

**ESTIMATION AND MODEL SELECTION OF SEMIPARAMETRIC COPULA-BASED  
MULTIVARIATE DYNAMIC MODELS UNDER COPULA MISSPECIFICATION**

by

Xiaohong Chen and Yanqin Fan



**Working Paper No. 04-W19**

September 2004

DEPARTMENT OF ECONOMICS  
VANDERBILT UNIVERSITY  
NASHVILLE, TN 37235

[www.vanderbilt.edu/econ](http://www.vanderbilt.edu/econ)

# Estimation and Model Selection of Semiparametric Copula-Based Multivariate Dynamic Models Under Copula Misspecification\*

Xiaohong Chen<sup>†</sup>      Yanqin Fan<sup>‡</sup>

First version: February 2004; This version: September 2004

## Abstract

Recently Chen and Fan (2003a) introduced a new class of semiparametric copula-based multivariate dynamic (SCOMDY) models. A SCOMDY model specifies the conditional mean and the conditional variance of a multivariate time series parametrically (such as VAR, GARCH), but specifies the multivariate distribution of the standardized innovation semiparametrically as a parametric copula evaluated at nonparametric marginal distributions. In this paper, we first study large sample properties of the estimators of SCOMDY model parameters under a misspecified parametric copula, and then establish pseudo likelihood ratio (PLR) tests for model selection between two SCOMDY models with possibly misspecified copulas. Finally we develop PLR tests for model selection between more than two SCOMDY models along the lines of the reality check of White (2000). The limiting distributions of the estimators of copula parameters and the PLR tests do not depend on the estimation of conditional mean and conditional variance parameters. Although the tests are affected by the estimation of unknown marginal distributions of standardized innovations, they have standard parametric rates and the limiting null distributions are very easy to simulate. Empirical applications to multiple daily exchange rate data indicate the simplicity and usefulness of the proposed tests. Although a SCOMDY model with Gaussian copula might be a reasonable model for some bivariate FX series, but a SCOMDY model with a copula which has (asymmetric) tail-dependence is generally preferred for tri-variate and higher dimensional FX series.

**JEL Classification:** C14; C22; G22

**KEY WORDS:** Multivariate dynamic models; Misspecified copulas; Multiple model selection; Semiparametric inference; Mixture copulas; t copula; Gaussian copula

---

\*We thank F. Diebold, G. Elliott, R. Engle, C. Granger, J. Hamilton, D. Hendry, A. Patton, N. Swanson and participants at the NYU/Stern seminar and the 2004 Predictive Methodology and Application in Economics and Finance Conference in honor of Clive Granger for helpful comments. We are grateful to Demian Pouzo for excellent research assistance in the empirical work. Chen acknowledges support from the National Science Foundation and the C.V. Starr Center at NYU. Fan acknowledges support from the National Science Foundation.

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

<sup>‡</sup>Yanqin Fan, Department of Economics, Vanderbilt University, VU Station B #351819, 2301 Vanderbilt Place, Nashville, TN 37235, USA. E-mail: yanqin.fan@vanderbilt.edu

# 1 Introduction

Economic and financial multivariate time series are typically nonlinear, non-normally distributed, and have nonlinear co-movements beyond the first two conditional moments. Granger (2002) points out that the classical linear multivariate modelling (based on the Gaussian distribution assumption) clearly fails to explain the stylized facts observed in economic and financial time series and that it is highly undesirable to perform various economic policy evaluations, financial forecasts, and risk managements based on the classical conditional (or unconditional) Gaussian modelling. The knowledge of the multivariate conditional distribution (especially fat-tails, asymmetry, positive or negative dependence) is essential in many important financial applications, including portfolio selection, option pricing, asset pricing models, Value-at-Risk (market risk, credit risk, liquidity risk) calculations and forecasting. Thus the entire conditional distribution of multivariate nonlinear economic and financial time series should be studied, see Granger (2002).

Recently Chen and Fan (2003a) introduce a new class of semiparametric copula-based multivariate dynamic (hereafter SCOMDY) models. A SCOMDY model specifies the multivariate conditional mean and conditional variance parametrically, but specifies the distribution of the (standardized) innovations semiparametrically as a parametric copula<sup>1</sup> evaluated at the nonparametric univariate marginals, where the copula function captures the concurrent dependence between the components of the multivariate innovation and the marginal distributions characterize their individual behaviors. Chen and Fan (2003a) demonstrate via examples the flexibility of SCOMDY models in capturing a wide range of nonlinear, asymmetric dependence structures and of the marginal behavior of a multivariate time series. In addition, a SCOMDY model allows for the estimation of multivariate conditional distribution semiparametrically, which, according to Granger (2002), is an important feature of a multivariate time series model.

There are three sets of unknown parameters associated with a SCOMDY model: the *dynamic* parameters (i.e., the finite-dimensional parameters of the conditional mean and conditional variance); the *copula dependence* parameters (i.e., the finite-dimensional parameters of the copula function of the standardized multivariate innovation); and the infinite-dimensional *marginal distributions* of individual components of the standardized innovation. Chen and Fan (2003a) provide simple estimators of the parameters in a correctly specified SCOMDY model and establish their asymptotic properties.

In this paper, we first study asymptotic properties of the estimators of the three sets of parameters associated with a SCOMDY model under a possibly misspecified parametric copula of the standardized innovation. This is motivated by the facts that financial theory and economic theory often shed little light on the specification of a parametric copula of the innovation and that the

---

<sup>1</sup>A copula is simply a multivariate probability distribution function with uniform marginals.

existing applications have typically used multiple choices of parametric copulas. While the true unknown dynamic parameters and the marginal distributions are still estimated consistently (at a root- $n$  rate,  $n$  is the sample size), the estimator of the copula dependence parameter will converge to the pseudo true value of the copula dependence parameter defined as the minimizer of the Kullback-Leibler Information Criterion (KLIC) between the candidate parametric copula density and the true unknown copula density. Interestingly, the limiting distribution of the estimator of the pseudo true value of the copula dependence parameters is not affected by the estimation of the dynamic parameters, albeit it does depend on the estimation of unknown marginal distributions.

As commonly used parametric copulas such as the Gaussian copula, the Gumbel copula, and the Clayton copula lead to SCOMDY models that may have very different dependence properties, one important issue in empirical implementation of any SCOMDY model is the choice of an appropriate parametric copula. Several papers have attempted to provide formal statistical model selection procedures for special cases of SCOMDY models. In modelling the dependence structure of bivariate high-frequency exchange rates data, Breymann, et al. (2003) and Dias and Embrechts (2004) applied the Akaike Information Criterion (AIC) to several parametric copulas and selected the one with the smallest AIC value. In modelling the dependence structure between risk factors in a portfolio, Junker and May (2002) applied a  $\chi^2$  goodness-of-fit test to each parametric copula and selected the one with smallest value of the test statistic. However, the important issue of data reuse has not been addressed in the context of copula model selection for general SCOMDY models.<sup>2</sup>

In this paper, we attempt to bridge this gap. In the case with only two models, we extend the likelihood ratio tests for model selection of parametric models in Vuong (1989) to SCOMDY models. Unlike Vuong (1989), the null hypothesis we entertain in this paper is: one copula model performs at least as well as the other in terms of KLIC, while in Vuong (1989), the null hypothesis is: the two models perform equally well. Our testing procedure is general, allowing both competing parametric copula models to be misspecified under the null and the alternative. Although our testing approach is similar to those in Vuong (1989), Sin and White (1996), Rivers and Vuong (2002), Marcellino (2002) and other work following Vuong (1989), we allow for infinite-dimensional nuisance parameters (marginal distributions) in our model selection criterion. Hence our test is really a pseudo- (or quasi-) likelihood ratio (hereafter PLR) test,<sup>3</sup> and the limiting distributions of our test statistics depend on the estimates of the unknown marginal distributions of the standardized innovations.

As we noted earlier, in empirical applications of copulas, it is more common to use several parametric copulas to fit the data and compare the results obtained from different models. To

---

<sup>2</sup>In Chen, et al. (2003) and Fermanian (2003), they respectively establish tests for the correct specification of a parametric copula for specific members of SCOMDY models. One drawback of these tests is that if the null hypothesis of correct specification is rejected, they provide no guidance as to which copula model to choose.

<sup>3</sup>Patton (2002) has applied Vuong (1989)'s likelihood ratio test in his study of purely parametric copula-based dynamic models. Our study differs from his since we do not specify marginal distributions of the standardized innovations.

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, one copula is taken as the benchmark copula and the rest candidate copulas: if no candidate model is closer to the true model (according to the KLIC distance) than the benchmark model, the benchmark model is chosen; otherwise, the candidate model that is closest to the true model will be selected. White (2000) proposes the reality check test for the superior predictive accuracy of at least one candidate parametric model over the benchmark parametric model.<sup>4</sup> Since the asymptotic null limiting distribution of the reality check test is complicated, White (2000) recommends to use either Monte Carlo simulation or bootstrap to compute the p-values. Although our PLR model selection tests involve the estimation of unknown marginal distributions of innovations, they have standard parametric rates and the limiting null distributions are very easy to simulate, thanks to the fact that the asymptotic distributions of the estimators of the pseudo true copula parameters are not affected by the estimation of the dynamic parameters in a SCOMDY model. To illustrate the simplicity and usefulness of our tests, we apply SCOMDY models and the model selection tests to a number of multivariate daily exchange rate series with dimensions 2, 3, and 10 respectively. Although a SCOMDY model with Gaussian copula might be a reasonable model for some (but not all) bivariate FX series, but a SCOMDY model with a copula which has (asymmetric) tail-dependence is generally preferred for three and ten dimensional FX series.

The rest of this paper is organized as follows. Section 2 briefly reviews the SCOMDY models. In Section 3, we study the large sample properties of the estimators of the SCOMDY model parameters proposed in Chen and Fan (2003a) under a possibly misspecified parametric copula. It is very interesting to note that the limiting distribution of the estimate of the copula dependence parameters is not affected by the estimation of the dynamic parameters of the conditional mean and conditional variance, although it does depend on the estimation of the unknown marginal distributions of the standardized innovations. This result is not only important in its own right, but also useful in establishing the asymptotic distribution of the PLR statistic under possibly misspecified copulas. In Section 4, we first present the null hypothesis and the PLR statistic for the model comparison of two SCOMDY models and then provide the limiting distributions of the PLR test statistics. Like the limiting distribution of the estimator of the copula dependence parameter, the limiting distribution of the PLR statistic does not depend on the dynamic parameter estimation, although it does depend on the estimation of marginal distributions of the standardized innovations of the SCOMDY models. Section 5 extends the above results to more than two competing SCOMDY models. Section 6 provides an empirical application and the last section briefly concludes. All technical proofs are gathered into the Appendix.

---

<sup>4</sup>Corradi and Swanson (2003, 2004), Su and White (2003) and Hansen (2003), among others, extend the reality check of White (2000) to different contexts.

## 2 Semiparametric Copula-based Multivariate Dynamic Models

Let  $\{(Y'_t, X'_t)\}_{t=1}^n$  be a vector stochastic process in which  $Y_t$  is of dimension  $d$ , and  $X_t$  is a vector of predetermined or exogenous variables distinct from the  $Y$ 's. Let  $\mathcal{I}_{t-1}$  denote the information set at time  $t$ , the sigma-field generated by  $\{Y_{t-1}, Y_{t-2}, \dots; X_t, X_{t-1}, \dots\}$ . In Chen and Fan (2003a), they specify the class of SCOMDY models as follows:

$$Y_t = \mu_t(\theta_{o1}) + \sqrt{H_t(\theta_o)}\epsilon_t, \quad (2.1)$$

where

$$\mu_t(\theta_{o1}) = (\mu_{1,t}(\theta_{o1}), \dots, \mu_{d,t}(\theta_{o1}))' = E\{Y_t | \mathcal{I}_{t-1}\}$$

is the true conditional mean of  $Y_t$  given  $\mathcal{I}_{t-1}$ , and is correctly parameterized up to a finite-dimensional unknown parameter  $\theta_{o1}$ ; and

$$H_t(\theta_o) = \text{diag.}(h_{1,t}(\theta_o), \dots, h_{d,t}(\theta_o)),$$

in which

$$h_{j,t}(\theta_o) = h_{j,t}(\theta_{o1}, \theta_{o2}) = E[(Y_{jt} - \mu_t(\theta_{o1}))^2 | \mathcal{I}_{t-1}], \quad j = 1, \dots, d,$$

is the true conditional variance of  $Y_{jt}$  given  $\mathcal{I}_{t-1}$ , and is correctly parameterized up to a finite-dimensional unknown parameter  $\theta_o = (\theta'_{o1}, \theta'_{o2})'$ , where  $\theta_{o1}$  and  $\theta_{o2}$  do not have common elements. The standardized multivariate innovations  $\{\epsilon_t \equiv (\epsilon_{1t}, \dots, \epsilon_{dt})' : t \geq 1\}$  in (2.1) are independent of  $\mathcal{I}_{t-1}$ , and are i.i.d. distributed with  $E(\epsilon_{jt}) = 0$  and  $E(\epsilon_{jt}^2) = 1$  for  $j = 1, \dots, d$ . Moreover,  $\epsilon_t = (\epsilon_{1t}, \dots, \epsilon_{dt})'$  has a distribution function  $F^o(\epsilon) = C^o(F_1^o(\epsilon_1), \dots, F_d^o(\epsilon_d))$ , where  $F_j^o(\cdot)$  is the true but unknown continuous marginal of  $\epsilon_{jt}$ ,  $j = 1, \dots, d$ , and  $C^o(u_1, \dots, u_d) = C^o(u_1, \dots, u_d; \alpha_o) : [0, 1]^d \rightarrow [0, 1]$  is the true copula function with unknown finite dimensional copula parameter  $\alpha_o$ .

Chen and Fan (2003a) provide many examples of SCOMDY models by combining different specifications of  $\mu_t(\theta_{o1})$ ,  $H_t(\theta_o)$  with  $C^o(u_1, \dots, u_d; \alpha_o)$ . Basically,  $\mu_t(\theta_{o1})$  and  $H_t(\theta_o)$  can be any commonly used conditional mean and conditional variance specifications such as ARCH, GARCH, VAR, Markov switching, etc., see Granger and Teräsvirta (1993), Hamilton (1994), Tsay (2002), and chapters on dependent processes in the *Handbook of Econometrics*, Vol. 4, edited by Engle and McFadden (1994). Similarly,  $C^o(u_1, \dots, u_d; \alpha_o)$  can be any parametric copula function such as the Normal (or Gaussian) copula, the Student's t-copula, the Frank copula, the Gumbel copula, and the Clayton copula, see Joe (1997) and Nelsen (1999) for examples and properties of copulas.

To illustrate, we present the following examples of SCOMDY models.

**Example 1 (GARCH(1,1)+Normal copula):** For  $j = 1, \dots, d$ ,

$$\begin{aligned} Y_{jt} &= X'_{jt}\delta_j + \sqrt{h_{jt}}\epsilon_{jt}, \\ h_{jt} &= \kappa_j + \beta_j h_{j,t-1} + \gamma_j (Y_{j,t-1} - X'_{j,t-1}\delta_j)^2, \end{aligned} \quad (2.2)$$

where

$$\kappa_j > 0, \beta_j \geq 0, \gamma_j \geq 0, \text{ and } \beta_j + \gamma_j < 1, j = 1, \dots, d.$$

In terms of our notation,  $\theta_1 = (\delta_1, \dots, \delta_d)'$ ,  $\theta_2 = (\kappa_1, \dots, \kappa_d; \beta_1, \dots, \beta_d; \gamma_1, \dots, \gamma_d)'$ ,  $\mu_t = (X'_{1t}\delta_1, \dots, X'_{dt}\delta_d)$  and  $H_t = \text{diag}\{h_{1t}, \dots, h_{dt}\}$ . The standardized multivariate innovations  $\{\epsilon_t \equiv (\epsilon_{1t}, \dots, \epsilon_{dt})' : t \geq 1\}$  are independent of  $\mathcal{I}_{t-1}$ , and are i.i.d. distributed with  $E(\epsilon_{jt}) = 0$  and  $E(\epsilon_{jt}^2) = 1$  for  $j = 1, \dots, d$ .

The copula of  $\epsilon_t = (\epsilon_{1t}, \dots, \epsilon_{dt})'$  is assumed to be the normal copula with unknown correlation matrix  $\Sigma$ . Let  $\Phi$  denote the univariate standard normal distribution and  $\Phi_{\Sigma, d}$  the  $d$ -dimensional normal distribution with correlation matrix  $\Sigma$ . Then the  $d$ -dimensional normal copula with correlation matrix  $\Sigma$  is

$$C(\mathbf{u}; \Sigma) = \Phi_{\Sigma, d}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d)),$$

whose copula density is

$$c(\mathbf{u}; \Sigma) = \frac{1}{\sqrt{\det(\Sigma)}} \exp \left\{ -\frac{(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d))' (\Sigma^{-1} - I_d) (\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d))}{2} \right\}.$$

The copula dependence parameter  $\alpha$  is the collection of all the unknown correlation coefficients in  $\Sigma$ . If  $\alpha \neq 0$ , then the corresponding Normal copula generates joint symmetric dependence, but no tail dependence (i.e., there is no joint extreme events).

**Example 2 (GARCH(1,1)+Student's t-copula):** The conditional mean and conditional variance of  $Y_{jt}$ ,  $j = 1, \dots, d$ , are specified in the same way as those in Example 1.

The copula of  $\epsilon_t = (\epsilon_{1t}, \dots, \epsilon_{dt})'$  is assumed to be a Student's t-copula with unknown correlation matrix  $\Sigma$ . Let  $T_\nu$  be the univariate Student's t distribution with  $\nu$  degrees of freedom, and  $T_{\Sigma, \nu}$  be the  $d$ -dimensional standardized Student's t distribution with  $\nu$  degrees of freedom and the correlation matrix  $\Sigma$ . Then the  $d$ -dimensional Student's t-copula with correlation matrix  $\Sigma$  is

$$C(\mathbf{u}; \Sigma, \nu) = T_{\Sigma, \nu}(T_\nu^{-1}(u_1), \dots, T_\nu^{-1}(u_d)).$$

The Student's  $t$  copula density is:

$$c(\mathbf{u}; \Sigma, \nu) = \frac{\Gamma(\frac{\nu+d}{2})[\Gamma(\frac{\nu}{2})]^{d-1}}{\sqrt{\det(\Sigma)}[\Gamma(\frac{\nu+1}{2})]^d} \left(1 + \frac{\mathbf{x}'\Sigma^{-1}\mathbf{x}}{\nu}\right)^{-\frac{\nu+d}{2}} \prod_{i=1}^d \left(1 + \frac{x_i^2}{\nu}\right)^{\frac{\nu+1}{2}},$$

$$\text{where } \mathbf{x} = (x_1, \dots, x_d)', \quad x_i = T_\nu^{-1}(u_i).$$

Let  $\alpha$  be the collection of correlation coefficients in  $\Sigma$ . For  $\nu < \infty$ , the Student's  $t$  copula with  $\alpha \neq 0$  can generate joint symmetric tail dependence, hence allow for *joint fat tails* (i.e., an increased probability of joint extreme events).

**Example 3 (GARCH(1,1)+Clayton copula):** The conditional mean and conditional variance of  $Y_{jt}$ ,  $j = 1, \dots, d$ , are specified in the same way as those in Example 1.

The copula of  $\epsilon_t = (\epsilon_{1t}, \dots, \epsilon_{dt})'$  is assumed to be the Clayton copula:

$$C(u_1, \dots, u_d; \alpha) = [u_1^{-\alpha} + \dots + u_d^{-\alpha} - d + 1]^{-1/\alpha}, \quad \text{where } \alpha > 0. \quad (2.3)$$

The copula density of the Clayton copula is given by

$$c(u_1, \dots, u_d; \alpha) = \{\prod_{j=1}^d [1 + (j-1)\alpha]\} \{\prod_{j=1}^d u_j^{-(\alpha+1)}\} \left[ \sum_{j=1}^d u_j^{-\alpha} - d + 1 \right]^{-(\alpha^{-1}+d)}, \quad \text{where } \alpha > 0.$$

Unlike the Gaussian and Student's  $t$  copulas, the Clayton copula can generate asymmetric dependence and lower tail dependence, but no upper tail dependence.

**Example 4 (GARCH(1,1)+Mixture copula):** The conditional mean and conditional variance of  $Y_{jt}$ ,  $j = 1, \dots, d$ , are specified in the same way as those in Example 1.

The copula of  $\epsilon_t = (\epsilon_{1t}, \dots, \epsilon_{dt})'$  is assumed to be a mixture copula:

$$C(u_1, \dots, u_d; \alpha) = \lambda C_1(u_1, \dots, u_d; \alpha_1) + (1 - \lambda) C_2(u_1, \dots, u_d; \alpha_2), \quad \lambda \in [0, 1], \quad (2.4)$$

where  $C_i(u_1, \dots, u_d; \alpha_i)$ ,  $i = 1, 2$ , are any two copula functions. For example, if we let  $C_1(\cdot; \alpha_1)$  be the Clayton copula, and  $C_2(\cdot; \alpha_2)$  be the Gumbel copula:

$$C_2(u_1, \dots, u_d; \alpha_2) = \exp \left( - \left[ \sum_{j=1}^d (-\ln u_j)^{\alpha_2} \right]^{1/\alpha_2} \right), \quad \text{where } \alpha_2 > 1,$$

then the mixture copula (2.4) will generate asymmetric lower and upper tail dependence when  $\lambda \in (0, 1)$ , lower tail dependence when  $\lambda = 1$  (Clayton copula), and upper tail dependence when  $\lambda = 0$  (Gumbel copula).

All four examples have been applied in empirical finance. For instance, Example 1 has been used in Hull and White (1998) for value-at-risk calculations for asset returns and exchange rates. It can be regarded as a special case of the DCC model proposed in Engle (2002) and Engle and Sheppard (2001). Examples 2 and 3 have been used in Breymann, et al. (2003) and Junker and May (2002), and Examples 2, 3 and 4 are used in Dias and Embrechts (2004) for joint tail dependence and risk management modelling for multivariate high frequency data.

### 3 Estimation Under Copula Misspecification

In this section, we first review the simple estimators of parameters in a SCOMDY model proposed in Chen and Fan (2003a) and then establish their large sample properties when the copula is misspecified.



### 3.1 Estimation of model parameters

Let  $\epsilon_t(\theta) \equiv [H_t(\theta)]^{-1/2}(Y_t - \mu_t(\theta_1))$  be the innovation function. The log-likelihood function for the SCOMDY model with a candidate copula function  $C(u_1, \dots, u_d; \alpha)$  is, (up to a constant term)

$$\begin{aligned} L_n(\theta, f; \alpha) &= \frac{1}{n} \sum_{t=1}^n l_t(\theta, f; \alpha) \\ &= \frac{1}{n} \sum_{t=1}^n \left\{ \frac{-\log |H_t(\theta)|}{2} + \sum_{j=1}^d \log f_j(\epsilon_{jt}(\theta)) + \log c(F_1(\epsilon_{1t}(\theta)), \dots, F_d(\epsilon_{dt}(\theta)); \alpha) \right\} \end{aligned}$$

where  $|H_t(\theta)|$  denotes the determinant of  $H_t(\theta)$ ,  $c(u_1, \dots, u_d; \alpha)$  is the copula density function associated with the copula function  $C(u_1, \dots, u_d; \alpha)$ , and  $f = (f_1, \dots, f_d)$  with  $f_j$  being the unknown probability density function (pdf) of  $F_j$ ,  $j = 1, \dots, d$ . As the marginal distributions  $F_j$  are completely unspecified, we normalize the mean and variance of the innovation  $\epsilon_{jt}$  such that  $E[\epsilon_{jt}(\theta)] = 0$  and  $Var[\epsilon_{jt}(\theta)] = 1$  for  $j = 1, \dots, d$ .

Throughout this paper, we let  $E^0[\cdot]$  denote the expectation of  $\cdot$  taken with respect to the true distribution  $C^o(F_1^o(\cdot), \dots, F_d^o(\cdot); \alpha_o)$ . We need to estimate three sets of parameters  $\theta_o$ ,  $(F_1^o, \dots, F_d^o)$  and  $\alpha^*$ , where  $\alpha^*$  is defined as

$$\begin{aligned} \alpha^* &\equiv \arg \max_{\alpha \in \mathcal{A}} E^0[\log c(F_1^o(\epsilon_{1t}(\theta_o)), \dots, F_d^o(\epsilon_{dt}(\theta_o)); \alpha)] \\ &= \arg \max_{\alpha \in \mathcal{A}} \int_{[0,1]^d} \log c(u_1, \dots, u_d; \alpha) c^o(u_1, \dots, u_d; \alpha_o) du_1 \cdots du_d, \end{aligned}$$

where  $c^o(u_1, \dots, u_d; \alpha_o)$  is the true copula density. If the copula density  $c(u_1, \dots, u_d; \alpha)$  correctly specifies the true copula density up to the copula parameter  $\alpha$ , then  $\alpha^*$  equals the true value  $\alpha_o$ . Otherwise, the copula density function  $c(u_1, \dots, u_d; \alpha^*)$  is the closest in the family of parametric copula densities  $\{c(u_1, \dots, u_d; \alpha) : \alpha \in \mathcal{A}\}$  to the true copula density in terms of the KLIC.

We first consider the estimation of  $\theta_o$  and  $(F_1^o, \dots, F_d^o)$ . The following estimators are proposed in Chen and Fan (2003a). The parameter  $\theta_{o1}$  is estimated by OLS:

$$\tilde{\theta}_1 = \arg \max_{\theta_1 \in \Theta_1} \left\{ \frac{-1}{2n} \sum_{t=1}^n [Y_t - \mu_t(\theta_1)]' [Y_t - \mu_t(\theta_1)] \right\};$$

and the parameter  $\theta_{o2}$  is estimated by QMLE:

$$\tilde{\theta}_2 = \arg \max_{\theta_2 \in \Theta_2} \frac{-1}{2n} \sum_{t=1}^n \sum_{j=1}^d \left\{ \frac{(Y_{jt} - \mu_t(\tilde{\theta}_1))^2}{h_{j,t}(\tilde{\theta}_1, \theta_2)} + \log h_{j,t}(\tilde{\theta}_1, \theta_2) \right\}.$$

Given the estimator  $\tilde{\theta} = (\tilde{\theta}_1, \tilde{\theta}_2)$ , one can estimate  $F_j^o$  using the rescaled empirical distribution of  $\{\epsilon_{jt}(\tilde{\theta})\}_{t=1}^n$ :

$$\tilde{F}_{nj}(x) = \frac{1}{n+1} \sum_{t=1}^n 1(\epsilon_{jt}(\tilde{\theta}) \leq x), \quad j = 1, \dots, d. \quad (3.1)$$

Since the estimators  $\tilde{\theta}$  and  $(\tilde{F}_{n1}, \dots, \tilde{F}_{nd})$  do not depend on the parametric copula specification, their asymptotic properties established in Chen and Fan (2003a) still hold. In particular, under mild regularity conditions,  $\tilde{\theta}$  is a  $\sqrt{n}$ -consistent estimator for  $\theta_o$  and  $(\tilde{F}_{n1}, \dots, \tilde{F}_{nd})$  is a  $\sqrt{n}$ -consistent estimator of  $(F_1^o, \dots, F_d^o)$ .

Given  $(\tilde{\theta}, \tilde{F}_{n1}, \dots, \tilde{F}_{nd})$ ,  $\alpha^*$  can be estimated by  $\hat{\alpha}$ :

$$\hat{\alpha} = \arg \max_{\alpha \in \mathcal{A}} \frac{1}{n} \sum_{t=1}^n \log c(\tilde{F}_{n1}(\epsilon_{1t}(\tilde{\theta})), \dots, \tilde{F}_{nd}(\epsilon_{dt}(\tilde{\theta})); \alpha). \quad (3.2)$$

Chen and Fan (2003a) establish the asymptotic properties of  $\hat{\alpha}$  when the parametric copula  $C(u_1, \dots, u_d; \alpha)$  correctly specifies the true copula. One surprising finding in Chen and Fan (2003a) is that the asymptotic distribution of  $\hat{\alpha}$  is not affected by the estimation of  $\theta_o$ , i.e., it is the same as that established in Genest, et al. (1995) for the case where  $\{\epsilon_t\}_{t=1}^n$  is observable.

### 3.2 Asymptotic properties of $\hat{\alpha}$ under copula misspecification

The difficulty in establishing the asymptotic properties of the estimator  $\hat{\alpha}$  arises from the fact that for many widely used copula functions including the Gaussian copula, the t-copula and the Clayton copula, the score function and its derivatives blow up to infinity. To handle this difficulty, Chen and Fan (2003a) first establish a weighted uniform CLT for the empirical process  $\sqrt{n}(\tilde{F}_{nj}(\cdot) - F_j^o(\cdot))$  based on pseudo-observations  $\{\epsilon_{jt}(\tilde{\theta})\}$  and then use it to prove the  $\sqrt{n}$ -consistency of  $\hat{\alpha}$  under the correct specification of parametric copula. In this section we modify their result to obtain the  $\sqrt{n}$ -consistent estimation of  $\alpha^*$  under misspecified copula.<sup>5</sup>

Let  $\mathcal{A}$  be the parameter space, a compact subset of  $\mathcal{R}^a$ . For  $\alpha \in \mathcal{A}$ , we use  $\|\alpha - \alpha^*\|$  to denote the usual Euclidean metric.

**Proposition 3.1** *Under Assumptions D and C stated in the Appendix, we have:  $\|\hat{\alpha} - \alpha^*\| = o_p(1)$ .*

Proposition 3.1 states that the estimator  $\hat{\alpha}$  is a consistent estimator of the pseudo true value  $\alpha^*$ . If the parametric copula correctly specifies the true copula, then  $\alpha^* = \alpha_o$  and  $\hat{\alpha}$  consistently estimate  $\alpha_o$ .

In the following we denote  $l(v_1, \dots, v_d, \alpha) = \log c(v_1, \dots, v_d, \alpha)$ ,  $l_\alpha(v_1, \dots, v_d, \alpha) = \frac{\partial l(v_1, \dots, v_d, \alpha)}{\partial \alpha}$ ,  $l_j(v_1, \dots, v_d, \alpha) = \frac{\partial l(v_1, \dots, v_d, \alpha)}{\partial v_j}$ ,  $l_{\alpha\alpha}(v_1, \dots, v_d; \alpha) = \frac{\partial^2 l(v_1, \dots, v_d; \alpha)}{\partial \alpha \partial \alpha'}$  and  $l_{\alpha j}(v_1, \dots, v_d; \alpha) = \frac{\partial^2 l(v_1, \dots, v_d; \alpha)}{\partial v_j \partial \alpha}$  for  $j = 1, \dots, d$ . Define  $U_{jt} \equiv F_j^o(\epsilon_{jt}(\theta_o))$  for  $j = 1, \dots, d$  and  $U_t = (U_{1t}, \dots, U_{dt})'$ . Denote

$$A_n^* \equiv \frac{1}{n} \sum_{s=1}^n \{l_\alpha(U_{1s}, \dots, U_{ds}, \alpha^*) + \sum_{j=1}^d Q_{\alpha j}(U_{js}; \alpha^*)\},$$

---

<sup>5</sup>Although White (1982) established the asymptotic properties of the maximum likelihood estimator under misspecified parametric models, his results are not directly applicable here since the estimation of the copula dependence parameter in a SCOMDY model under copula misspecification depends on the estimates of the unknown marginal distributions.

where

$$Q_{\alpha_j}(U_{js}; \alpha^*) \equiv E^0\{l_{\alpha_j}(U_t; \alpha^*)[I\{U_{js} \leq U_{jt}\} - U_{jt}]|U_{js}\}.$$

We also denote  $B \equiv -E^0[l_{\alpha\alpha}(U_t; \alpha^*)]$  and  $\Sigma \equiv Var^0[l_{\alpha}(U_s; \alpha^*) + \sum_{j=1}^d Q_{\alpha_j}(U_{js}; \alpha^*)]$ , and assume that both  $B$  and  $\Sigma$  are finite, positive definite.

**Proposition 3.2** *Let  $\alpha^* \in \text{int}(\mathcal{A})$ . Under Assumptions D and N stated in the Appendix, we have:*

(1)  $\hat{\alpha} - \alpha^* = B^{-1}A_n^* + o_p(n^{-1/2})$ ; (2)  $\sqrt{n}(\hat{\alpha} - \alpha^*) \rightarrow \mathcal{N}(0, B^{-1}\Sigma B^{-1})$  in distribution.

The additional terms  $Q_{\alpha_j}(U_{js}; \alpha^*)$  in  $A_n^*$  are introduced by the need to estimate the marginal distribution functions  $F_j^o(\cdot)$  ( $j = 1, \dots, d$ ). In the case where the distributions  $F_j^o(\cdot)$  ( $j = 1, \dots, d$ ) are completely known, these terms will disappear from  $A_n^*$ . It is interesting to note that the asymptotic distribution of  $\hat{\alpha}$  is not affected by the estimation of the dynamic parameters  $\theta_o$  and its asymptotic variance does not depend on the functional forms of the marginal distributions  $F_j^o$  ( $j = 1, \dots, d$ ).

**Remark:** Let  $\tilde{U}_t = (\tilde{U}_{1t}, \dots, \tilde{U}_{dt})' = (\tilde{F}_{n1}(\epsilon_{1t}(\tilde{\theta})), \dots, \tilde{F}_{nd}(\epsilon_{1t}(\tilde{\theta})))'$ . The asymptotic variance of  $\hat{\alpha}$  can be consistently estimated by  $\hat{B}^- \hat{\Sigma} \hat{B}^-$ , where  $\hat{B}^-$  is the generalized inverse of  $\hat{B} = -n^{-1} \sum_{t=1}^n l_{\alpha\alpha}(\tilde{U}_t; \hat{\alpha})$ , where

$$\hat{\Sigma} = \frac{1}{n} \sum_{t=1}^n [l_{\alpha}(\tilde{U}_t; \hat{\alpha}) + \sum_{j=1}^d \hat{Q}_{\alpha_j}(\tilde{U}_{jt}; \hat{\alpha})][l_{\alpha}(\tilde{U}_t; \hat{\alpha}) + \sum_{j=1}^d \hat{Q}_{\alpha_j}(\tilde{U}_{jt}; \hat{\alpha})]'$$

with

$$\hat{Q}_{\alpha_j}(U_{jt}; \hat{\alpha}) = \frac{1}{n} \sum_{s=1, s \neq t}^n \left( l_{\alpha_j}(U_s; \hat{\alpha}) \{I_{\{U_{jt} \leq U_{js}\}} - U_{js}\} \right).$$

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

## 4 Model Selection Between Two SCOMDY Models

In this section we first introduce the appropriate PLR statistic for testing model selection between two SCOMDY models along the lines of Vuong (1989). We then establish the limiting distribution of the PLR statistic.

### 4.1 The PLR statistic and its Asymptotic Properties

For each  $i = 1, 2$ , let  $\{C_i(u_1, \dots, u_d; \alpha_i) : \alpha_i \in \mathcal{A}_i \subset \mathcal{R}^{a_i}\}$  be a class of parametric copulas. Assuming that the conditional mean  $\mu_t$  and the conditional variance  $H_t$  are correctly specified, we

are interested in selecting a parametric copula such that the resulting SCOMDY model is closer to the true data generating process. Let

$$\ell_{t,i}(\alpha) = -\frac{1}{2} \log |H_t(\theta_o)| + \sum_{j=1}^d \log f_j^o(\epsilon_{jt}(\theta_o)) + \log c_i(F_1^o(\epsilon_{1t}(\theta_o)), \dots, F_d^o(\epsilon_{dt}(\theta_o)); \alpha_i),$$

in which  $c_i(\cdot; \alpha_i)$  is the density function of the copula  $C_i(\cdot; \alpha_i)$ , and  $f_j^o(\cdot)$  is the density function of the true marginal cdf  $F_j^o(\cdot)$  of  $\epsilon_{jt}(\theta_o)$ ,  $j = 1, \dots, d$ . Denote  $\alpha_i^* = \arg \max_{\alpha_i \in \mathcal{A}_i} E^0[\ell_{t,i}(\alpha_i)]$  as the pseudo true value associated with the copula model  $i = 1, 2$ . It depends on both the parametric copula  $c_i(u_1, \dots, u_d; \alpha)$  and the true copula  $c^o(u_1, \dots, u_d)$ .

Following Vuong (1989), we measure the closeness of a SCOMDY model to the true model by the minimum of the KLIC over the distributions in the copula model or equivalently by the maximum of  $E^0[\ell_{t,i}(\alpha_i)]$ . Since only the third term in the expression for  $\ell_{t,i}(\alpha_i)$  depends on the copula, an equivalent measure of the closeness of the  $i$ -th copula model to the true copula model is  $E^0 \log [c_i(F_1^o(\epsilon_{1t}(\theta_o)), \dots, F_d^o(\epsilon_{dt}(\theta_o)); \alpha_i^*)]$ ,  $i = 1, 2$ ; the larger  $E^0 \log [c_i(F_1^o(\epsilon_{1t}(\theta_o)), \dots, F_d^o(\epsilon_{dt}(\theta_o)); \alpha_i^*)]$ , the closer is the model to the true model. This motivates the following hypotheses: For pseudo true values  $\alpha_1^*$  and  $\alpha_2^*$ , the null hypothesis is

$$H_0 : E^0 \left\{ \log \frac{c_2(F_1^o(\epsilon_{1t}(\theta_o)), \dots, F_d^o(\epsilon_{dt}(\theta_o)); \alpha_2^*)}{c_1(F_1^o(\epsilon_{1t}(\theta_o)), \dots, F_d^o(\epsilon_{dt}(\theta_o)); \alpha_1^*)} \right\} \leq 0,$$

meaning that the copula model with the copula  $C_1(\cdot; \alpha_1)$  is not worse than the copula model with the copula  $C_2(\cdot; \alpha_2)$ , and the alternative hypothesis is

$$H_1 : E^0 \left\{ \log \frac{c_2(F_1^o(\epsilon_{1t}(\theta_o)), \dots, F_d^o(\epsilon_{dt}(\theta_o)); \alpha_2^*)}{c_1(F_1^o(\epsilon_{1t}(\theta_o)), \dots, F_d^o(\epsilon_{dt}(\theta_o)); \alpha_1^*)} \right\} > 0,$$

meaning that the copula model with  $C_1(\cdot; \alpha_1)$  is worse than the copula model with  $C_2(\cdot; \alpha_2)$ .

In the above formulation, one can take the copula model  $C_1(\cdot; \alpha_1)$  as the benchmark model and the model  $C_2(\cdot; \alpha_2)$  as a candidate model. Given the prevalence of the Gaussian distribution in multivariate financial time series modelling, it is natural to take the Gaussian copula model as the benchmark model; the Gaussian copula model will be retained unless the test strongly suggests that the candidate model outperforms the Gaussian copula model. In Junker and May (2002), the benchmark model is either the Student's t-copula or the Cook-Johnson copula.

Define

$$LR_n(\theta_o, F_1^o, \dots, F_d^o; \alpha_2^*, \alpha_1^*) = \frac{1}{n} \sum_{t=1}^n \left\{ \log \frac{c_2(F_1^o(\epsilon_{1t}(\theta_o)), \dots, F_d^o(\epsilon_{dt}(\theta_o)); \alpha_2^*)}{c_1(F_1^o(\epsilon_{1t}(\theta_o)), \dots, F_d^o(\epsilon_{dt}(\theta_o)); \alpha_1^*)} \right\}.$$

Let  $\hat{\alpha}_i$  denote the two-step estimator of  $\alpha_i^*$  for the SCOMDY model with copula  $C_i(u_1, \dots, u_d; \alpha_i)$ ,  $i = 1, 2$ . Our tests will be based on the following PLR statistic:

$$LR_n(\tilde{\theta}, \tilde{F}_{n1}, \dots, \tilde{F}_{nd}; \hat{\alpha}_2, \hat{\alpha}_1) = \frac{1}{n} \sum_{t=1}^n \left\{ \log \frac{c_2(\tilde{F}_{n1}(\epsilon_{1t}(\tilde{\theta})), \dots, \tilde{F}_{nd}(\epsilon_{dt}(\tilde{\theta})); \hat{\alpha}_2)}{c_1(\tilde{F}_{n1}(\epsilon_{1t}(\tilde{\theta})), \dots, \tilde{F}_{nd}(\epsilon_{dt}(\tilde{\theta})); \hat{\alpha}_1)} \right\}.$$

We first obtain the probability limit of the PLR statistic:

**Proposition 4.1** *Suppose for  $i = 1, 2$ , the copula model  $i$  satisfies assumptions of Proposition 3.1 and C6 in the Appendix. Then:  $LR_n(\tilde{\theta}, \tilde{F}_{n1}, \dots, \tilde{F}_{nd}; \hat{\alpha}_2, \hat{\alpha}_1) - E^0 \left[ \log \frac{c_2(U_{1t}, \dots, U_{dt}; \alpha_2^*)}{c_1(U_{1t}, \dots, U_{dt}; \alpha_1^*)} \right] = o_p(1)$ .*

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  $\{(v_1, \dots, v_d) : c_1(v_1, \dots, v_d; \alpha_1^*) \neq c_2(v_1, \dots, v_d; \alpha_2^*)\}$  has positive Lebesgue measure;*

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

It is important to note that as the closest copula in a parametric class of copulas depends on the true copula, it is not obvious a priori whether two parametric classes of copulas are generalized non-nested or generalized nested. However commonly used parametric classes of copulas such as the Clayton copula, the Gumbel copula, and the Gaussian copula can be shown to be generalized non-nested unless the closest member to the true copula in each class is the independence copula.

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

$$\Sigma_{12} = Cov^0[l_{1,\alpha}(U_s; \alpha_1^*) + \sum_{j=1}^d Q_{1,\alpha_j}(U_{js}; \alpha_1^*), l_{2,\alpha}(U_s; \alpha_2^*) + \sum_{j=1}^d Q_{2,\alpha_j}(U_{js}; \alpha_2^*)],$$

and

$$\sigma^2 = Var^0 \left[ \log \frac{c_2(U_{1t}, \dots, U_{dt}; \alpha_2^*)}{c_1(U_{1t}, \dots, U_{dt}; \alpha_1^*)} + \sum_{j=1}^d \{Q_{2,j}(U_{jt}; \alpha_2^*) - Q_{1,j}(U_{jt}; \alpha_1^*)\} \right] \quad (4.1)$$

where for  $i = 1, 2$ ,

$$Q_{i,j}(U_{js}, \alpha_i^*) \equiv E^0 \{l_{i,j}(U_t; \alpha_i^*) [I\{U_{js} \leq U_{jt}\} - U_{jt}] | U_{js}\} \quad \text{for } j = 1, \dots, d. \quad (4.2)$$

**Theorem 4.2** *Suppose for  $i = 1, 2$ , the copula model  $i$  satisfies assumptions of Proposition 3.2. Then:*

(1) *for the generalized non-nested case,*

$$n^{1/2} \left\{ LR_n(\tilde{\theta}, \tilde{F}_{n1}, \dots, \tilde{F}_{nd}; \hat{\alpha}_2, \hat{\alpha}_1) - E^0 \left[ \log \frac{c_2(U_{1t}, \dots, U_{dt}; \alpha_2^*)}{c_1(U_{1t}, \dots, U_{dt}; \alpha_1^*)} \right] \right\} \rightarrow \mathcal{N}(0, \sigma^2);$$

(2) for the generalized nested case,

$$\begin{aligned} 2nLR_n(\tilde{\theta}, \tilde{F}_{n1}, \dots, \tilde{F}_{nd}; \hat{\alpha}_2, \hat{\alpha}_1) &= n(\alpha_2^* - \hat{\alpha}_2)' B_2(\alpha_2^* - \hat{\alpha}_2) - n(\alpha_1^* - \hat{\alpha}_1)' B_1(\alpha_1^* - \hat{\alpha}_1) + o_p(1) \\ &\rightarrow M_{a_1+a_2}(\cdot; \lambda^*), \end{aligned}$$

where  $M_{a_1+a_2}(\cdot; \lambda^*)$  is the distribution of a weighted sum of independent  $\chi_{[1]}^2$  random variables in which the weight  $\lambda^* = (\lambda_1^*, \dots, \lambda_{a_1+a_2}^*)'$  is the  $(a_1 + a_2) \times 1$  vector of eigenvalues of the matrix  $W$  defined as

$$W = \begin{bmatrix} \Sigma_2 B_2^{-1} & -\Sigma'_{12} B_1^{-1} \\ \Sigma_{12} B_2^{-1} & -\Sigma_1 B_1^{-1} \end{bmatrix}.$$

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  $\sum_{j=1}^d \{Q_{2,j}(U_{jt}; \alpha_2^*) - Q_{1,j}(U_{jt}; \alpha_1^*)\}$ . These are introduced by the first step estimation of the unknown marginal distributions  $F_j^o$ ,  $j = 1, \dots, d$ . However, like the asymptotic distribution of  $\hat{\alpha}$ , the asymptotic distribution of the PLR statistic depends neither on the functional forms of the unknown marginal distributions  $F_j^o$ ,  $j = 1, \dots, d$ , nor on the estimation of the dynamic parameters  $\theta_o$ .

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  $n^{1/2}LR_n(\tilde{\theta}, \tilde{F}_{n1}, \dots, \tilde{F}_{nd}; \hat{\alpha}_2, \hat{\alpha}_1)$  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 4.3** *Let  $\sigma_a^2 = \text{Var}^0[\log \frac{c_2(U_{1t}, \dots, U_{dt}; \alpha_2^*)}{c_1(U_{1t}, \dots, U_{dt}; \alpha_1^*)}]$  and  $\sigma^2$  be given by (4.1). Under conditions of Theorem 4.2,  $\sigma^2 = 0$  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.*

## 4.2 PLR Tests

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 4.2(1). The usual approach to handling this problem is based on the Least Favorable Configuration (hereafter LFC) which is the point least favorable to the alternative. In our case, the LFC satisfies  $E^0 \left[ \log \frac{c_2(U_{1t}, \dots, U_{dt}; \alpha_2^*)}{c_1(U_{1t}, \dots, U_{dt}; \alpha_1^*)} \right] = 0$ . Under the LFC, Theorem 4.2(1) implies that  $n^{1/2}LR_n(\tilde{\theta}, \tilde{F}_{n1}, \dots, \tilde{F}_{nd}; \hat{\alpha}_2, \hat{\alpha}_1) \rightarrow \mathcal{N}(0, \sigma^2)$ . Moreover,  $\sigma^2 > 0$  by Proposition 4.3. We now provide a consistent estimator of  $\sigma^2$ .

First, by the definition of  $Q_{i,j}(U_{js}; \alpha_i^*)$  in (4.2), we have  $E^0\{Q_{i,j}(U_{js}; \alpha_i^*)\} = 0$  for  $i = 1, 2$  and  $j = 1, \dots, d$ . Moreover given  $U_{js}$ ,  $Q_{i,j}(U_{js}; \alpha_i^*)$  can be estimated by

$$\hat{Q}_{i,j}(U_{js}, \hat{\alpha}_i) = \frac{1}{n} \sum_{t=1, t \neq s}^n [l_{i,j}(U_t; \hat{\alpha}_i) \{I_{\{U_{js} \leq U_{jt}\}} - U_{jt}\}]. \quad (4.3)$$

Then a consistent estimator of  $\sigma^2$  is given by  $\hat{\sigma}^2 =$

$$\frac{1}{n} \sum_{t=1}^n \left[ \log \frac{c_2(\tilde{U}_t; \hat{\alpha}_2)}{c_1(\tilde{U}_t; \hat{\alpha}_1)} - \frac{1}{n} \sum_{s=1}^n \log \frac{c_2(\tilde{U}_s; \hat{\alpha}_2)}{c_1(\tilde{U}_s; \hat{\alpha}_1)} + \sum_{j=1}^d \{ \hat{Q}_{2,j}(\tilde{U}_{jt}; \hat{\alpha}_2) - \hat{Q}_{1,j}(\tilde{U}_{jt}; \hat{\alpha}_1) \} \right]^2. \quad (4.4)$$

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

$$T_n^N = \frac{n^{1/2} LR_n(\tilde{\theta}, \tilde{F}_{n1}, \dots, \tilde{F}_{nd}; \hat{\alpha}_2, \hat{\alpha}_1)}{\hat{\sigma}}, \quad (4.5)$$

where the superscript “ $N$ ” in  $T_n^N$  is meant for non-nested models and normal limiting distributions.

**Theorem 4.4** *Suppose the conditions of Proposition 4.1 and Theorem 4.2 hold and the two models are generalized non-nested. Then under the LFC,  $T_n^N \rightarrow \mathcal{N}(0, 1)$ .*

Proposition 4.1 and Theorem 4.4 suggest the following directional test: Given a significance level  $\alpha$ , reject  $H_0$  in favor of  $H_1$  if  $T_n^N > Z_\alpha$ , where  $Z_\alpha$  is the upper  $\alpha$ -percentile of the standard normal distribution (i.e.,  $Z_\alpha$  is the value of the inverse standard normal distribution evaluated at  $1 - \alpha$ ).

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. Define the test statistic:

$$T_n^Q = 2n LR_n(\tilde{\theta}, \tilde{F}_{n1}, \dots, \tilde{F}_{nd}; \hat{\alpha}_2, \hat{\alpha}_1), \quad (4.6)$$

where the superscript “ $Q$ ” in  $T_n^Q$  is meant for nested models and quadratic limiting statistics.

Theorem 4.2(2) implies that in this case the null limiting distribution of the PLR statistic  $T_n^Q$  is not distribution-free; it depends on both the parametric copulas and the true distribution function. However, similar to the estimation of the asymptotic variance of  $\hat{\alpha}$ , one can easily estimate the matrix  $W$  and hence its eigenvalues to get asymptotic critical values. Alternatively, one can use the method of bootstrap to approximate the critical values of the test. Typically, in order for a bootstrap test to work, the bootstrap sample must satisfy the null model. In our case, both parametric copulas can be misspecified and hence the null hypothesis does not specify a complete null model. In the case where the  $\epsilon_t$ 's are observable, Chen and Fan (2003b) propose to bootstrap  $2nD_n$  instead of  $T_n^Q$  via Efron's naive bootstrap, where

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

We will explore this possibility in the current setting in future work.

The tests we have proposed so far are general in the sense that they apply to cases where both parametric copula models could be misspecified.<sup>6</sup> However, one needs to know whether the two

---

<sup>6</sup>Our test procedure follows the approach of Vuong (1989) and Rivers and Vuong (2002), which contrasts with Cox's (1962) non-nested testing procedure by not requiring one of the competing models to be correct under the null hypothesis.

models are generalized non-nested. As the pseudo-values  $\alpha_1^*$  and  $\alpha_2^*$  are unknown, it is unknown a priori if this is the case. Vuong (1989) suggests 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.

The null hypothesis of generalized nested models can be tested by testing the null hypothesis  $\sigma_a^2 = 0$ . A consistent estimator of  $\sigma_a^2$  is given by

$$\hat{\sigma}_a^2 = \frac{1}{n} \sum_{t=1}^n \left[ \log \left\{ \frac{c_2(\tilde{U}_t; \hat{\alpha}_2)}{c_1(\tilde{U}_t; \hat{\alpha}_1)} \right\} - \frac{1}{n} \sum_{s=1}^n \log \left\{ \frac{c_2(\tilde{U}_s; \hat{\alpha}_2)}{c_1(\tilde{U}_s; \hat{\alpha}_1)} \right\} \right]^2. \quad (4.7)$$

Let  $\lambda^{*2} = (\lambda_1^{*2}, \dots, \lambda_{a_1+a_2}^{*2})'$ , the vector of squares of  $\lambda^* = (\lambda_1^*, \dots, \lambda_{a_1+a_2}^*)'$ , the eigenvalue weights in Theorem 4.2(2).

**Theorem 4.5** *Under the conditions of Proposition 4.1 and Theorem 4.2, we have:*

- (1)  $\hat{\sigma}_a^2$  given in (4.7) is a consistent estimator of  $\sigma_a^2$ ;
- (2) When  $\sigma_a^2 = 0$ ,  $n\hat{\sigma}_a^2 \rightarrow M_{a_1+a_2}(\cdot; \lambda^{*2})$  in distribution.

Theorem 4.5 and Proposition 4.3 suggest that a sequential test can be constructed in our case as well. First, 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. Estimates of the eigenvalues of the matrix  $W$  or a bootstrap procedure can be used to implement the test.

## 5 Multiple SCOMDY 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, see e.g. Breyman et al. (2003), Junker and May (2002) and Dias and Embrechts (2004). 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). In this case, all the candidate copula models are compared with a benchmark copula model. If no candidate model is closer to the true model than the benchmark model according to the KLIC distance, the benchmark model is chosen; otherwise, the candidate model that is closest to the true model will be selected. As mentioned earlier, one natural benchmark model is the Gaussian copula model, although the test applies to any benchmark model.

Let  $\{C_i(u_1, \dots, u_d; \alpha_i) : \alpha_i \in \mathcal{A}_i \subset \mathcal{R}^{\alpha_i}\}$  be a class of parametric copulas with  $i = 1, 2, \dots, M$ . As in the previous sections, we are interested in selecting a parametric copula such that the resulting SCOMDY model with copula  $C_i(u_1, \dots, u_d; \alpha_i)$  is closest to the true SCOMDY model with unknown copula  $C^o(u_1, \dots, u_d)$ . This can be formulated as follows. Let  $C_1(u_1, \dots, u_d; \alpha_1)$  be the benchmark



model and  $\{C_i(u_1, \dots, u_d; \alpha_i)\}_{i=2}^M$  be the candidate models. We are interested in testing if the best candidate model outperforms the benchmark model according to the KLIC distance. Hence, for pseudo true values  $\alpha_i^*$ ,  $i = 1, \dots, M$ , the null hypothesis is

$$H_0^M : \max_{i=2, \dots, M} E^0 \left\{ \log \frac{c_i(F_1^o(\epsilon_{1t}(\theta_o)), \dots, F_d^o(\epsilon_{dt}(\theta_o)); \alpha_i^*)}{c_1(F_1^o(\epsilon_{1t}(\theta_o)), \dots, F_d^o(\epsilon_{dt}(\theta_o)); \alpha_1^*)} \right\} \leq 0,$$

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

$$H_1^M : \max_{i=2, \dots, M} E^0 \left\{ \log \frac{c_i(F_1^o(\epsilon_{1t}(\theta_o)), \dots, F_d^o(\epsilon_{dt}(\theta_o)); \alpha_i^*)}{c_1(F_1^o(\epsilon_{1t}(\theta_o)), \dots, F_d^o(\epsilon_{dt}(\theta_o)); \alpha_1^*)} \right\} > 0,$$

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

Our test will be based on the following PLR statistics ( $i = 2, \dots, M$ ):

$$LR_n(\tilde{\theta}, \tilde{F}_{n1}, \dots, \tilde{F}_{nd}; \hat{\alpha}_i, \hat{\alpha}_1) = \frac{1}{n} \sum_{t=1}^n \left\{ \log \frac{c_i(\tilde{F}_{n1}(\epsilon_{1t}(\tilde{\theta})), \dots, \tilde{F}_{nd}(\epsilon_{dt}(\tilde{\theta})); \hat{\alpha}_i)}{c_1(\tilde{F}_{n1}(\epsilon_{1t}(\tilde{\theta})), \dots, \tilde{F}_{nd}(\epsilon_{dt}(\tilde{\theta})); \hat{\alpha}_1)} \right\}.$$

In the following we denote  $\Omega = (\sigma_{ik})_{i,k=2}^M$  in which

$$\begin{aligned} \sigma_{ik} &= \text{Cov}^0 \left[ \log \frac{c_i(U_{1t}, \dots, U_{dt}; \alpha_i^*)}{c_1(U_{1t}, \dots, U_{dt}; \alpha_1^*)} + \sum_{j=1}^d \{Q_{i,j}(U_{jt}; \alpha_i^*) - Q_{1,j}(U_{jt}; \alpha_1^*)\}, \right. \\ &\quad \left. \log \frac{c_k(U_{1t}, \dots, U_{dt}; \alpha_k^*)}{c_1(U_{1t}, \dots, U_{dt}; \alpha_1^*)} + \sum_{j=1}^d \{Q_{k,j}(U_{jt}; \alpha_k^*) - Q_{1,j}(U_{jt}; \alpha_1^*)\} \right], \end{aligned}$$

where for  $i = 1, 2, \dots, M$  and  $j = 1, \dots, d$ ,

$$Q_{i,j}(U_{jt}, \alpha_i^*) \equiv E^0 \{l_{i,j}(U_s; \alpha_i^*) [I\{U_{jt} \leq U_{js}\} - U_{js}] | U_{jt}\}.$$

**Proposition 5.1** *Suppose that for  $i = 1, 2, \dots, M$ , the copula model  $i$  satisfies conditions of Theorem 4.2. Suppose that  $\Omega$  is positive semi-definite and its largest eigenvalue is positive. Then jointly*

$$n^{1/2} \left\{ LR_n(\tilde{\theta}, \tilde{F}_{n1}, \dots, \tilde{F}_{nd}; \hat{\alpha}_i, \hat{\alpha}_1) - E^0 \left[ \log \frac{c_i(U_{1t}, \dots, U_{dt}; \alpha_i^*)}{c_1(U_{1t}, \dots, U_{dt}; \alpha_1^*)} \right] \right\}_{i=2, \dots, M} \rightarrow (Z_2, \dots, Z_M)',$$

in distribution, where  $(Z_2, \dots, Z_M)' \sim \mathcal{N}(0, \Omega)$ . Hence

$$\max_{i=2, \dots, M} n^{1/2} \left\{ LR_n(\tilde{\theta}, \tilde{F}_{n1}, \dots, \tilde{F}_{nd}; \hat{\alpha}_i, \hat{\alpha}_1) - E^0 \left[ \log \frac{c_i(U_t; \alpha_i^*)}{c_1(U_t; \alpha_1^*)} \right] \right\} \rightarrow \max_{i=2, \dots, M} Z_i \text{ in dist.}$$

Define

$$T_n^M = \max_{i=2, \dots, M} [n^{1/2} LR_n(\tilde{\theta}, \tilde{F}_{n1}, \dots, \tilde{F}_{nd}; \hat{\alpha}_i, \hat{\alpha}_1)].$$

Proposition 5.1 implies that under the LFC (i.e.,  $E^0 \left\{ \log \frac{c_i(U_{1t}, \dots, U_{dt}; \alpha_i^*)}{c_1(U_{1t}, \dots, U_{dt}; \alpha_1^*)} \right\} = 0$  for  $i = 2, \dots, M$ ),  $T_n^M \rightarrow \max_{i=2, \dots, M} Z_i$  in distribution. Since the asymptotic variance  $\Omega$  of  $(Z_2, \dots, Z_M)$  depends on  $\alpha_1^*, \dots, \alpha_M^*$ , the distribution of  $\max_i Z_i$  is unknown. Following White (2000), one can use either “Monte Carlo RC” p-value or “bootstrap RC” p-value to implement the test. In this paper, we use the Monte Carlo p-value approach. Let  $\widehat{\Omega} = (\widehat{\sigma}_{ik})_{i,k=2}^M$  where

$$\widehat{\sigma}_{ik} = \frac{1}{n} \sum_{t=1}^n \left[ \begin{array}{l} \left( \log \frac{c_i(\tilde{U}_t; \hat{\alpha}_i)}{c_1(\tilde{U}_t; \hat{\alpha}_1)} - \frac{1}{n} \sum_{s=1}^n \log \frac{c_i(\tilde{U}_s; \hat{\alpha}_i)}{c_1(\tilde{U}_s; \hat{\alpha}_1)} + \sum_{j=1}^d \{ \widehat{Q}_{i,j}(\tilde{U}_{jt}; \hat{\alpha}_i) - \widehat{Q}_{1,j}(\tilde{U}_{jt}; \hat{\alpha}_1) \} \right) \\ \times \left( \log \frac{c_k(\tilde{U}_t; \hat{\alpha}_k)}{c_1(\tilde{U}_t; \hat{\alpha}_1)} - \frac{1}{n} \sum_{s=1}^n \log \frac{c_k(\tilde{U}_s; \hat{\alpha}_k)}{c_1(\tilde{U}_s; \hat{\alpha}_1)} + \sum_{j=1}^d \{ \widehat{Q}_{k,j}(\tilde{U}_{jt}; \hat{\alpha}_k) - \widehat{Q}_{1,j}(\tilde{U}_{jt}; \hat{\alpha}_1) \} \right) \end{array} \right]$$

where  $\widehat{Q}_{i,j}$  is a consistent estimator of  $Q_{i,j}$  and is computed the same way as in (4.3). Then it is easy to show that  $\widehat{\Omega} = \Omega + o_p(1)$ . Next we compute the Cholesky decomposition of  $\widehat{\Omega} = \widehat{C}\widehat{C}'$ , and form  $\widehat{Z}_t = (\widehat{Z}_{t,2}, \dots, \widehat{Z}_{t,M})' = \widehat{C}\eta_t$  where  $\eta_t = N(0, I_{M-1})$  is a  $(M-1)$ -dimensional standard normal. Then compute the largest order statistics  $\zeta_{t,M-1} = \max_{k=2, \dots, M} \widehat{Z}_{t,k}$ . Repeat this for a large number of times (say 100,000) and compute the Monte Carlo p-value from the largest order statistics  $\zeta_{t,M-1}$ .

**Remark:** In the context of observable  $\epsilon_t$ 's, Chen and Fan (2003b) show the validity of a bootstrap procedure for tests closely related to the above test.

## 6 Empirical Applications

In this section, we estimate SCOMDY models and apply our copula model selection tests to numerous daily exchange rate data. We have considered SCOMDY models of dimensions 2, 3 and 10 respectively. Throughout this section, the benchmark model is taken as the normal copula, and the Monte Carlo p-values are determined by 100,000 simulations,<sup>7</sup> unless otherwise stated.

For all cases, we used an AR(p)-GARCH(r,s) model for the conditional mean and the conditional variance of each series. Specifically, for  $j = 1, \dots, d$ , let  $Y_{jt}$  be the logarithm difference of the  $j$ th exchange rate between  $t$  and  $t+1$ . The AR(p)-GARCH(r,s) model is given by

$$\begin{aligned} Y_{jt} &= c_j + \Theta_{1j}(L)Y_{jt-1} + \sigma_{j,t}\varepsilon_{j,t}, \\ \sigma_{j,t}^2 &= \omega_j + \Theta_{2j}(L)\sigma_{j,t-1}^2 + \Theta_{3j}(L)\sigma_{j,t-1}^2\varepsilon_{j,t-1}^2, \end{aligned}$$

where  $\Theta_{1j}(L), \Theta_{2j}(L), \Theta_{3j}(L)$  are autoregressive lag polynomials of orders  $p, r, s$  respectively.

The basic findings are as follows: (1) for two series cases, although AIC selects t-copula, but our model selection test cannot reject normal benchmark for some bivariate series; (2) for three series and ten series cases, our model selection tests strongly reject normal benchmark in favor of the copulas which have (asymmetric) tail dependence.

<sup>7</sup>The critical values of the tests are all determined by 100,000 simulations and are available upon request.

## 6.1 Two Series Case

We worked with two groups of daily exchange rates from 28-Apr-1986 to 26-Oct-1998. The first group consists in data from Japan/USA(JAP), Germany/USA(DEM), while the second group consists in data from Germany/USA and France/USA(FRA). Data are gathered from the Fed. We worked with seasonal adjusted data (using moving average filters). Table 1 summarizes the estimation results of an  $AR(1) - GARCH(1, 1)$  model for each of the three different series, where numbers with a \*\* means they are significant at 1%, numbers with a \* means they are significant at 5%, and those with no \* means they are not significant at 10%.

	$c \times 10^{-5}$	$\theta_1$	$\omega \times 10^{-7}$	$\theta_2$	$\theta_3$
DEM	-5.04	0.04*	11.00**	0.06**	0.92**
JAP	-7.07	0.05**	17.50**	0.07**	0.90**
FRA	-7.18	0.04*	9.80**	0.06**	0.91**

Table 1: Summary of  $AR - GARCH$  coefficients

Following Dias and Embrechts (2004) we considered the following ten copula functions: Clayton, Frank, Plackett, Gumbel, t-copula, Gaussian, mixture of Gumbel and Clayton, mixture of Clayton and Survival Clayton, mixture of Survival Clayton and Survival Gumbel, and mixture of Gumbel and Survival Gumbel. The estimates of the copula parameters are presented in Table 2, where  $\hat{\alpha}_1$  corresponds to the parameter in the first copula of the mixture (or the copula if there is no mixture) and  $\hat{\alpha}_2$  corresponds to the second copula of the mixture, and  $\hat{\lambda}$  is the weight of the first copula with the exception of the t-copula for which  $\hat{\alpha}_1$  corresponds to the degree of freedom and  $\hat{\alpha}_2$  corresponds to the correlation coefficient.

Copula	JAP-DEM			FRA-DEM		
	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\lambda}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\lambda}$
t-copula	6.761	0.609		2.114	0.975	
Gaussian	0.604			0.965		
Clayton	0.961			7.786		
Gumbel	1.663			6.713		
Frank	-4.471			-26.111		
Plackett	7.473			243.620		
Mix. Clay and Gumb	1.574	1.681	0.284	10.976	7.234	0.278
Mix. Clay and S-Clay	1.374	1.300	0.466	10.510	9.699	0.515
Mix. S-Gumb and S-Clay	1.724	1.351	0.628	8.092	6.964	0.798
Mix. S-Gumb and Gumb	1.545	1.663	0.000	9.914	4.976	0.628

Table 2: Estimates of Copula Parameters

Table 3 shows the value of the AIC for each copula. For JAP-DEM, AIC clearly chose t-copula followed by the mixture of Clayton and Gumbel. Gaussian copula is ranked fifth according to AIC,

but the difference between the value of AIC for Gaussian copula and for the mixture of Clayton and Gumbel is not as big as the difference between the value of AIC for Gaussian and for t-copula. For FRA-DEM, again AIC chose t-copula. Gaussian copula is ranked the 8-th for this case and it results in a value of the AIC that differs from the AIC for t-copula by a much larger amount than for JAP-DEM. This difference in the values of the AIC for Gaussian and t-copulas for the two cases JAP-DEM and FRA-DEM is picked up by the model selection test proposed in this paper. The values of the test statistic are 0.561 and 8.859 respectively for JAP-DEM and FRA-DEM resulting in Monte Carlo p-values 0.491 and 0.022 for JAP-DEM and FRA-DEM respectively. Hence the null hypothesis of the Gaussian copula is not rejected for JAP-DEM but rejected for FRA-DEM (in favor of the t-copula), even though AIC chose t-copula for both cases. The estimates of parameters of the t-copula for JAP-DEM and FRA-DEM also reveal much stronger tail dependence for FRA-DEM than for JAP-DEM as reflected in the estimates of the degree-of-freedom parameter, see Table 2.

Copula	AIC (JAP-DEM)	AIC (FRA-DEM)
t-copula	-1502.094	-9727.016
Gaussian	-1441.278	-8766.322
Clayton	-1100.545	-7839.061
Gumbel	-1409.238	-9136.888
Frank	-1327.923	-8539.438
Plackett	-1407.887	-9464.102
Mix Clay and Gumb	-1485.133	-9471.528
Mix Clay and S-Clay	-1443.325	-9130.872
Mix S-Gumb and S-Clay	-1480.082	-9468.231
Mix S-Gumb and Gumb	-1405.239	-9581.344

Table 3: AIC

## 6.2 Three Series Case

Data are again gathered from the Fed. We worked with three groups of daily exchange rates from 28-Apr-1986 to 26-Oct-1998.. The first group consists in data from Japan/USA, Germany/USA and France/USA used in the two series case. The second one is formed by Japan/USA, Germany/USA and Canada/USA(CAN). The third group is Germany/USA, Italy/USA(ITA) and France/USA. We worked with seasonal adjusted data (using moving average filters). Since the results for all three groups are very similar, we will only report them for the first group (Jap-Dem-Fra).

The estimation results for the conditional mean and the conditional variance for each series are reported in Table 1. The estimates of the copula parameters are presented in Table 4 except the

correlation matrix for t-copula and Gaussian copula. They are respectively

$$\begin{pmatrix} 1.000 & 0.609 & 0.975 \\ 0.609 & 1.000 & 0.593 \\ 0.975 & 0.593 & 1.000 \end{pmatrix} \text{ and } \begin{pmatrix} 0.998 & 0.604 & 0.965 \\ 0.604 & 0.998 & 0.590 \\ 0.965 & 0.590 & 0.998 \end{pmatrix}.$$

Copula	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\lambda}$		
Clayton	1.489				
Gumbel	1.998				
Frank	6.319				
t-copula	4.168				
Mix. Clay and Gumbel	0.774	3.431	0.426		
Mix. Clay and Surv. Clay.	2.643	2.203	0.498		
Mix. Surv. Gumbel and Surv. Clay.	1.805	3.963	0.610		
Mix. Surv. Gumbel and Gumbel	1.669	3.408	0.501		

Table 4: Estimates of Copula parameters

Table 5 shows the value of the AIC for each copula. The value of the test statistic is 7.418 with p-value 0.0947. Thus we strongly reject the null hypothesis of Gaussian copula in favor of the student's t-copula.

Copula	AIC
t-copula	-11039.808
Gaussian	-10210.469
Clayton	-4421.678
Gumbel	-5145.609
Frank	-4951.188
Mix. Clay and Gumbel	-5836.897
Mix. Clay and Surv. Clay.	-6096.045
Mix. Surv. Gumbel and Surv. Clay.	-5838.349
Mix. Surv. Gumbel and Gumbel	-6062.988

Table 5: AIC

### 6.3 Ten Series Case

We gathered the data from Hull's homepage. The data consist of 10 exchange rate against the US Dollar from: Australia (AUD), Belgium (BEF), Switzerland (CHF), Germany (DEM), Denmark (DKK), Spain (ESP), France (FRF), Great Britain (GBP), Holland (NLG) and Sweden (SEK). The dataset ranges from 04-Jan-94 to 15-Aug-97, daily frequency.

To save space, we omit the estimation results for the conditional mean and conditional variance of each series. We also omit the estimation of correlation matrix for t-copula and normal copula, and only report results for the rest of copula parameter estimation in Table 6.

Copula	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\lambda}$		
Clayton	0.612				
Gumbel	1.385				
t-Copula	7.885				
Mix. Clayton and Gumbel	1.299	1.096	0.686		
Mix. t-copula and Clayton	9.656	0.204	0.984		
Mix. t-copula and Gumbel	9.634	1.323	0.983		

Table 6: Estimates of Copula Parameters

Copula	AIC
t-copula	-12576.168
Gaussian	-11580.732
Clayton	-3661.3712
Gumbel	-3287.8690
Mixture Clayton and Gumbel	-4527.1653
Mixture Clayton and t-student	-12706.728
Mixture Gumbel and t-student	-12713.121

Table 7: AIC

The values of AIC in Table 7 show strong dominance of t-copula over Gaussian copula indicating the presence of joint tail dependence in the data and t-copula is slightly dominated by the mixture of Clayton and t-copula and the mixture of Gumbel and t-copula implying that in addition to joint tail dependence, the data may exhibit asymmetric dependence. Not surprisingly, the Gaussian copula is rejected by our model selection test (the value of the test statistic is 18.931 with a p-value 0.06).

## 7 Conclusion

In this paper, we establish the large sample properties of the estimators of SCOMDY model parameters proposed in Chen and Fan (2003a) under copula misspecification. One interesting finding is that the limiting distribution of the estimator of the pseudo true value of the copula dependence parameter is not affected by the estimation of the dynamic parameters, albeit it does depend on the estimation of unknown marginal distributions. Therefore, the common practice in empirical finance of ignoring the estimation error of the dynamic parameters is theoretically justified by our first order large sample theory. Nevertheless, our results show that the statistical uncertainty of the goodness-of-fit model selection criterion cannot be ignored. We then establish PLR tests for model selection of two SCOMDY models with possibly misspecified parametric copulas for both generalized nonnested copulas and generalized nested copulas. Finally we consider the PLR test for model selection between more than two SCOMDY models in which one is the benchmark model and the rest are candidate models. Here we assume that the benchmark and at least one of the

candidate models are generalized nonnested.

We are currently working on several extensions of the results reported in this paper. First, instead of the in-sample PLR model comparison, we could consider out-of-sample PLR model comparison. Second, we could follow the encompassing approach to perform in-sample and/or out-of-sample multiple SCOMDY model comparison. See Hendry and Richard (1982), Mizon and Richard (1986), Diebold (1989), White (1994), Clements and Hendry (1998), West (2001) and many others for the encompassing tests of model comparison. Third, since some of the copula applications are in terms of portfolio value-at-risk calculation, option pricing and forecasting, we could consider alternative loss functions instead of the KLIC, see e.g. Machina and Granger (2000), Elliott and Timmermann (2002) and Su and White (2003). In particular, we could follow the idea in Patton (2004) and use some economic measure to distinguish between competing copula models that may all be possibly misspecified. Third, we could consider the misspecification and model comparison of conditional mean, conditional variance, and copula specifications jointly. This will be related to the model comparison of semiparametric multivariate conditional distributions. The ideas in Diebold, et al. (1999), Giacomini and White (2004), and Corradi and Swanson (2004) might be useful here.

One limitation of the class of SCOMDY models arises from the assumption of a parametric copula specification where the copula parameters are constant. In the case of fully parametric models, Granger, et al. (2003) and Patton (2002) allow the copula parameters to change over time. We will explore the possibility of extending the results in this paper to allow for time varying copulas in future work.

# Appendix: Assumptions and Proofs

The following Assumption D is on the dynamic part of the SCOMDY model.

**Assumption D:**

**D1:**  $\{(Y'_t, X'_t)'\}_{t=1}^n$  satisfies the SCOMDY specification, and is stationary beta-mixing with decay rate  $\beta_t$  such that  $\sum_{t=1}^{\infty} \beta_t < \infty$ ;

**D2:**  $\sqrt{n}(\tilde{\theta} - \theta_o) = \frac{1}{\sqrt{n}} \sum_{t=1}^n \psi_t(\theta_o) + o_p(1) \rightarrow Z^*$  in distribution, where  $Z^* = \mathcal{N}(0, V_\theta)$  is a multivariate normal random variable with mean zero and variance  $V_\theta$ ;

**D3:** for  $j = 1, \dots, d$ , (i)  $\epsilon_{jt}(\theta) \equiv [h_{jt}(\theta)]^{-1/2}(Y_{jt} - \mu_{jt}(\theta_1))$  is continuously differentiable at the neighborhood of  $\theta_o$  for all  $t \geq 1$ ; (ii)  $E^0\left\{\frac{\nabla_\theta \mu_{jt}(\theta_{o1})}{\sqrt{h_{jt}(\theta_o)}}\right\} < \infty$  and  $E^0\left\{\frac{\nabla_\theta h_{jt}(\theta_o)}{h_{jt}(\theta_o)}\right\} < \infty$ .

In the following we denote the infeasible estimator of  $F_j^o$  as  $F_{nj}$ , which is the rescaled empirical distributions of  $\epsilon_{jt}(\theta_o)$  as if  $\theta_o$  were known:

$$F_{nj}(x) = \frac{1}{n+1} \sum_{t=1}^n 1(\epsilon_{jt}(\theta_o) \leq x), \quad j = 1, \dots, d.$$

Also denote  $h_{jt}^o = h_{jt}(\theta_o)$ ,  $\nabla_\theta \mu_{jt}^o = \nabla_\theta \mu_{jt}(\theta_{o1})$ ,  $\nabla_\theta h_{jt}^o = \nabla_\theta h_{jt}(\theta_o)$ .

**Lemma 5.2** (Chen and Fan 2003a, theorem 3.1) *Let  $w(\cdot)$  be a positive continuous weight function on  $(0, 1)$ , symmetric at  $v = 1/2$  and increasing on  $(0, 1/2]$ . Under Assumption D, we have:*

(1) *If  $\beta_t = O(t^{-b})$  for some  $b > 0$  and  $\int_0^1 \frac{1}{w(v)} \log(1 + \frac{1}{w(v)}) dv < \infty$ , then*

$$\sup_{y_j \in \mathcal{R}} \left| \frac{\tilde{F}_{nj}(y_j) - F_j^o(y_j)}{w(F_j^o(y_j))} \right| = o_{a.s.}(1).$$

(2) *If either (i)  $\beta_t = O(t^{-b})$  for some  $b > \gamma/(\gamma - 1)$  with  $\gamma > 1$  and  $\int_0^1 (\frac{1}{w(v)})^{2\gamma} dv < \infty$ ; or (ii)  $\beta_t = O(b^{-t})$  for some  $b > 1$  and  $\int_0^1 (\frac{1}{w(v)})^2 \log(1 + \frac{1}{w(v)}) dv < \infty$ , then*

$$\sqrt{n} \sup_{y_j \in \mathcal{R}} \left| \frac{\tilde{F}_{nj}(y_j) - F_j^o(y_j)}{w(F_j^o(y_j))} \right| = O_p(1).$$

(3) *uniformly over  $x \in \mathcal{R}$ ,*

$$\begin{aligned} & \tilde{F}_{nj}(x) - F_j^o(x) \\ &= F_{nj}(x) - F_j^o(x) + f_j^o(x) \left( E^0 \left\{ \frac{\nabla_\theta \mu_{jt}^o}{\sqrt{h_{jt}^o}} \right\} + x E^0 \left\{ \frac{\nabla_\theta h_{jt}^o}{2h_{jt}^o} \right\} \right)' (\tilde{\theta} - \theta_o) + o_p(n^{-1/2}). \end{aligned}$$

**Lemma 5.3** (Chen and Fan 2003a) *For a square integrable function  $g$  of  $U_t$ , we have:*

$$\begin{aligned} & E^0 \{ g(U_t) [\tilde{F}_{nj}(\epsilon_{jt}(\tilde{\theta})) - F_j^o(\epsilon_{jt}(\theta_o))] \} \\ &= \frac{1}{n} \sum_{s=1}^n E^0 \{ g(U_t) [1\{U_{js} \leq U_{jt}\} - U_{jt}] \mid U_{js} \} + o_p(n^{-1/2}). \end{aligned}$$



The following Assumption C is sufficient to ensure the consistency of  $\hat{\alpha}$  to  $\alpha^*$ . First we define  $\mathcal{F}_j$  as the space of probability distributions over the support of  $\epsilon_{jt}(\theta_o)$  [say  $\mathcal{R}$ ]. For any  $F_j \in \mathcal{F}_j$  we let  $\|F_j - F_j^o\|_{\mathcal{F}_j} = \sup_y |[F_j(y) - F_j^o(y)]/w(F_j^o(y))|$  with  $w(v) \equiv [v(1-v)]^{1-\xi}$  for all  $v \in (0, 1)$  and for some  $\xi \in (0, 1)$ . Let  $\mathcal{F}_{j\delta} = \{F_j \in \mathcal{F}_j : \|F_j - F_j^o\|_{\mathcal{F}_j} \leq \delta\}$  for a small  $\delta > 0$ . Obviously  $F_j^o \in \mathcal{F}_j$ , and  $\tilde{F}_{nj}$  given in (3.1) will belong to  $\mathcal{F}_j$  with probability approaching one. Let  $\mathcal{F}_\delta = \{F = (F_1, \dots, F_d) : F_j \in \mathcal{F}_{j\delta}, j = 1, \dots, d\}$ . We also use the short-hand notation  $F(\epsilon_t(\theta)) = (F_1(\epsilon_{1t}(\theta)), \dots, F_d(\epsilon_{dt}(\theta)))$ . Let  $N_o$  denote a small compact neighborhood of  $\theta_o$ .

**Assumption C:**

**C1.** There exists  $\alpha^* \in \mathcal{A}$ , a compact subset of  $\mathcal{R}^a$ , such that  $E^0[l_\alpha(F^o(\epsilon_t(\theta_o)); \alpha)] = 0$  if and only if  $\alpha = \alpha^*$ ;

**C2.** (i)  $l_\alpha(v_1, \dots, v_d; \alpha)$  is well-defined and is Lipschitz continuous in  $(v_1, \dots, v_d, \alpha) \in (0, 1)^d \times \mathcal{A}$ ;

(ii)  $l_{\alpha_j}(v_1, \dots, v_d; \alpha)$ ,  $j = 1, \dots, d$ , are well-defined and continuous in  $(v_1, \dots, v_d, \alpha) \in (0, 1)^d \times \mathcal{A}$ ;

**C3.**  $\{(Y_t, X_t) : t \geq 1\}$  is stationary  $\beta$ -mixing with the mixing decay rate  $\beta_t = O(t^{-b})$  for some  $b > 0$ ;

**C4.**  $E\{\sup_{\bar{\alpha} \in \mathcal{A}, \theta \in N_o} \|l_\alpha(F^o(\epsilon_t(\theta)); \bar{\alpha})\| \log^+ \|l_\alpha(F^o(\epsilon_t(\theta)), \bar{\alpha})\|\} < \infty$ ;

**C5.**  $E^0\{\sup_{\bar{\alpha} \in \mathcal{A}, F \in \mathcal{F}_\delta, \theta \in N_o} \|l_{\alpha_j}(F(\epsilon_t(\theta)); \bar{\alpha})\| w(F_j^o(\epsilon_{jt}(\theta)))\} < \infty$  for  $j = 1, \dots, d$ , where  $w(\cdot)$  satisfies the condition in Lemma A.1(1).

**Proof.** (**Proposition 3.1**) can be proved in the same way as that for Theorem 3.2 in Chen and Fan (2003a), except to replace their true copula parameter value  $\alpha_o$  by the pseudo-true value  $\alpha^*$ . □

The following Assumption N is sufficient to ensure the  $\sqrt{n}$ -asymptotic normality of  $\hat{\alpha}$ . First we denote  $\mathcal{G}_\delta = \{(\alpha, F) \in \mathcal{A} \times \mathcal{F}_\delta : \|\alpha - \alpha^*\| \leq \delta\}$  for a small  $\delta > 0$ . Let  $\{(\alpha_\eta, F_{1\eta}, \dots, F_{d\eta}) : \eta \in [0, 1]\} \subset \mathcal{G}_\delta$  be an one-dimensional smooth path in  $\mathcal{G}_\delta$  with  $(\alpha_\eta, F_{1\eta}, \dots, F_{d\eta})|_{\eta=0} = (\alpha^*, F_1^o, \dots, F_d^o)$ .

**Assumption N:**

**N1.** (i) C1 holds with  $\alpha^* \in \text{int}(\mathcal{A})$ ; (ii)  $\|\hat{\alpha} - \alpha^*\| = o_p(1)$ ; (iii)  $B \equiv -E^0[l_{\alpha\alpha}(U_t; \alpha^*)]$  is positive definite; (iv)  $\Sigma \equiv \text{Var}^0[\sqrt{n}A_n^*]$  is finite positive definite; (v) for  $j = 1, \dots, d$ ,  $\sup_x |\{\tilde{F}_{nj}(x) - F_j^o(x)\}/w_2(F_j^o(x))| = O_p(n^{-1/2})$ , where  $w_2(\cdot)$  satisfies the condition in Lemma A.1(2);

**N2.**  $l_{\alpha\alpha}(v; \alpha)$  is well-defined and continuous in  $(v, \alpha) \in (0, 1)^d \times \text{int}(\mathcal{A})$ ;

**N3.** the interchange of differentiation and integration of  $l_\alpha(F_{1\eta}(x_1), \dots, F_{d\eta}(x_d), \alpha_\eta)$  with respect to  $\eta \in (0, 1)$  is valid;

**N4.** (i)  $\{(Y_t, X_t) : t \geq 1\}$  is stationary  $\beta$ -mixing with the mixing decay rate  $\beta_t = O(t^{-b})$  for some  $b > \gamma/(\gamma - 1)$ , in which  $\gamma > 1$ ; (ii)  $E\{\|\sum_{j=1}^d Q_{\alpha_j}(U_{js}; \alpha^*)\|^2\} < \infty$  for some  $\gamma > 1$ ;

(iii)  $E\{\sup_{(\bar{\alpha}, F) \in \mathcal{G}_\delta, \theta \in N_o} \|l_\alpha(F(\epsilon_t(\theta)), \bar{\alpha})\|\}^{2\gamma} < \infty$  for some  $\gamma > 1$ ;

**N4'.** (i)  $\{(Y_t, X_t) : t \geq 1\}$  is stationary  $\beta$ -mixing with the mixing decay rate  $\beta_t = O(b^{-t})$  for some  $b > 1$ ; (ii)  $E\{\|\sum_{j=1}^d Q_{\alpha_j}(U_{js}; \alpha^*)\|^2 \log[1 + \|\sum_{j=1}^d Q_{\alpha_j}(U_{js}; \alpha^*)\|\}\} < \infty$ ;

(iii)  $E\{\sup_{(\bar{\alpha}, F) \in \mathcal{G}_\delta, \theta \in N_o} \|l_\alpha(F(\epsilon_t(\theta)), \bar{\alpha})\|^2 \log[1 + \|l_\alpha(F(\epsilon_t(\theta)), \bar{\alpha})\|\}\} < \infty$ ;

**N5.**  $E\{\sup_{(\bar{\alpha}, F) \in \mathcal{G}_\delta, \theta \in N_o} \|l_{\alpha\alpha}(F(\epsilon_t(\theta)), \bar{\alpha})\|\}^2 < \infty$ ;

**N6.**  $E\{\sup_{(\bar{\alpha}, F) \in \mathcal{G}_\delta, \theta \in N_o} \|l_{\alpha_j}(F(\epsilon_t(\theta)), \bar{\alpha})\| w(F_j^o(\epsilon_{jt}(\theta)))\}^2 < \infty$  for  $j = 1, \dots, d$ , where  $w(\cdot)$  satisfies the condition in Lemma A.1(1) and  $E\{\sup_{\theta \in N_o} [\frac{w_2(F_j^o(\epsilon_t(\theta)))}{w(F_j^o(\epsilon_t(\theta)))}]^2\} < \infty$ .

**Proof.** (**Proposition 3.2**) can be proved in the same way as that for Theorem 3.3 in Chen and Fan (2003a), except to replace their true copula parameter value  $\alpha_o$  by the pseudo-true value  $\alpha^*$ .

□

**Assumption C6:**  $E^0\{\sup_{(\bar{\alpha}, F) \in \mathcal{G}_\delta, \theta \in N_o} \|l_j(F(\epsilon_t(\theta)); \bar{\alpha})\| w(F_j^o(\epsilon_{jt}(\theta)))\} < \infty$  for  $j = 1, \dots, d$ , where  $w(\cdot)$  satisfies the condition in Lemma A.1(1).

**Proof. (Proposition 4.1)** Directly follows from Lemma A.1(1), Proposition 3.1, Assumptions D2 and C6.

□

**Proof. (Theorem 4.2):** Let  $l_i(\tilde{F}_{n1}(\epsilon_{1t}(\tilde{\theta})), \dots, \tilde{F}_{nd}(\epsilon_{dt}(\tilde{\theta})); \hat{\alpha}_i) = \log c_i(\tilde{F}_{n1}(\epsilon_{1t}(\tilde{\theta})), \dots, \tilde{F}_{nd}(\epsilon_{dt}(\tilde{\theta})); \hat{\alpha}_i)$  for  $i = 1, 2$ . By the definition of  $\hat{\alpha}_i$ , we have  $\sum_{t=1}^n l_{i,\alpha}(\tilde{F}_{n1}(\epsilon_{1t}(\tilde{\theta})), \dots, \tilde{F}_{nd}(\epsilon_{dt}(\tilde{\theta})); \hat{\alpha}_i) = 0$ . Hence,

$$\begin{aligned} & \sum_{t=1}^n l_i(\tilde{F}_{n1}(\epsilon_{1t}(\tilde{\theta})), \dots, \tilde{F}_{nd}(\epsilon_{dt}(\tilde{\theta})); \alpha_i^*) - \sum_{t=1}^n l_i(\tilde{F}_{n1}(\epsilon_{1t}(\tilde{\theta})), \dots, \tilde{F}_{nd}(\epsilon_{dt}(\tilde{\theta})); \hat{\alpha}_i) \\ &= \frac{(\alpha_i^* - \hat{\alpha}_i)' \sum_{t=1}^n l_{i,\alpha}(\tilde{F}_{n1}(\epsilon_{1t}(\tilde{\theta})), \dots, \tilde{F}_{nd}(\epsilon_{dt}(\tilde{\theta})); \bar{\alpha}_i)(\alpha_i^* - \hat{\alpha}_i)}{2}, \end{aligned}$$

where  $\bar{\alpha}_i$  is between  $\alpha_i^*$  and  $\hat{\alpha}_i$ . By conditions of Proposition 3.2, we have:

$$\begin{aligned} & \frac{(\alpha_i^* - \hat{\alpha}_i)' \sum_{t=1}^n l_{i,\alpha}(\tilde{F}_{n1}(\epsilon_{1t}(\tilde{\theta})), \dots, \tilde{F}_{nd}(\epsilon_{dt}(\tilde{\theta})); \bar{\alpha}_i)(\alpha_i^* - \hat{\alpha}_i)}{2n} \\ &= -\frac{(\alpha_i^* - \hat{\alpha}_i)' B_i(\alpha_i^* - \hat{\alpha}_i)}{2} + o_p\left(\frac{1}{n}\right). \end{aligned}$$

Hence,

$$\begin{aligned} & \frac{1}{n} \sum_{t=1}^n l_i(\tilde{F}_{n1}(\epsilon_{1t}(\tilde{\theta})), \dots, \tilde{F}_{nd}(\epsilon_{dt}(\tilde{\theta})); \hat{\alpha}_i) - \frac{1}{n} \sum_{t=1}^n l_i(F_1^o(\epsilon_{1t}(\theta_o)), \dots, F_d^o(\epsilon_{dt}(\theta_o)); \alpha_i^*) \\ &= \frac{(\alpha_i^* - \hat{\alpha}_i)' B_i(\alpha_i^* - \hat{\alpha}_i)}{2} + o_p\left(\frac{1}{n}\right) + \frac{1}{n} \sum_{t=1}^n \sum_{j=1}^d l_{i,j}(U_t; \alpha_i^*) [\tilde{F}_{nj}(\epsilon_{jt}(\tilde{\theta})) - F_j^o(\epsilon_{jt}(\theta_o))] \\ & \quad + \frac{1}{n} \sum_{t=1}^n \sum_{j=1}^d \{l_{i,j}(\bar{F}_1(\epsilon_{1t}(\bar{\theta})), \dots, \bar{F}_d(\epsilon_{dt}(\bar{\theta})); \alpha_i^*) - l_{i,j}(U_t; \alpha_i^*)\} [\tilde{F}_{nj}(\epsilon_{jt}(\tilde{\theta})) - F_j^o(\epsilon_{jt}(\theta_o))], \end{aligned}$$

where  $\bar{F}_k(\epsilon_{kt}(\bar{\theta}))$  is some random value between  $\tilde{F}_{nk}(\epsilon_{kt}(\tilde{\theta}))$  and  $F_k^o(\epsilon_{kt}(\theta_o))$  for  $k = 1, \dots, d$ .

As a result, we get

$$\begin{aligned} & LR_n(\tilde{\theta}, \tilde{F}_{n1}, \dots, \tilde{F}_{nd}; \hat{\alpha}_2, \hat{\alpha}_1) - E^0[LR_n(\theta_o, F_1^o, \dots, F_d^o; \alpha_2^*, \alpha_1^*)] \\ &= D1_n + D2_n + D3_n + D_n + o_p\left(\frac{1}{n}\right), \end{aligned}$$

where

$$\begin{aligned}
D1_n &\equiv LR_n(\theta_o, F_1^o, \dots, F_d^o; \alpha_2^*, \alpha_1^*) - E^0[LR_n(\theta_o, F_1^o, \dots, F_d^o; \alpha_2^*, \alpha_1^*)] \\
D2_n &\equiv \frac{1}{n} \sum_{t=1}^n \sum_{j=1}^d [l_{2,j}(U_t; \alpha_2^*) - l_{1,j}(U_t; \alpha_1^*)] [\tilde{F}_{nj}(\epsilon_{jt}(\tilde{\theta})) - F_j^o(\epsilon_{jt}(\theta_o))] \\
D3_n &\equiv \frac{1}{n} \sum_{t=1}^n \sum_{j=1}^d \{l_{2,j}(\bar{F}_1(\epsilon_{1t}(\bar{\theta})), \dots, \bar{F}_d(\epsilon_{dt}(\bar{\theta})); \alpha_2^*) - l_{2,j}(U_t; \alpha_2^*)\} [\tilde{F}_{nj}(\epsilon_{jt}(\tilde{\theta})) - F_j^o(\epsilon_{jt}(\theta_o))] \\
&\quad - \frac{1}{n} \sum_{t=1}^n \sum_{j=1}^d \{l_{1,j}(\bar{F}_1(\epsilon_{1t}(\bar{\theta})), \dots, \bar{F}_d(\epsilon_{dt}(\bar{\theta})); \alpha_1^*) - l_{1,j}(U_t; \alpha_1^*)\} [\tilde{F}_{nj}(\epsilon_{jt}(\tilde{\theta})) - F_j^o(\epsilon_{jt}(\theta_o))] \\
D_n &\equiv \frac{1}{2}(\hat{\alpha}_2 - \alpha_2^*)' B_2(\hat{\alpha}_2 - \alpha_2^*) - \frac{1}{2}(\hat{\alpha}_1 - \alpha_1^*)' B_1(\hat{\alpha}_1 - \alpha_1^*).
\end{aligned}$$

(1) For generalized non-nested models, the terms  $D1_n$  and  $D2_n$  are of the order  $O_p(n^{-1/2})$ , while  $D3_n = o_p(n^{-1/2})$  by Lemma A.1(2) and conditions C2(ii) and C5, and  $D_n = O_p(n^{-1})$  by Proposition 3.2. Thus

$$LR_n(\tilde{\theta}, \tilde{F}_{n1}, \dots, \tilde{F}_{nd}; \hat{\alpha}_2, \hat{\alpha}_1) - E^0[LR_n(\theta_o, F_1^o, \dots, F_d^o; \alpha_2^*, \alpha_1^*)] = D1_n + D2_n + o_p\left(\frac{1}{\sqrt{n}}\right).$$

Further by the same proof as that for Proposition 3.2, and applying Lemma A.2 with  $g(U_t) = l_{2,j}(U_t; \alpha_2^*) - l_{1,j}(U_t; \alpha_1^*)$ , we have

$$\begin{aligned}
D2_n &= E^0\left\{\sum_{j=1}^d [l_{2,j}(U_t; \alpha_2^*) - l_{1,j}(U_t; \alpha_1^*)] [\tilde{F}_{nj}(\epsilon_{jt}(\tilde{\theta})) - F_j^o(\epsilon_{jt}(\theta_o))]\right\} + o_p\left(\frac{1}{\sqrt{n}}\right) \\
&= \sum_{j=1}^d \frac{1}{n} \sum_{s=1}^n E^0\{[l_{2,j}(U_t; \alpha_2^*) - l_{1,j}(U_t; \alpha_1^*)] [1\{U_{js} \leq U_{jt}\} - U_{jt}] | U_{js}\} + o_p\left(\frac{1}{\sqrt{n}}\right) \\
&\equiv \frac{1}{n} \sum_{s=1}^n \sum_{j=1}^d [Q_{2,j}(U_{js}; \alpha_2^*) - Q_{1,j}(U_{js}; \alpha_1^*)] + o_p\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

Hence

$$\begin{aligned}
LR_n(\tilde{\theta}, \tilde{F}_{n1}, \dots, \tilde{F}_{nd}; \hat{\alpha}_2, \hat{\alpha}_1) &- E^0 \log \left\{ \frac{c_2(F_1^o(\epsilon_{1t}(\theta_o)), \dots, F_d^o(\epsilon_{dt}(\theta_o)); \alpha_2^*)}{c_1(F_1^o(\epsilon_{1t}(\theta_o)), \dots, F_d^o(\epsilon_{dt}(\theta_o)); \alpha_1^*)} \right\} \\
&= \frac{1}{n} \sum_{s=1}^n \left\{ \log \frac{c_2(U_s; \alpha_2^*)}{c_1(U_s; \alpha_1^*)} - E^0 \log \frac{c_2(U_s; \alpha_2^*)}{c_1(U_s; \alpha_1^*)} + \sum_{j=1}^d [Q_{2,j}(U_{js}; \alpha_2^*) - Q_{1,j}(U_{js}; \alpha_1^*)] \right\} + o_p\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

(2) For generalized nested models, the terms  $D1_n$ ,  $D2_n$  and  $D3_n$  become zero almost surely, we have

$$LR_n(\tilde{\theta}, \tilde{F}_{n1}, \dots, \tilde{F}_{nd}; \hat{\alpha}_2, \hat{\alpha}_1) - E^0[LR_n(\theta_o, F_1^o, \dots, F_d^o; \alpha_2^*, \alpha_1^*)] = D_n + o_p\left(\frac{1}{n}\right),$$

where by Proposition 3.2,  $D_n = O_p(n^{-1})$  and

$$\begin{aligned}
2nD_n &= nA_{n2}^* B_2^{-1} A_{n2}^* - nA_{n1}^* B_1^{-1} A_{n1}^* \\
&= \begin{pmatrix} \sqrt{n}A_{n2}^* \\ \sqrt{n}A_{n1}^* \end{pmatrix}' \begin{pmatrix} B_2^{-1} & 0 \\ 0 & -B_1^{-1} \end{pmatrix} \begin{pmatrix} \sqrt{n}A_{n2}^* \\ \sqrt{n}A_{n1}^* \end{pmatrix}
\end{aligned}$$

is distributed as the weighted chi-squares as stated in the Proposition in which

$$W = \begin{bmatrix} \Sigma_2 & \Sigma'_{12} \\ \Sigma_{12} & \Sigma_1 \end{bmatrix} \begin{bmatrix} B_2^{-1} & 0 \\ 0 & -B_1^{-1} \end{bmatrix}.$$

□

**Proof. (Proposition 4.3)** 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 the two copula models are generalized nested. **(i)** Define

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

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

$$\begin{aligned} a_n &= \int_{[0,1]^d} \log \frac{c_2(u_1, \dots, u_d; \alpha_2^*)}{c_1(u_1, \dots, u_d; \alpha_1^*)} dC_n(u_1, \dots, u_d) \\ &= \frac{1}{n} \sum_{t=1}^n \log \frac{c_2(U_{1t}, \dots, U_{dt}; \alpha_2^*)}{c_1(U_{1t}, \dots, U_{dt}; \alpha_1^*)}, \end{aligned}$$

where  $C_n(u_1, \dots, u_d)$  is the empirical cdf estimator of the true copula function  $C^o(u_1, \dots, u_d)$ . It is obvious that

$$\sqrt{n}(a_n - a^*) \rightarrow \mathcal{N}(0, \sigma_a^2) \quad \text{with} \quad \sigma_a^2 = \text{Var}^0 \left[ \log \frac{c_2(U_{1t}, \dots, U_{dt}; \alpha_2^*)}{c_1(U_{1t}, \dots, U_{dt}; \alpha_1^*)} \right].$$

Now suppose that we do not observe an i.i.d. random sample  $\{U_t = (U_{1t}, \dots, U_{dt})\}_{t=1}^n$ , instead we observe an i.i.d. random sample  $\{\epsilon_t^o = (\epsilon_{1t}^o, \dots, \epsilon_{dt}^o)\}_{t=1}^n$  where  $\epsilon_t^o$  is distributed according to the unknown true multivariate cdf function  $F^o(\epsilon) = C^o(F_1^o(\epsilon_1), \dots, F_d^o(\epsilon_d))$ , with  $F_j^o(\cdot)$  being the unknown marginal cdf of  $\epsilon_{jt}^o$ . Let  $F_{nj}(x) = \frac{1}{n+1} \sum_{t=1}^n 1(\epsilon_{jt}^o \leq x)$ ,  $j = 1, \dots, d$  be the rescaled empirical cdf estimator of  $F_j^o(\cdot)$ . Then

$$\tilde{a} = \frac{1}{n} \sum_{t=1}^n \log \frac{c_2(F_{n1}(\epsilon_{1t}^o), \dots, F_{nd}(\epsilon_{dt}^o); \alpha_2^*)}{c_1(F_{n1}(\epsilon_{1t}^o), \dots, F_{nd}(\epsilon_{dt}^o); \alpha_1^*)}$$

will be another estimator of  $a^*$ , moreover

$$\sqrt{n}(\tilde{a} - a^*) \rightarrow \mathcal{N}(0, \sigma^2) \quad \text{with} \quad \sigma^2 \text{ given in (4.1).}$$

Hence  $\sigma^2 \geq \sigma_a^2$ . Since  $\sigma_a^2 \geq 0$ , we have that  $\sigma^2 = 0$  implies  $\sigma_a^2 = 0$ .

**(ii)** Next,  $\sigma_a^2 = 0$  implies  $\log \left\{ \frac{c_2(U_{1t}, \dots, U_{dt}; \alpha_2^*)}{c_1(U_{1t}, \dots, U_{dt}; \alpha_1^*)} \right\} = K$  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. (Theorem 4.4)** Theorem 4.2(1) and Proposition 4.3 imply that  $\sigma > 0$  and

$$\frac{n^{1/2}LR_n(\tilde{\theta}, \tilde{F}_{n1}, \dots, \tilde{F}_{nd}; \hat{\alpha}_2, \hat{\alpha}_1)}{\sigma} \rightarrow \mathcal{N}(0, 1).$$

Now under the conditions of Proposition 4.1 and Theorem 4.2(1), we have  $\hat{\sigma}^2 = \sigma^2 + o_p(1)$ .

□

**Proof. (Theorem 4.5)** (1) trivially true under the conditions of Proposition 4.1 and Theorem 4.2. (2) By Theorem 4.2(2) and Proposition 4.3, Taylor series expansion and the fact that  $C_1$  and  $C_2$  are generalized nested copula models, we get

$$n\hat{\sigma}_a^2 = \sum_{t=1}^n [l_{2,\alpha}(\tilde{U}_t; \alpha_2^*)(\hat{\alpha}_2 - \alpha_2^*) - l_{1,\alpha}(\tilde{U}_t; \alpha_1^*)(\hat{\alpha}_1 - \alpha_1^*)] + o_p(1) = nV_d + o_p(1),$$

where  $V_d =$

$$\begin{bmatrix} \hat{\alpha}_2 - \alpha_2^* \\ \hat{\alpha}_1 - \alpha_1^* \end{bmatrix}' \begin{bmatrix} E^0 l_{2,\alpha}(U_t; \alpha_2^*)' l_{2,\alpha}(U_t; \alpha_2^*) & -E^0 l_{2,\alpha}(U_t; \alpha_2^*)' l_{1,\alpha}(U_t; \alpha_1^*) \\ -E^0 l_{1,\alpha}(U_t; \alpha_1^*)' l_{2,\alpha}(U_t; \alpha_2^*) & E^0 l_{1,\alpha}(U_t; \alpha_1^*)' l_{1,\alpha}(U_t; \alpha_1^*) \end{bmatrix} \begin{bmatrix} \hat{\alpha}_2 - \alpha_2^* \\ \hat{\alpha}_1 - \alpha_1^* \end{bmatrix}$$

The result now follows.

□

**Proof. (Proposition 5.1)** Essentially the same proof as that for Theorem 4.2(1).

□

## References

- [1] Breymann, W. , A. Dias, and Embrechts, P. (2003), “Dependence Structures for Multivariate High-Frequency Data in Finance,” *Quantitative Finance* 3(1), 1-16.
- [2] Chen, X. and Y. Fan (2003a), “Specification and Estimation of A New Class of Semiparametric Copula-based Multivariate Dynamic Models,” Manuscript, New York University and Vanderbilt University.
- [3] Chen, X. and Y. Fan (2003b), “Pseudo-Likelihood Ratio Tests for Model Selection in Semiparametric Multivariate Copula Models,” Manuscript, New York University and Vanderbilt University.
- [4] Chen, X., Y. Fan and A. Patton (2003), “Simple Tests for Models of Dependence Between Multiple Financial Time Series: with Applications to U.S. Equity Returns and Exchange Rates,” Manuscript, New York University, Vanderbilt University and London School of Economics.
- [5] Clements, M. and D. Hendry (1998), *Forecasting Economic Time Series*, Cambridge University Press.
- [6] Corradi, V. and N. Swanson (2003), “A Test for Comparing Multiple Misspecified Conditional Distributions,” Manuscript, Rutgers University.
- [7] Corradi, V. and N. Swanson (2004), “Predictive Density Accuracy Tests,” Manuscript, Rutgers University.
- [8] Cox, D. (1962), “Further Results on Tests of Separate Families of Hypotheses,” *J. R. Statist. Soc. B* 24, 406-424.
- [9] Dias, A. and P. Embrechts (2004), “Dynamic Copula Models for Multivariate High-Frequency Data in Finance,” Manuscript, ETH Zurich.
- [10] Diebold, F.X. (1989), “Forecast Combination and Encompassing: Reconciling Two Divergent Literature,” *International Journal of Forecasting* 5, 589-592.
- [11] Diebold, F.X., J. Hahn and A. Tay (1999), “Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High Frequency Returns on Foreign Exchange,” *Review of Economics and Statistics* 81, 661-673.
- [12] Elliott, G. and A. Timmermann (2002), “Optimal Forecast Combinations under General Loss Functions and Forecast Error Distributions,” Manuscript, UCSD.
- [13] Engle, R. (2002), “Dynamic Conditional Correlation-A Simple Class of Multivariate GARCH Models,” Forthcoming in *Journal of Business and Economic Statistics*.
- [14] Engle, R. and D. McFadden (1994), *Handbook of Econometrics, Vol. 4*, Elsevier: New York.
- [15] Engle, R. and K. Sheppard (2001), “Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH,” Working paper, UCSD and New York University.
- [16] Fermanian, J.-D. (2003), “Goodness of Fit Tests for Copulas,” Mimeo, CREST.

- [17] Genest, C. , K. Ghouli, and L.-P. Rivest (1995), “A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions,” *Biometrika* 82 (3), 543-552.
- [18] Giacomini, R. and H. White (2004), “Tests of Conditional Predictive Ability,” Working paper, UCSD.
- [19] Granger, C.W.J. (2002), “Time Series Concept for Conditional Distributions,” Manuscript, UCSD.
- [20] Granger, C.W.J. and T. Teräsvirta (1993), *Modelling Nonlinear Economic Relationships*. Oxford: Oxford University Press.
- [21] Granger, C.W.J., T. Teräsvirta and A. Patton (2003), “Common Factors in Conditional Distributions,” Forthcoming in *Journal of Econometrics*.
- [22] Hamilton, J. (1994), *Time Series Analysis*. Princeton: Princeton University Press.
- [23] Hansen, R. P. (2003), “A Test for Superior Predictive Ability,” Manuscript, Brown University.
- [24] Hendry, D. and J. Richard (1982), “On the Formulation of Empirical Models in Dynamic Econometrics,” *Journal of Econometrics* 20, 3-33.
- [25] Hull, J. and A. White (1998): “Value at Risk when Daily Changes in Market Variables are not Normally Distributed,” *Journal of Derivatives* 5, 9-19.
- [26] Joe, H. (1997), *Multivariate Models and Dependence Concepts*, Chapman & Hall/CRC.
- [27] Junker, M. and A. May (2002), “Measurement of Aggregate Risk With Copulas,” Preprint, Caesar.
- [28] Machina, M. and C. Granger (2000), “Forecast Comparison Without the Use of Cost Functions,” Working paper, UCSD.
- [29] Marcellino, M. (2002), “Model Selection for Nonlinear Dynamic Models,” Manuscript.
- [30] Mizon, G. and J. Richard (1986), “The Encompassing Principle and its Application to Testing Non-nested Hypotheses,” *Econometrica* 54, 657-678.
- [31] Nelsen, R. B. (1999), *An Introduction to Copulas*, Springer.
- [32] Patton, A.J. (2002), “Modeling Time-Varying Exchange Rate Dependence Using the Conditional Copula,” Working Paper 01-09, University of California, San Diego.
- [33] Patton, A. J. (2004), “On the Out-of-Sample Importance of Skewness and Asymmetric Dependence for Asset Allocation,” *Journal of Financial Econometrics* 2(1), 130-168.
- [34] Rivers, D. and Q. H. Vuong (2002), “Model Selection Tests for Nonlinear Dynamic Models,” *Econometric Review*.
- [35] Sin, C. and H. White (1996), “Information Criteria for Selecting Possibly Misspecified Parametric Models,” *Journal of Econometrics* 71, 207-225.
- [36] Su, L. and H. White (2003), “Multiple Forecast Comparison under General Loss Functions:  $\bar{\mathcal{X}}^2$  and Resampling Methods,” Working paper, UCSD.

- [37] Tsay, R. (2002), *Analysis of Financial Time Series*, Wiley Interscience: New York.
- [38] Vuong, Q. H. (1989), "Likelihood Ratio Test for Model Selection and Non-nested Hypotheses," *Econometrica* 57, 307-333.
- [39] West, K. (2001), "Encompassing Tests When no Model is Encompassing," *Journal of Econometrics* 105, 287-308.
- [40] White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica* 50, 143-161.
- [41] White, H. (1994), *Estimation, Inference and Specification Analysis*, Cambridge University Press, Cambridge, UK.
- [42] White, H. (2000), "A Reality Check for Data Snooping," *Econometrica* 68, 1097-1126.