Selection of Copulas with Applications in Finance*

Zongwu Cai\textsuperscript{a,b}, Xiaohong Chen\textsuperscript{c}, Yanqin Fan\textsuperscript{d}, and Xian Wang\textsuperscript{a}

\textsuperscript{a}Department of Mathematics & Statistics, University of North Carolina at Charlotte, Charlotte, NC 28223, USA, E-mails: zcai@uncc.edu (Z. Cai) and xiwang@uncc.edu (X. Wang)

\textsuperscript{b}The Wang Yanan Institute for Studies in Economics, Xiamen University, China

\textsuperscript{c}Department of Economics, Yale University, New Haven, CT 06520, USA, E-mail: xiaohong.chen@yale.edu

\textsuperscript{d}Department of Economics, Vanderbilt University, Nashville, TN 37240, USA, E-mail: yanqin.fan@vanderbilt.edu

First Draft: March 5, 2008

A fundamental issue of applying copula method in applications is how to choose an appropriate copula function. In this article we address this issue by proposing a new copula selection approach via penalized likelihood. The proposed method selects the appropriate copula functions and estimates copula coefficients simultaneously. The asymptotic properties, including the rate of convergence and asymptotic normality and abnormality, are established for the proposed penalized likelihood estimator. Particularly, when the true coefficient parameters may be on the boundary of the parameter space and the dependence parameters are in an unidentifiable subset of the parameter space, it shows that the limiting distribution for boundary parameters is abnormal and the penalized likelihood estimator for unidentified parameters converges to an arbitrary value. Moreover, the EM algorithm is proposed for optimizing penalized likelihood function. Finally, Monte Carlo simulation studies are carried out to illustrate the finite sample performance of the proposed method and the proposed method is used to investigate the correlation structure and comovement of financial stock markets.

Keywords: EM algorithm; Mixed copula; Penalized likelihood; SCAD; Variable selection.

*The authors thank Yoon-Jae Whang and other audiences at Sino-Korean Workshop 2007 in Xiamen for their constructive comments and suggestions.
1 Introduction

Copula approach has been used in many financial and economic fields recently, such as, asset pricing, risk management, portfolio and market value-at-risk calculations, and correlation structure and movement of the financial market. For example, Li (2000) was the first to study the default correlations in credit risk models by using the copula function to model the joint distribution between two default times. Bouye, Nikeghbali, Riboulet and Roncalli (2001), Embrechts, Lindskog and McNell (2003), and Cherubini, Vecchiato and Luciano (2004) applied the copula functions to measure the portfolio value-at-risk. Longin and Solnik (2001) examined cross-national dependence structure of asset returns in international financial market. Ang and Chen (2002) discovered asymmetric dependence between two asset returns during market downturns and market upturns. Although copula method has been applied to financial fields recently, it grows swiftly due to its several advantages. First, copulas are invariant to strictly increasing transformations of random variables. This property is very useful as transformations, such as log-transformation, are commonly used in financial analysis. Secondly, nonparametric dependence measures, like Kendall’s $\tau$ and Spearman’s $\rho$, are properties of copula functions. It is well known that economic and financial multivariate time series are typically nonlinear, abnormally distributed, and have nonlinear comovement. By using nonparametric measures, the assumptions like Gaussian distribution or linear dependence required for the correlation coefficient can be relaxed. Thirdly, asymptotic tail dependence measures dependence between extreme values. This property implies that copula functions enable us to model different patterns of dependence structures. Finally, by Sklar’s (1959) theorem, any multidimensional joint distribution function may be decomposed into its marginal distributions and a copula function which completely describes the dependence structure. Indeed, copulas allow us to model fat tails of the marginal distributions and tail dependence separately.

As aforementioned, copulas can be used to model the different patterns of dependence structures. For example, a Gaussian copula model has zero tail dependence, which means the probability that both variables are in their extremes is asymptotically zero unless their linear correlation coefficient
is unit. A Gumbel copula has a positive right tail dependence, i.e., the probability that both variables are in the right tails is positive. To capture different shapes of distributions and different patterns of dependence structure, researchers have considered a mixed copula which is a linear combination of several copula families. For example, Hu (2003) used a mixed copula approach with predetermined coefficients to measure the dependence pattern across financial markets. The mixture copula used in Hu (2003) is composed of a Gaussian copula, a Gumbel copula and a Gumbel survival copula. Clearly, a Gaussian copula is chosen based on the traditional approach and the other two copulas are selected to take into account of possible left and right tail dependence. Chollete, Pena, and Lu (2005) analyzed the co-movement of international financial markets by using a mixed copula model. They considered the same mixture as Hu (2003) and another mixture by replacing Gaussian copula with t-copula. The biggest advantage of using a mixed copula model is that it can nest different copula shapes, i.e., mixed copulas can capture different patterns of dependence structure. For example, if we consider a mixed copula which includes Gaussian and Gumbel copulas, it can improve a single Gaussian dependence structure by allowing possible right tail dependence. Therefore, empirically, a mixed copula is more flexible to model the dependence and can deliver better descriptions of dependence structure than an individual copula.

For applying the copula approach to solve real problems, it is important to choose appropriate parametric or nonparametric copulas since the distribution from which each data point is drawn is unknown. To attenuate this problem, there have been some efforts in literature to choose an appropriate individual copula, while Chen, Fan and Patton (2003) and Fermanian (2003) developed goodness-of-fit tests of an individual parametric copula. But there is no guidance as to which copula model should be used if these tests are used and the null hypothesis of correct parametric specification is rejected. Further, Hu (2003) considered a mixed copula by deleting the component if the corresponding weight is less than 0.1 or if the corresponding dependence measure is close to independence. To the best of our knowledge, so far no work with theoretical support has been attempted to choose the suitable copulas for a mixed copula.
In this paper, we propose a copula selection method via penalized likelihood. Initially, a large number of candidate copula families are of interest and their contributions to the dependence structure vary from one component to another. Our task is to select an appropriate copula or several copulas which can capture the dependence structure of the given data set among the candidate copulas. As we stated earlier, copula function can be used to describe the joint distribution. Therefore, this question can be solved by searching a copula or the mixture of several copulas that produce the highest likelihood. To select and estimate a mixed copula simultaneously, we formulate the problem of copula selection as the problem of variable selection. When the fitted mixed copula contains some component copulas which have small weight (small contribution to the dependence structure), we expect these components would not be in the mixed copula. This idea is in principle similar to the approach in Fan and Li (2001) and Fan and Peng (2004) based on a penalty function to delete the insignificant variables and to estimate the coefficients of significant variables in the context of regression settings. Chen and Khalili (2006) applied the variable selection method to the order selection in finite mixture models. In contrast, our approach can be formulated as the penalized likelihood function with an appropriate penalty function. By maximizing the penalized likelihood function, copulas with small weights are removed by a thresholding rule and parameters remained are estimated. In such a way, the model selection and parameter estimation can be done simultaneously. We show that this method is less computing intensive than many other existing methods which select an appropriate copula. Also the new method has a high probability of selecting an appropriate mixed copula model. Further, we establish the asymptotic properties of the proposed estimators. It is interesting to note that the general mathematical derivation for the asymptotic properties for the maximum likelihood estimator is not applicable here since the parameters may not be an interior point of the parameter space, which was addressed by Andrew (1999) for the case when the parameter is on a boundary for iid sample. Therefore, the challenges we face here are not only that the coefficient parameters are on the boundary of the parameter space but also that the dependent parameters are in a non-identifiable subset of the parameter
space. Thus, another two main contributions are, under this non-standard situation, we show that the estimate of non-identifiable dependent parameters converge to arbitrary value and we establish the abnormality of the boundary coefficient parameter estimate. Finally, to make the proposed methodology practically useful and applicable, we propose using the EM algorithm to find the penalized likelihood estimators. Also, we consider the data-driven methods for finding the threshold parameters in the penalty function. Further, we suggest an ad hoc method for a consistent estimator of the standard error.

The rest of this paper is organized as follows. Section 2 briefly reviews some basic facts about copulas and discusses the identification of the mixed copula model. In Section 3, we introduce the selection procedure based on the penalized likelihood. In Section 4, we list the regularity conditions and develop the asymptotic properties of the proposed estimators. The EM algorithm is outlined for finding the penalized likelihood estimators and the data-driven methods for finding the threshold parameters are discussed in Section 5, together with an ad hoc method for proposing a consistent estimator of the standard error. In Section 6, the results of a Monte Carlo study are reported to demonstrate the finite sample performance of the proposed methods, together with empirical analyses of real financial data. The technical proofs are collected in Section 7.

2 Mixed Copulas

2.1 Copulas

For simplicity of presentation, we only consider the two-dimensional case. Indeed, all methods and theory developed here continue to hold for multivariate case. The only difference for multivariate is the computing issue. We consider an i.i.d sequence \(\{(X_t, Y_t), t \in \mathbb{Z}\}\) taking values in \(\mathbb{R}^2\) with a realization of \(\{(X_t, Y_t); 1 \leq t \leq T\}\). We denote by \(f(x, y), F(x, y)\), the joint density and distribution of \(\{(X_t, Y_t)\}\). The joint distribution \(F(\cdot, \cdot)\) completely determine the probabilistic properties of \(\{(X_t, Y_t)\}\). \(F(\cdot, \cdot)\) can be expressed as a copula function between \(X_t\) and \(Y_t\) and marginal distribution functions of \(X_t\) and \(Y_t\). Let \(f_x(x)\) and \(f_y(y)\) be the marginal density of \(X_t\) and \(Y_t\). The marginal distribution of \(X_t\) and \(Y_t\) are written as \(F_x(x)\) and \(F_y(y)\), respectively. Now we would like to review
some formal copula and useful dependence concepts. For more discussions, the reader is referred to the books by Nelsen (1999) and Joe (1997).

**Definition 1:** A two-dimensional copula is a function $C(\cdot, \cdot): [0,1]^2 \rightarrow [0,1]$ with the following properties:

1. $C$ is grounded, i.e., for every $(u,v)$ in $[0,1]^2$, $C(u,v) = 0$ if at least one coordinate is 0.
2. $C$ is two-increasing, i.e., for every $a$ and $b$ in $[0,1]^2$ such that $a \leq b$, the C-volume $V_C([a,b])$ of the box $[a,b]$ is positive.
3. $C(u,1) = u$ and $C(1,v) = v$ for every $(u,v) \in [0,1]^2$.

One can think of a copula as a function which assigns any point in the unit square $[0,1] \times [0,1]$ a number in the interval $[0,1]$. From a probabilistic point of view, a copula function is a joint distribution whose marginal distributions are uniform. Next, we would like to state the famous Sklar’s (1959) theorem which links the univariate margins and the multivariate dependence structure.

**Sklar’s Theorem:** (Sklar (1959)). Let $F(x,y)$ be a bivariate distribution function with margins $F_x(x)$ and $F_y(y)$. Then, there exists a bivariate copula $C$ such that for all $(x,y)$ in $R^2$, 

$$F(x,y) = C(F_x(x), F_y(y)).$$

(1)

If $F_x(x)$ and $F_y(y)$ are all continuous, then $C$ is uniquely defined. Otherwise, $C$ is uniquely determined on $\text{Range}F_x(x) \times \text{Range}F_y(y)$. Conversely, if $C$ is a bivariate copula and $F_x(x)$ and $F_y(y)$ are distribution functions, then the function $F(x,y)$ defined by (1) is a bivariate distribution function with margins $F_x(x)$ and $F_y(y)$.

As an immediate result of Sklar’s theorem, $F(x,y)$ can be expressed in terms of the marginal distribution functions of $X_t$ and $Y_t$ and the copula function of $X_t$ and $Y_t$. That is, 

$$F(x,y) = C(F_x(x), F_y(y)).$$
More specifically, if we consider a parametric model in which the marginal distributions $F_x(x)$ and $F_y(y)$ depend on the distinct parameters $\alpha$ and $\beta$, we have

$$F(x, y; \alpha, \beta, \theta) = C(F_x(x; \alpha), F_y(y; \beta); \theta),$$  \hspace{1cm} (2)

where $\theta$ is the dependence parameter in copula. To avoid confusion, we need to note that $x$ and $y$ are used to denote the actual observations, such as the returns in different markets. And we use $u$ and $v$ to denote the marginal distribution of $x$ and $y$. So $x$ and $y$ are any real numbers but $u, v \in [0, 1]$.

Another useful concept in copula field is tail dependence. It measures the dependence between two variables in the upper and lower joint tail values of two variables.

**Definition 2:** Let $X$ and $Y$ be continuous random variables with copula $C$ and marginal distribution functions $F_x(x)$ and $F_y(y)$. The coefficients of upper and lower tail dependence of $(X_t, Y_t)$ are defined as

$$\tau_U = \lim_{v \to 1} Pr(F_y(y) > v | F_x(x) > v) = \lim_{v \to 1} \frac{1 - 2v + C(v, v)}{1 - v}$$

and

$$\tau_L = \lim_{v \to 0} Pr(F_y(y) \leq v | F_x(x) \leq v) = \lim_{v \to 0} \frac{C(v, v)}{v}$$

provided that the limit $\tau_U \in [0, 1]$ and $\tau_L \in [0, 1]$ exist.

Tail dependence is a copula property and the amount of tail dependence is invariant under strictly increasing transformations of $X$ and $Y$. If $\tau_U \in (0, 1]$, $X$ and $Y$ are asymptotically dependent in the upper tail; if $\tau_U = 0$, $X$ and $Y$ are asymptotically independent in the upper tail. Similarly, if $\tau_L \in (0, 1]$, $X$ and $Y$ are asymptotically dependent in the lower tail; if $\tau_L = 0$, $X$ and $Y$ are asymptotically independent in the lower tail.

Now we consider the case that the copula function in equation (2) is a mixture of several copula families. Then the mixed copula function can be formulated as follows:

$$C(u, v; \theta_c) = \sum_{k=1}^{s} \lambda_k C_k(u, v; \theta_k),$$  \hspace{1cm} (3)
where $\{C_k(u, v, \theta_k)\}$ is a sequence of known copulas with unknown parameters $\{\theta_k\}$ and $\{\lambda_k\}_{k=1}^s$ are the weights satisfying $0 \leq \lambda_k \leq 1$ and $\sum_{k=1}^s \lambda_k = 1$ for $k = 1, \ldots, s$. Denote the whole parameters in the mixed copula as $\theta_c$, that is, $\theta_c = (\theta_1, ..., \theta_p, \lambda_1, ..., \lambda_s)^T$. It is easy to see that $C(u, v)$ is also a copula. Let $\theta = (\theta_1, ..., \theta_s)^T$ be the associate parameters in the mixture which present the degree of dependence, and $\lambda = (\lambda_1, ..., \lambda_s)^T$ be weights or shape parameters reflecting the credence we should place in the corresponding copula. For example, if the $m$-th copula is the most appropriate structure for the data, then we would expect that $\lambda_m$ is close to 1.

Let $r$ be the true number of copulas included in the mixed copula. The true value of $r$ is the smallest possible values such that all the component copulas are different and the mixing proportion $\lambda_k$’s are non-zero. Thus, we can denote the true mixed copula as

$$C_0(u, v; \theta_{c0}) = \sum_{k=1}^r \lambda_{0k} C_k(u, v; \theta_{0k}), \quad (4)$$

where $\theta_{0k}$ is the true dependence parameter and $\lambda_{0k}$ is the true weight. Note that when $r = 1$, we just consider a single parametric copula. A potential issue of the mixed copulas is their identifiability.

### 2.2 Identification

**Definition 3:** For any two mixed copulas

$$C(u, v; \theta_c) = \sum_{k=1}^s \lambda_k C_k(u, v; \theta_k), \quad \text{and} \quad C^*(u, v; \theta_c^*) = \sum_{k=1}^{s^*} \lambda_k^* C_k^*(u, v; \theta_k^*)$$

with parameters $\theta_c = (\theta, \lambda)^T$ and $\theta_c^* = (\theta^*, \lambda^*)^T$, the mixed copula model is said to be identified if

$$C(u, v; \theta_c) \equiv C^*(u, v; \theta_c^*)$$

if and only if $s = s^*$ and we can order the summations such that

$$\lambda_k = \lambda_k^* \quad \text{and} \quad C_k(u, v; \theta_k) = C_k^*(u, v; \theta_k^*), \quad k = 1, \ldots, s.$$
It is obvious that mixed copula is similar to the standard finite mixture model. For more discussions about finite mixture model, we refer to the book by McLachlan and Peel (2000). But we need to note that mixed copula is different from finite mixture model in that mixed copula contains different copula families with different density forms. Similar to finite mixture model, identifiability of mixed copula model depends on the component copulas and the number of components. In this paper, the identifiability is not a key issue for us. We assume throughout the paper that the copula model is identifiable.

3 Copula Selection via Penalized likelihood

In this section, we present the selection and estimation procedures of mixed copula. First, we assume that $F_x(x)$ and $F_y(y)$ are known and have a finite number of unknown parameters, denoted by $F_x(x, \alpha)$ and $F_y(y, \beta)$, respectively.

3.1 Penalized Likelihood

In what follows, we use $\phi$ to denote all the parameters. That is, $\phi = (\alpha^T, \beta^T, \lambda^T, \theta^T)^T$. According to the mixed copula defined in (3), the distribution function based on mixture copula can be written as

$$F(x, y; \phi) = \sum_{k=1}^s \lambda_k C_k(F_x(x; \alpha), F_y(y; \beta); \theta_k)$$

and the joint density function is given by

$$f(x, y; \phi) = f_x(x; \alpha)f_y(y; \beta) \left[ \sum_{k=1}^s \lambda_k c_k(F_x(x; \alpha), F_y(y; \beta); \theta_k) \right],$$

where $c_k(u, v; \theta_k) = \partial C_k(u, v; \theta_k)/\partial u\partial v$ is the mixed partial derivative of the copula $C$ and $c(\cdot, \cdot)$ is called the copula density. Then we can re-write the log-likelihood function of observed random pairs $(x_t, y_t), t = 1, \ldots, T$ as follows:

$$L(\phi) = \sum_{t=1}^T [\ln f_x(x_t; \alpha) + \ln f_y(y_t; \beta)] + \sum_{t=1}^T \ln \left[ \sum_{k=1}^s \lambda_k c_k(F_x(x_t; \alpha), F_y(y_t; \beta); \theta_k) \right].$$

Clearly, the first summand is the logarithm of the likelihood for the iid sample for marginal parameters and the second one is the logarithm of the likelihood for the dependence parameters. The
penalized log-likelihood takes the following form with a Lagrange multiplier term

\[ Q(\phi) = L(\phi) - T \sum_{k=1}^{s} p_{\gamma T}(\lambda_k) + \delta \left( \sum_{k=1}^{s} \lambda_k - 1 \right), \]  

where \( p_{\gamma T}(\cdot) \) is a non-concave penalized function and \( \{\gamma_T\} \) are the tuning parameters. \( \{\gamma_T\} \) control the complexity of model and can be selected by the data-driven methods, such as the cross-validation (CV) and the generalized cross-validation (GCV); see, for example, Fan and Li (2001). To avoid overfitting, the penalty functions are applied only to weight parameters \( \lambda_k \) since some weight parameters might be estimated as zero if they are not significant in the copula model. Therefore, we can delete the corresponding copula functions with very small values of weight parameters, whereas others are not. In such a way, the copulas are selected and the parameters of copulas are estimated simultaneously. We show that the estimator achieves the so-called “oracle” and “sparsity” properties; see later for details. The third term in equation (5) is added for the constraint on \( \lambda_k \).

For simplicity of presentation, we assume that the penalty function \( p_{\gamma T}(\cdot) \) is the same for all \( \lambda_k \). In the literature, there are many penalty functions available, like the popular family of \( L_p \) \((p \geq 0)\) type penalty functions. For example, \( p = 0 \) corresponds to the entropy penalty, the \( L_1 \) penalty \( p_1(|\eta|) = \gamma |\eta| \) yields the least absolute shrinkage and selection operator (LASSO) proposed by Tibshirani (1996), and \( L_2 \) penalty \( p_2(|\eta|) = \gamma |\eta|^2 \) results in a ridge regression initiated by Frank and Friedman (1993) and Fu (1998). Hard thresholding penalty function

\[ p_\gamma(|\eta|) = \gamma^2 - (|\eta| - \gamma)^2I(|\eta| < \gamma) \]

was used by Antoniadis (1997) and Fan (1997). Finally, the smoothly clipped absolute deviation (SCAD) penalty was proposed by Fan (1997) and is given by

\[ p'_\gamma(\eta) = \gamma I(\eta \leq \gamma) + \frac{(a\gamma - \eta)_+}{(a - 1)} I(\eta > \gamma) \]

for some \( a > 2 \) and \( \eta > 0 \), and \( p_\gamma(0) = 0 \). As shown in Fan and Li (2001), SCAD penalty function leads to estimator with the three desired properties which can not be achieved by either \( Q \) penalty function or hard penalty function. Three properties are: unbiasedness for large true coefficient to
avoid unnecessary estimation bias, sparsity of estimating a small coefficient as 0 to reduce model complexity, and continuity of resulting estimator to avoid unnecessary variation in model prediction. More discussions about these properties can be found in the papers by Antoniadis and Fan (2001) and Fan and Li (2001). Therefore, we use the SCAD penalty in this paper.

3.2 Estimation

To estimate the parameters, we propose using a full maximum likelihood approach. We maximize the logarithm of penalized likelihood function $Q(\phi)$ with respect to all the parameters $\phi$. The full maximum likelihood estimate is denoted as $\hat{\phi}$. Since the maximum likelihood estimate does not have a close form, an iterative algorithm is used to compute the numerical solution and an initial value is suggested as follows. A two-step estimation is used as the initial value of the maximum likelihood procedure. A two-step procedure is to estimate the univariate parameters from univariate likelihood and then to optimize the multivariate likelihood with univariate parameter replaced by estimators from the first step. More specifically, at the first step, we estimate the marginal parameters by maximizing the likelihood corresponding to the marginal models:

$$
\sum_{t=1}^{T} \ln f_x(X_t; \alpha) + \sum_{t=1}^{T} \ln f_y(Y_t; \beta),
$$

which is not affected by the copula parameters $\theta_c$. Let $\hat{\theta}_m = (\hat{\alpha}_m, \hat{\beta}_m)^T$ denote the solution of the above optimization problem. As demonstrated in Joe (1997), $\hat{\alpha}_m$ and $\hat{\beta}_m$ are $\sqrt{T}$ consistent. Then at the second step, we substitute $\alpha$ and $\beta$ by their estimators in equation (5). Hence we obtain a penalized likelihood function

$$
Q(\theta_c) = L(\hat{\theta}_m, \theta_c) - T \sum_{k=1}^{s} p_{\gamma T}(\lambda_k) + \delta \left( \sum_{k=1}^{s} \lambda_k - 1 \right).
$$

Maximizing $Q$ with respect to $\theta$ results in penalized likelihood estimators $\hat{\theta}_c^T = (\hat{\theta}^T, \hat{\lambda}^T)$. In next section, we demonstrate the consistency of $\hat{\phi}$ and derive the limiting distribution of $\hat{\phi}$. 


4 Asymptotic Properties

We next investigate the asymptotic behavior of the penalized likelihood estimators. An outline of regularity conditions and asymptotic results are given in the following subsection.

4.1 Notations and Assumptions

Before we proceed with the asymptotic theory of the proposed estimators, we list all assumptions used for the asymptotic theory. The proof of the theorems presented in this section can be found in Section 7 with some lemmas and their detailed proof. Notation used in the following section is defined here.

We define \( \Lambda = \{ (\lambda_1, \ldots, \lambda_s) \mid \lambda_i \geq 0, \sum_{i=1}^{s} \lambda_i = 1, \ i = 1, \ldots, s \} \), which is a subset of \([0, 1]^s\), and \( \Theta = \Theta_1 \times \cdots \times \Theta_s \), where \( \Theta_i \subset \mathbb{R}^{d_i} \) for \( i = 1, \ldots, s \). Then, \( \Theta \subset \mathbb{R}^{d_\theta} \), where \( d_\theta = \sum_{i=1}^{s} d_i \).

The parameter space is given by \( \Gamma = \Theta_\alpha \times \Theta_\beta \times \Lambda \times \Theta \), where \( \Theta_\alpha \subset \mathbb{R}^{d_\alpha} \) and \( \Theta_\beta \subset \mathbb{R}^{d_\beta} \). Thus, \( \Gamma \subset \mathbb{R}^{d_\gamma} \), where \( d_\gamma = d_\alpha + d_\beta + s + d_\theta \).

We use \( \alpha_0, \beta_0, \) and \( \lambda_0 \) to denote the true value of \( \alpha, \beta, \) and \( \lambda \). Further, we partition \( \lambda_0 = (\lambda_{01}, \ldots, \lambda_{0s})^T \) as \( \lambda_{10} = (\lambda_{01}, \ldots, \lambda_{0r})^T \) and \( \lambda_{20} = (\lambda_{0(r+1)}, \ldots, \lambda_{0s})^T \). Without loss of generality, we assume that \( \lambda_{10} \) consists of all nonzero components of \( \lambda_0 \), and \( \lambda_{20} \) contains all zero components and is on the boundary of the parameter space. Similarly, \( \theta_{10} = (\theta_{01}, \ldots, \theta_{0r})^T \) is a vector of true dependence parameters corresponding to \( \lambda_{10} \) and \( \theta^*_2 = (\theta^*_1, \ldots, \theta^*_s)^T \) is a vector of dependence parameters corresponding to \( \lambda_{20} \) and can be arbitrary values in the parameter space. Thus, we define

\[
\phi_0 = (\alpha_0^T, \beta_0^T, \lambda_0^T, \theta_0^T)^T,
\]

where \( \lambda_0 \) is the true weights corresponding to the copulas and \( \theta_0 \) includes the true dependent parameters corresponding to the non-zero weight copulas and arbitrary values related to the zero weight copulas. We use \( \Gamma_0 \) to denote the collection of all the true values of \( \phi_0 \).

In the general framework of the likelihood estimator, the true parameter is commonly assumed to be an interior point of the parameter space. The classical assumptions and standard procedures
can be applied to derive the asymptotic properties. In our model, the true parameter may not be an interior point of the parameter space, thus we can not follow the standard procedures to derive the asymptotic properties. Now, let us take a look at the true parameter in our case. Firstly, we consider $r = 1$, that is, the mixed copula includes only one component. In this case, we can find that $\lambda_0 = \{\lambda_{01} = 1, \lambda_{02} = \cdots = \lambda_{0s} = 0\}$ and $\theta_0 = \{\theta_{01}, \theta^*_2 \in \Theta_2, \cdots, \theta^*_s \in \Theta_s\}$. It is obvious that $\lambda_0$ is on the boundary of parameter space $\Lambda$ and the dependence parameters associated to the zero weight corresponds to non-identifiable subset of the parameter space $\Theta$. Secondly, if $s > r > 1$, we have

$$\lambda_0 = \left\{\lambda_{01}, \ldots, \lambda_{0r}, \lambda_{0(r+1)} = \cdots = \lambda_{0s} = 0, \lambda_{0i} > 0, \sum_{i=1}^{r} \lambda_{0i} = 1, i = 1, \ldots, r\right\}$$

and $\theta_0 = \{\theta_{01}, \cdots, \theta_{0r}, \theta^*_{(r+1)} \in \Theta_{r+1}, \cdots, \theta^*_s \in \Theta_s\}$. We can observe that a subset of $\lambda_0$ is on the boundary of the parameter space and a subset of $\theta_0$ is non-identifiable. Finally, if $r = s$, $\phi_0$ is an interior point of parameter space $\Gamma$. In the first two cases, the classical assumptions about the asymptotic properties of the maximum likelihood estimator are not valid and the standard asymptotic theory can not be applied here. We extend the standard results by considering a situation where the true parameter lies in a non-identifiable subset, and this subset may be on the boundary of the parameter space.

We need the following assumptions, with $\phi_0$ an arbitrary fixed point in $\Gamma_0$:

**Assumption A:**

A1. $(X_t, Y_t)$ has a joint density $f(x, y; \phi) = f_x(x; \alpha)f_y(y; \beta)c(x, y; \lambda, \theta)$ and $f(x, y; \phi)$ has a common support as well as the model is identified.

A2. $f(x, y, \phi_0) = f(x, y, \phi_0^*)$ for all $\phi_0, \phi_0^* \in \Gamma_0$.

A3. There exists an open subset $Q_\varepsilon \subset R^p$ containing $\Gamma_0$ such that, for almost all $(x, y)$, $f_x(x; \alpha), f_y(y; \beta)$, and $f(x, y; \phi)$ admit all third derivatives with respect to $\alpha, \beta$, and $\phi$, respectively. Also, we suppose that there exist functions $M_{jkl}(x, y; \phi)$ such that

$$\left| \frac{\partial^3}{\partial \phi_j \partial \phi_k \partial \phi_l} \{\log f(x, y; \phi)\} \right| \leq M_{jkl}(x, y; \phi)$$
for all j, k, and l, and x and y, and there exists a constant C such that \( m_{jkl}(x, y; \phi) = E_{\phi_0} \left[ M^2_{jkl}(x, y; \phi) \right] < C \) for any fixed \( \phi_0 \in \Gamma_0 \).

A4. For any \( \phi_0 \in \Gamma_0 \), the second logarithmic derivatives of \( f(x, y; \phi) \) satisfy the equations

\[
I_{jk}(\phi_0) = E_{\phi_0} \left[ \frac{\partial}{\partial \phi_j} \left\{ \log f(x, y; \phi) \right\} \frac{\partial}{\partial \phi_k} \left\{ \log f(x, y; \phi) \right\} \right] = -E_{\phi_0} \left[ \frac{\partial^2}{\partial \phi_j \partial \phi_k} \left\{ \log f(x, y; \phi) \right\} \right],
\]

and \( I_{jk}(\phi_0) \) is finite. Furthermore, the Fisher information matrix

\[
I(\phi) = E \left\{ \left[ \frac{\partial}{\partial \phi_j} \log f(x, y; \phi) \right] \left[ \frac{\partial}{\partial \phi_k} \log f(x, y; \phi) \right]^T \right\}
\]

is positive definite at \( \phi = \phi_0, \phi_0 \in \Gamma_0 \).

**Remark 1:** In general, \( Q_\varepsilon \) can be expressed as \( Q_\varepsilon \equiv \bigcup_{\phi_0 \in \Gamma_0} B_\varepsilon(\phi_0) \) for each \( \varepsilon(\phi_0) > 0 \) depending on \( \phi_0 \), and \( B_\varepsilon(\phi) \) is an open ball of radius \( \varepsilon \) centered at \( \phi \).

**Remark 2:** We can show that \( E_{\phi_0} \left[ \frac{\partial}{\partial \phi_j} \left\{ \log f(x, y, \phi) \right\} \right] = 0 \) for \( j = 1, \cdots, s \) and all \( \phi_0 \in \Gamma_0 \).

### 4.2 Large Sample Theory

Firstly, we establish the convergence rate of the penalized likelihood estimator. To this end, define

\[
b_T = \max_{1 \leq k \leq s} \{ p'_{\gamma k}(\lambda_{0k}), \lambda_{0k} \neq 0 \}. \quad \text{Also, denote}
\]

\[
\Sigma = \text{diag}\{0, \cdots, 0, p''_{\gamma 1}(\lambda_{01}), \cdots, p''_{\gamma s}(\lambda_{0s}), 0, \cdots, 0\}_{d_\gamma \times d_\gamma},
\]

and

\[
b = (0, \cdots, 0, p'_{\gamma 1}(\lambda_{01}), \cdots, p'_{\gamma s}(\lambda_{0s}), 0, \cdots, 0)_{d_\gamma \times 1}^T.
\]

Furthermore, we define

\[
\phi_I = (\alpha^T, \beta^T, \lambda^T, \theta_1^T)^T,
\]

where \( \phi_I \) includes all the identified parameters. Thus, we have

\[
\phi = (\phi_I^T, \theta_2^T)^T.
\]
Theorem 1: Under the regularity conditions A1 – A4, if \( \max_{1 \leq k \leq s} \{ |\beta_T^\gamma(\lambda_0)|, \lambda_0 \neq 0 \} \) tends to 0, for any given point \( \phi_0 \in \Gamma_0 \), then, there exists a local maximizer \( \hat{\phi} = (\hat{\phi}_I^T, \hat{\theta}_2^T)^T \) of \( Q(\phi) \) defined in equation (6) such that

\[
\hat{\phi}_I \rightarrow \phi_{I0}, \text{ in probability},
\]

where \( \phi_{I0} \) is the true value of \( \phi_I \), and for arbitrary point \( \theta_2^* \),

\[
\hat{\theta}_2 \rightarrow \theta_2^*, \text{ in probability},
\]

Moreover, rate of convergence of \( \hat{\phi} \) is \( T^{-1/2} + b_T \). 

Remark 3: We need to notice that, for any given point \( \phi_0 = (\phi_{I0}^T, \theta_2^T)^T \in \Gamma_0 \), \( \phi_{I0} \) is the true value and \( \theta_2^* \) can be arbitrary value in a non-identified subset. Therefore, non-identifiable dependence parameter estimate \( \hat{\theta}_2 \) which correspond to zero weights converge to arbitrary value and \( \hat{\phi}_I \) converge to the true value \( \phi_{I0} \).

Theorem 1 demonstrates that the estimator can achieve the root-\( n \) convergence rate when \( b_T = O_p(T^{-1/2}) \). When \( \gamma_T \rightarrow 0 \), \( b_T = 0 \) for SCAD penalty function. therefore, the penalized likelihood estimate correctly estimates some weights 0 with positive probability. This leads to the oracle property, which introduced by Donoho and Johnstone (1994).

Next, we present the oracle property for the penalized likelihood estimator. Before we proceed to the oracle property, we need the following notations. Let \( \phi_1 = (\alpha^T, \beta^T, \lambda_1^T, \theta_1^T, \theta_2^T)^T \) without zero weights. We use \( \phi_r \) to denote the parameters after the parameters corresponding to the zero weights are removed and they are re-ordered. That is, \( \phi_r = (\lambda_1^T, \theta_1^T, \alpha^T, \beta^T)^T \) without zero weights and unidentified parameters. And let \( q \) be the number of parameters included in \( \phi_r \), that is \( q = r + \sum_{k=1}^{r} d_k + d_\alpha + d_\beta \). Also, denote,

\[
\Sigma_1 = diag\{ p_{\gamma r}^\gamma(\lambda_{01}), \ldots, p_{\gamma r}^\gamma(\lambda_{0r}), 0, \ldots, 0 \}_{q \times q}
\]

and

\[
b_1 = (p_{\gamma r}^\gamma(\lambda_{01}), \ldots, p_{\gamma r}^\gamma(\lambda_{0r}), 0, \ldots, 0)^T_{q \times 1}.
\]
Definition 4: The set $\Gamma$ is approximated at $\phi_0$ by a cone $C_\Gamma$ with vertex at $\phi_0$ if
\[
\inf_{x \in C_\Gamma} \|x - y\| = o(\|y - \phi_0\|) \text{ for all } y \in \Gamma \quad \text{and} \quad \inf_{y \in \Gamma} \|x - y\| = o(\|x - \phi_0\|) \text{ for all } x \in C_\Gamma.
\]

Theorem 2: Under the regularity conditions A1 – A4, if $b_T = O_p(T^{-1/2})$, $\sqrt{T}\gamma_T \to \infty$, and
\[
\lim \inf_{T \to \infty} \lim \inf_{\lambda_k \to 0^+} p'_{\gamma_T}(\lambda_k)/\gamma_T > 0,
\]
the root-$T$ consistent estimator $\hat{\phi} = (\hat{\phi}_1, \hat{\lambda}_2)^T$ in Theorem 1 must satisfy the sparsity, that is
\[
\hat{\lambda}_2 = 0.
\]

(i) When $\lambda_{10} = (\lambda_{01}, \ldots, \lambda_{0r})$, $r > 1$ and $\lambda_{0i} \in (0, 1)$ for $i = 1, \ldots, r$,
\[
\sqrt{T}[\{I_1(\phi_{r0}) + \Sigma_1\}(\hat{\phi}_r - \phi_{r0}) + b_1] \to N(0, I_1(\phi_{r0})).
\]

(ii) When $\lambda_{10} = \lambda_{01} = 1$,
\[
\sqrt{T}[I_1(\phi_{r0}) + \Sigma_1](\hat{\phi}_r - \phi_{r0}) = \hat{\phi}_r + o_p(1),
\]
where the limiting random variable $\hat{\phi}_r$ has the following representation
\[
\begin{pmatrix}
Z_{11} - b_{11} \\
Z_{12} - b_{12} \\
\vdots \\
Z_{1q} - b_{1q}
\end{pmatrix} I\{Z_{11} > b_{11}\} + \begin{pmatrix}
Z_{12} - b_{12} - (I_{11}^{21}/I_{11})(Z_{11} - b_{11}) \\
\vdots \\
Z_{1q} - b_{1q} - (I_{11}^{q1}/I_{11})(Z_{11} - b_{11})
\end{pmatrix} I\{Z_{11} < b_{11}\},
\]
where $Z_1 = (Z_{11}, Z_{12}, \ldots, Z_{1q})^T$ be a random variable with multivariate Gaussian distribution with mean 0 and covariance matrix $I_1(\phi_{r0})$, $\phi_r \in C_\Gamma - \phi_{r0}$, $I_{1i}^{ij} = I_{11}^{ij}(\phi_{r0})$ are elements of matrix $I_1^{-1}(\phi_{r0}) + \Sigma_1$ and $I_1(\phi_r)$ is the Fisher information when all zero effects are removed.

Now, we examine the asymptotic behavior for boundary parameters. When $r = 1$, it follows from Theorem 2 that $\hat{\phi}_{r1} = (Z_{11} - b_{11})I\{Z_{11} > b_{11}\}$, which is half-normal. In other words, when $Z_{11} > b_{11}$, the first component of the limiting distribution is normal and zero otherwise. Further, if $I_{11}^{21} = 0$, the second component of the previous equation is normal as well as the last two components. For the SCAD thresholding penalty function, $b_T = 0$ if $\gamma_T \to 0$. Thus, when
\( \sqrt{T} \gamma T \to \infty \), by Theorem 2, the corresponding penalized likelihood estimators preserve the oracle property and perform as well as the maximum likelihood estimators for estimating \( \lambda_1 \) if we would know \( \lambda_2 = 0 \) in advance.

**Remark 4:** We comment that all the asymptotic results in the above theorems would hold true for time series context with a change of the asymptotic variance. For details, see Cai and Wang (2008) for a more general setting under time series context.

## 5 Numerical Solutions

### 5.1 Choice of the tuning parameters

To implement the approach proposed in Section 3, we need to choose the appropriate tuning parameters \( \gamma = \gamma T \) and \( a \) in the SCAD penalty function. In implementation, we suggest using the multi-fold cross-validation method of estimating \( \gamma \) and \( a \) as suggested by Cai, Fan and Yao (2000) and Fan and Li (2001) for regression settings. Next, we describe the details of the cross-validation procedure.

Denote \( D \) as the full data set. Let \( D_i, i = 1, \ldots, m \) be the subset of \( D \). We treat \( D - D_i \) as the cross-validation training set and \( D_i \) as test set. That is, for each pair of \( (\gamma, a) \), \( D - D_i \) is used to estimate \( \phi \) and \( D_i \) is used for evaluation. The penalized maximum likelihood estimator \( \hat{\phi}_i \) can be used to construct the following cross-validation criterion based on the test data \( D_i \)

\[
CV(\gamma, a) = \sum_{i=1}^{m} \sum_{(x_t, y_t) \in D_i} L(\hat{\phi}_i).
\]

By maximizing \( CV(\gamma, a) \), the data-driven choice of tuning parameters is selected. As suggested by Cai, Fan and Yao (2000) and Fan and Li (2001), we may choose \( m \) to 4 or 5 in implementation and we take \( m = 5 \) in our empirical studies in Section 6.

### 5.2 Maximization of penalized likelihood function

It is clear that there is not an explicit expression for maximum likelihood estimator of equation (5). We need to use numerical methods in implementation. One of the most popular algorithms for
maximum likelihood estimation of mixture models is the expectation maximization (EM) algorithm of Dempster, Laird and Rubin (1977). EM algorithm is numerically simple to implement. We now explain this algorithm in detail. Note that our likelihood expression is quite different from the standard mixture models since copulas come from different families. The main idea of EM algorithm is to decompose the optimization step into two steps: E-step computes and updates the conditional probability that our observations come from each component copula, and M-step maximizes the penalized log likelihood to estimate the dependence parameters.

In order to maximize equation (5), we take the derivative and set it equal to zero. The only closed-form expression we can get is the weight expression. We consider the estimate of weights now, we can express the first order condition of weights as follow:

$$\frac{\partial Q(\phi)}{\partial \lambda_k} = \sum_{t=1}^{T} \frac{c_k(u, v; \theta_k)}{f(x, y; \phi)} - T p'_{\gamma_T}(\lambda_k) - \delta = 0,$$

which implies that

$$\sum_{t=1}^{T} \frac{\lambda_k c_k(u, v; \theta_k)}{f(x, y; \phi)} - T \lambda_k p'_{\gamma_T}(\lambda_k) - \lambda_k \delta = 0.$$

By summing over equation (7) for all $k$, we get

$$\delta = T \left[ 1 - \sum_{k=1}^{s} \lambda_k p'_{\gamma_T}(\lambda_k) \right].$$

Plugging $\delta$ back to equation (7) leads to

$$\lambda_k = \left[ z_k - T \lambda_k p'_{\gamma_T}(\lambda_k) \right] / \delta,$$

where $z_k = \sum_{t=1}^{T} c_k(u, v; \theta_k) / f(x, y; \phi)$.

### 5.2.1 E-step

Let $\hat{\phi}^{(0)}$ and $\{\hat{\phi}^{(m)}\}$ be the initial value and a sequence of estimates of the parameters at each iteration. At the expectation step, we can estimate the shape parameters by

$$\lambda_k^{(m+1)} = \left[ \sum_{t=1}^{T} \frac{c_k(u, v; \theta_k^{(m)})}{f(x, y; \phi^{(m)})} - T \lambda_k p'_{\gamma_T}(\lambda_k^{(m)}) \right] \delta^{-1}.$$
5.2.2 M-step

What distinguishes the estimation of the mixed copula from most of the other mixture models is the M-step. The marginal parameters and dependent parameters are updated by solving the following equations for any given estimates $\lambda_{k}^{(m)}$

$$\frac{\partial Q(u,v;\theta_{m},\theta,\lambda_{k}^{(m)})}{\partial \theta_{k}} = 0.$$ 

Newton-Raphson method is used here to estimate the marginal parameters and dependence parameters since no close-form is available for the estimate of $\theta_{m}$ and $\theta$.

5.3 Estimation of Standard Error

In this section, we discuss the subsample procedure proposed Politis and Romano (1994) and addressed by Andrews (1999) to obtain consistent standard error estimators whether the parameter is on a boundary or not. This method is applicable in iid context, see Section 2 of Politis and Romano (1994), as well as in stationary time series contexts which was discussed in Section 3 of Politis and Romano (1994). Under certain assumptions of the subsample size $b$ (which require $b \to \infty$ and $b/T \to 0$), the recomputed estimate over the subsamples of the data set can be used to approximate the true limit distribution. See Politis and Romano (1994) for the detailed discussion.

6 Monte Carlo Simulations and Real Applications

6.1 Simulated Examples

To illustrate the methods proposed earlier, we consider two simulated examples. Firstly, we describe the data generating process (DGP). The DGP is a process that the bivariate joint distribution has a form of copula function and individual variables are normally distributed. The bivariate joint distribution can be specified as

$$(u_t, v_t) \sim iid \ C(u,v;\theta_c),$$

where $C(u,v,\theta_c)$ can be a single copula function or mixed copula. The marginal distributions are normally distributed with marginal parameters $(\mu_x, \sigma_x)$ and $(\mu_y, \sigma_y)$. 
Our candidate copula families include four commonly used copulas: Gaussian, Clayton, Gumbel, and Frank copulas. The Gaussian copula is widely used in financial fields and has symmetric dependence in both tails of the distribution. Similarly, variables drawn from the Frank copula also exhibit symmetric dependence in both tails. However, compared to the Gaussian copula, dependence in the Frank copula is weaker in both tails and stronger in the center of the distribution, as is evident from the fanning out in the tails. This suggests that the Frank copula is best suited for applications in which tail dependence is relatively weak. In contrast to the Gaussian and Frank copulas, the Clayton and Gumbel copulas exhibit asymmetric dependence. Clayton dependence is strong in the left tail. The implication is that the Clayton copula is best suited for applications in which two variables are likely to decrease together. On the other hand, the Gumbel copula exhibits strong right tail dependence. Consequently, as is well-known, Gumbel is an appropriate modeling choice when two variables are likely to simultaneously increase. The simulation results show that the right copula can be selected although some of them have the same type of dependence structure.

The mixed copula based on the candidate copulas can be written as:

$$C(u, v; \theta_c) = \lambda_1 C_{Ga}(u, v; \theta_1) + \lambda_2 C_{Cl}(u, v; \theta_2) + \lambda_3 C_{Gum}(u, v; \theta_3) + \lambda_4 C_{Fra}(u, v; \theta_4).$$

**Example 1:** In this example, we consider the case that the joint distribution can be expressed by a single copula function. The marginal distributions are Gaussian with parameters $(\mu_x, \sigma_x)$ and $(\mu_y, \sigma_y)$ with values given in Table 1. Three sample sizes, 400, 700, and 1000, are considered and the observations are simulated from single Gaussian copula by following the GDP proposed above and the penalized maximum likelihood estimators are estimated. The simulation is repeated 100 times. The same procedures are performed for the other three copulas: Clayton, Gumbel, and Frank. Thus, the $j$th model in this example has the form of equation (9) with $\lambda_j = 1$ and $\lambda_i = 0$ for all $i \neq j$.

The simulated results are summarized in Tables 1, 2, 3, and 4. If the estimated weight of the copula component is not zero, this copula component is selected, otherwise it is removed from the mixed copula. In Table 3, the values without parentheses correspond to the percentages that
correct copulas are chosen, i.e., to the number of replications out of 100 when \( \lambda_j = 1 \) and \( \hat{\lambda}_j \neq 0 \). The values with parentheses correspond to the percentages that copulas are selected incorrectly, i.e., to the number of replications out of 100 when \( \hat{\lambda}_j \neq 0 \), but actually \( \lambda_j = 0 \). Tables 1 and 2 show the true values of the marginal parameters and dependence parameters from the copula, biases and MSEs of the penalization maximum likelihood estimators. It shows clearly that the bias becomes smaller when the sample size is getting larger. Table 4 shows the initial values and the estimated values of unidentified parameters and we can find the estimators converge to the initial values. In other words, the penalized likelihood estimate of an unidentified parameter converges to an arbitrary value. This is in line with our theory.

We can observe that performance of proposed method is good for choosing an appropriate copula from a mixed model. For each model, we have the 100% of chance to choose the correct copula function from which each data point is drawn. The percentage that incorrect copula is selected is small. For the Gaussian copula, only 3% of chance to select Frank copula when \( n = 1000 \) and 8% of chance to choose the Gaussian copula for those data points generated from the Frank copula. This is not surprising because the Frank and Gaussian copulas have the same type of dependence structure as that it is not easy to distinguish them.

**Example 2:** Now we consider the performance of the estimators when the joint distribution has a form of mixed copula. We use the same marginal parameter values as those in Example 1. We consider several mixed copulas with two components. The true values of parameters for mixed copula are given in Table 5. We generate the data values for each mixed copula for three different sample sizes: 400, 700, and 1000, thus the penalized maximum likelihood estimators are estimated. We repeat this experiment 100 times for each model.

The simulation results are presented in Tables 6, 7, 8, and 9. Same as the format of Table 3, Table 7 shows the percentages corresponding to the correct selected copula and incorrect selected copula. Tables 6 and 8 display the true values of the marginal parameters and dependence parameters from the copula, biases and MSEs of the penalization maximum likelihood estimators and
Table 9 shows the initial values and the estimators of unidentified parameters.

Models simulated in this example can capture different dependence structures. Let’s consider the cases when \( n = 1000 \). Model 1 is a mixture of Gaussian and Clayton copulas and describes the lower tail dependence. We can see that there is a large probability to select the appropriate copula in Model 1. For Model 2 which consists of Clayton and Gumbel copulas, Clayton is selected in all replications and Gumbel is chosen 80 times out of 100 replications. The results from Model 3 are very promising since Gaussian copula is selected 100 times and Frank copula is selected 98 times.

6.2 Empirical Examples

It is well known that the following real financial data are time series. As advocated in Remark 4, the asymptotic theory developed in this paper holds true for time series too with a change of standard error. To estimate the standard error consistently, we suggest using a subsampling technique in Section 5.3, which is valid for time series context.

We conduct the empirical studies on the co-movement of returns among Asian markets, Chinese markets and International markets in the following examples. The marginal distributions are assumed to follow the t-distribution. The Kolmogorov-Smirnov (KS) tests are performed and the results show that t-distribution is appropriate fit of the marginal distributions. The proposed estimator is calculated and the standard errors are estimated based on the procedure suggested in Section 5.3 with \( b = T^{0.9} \), where \( T \) is the number of observations. Due to the limited space, we only present the results of proposed estimate.

Example 3: The empirical analysis is carried out on returns on equity indices for six Asian markets: China (CH), Hong Kong (HK), Taiwan (TW), Japan (JP), Singapore (SI), and South Korea (KR). The data consists of 93 monthly observations from February 2000 to September 2007. The descriptive statistics are presented in Table 10. The correlation matrix between the six markets are given in Table 11. The working model is model (9) for each pair. That is, a mixed copula contains four well known copulas: Gaussian, Clayton, Gumbel, and Frank. Table 12 reports
the estimates of 15 models. We can notice several facts from Table 12. First, it can be seen clearly that the dependence between Hong Kong and Singapore markets are strongest and that between Singapore and China markets are weakest among all pairs. Further, only one pair, KR-JA, which takes no weight on Clayton copula, and the other pairs put either zero weights or smaller weights on Clayton copula than on Gumbel copula. This implies that the stock markets we consider have bigger right tail dependence than left tail dependence. To explain this asymmetry, one hypothesis is that investors are more sensitive to bad news than good news in other markets. When a crash takes place in a foreign market, investors in domestic markets tend to pay much attention on it and may result in some actions. These actions may pull the domestic market down. In the contrary, people may not pay too much attention on the boom of another market. Moreover, 12 out of 15 pairs put larger weights on Gaussian than Frank copula. This finding suggests that the dependence structure for both markets are more likely Gaussian type. That is, it is less possible that the strongest dependence is centered in the middle of the distribution. Finally, ten pairs take more weight on Clayton than Gaussian or Frank copula. We can conclude that the left tail dependence describes the dependence structure best in our data set.

**Example 4:** We consider the Chinese markets A, B, and H daily indices and study their correlation structures. Data period is from January 7, 2002 to September 12, 2007. The estimation results are displayed in Table 13. We summarize several findings from the results: A index and B index are more likely to boom together, A index and H index are more likely to crash together, and B index and H index are more likely to crash together. We conclude that H may be in the lead position such that the decrease in H index results in the decrease in A and B indices.

**Example 5:** We consider four stock market indices, S&P500 (US), FTSE 100 (UK), Nikkei (Japan), and Hang Seng (Hong Kong). The data set is monthly returns from January 1987 to February 2007 and includes total 242 observations. Table 14 reports the estimates of models. Note that the weights on Gumbel copula are zeros for all the pairs which indicates that no right tail dependence appears for all pairs. All the pairs can be detected the left tail dependence which means that any two
different markets crash together. Hu (2003) proposed to use the Gaussian, Gumbel, and Survival Gumbel mixture model to model the correlation structure among these four markets. The above two findings we have are similar to the results of Hu (2003). But our results show somewhat the Gaussian type of dependence which is in contrast to that in Hu (2003), who generally found a weight of close to 0 on the Gaussian copula. This suggests that our method detects both linear and nonlinear dependence, while Hu (2003) detected only nonlinear dependence.

7 Proofs of Theorems

In this section, we present the detailed proofs of Theorems 1 and 2.

Proof of Theorem 1: For any fixed point \( \phi_0 \in \Gamma_0 \), let \( B_{\varepsilon}(\phi_0) \) be an open ball of radius \( \varepsilon \) centered at \( \phi_0 \). We know \( \phi_0 = (\phi_0^T, \theta_2^T)^T \), \( \phi_{f0} \) corresponds to the true values of the parameters and it is unique for all the \( \phi_0 \), but the values of \( \theta_2^* \) are different for different points \( \phi_0 \). Therefore, in order to show that

\[ \hat{\phi}_f \rightarrow \phi_{f0}, \text{ in probability,} \]

and

\[ \hat{\theta}_2 \rightarrow \theta_2^*, \text{ in probability,} \]

it is sufficient to show that, for fixed point \( \phi_0 \),

\[ Q(\phi) < Q(\phi_0), \text{ in probability,} \]

for any sufficiently small \( \varepsilon > 0 \) at all points \( \phi \) on the surface of intersection of \( B_{\varepsilon}(\phi) \) and \( \Gamma \). That is, there exists a local maximum in the interior of \( B_{\varepsilon}(\phi) \).

\[
Q(\phi) - Q(\phi_0) = [L(\phi) - L(\phi_0)] - T \sum_{k=1}^{s} [p_{\gamma_T}(\lambda_k) - p_{\gamma_T}(\lambda_{0k})] + \delta \sum_{k=1}^{s} [\lambda_k - \lambda_{0k}]
\]

\[
\equiv I_1 - I_2 + I_3.
\]

Applying Taylor’s expansion to \( L(\phi) \) at point \( \phi_0 \), we have

\[
I_1 = \left[ \frac{\partial L(\phi_0)}{\partial \phi} \right]^T (\phi - \phi_0) + \frac{1}{2} (\phi - \phi_0)^T \frac{\partial^2 L(\phi_0)}{\partial \phi \partial \phi^T} (\phi - \phi_0)
\]
\[
\begin{align*}
&+ \frac{1}{6} \frac{\partial}{\partial \phi} \left[ (\phi - \phi_0)^T \frac{\partial^2 L(\phi^*)}{\partial \phi \partial \phi^T} (\phi - \phi_0) \right] (\phi - \phi_0) \\
\equiv I_{11} + I_{12} + I_{13},
\end{align*}
\]

where \( \phi^* \) lies between \( \phi \) and \( \phi_0 \). By the law of large numbers,

\[
\frac{1}{T} \frac{\partial L(\phi_0)}{\partial \phi} \to 0 \quad \text{in probability,}
\]

then, for any given \( \varepsilon \),

\[
\left| \frac{1}{T} I_{11} \right| < \varepsilon^2.
\]

Thus,

\[
\left| \frac{1}{T} I_{11} \right| < \sum_{j=1}^{p} \varepsilon^2 (\phi_j - \phi_0)(\phi_j) = p\varepsilon^3.
\]

Next, we consider \( I_{12} \). It is easy to see that

\[
\frac{1}{T} I_{12} = \frac{1}{2T} (\phi - \phi_0)^T \left[ \frac{\partial^2 L(\phi_0)}{\partial \phi \partial \phi^T} - E \left\{ \frac{\partial^2 L(\phi_0)}{\partial \phi \partial \phi^T} \right\} \right] (\phi - \phi_0)
\]

\[
+ \left[ -\frac{1}{2} (\phi - \phi_0)^T I(\phi_0)(\phi - \phi_0) \right]
\]

\[
\equiv K_1 + K_2.
\]

By the law of large numbers again, we have

\[
\frac{1}{T} \frac{\partial^2 L(\phi_0)}{\partial \phi \partial \phi^T} = \frac{1}{T} E \left\{ \frac{\partial^2 L(\phi_0)}{\partial \phi \partial \phi^T} \right\} + o_p(1) = -I(\phi_0) + o_p(1).
\]

Therefore, for any given \( \varepsilon \), we have \( K_1 < p^2 \varepsilon^3 \). The second term \( K_2 \) is a negative quadratic form of \( (\phi - \phi_0) \) and it is easy to show that \( K_2 < -q_1 \varepsilon^2 \), where \( q_1 \) is the maximum eigenvalue of information matrix \( I(\phi_0) \) and \( q_1 < 0 \). Combining the first term \( K_1 \) and second term \( K_2 \), we obtain that there exists a constant \( C > 0 \) such that

\[
\frac{1}{T} I_{12} < -C \varepsilon^2 \quad \text{in probability.}
\]

For \( I_{13} \), we have

\[
\left| \frac{1}{T} \sum_{t=1}^{T} M_{jkl}(X_t, Y_t) \right| < 2m_{jkl} \quad \text{in probability.}
\]

Hence, by assumption \( A3 \),

\[
\left| \frac{1}{T} I_{13} \right| = \left| \frac{1}{6T} \sum_{j,k,l=1}^{p} \frac{\partial^3 L(\phi^*)}{\partial \phi_j \partial \phi_k \partial \phi_l} (\phi_j - \phi_0)(\phi_k - \phi_0)(\phi_l - \phi_0) \right|
\]

\[
\leq \frac{1}{6T} \varepsilon^3 \sum_{t=1}^{T} \left\{ \sum_{j,k,l=1}^{p} M_{jkl}^2(X_t, Y_t) \right\}^{1/2} \leq \frac{1}{6} \varepsilon^3 C = C \varepsilon^3,
\]
where $a = \frac{1}{6}p^3C$. Thus, we get

$$\frac{1}{T}I_1 < -C\varepsilon^2 + (p + a)\varepsilon^3. \quad (10)$$

Now, we deal with the penalty term,

$$\frac{1}{T}I_2 = \sum_{k=1}^{s} [p_{\gamma T}(\lambda_k) - p_{\gamma T}(\lambda_0 k)]$$

$$= \sum_{k=1}^{s} p_{\gamma T}'(\lambda_0 k)(\lambda_k - \lambda_0 k) + \frac{1}{2} \sum_{k=1}^{s} p_{\gamma T}''(\lambda_0 k)(\lambda_k - \lambda_0 k)^2 \{1 + o(1)\}$$

$$\leq s\varepsilon \max_{1 \leq k \leq s} \{p_{\gamma T}'(\lambda_0 k)\} + \frac{1}{2} s\varepsilon^2 \max_{1 \leq k \leq s} \{p_{\gamma T}''(\lambda_0 k)\} \{1 + o(1)\}$$

$$\leq s\varepsilon b_T + \frac{1}{2} s\varepsilon^2 \max_{1 \leq k \leq s} \{p_{\gamma T}''(\lambda_0 k)\} \{1 + o(1)\},$$

where $\lambda_{0k}^*$ is between $\lambda_k$ and $\lambda_{0k}$. By the fact that $b_T \to 0$, the first term in $\frac{1}{T}I_2$ is less than $s\varepsilon^3$ for given $\varepsilon$. If $\max\{p_{\gamma T}''(\lambda_0 k)\}$ goes to zero, the second term in $\frac{1}{T}I_2$ is bounded by $\frac{1}{2} s\varepsilon^3$. That is

$$\frac{1}{T}I_2 = \frac{3}{2} s\varepsilon^3. \quad (11)$$

By solving the Lagrange multiplier, we can find that $\delta = T[1 - \sum_{k=1}^{s} \lambda_k p_{\gamma T}'(\lambda_k)]$ and

$$\frac{1}{T}I_3 = [1 - \sum_{k=1}^{s} \lambda_k p_{\gamma T}'(\lambda_k)] \sum_{k=1}^{s} (\lambda_k - \lambda_{0k}) = 0. \quad (12)$$

Combining the equations (10), (11), and (12), we have

$$\frac{1}{T}(I_1 + I_2 + I_3) < -C\varepsilon^2 + (p + a)\varepsilon^3 + \frac{3}{2} s\varepsilon^3.$$

Therefore,

$$Q(\phi) < Q(\phi_0).$$

Now, we embark on the proof of the second part.

$$0 \leq Q(\hat{\phi}) - Q(\phi_0)$$

$$= L(\hat{\phi}) - L(\phi_0) - T \sum_{k=1}^{s} [p_{\gamma T}(\hat{\lambda}_k) - p_{\gamma T}(\lambda_0 k)] + \delta \sum_{k=1}^{s} (\hat{\lambda}_k - \lambda_{0k})$$

$$= \left[ \frac{\partial L(\phi_0)}{\partial \phi} \right]^T (\hat{\phi} - \phi_0) + \frac{1}{2} (\hat{\phi} - \phi_0)^T \left[ \frac{\partial^2 L(\phi_0)}{\partial \phi \partial \phi^T} \right] (\hat{\phi} - \phi_0) - T \sum_{k=1}^{s} p_{\gamma T}'(\lambda_0 k)(\hat{\lambda}_k - \lambda_{0k})$$

$$= \left[ \frac{\partial L(\phi_0)}{\partial \phi} \right]^T (\hat{\phi} - \phi_0) + \frac{1}{2} (\hat{\phi} - \phi_0)^T \left[ \frac{\partial^2 L(\phi_0)}{\partial \phi \partial \phi^T} \right] (\hat{\phi} - \phi_0) - T \sum_{k=1}^{s} p_{\gamma T}'(\lambda_0 k)(\hat{\lambda}_k - \lambda_{0k})$$

25
\[-\frac{1}{2} \sum_{k=1}^{s} \gamma_k (\hat{\lambda}_k - \lambda_k)^2 + \| \hat{\phi} - \phi_0 \|^3 \cdot O_p(1) \]
\[= \left[ \frac{\partial L(\phi_0)}{\partial \phi} \right]^T (\hat{\phi} - \phi_0) - \frac{1}{2} (\hat{\phi} - \phi_0)^T I(\phi_0) (\hat{\phi} - \phi_0) - T b'(\hat{\phi} - \phi_0) \]
\[-\frac{1}{2} (\hat{\phi} - \phi_0)^T \Sigma (\hat{\phi} - \phi_0) + \| \hat{\phi} - \phi_0 \|^3 \cdot O_p(1). \]

Note that $T^{-1/2} \frac{\partial L(\phi_0)}{\partial \phi} = O_p(1)$ and $\| \hat{\phi} - \phi_0 \| = o_p(1)$. For each $\varepsilon$, there is a sequence $c_{T\varepsilon} \to 0$ and a $K_\varepsilon$ such that with probability greater than $1 - \varepsilon$,
\[\left| \frac{1}{T} \frac{\partial L(\phi_0)}{\partial \phi} \right| < K_\varepsilon \frac{\sqrt{T}}{\sqrt{T}}\]
and
\[\| \hat{\phi} - \phi_0 \| < c_{T\varepsilon}.

Then, we have
\[0 \leq \frac{1}{T}(Q(\hat{\phi}) - Q(\phi_0)) \leq \frac{K_\varepsilon}{\sqrt{T}} \| \hat{\phi} - \phi_0 \| - \sqrt{b} T \| \hat{\phi} - \phi_0 \| - \frac{1}{2} (\hat{\phi} - \phi_0)^T (I(\phi_0) + \Sigma) (\hat{\phi} - \phi_0) + \| \hat{\phi} - \phi_0 \|^3 \leq \left( \frac{K_\varepsilon}{\sqrt{T}} - \sqrt{b} T \right) \| \hat{\phi} - \phi_0 \| - \frac{\sqrt{b}}{2} \max_{1 \leq i \leq p} (h_i) \| \hat{\phi} - \phi_0 \|^2 + \frac{1}{2} c_{T\varepsilon} \| \hat{\phi} - \phi_0 \|^2,

where $h_i$ is the eigenvalue of $I(\phi_0) + \Sigma$. By solving this inequality, we can get
\[\| \hat{\phi} - \phi_0 \| < \frac{2(K_\varepsilon/\sqrt{T} + \sqrt{b} T)}{\sqrt{b} \max_{1 \leq i \leq p} (h_i) + c_{T\varepsilon}} < K_\varepsilon^* \cdot (1/\sqrt{T} + b T).

This completes the proof of the theorem. \qed

Lemma 1: Assume the conditions in Theorem 1, if $\lim \inf_{T \to \infty} \lim \inf_{\lambda_k \to 0^+} p_{\gamma k}(\lambda_k)/\gamma T > 0$ and $\sqrt{T} \gamma T \to \infty$, for any given $\phi$ such that $\| \phi - \phi_0 \| = O_p(T^{-1/2})$ and any constant $C$, then,
\[Q(\phi_1, 0) \geq Q(\phi_1, \lambda_2) \text{ in probability.} \]

Proof: It suffices to show that as $n \to \infty$, for any $\phi_2$ satisfying $\| \phi_2 - \phi_20 \| = O_p(T^{-1/2})$ and for small $\varepsilon_T = CT^{-1/2}$ and $j = r + 1, \ldots, s$,
\[\frac{\partial Q(\phi)}{\partial \lambda_j} < 0, \text{ for } 0 < \lambda_j < \varepsilon_T.\]
By Taylor expansion, we have

\[
\frac{\partial Q(\phi)}{\partial \lambda_j} = \frac{\partial L(\phi)}{\partial \lambda_j} - Tp_{\tau}^\prime(\lambda_j) - \delta
\]

\[
= \frac{\partial L(\phi_0)}{\partial \lambda_j} + \sum_{i=1}^{p} \frac{\partial^2 L(\phi_0)}{\partial \lambda_j \partial \phi_i} (\phi_i - \phi_{i0}) + \sum_{i=1}^{p} \sum_{k=1}^{p} \frac{\partial^3 L(\phi^*_{k})}{\partial \lambda_j \partial \phi_i \partial \phi_k} (\phi_i - \phi_{i0})(\phi_k - \phi_{k0}) - Tp_{\tau}^\prime(\lambda_j) - (3)
\]

It is easy to show that

\[
\frac{1}{T} \frac{\partial L(\phi_0)}{\partial \lambda_j} = O_p(T^{-1/2})
\]

and

\[
\frac{1}{T} \frac{\partial^2 L(\phi_0)}{\partial \lambda_j \partial \phi_i} = \frac{1}{T} E \left[ \frac{\partial^2 L(\phi_0)}{\partial \lambda_j \partial \phi_i} \right] + o_p(1).
\]

For the second term in (13),

\[
\sum_{i=1}^{p} \frac{\partial^2 L(\phi_0)}{\partial \lambda_j \partial \phi_i} (\phi_i - \phi_{i0})
\]

\[
= \sum_{i=1}^{p} \left[ \frac{\partial^2 L(\phi_0)}{\partial \lambda_j \partial \phi_i} - E \left( \frac{\partial^2 L(\phi_0)}{\partial \lambda_j \partial \phi_i} \right) \right] (\phi_i - \phi_{i0}) + \sum_{i=1}^{p} E \left( \frac{\partial^2 L(\phi_0)}{\partial \lambda_j \partial \phi_i} \right) (\phi_i - \phi_{i0})
\]

\[
\equiv K_1 + K_2
\]

By the assumption \(\|\phi - \phi_0\| = O_p(T^{-1/2})\), we have

\[
|K_2| = \left| T \sum_{i=1}^{p} I(\phi_0)(\phi_i - \phi_{i0}) \right| \leq TO_p(T^{-1/2}) \left\{ \sum_{i=1}^{p} I^2(\phi_0) \right\}^{1/2}.
\]

Based on assumption (A4), we have

\[
K_2 = O_p(\sqrt{T}).
\]

As for the term \(K_1\), by the Cauchy-Schwarz inequality we have

\[
|K_1| \leq \|\phi_i - \phi_{i0}\| \left[ \sum_{i=1}^{p} \left\{ \frac{\partial^2 L(\phi_0)}{\partial \lambda_j \partial \phi_i} - E \left( \frac{\partial^2 L(\phi_0)}{\partial \lambda_j \partial \phi_i} \right) \right\} \right]^{1/2} = O_p(T^{-1/2}) o_p(T) = o_p(\sqrt{T}).
\]

For the third term in (13), we have

\[
\sum_{i=1}^{p} \frac{\partial^3 L(\phi^*_{k})}{\partial \lambda_j \partial \phi_i \partial \phi_k} (\phi_i - \phi_{i0})(\phi_k - \phi_{k0})
\]

\[
= \sum_{i=1}^{p} \left[ \frac{\partial^3 L(\phi^*_{k})}{\partial \lambda_j \partial \phi_i \partial \phi_k} - E \left( \frac{\partial^3 L(\phi^*_{k})}{\partial \lambda_j \partial \phi_i \partial \phi_k} \right) \right] (\phi_i - \phi_{i0})(\phi_k - \phi_{k0}) + \sum_{i=1}^{p} E \left( \frac{\partial^3 L(\phi^*_{k})}{\partial \lambda_j \partial \phi_i \partial \phi_k} \right) (\phi_i - \phi_{i0})(\phi_k - \phi_{k0})
\]

\[
\equiv K_3 + K_4.
\]
By assumption (A3),

\[ |K_4| \leq C^{1/2} T_p \| \phi - \phi_0 \|^2 = O_p(1) = O_p(\sqrt{T}). \]

It is easy to show that

\[ |K_3| \leq o_p(T) \| \phi - \phi_0 \|^2 = O_p(1) = O_p(\sqrt{T}). \]

Then,

\[ \frac{\partial Q(\phi)}{\partial \lambda_j} = O_p(T^{1/2}) - T p'_{\gamma_T}(\lambda_j) - \delta = T \gamma_T \left( - \frac{p'_{\gamma_T}(\lambda_j)}{\gamma_T} - \frac{\delta}{T \gamma_T} + O_p \left( \frac{T^{-1/2}}{\gamma_T} \right) \right). \]

When \( \lim \inf_{T \to \infty} \lim \inf_{\lambda_k \to 0^+} p'_{\gamma_T}(\lambda_k) / \gamma_T > 0 \), \( \delta / (T \gamma_T) = \left[ 1 - \sum_{k=1}^s \lambda_k p'_{\gamma_T}(\lambda_k) \right] \gamma_T^{-1} > 0 \), and \( \sqrt{T} \gamma_T \to \infty \), we can show that \( O_p(T^{1/2}) \lambda_k - \delta \lambda_k < T p_{\gamma_T}(\lambda_k) \) in probability in a neighborhood of 0. That is,

\[ Q(\phi_1, 0) - Q(\phi_1, \lambda_2) \geq 0. \]

This completes the proof. \( \square \)

**Lemma 2:** Under the regularity conditions, if \( b_T = O_p(T^{-1/2}) \), then

\[ \frac{2}{T} (Q(\phi) - Q(\phi_0)) = g_T(\phi) + (Z_T - b)^T (I(\phi_0) + \Sigma)^{-1} (Z_T - b) + O_p \left( \| \phi - \phi_0 \|^2 \right), \]

where

\[ g_T(\phi) = - \left[ (I(\phi_0) + \Sigma)^{-1} (Z_T - b) - (\phi - \phi_0) \right]^T (I(\phi_0) + \Sigma) \left[ (I(\phi_0) + \Sigma)^{-1} (Z_T - b) - (\phi - \phi_0) \right] \]

and \( Z_T = \frac{1}{T} \frac{\partial L(\phi_0)}{\partial \phi} \). Moreover, if \( \Gamma \) is convex in a neighborhood of \( \phi_0 \), then

\[ \| \hat{\phi} - \tilde{\phi} \| = o_p(T^{-1/2}), \]

where \( \tilde{\phi} \) is the maximized value of quadratic form \( g_T(\phi) \) over \( \Gamma \), i.e., \( \tilde{\phi} = \max_{\Gamma} g_T(\phi - \phi_0) \).

**Proof:** It is easy to verify that

\[ Q(\phi) - Q(\phi_0) \]
\[ \begin{align*}
&= L(\phi) - L(\phi_0) - T \left[ \sum_{k=1}^{s} p_{\gamma T}(\lambda_k) - \sum_{k=1}^{s} p_{\gamma T}(\lambda_0k) \right] + \delta \left( \sum_{k=1}^{s} \lambda_k - \sum_{k=1}^{s} \lambda_0k \right) \\
&= \left[ \frac{\partial L(\phi_0)}{\partial \phi} \right]^T (\phi - \phi_0) + \frac{1}{2} (\phi - \phi_0)^T \frac{\partial^2 L(\phi_0)}{\partial \phi \partial \phi}(\phi - \phi_0) - T \sum_{k=1}^{s} p_{\gamma T}'(\lambda_0k)(\lambda_k - \lambda_0k) \\
& \quad - \frac{1}{2} T \sum_{k=1}^{s} p_{\gamma T}'(\lambda_0k)(\lambda_k - \lambda_0k)^2 + O_p \left( \|\phi - \phi_0\|^3 \right) \\
&= T(Z_T - b)^T (\phi - \phi_0) - \frac{1}{2} T(\phi - \phi_0)^T (I(\phi_0) + \Sigma)(\phi - \phi_0) + O_p(\|\phi - \phi_0\|^3) \\
&= -T \left[ (I(\phi_0) + \Sigma)^{-1} (Z_T - b) - (\phi - \phi_0) \right]^T (I(\phi_0) + \Sigma) \left[ (I(\phi_0) + \Sigma)^{-1} (Z_T - b) - (\phi - \phi_0) \right] \\
& \quad + (Z_T - b)^T (I(\phi_0) + \Sigma)^{-1} (Z_T - b) + O_p \left( \|\phi - \phi_0\|^3 \right). 
\end{align*} \]

To show \( \sqrt{T} \|\hat{\phi} - \tilde{\phi}\| = o_p(1) \), it is equivalent to showing that \( |g_T(\hat{\phi}) - g_T(\tilde{\phi})| = o_p(T^{-1}) \) since \( g_T(\phi) \) is a quadratic function. Let \( R_T(\phi) \) denote the last term in (14). It is clear that

\[ 0 \leq \frac{1}{T} \left( Q(\hat{\phi}) - Q(\tilde{\phi}) \right) = g_T(\hat{\phi}) - g_T(\tilde{\phi}) + R_T(\hat{\phi}) - R_T(\tilde{\phi}), \]

and that \( g_T(\hat{\phi}) - g_T(\tilde{\phi}) \) is negative since \( \tilde{\phi} \) is the maximum of \( g_T(\phi) \). Then,

\[ |g_T(\hat{\phi}) - g_T(\tilde{\phi})| \leq |R_T(\hat{\phi}) - R_T(\tilde{\phi})|. \]

We know \( \|\hat{\phi} - \phi_0\| = O_p(T^{-1/2}) \) and we can show that \( \|\tilde{\phi} - \phi_0\| = O_p(T^{-1/2}) \) by using the same arguments in Theorem 1. Thus,

\[ |R_T(\hat{\phi}) - R_T(\tilde{\phi})| = O_p(T^{-\frac{3}{2}}), \]

then,

\[ |g_T(\hat{\phi}) - g_T(\tilde{\phi})| = o_p(T^{-1}). \]

This completes the proof.

**Lemma 3:** Let \( \bar{\phi} = \max_{C_T} g_T(\phi - \phi_0) \), where \( C_T \) is a cone that satisfies

\[ \inf_{x \in C_T} \|x - y\| = o(\|y - \phi_0\|) \text{ for all } y \in \Gamma, \quad \text{and} \quad \inf_{y \in \Gamma} \|x - y\| = o(\|x - \phi_0\|) \text{ for all } x \in C_T. \]

Then, \( \|\bar{\phi} - \tilde{\phi}\| = o_p(T^{-1/2}). \)
**Proof:** This lemma can be justified by the square root-T consistency of $\tilde{\phi}$ and the definition of cone. Define

$$q_T(W_T, \phi) = \left[ (I(\phi_0) + \Sigma)^{-1}(Z_T - b) - (\phi - \phi_0) \right]^T \left( I(\phi_0) + \Sigma \right) \left[ (I(\phi_0) + \Sigma)^{-1}(Z_T - b) - (\phi - \phi_0) \right],$$

where $W_T = (I(\phi_0) + \Sigma)^{-1}(Z_T - b)$. According to the definitions of $\tilde{\phi}$ and $\bar{\phi}$, we have

$$\tilde{\phi} = \inf_{\phi \in C_T} q_T(W_T, \phi)$$

and

$$\bar{\phi} = \inf_{\phi \in \Gamma} q_T(W_T, \phi).$$

Let $W_T^* \in \Gamma$ such that $\inf_{\Gamma} q_T(W_T, \phi) = \inf_{\Gamma} q_T(W_T, W_T^*) + o_P(T^{-1/2})$. Thus,

$$\|\tilde{\phi} - \bar{\phi}\| = \inf_{C_T} q_T(W_T, \phi) - \inf_{\Gamma} q_T(W_T, \phi) \leq \inf_{\Gamma} q_T(W_T, W_T^*) + \inf_{C_T} q_T(W_T^*, \phi) - \inf_{\Gamma} q_T(W_T, \phi) \leq \inf_{\Gamma} q_T(W_T^*, \phi) + o_P(T^{-1/2}) = o_p(\|\tilde{\phi} - \phi_0\|) + o_P(T^{-1/2}) = o_P(T^{-1/2}).$$

By the definition of cone, we have

$$\|\tilde{\phi} - \bar{\phi}\| \geq o_p(\|\tilde{\phi} - \phi_0\|) = o_P(T^{-1/2}).$$

This completes the proof.

**Proof of Theorem 2:** In the following proof, we consider the partition $\phi = \left( \begin{array}{c} \phi_1 \\ \lambda_2 \end{array} \right)$. By assuming that $\left( \begin{array}{c} \hat{\phi}_1 \\ 0 \end{array} \right)$ is the local maximizer of the penalized log-likelihood function $Q(\phi_1, 0)$, it suffices to show that as $n \to \infty$,

$$P \left\{ Q(\phi_1, \lambda_2) < Q(\hat{\phi}_1, 0) \right\} \to 1.$$ 

Firstly, by the assumption above, we have the following expressions,

$$Q(\phi_1, \lambda_2) - Q(\hat{\phi}_1, 0) = Q(\phi_1, \lambda_2) - Q(\phi_1, 0) + Q(\phi_1, 0) - Q(\hat{\phi}_1, 0) \leq Q(\phi_1, \lambda_2) - Q(\phi_1, 0).$$
Note that by Lemma 1, the last expression is negative with probability tending to one as \( n \) increases to infinity. This completes the proof of part (a).

Now, we proceed to the proof of part (b). Using the similar proof as in Theorem 1, it can be shown easily that there exists a root-\( n \) consistent estimator, say \( \tilde{\phi}_r \), which is the local maximizer of \( Q\{ (\phi_r, 0) \} \). Note that

\[
\frac{1}{T} [Q(\phi_r, 0) - Q(\phi_{r0}, 0)] \\
= \frac{1}{T} [L(\phi_r, 0) - L(\phi_{r0}, 0)] - \sum_{k=1}^{r} [p_{\gamma_T}(\lambda_k) - p_{\gamma_T}(\lambda_{0k})] + \frac{\delta}{T} \left[ \sum_{k=1}^{r} (\lambda_k - \lambda_{0k}) \right] \\
= g_T(\phi_r) + R_T(\phi_r) - (Z_{T1} - b_1)^T (I(\phi_{r0}) + \Sigma_1)^{-1} (Z_{T1} - b_1),
\]

where \( Z_{T1} = T^{-1} \partial L(\phi_{r0})/\partial \phi_r \). Let \( \bar{\phi}_r = \max_T g_T(\phi_r - \phi_{r0}) \). It follows from Lemmas 2 and 3 that

\[
\| \bar{\phi}_r - \tilde{\phi}_r \| = o_p(T^{-1/2}), \quad \text{and} \quad \| \bar{\phi}_r - \bar{\phi}_r \| = o_p(T^{-1/2}),
\]

where \( \tilde{\phi}_r = \max_{C_1} g_T(\phi_r - \phi_{r0}) \). Then, combining above two equations, we have

\[
\sqrt{T} \| \hat{\phi} - \phi_{r0} \| \leq \sqrt{T} \left( \| \bar{\phi}_r - \tilde{\phi}_r \| + \| \bar{\phi}_r - \tilde{\phi}_r \| + \| \bar{\phi}_r - \phi_{r0} \| \right) = \sqrt{T} \| \bar{\phi}_r - \phi_{r0} \| + o_p(1).
\]

Next, to finish the proof, we consider two cases. Case I. Suppose that \( \phi_{r0} \) is an interior point of the subset of \( \Gamma \). Then, \( C_\Gamma = R^q \). According to the definition of \( \tilde{\phi}_r \),

\[
\tilde{\phi}_r - \phi_{r0} = (I_1(\phi_{r0}) + \Sigma_1)^{-1} (Z_{T1} - b_1).
\]

Therefore,

\[
\sqrt{T} (\hat{\phi}_r - \phi_{r0}) = \sqrt{T} (I_1(\phi_{r0}) + \Sigma_1)^{-1} (Z_{T1} - b_1) + o_p(1).
\]

Therefore,

\[
\sqrt{T} \left[ \hat{\phi}_r - \phi_{r0} + (I_1(\phi_{r0}) + \Sigma_1)^{-1} b_1 \right] = \sqrt{T} (I_1(\phi_{r0}) + \Sigma_1)^{-1} Z_{T1} + o_p(1).
\]

Case II. Suppose that \( \phi_{r0} \) is on the boundary of the subset of \( \Gamma \). That is, \( \phi_{r0} = (\lambda_{10}, \theta_{10}, \alpha_0^T, \beta_0^T)^T = (1, \theta_{10}, \alpha_0^T, \beta_0^T)^T \). Then \( C_\Gamma = [0, \infty) \times R^{d_1 + d_\alpha + d_\beta} \). In this case, \( q = 1 + d_1 + d_\alpha + d_\beta \), and, by
maximizing the quadratic form $g_T(\phi_r)$ over $C_r$, $\tilde{\phi}_r$ has the representation as

$$\sqrt{T} [I_1(\phi_{r0}) + \Sigma_1] (\tilde{\phi}_r - \phi_{r0}) = \phi_r + o_p(1),$$

where the limiting random variable $\tilde{\phi}_r$ has the following representation

$$
\begin{pmatrix}
Z_{11} - b_{11} \\
Z_{12} - b_{12} \\
\vdots \\
Z_{1q} - b_{1q}
\end{pmatrix}
I\{Z_{11} > b_{11}\} +
\begin{pmatrix}
0 \\
Z_{12} - b_{12} - (I_{11}^{21}/I_{11})(Z_{11} - b_{11}) \\
\vdots \\
Z_{1q} - b_{1q} - (I_{11}^{q1}/I_{11})(Z_{11} - b_{11})
\end{pmatrix}
I\{Z_{11} < b_{11}\},
$$

where $Z_1 = (Z_{11}, Z_{12}, \ldots, Z_{1q})^T$ be a random variable with multivariate Gaussian distribution with mean 0 and covariance matrix $I_1(\phi_{r0})$, $\phi_r \in C_r - \phi_{r0}$, $I_1^{ij} = I_1^{ij}(\phi_{r0})$ are elements of matrix $I_1^{-1}(\phi_{r0}) + \Sigma_1$ and $I_1(\phi_r)$ is the Fisher information when all zero effects are removed. This completes the proof.

\[\square\]

**References**


<table>
<thead>
<tr>
<th>Model</th>
<th>$n$</th>
<th>$\mu_x$ Bias MSE</th>
<th>$\sigma_x$ Bias MSE</th>
<th>$\mu_y$ Bias MSE</th>
<th>$\sigma_y$ Bias MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>400</td>
<td>-0.003 0.0001</td>
<td>0.099 0.010</td>
<td>0.015 0.012</td>
<td>-0.002 0.006</td>
</tr>
<tr>
<td></td>
<td>700</td>
<td>0.001 0.000</td>
<td>0.098 0.010</td>
<td>-0.012 0.007</td>
<td>-0.005 0.002</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.003 0.000</td>
<td>0.090 0.010</td>
<td>0.002 0.004</td>
<td>-0.001 0.001</td>
</tr>
<tr>
<td>2</td>
<td>400</td>
<td>0.002 0.000</td>
<td>0.102 0.010</td>
<td>0.002 0.008</td>
<td>0.002 0.005</td>
</tr>
<tr>
<td></td>
<td>700</td>
<td>0.002 0.000</td>
<td>0.102 0.011</td>
<td>0.002 0.005</td>
<td>0.002 0.003</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.000 0.000</td>
<td>0.100 0.010</td>
<td>-0.002 0.003</td>
<td>-0.000 0.002</td>
</tr>
<tr>
<td>3</td>
<td>400</td>
<td>0.005 0.001</td>
<td>0.097 0.010</td>
<td>0.007 0.011</td>
<td>-0.007 0.005</td>
</tr>
<tr>
<td></td>
<td>700</td>
<td>0.002 0.000</td>
<td>0.096 0.010</td>
<td>0.008 0.004</td>
<td>-0.007 0.002</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.000 0.000</td>
<td>0.090 0.010</td>
<td>0.000 0.004</td>
<td>-0.000 0.002</td>
</tr>
<tr>
<td>4</td>
<td>400</td>
<td>0.004 0.001</td>
<td>0.100 0.010</td>
<td>-0.008 0.010</td>
<td>0.004 0.005</td>
</tr>
<tr>
<td></td>
<td>700</td>
<td>-0.004 0.000</td>
<td>0.100 0.010</td>
<td>-0.005 0.005</td>
<td>0.005 0.001</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.003 0.000</td>
<td>0.093 0.010</td>
<td>-0.000 0.003</td>
<td>-0.002 0.002</td>
</tr>
</tbody>
</table>

Table 1: Bias and MSE of estimators of marginal parameters in Example 1 together with their true values.

<table>
<thead>
<tr>
<th>$n$</th>
<th>Model 1 ($\lambda_1, \theta_1$)</th>
<th>Model 2 ($\lambda_2, \theta_2$)</th>
<th>Model 3 ($\lambda_3, \theta_3$)</th>
<th>Model 4 ($\lambda_4, \theta_4$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>400</td>
<td>Bias MSE (-0.026, -0.004)</td>
<td>(0.008, -0.001)</td>
<td>(0.000, 0.011)</td>
<td>(0.000, -0.518)</td>
</tr>
<tr>
<td>700</td>
<td>Bias MSE (-0.009, -0.002)</td>
<td>(0.002, 0.000)</td>
<td>(0.000, -0.003)</td>
<td>(0.000, -0.358)</td>
</tr>
<tr>
<td>1000</td>
<td>Bias MSE (-0.010, -0.002)</td>
<td>(0.004, 0.000)</td>
<td>(0.000, -0.002)</td>
<td>(0.000, -0.236)</td>
</tr>
</tbody>
</table>

Table 2: Bias and MSE of estimators of copula parameters in Example 1 together with their true values.

<table>
<thead>
<tr>
<th>Model</th>
<th>$n$</th>
<th>Gaussian</th>
<th>Clayton</th>
<th>Gumbel</th>
<th>Frank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>400</td>
<td>1</td>
<td>(0)</td>
<td>(0)</td>
<td>(0.09)</td>
</tr>
<tr>
<td></td>
<td>700</td>
<td>1</td>
<td>(0)</td>
<td>(0)</td>
<td>(0.04)</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1</td>
<td>(0)</td>
<td>(0)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>2</td>
<td>400</td>
<td>(0.01)</td>
<td>1</td>
<td>(0)</td>
<td>(0)</td>
</tr>
<tr>
<td></td>
<td>700</td>
<td>(0)</td>
<td>1</td>
<td>(0)</td>
<td>(0)</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>(0)</td>
<td>1</td>
<td>(0)</td>
<td>(0)</td>
</tr>
<tr>
<td>3</td>
<td>400</td>
<td>(0.03)</td>
<td>(0)</td>
<td>1</td>
<td>(0)</td>
</tr>
<tr>
<td></td>
<td>700</td>
<td>(0)</td>
<td>(0)</td>
<td>1</td>
<td>(0)</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>(0)</td>
<td>(0)</td>
<td>1</td>
<td>(0)</td>
</tr>
<tr>
<td>4</td>
<td>400</td>
<td>(0.29)</td>
<td>(0)</td>
<td>(0)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>700</td>
<td>(0.07)</td>
<td>(0)</td>
<td>(0)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>(0.08)</td>
<td>(0)</td>
<td>(0)</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3: Percentage that the corresponding copula was chosen correctly(incorrectly) in Example 1. Values without parentheses are percentages that copula in mixed copula was chosen correctly. Values with parentheses are percentages that copula not in mixed copula was chosen incorrectly.
Table 4: Estimators of unidentified copula parameters in Example 1 when n=1000.

<table>
<thead>
<tr>
<th>model</th>
<th>(λ_1, θ_1)</th>
<th>(λ_2, θ_2)</th>
<th>(λ_3, θ_3)</th>
<th>(λ_4, θ_4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(1/2, 0.6)</td>
<td>(1/2, 5)</td>
<td>(1/2, 6)</td>
<td>(1/2, 6)</td>
</tr>
<tr>
<td>2</td>
<td>(1/2, 5)</td>
<td>(1/2, 6)</td>
<td>(1/2, 6)</td>
<td>(1/2, 6)</td>
</tr>
<tr>
<td>3</td>
<td>(1/2, 0.6)</td>
<td>(1/2, 5)</td>
<td>(1/2, 6)</td>
<td>(1/2, 6)</td>
</tr>
<tr>
<td>4</td>
<td>(1/2, 0.6)</td>
<td>(1/2, 6)</td>
<td>(1/2, 6)</td>
<td>(1/2, 6)</td>
</tr>
</tbody>
</table>

Table 5: True parameters of the mixed copulas in Example 2

Table 6: Bias and MSE of estimators of marginal parameters in Example 2 together with their true values.
Table 7: Percentage that the corresponding copula was chosen correctly (incorrectly) in Example 2. Values without parentheses are percentages that copula in mixed copula was chosen correctly. Values with parentheses are percentages that copula not in mixed copula was chosen incorrectly.

<table>
<thead>
<tr>
<th>Model($\times 100$)</th>
<th>$n$</th>
<th>Gaussian</th>
<th>Clayton</th>
<th>Gumbel</th>
<th>Frank</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>400</td>
<td>0.95</td>
<td>1</td>
<td>(0)</td>
<td>(0.14)</td>
</tr>
<tr>
<td></td>
<td>700</td>
<td>1</td>
<td>1</td>
<td>(0)</td>
<td>(0.06)</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1</td>
<td>1</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>2</td>
<td>400</td>
<td>(0.52)</td>
<td>1</td>
<td>0.53</td>
<td>(0)</td>
</tr>
<tr>
<td></td>
<td>700</td>
<td>(0.38)</td>
<td>1</td>
<td>0.60</td>
<td>(0)</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>(0.20)</td>
<td>1</td>
<td>0.80</td>
<td>(0)</td>
</tr>
<tr>
<td>3</td>
<td>400</td>
<td>0.99</td>
<td>(0)</td>
<td>(0.01)</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>700</td>
<td>1</td>
<td>(0)</td>
<td>(0)</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1</td>
<td>(0)</td>
<td>(0)</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Table 8: Bias and MSE of estimators of copula parameters in Example 2 together with their true values.

<table>
<thead>
<tr>
<th>Model</th>
<th>$n$</th>
<th>$(\lambda_1, \theta_1)$</th>
<th>$(\lambda_2, \theta_2)$</th>
<th>$(\lambda_3, \theta_3)$</th>
<th>$(\lambda_4, \theta_4)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>400</td>
<td>Bias (-0.035, 0.015)</td>
<td>Bias (-0.015, -0.045)</td>
<td>Bias (-0.005, -0.050)</td>
<td>Bias (-0.387, -0.852)</td>
</tr>
<tr>
<td></td>
<td>700</td>
<td>MSE (0.019, 0.006)</td>
<td>MSE (0.003, 0.107)</td>
<td>MSE (0.002, 0.064)</td>
<td>MSE (0.171, 3.367)</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>Bias (-0.009, 0.009)</td>
<td>Bias (-0.003, 0.013)</td>
<td>Bias (-0.001, 0.046)</td>
<td>Bias (-0.294, -0.717)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MSE (0.004, 0.002)</td>
<td>MSE (0.002, 0.064)</td>
<td>MSE (0.001, 0.046)</td>
<td>MSE (0.133, 1.733)</td>
</tr>
<tr>
<td>2</td>
<td>400</td>
<td>Bias (0.094, 0.541)</td>
<td>Bias (-0.036, -0.253)</td>
<td>Bias (-0.001, -0.208)</td>
<td>Bias (-0.241, -0.690)</td>
</tr>
<tr>
<td></td>
<td>700</td>
<td>MSE (0.050, 0.494)</td>
<td>MSE (0.013, 0.171)</td>
<td>MSE (0.013, 0.144)</td>
<td>MSE (0.108, 1.787)</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>Bias (-0.036, -0.253)</td>
<td>Bias (-0.001, -0.208)</td>
<td>Bias (-0.001, -0.208)</td>
<td>Bias (-0.241, -0.690)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MSE (0.004, 0.002)</td>
<td>MSE (0.002, 0.064)</td>
<td>MSE (0.001, 0.046)</td>
<td>MSE (0.133, 1.733)</td>
</tr>
<tr>
<td>3</td>
<td>400</td>
<td>Bias (0.083, -0.286)</td>
<td>Bias (-0.036, -0.253)</td>
<td>Bias (-0.001, -0.208)</td>
<td>Bias (-0.085, 0.069)</td>
</tr>
<tr>
<td></td>
<td>700</td>
<td>MSE (0.054, 0.101)</td>
<td>MSE (0.013, 0.171)</td>
<td>MSE (0.013, 0.144)</td>
<td>MSE (0.054, 0.061)</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>Bias (0.122, -0.272)</td>
<td>Bias (-0.001, -0.208)</td>
<td>Bias (-0.001, -0.208)</td>
<td>MSE (0.002, 0.033)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MSE (0.049, 0.076)</td>
<td>MSE (0.013, 0.171)</td>
<td>MSE (0.013, 0.144)</td>
<td>MSE (0.049, 0.047)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bias (0.084, -0.272)</td>
<td>Bias (-0.001, -0.208)</td>
<td>Bias (-0.001, -0.208)</td>
<td>MSE (0.002, 0.033)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MSE (0.028, 0.076)</td>
<td>MSE (0.013, 0.171)</td>
<td>MSE (0.013, 0.144)</td>
<td>MSE (0.002, 0.033)</td>
</tr>
</tbody>
</table>
Table 9: Estimators of unidentified copula parameters in Example 2.

<table>
<thead>
<tr>
<th>Model</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>$\theta_3$</th>
<th>$\theta_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Initial Value</td>
<td>2.600</td>
<td>4.000</td>
<td>2.010</td>
</tr>
<tr>
<td></td>
<td>Estimator</td>
<td>2.000</td>
<td>3.000</td>
<td>2.705</td>
</tr>
<tr>
<td></td>
<td>Initial Value</td>
<td>3.200</td>
<td>5.000</td>
<td>2.910</td>
</tr>
<tr>
<td></td>
<td>Estimator</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Initial Value</td>
<td>0.600</td>
<td>4.000</td>
<td>0.839</td>
</tr>
<tr>
<td></td>
<td>Estimator</td>
<td>0.400</td>
<td>3.000</td>
<td>0.272</td>
</tr>
<tr>
<td></td>
<td>Initial Value</td>
<td>0.800</td>
<td>5.000</td>
<td>0.692</td>
</tr>
<tr>
<td></td>
<td>Estimator</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Initial Value</td>
<td>5.000</td>
<td>2.600</td>
<td>4.928</td>
</tr>
<tr>
<td></td>
<td>Estimator</td>
<td>4.928</td>
<td>2.189</td>
<td>4.000</td>
</tr>
<tr>
<td></td>
<td>Initial Value</td>
<td>3.928</td>
<td>2.885</td>
<td>6.000</td>
</tr>
<tr>
<td></td>
<td>Estimator</td>
<td>5.925</td>
<td>2.559</td>
<td></td>
</tr>
</tbody>
</table>

Table 10: Summary statistics for five Asian markets.

<table>
<thead>
<tr>
<th></th>
<th>China</th>
<th>Hong Kong</th>
<th>Taiwan</th>
<th>Korea</th>
<th>Singapore</th>
<th>Japan</th>
<th>Japan</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>0.0122</td>
<td>0.0045</td>
<td>-0.0012</td>
<td>0.0077</td>
<td>0.0046</td>
<td>-0.0020</td>
<td>-0.0020</td>
</tr>
<tr>
<td>stdev</td>
<td>0.0740</td>
<td>0.0787</td>
<td>0.0769</td>
<td>0.1016</td>
<td>0.0779</td>
<td>0.0536</td>
<td>0.0536</td>
</tr>
<tr>
<td>skewness</td>
<td>0.8789</td>
<td>-0.4253</td>
<td>-0.0061</td>
<td>0.2463</td>
<td>-0.2286</td>
<td>-0.3647</td>
<td>-0.3647</td>
</tr>
<tr>
<td>kurtosis</td>
<td>0.1052</td>
<td>3.2174</td>
<td>0.5143</td>
<td>1.4943</td>
<td>2.2949</td>
<td>3.1228</td>
<td>3.1228</td>
</tr>
</tbody>
</table>

Table 11: Linear Correlation Coefficients across five Asian markets.

<table>
<thead>
<tr>
<th></th>
<th>Taiwan</th>
<th>Korea</th>
<th>Singapore</th>
<th>Japan</th>
<th>China</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hong Kong</td>
<td>0.5133</td>
<td>0.5088</td>
<td>0.7613</td>
<td>0.4082</td>
<td>0.2640</td>
</tr>
<tr>
<td>Taiwan</td>
<td>0.4722</td>
<td>0.4934</td>
<td>0.4123</td>
<td>0.1146</td>
<td></td>
</tr>
<tr>
<td>Korea</td>
<td>0.4781</td>
<td>0.5352</td>
<td>0.0742</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Singapore</td>
<td>0.4098</td>
<td>-0.020</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>China</td>
<td></td>
<td></td>
<td>0.075</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>Normal</td>
<td>0.543</td>
<td>0.579</td>
<td>0.518</td>
<td>0.236</td>
</tr>
<tr>
<td>Clayton</td>
<td>0.101</td>
<td>0.246</td>
<td>0.359</td>
<td>0.637</td>
<td>0.603</td>
</tr>
<tr>
<td>Gumbel</td>
<td>0.356</td>
<td>0.175</td>
<td>0</td>
<td>0.126</td>
<td>0</td>
</tr>
<tr>
<td>Frank</td>
<td>0.101</td>
<td>0.246</td>
<td>0.359</td>
<td>0.637</td>
<td>0.603</td>
</tr>
<tr>
<td>( \theta )</td>
<td>Normal</td>
<td>0.746</td>
<td>0.777</td>
<td>0.88</td>
<td>0.321</td>
</tr>
<tr>
<td>Clayton</td>
<td>0.567</td>
<td>1.509</td>
<td>0.706</td>
<td>0.647</td>
<td>0.446</td>
</tr>
<tr>
<td>Gumbel</td>
<td>2.345</td>
<td>0.002</td>
<td>3.325</td>
<td>2.134</td>
<td>0.002</td>
</tr>
<tr>
<td>Frank</td>
<td>0.567</td>
<td>1.509</td>
<td>0.706</td>
<td>0.647</td>
<td>0.446</td>
</tr>
</tbody>
</table>

Table 12: Asia Markets. The values in parentheses are the standard errors of the estimates.
<table>
<thead>
<tr>
<th>Copula</th>
<th>A-B</th>
<th>A-H</th>
<th>B-H</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>0.150</td>
<td>0.363</td>
<td>0.391</td>
</tr>
<tr>
<td>Clayton</td>
<td>0.244</td>
<td>0.256</td>
<td>0.308</td>
</tr>
<tr>
<td>Gumbel</td>
<td>0.321</td>
<td>0.131</td>
<td>0.000</td>
</tr>
<tr>
<td>Frank</td>
<td>0.2851</td>
<td>0.250</td>
<td>0.301</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>θ</th>
<th>Normal</th>
<th>-0.531</th>
<th>0.293</th>
<th>0.138</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clayton</td>
<td>0.049</td>
<td>0.136</td>
<td>0.097</td>
<td></td>
</tr>
<tr>
<td>Gumbel</td>
<td>1.036</td>
<td>1.199</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frank</td>
<td>0.136</td>
<td>0.858</td>
<td>0.756</td>
<td></td>
</tr>
</tbody>
</table>

Table 13: Chinese Markets. The values in parentheses are the standard errors of the estimates.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>λ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>0.418</td>
<td>0.125</td>
<td>0.528</td>
<td>0.113</td>
<td>0.689</td>
<td>0.139</td>
</tr>
<tr>
<td>Clayton</td>
<td>0.582</td>
<td>0.472</td>
<td>0.504</td>
<td>0.731</td>
<td>0.311</td>
<td>0.772</td>
</tr>
<tr>
<td>Gumbel</td>
<td>0</td>
<td>0.061</td>
<td>0</td>
<td>0.156</td>
<td>0</td>
<td>0.089</td>
</tr>
<tr>
<td>Frank</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

| θ     | Normal | 0.849 | 0.840 | 0.536 | 0.923 | 0.558 | 0.433 |
|       | Clayton | 1.525 | 0.491 | 1.703 | 0.449 | 1.320 | 0.428 |
|       | Gumbel | 0.008 | 0.142 |       |       |       |       |
|       | Frank  |       |       |       |       |       | 0.030 |

Table 14: International Markets. The values in parentheses are the standard errors of the estimates.