* = (\lambda_1^*, \ldots, \lambda_{p_1+p_2}^*)^\top$, the eigenvalue weights in Theorem 2 (b).

THEOREM 5. *Assume that the conditions of Theorem 2 hold. Then*

(a) $\hat{\sigma}_a^2$ *is a consistent estimator of* σ_a^2.

(b) *When* $\sigma_a^2 = 0$, $n\hat{\sigma}_a^2 \to M_{a_1+a_2}(\cdot; \lambda^{*2})$ *in distribution.*

Theorem 5 suggests that a sequential test can be constructed in our case as well. First, one tests the null hypothesis that the two copula models are generalized nested by using the test statistic $n\hat{\sigma}_a^2$; if the pretest suggests that the two models are generalized nested, then stop; otherwise proceed to use the test T_n^N for H_0.

Like the null limiting distribution of T_n^Q, that of $n\hat{\sigma}_a^2$ is not distribution-free. We propose to use bootstrap to approximate its null distribution. The proof of Theorem 5 reveals that the null limiting distribution of $\hat{\sigma}_a^2$ is given by the limiting distribution of

$$V = \begin{pmatrix} \hat{\alpha}_2 - \alpha_2^* \\ \hat{\alpha}_1 - \alpha_1^* \end{pmatrix}^\top \begin{bmatrix} \mathrm{E}^0 \ell_{2,\alpha}(U_t; \alpha_2^*)^\top \ell_{2,\alpha}(U_t; \alpha_2^*) & -\mathrm{E}^0 \ell_{2,\alpha}(U_t; \alpha_2^*)^\top \ell_{1,\alpha}(U_t; \alpha_1^*) \\ -\mathrm{E}^0 \ell_{1,\alpha}(U_t; \alpha_1^*)^\top \ell_{2,\alpha}(U_t; \alpha_2^*) & \mathrm{E}^0 \ell_{1,\alpha}(U_t; \alpha_1^*)^\top \ell_{1,\alpha}(U_t; \alpha_1^*) \end{bmatrix} \begin{pmatrix} \hat{\alpha}_2 - \alpha_2^* \\ \hat{\alpha}_1 - \alpha_1^* \end{pmatrix} \tag{7}$$

This motivates us to bootstrap nV. Specifically, the bootstrap procedure consists of Step 1 and Step 2 in Section 4.2 and Steps 3' and 4' below.

Step 3': Compute $V^b =$

$$\begin{bmatrix} \hat{\alpha}_2^b - \hat{\alpha}_2 \\ \hat{\alpha}_1^b - \hat{\alpha}_1 \end{bmatrix}^\top \begin{bmatrix} \dfrac{1}{n}\displaystyle\sum_{t=1}^n \ell_{2,\alpha}(\tilde{U}_t; \hat{\alpha}_2)^\top \ell_{2,\alpha}(\tilde{U}_t; \hat{\alpha}_2) & \dfrac{-1}{n}\displaystyle\sum_{t=1}^n \ell_{2,\alpha}(\tilde{U}_t; \hat{\alpha}_2)^\top \ell_{1,\alpha}(\tilde{U}_t; \hat{\alpha}_1) \\ \dfrac{-1}{n}\displaystyle\sum_{t=1}^n \ell_{1,\alpha}(\tilde{U}_t; \hat{\alpha}_1)^\top \ell_{2,\alpha}(\tilde{U}_t; \hat{\alpha}_2) & \dfrac{1}{n}\displaystyle\sum_{t=1}^n \ell_{1,\alpha}(\tilde{U}_t; \hat{\alpha}_1)^\top \ell_{1,\alpha}(\tilde{U}_t; \hat{\alpha}_1) \end{bmatrix} \begin{bmatrix} \hat{\alpha}_2^b - \hat{\alpha}_2 \\ \hat{\alpha}_1^b - \hat{\alpha}_1 \end{bmatrix}.$$

Step 4': Repeat Steps 1, 2 and 3' a large number of times and use the empirical distribution of the resulting values of nV^b to approximate the null distribution of the test statistic $n\hat{\sigma}_a^2$.

THEOREM 6. *Under the conditions of Theorem 2,* $\mathrm{P}^*(nV^b \le x | X_1, \ldots, X_n)$ *converges in probability to* $M_{p_1+p_2}(x; \lambda^{*2})$.

6. EXTENSION TO MORE THAN TWO MODELS

In empirical applications of copulas, several parametric copulas are often used to fit the data and the results from models based on these copulas are then compared. The PLR tests developed in the previous sections can be extended to the comparison of more than two copulas along the lines of White (2000).

For $1 \le i \le M$, let $\{C_i(u_1, \ldots, u_d; \alpha_i) : \alpha_i \in \mathcal{A}_i \subset \mathbb{R}^{p_i}\}$ be a class of parametric copulas. Let $C_1(u_1, \ldots, u_d; \alpha_1)$ be the benchmark model. For pseudo true values $\alpha_1^*, \ldots, \alpha_M^*$, the null hypothesis is

$$H_0^M: \max_{2 \le i \le M} \mathrm{E}^0 \left[\log \left\{ \frac{c_i(U_{1t}, \ldots, U_{dt}; \alpha_i^*)}{c_1(U_{1t}, \ldots, U_{dt}; \alpha_1^*)} \right\} \right] \le 0,$$

meaning that no candidate copula model is closer to the true model than the benchmark model, and the alternative is

$$H_1^M: \max_{2\le i\le M} \mathrm{E}^0\left[\log\left\{\frac{c_i(U_{1t},\ldots,U_{dt};\alpha_i^*)}{c_1(U_{1t},\ldots,U_{dt};\alpha_1^*)}\right\}\right] > 0,$$

meaning that there exists a candidate copula model that is closer to the true model than the benchmark model.

For $2 \le i \le M$, define

$$LR_n\left(\widetilde{F}_1,\ldots,\widetilde{F}_d;\hat{\alpha}_i,\hat{\alpha}_1\right) = \frac{1}{n}\sum_{t=1}^n \log\left[\frac{c_i\{\widetilde{F}_1(X_{1t}),\ldots,\widetilde{F}_d(X_{dt});\hat{\alpha}_i\}}{c_1\{\widetilde{F}_1(X_{1t}),\ldots,\widetilde{F}_d(X_{dt});\hat{\alpha}_1\}}\right].$$

Also define $\Omega = (\sigma_{ik})_{i,k=2}^m$ in which

$$\begin{aligned}\sigma_{ik} &= \mathrm{cov}^0\Bigg[\{\ell_i(U_t;\alpha_i^*) - \ell_1(U_t;\alpha_1^*)\} + \sum_{j=1}^d \{Q_{i,j}(U_{jt};\alpha_i^*) - Q_{1,j}(U_{jt};\alpha_1^*)\},\\ &\qquad \{\ell_k(U_t;\alpha_k^*) - \ell_1(U_t;\alpha_1^*)\} + \sum_{j=1}^d \{Q_{k,j}(U_{jt};\alpha_k^*) - Q_{1,j}(U_{jt};\alpha_1^*)\}\Bigg],\end{aligned}$$

where $Q_{i,j}(U_{jt};\alpha_i^*)$ is defined in (4).

The following proposition follows immediately from the proof of Theorem 2 (a) and provides the basis for the subsequent tests.

PROPOSITION 4. *Suppose that Copula Models* $1,\ldots,M$ *satisfy conditions of Theorem 2. If* $\Omega = (\sigma_{ik})_{i,k=2}^m$ *is finite and its largest eigenvalue is positive, then*

$$\sqrt{n}\left[LR_n(\tilde{F}_1,\ldots,\tilde{F}_d;\hat{\alpha}_i,\hat{\alpha}_1) - \mathrm{E}^0\left[\log\left\{\frac{c_i(U_{1t},\ldots,U_{dt};\alpha_i^*)}{c_1(U_{1t},\ldots,U_{dt};\alpha_1^*)}\right\}\right]\right]_{i=2,\ldots,M} \to (Z_2,\ldots,Z_M)^\top$$

in distribution, where $(Z_2,\ldots,Z_M)^\top \sim N(0,\Omega)$.

6.1. The bootstrap test.

Proposition 4 and continuous mapping theorem imply that under the LFC of

$$H_0^M: \mathrm{E}^0\left[\log\left\{\frac{c_i(U_{1t},\ldots,U_{dt};\alpha_i^*)}{c_1(U_{1t},\ldots,U_{dt};\alpha_1^*)}\right\}\right] = 0, \quad \text{for } i = 2,\ldots,M$$

one has

$$\max_{2\le i\le M}\left\{\sqrt{n}\, LR_n\left(\tilde{F}_1,\ldots,\tilde{F}_d;\hat{\alpha}_i,\hat{\alpha}_1\right)\right\} \to \max_{2\le i\le M} Z_i \quad \text{in distribution.}$$

This could be used to construct the original reality check (RC) test of White (2000) for H_0^M: Reject H_0^M if

$$\max_{2\le i\le M}\left\{\sqrt{n}\, LR_n\left(\tilde{F}_1,\ldots,\tilde{F}_d;\hat{\alpha}_i,\hat{\alpha}_1\right)\right\} > \bar{z},$$

where $\bar{z}$ is the appropriate critical value of the distribution of $\max_{2\le i\le M} Z_i$. See White (2000) for details on the implementation of the RC test.

For power purposes, we propose the following modified test statistic:

$$T_{nI} = \max_{2\le i\le M}\left\{\frac{\sqrt{n}}{\sqrt{\hat{\sigma}_{ii}}}\, LR_n\left(\tilde{F}_1,\ldots,\tilde{F}_d;\hat{\alpha}_i,\hat{\alpha}_1\right) G_b(\hat{\sigma}_{ii}), 0\right\},$$

where $\hat{\sigma}_{ii}^2$ is the consistent estimator of σ_{ii}^2 defined in the same way as $\hat{\sigma}^2$ in Section 4.1, $b = b_n \to 0$ as $n \to \infty$ and G_b is a smoothed trimming function which trims out small $\hat{\sigma}_{ii}$.

We use the following smoothed trimming that has recently been used by Andrews (1995), Ai (1997) and Linton & Xiao (2001). Let g be a density function that has support $[0,1]$, $g(0) = g(1) = 0$, and let $g_b(x) = b^{-1} g(b^{-1}x - 1)$. Then $g_b(x)$ has support on $[b, 2b]$. Let

$$G_b(x) = 1(b \le x \le 2b) \int_{-\infty}^{x} g_b(z)dz + 1(x > 2b).$$

As an example we consider the Beta density: $g(z) = B(a+1)^{-1} z^a (1-z)^a$, $z \in [0,1]$ for some positive integer a, where $B(a) = \Gamma(a)^2/\Gamma(2a)$ is the beta function and $\Gamma(a)$ is the Euler gamma function. Then the resulting function $G_b(x)$ is $(a+1)$-times continuously differentiable on $[0,1]$; see Linton & Xiao (2001). We will suppose that $a \ge 1$.

THEOREM 7. *Suppose that Copula Models* $1, \ldots, M$ *satisfy the conditions of Proposition 4. If* $b \to 0$ *and* $nb \to \infty$*, then under the null hypothesis* H_0^M*, the limiting distribution of* T_{nI} *is given by that of* $\max_{i \in S_{NB}} \left(Z_i/\sqrt{\sigma_{ii}}\ , 0 \right)$*, where*

$$S_{NB} = \left\{ i \in \{2, \ldots, M\} :\ \sigma_{ii} > 0 \text{ and } \mathrm{E}^0 \left[\log c_i(U_{1t}, \ldots, U_{dt}; \alpha_i^*) \right] = \mathrm{E}^0 \left[\log c_1(U_{1t}, \ldots, U_{dt}; \alpha_1^*) \right] \right\}.$$

Theorem 7 implies that the asymptotic null distribution of T_{nI} depends on models that are generalized non-nested with the benchmark and satisfy

$$\mathrm{E}^0 \left\{ \log c_i(U_{1t}, \ldots, U_{dt}; \alpha_i^*) \right\} = \mathrm{E}^0 \left\{ \log c_1(U_{1t}, \ldots, U_{dt}; \alpha_1^*) \right\},$$

and hence is unknown. We propose the following bootstrap procedure to approximate it:

Step 1. Let $X_1^b, \ldots, X_n^b$ be a random sample with replacement from the original data $X_1, \ldots, X_n$ and let $\tilde{F}_j^b$ and $\hat{\alpha}_i^b$ be the bootstrapped counterparts of $\tilde{F}_j$ and $\hat{\alpha}_i$, $j = 1, \ldots, d$ and $i = 1, \ldots, M$.

Step 2. Compute the bootstrap value

$$T_{in}^b \equiv LR_n \left(\tilde{F}_1^b, \ldots, \tilde{F}_d^b; \hat{\alpha}_i^b, \hat{\alpha}_1^b \right)$$

of $T_{in} \equiv LR_n(\tilde{F}_1, \ldots, \tilde{F}_d; \hat{\alpha}_i, \hat{\alpha}_1)$, $i = 2, \ldots, M$, and define its recentered value as $T_{inC}^b = T_{in}^b - T_{in} I(T_{in} \ge -a_n)$, where $a_n \to 0$ is a small positive (possibly random) number such that $\sqrt{n} a_n \to \infty$.

Step 3. Compute the bootstrap value of T_{nI} as

$$T_{nI}^b = \max_{2 \le i \le M} \left\{ \frac{\sqrt{n}\, T_{inC}^b}{\sqrt{\hat{\sigma}_{ii}}}\ G_b(\hat{\sigma}_{ii}), 0 \right\};$$

Step 4. Repeat Steps 1–3 for a large number of times and use the empirical distribution function of the resulting values T_{nI}^b to approximate the null distribution of T_{nI}.

It is worthwhile to summarize the roles played by the two trimming functions: G_b in the definition of the test statistic T_{nI} and $I(T_{in} \ge -a_n)$ in Step 2 of the bootstrap procedure. The trimming function G_b removes the effect of candidate models that are generalized nested with the benchmark model on the null limiting distribution of T_{nI}, and the other trimming $I(T_{in} \ge -a_n)$ in Step 2 identifies among the class of candidate models satisfying $\sigma_{ii} > 0$ the ones that are not strictly dominated by the benchmark model under the null hypothesis.

This test is similar to the standardized test of Hansen (2003) with an important difference: our test allows some candidate models to be generalized nested with the benchmark model, since the trimming $G_b(\hat{\sigma}_{ii})$ in T_{nI} removes the effect of generalized nested models (with the benchmark model) on its limiting distribution.

THEOREM 8. *Under the conditions of Theorem 7, the conditional distribution of T_{nI}^b given the original data $X_1, \ldots, X_n$ converges in probability to the null limiting distribution of T_{nI}.*

Remark 6. To take into account dimensions of the parametric copulas, we modify the test statistic T_{nI} as follows:

$$T_{nI}^P = \max_{2\leq i\leq M} \left[\frac{\sqrt{n}}{\sqrt{\hat{\sigma}_{ii}}} \left\{ LR_n(\tilde{F}_1, \ldots, \tilde{F}_d; \hat{\alpha}_i, \hat{\alpha}_1) + (p_1 - p_i)/n \right\} G_b(\hat{\sigma}_{ii}), 0 \right].$$

It can be shown that the above bootstrap procedure still works provided that T_{in} and T_{in}^b are replaced with

$$T_{in}^P = LR_n\left(\tilde{F}_1, \ldots, \tilde{F}_d; \hat{\alpha}_i, \hat{\alpha}_1\right) + \frac{1}{n}(p_1 - p_i)$$

and

$$T_{in}^{bP} = LR_n\left(\tilde{F}_1^b, \ldots, \tilde{F}_d^b; \hat{\alpha}_i^b, \hat{\alpha}_1^b\right) + \frac{1}{n}(p_1 - p_i).$$

6.2. An application.

Here, we illustrate our testing procedure for selection of multiple copula models by using insurance company data on losses and ALAEs. The particular data set we use were collected by the US Insurance Services Office and have been analysed in some detail in Frees & Valdez (1998), Genest, Ghoudi & Rivest (1998), Klugman & Parsa (1999), and Denuit, Purcaru & Van Keilegom (2004). We refer readers to Frees & Valdez (1998) for a detailed description of this data set.

In Frees & Valdez (1998) and Klugman & Parsa (1999), joint parametric modeling of the loss and ALAE has been examined; using various model selection techniques including AIC/BIC, Frees & Valdez (1998) chose Pareto marginals and a Gumbel–Hougaard copula, while Klugman & Parsa (1999) chose inverse paralogistic for loss, inverse Burr for ALAE and Frank copula. Denuit, Purcaru & Van Keilegom (2004) adopted a semiparametric framework in which the marginal distributions of loss and ALAE are left unspecified, but their copula is modelled parametrically via an Archimedean copula. Their model selection procedure is similar to that in Wang & Wells (2000) except that the marginal distributions of loss and ALAE are estimated differently. They examined four Archimedean copulas, namely the Gumbel–Hougaard, Clayton, Frank and Joe copulas. They then chose the Gumbel–Hougaard copula, as Frees & Valdez (1998) and Genest, Ghoudi & Rivest (1998). In addition, they concluded that the limited amount of censored points present in the data does not seem to affect the selection result.

As opposed to Denuit, Purcaru & Van Keilegom (2004), we do not restrict the parametric copulas to be Archimedean. In addition, our test takes into account the randomness of the selection criterion. However, the current form of the test is applicable to uncensored data only. We thus restrict our analysis to the 1466 complete data. Extensions of the proposed tests in this paper to censored data are possible by using the two-step estimator proposed by Shih & Louis (1995). We will report them in a separate paper.

The scatterplots for loss and ALAE presented in Frees & Valdez (1998) and Denuit, Purcaru & Van Keilegom (2004) reveal positive right-tail dependence between loss and ALAE: large losses tend to be associated with large ALAE's. This is because expensive claims generally need some time to be settled and induce considerable costs for the insurance company. Actuaries therefore expect positive dependence between large losses and large ALAE's. On the other hand, these plots do not reveal any visible left-tail dependence between the two variables. As a result, it is not surprising that the Gumbel–Hougaard copula was chosen in Frees & Valdez (1998), Genest, Ghoudi & Rivest (1998), and Denuit, Purcaru & Van Keilegom (2004).

To shed some light on the robustness of this result to the set of copula families being considered, we added four more copula families to the set considered in Denuit, Purcaru & Van Keilegom (2004): survival Clayton, mixture of Clayton and Gumbel–Hougaard copula, Gaussian copula, and Student's t-copula. Survival Clayton has right-tail dependence and the mixture of Clayton and Gumbel–Hougaard exhibits both left-tail and right-tail dependence unless the weights are degenerate. Gaussian copula does not have tail dependence and is thus expected to fit poorly. Although Student's t-copula has both left-tail and right-tail dependence, but is symmetric and hence is not expected to fit well either. They are included here in the set of copulas to see if the power of the test is adversely affected by the presence of poor copula candidates in the selection set.

For each copula, we estimated its parameter(s) by the two-step procedure and computed the value of AIC. These are reported in Table 1 below. To apply our model selection test based on T_{nI}^{P}, we need to choose a benchmark model. In view of the existing results, we first used the Gumbel–Hougaard copula as the benchmark. Since the test depends on the two parameters a and b in the trimming function and the trimming parameter a_n, we experimented with various values of (a, b, a_n) and found the results not very sensitive to the values of these parameters. We thus fixed (a, b, a_n) at $(1, n^{-1/2}, 0.025n^{-0.5}\log\{\log(n)\})$. The number of bootstrap repetitions is 1000. For the Gumbel–Hougaard benchmark, we found the P-value of the test to be 1, which provides strong evidence that none of the other seven copulas performs significantly better than the Gumbel–Hougaard copula for the loss-ALAE data. This is consistent with the selection result based on comparing the values of AIC only; Gumbel–Hougaard followed by survival Clayton and then by mixture of Clayton and Gumbel–Hougaard. The parameter estimates for the mixture of Clayton and Gumbel–Hougaard provide additional evidence in favour of the Gumbel–Hougaard copula; the estimate of the weight on Clayton is only 0.0921 and the estimate (1.4606) of the parameter in the Gumbel–Hougaard copula is very close to that (1.4281) obtained from fitting the Gumbel–Hougaard copula alone.

Benchmark	P-value	AIC	Two-step estimate
Gumbel–Hougaard	1.0000	-0.2611	1.4281
Clayton	0.0000	-0.1227	0.5118
Frank	0.0050	-0.2199	0.0488
Joe	0.0030	-0.2403	1.6145
Gaussian	0.0030	-0.2336	0.4626
Student's t	0.0000	-0.2123	(0.2991,6.9914)
Survival Clayton	0.1260	-0.2597	0.749
Mix. Clayton and Gumbel–Hougaard	0.0083	-0.2517	(0.2772,1.4606,0.0921)

Table 1: P-values of the test

To see whether the test result is sensitive to the choice of the benchmark model, we also used each of the remaining seven copulas as the benchmark model in applying our test. The corresponding P-values are reported in Table 1. They overwhelmingly support the conclusion that the Gumbel-Hougaard copula fits the loss-ALAE data the best among the eight copulas we considered.

7. CONCLUSION

In this paper, we have provided a complete treatment of the issue of model selection for semiparametric copula models for uncensored i.i.d. data. In the case with more than two models, our test is established using a novel idea of trimming out the effect of generalized nested models on the test statistic asymptotically. This idea is applicable to other settings such as those in White (2000) and in many works following White's approach.

APPENDIX: TECHNICAL PROOFS

Define the rescaled empirical copula of $X_1, \ldots, X_n$ as:

$$\tilde{C}(u_1, \ldots, u_d) = \frac{1}{n+1} \sum_{t=1}^{n} 1\left\{\tilde{F}_1(X_{1t}) \leq u_1, \ldots, \tilde{F}_d(X_{dt}) \leq u_d\right\}.$$

Let $\tilde{C}^b(u_1, \ldots, u_d)$ be the bootstrap version of $\tilde{C}(u_1, \ldots, u_d)$. In the following we denote G_{C^0} as a Gaussian process in $\ell^\infty([0,1]^d)$ defined as

$$G_{C^0}(u_1, \ldots, u_d) = B_{C^0}(u_1, \ldots, u_d) - \sum_{j=1}^{d} \frac{\partial C^0(u_1, \ldots, u_d)}{\partial u_j} B_{C^0}(1, \ldots, 1, u_j, 1, \ldots, 1),$$

in which B_{C^0} is a Brownian bridge on $[0,1]^d$ with covariance function

$$\mathrm{E}\left\{B_{C^0}(u_1, \ldots, u_d) B_{C^0}(u_1^\top, \ldots, u_d^\top)\right\} = C^0(u_1 \wedge u_1^\top, \ldots, u_d \wedge u_d^\top) - C^0(u_1, \ldots, u_d) C^0(u_1^\top, \ldots, u_d^\top)$$

for each $0 \leq u_1, \ldots, u_d, u_1^\top, \ldots, u_d^\top \leq 1$.

Lemma 1. *Under conditions C2 and C3, we have:*

(a) *The rescaled empirical copula process $\{\sqrt{n}\{\tilde{C}(u) - C^0(u)\} : u \in [0,1]^d\}$ converges weakly to the Gaussian process $\{G_{C^0}(u) : u \in [0,1]^d\}$ in $\ell^\infty([0,1]^d)$.*

(b) *The conditional distribution of $\{\sqrt{n}\{\tilde{C}^b(u) - \widetilde{C}(u)\} : u \in [0,1]^d\}$ converges to the same limiting Gaussian process as $\{\sqrt{n}\{\tilde{C}(u) - C^0(u)\} : u \in [0,1]^d\}$ in $\ell^\infty([0,1]^d)$ in probability.*

(c) *Let $\mathcal{H}$ be a class of functions $h : [0,1]^d \to (-\infty, \infty)$ which satisfies: for every $\delta > 0$, $N_{[]}(\delta, \mathcal{H}, L_1(C^0)) < \infty$, where $N_{[]}(\delta, \mathcal{H}, L_1(C^0))$ is the $L_1(C^0)$ bracketing number of the class $\mathcal{H}$. Then:*

$$\sup_{h \in \mathcal{H}} \left| \int_{u \in [0,1]^d} h(u) d\{\tilde{C}(u) - C^0(u)\} \right| \to 0 \textit{ almost surely.}$$

Proof of Lemma 1: Parts (a) and (b) are respectively the extensions of Theorem 3 and Theorem 5 in Fermanian, Radulović & Wegkamp (2004) for the bivariate empirical copula process to the multivariate rescaled empirical copula process. By examining the proofs of their Theorems 3 and 5, one can easily show that under conditions C2 and C3, their theorems still hold for the rescaled empirical copula process $\sqrt{n}(\tilde{C} - C^0)$ and for the bootstrapped rescaled empirical copula process $\sqrt{n}(\tilde{C}^b - \widetilde{C})$.

As for part (c), it is the extension of Theorem 17 in the working paper version of Fermanian, Radulović & Wegkamp (2004) to the rescaled empirical copula process, and can be established in the same way.

We note that a slightly different version of Lemma 1 (a), i.e., $\{\sqrt{n}\{\tilde{C}(u) - C^0(u)\} : u \in [0,1]^d\}$ converges weakly to a continuous Gaussian process $\{G_{C^0}(u) : u \in [0,1]^d\}$ in $D([0,1]^d)$, can also be proved by following the proof of Gänssler & Stute (1987) and Ghoudi & Rémillard (2004).

Proof of Proposition 1: (1) We rewrite the definition of $\hat{\alpha}$ in terms of $\tilde{C}(u_1, \ldots, u_d)$:

$$\hat{\alpha} = \text{argmax}_{\alpha \in \mathcal{A}} \left\{ \int_0^1 \cdots \int_0^1 \ell(u_1, \ldots, u_d; \alpha) d\tilde{C}(u_1, \ldots, u_d) \right\}.$$

Let

$$\tilde{Q}(\alpha) = \int_0^1 \cdots \int_0^1 \ell(u_1, \ldots, u_d; \alpha) d\tilde{C}(u_1, \ldots, u_d)$$

and

$$Q^0(\alpha) = \int_0^1 \cdots \int_0^1 \ell(u_1, \ldots, u_d; \alpha) dC^0(u_1, \ldots, u_d).$$

By Example 19.8 in van der Vaart (1998), condition C4 implies $N_{[]}(\delta, \mathcal{H}, L_1(C^0)) < \infty$ for the class of functions $\mathcal{H} = \{\ell(u_1, \ldots, u_d; \alpha) : \alpha \in \mathcal{A}\}$. Hence under conditions C2–C4, Lemma 1 (c) applies and we obtain: $\sup_{\alpha \in \mathcal{A}} |\tilde{Q}(\alpha) - Q^0(\alpha)| \to 0$ almost surely. By Theorem 3.4 in White (1994), it follows that under conditions C1–C4, $\hat{\alpha}$ converges *a.s.* to α^*.

(2) Noting that

$$\frac{1}{n} \sum_{t=1}^{n} \ell \left\{ \tilde{F}_1(X_{1t}), \ldots, \tilde{F}_d(X_{dt}); \hat{\alpha} \right\} = \frac{n+1}{n} \tilde{Q}(\hat{\alpha}),$$

the result follows immediately from Proposition 1 (a) and the uniform almost sure convergence of $\tilde{Q}(\alpha)$.

LEMMA 2. *Suppose condition C2 holds. For $j = 1, \ldots, d$, let w_j be a continuous function on $[0, 1]$, positive on $(0, 1)$, symmetric at $1/2$, increasing on $(0, 1/2]$ and such that*

$$\int_0^1 \left\{ \frac{1}{w_j(v)} \right\}^2 dv < \infty.$$

Let $J : [0, 1]^d \to (-\infty, \infty)$ be a continuous function and have continuous partial derivatives $J_j(u) = \partial J(u)/\partial u_j$ on $(0, 1)^d$ for $j = 1, \ldots, d$. Suppose:

(i) $|J(u)| \leq \text{constant} \times \prod_{j=1}^{d} \{u_j(1 - u_j)\}^{-a_j}$ *for some $a_j \geq 0$ such that*

$$\mathrm{E}^0 \left[\prod_{j=1}^{d} \{U_{jt}(1 - U_{jt})\}^{-2a_j} \right] < \infty;$$

(ii) $|J_k(u)| \leq \text{constant} \times \{u_k(1 - u_k)\}^{-b_k} \prod_{j=1, j \neq k}^{d} \{u_j(1 - u_j)\}^{-a_j}$ *for some $b_k > a_k$ such that*

$$\mathrm{E}^0 \left[w_k(U_{kt}) \{U_{kt}(1 - U_{kt})\}^{-b_k} \prod_{j=1, j \neq k}^{d} \{U_{jt}(1 - U_{jt})\}^{-a_j} \right] < \infty$$

for $k = 1, \ldots, d$.

Then

$$\sqrt{n} \int_{u \in [0,1]^d} J(u) d\left\{ \tilde{C}(u) - C^0(u) \right\} \to \mathcal{N}(0, \sigma_J^2) \text{ in distribution,}$$

where

$$\sigma_J^2 = \text{var}^0 \left\{ J(U_{1t}, \ldots, U_{dt}) + \sum_{j=1}^{d} \int_{u \in [0,1]^d} J_j(u) \times I(U_{jt} \leq u_j) dC^0(u) \right\}.$$

Proof of Lemma 2: The proof mimics that of Theorem 2.1 in Ruymgaart (1974). See also the proofs of Theorem 2.1 in Ruymgaart, Shorack & Zwet (1972) or Proposition A.1 (ii) in Genest, Ghoudi & Rivest (1995).

We note that when J is a function of bounded variation, Lemma 2 becomes Theorem 6 in Fermanian, Radulović & Wegkamp (2004). However, the bounded variation assumption is violated by many popular copula functions with unbounded copula score functions.

Proof of Proposition 2: Since $||\widehat{\alpha} - \alpha^*|| = o_p(1)$ by Proposition 1, under conditions A1(i) and A4(i), we have by first order condition and Taylor expansion,

$$\begin{aligned} 0 &= \frac{1}{n} \sum_{t=1}^{n} \ell_\alpha \left\{ \tilde{F}_1(X_{1t}), \ldots, \tilde{F}_d(X_{dt}); \hat{\alpha} \right\} \\ &= \frac{1}{n} \sum_{t=1}^{n} \ell_\alpha \left\{ \tilde{F}_1(X_{1t}), \ldots, \tilde{F}_d(X_{dt}); \alpha^* \right\} + \frac{1}{n} \sum_{t=1}^{n} \ell_{\alpha\alpha} \left\{ \tilde{F}_1(X_{1t}), \ldots, \tilde{F}_d(X_{dt}); \overline{\alpha} \right\} (\widehat{\alpha} - \alpha^*), \end{aligned}$$

where $\bar{\alpha}$ is between α^* and $\hat{\alpha}$. Since

$$\begin{aligned} &\frac{1}{n} \sum_{t=1}^{n} \ell_{\alpha\alpha} \left\{ \tilde{F}_1(X_{1t}), \ldots, \tilde{F}_d(X_{dt}); \alpha \right\} - \mathrm{E}^0 \left\{ \ell_{\alpha\alpha}(U_{1t}, \ldots, U_{dt}; \alpha) \right\} \\ = &\frac{n+1}{n} \int_{u \in [0,1]^d} \ell_{\alpha\alpha}(u; \alpha) d \left\{ \tilde{C}(u) - C^0(u) \right\}, \end{aligned}$$

we have, under conditions C2–C3, A4 and Lemma 1 (c) with

$$\mathcal{H} = \left\{ \ell_{\alpha\alpha}(u; \alpha) : \alpha \in \mathcal{A} : ||\alpha - \alpha^*|| = o(1) \right\},$$

that

$$\sup_{\alpha \in \mathcal{A}: ||\alpha - \alpha^*|| = o(1)} || \frac{1}{n} \sum_{t=1}^{n} \ell_{\alpha\alpha} \left\{ \tilde{F}_1(X_{1t}), \ldots, \tilde{F}_d(X_{dt}); \alpha \right\} - \mathrm{E}^0 \left\{ \ell_{\alpha\alpha}(U_{1t}, \ldots, U_{dt}; \alpha^*) \right\} || = o_p(1).$$

By conditions A1(i), C2, A1(iii), A2–A3 and Lemma 2 with $J(u) = l_\alpha(u; \alpha^*)$, we have that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \ell_\alpha \{ \tilde{F}_1(X_{1t}), \ldots, \tilde{F}_d(X_{dt}); \alpha^* \} = \left(1 + \frac{1}{n} \right) \sqrt{n} \int_{u \in [0,1]^d} \ell_\alpha(u; \alpha^*) d\{ \tilde{C}(u) - C^0(u) \} \qquad (8)$$

converges in distribution to a $\mathcal{N}(0, \Sigma)$. Hence under condition A1(ii)(iii), $B(\widehat{\alpha} - \alpha^*) + o_p\left(||\widehat{\alpha} - \alpha^*||\right) = O_p(n^{-1/2})$ and $\sqrt{n}(\widehat{\alpha} - \alpha^*) \to \mathcal{N}(0, B^{-1}\Sigma B^{-1})$ in distribution.

Proof of Theorems 1 and 2: Let

$$\ell_i \left\{ \widetilde{F}_1(X_{1t}), \ldots, \widetilde{F}_d(X_{dt}); \hat{\alpha}_i \right\} = \log c_i \left\{ \widetilde{F}_1(X_{1t}), \ldots, \widetilde{F}_d(X_{dt}); \hat{\alpha}_i \right\}, \quad i = 1, 2.$$

Then by the definition of $\widehat{\alpha}_i$, we have

$$\sum_{t=1}^{n} \ell_{i,\alpha} \left\{ \widetilde{F}_1(X_{1t}), \ldots, \widetilde{F}_d(X_{dt}); \hat{\alpha}_i \right\} = 0.$$

Hence,

$$\begin{aligned}\sum_{t=1}^{n} \ell_i \left\{ \widetilde{F}_1(X_{1t}), \ldots, \widetilde{F}_d(X_{dt}); \alpha_i^* \right\} &= \sum_{t=1}^{n} \ell_i \left\{ \widetilde{F}_1(X_{1t}), \ldots, \widetilde{F}_d(X_{dt}); \hat{\alpha}_i \right\} \\ &+ \frac{1}{2} (\alpha_i^* - \hat{\alpha}_i)^\top \sum_{t=1}^{n} \ell_{i,\alpha\alpha} \left\{ \widetilde{F}_1(X_{1t}), \ldots, \widetilde{F}_d(X_{dt}); \bar{\alpha}_i \right\} (\alpha_i^* - \hat{\alpha}_i),\end{aligned}$$

where $\bar{\alpha}_i$ is between α_i^* and $\hat{\alpha}_i$. By conditions C2–C4, A1–A4, Proposition 2 and Lemma 1 (c), we have

$$\begin{aligned}&\frac{1}{2n} (\alpha_i^* - \hat{\alpha}_i)^\top \sum_{t=1}^{n} \ell_{i,\alpha\alpha} \left\{ \widetilde{F}_1(X_{1t}), \ldots, \widetilde{F}_d(X_{dt}); \bar{\alpha}_i \right\} (\alpha_i^* - \hat{\alpha}_i) \\ &\qquad = -\frac{1}{2} (\alpha_i^* - \hat{\alpha}_i)^\top B_i(\alpha_i^* - \hat{\alpha}_i) + o_p(1/n).\end{aligned}$$

Hence,

$$\begin{aligned}&\frac{1}{n} \sum_{t=1}^{n} \ell_i \left\{ \widetilde{F}_1(X_{1t}), \ldots, \widetilde{F}_d(X_{dt}); \hat{\alpha}_i \right\} \\ &= \frac{1}{n} \sum_{t=1}^{n} \ell_i \left\{ \widetilde{F}_1(X_{1t}), \ldots, \widetilde{F}_d(X_{dt}); \alpha_i^* \right\} + \frac{1}{2} (\alpha_i^* - \hat{\alpha}_i)^\top B_i(\alpha_i^* - \hat{\alpha}_i) + o_p(1/n).\end{aligned}$$

As a result, we get

$$LR_n \left(\tilde{F}_1, \ldots, \tilde{F}_d; \hat{\alpha}_2, \hat{\alpha}_1 \right) - \mathrm{E}^0 \left\{ LR_n \left(F_1^0, \ldots, F_d^0; \alpha_2^*, \alpha_1^* \right) \right\} = D1 + D_n + o_p(1/n),$$

where

$$\begin{aligned}D1 &\equiv LR_n(\tilde{F}_1, \ldots, \tilde{F}_d; \alpha_2^*, \alpha_1^*) - \mathrm{E}^0 \{ LR_n(F_1^0, \ldots, F_d^0; \alpha_2^*, \alpha_1^*) \}, \\ D_n &\equiv \frac{1}{2}(\hat{\alpha}_2 - \alpha_2^*)^\top B_2(\hat{\alpha}_2 - \alpha_2^*) - \frac{1}{2}(\hat{\alpha}_1 - \alpha_1^*)^\top B_1(\hat{\alpha}_1 - \alpha_1^*).\end{aligned}$$

By Proposition 2, we have $D_n = O_p(n^{-1})$.

For generalized non-nested models, by conditions C2, A2–A3 and Lemma 2 with $J(u) = \ell_2(u; \alpha_2^*) - \ell_1(u; \alpha_1^*)$, we have

$$\sqrt{n} \times D1 = \left(1 + \frac{1}{n} \right) \times \sqrt{n} \int_{u \in [0,1]^d} \{ \ell_2(u; \alpha_2^*) - \ell_1(u; \alpha_1^*) \} d \left\{ \tilde{C}(u) - C^0(u) \right\}$$

converges in distribution to a $\mathcal{N}(0, \sigma^2)$. Hence the term $D1$ is of the order $O_p(n^{-1/2})$, and

$$LR_n \left(\tilde{F}_1, \ldots, \tilde{F}_d; \hat{\alpha}_2, \hat{\alpha}_1 \right) - \mathrm{E}^0 \left\{ LR_n \left(F_1^0, \ldots, F_d^0; \alpha_2^*, \alpha_1^* \right) \right\} = D1 + o_p \left(1/\sqrt{n} \right).$$

Therefore,

$$\sqrt{n} \left[LR_n \left(\tilde{F}_1, \ldots, \tilde{F}_d; \hat{\alpha}_2, \hat{\alpha}_1 \right) - \mathrm{E}^0 \left[\log \left\{ \frac{c_2(U_{1t}, \ldots, U_{dt}; \alpha_2^*)}{c_1(U_{1t}, \ldots, U_{dt}; \alpha_1^*)} \right\} \right] \right]$$

converges in distribution to a $\mathcal{N}(0,\sigma^2)$.

For generalized nested models, the term $D1$ becomes zero almost surely, we have

$$LR_n\left(\tilde{F}_1,\ldots,\tilde{F}_d;\hat{\alpha}_2,\hat{\alpha}_1\right) - \mathrm{E}^0\left\{LR_n\left(F_1^0,\ldots,F_d^0;\alpha_2^*,\alpha_1^*\right)\right\} = D_n + o_p(1/n),$$

where by Proposition 2, $2nD_n$ is distributed as a weighted sum of independent $\chi^2_{[1]}$ random variables.

Proof of Proposition 3: It is obviously true that the generalized nested copula models imply $\sigma^2 = 0$ and $\sigma_a^2 = 0$. It remains to show that $\sigma^2 = 0$ implies $\sigma_a^2 = 0$, which in turn implies that the two copula models are generalized nested.

Part (a): Define

$$a^* = \mathrm{E}^0\left[\log\left\{\frac{c_2(U_{1t},\ldots,U_{dt};\alpha_2^*)}{c_1(U_{1t},\ldots,U_{dt};\alpha_1^*)}\right\}\right] = \int_{[0,1]^d}\log\left\{\frac{c_2(u_1,\ldots,u_d;\alpha_2^*)}{c_1(u_1,\ldots,u_d;\alpha_1^*)}\right\}dC^0(u_1,\ldots,u_d).$$

Obviously a^* is a smooth functional of the true unknown copula function $C^0(u_1,\ldots,u_d)$. If we could observe an i.i.d. random sample $\{U_t = (U_{1t},\ldots,U_{dt})\}_{t=1}^n$ where U_t is distributed according to $C^0(u_1,\ldots,u_d)$, then an efficient estimator of a^* will simply be

$$\hat{a}_n = \int_{[0,1]^d}\log\left\{\frac{c_2(u_1,\ldots,u_d;\alpha_2^*)}{c_1(u_1,\ldots,u_d;\alpha_1^*)}\right\}dC_n(u_1,\ldots,u_d) = \frac{1}{n}\sum_{t=1}^n\log\left\{\frac{c_2(U_{1t},\ldots,U_{dt};\alpha_2^*)}{c_1(U_{1t},\ldots,U_{dt};\alpha_1^*)}\right\},$$

where $C_n(u_1,\ldots,u_d)$ is the empirical distribution estimator of the true copula function $C^0(u_1,\ldots,u_d)$. It is obvious that $\sqrt{n}(\hat{a}_n - a^*) \to \mathcal{N}(0,\sigma_a^2)$.

Now in fact we do not observe $\{U_t = (U_{1t},\ldots,U_{dt})\}_{t=1}^n$, but only observe an i.i.d. random sample $\{X_t = (X_{1t},\ldots,X_{dt})\}_{t=1}^n$ where X_t is distributed according to the unknown true multivariate distribution function $F^0(x) = C^0\{F_1^0(x_1),\ldots,F_d^0(x_d)\}$, with F_j^0 being the unknown marginal distribution of X_{1t}. Let

$$\widetilde{a} = \frac{1}{n}\sum_{t=1}^n\log\left[\frac{c_2\{\tilde{F}_1(X_{1t}),\ldots,\tilde{F}_d(X_{dt});\alpha_2^*\}}{c_1\{\tilde{F}_1(X_{1t}),\ldots,\tilde{F}_d(X_{dt});\alpha_1^*\}}\right].$$

Clearly it satisfies $\sqrt{n}(\widetilde{a} - a^*) \to \mathcal{N}(0,\sigma^2)$. Hence $\sigma^2 \geq \sigma_a^2$. Since $\sigma_a^2 \geq 0$, it must be true that $\sigma^2 = 0$ implies $\sigma_a^2 = 0$.

Part (b): Next, $\sigma_a^2 = 0$ implies $\log\{c_2(U_{1t},\ldots,U_{dt};\alpha_2^*)/c_1(U_{1t},\ldots,U_{dt};\alpha_1^*)\} = K$ almost surely for some constant K. As both c_1 and c_2 are density functions, we have $K = 0$, hence both copula models are generalized nested.

Proof of Theorems 4 and 6: For both theorems, it suffices to show that for $i = 1,2$, the conditional distribution of $\sqrt{n}\,(\hat{\alpha}_i^b - \hat{\alpha}_i)$ given the original data converges in probability to $\mathcal{N}(0, B_i^{-1}\Sigma_i B_i^{-1})$. For notational convenience, we'll ignore the $i = 1,2$ subscript. Let

$$\tilde{Q}^b(\alpha) = \int_0^1\cdots\int_0^1 \ell(u_1,\ldots,u_d;\alpha)d\tilde{C}^b(u_1,\ldots,u_d)$$

and $\hat{\alpha}^b = \arg\max_{\alpha\in\mathcal{A}}\tilde{Q}^b(\alpha)$. We can apply Lemma A.3 in Goncalves & White (2004) to show that the conditional distribution of $\sqrt{n}\,(\hat{\alpha}^b - \hat{\alpha})$ converges in probability to the asymptotic distribution of $\sqrt{n}\,(\hat{\alpha} - \alpha^*)$. The following proof mimics that of Proposition 2.

First, we show that $\hat{\alpha}^b - \hat{\alpha} = o_{p^*}(1)$ prob-P. Using Lemma A.2 in Goncalves & White (2004), it suffices to show that $\sup_{\alpha \in \mathcal{A}} |\tilde{Q}^b(\alpha) - \tilde{Q}(\alpha)| = o_{p^*}(1)$ prob-P, which directly follows from conditions C2–C4 and our Lemma 1 (b,c).

Next, we verify conditions (b3) and (b4) of Lemma A.3 in Goncalves & White (2004), which are: (b3) The conditional distribution of $\sqrt{n}\,\tilde{Q}^b_\alpha(\hat{\alpha})$ converges to $\mathcal{N}(0,\Sigma)$ in probability; (b4) $\sup_{\alpha \in \mathcal{A}: ||\alpha-\alpha^*||=o(1)} ||\tilde{Q}^b_{\alpha\alpha}(\alpha) - \tilde{Q}_{\alpha\alpha}(\alpha)|| = o_{p^*}(1)$ prob.-P.

For (b3), since $\tilde{Q}_\alpha(\hat{\alpha}) = 0$ we have

$$\sqrt{n}\,\tilde{Q}^b_\alpha(\hat{\alpha}) = \sqrt{n}\int_{[0,1]^d} \ell_\alpha(u;\hat{\alpha})d\left(\tilde{C}^b - \tilde{C}\right)(u).$$

We now apply Theorem 3.9.11 in van der Vaart & Wellner (1996, p. 378). Note that our Lemma 1 (a,b) imply that their condition (3.9.9) is satisfied with their $\widehat{P}_n - P_n$ being our $\tilde{C}^b - \tilde{C}$ and their G being our G_{C^0}. Assumption A3(i) implies that the mapping $\int_{[0,1]^d} \ell_\alpha(u;\alpha^*)dC(u)$ is Hadamard-differentiable at C^0. By Lemma 2 with $J(u) = \ell_\alpha(u;\alpha^*)$, the relation (8), and Theorem 3.9.11 in van der Vaart & Wellner (1996), we obtain

$$\sqrt{n}\int_{[0,1]^d} \ell_\alpha(u;\alpha^*)d\left(\tilde{C}^b - \tilde{C}\right)(u) \to \mathcal{N}(0,\Sigma) \text{ in distribution.}$$

Under conditions C2–C3, A4 and our Lemma 1 (b), we have

$$\sup_{\alpha \in \mathcal{A}: ||\alpha-\alpha^*||=o(1)} ||\sqrt{n}\left[\int_{[0,1]^d} \ell_\alpha(u;\alpha)d\left(\tilde{C}^b - \tilde{C}\right)(u) - \int_{[0,1]^d} \ell_\alpha(u;\alpha^*)d\left(\tilde{C}^b - \tilde{C}\right)(u)\right]||$$

is $o_{p^*}(1)$ prob.-P. These and $\hat{\alpha}^b - \hat{\alpha} = o_{p^*}(1)$ prob-P. imply (b3).

As for (b4), it follows from conditions C2–C3, A4 and our Lemma 1 (c).

Proof of Theorem 5. (a) is trivially true. To prove (b), observe that by Theorem 2 (b), Proposition 3, Taylor expansion and the fact that C_1 and C_2 are generalized nested models, we get

$$n\hat{\sigma}_a^2 = \sum_{t=1}^n \left\{\ell_{2,\alpha}(\tilde{U}_t;\alpha_2^*)(\hat{\alpha}_2 - \alpha_2^*) - \ell_{1,\alpha}(\tilde{U}_t;\alpha_1^*)(\hat{\alpha}_1 - \alpha_1^*)\right\}^2 + o_p(1) = nV + o_p(1),$$

where V is given in (7). The result follows by mimicking the proof of Theorem 2 (b).

Proof of Theorem 7: The proof consists of two steps: Under the null hypothesis, we first show that (i) $T_{nI} = T_{nM'} + o_p(1)$, where

$$T_{nM'} = \max_{i \in S_N}\left\{\frac{\sqrt{n}}{\sqrt{\hat{\sigma}_{ii}}} LR_n\left(\tilde{F}_1,\ldots,\tilde{F}_d;\hat{\alpha}_i,\hat{\alpha}_1\right), 0\right\},$$

in which $S_N = \{i \in \{2,\ldots,M\} : \sigma_{ii} > 0\}$ and M' is the number of candidate models in S_N. Then we show that (ii) the null limiting distribution of $T_{nM'}$ is given by that of $\max_{i \in S_{NB}}(Z_i/\sqrt{\sigma_{ii}}, 0)$.

To prove (i), note that by Taylor expansion,

$$\frac{\sqrt{n}}{\sqrt{\hat{\sigma}_{ii}}} LR_n\left(\widetilde{F}_1,\ldots,\widetilde{F}_d;\hat{\alpha}_i,\hat{\alpha}_1\right) G_b(\hat{\sigma}_{ii}) = \frac{\sqrt{n}}{\sqrt{\hat{\sigma}_{ii}}} LR_n\left(\widetilde{F}_1,\ldots,\widetilde{F}_d;\hat{\alpha}_i,\hat{\alpha}_1\right) G^*_{bi},$$

where σ^*_{ii} lies between σ_{ii} and $\hat{\sigma}_{ii}$, and

$$G^*_{bi} \equiv G_b(\sigma_{ii}) + G'_b(\sigma_{ii})(\hat{\sigma}_{ii} - \sigma_{ii}) + \frac{1}{2}G''_b(\sigma^*_{ii})(\hat{\sigma}_{ii} - \sigma_{ii})^2.$$

We consider two cases: (1) the model i and the benchmark model are generalized nested, and (2) the model i and the benchmark model are generalized non-nested. In case (1), Proposition 3 implies that $\sigma_{ii} = 0$. Hence

$$G_{bi}^* = \frac{1}{2}\, G_b''(\sigma_{ii}^*)(\hat{\sigma}_{ii} - \sigma_{ii})^2 = \begin{cases} 0, & \text{if } \sigma_{ii}^* < b \text{ or } \sigma_{ii}^* > 2b \\ \frac{1}{2b^2} g'(\sigma_{ii}^*/b - 1)(\hat{\sigma}_{ii} - \sigma_{ii})^2, & \text{if } b \le \sigma_{ii}^* \le 2b. \end{cases}$$

Similar to Theorem 5 (b), one can show that in case (1), $\hat{\sigma}_{ii} = O_p(n^{-1})$. Since $nb \to \infty$ we obtain $G_{bi}^* = o_p(1)$. In case (2), since $\sigma_{ii} > 0$ and $b \to 0$, similar to Theorem 5 (a), we have for large enough n, $G_{bi}^* = G_b(\sigma_{ii}) + o_p(1) = 1 + o_p(1)$. In summary,

$$\frac{\sqrt{n}}{\sqrt{\hat{\sigma}_{ii}}}\, LR_n\left(\widetilde{F}_1, \ldots, \widetilde{F}_d; \hat{\alpha}_i, \hat{\alpha}_1\right) G_b(\hat{\sigma}_{ii}) = \frac{\sqrt{n}}{\sqrt{\hat{\sigma}_{ii}}}\, LR_n\left(\widetilde{F}_1, \ldots, \widetilde{F}_d; \hat{\alpha}_i, \hat{\alpha}_1\right)\{1 + o_p(1)\}$$

if models i and 1 are not nested, and is $o_p(1)$ otherwise.

As for part (ii), it follows from our Proposition 4, Theorem 7, and Theorem 1 of Hansen (2003).

Proof of Theorem 8: By following the proof of Theorem 7, one can show that the conditional limiting distribution of T_{nI}^b given the data is the same as that of $T_{nM'}^b$, where $T_{nM'}^b$ is the bootstrap value of $T_{nM'}$:

$$T_{nM'}^b = \max_{i \in S_N}\left[\frac{\sqrt{n}}{\sqrt{\hat{\sigma}_{ii}}}\{T_{in}^b - T_{in} I(T_{in} \ge -a_n)\}, 0\right].$$

So it remains to show that the conditional distribution of $T_{nM'}^b$ given the data converges in probability to the null limiting distribution of $T_{nM'}$, which is the distribution of $\max_{i \in S_{NB}}(Z_i/\sqrt{\sigma_{ii}}, 0)$.

To show that the bootstrap works for $T_{nM'}$, it suffices to show that

$$\sup_{z \in R^{M-1}}\left| \mathrm{P}^*\left(\begin{bmatrix} \sqrt{n}\,(T_{2n}^b - T_{2n}) \\ \vdots \\ \sqrt{n}\,(T_{Mn}^b - T_{Mn}) \end{bmatrix} \le z\right) - \mathrm{P}\left(\begin{bmatrix} \sqrt{n}\,(T_{2n} - \mathrm{E}^0 \log \frac{c_2(U_{1t},\ldots,U_{dt};\alpha_2^*)}{c_1(U_{1t},\ldots,U_{dt};\alpha_1^*)}) \\ \vdots \\ \sqrt{n}\,(T_{Mn} - \mathrm{E}^0 \log \frac{c_M(U_{1t},\ldots,U_{dt};\alpha_M^*)}{c_1(U_{1t},\ldots,U_{dt};\alpha_1^*)}) \end{bmatrix} \le z\right)\right|$$

is $o_p(1)$.

Note that one can write T_{in} and T_{in}^b as

$$\begin{aligned} T_{in} &= \frac{n+1}{n}\int_0^1 \cdots \int_0^1 \{\ell_i(u_1, \ldots, u_d; \hat{\alpha}_i) - \ell_1(u_1, \ldots, u_d; \hat{\alpha}_1)\} d\tilde{C}(u_1, \ldots, u_d), \\ T_{in}^b &= \frac{n+1}{n}\int_0^1 \cdots \int_0^1 \{\ell_i(u_1, \ldots, u_d; \hat{\alpha}_i^b) - \ell_1(u_1, \ldots, u_d; \hat{\alpha}_1^b)\} d\tilde{C}^b(u_1, \ldots, u_d). \end{aligned}$$

For $i = 1, \ldots, d$, by the definition of $\widehat{\alpha}_i^b$, we have

$$\frac{n+1}{n}\int \ell_{i,\alpha}\left(u; \hat{\alpha}_i^b\right) d\tilde{C}^b(u) = 0.$$

Hence,

$$\begin{aligned} &\frac{n+1}{n}\int \ell_i(u; \hat{\alpha}_i) d\tilde{C}^b(u) \\ &= \frac{n+1}{n}\int \ell_i\left(u; \hat{\alpha}_i^b\right) d\tilde{C}^b(u) + \frac{1}{2}\left(\hat{\alpha}_i - \hat{\alpha}_i^b\right)^\top \frac{n+1}{n}\int \ell_{i,\alpha\alpha}(u; \bar{\alpha}_i) d\tilde{C}^b(u)\left(\hat{\alpha}_i - \hat{\alpha}_i^b\right), \end{aligned}$$

where $\bar{\alpha}_i$ is between $\widehat{\alpha}_i^b$ and $\hat{\alpha}_i$. Following the proofs of Theorems 2 and 4, we have for $i = 1, \ldots, d$:

$$\frac{n+1}{n} \int \ell_i \left(u; \hat{\alpha}_i^b\right) d\tilde{C}^b(u) - \frac{n+1}{n} \int \ell_i \left(u; \hat{\alpha}_i\right) d\tilde{C}^b(u) = O_{p^*}(1/n).$$

Hence,

$$T_{in}^b - \frac{n+1}{n} \int \{\ell_i(u; \hat{\alpha}_i) - \ell_1(u; \hat{\alpha}_1)\} d\tilde{C}^b(u) = O_{p^*}(1/n),$$

and

$$T_{in}^b - T_{in} = \frac{n+1}{n} \int \{\ell_i(u; \hat{\alpha}_i) - \ell_1(u; \hat{\alpha}_1)\} d\left\{\tilde{C}^b(u) - \tilde{C}(u)\right\} + O_{p^*}(1/n).$$

Again by applying our Proposition 4, Lemma 1 (a, b) and Lemma 2, Theorem 3.9.11 in van der Vaart & Wellner (1996), one can show that

$$\sqrt{n} \left[\int \{\ell_i(u; \alpha_i^*) - \ell_1(u; \alpha_1^*)\} d\{\tilde{C}^b - \tilde{C}\}(u)\right]_{i=2,\ldots,M} \to (Z_2, \ldots, Z_M)^\top \text{ in distribution.}$$

Under conditions C2–C3, A4, our Lemma 1 (b) and following the proof of Theorem 4, we conclude that

$$\sup_{\alpha_i \in \mathcal{A}_i : ||\alpha_i - \alpha_i^*|| = o(1)} ||\sqrt{n} \left[\int \ell_i(u; \alpha_i) d\{\tilde{C}^b - \tilde{C}\}(u) - \int \ell_i(u; \alpha_i^*) d\{\tilde{C}^b - \tilde{C}\}(u)\right]||$$

is $o_{p^*}(1)$ prob.-P, and the result follows.

ACKNOWLEDGEMENTS

We thank Professors Frees and Valdez for kindly providing the loss-ALAE data, which were collected by the US Insurance Services Office (ISO). We also thank an Associate Editor, two anonymous referees and especially the guest Editor for detailed suggestions which greatly improved this paper. We owe special thanks to Christian Genest who carefully edited our paper. We are grateful to participants at 2004 DeMoSTAFI conference for helpful comments and to Demian Pouzo for excellent research assistance in the empirical application. Both authors gratefully acknowledge financial support from the National Science Foundation. Chen also acknowledges financial support from the C. V. Starr Center at NYU.

REFERENCES

C. Ai (1997). A semiparametric maximum likelihood estimator. *Econometrica*, 65, 933–963.

D. Andrews (1995). Nonparametric kernel estimation for semiparametric models. *Econometric Theory*, 11, 560–596.

W. Breymann, A. Dias & P. Embrechts (2003). Dependence structures for multivariate high-frequency data in finance. *Quantitative Finance*, 3, 1–14.

A. Cebrián, M. Denuit & O. Scaillet (2003). Testing for concordance ordering. Forthcoming in *Astin Bulletin*.

X. Chen & Y. Fan (2002). Estimation of copula-based semiparametric time series models. Forthcoming in *Journal of Econometrics*.

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

D. R. Cox (1961). Tests of separate families of hypotheses. In *Proc. 4th Berkeley Symp. 1*, J. Neyman (ed.), pp. 105–123, University of California Press, Berkeley, CA.

M. Denuit, O. Purcaru & I. Van Keilegom (2004). Bivariate Archimedean copula modelling for loss-ALAE data in non-life insurance. Working paper, UCL.

J.-D. Fermanian (2003). Goodness-of-fit Tests for Copulas. Forthcoming in *Journal of Multivariate Analysis*.

J.-D. Fermanian, D. Radulović & M. J. Wegkamp (2004). Weak convergence of empirical copula processes. *Bernoulli*, 10, 847–860.

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

P. Gänssler & W. Stute (1987). *Seminar on Empirical Processes*. DMV Seminar 9, Birkhäuser, Basel.

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

C. Genest, K. Ghoudi & L.-P. Rivest (1998). Comment on the paper by E. W. Frees and E. A. Valdez entitled "Understanding relationships using copulas." *North American Actuarial Journal*, 2, 143–149.

C. Genest, J-F. Quessy & B. Rémillard (2005). Goodness-of-fit procedures for copula models based on the integral probability transformation. In press.

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

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

K. Ghoudi & B. Rémillard (2004). Empirical processes based on pseudo-observations II: The multivariate case. In *Asymptotic Methods in Stochastics: Festschrift for Miklós Csörgő*, L. Horváth and B. Szyszkowicz (eds), The Fields Institute Communications Series 44, pp. 381–406, American Mathematical Society, Providence, RI.

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

R. P. Hansen (2003). A Test for Superior Predictive Ability. Manuscript, Brown University, Providence, RI.

H. Joe (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall, London.

M. Junker & A. May (2002). Measurement of aggregate risk with copulas. Preprint, Caesar.

C. A. J. Klaassen & J. A. Wellner (1997). Efficient estimation in the bivariate Normal copula model: Normal margins are least-favourable.. *Bernoulli*, 3, 55–77.

S. A. Klugman & R. Parsa (1999). Fitting bivariate loss distributions with copulas. *Insurance: Mathematics and Economics*, 24, 139–148.

O. Linton & Z. Xiao (2001). A Nonparametric Regression Estimator that Adapts to Error Distribution of Unknown Form. Manuscript, London School of Economics.

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

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

D. Oakes (1994). Multivariate survival distributions. *Journal of Nonparametric Statistics*, 3, 343–354.

F. Ruymgaart (1974). Asymptotic normality of nonparametric tests for independence. *The Annals of Statistics*, 2, 892–910.

F. Ruymgaart, G. Shorack & W. Zwet (1972). Asymptotic normality of nonparametric tests for independence. *The Annals of Mathematical Statistics*, 43, 1122–1135.

J. H. Shih & T. A. Louis (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, 51, 1384-1399.

A. Sklar (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de statistique de l'Université de Paris*, 8, 229–231.

A. van der Vaart (1998). *Asymptotic Statistics*. Cambridge University Press.

A. van der Vaart & J. A. Wellner (1996). *Weak Convergence and Empirical Processes: with Applications to Statistics*. Springer, New York.

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

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

H. White (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50, 143–161.

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

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

Received 1 July 2004
Accepted 11 February 2005

Xiaohong CHEN: `xiaohong.chen@nyu.edu`
Department of Economics, New York University
269 Mercer Street, 7th Floor
New York, NY 10003, USA

Yanqin FAN: `yanqin.fan@vanderbilt.edu`
Department of Economics, Vanderbilt University
VU Station B no 351819, 2301 Vanderbilt Place
Nashville, TN 37235, USA