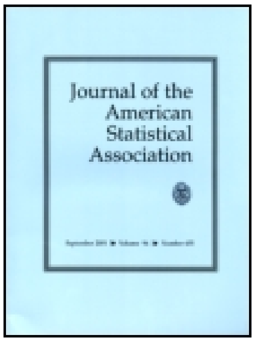

Taylor & Francis
Taylor & Francis Group

Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Selection of Mixed Copula Model via Penalized Likelihood

Zongwu Cai & Xian Wang

To cite this article: Zongwu Cai & Xian Wang (2014) Selection of Mixed Copula Model via Penalized Likelihood, Journal of the American Statistical Association, 109:506, 788-801, DOI: 10.1080/01621459.2013.873366

To link to this article: https://doi.org/10.1080/01621459.2013.873366

Published online: 13 Jun 2014.

Selection of Mixed Copula Model via Penalized Likelihood

Zongwu CAI and Xian WANG

A fundamental issue of applying a copula method in applications is how to choose an appropriate copula function for a given problem. In this article we address this issue by proposing a new copula selection approach via penalized likelihood plus a shrinkage operator. The proposed method selects an appropriate copula function and estimates the related parameters simultaneously. We establish the asymptotic properties of the proposed penalized likelihood estimator, including the rate of convergence and asymptotic normality and abnormality. Particularly, when the true coefficient parameters may be on the boundary of the parameter space and the dependence parameters are in an unidentified subset of the parameter space, we show that the limiting distribution for boundary parameter estimator is half-normal and the penalized likelihood estimator for unidentified parameter converges to an arbitrary value. Finally, Monte Carlo simulation studies are carried out to illustrate the finite sample performance of the proposed approach and the proposed method is used to investigate the correlation structure and comovement of financial stock markets.

KEY WORDS: Boundary and unidentified parameters; Comovement; EM algorithm; LASSO; SCAD; Shrinkage operator; Variable selection.

1. INTRODUCTION

To capture the complex dependence structure among variables, a copula approach has been used recently in many applied fields, in particular, in finance and economics. Li (2000) was the first to investigate the default correlations in credit risk models; Frey and McNeil (2003), Mashal, Naldi, and Zeevi (2003), and Hamerle and Rösch (2005) studied the impact of copula in option pricing; Denuit and Scaillet (2004) applied copula modeling to hedge fund indices to study positive quadrant dependence found in financial indices; Kole, Koedijk, and Verbeek (2007) discussed the applications of copula and addressed the importance of selecting copulas in risk management; Bouyé et al. (2001), Embrechts, Lindskog, and McNell (2003), and Cherubini, Vecchiato, and Luciano (2004) used a copula function to measure the portfolio value-at-risk; Junker, Szimayer, and Wagner (2006) used a copula approach to model the term structure of interest rates; Jondeau and Rockinger (2006) and Lee and Long (2009) established copula-GARCH models to study the dependence among international stock markets and exchange rate time series; Longin and Solnik (2001) examined cross-national dependence structure of asset returns in international financial markets; and Ang and Chen (2002) discovered asymmetric dependence between two asset returns during market downturns and market upturns.

To describe the dependence structure more flexibly, researchers have proposed using a mixed copula which is a linear combination of several copula families. Hu (2006) discussed a combination of Gaussian, Gumbel, and survival Gumbel copula to measure the dependence patterns across financial markets, whereas Chollete, Pena, and Lu (2005) used mixed copula to analyze the comovement of international financial markets. The biggest advantage of mixed copula model is that it can nest different copula shapes. For instance, Gaussian and Gumbel mixed copula can improve a single Gaussian dependence structure by allowing possible right tail dependence. Therefore, empirically, a mixed copula is more flexible to model dependence structure and can deliver better descriptions of dependence structure than an individual copula.

A motivation of this study comes from an analysis of real financial data, consisting of monthly measurements of international stock market indices: S&P500 (U.S.), FTSE 100 (UK), Nikkei (Japan), and Hang Seng (Hong Kong) (1987:01–2007:02) from the Center for Research in Security Prices (CRSP). Of interest examining the existence of the comovement of returns among these four international stock markets, which is one of the popular topics in financial econometrics; see Hu (2006) and Chollete, Pena, and Lu (2005) for more details. Therefore, to study the comovement with various dependence structures, a copula approach is appropriate. The detailed analysis of this dataset is reported in Section 5.2.

When applying a copula approach to solve real problems, a natural question is how to choose an appropriate parametric copula because the distribution from which each data point is drawn is unknown. To attenuate this problem, there have been some efforts in the literature to choose an appropriate individual copula. Chen, Fan, and Patton (2003) and Fermanian (2005) developed goodness-of-fit tests; Kole, Koedijk, and Verbeek (2007) used Kolmogorov-Smirnov and Anderson and Darling (1952) type tests; Scaillet (2007) proposed a kernel-based goodness-of-fit

Zongwu Cai is The Charles Osward Professor of Econometrics and Professor of Economics, Department of Economics, University of Kansas, Lawrence, KS 66945 (E-mail: *caiz@ku.edu*) and Adjunct Professor of Economics, The Wang Yanan Institute for Studies in Economics, Xiamen University, Fujian 361005, China. Xian Wang is Consumer and Production Strategy Analyst III, Detection Analytic, Bank of America, 150 N College Street, Charlotte, NC 28277 (E-mail: *xian.wang@bankofamerica.com*). Cai's research was supported, in part, by the National Nature Science Foundation of China grants #71131008 (Key Project) and #70971113. The authors thank Xiaohong Chen and Yanqin Fan for their useful discussions and suggestions at the early stage of this article, and our thanks also go to the Editor (Xuming He), the Associate Editor, the referees, Qi Li, Runze Li, Yoon-Jae Whang, and other audiences at Sino-Korean Workshop in December 2007 in Xiamen, the Econometrics Workshop at Indiana University in March 2008, the SETA Conference in May 2008 in Seoul, and the FERM conference in June 2010 in Taipei, for their constructive comments and suggestions.

Journal of the American Statistical Association
June 2014, Vol. 109, No. 506, Theory and Methods
DOI: 10.1080/01621459.2013.873366

test; and omnibus tests proposed by Genest, Rémillard, and Beaudoin (2009) can be used to test the existence of copulas. But there is no guidance as to which copula model should be used if the null hypotheses of correct parametric specification in those tests are rejected. Furthermore, based on the idea for variable selection in a classical regression model, Hu (2006) 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. However, Hu (2006) did not provide a theoretical foundation. To the best of our knowledge, so far no work with a theoretical support has been attempted to choose a suitable mixed copula based on a data-driven method although there are a vast amount of papers on variable selection in regression models; see the review paper by Fan and Lv (2010) for details.

This article makes the following major contributions to the literature. The first is to propose a data-driven copula selection method via penalized likelihood plus a shrinkage operator, which is in spirit similar to the least absolute shrinkage and selection operator (LASSO) proposed by Tibshirani (1996) and studied by Fan and Li (2001) for variable selection in a classical linear regression model. Ideally, a large number of candidate copula families are of interest and their contributions to dependence structure vary from one component to another. The main goal is to select a best mixed copula among all candidate copulas to capture dependence structure of the given data and to estimate the corresponding parameters. The best mixed copula can be selected by choosing the one with the highest likelihood. When a fitted mixed copula contains some component copulas with small weights which imply small contribution to dependence structure, as expected, these components should not be included in the mixed copula. To filter out the components with small weights, some constraints should be added on all weights, such as penalty functions. Although 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, our setting has some extra difficulties associated with the boundary and unidentified parameters. In another context, Chen and Khalili (2008) applied a variable selection method to the order selection in finite mixture models. Therefore, a likelihood function can be formulated as a form of a mixed copula including all candidate copulas. Furthermore, a penalized likelihood function is constructed by adding some appropriate penalty functions and constraints for weights. By maximizing the penalized likelihood function, copulas with small weights can be removed by a thresholding rule (shrinkage operator) and parameters remained are estimated. In such a way, model selection and parameter estimation can be done simultaneously.

Another big contribution of this article is to establish the asymptotic properties of proposed estimator under nonstandard situations. It is interesting to note that a general mathematical derivation of asymptotic properties for a maximum likelihood estimator is not applicable here because some parameters may be on a boundary of the parameter space. Andrews (1999) addressed this issue and discussed the asymptotic properties when a parameter is on a boundary for iid sample under a regression setting. Another challenge is that the parameters corresponding to the removed copula components with small weights are in an unidentified subset of the parameter space. Under these nonstandard situations, it is shown that the estimator of unidentified dependent parameter converges to an arbitrary value and the abnormality of the boundary parameter estimator has been established. Finally, to make the proposed methodology practically useful and applicable, the EM algorithm is used to find the penalized likelihood estimator. Also, a data-driven method is used to find the tuning and thresholding parameters in the penalty function. Our simulation results show that our new method has a high probability of selecting an appropriate mixed copula model.

The rest of the article is organized as follows. Section 2 briefly reviews some concepts and facts about mixed copula models. Section 3 introduces the selection procedure based on the penalized likelihood plus a shrinkage operator. Section 4 lists the regularity conditions and develops the asymptotic properties of the proposed estimator. In Section 5, the results of Monte Carlo studies are reported to demonstrate the finite sample performance of the proposed method, together with an empirical analysis of a real financial dataset. Some concluding remarks are provided in Section 6. Finally,the EM algorithm for finding the penalized likelihood estimator and the data-driven methods for finding the threshold parameters are discussed in Appendix A and the technical proofs are relegated to Appendix B.

2. A REVIEW OF MIXED COPULA

Let $\{\mathbf{X}_t\}_{t=1}^T$ be independent p-dimensional vectors of random variables with $\mathbf{X}_t = (X_{t1}, \ldots, X_{tp})^\top$, where $A^\top$ denotes the transpose of a matrix or vector A. Let $f(\mathbf{x})$ and $F(\mathbf{x})$ be the joint density and distribution of $\mathbf{X}$, respectively, and $f_j(x_j)$ and $F_j(x_j)$, $1 \le j \le p$, be the marginal density and distribution of X_j, respectively. Next, we briefly review some basic facts about mixed copula. For more copula concepts and the related properties, the reader is referred to the books by Nelsen (1998) and Joe (1997).

A mixed copula is a linear combination of several copula families. Mathematically, a mixed copula function is formulated as

$$C(\mathbf{u};\boldsymbol{\theta}) = \sum_{k=1}^{s} \lambda_k\, C_k(\mathbf{u};\theta_k) = \sum_{k=1}^{s} \lambda_k\, C_k(F_1(x_1;\alpha_1), \ldots, F_p(x_p;\alpha_p);\theta_k), \tag{1}$$

where $\{C_1(\cdot), \ldots, C_s(\cdot)\}$ is a set of basis copulas, which is a sequence of known copulas with unknown parameters $\{\theta_k\}$, $\{\lambda_k\}_{k=1}^s$ are the weights satisfying $0 \le \lambda_k \le 1$ and $\sum_{k=1}^s \lambda_k = 1$, and s is the number of candidate copulas. A single copula is a special case of the mixed copula when only one component is included in mixed copula. Generally speaking, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_s)^\top$ is the vector of *associate parameters* in the mixture representing the degree of dependence, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_s)^\top$ is the vector of weights or *shape parameters* reflecting the credence in the corresponding copula and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)^\top$ is the vector of *marginal parameters* for marginal distributions. In what follows, let $\boldsymbol{\phi} = (\boldsymbol{\alpha}^\top, \boldsymbol{\lambda}^\top, \boldsymbol{\theta}^\top)^\top$ be a vector of all the parameters involved.

It is easy to show that $C(\mathbf{u})$ in Equation (1) is also a copula. Therefore, all the copula properties hold for a mixed

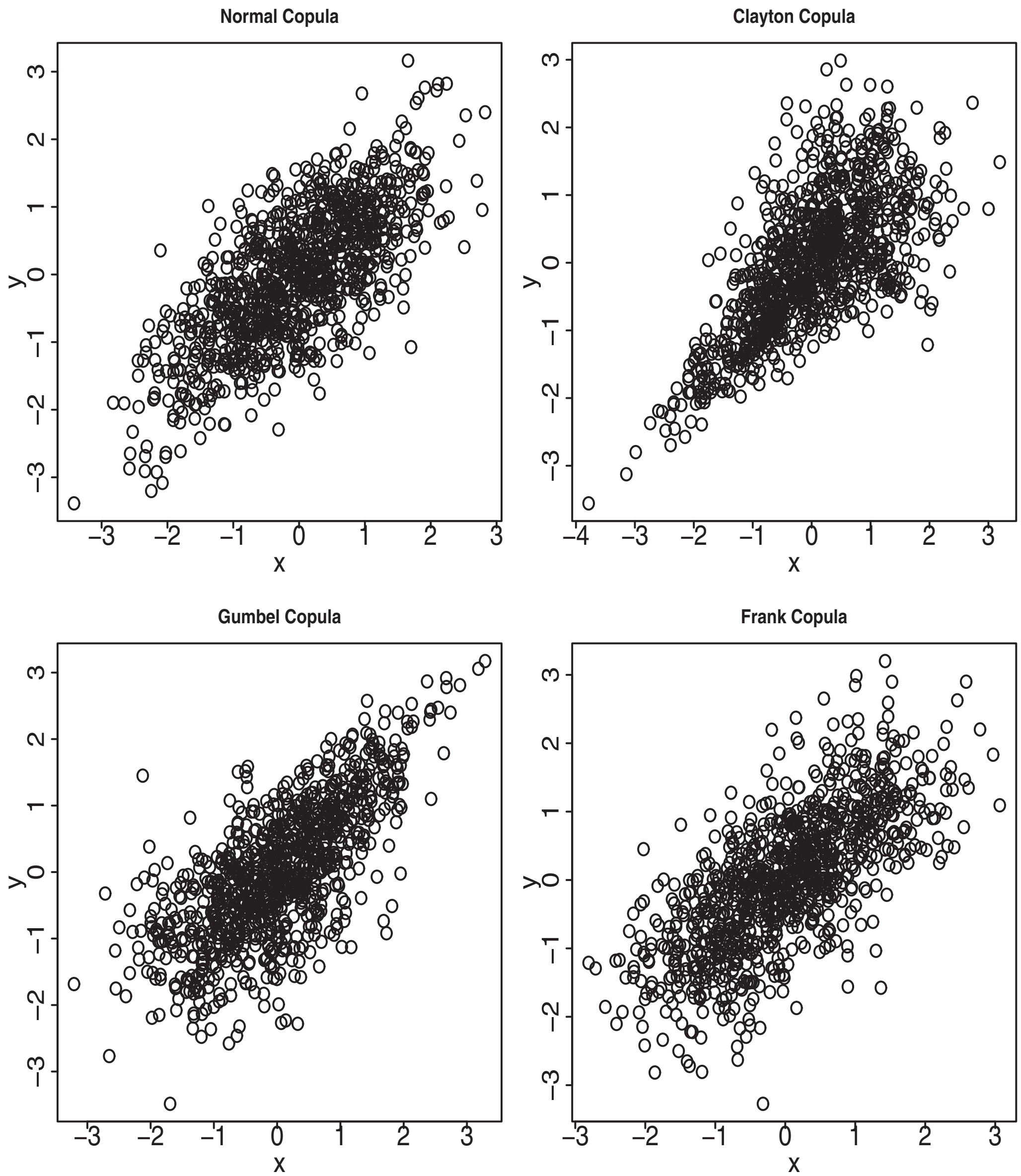

Figure 1. Scatterplots for four different copulas with standard normal marginal distributions and Kendall's $\tau = 0.5$.

copula. Also, for any two random variables X and Y, copula-dependent parameters $\boldsymbol{\theta}$ can be transformed to the Kendall's τ proposed by Kendall (1938) and the Spearman's ρ proposed by Spearman (1904), where $\tau = 4E[C(F_x(X), F_y(Y); \boldsymbol{\theta})] - 1$ and $\rho_s = \rho(F_x(X), F_y(Y))$, the correlation coefficient between two marginal distributions $F_x(X)$ and $F_y(Y)$. For details, see the books by Cherubini, Vecchiato, and Luciano (2004) and Nelsen (1999).

For illustration, Figure 1 gives a visual view of the tail dependence of copulas by showing scatterplots of 300 iid samples generated from four popular copula functions: Gaussian, Gumbel, Clayton, and Frank. Figure 2 presents the scatterplots of 300 iid samples generated from different mixed copula functions with equal weights on each component. The dependence parameters are the same for all the components with Kendall's $\tau = 0.5$ and marginal distributions are generated from the standard normal distributions.

Finally, a mixture copula given in (1) can be intuitively regarded as a finite approximation of an unknown complex copula, where the known copulas $\{C_k(\cdot)\}$ could serve as basis functions. Theoretically or ideally, the basis copulas $\{C_k(\cdot)\}$ should be chosen as many as possible to characterize all possible dependence structures. But in practice, the question is how to choose them efficiently and optimally and to be a better approximation of a true copula. One of the main purposes of the current article is to answer this question by proposing a penalized likelihood method plus a shrinkage operator, described in Section 3 (see later) in detail.

Definition 1. Two mixed copulas

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

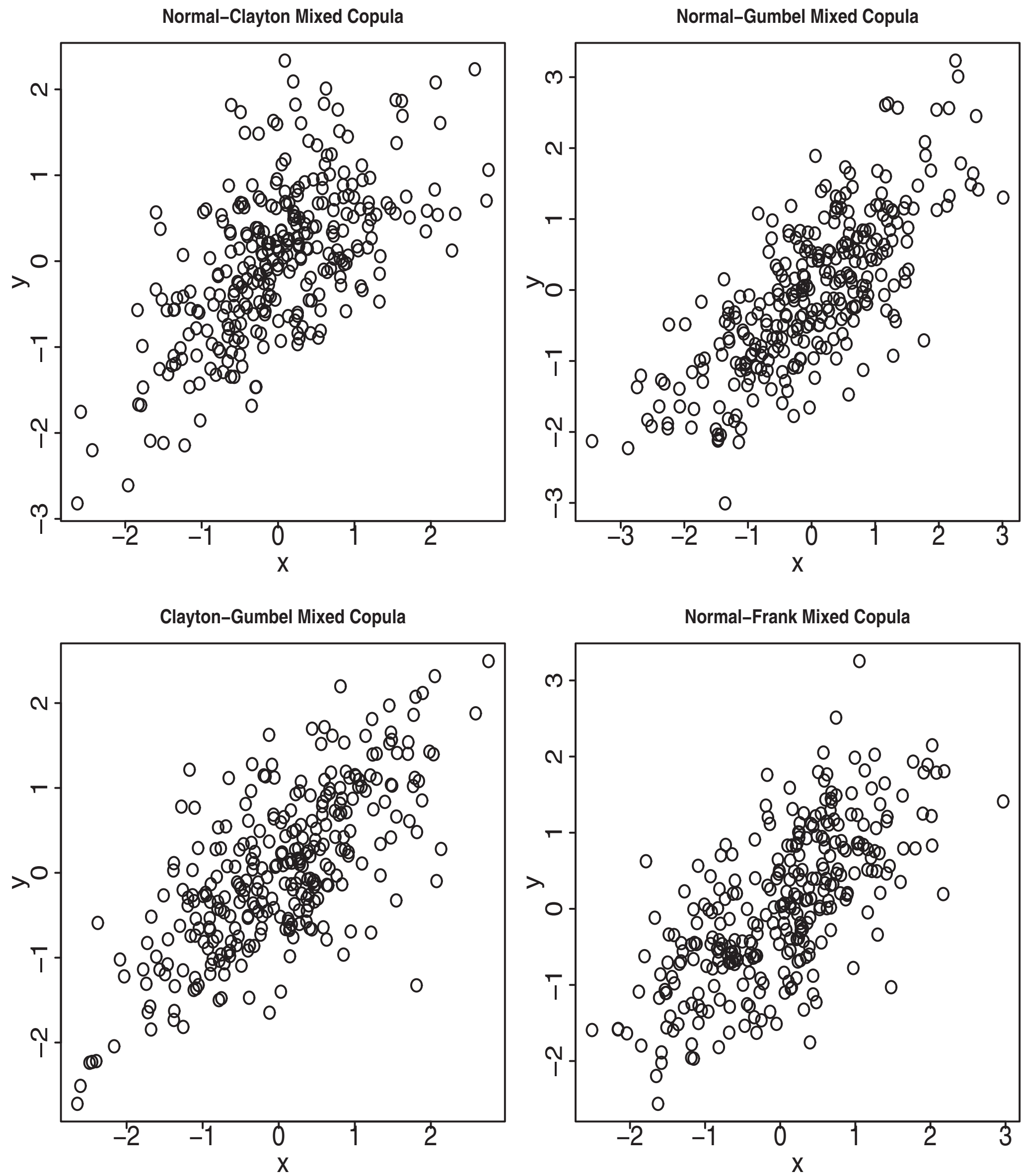

Figure 2. Scatterplots for mixed copulas with standard normal marginal distributions and Kendall's $\tau = 0.5$ as well as $\lambda_k = 0.5$.

are said to be identified, $C(\mathbf{u};\boldsymbol{\theta}) \equiv C^*(\mathbf{u};\boldsymbol{\theta}^*)$, if and only if $s = s^*$ and we can order the summations such that $\lambda_k = \lambda_k^*$ and $C_k(\mathbf{u};\theta_k) = C_k^*(\mathbf{u};\theta_k^*)$ for all possible values of $\mathbf{u}$ and $k = 1, \ldots, s$. Without loss of generality, it is assumed throughout the article that the mixed copula model under our study is identified, so that the identification issue is out of our concern.

Note that if the marginal distributions are not specified as parametric forms, model (1) becomes the following semiparametric copula

$$C(\mathbf{u};\boldsymbol{\theta},\boldsymbol{\lambda}) = \sum_{k=1}^{s} \lambda_k\, C_k(\mathbf{u};\theta_k) = \sum_{k=1}^{s} \lambda_k\, C_k(F_1(x_1), \ldots, F_p(x_p);\theta_k) \qquad (2)$$

with the marginal distributions $\{F_j(\cdot)\}$ in nonparametric form. This is an interesting feature of copula modeling: see Fermanian and Scaillet (2005) for several examples of practical relevance. The proposed copula selection procedure described in Section 3 (see later) can be applied to selecting the model given in (2) and this is an interesting future research topic. Indeed, it is under investigation by Cai and Wang (2013) and, hopefully, it will be available very soon in a separate article.

3. COPULA SELECTION VIA PENALIZED LIKELIHOOD WITH A SHRINKAGE

In this section, we present the selection and estimation procedures for a mixed copula model.

3.1 Penalized Likelihood

In this study, marginal distributions are assumed to be known and have a finite number of unknown parameters. Applying Sklar's theorem (1959) to Equation (1), the distribution function can be written as

$$F(\mathbf{x};\boldsymbol{\phi}) = \sum_{k=1}^{s} \lambda_k\, C_k(F_1(x_1;\alpha_1), \ldots, F_p(x_p;\alpha_p);\theta_k)$$

and the joint density function is given by

$$f(\mathbf{x};\boldsymbol{\phi}) = \prod_{j=1}^{p} f_j(x_j;\alpha_j) \sum_{k=1}^{s} \lambda_k\, c_k(F_1(x_1;\alpha_1), \ldots, F_p(x_p;\alpha_p);\theta_k),$$

where $c_k(\mathbf{u};\theta_k) = \partial^p C_k(\mathbf{u};\theta_k)/\partial \mathbf{u}$ is the mixed partial derivative of the copula $C(\cdot)$ and we assume these copula densities $c_1(\cdot), \ldots, c_s(\cdot)$ exist. When the sample is iid, the penalized log-likelihood takes the following form with a Lagrange multiplier term

$$Q(\boldsymbol{\phi}) = \sum_{t=1}^{T}\sum_{j=1}^{p} \ln f_j(X_{jt};\alpha_j) + \sum_{t=1}^{T} \ln \times \left[\sum_{k=1}^{s} \lambda_k\, c_k(F_1(X_{1t};\alpha_1), \ldots, F_p(X_{pt};\alpha_p);\theta_k) \right] - T\sum_{k=1}^{s} p_{\gamma_T}(\lambda_k) + \delta\left(\sum_{k=1}^{s}\lambda_k - 1\right). \quad (3)$$

Clearly, the first summand is the logarithm of the likelihood for marginal parameters and the second one is the logarithm of the likelihood for dependence parameters. The penalty function $p_{\gamma_T}(\cdot)$ in (3) is assumed to be nonconcave and γ_T is the tuning parameter, which controls the complexity of model and can be selected by some data-driven methods, such as the cross-validation (CV) and the generalized cross-validation (GCV); see Fan and Li (2001). To avoid overfitting, the penalty function is applied only to the weight parameters $\{\lambda_k\}$ because some of them might be estimated as zero if they are insignificant in the 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 copula components are selected and the corresponding parameters are estimated simultaneously. It will be shown that the estimator achieves the so-called *oracle* and *sparsity* properties; see Fan and Li (2001) for details. The last term in Equation (3) is for the constraint on $\{\lambda_k\}$.

For simplicity of presentation, it is assumed that the penalty function $p_{\gamma_T}(\cdot)$ is the same for all weight coefficients $\{\lambda_k\}$. The smoothly clipped absolute deviation (SCAD) penalty proposed by Fan (1997) is given by $p'_\gamma(\eta) = \gamma I(\eta \le \gamma) + (a\gamma - \eta)_+ I(\eta > \gamma)/(a-1)$ for some $a > 2$ and $\eta > 0$ with $p_\gamma(0) = 0$. The SCAD penalty function is applied to our simulated and empirical examples in Section 5 (see later) due to its good properties although other penalty functions may be applicable too in (3); see Fan and Lv (2010) for more discussions on the choice of various penalty functions.

3.2 Estimation Procedures

To estimate parameters, a full maximum likelihood approach is used. That is to maximize the logarithm of penalized likelihood function $Q(\boldsymbol{\phi})$ with respect to $\boldsymbol{\phi}$. The full maximum likelihood estimator is denoted by $\hat{\boldsymbol{\phi}}$. Because the maximum likelihood estimator does often not have a closed form, an iterative algorithm is suggested to compute the numerical solution.

It is well known that an initial value is a potential key for implementing an iterative algorithm. To this end, we propose using the two-step estimation as the initial value of the full maximum likelihood procedure. A two-step procedure is to estimate the marginal parameters from marginal likelihood and then to optimize the full likelihood with marginal parameters replaced by their estimators from the first step. More specifically, at the first step, the marginal parameters $\boldsymbol{\alpha}$ is estimated by maximizing the following likelihood corresponding to the marginal models $\sum_{t=1}^{T}\sum_{j=1}^{p} \ln f_j(x_{jt};\alpha_j)$, which is not affected by the copula parameters $\boldsymbol{\theta}$. Let $\bar{\boldsymbol{\alpha}}$ denote the solution of the above optimization problem. As demonstrated in Joe (1997), $\bar{\boldsymbol{\alpha}}$ is $\sqrt{T}$ consistent. Then at the second step, $\boldsymbol{\alpha}$ is substituted by its estimator obtained in Equation (3). Hence, a penalized likelihood function is given by

$$Q(\bar{\boldsymbol{\alpha}}, \boldsymbol{\theta}) = L(\bar{\boldsymbol{\alpha}}, \boldsymbol{\theta}) - T\sum_{k=1}^{s} p_{\gamma_T}(\lambda_k) + \delta\left(\sum_{k=1}^{s}\lambda_k - 1\right),$$

where $L(\cdot)$ is the sum of first two terms in Equation (3). Maximizing $Q(\bar{\boldsymbol{\alpha}}, \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ results in two-step penalized likelihood estimator $\bar{\boldsymbol{\theta}} = (\bar{\boldsymbol{\theta}}^\top, \bar{\boldsymbol{\lambda}}^\top)^\top$. In the next section, we demonstrate the consistency of our full maximum penalized likelihood estimator $\hat{\boldsymbol{\phi}}$ and derive its asymptotic distribution. Finally, the practical implementation issues are presented in Appendix A.

4. STATISTICAL PROPERTIES

In this section, we investigate the asymptotic behavior of the penalized likelihood estimator. A collection of some regularity conditions and asymptotic results are given in the following sections. The detailed proofs of the theorems, presented in this section, can be found in Appendix B with some lemmas and their proofs.

4.1 Notations and Assumptions

Before proceeding to the asymptotic theories of the proposed estimator, some notations are introduced and all possible assumptions are listed. The parameter space can be written by $\Gamma = \Theta_\alpha \times \Theta \times \Lambda$, where $\Theta_\alpha \subset \mathbb{R}^{d_\alpha}$, $\Theta = \Theta_1 \times \cdots \times \Theta_s \subset \mathbb{R}^{d_\theta}$ with $\Theta_i \subset \mathbb{R}^{d_i}$ and $d_\theta = \sum_{i=1}^{s} d_i$, and $\Lambda = \left\{(\lambda_1, \ldots, \lambda_s),\ \ \lambda_i \ge 0,\ \ \sum_{i=1}^{s}\lambda_i = 1,\ i = 1, \ldots, s\right\}$.

Obviously, $\Gamma \subset \mathbb{R}^{d_\gamma}$, where $d_\gamma = d_\alpha + s + d_\theta$. Let $\boldsymbol{\alpha}_0$ and $\boldsymbol{\lambda}_0$ be the unknown true parameters. Furthermore, $\boldsymbol{\lambda}_0 = (\lambda_{01}, \ldots, \lambda_{0s})^\top$ is partitioned as $\boldsymbol{\lambda}_{1,0} = (\lambda_{01}, \ldots, \lambda_{0r})^\top$ and $\boldsymbol{\lambda}_{2,0} = (\lambda_{0(r+1)}, \ldots, \lambda_{0s})^\top$, where r is the number of actual components. Without loss of generality, it is assumed that $\boldsymbol{\lambda}_{1,0}$ consists of all nonzero components of $\boldsymbol{\lambda}_0$ and $\boldsymbol{\lambda}_{2,0}$ contains all zero components. Obviously, $\boldsymbol{\lambda}_{2,0}$ is on the boundary of the parameter space. Similarly, $\boldsymbol{\theta}_{1,0} = (\theta_{01}, \ldots, \theta_{0r})^\top$ is a vector of true dependence parameters corresponding to $\boldsymbol{\lambda}_{1,0}$ and $\boldsymbol{\theta}_2^* = (\theta_{(r+1)}^*, \ldots, \theta_s^*)^\top$ is a vector of dependence parameters corresponding to $\boldsymbol{\lambda}_{2,0}$. It is worth to mention that $\boldsymbol{\theta}_2^*$ can be arbitrary (unidentified) because the corresponding weights are 0. Thus, the true parameters can be represented as

$$\boldsymbol{\phi}_0 = (\boldsymbol{\alpha}_0^\top, \boldsymbol{\lambda}_0^\top, \boldsymbol{\theta}_0^\top)^\top = (\boldsymbol{\alpha}_0^\top, \boldsymbol{\lambda}_{1,0}^\top, 0, \ldots, 0, \boldsymbol{\theta}_{1,0}^\top, \boldsymbol{\theta}_2^{*\top})^\top.$$

Finally, let Γ_0 be the collection of all the possible values of $\boldsymbol{\phi}_0$.

While deriving the asymptotic properties, it is typical to assume that the true parameter is an interior point of the parameter space. However, it is not always the case in this study. We would like to discuss the parameter space before we proceed to the assumptions.

Case I: Data are generated from a single copula. Obviously, $\boldsymbol{\lambda}_0 = (1, 0, \ldots, 0)^\top$ which is on the boundary of the parameter space Λ; $\boldsymbol{\theta}_0 = \left(\theta_{01}, \theta_2^* \in \Theta_2, \ldots, \theta_s^* \in \Theta_s\right)^\top$ and dependence parameters associated to the zero weights belong to an unidentified subset of the parameter space Θ.

Case II: True copula components are only a part of the candidate copulas when $1 < r < s$. In other words, $\boldsymbol{\lambda}_0 = (\lambda_{01}, \ldots, \lambda_{0r}, 0, \ldots, 0)^\top$ and $\boldsymbol{\theta}_0 = (\theta_{01}, \ldots, \theta_{0r}, \theta^*_{(r+1)} \in \Theta_{r+1}, \ldots, \theta_s^* \in \Theta_s)^\top$. It is easy to see that a subset of $\boldsymbol{\lambda}_0$ is on the boundary of the parameter space and a subset of $\boldsymbol{\theta}_0$ is unidentified. Therefore, to derive the asymptotic properties under nonregular situations, we need the following assumptions, with $\boldsymbol{\phi}_0$ an arbitrary fixed point in Γ_0.

Assumptions:

(A1) $\{\mathbf{X}_t\}$ is independent and identically distributed with the joint density $f(\mathbf{x}; \boldsymbol{\phi}) = \prod_{j=1}^p f_j(x_j; \alpha_j)\ c(\mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\theta})$, where $f(\mathbf{x}; \boldsymbol{\phi})$ has a common support and

$$c(\mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\theta}) = \sum_{k=1}^{s} \lambda_k\, c_k(F_1(x_1; \alpha_1), \ldots, F_p(x_p; \alpha_p); \theta_k).$$

Assume that the model is identified.

(A2) There exists an open subset $Q_\varepsilon \subset \mathbb{R}^{d_\gamma}$ containing Γ_0 such that, for almost all $\mathbf{x}$, $f_j(x_j; \alpha_j)$, $1 \le j \le p$, and $f(\mathbf{x}; \boldsymbol{\phi})$ admit all third derivatives with respect to $\boldsymbol{\phi}$, respectively. Also, we suppose that there exist functions $M_{jkl}(\mathbf{x}; \boldsymbol{\phi})$ such that for all j, k, and l, and $\mathbf{x}$,

$$\left| \frac{\partial^3}{\partial\phi_j \partial\phi_k \partial\phi_k} \{\ln f(\mathbf{x}; \boldsymbol{\phi})\} \right| \le M_{jkl}(\mathbf{x}; \boldsymbol{\phi})$$

and there exists a constant B such that $m_{jkl}(\mathbf{x}; \boldsymbol{\phi}) = E_{\phi_0}[M^2_{jkl}(\mathbf{x}; \boldsymbol{\phi})] < B$ for any fixed $\boldsymbol{\phi}_0 \in \Gamma_0$.

(A3) For any $\boldsymbol{\phi}_0 \in \Gamma_0$, the second logarithmic derivatives of $f(\mathbf{x}; \boldsymbol{\phi})$ satisfy the equations

$$I_{jk}(\boldsymbol{\phi}_0) = E_{\boldsymbol{\phi}_0}\left[\frac{\partial}{\partial\phi_j}\{\ln f(\mathbf{x}; \boldsymbol{\phi})\} \frac{\partial}{\partial\phi_k}\{\ln f(\mathbf{x}; \boldsymbol{\phi})\} \right] = -E_{\boldsymbol{\phi}_0}\left[\frac{\partial^2}{\partial\phi_j \partial\phi_k}\{\ln f(\mathbf{x}; \boldsymbol{\phi})\} \right],$$

and $I_{jk}(\boldsymbol{\phi}_0)$ is finite. Furthermore, the Fisher information matrix

$$\boldsymbol{I}(\boldsymbol{\phi}) = E\left\{ \left[\frac{\partial}{\partial\boldsymbol{\phi}} \ln f(\mathbf{x}; \boldsymbol{\phi})\right] \left[\frac{\partial}{\partial\boldsymbol{\phi}} \ln f(\mathbf{x}; \boldsymbol{\phi})\right]^\top \right\}$$

is positive definite at $\boldsymbol{\phi} = \boldsymbol{\phi}_0$, $\boldsymbol{\phi}_0 \in \Gamma_0$.

Remark 1. Note that the independence assumption on $\{\mathbf{X}_t\}$ given in Assumption A1 can be relaxed to a stationary time series case; see Cai and Wang (2008) for more details. In general, Q_ε can be expressed as $Q_\varepsilon \equiv \cup_{\phi_0 \in \Gamma_0} B_{\varepsilon(\phi_0)}(\boldsymbol{\phi}_0)$ for each $\varepsilon(\boldsymbol{\phi}_0) > 0$ depending on $\boldsymbol{\phi}_0$, and $B_\varepsilon(\boldsymbol{\phi})$ is an open ball of radius ε centered at $\boldsymbol{\phi}$.

Remark 2. It is easy to show that $E_{\phi_0}\left[\partial \ln f(\mathbf{x}; \boldsymbol{\phi})/\partial\boldsymbol{\phi}_j\right] = 0$ for $j = 1, \ldots, d_\gamma$ and all $\boldsymbol{\phi}_0 \in \Gamma_0$.

4.2 Large Sample Theory

First, we establish the convergence rate of the penalized likelihood estimator. Toward this end, define $b_T = \max_{1\le k\le s}\{p'_{\gamma_T}(\lambda_{0k}), \lambda_{0k} \ne 0\}$.

Theorem 1. Under the regularity conditions A1–A3, if $\max_{1\le k\le s}\{|p''_{\gamma_T}(\lambda_{0k})|, \lambda_{0k} \ne 0\} \to 0$, there exists a local maximizer $\widehat{\boldsymbol{\phi}}$ of $Q(\boldsymbol{\phi})$ defined in Equation (3), which has the property that there exists a $\boldsymbol{\phi}_0(\widehat{\boldsymbol{\phi}}) \in \Gamma_0$ which depends on $\widehat{\boldsymbol{\phi}}$ such that

$$\|\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0(\widehat{\boldsymbol{\phi}})\| = O_p(T^{-1/2} + b_T),$$

where $\|\cdot\|$ represents the Euclidean norm.

Remark 3. We can partition $\boldsymbol{\phi}$ as an identified subset $\boldsymbol{\phi}_I$ and $\boldsymbol{\theta}$ as an unidentified subset $\boldsymbol{\theta}_2$, where $\boldsymbol{\phi}_I = (\boldsymbol{\alpha}^\top, \boldsymbol{\lambda}^\top, \boldsymbol{\theta}_1^\top)^\top$. There exists a unique $\boldsymbol{\phi}_{I0}$ such that $\widehat{\boldsymbol{\phi}_I} \to \boldsymbol{\phi}_{I0}$ with probability 1. While the unidentified dependence parameter estimate $\widehat{\boldsymbol{\theta}}_2$ which corresponds to the zero weight converges to a point in an unidentifiable subset, which is a set including all the arbitrary fixed point $\boldsymbol{\theta}_0 \in \Gamma_0$. Also, Theorem 1 demonstrates that the estimator can achieve the square root-T convergence rate when $\gamma_T = O(T^{-1/2})$. When $\gamma_T \to 0$, $b_T = 0$ for the hard and SCAD penalty function. Therefore, $\hat{\boldsymbol{\phi}}$ is $\sqrt{T}$ consistent when hard and the SCAD penalty function is used to the penalized likelihood function. However, for the L_1 penalty function, $b_T = \gamma_T$, hence, the square root-T consistency requires that $\gamma_T = O(T^{-1/2})$.

Next, we present the oracle property of the penalized likelihood estimator. Before doing so, the following notations are needed. Let $\boldsymbol{\phi}_1 = (\boldsymbol{\alpha}^\top, \boldsymbol{\lambda}_1^\top, \boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$ without zero weights. In addition, we use $\boldsymbol{\phi}_r$ to denote the parameters after the parameters corresponding to the zero weights are removed and we reorder them. That is, $\boldsymbol{\phi}_r = (\boldsymbol{\lambda}_1^\top, \boldsymbol{\theta}_1^\top, \boldsymbol{\alpha}^\top)^\top$ without zero weights and unidentified parameters. Let $\Gamma_r \subset \mathbb{R}^q$ be the parameter space of $\boldsymbol{\phi}_r$, where $q = r + \sum_{k=1}^r d_k + d_{\alpha_1}$. Thus, $\Gamma_r \subset \Gamma$. Furthermore, denote, $\boldsymbol{\Sigma}_1 = \operatorname{diag}\{p''_{\gamma_T}(\lambda_{01}), \ldots, p''_{\gamma_T}(\lambda_{0r}), 0, \ldots, 0\}_{q\times q}$ and $\boldsymbol{b}_1 = (p'_{\gamma_T}(\lambda_{01}), \ldots, p'_{\gamma_T}(\lambda_{0r}), 0, \ldots, 0)^\top_{q\times 1}$. Finally, let $\boldsymbol{c} = (c_1, \ldots, c_r, 0, \ldots, 0)^\top_{q\times 1}$, where $c_i = \lim_{T\to\infty} p'_{\gamma_T}(\lambda_{0i})$.

Definition 2. The cone C_Γ with vertex at $\boldsymbol{\phi}_0$ is a local approximation of the set Γ at $\boldsymbol{\phi}_0$ if $\inf_{y\in C_\Gamma}\|x - y\| = o(\|x - \boldsymbol{\phi}_0\|)$ for all $x \in \Gamma$ such that $\|x - \boldsymbol{\phi}_0\| \to 0$ and $\inf_{y\in\Gamma}\|z - y\| = o(\|z - \boldsymbol{\phi}_0\|)$ for all $z \in C_\Gamma$ such that $\|z - \boldsymbol{\phi}_0\| \to 0$.

Theorem 2. Under the regularity conditions A1–A3, if $b_T = O(T^{-1/2})$, $\sqrt{T}\gamma_T \to \infty$, and $\liminf_{T\to\infty} \liminf_{\lambda_k\to 0+} p'_{\gamma_T}(\lambda_k)/\gamma_T > 0$, the root-$T$ consistent estimator $\widehat{\boldsymbol{\phi}} = (\widehat{\boldsymbol{\phi}}_1, \widehat{\boldsymbol{\lambda}}_2)^\top$ in Theorem 1 must satisfy:

(a) Sparsity: $\widehat{\boldsymbol{\lambda}}_2 = 0$.

(b) Asymptotic distribution:

(i) When $\boldsymbol{\lambda}_{10} = (\lambda_{01}, \ldots, \lambda_{0r})^\top$, $\lambda_{0i} \in (0, 1)$ for $i = 1, \ldots, r$ and $r > 1$,

$$\sqrt{T}[\{\boldsymbol{I}_1(\boldsymbol{\phi}_{r0}) + \boldsymbol{\Sigma}_1\}(\widehat{\boldsymbol{\phi}}_r - \boldsymbol{\phi}_{r0}) + \boldsymbol{b}_1] \to N(0,\ \boldsymbol{I}_1(\boldsymbol{\phi}_{r0})),$$

where $\boldsymbol{I}_1(\boldsymbol{\phi}_r)$ is the Fisher information when all zero effects are removed.

(ii) When $\boldsymbol{\lambda}_{10} = \lambda_{01} = 1$,

$$\sqrt{T}\left[\boldsymbol{I}_1(\boldsymbol{\phi}_{r0}) + \boldsymbol{\Sigma}_1\right](\widehat{\boldsymbol{\phi}}_r - \boldsymbol{\phi}_{r0}) \to \check{\boldsymbol{\phi}}_r,$$

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

$$\begin{pmatrix} Z_{11} - c_1 \\ Z_{12} - c_2 \\ \vdots \\ Z_{1q} - c_q \end{pmatrix} I\{Z_{11} > c_1\} + \begin{pmatrix} 0 \\ Z_{12} - c_2 - (\boldsymbol{I}_1^{21}/\boldsymbol{I}_1^{11})(Z_{11} - c_1) \\ \vdots \\ Z_{1q} - c_q - (\boldsymbol{I}_1^{q1}/\boldsymbol{I}_1^{11})(Z_{11} - c_1) \end{pmatrix} \times I\{Z_{11} < c_1\}$$

with $Z_1 = (Z_{11}, Z_{12}, \ldots, Z_{1q})^\top$ being a random variable with multivariate Gaussian distribution with mean 0 and covariance matrix $\boldsymbol{I}_1(\boldsymbol{\phi}_{r0})$, $\boldsymbol{\phi}_r \in C_{\Gamma_r} - \boldsymbol{\phi}_{r0}$, and $\boldsymbol{I}_1^{ij} = \boldsymbol{I}_1^{ij}(\boldsymbol{\phi}_{r0})$ being elements of matrix $\boldsymbol{I}_1^{-1}(\boldsymbol{\phi}_{r0})$.

The part (a) of Theorem 2 shows that the penalized likelihood estimator correctly estimates some weights 0 with positive probability. This leads to the so-called sparsity property. The oracle property is presented in the part (b) of Theorem 2. 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, introduced originally by Donoho and Johnstone (1994), and perform as well as the maximum likelihood estimators for estimating $\boldsymbol{\lambda}_1$ if $\boldsymbol{\lambda}_2 = 0$ would be known in advance.

Now, we examine the asymptotic behavior for boundary parameters. When $r = 1$, it follows from Theorem 2 that $\check{\boldsymbol{\phi}}_{r1} = (Z_{11} - c_1)I\{Z_{11} > c_1\}$. Furthermore, if $\boldsymbol{I}_1^{21} = 0$, the second component of $\check{\boldsymbol{\phi}}_r$ is normal as well as the last two components.

Remark 4. We comment that when $\{\mathbf{X}_t\}$ is a stationary time series satisfying some regularity conditions, all the asymptotic results in Theorems 1 and 2 hold true with a change of the asymptotic covariance, which has the form of $\Sigma^* = \text{var}(l_\phi(\mathbf{x}_1; \boldsymbol{\phi}_{r0}, 0)) + 2\sum_{t=2}^{\infty} \text{cov}(l_\phi(\mathbf{x}_1, \boldsymbol{\phi}_{r0}, 0), l_\phi(\mathbf{x}_t, \boldsymbol{\phi}_{r0}, 0))$, where $l(x, \phi) \equiv \ln f(\mathbf{x}, \phi)$ and $l_\phi(\mathbf{x}, \phi) = \partial l(\mathbf{x}, \phi)/\partial\phi$. For the detailed descriptions and proofs, see Cai and Wang (2008) for a more general setting under time series context. It is noted that a consistent estimation of the asymptotic covariance Σ^*, such as heteroscedasticity and autocorrelation consistent estimator of Newey and West (1987), is particularly important in regard of the empirical example.

5. EMPIRICAL STUDIES

To illustrate the proposed methods, two simulated examples and a real example are considered in this section.

5.1 Simulated Examples

The first simulated example considers some single copula functions and mixed copulas and the second example discusses the misspecified model which one or more actual copula components are missing from the candidate copula families. The first example demonstrates that indeed, the proposed estimation procedures work reasonably well in the finite sample case. The second example illustrates that even the working is misspecified, the proposed model selection procedure can find a *best* (which means as close as possible) model to approximate the true model.

Our data-generating process (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 $C(u, v; \boldsymbol{\theta})$. Considering the popularity of the normal distribution and computational convenience, the marginal distributions are normally distributed and marginal parameters $(\mu_x, \sigma_x) = (1, 0.5)$ and $(\mu_y, \sigma_y) = (0, 2)$. Four commonly used copulas, Gaussian, Clayton, Gumbel, and Frank, consist of our candidate copula families. Indeed, all possible combinations of these four copulas have an ability to capture most of the possible dependence structures. The Gaussian copula is widely used in financial fields, while the Frank copula exhibits no tail dependence like the Gaussian copula. However, compared to the Gaussian copula, dependence in the Frank copula is stronger in the center of the distribution, as evidenced from the fanning out in the tails. In contrast to the Gaussian and Frank copulas, the Clayton and Gumbel copulas exhibit asymmetric dependence. The Clayton dependence is strong in the left tail, which implies 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, the Gumbel is an appropriate modeling choice when two variables are likely to simultaneously increase. Therefore, the working mixed copula can be written as

$$C(u, v; \boldsymbol{\theta}) = \lambda_1 C_{\text{Ga}}(u, v; \theta_1) + \lambda_2 C_{\text{Cl}}(u, v; \theta_2) + \lambda_3 C_{\text{Gu}}(u, v; \theta_3) + \lambda_4 C_{\text{Fr}}(u, v; \theta_4),$$

where $C_{\text{Ga}}(\cdot)$ denotes the Gaussian copula, $C_{\text{Cl}}(\cdot)$ is for the Clayton, $C_{\text{Gu}}(\cdot)$ stands for the Gumbel, and $C_{\text{Fr}}(\cdot)$ depicts the Frank.

Example 1. This simulated example considers three sample sizes, $T = 400, 700$, and 1000. Observations are simulated from different copula models by following the GDP described above and the penalized maximum likelihood estimators are computed. The simulation is repeated 1000 times.

First, data are generated from four single copulas, Gaussian, Clayton, Gumbel, and Frank, separately. Thus, the weight corresponding to the actual single copula is 1 and 0 otherwise. Due to the limited space, we omit the results of the marginal parameters. The simulated results for copula parameters are summarized in Tables 1, 2, and 3. Table 1 shows biases and mean squared errors (MSEs) of the penalization maximum likelihood estimators for dependence parameters λ_j and θ_j. It is clear that the bias becomes smaller when the sample size is getting larger. Table 2 shows the initial values and the estimated values of unidentified parameters. One can easily see that the estimates are very close to the arbitrarily given initial values. In other words, the penalized likelihood estimator of unidentified parameter converges to an arbitrary value. This is in line with our theory. In Table 3, the values without parentheses correspond to the percentages

Table 1. Biases and MSEs of the single copula parameter estimates in Example 1

T		Gaussian (λ_1, θ_1)	Clayton (λ_2, θ_2)	Gumbel (λ_3, θ_3)	Frank (λ_4, θ_4)
400	Bias	(−0.026, −0.004)	(0.000, 0.011)	(0.000, −0.518)	(−0.081, −0.104)
	MSE	(0.008, 0.001)	(0.001, 0.117)	(0.001, 0.604)	(0.026, 0.124)
700	Bias	(−0.009, −0.002)	(0.000, −0.003)	(0.001, −0.358)	(−0.020, −0.054)
	MSE	(0.002, 0.001)	(0.001, 0.079)	(0.001, 0.304)	(0.006, 0.024)
1000	Bias	(−0.010, −0.002)	(0.000, −0.002)	(0.000, −0.236)	(−0.015, −0.038)
	MSE	(0.004, 0.000)	(0.001, 0.034)	(0.001, 0.147)	(0.003, 0.014)

of correctly selected copulas and the values within parentheses correspond to the percentages that the copulas are selected incorrectly. One can observe from Table 3 that the performance of the proposed method is reasonably well for choosing an appropriate individual copula from a mixed model. There is 100% of chance to choose the correct single copula functions from which each data point is drawn. The percentage that the incorrect copula is selected is small. For example, when $T = 1000$, only 3.2% of chance to select the Frank copula when data are drawn from a Gaussian copula and 7.8% 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 almost same type of dependence structure, so it is not easy to distinguish them. In sum, in view of the above empirical results, the proposed method works reasonably well.

Second, we simulate three mixed copulas with only two components. Table 4 displays the true values of parameters in the mixed copula. The simulation results are presented in Tables 5, 6, and 7. Table 5 shows the percentages corresponding to the correctly and incorrectly (in parentheses) selected copula, Tables 6 displays biases and MSEs of the penalized maximum likelihood estimates, and Table 7 gives the initial values and the estimates of unidentified parameters. We only comment on the results for $T = 1000$ from Table 5. For the mixture of the Gaussian and Clayton copulas which describes lower tail dependence, there is 100% probability of selecting the appropriate copulas. For the mixed copula consisting of the Clayton and Gumbel copulas, the Clayton is selected in all replications and the Gumbel is chosen 80.4%. The results for the last one are very promising because the Gaussian copula is selected with 100% and the Frank copula is selected with 98.2%. Therefore, one can conclude that the proposed methods for the models simulated in this example can capture well different dependence structures.

Table 2. Average estimated values of the unidentified copula parameters for single copula in Example 1 when $T = 1000$

Model	θ_{10}	θ_{20}	θ_{30}	θ_{40}	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
Gaussian		5.000	6.100	6.900		4.842	6.063	6.922
		4.000	5.100	5.900		3.826	5.043	5.953
		6.000	7.100	7.900		5.846	7.076	7.908
Clayton	0.600		6.100	6.900	0.948		5.811	6.736
	0.400		5.100	5.900	0.745		4.700	5.933
	0.800		7.100	7.900	0.847		6.942	7.906
Gumbel	0.600	5.000		6.900	0.683	5.109		6.912
	0.400	4.000		5.900	0.349	4.198		5.928
	0.800	6.000		7.900	0.466	6.050		7.894
Frank	0.600	5.000	6.100		0.737	4.918	5.953	
	0.400	4.000	5.100		0.860	3.979	4.813	
	0.800	6.000	7.100		0.263	5.885	7.014	

NOTE: θ_{i0} is initial values and $\hat{\theta}_i$ is estimate for $i = 1$ to 4.

Table 3. Percentage that the corresponding copula was chosen correctly (incorrectly) in Example 1

Model (×100)	T	Gaussian	Clayton	Gumbel	Frank
Gaussian	400	1.000	(0.000)	(0.000)	(0.092)
	700	1.000	(0.000)	(0.000)	(0.038)
	1000	1.000	(0.000)	(0.000)	(0.032)
Clayton	400	(0.010)	1.000	(0.000)	(0.000)
	700	(0.000)	1.000	(0.000)	(0.000)
	1000	(0.000)	1.000	(0.000)	(0.000)
Gumbel	400	(0.028)	(0.000)	1.000	(0.000)
	700	(0.000)	(0.000)	1.000	(0.000)
	1000	(0.000)	(0.000)	1.000	(0.000)
Frank	400	(0.289)	(0.000)	(0.000)	1.000
	700	(0.074)	(0.000)	(0.000)	1.000
	1000	(0.078)	(0.000)	(0.000)	1.000

NOTE: Values without parentheses are the percentages that copula in the single copula was chosen correctly. Values with parentheses are the percentages that copula not in the single copula was chosen incorrectly.

Example 2. In this example, the purpose is to see how the proposed selection method works if the working model is misspecified. To this end, we consider the following three misspecified models when our candidate copulas are not included in all the actual components.

I: Data are generated from the single Gaussian copula but working model is

$$C(u, v) = \lambda_1 C_{\mathrm{Cl}}(u, v) + \lambda_2 C_{\mathrm{Gu}}(u, v) + \lambda_3 C_{\mathrm{Fr}}(u, v).$$

II: Data are generated from the single Clayton copula but the working model is

$$C(u, v) = \lambda_1 C_{\mathrm{Ga}}(u, v) + \lambda_2 C_{\mathrm{Sg}} C(u, v) + \lambda_3 C_{\mathrm{Gu}}(u, v) + \lambda_4 C_{\mathrm{Fr}}(u, v).$$

Table 4. Actual values of the mixed copula parameters in Example 1

Model	(λ_1, θ_1)	(λ_2, θ_2)	(λ_3, θ_3)	(λ_4, θ_4)
Gaussian + Clayton	(1/2,0.6)	(1/2,5)		
Clayton + Gumbel		(1/2,5)	(1/2,2.6)	
Gaussian + Frank	(1/2,0.6)			(1/2,4)

Table 5. Percentage that the corresponding copula was chosen correctly (incorrectly) in Example 1

Model (×100)	T	Gaussian	Clayton	Gumbel	Frank
Gaussian + Clayton	400	0.952	1.000	(0.000)	(0.141)
	700	0.995	1.000	(0.000)	(0.062)
	1000	1.000	1.000	(0.009)	(0.010)
Clayton + Gumbel	400	(0.522)	1.000	0.528	(0.000)
	700	(0.378)	1.000	0.603	(0.000)
	1000	(0.203)	1.000	0.804	(0.000)
Gaussian + Frank	400	0.992	(0.000)	(0.011)	0.883
	700	1.000	(0.000)	(0.000)	0.904
	1000	1.000	(0.000)	(0.000)	0.982

NOTE: Values without parentheses are the percentages that copula in the mixed copula was chosen correctly. Values with parentheses are the percentages that copula not in the mixed copula was chosen incorrectly.

Table 7. Average estimated values of the unidentified copula parameters for mixed copula in Example 1

Model	θ_{10}	θ_{20}	θ_{30}	θ_{40}	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
Gaussian + Clayton			2.600	4.000			2.010	3.938
			2.000	3.000			2.705	4.009
			3.200	5.000			2.910	5.049
Clayton + Gumbel	0.600			4.000	0.839			4.803
	0.400			3.000	0.272			4.018
	0.800			5.000	0.692			4.940
Gaussian + Frank		5.000	2.600			4.928	2.189	
		4.000	2.000			3.928	2.885	
		6.000	3.200			5.925	2.559	

NOTE: θ_{i0} is initial values and $\hat{\theta}_i$ is estimate for $i = 1$ to 4.

III: Data are generated from the mixed Clayton and Gumbel copula with equal weights but the working model is

$$C(u, v) = \lambda_1 C_{\mathrm{Ga}}(u, v) + \lambda_2 C_{\mathrm{Sg}}(u, v) + \lambda_3 C_{\mathrm{Gu}}(u, v) + \lambda_4 C_{\mathrm{Fr}}(u, v),$$

where $C_{\mathrm{Sg}}(\cdot)$ means a survival Gumbel copula function. The Clayton copula is absent from both Models II and III, while the survival Gumbel copula which exhibits the same left tail dependence pattern as the Clayton copula is added to these two models. For sample size of $T = 1000$, the simulation is replicated 1000 times. Table 8 reports the results of the misspecified models, including the percentage that the corresponding copula function is selected and the estimated value of the related copula parameter in parentheses.

From Table 8, one can observe that in Model I, the Frank copula which is similar to the Gaussian copula is selected with 100% for the missed Gaussian copula. In Model II, the working model includes the Gaussian, survival Gumbel, Gumbel, and Frank copulas. But the true model includes only the Clayton copula. The results show that the survival Gumbel copula exhibiting left tail dependence like the Clayton copula is chosen for all replications. Clearly, the similar result can be observed for Model III. The survival Gumbel is selected to replace the Clayton copula and the Gumbel copula appears with 88.4%. For these misspecified models, copulas exhibiting the same dependence patterns are selected. In other words, the proposed selection method can choose the *best* copula based on a combination of several copulas to approximate the actual dependence structure.

5.2 A Real Example

Example 3. We consider four international stock market indices, S&P500 (U.S.), FTSE100 (UK), Nikkei (JP), and Hang Seng (HK). The dataset is a collection of the monthly returns from January 1987 to February 2007 and includes total 242 observations, as mentioned in Section 1. The main purpose of the empirical analysis of this dataset is to examine the comovement of returns among these four markets. First, the Kolmogorov–Smirnov (KS) tests are performed and the results show that t-distribution is an appropriate fit of the marginal distributions. So, the marginal distributions are assumed to

Table 6. Biases and MSEs of the copula parameter estimates for mixed copula in Example 1

Model	T		(λ_1, θ_1)	(λ_2, θ_2)	(λ_3, θ_3)	(λ_4, θ_4)
Gaussian + Clayton	400	Bias	(−0.035, 0.015)	(−0.015, −0.045)		
		MSE	(0.019, 0.006)	(0.003, 0.107)		
	700	Bias	(−0.009, 0.009)	(−0.005, −0.050)		
		MSE	(0.004, 0.002)	(0.002, 0.064)		
	1000	Bias	(−0.002, 0.000)	(−0.003, 0.013)		
		MSE	(0.002, 0.001)	(0.001, 0.046)		
Clayton + Gumbel	400	Bias		(0.094, 0.541)	(−0.387, −0.852)	
		MSE		(0.050, 0.494)	(0.171, 3.367)	
	700	Bias		(−0.036, −0.253)	(−0.294, −0.717)	
		MSE		(0.013, 0.171)	(0.133, 1.733)	
	1000	Bias		(−0.001, −0.208)	(−0.241, −0.690)	
		MSE		(0.013, 0.144)	(0.108, 1.787)	
Gaussian + Frank	400	Bias	(0.083, −0.286)			(−0.085, 0.069)
		MSE	(0.054, 0.101)			(0.054, 0.061)
	700	Bias	(0.122, −0.272)			(−0.122, 0.144)
		MSE	(0.049, 0.076)			(0.049, 0.047)
	1000	Bias	(0.084, −0.272)			(−0.084, 0.130)
		MSE	(0.028, 0.076)			(0.028, 0.033)

Table 8. Percentage that the corresponding copula function is selected and the estimated value of the related copula parameter θ in parentheses for Example 2

Model I	Clayton	Gumbel	Frank	
Percentage (Estimate of θ)	0.000	0.000	1.000 (7.107)	
Model II	Gaussian	SGumbel	Gumbel	Frank
Percentage (Estimate of θ)	0.000	0.996 (3.411)	0.000	0.152 (4.811)
Model III	Gaussian	SGumbel	Gumbel	Frank
Percentage (Estimate of θ)	0.000	0.994 (3.320)	0.884 (4.501)	0.000

Table 9. Linear correlation coefficient and Kendall's τs in parentheses across four international markets

	UK	JP	HK
US	0.780 (0.509)	0.423 (0.262)	0.614 (0.408)
UK		0.389 (0.235)	0.632 (0.385)
JP			0.343 (0.216)

follow the t-distribution. Second, the linear correlation coefficients and Kendall's τ's (in parentheses) across the four markets are computed and displayed in Table 9. One can see from Table 9 that the U.S. and U.K. markets have the strongest correlation based on both linear correlation and Kendall's τ. Finally, the penalized likelihood estimators are computed for all six pairs of markets. Due to space limitations, we only present the results of the copula parameter estimator.

Table 10 reports the estimation results along with the 95% confidence intervals for all nonzero parameters in all models. It can be seen from Table 10 that all 95% confidence intervals for $\{\lambda_k\}$ on Gaussian and Clayton copulas as well as three 95% confidence intervals for $\{\lambda_k\}$ on Frank copula do not contain zero. This implies that indeed, they are statistically significant away from zero. Interestingly, the weight on Gumbel copula is zero for all the pairs, which indicates that no right tail dependence appears for all pairs. In other words, the chance that two markets boom together is close to zero or very small. Moreover, the weight on Frank copula is zero for three pairs (U.S.-HK, U.S.-UK, and UK-HK) without JP. This means that the Gaussian copula is enough to characterize the central dependence for these three pairs. Furthermore, it is interesting to see that all coefficients $\{\lambda_k\}$ on Clayton copula are statistically significant away from zero. This implies that all the pairs have the left tail dependence, which means that any two different markets crash together although the degrees may be different. Therefore, the international stock markets have asymmetric dependence structure. To explain this asymmetry, one may conclude that investors are more sensitive to bad news than good news in other markets. When a market suffers the loss from the crash, investors in other markets may react immediately and make some moves to avoid the possible loss on the stock. These reactions may drag the market down. On the contrary, people may not response too much on the boom of another market. Hu (2006) proposed to use the Gaussian, Gumbel, and survival Gumbel mixture model to model the correlation structure among these four markets. The above two findings (the right tail dependence does not exist but the left tail dependence exists) are similar to the results of Hu (2006). But our results support the Gaussian or/and Frank type of dependence which is in contrast to that in Hu (2006), who generally found a weight of close to 0 on a Gaussian copula. This suggests that our method detects both linear and nonlinear dependence structures, while Hu (2006) detected only nonlinear dependence.

6. CONCLUSION

In this article, we proposed a new data-driven copula selection approach through penalized likelihood plus a shrinkage operator to select a mixture copula with applications in risk management. The proposed method selects an appropriate copula function and estimates the related parameters simultaneously. We derived the asymptotic properties of the proposed penalized likelihood estimator, including the rate of convergence and asymptotic normality and abnormality. Particularly, when the true coefficient parameters may be on the boundary of the parameter space and the dependence parameters are in an unidentified subset of

Table 10. Estimates of copula parameters with the 95% confidence interval in parentheses for international markets in the empirical example

	Copulas	Gaussian	Clayton	Gumbel	Frank
$\boldsymbol{\lambda}$	U.S.-UK	0.418(0.311,0.525)	0.582(0.472,0.692)	0	0
	U.S.-JP	0.125(0.015,0.235)	0.814(0.725,0.903)	0	0.061(0.045,0.077)
	U.S.-HK	0.528(0.414,0.642)	0.472(0.358,0.586)	0	0
	UK-JP	0.121(0.044,0.198)	0.771(0.677,0.865)	0	0.107(0.022,0.192)
	UK-HK	0.689(0.536,0.842)	0.311(0.161,0.461)	0	0
	JP-HK	0.139(0.049,0.223)	0.772(0.659,0.884)	0	0.089(0.064,0.114)
$\boldsymbol{\theta}$	U.S.-UK	0.849(0.845,0.853)	1.525(1.472,1.577)		
	U.S.-JP	0.840(0.823,0.857)	0.491(0.457,0.525)		0.008(0.007,0.009)
	U.S.-HK	0.536(0.490,0.582)	1.703(1.641,1.765)		
	UK-JP	0.919(0.914,0.924)	0.475(0.440,0.510)		0.500(0.224,0.776)
	UK-HK	0.558(0.550,0.565)	1.320(1.270,1.370)		
	JP-HK	0.433(0.282,0.584)	0.428(0.394,0.462)		0.030(0.023,0.037)

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. We proposed an ad hoc computational algorithm (EM algorithm) to compute the optimization of the penalized likelihood function. Finally, we conducted Monte Carlo simulation studies to illustrate the finite sample performance of the proposed method and investigated a real dataset.

Some interesting future research topics related to this article should be mentioned. First, we believe that it is not difficult to generalize the proposed method and the related theory to the time series data and a more general setting like finite mixture distribution. Second, it would be very useful to investigate the case that the number of basis copulas s is large, say $s = s_T$ going to infinity in a certain rate, theoretically and empirically. Third, it would be very interesting to consider the case that the marginal distributions are purely nonparametric, so that the model becomes a semiparametric setting. Furthermore, the computational implementation should be addressed when the dimension of copula function is high. Finally, the proposed method potentially can be applied to the analysis of multivariate financial data, such as multivariate GARCH models studied by Lee and Long (2009) and their extensions, and predictor selection in portfolio choice investigated by Cai, Peng, and Ren (2011).

APPENDIX A: PRACTICAL COMPUTATIONAL ISSUES

Here, we need to address some computational issues in practice to make implementation easily.

A.1 Choice of the Tuning Parameters

To implement the approach proposed in Section 3 practically, one needs to choose the appropriate tuning parameters $\gamma = \gamma_T$ and a in the SCAD penalty function. In practical implementation, it is suggested using the multifold cross-validation method of estimating γ and a as suggested by Cai, Fan, and Yao (2000) and Fan and Li (2001) for regression settings, and it is briefly described next.

Denote D as the full dataset. 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 (γ, a), $D - D_i$ is used to estimate $\boldsymbol{\phi}$ and D_i is used for evaluation. The penalized maximum likelihood estimator $\widehat{\boldsymbol{\phi}}_i$ can be used to construct the following cross-validation criterion based on the test data D_i

$$\mathrm{CV}(\gamma, a) = \sum_{i=1}^{m} \sum_{(x_t, y_t)\in D_i} Q(\widehat{\boldsymbol{\phi}}_i).$$

By maximizing $\mathrm{CV}(\gamma, a)$, the data-driven choice of tuning parameters is selected. As suggested by Cai, Fan, and Yao (2000) and Fan and Li (2001), m can be chosen as 4 or 5 in a real implementation and $m = 5$ is taken in our empirical studies in Section 5.

A.2 Maximization of Penalized Likelihood Function

It is clear that there is not an explicit expression for maximum likelihood estimator of Equation (3) so it needs a numerical method in practical implementation. One of the most popular algorithms for finding the maximum likelihood estimation of the finite mixture model is the expectation maximization (EM) algorithm of Dempster, Laird, and Rubin (1977) due to its easy numerical computation. The EM algorithm is described below in detail. The main idea of the 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.

To maximize Equation (3), we take the derivative and set it equal to zero. The only closed-form expression we can get is the weight expression. The first-order condition of weights can be expressed as follows:

$$\frac{\partial Q(\boldsymbol{\phi})}{\partial \lambda_k} = \sum_{t=1}^{T} \frac{f_0(\mathbf{x}_t, \boldsymbol{\alpha}) c_k(\mathbf{u}_t; \alpha_k, \theta_k)}{f(\mathbf{x}_t; \boldsymbol{\phi})} - T p'_{\gamma_T}(\lambda_k) + \delta = 0, \quad \text{(A.1)}$$

where $f_0(\mathbf{x}, \boldsymbol{\alpha}) = \prod_{j=1}^{p} f_j(x_j; \alpha_j)$. By multiplying λ_k on both sides of Equation (A.1), one has

$$\sum_{t=1}^{T} \frac{f_0(\mathbf{x}_t, \boldsymbol{\alpha}) \lambda_k c_k(\mathbf{u}_t; \alpha_k, \theta_k)}{f(\mathbf{x}_t; \boldsymbol{\phi})} - T \lambda_k p'_{\gamma_T}(\lambda_k) + \lambda_k \delta = 0. \quad \text{(A.2)}$$

Then, by summing over Equation (A.2) for all k, one obtains $\delta = T\left[\sum_{k=1}^{s} \lambda_k p'_{\gamma_T}(\lambda_k) - 1\right]$. Therefore, plugging δ back to Equation (A.2) leads to

$$\lambda_k = \left[T \lambda_k p'_{\gamma_T}(\lambda_k) - \sum_{t=1}^{T} \frac{f_0(\mathbf{x}_t, \boldsymbol{\alpha}) \lambda_k c_k(\mathbf{u}_t; \alpha_k, \theta_k)}{f(\mathbf{x}_t; \boldsymbol{\phi})}\right] \delta^{-1}.$$

A.3 E-step and M-step

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

$$\lambda_k^{(m)} = \left[\lambda_k p'_{\gamma_T}(\lambda_k^{(m-1)}) - \frac{1}{T} \times \sum_{t=1}^{T} \frac{f_0(\mathbf{x}_t, \boldsymbol{\alpha}^{(m-1)}) \lambda_k^{(m-1)} c_k(\mathbf{u}_t; \alpha_k^{(m-1)}, \theta_k^{(m-1)})}{f(\mathbf{x}_t; \boldsymbol{\phi}^{(m-1)})}\right] \times \left[\sum_{k=1}^{s} \lambda_k p'_{\gamma_T}(\lambda_k) - 1\right]^{-1}$$

until it converges. What distinguishes the estimation of a 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 $\{\lambda_k^{(m)}\}$:

$$\frac{\partial Q(\boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\lambda}^{(m)})}{\partial \boldsymbol{\theta}} = 0 \quad \text{and} \quad \frac{\partial Q(\boldsymbol{\alpha}, \boldsymbol{\theta}, \boldsymbol{\lambda}^{(m)})}{\partial \boldsymbol{\alpha}} = 0.$$

The Newton–Raphson (iterative) method is used here to estimate the marginal parameters and dependence parameters because no close-form is available for the estimate of $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$.

APPENDIX B: PROOFS OF THEOREMS

In this appendix, we present the brief proofs of Theorems 1 and 2.

Proof of Theorem 1. Let $d_T = T^{-1/2} + b_T$. As defined in Assumption A2, $B_\varepsilon(\boldsymbol{\phi}_0)$ is an open ball of radius ε centered at $\boldsymbol{\phi}_0$, for any fixed point $\boldsymbol{\phi}_0 \in \Gamma_0$. Let $Q_\varepsilon \equiv \cup_{\phi_0 \in \Gamma_0} B_{\varepsilon(\phi_0)}(\boldsymbol{\phi}_0)$. For any fixed $\boldsymbol{\phi}$ on the boundary of $Q_\varepsilon(\boldsymbol{\phi}_0)$, we define $\boldsymbol{\phi}_0(\boldsymbol{\phi})$ such that $|\boldsymbol{\phi} - \boldsymbol{\phi}_0(\boldsymbol{\phi})| \le |\boldsymbol{\phi} - \boldsymbol{\phi}_0|$ for all $\boldsymbol{\phi}_0 \in \Gamma_0$. That is, $\boldsymbol{\phi}_0(\boldsymbol{\phi})$ is the point in Γ_0 closest to $\boldsymbol{\phi}$. We want to show that for any given $\varepsilon > 0$, there exists a large constant M such that

$$\lim_{n\to\infty} P\left\{\sup_{\|\mathbf{u}\|=M} Q(\boldsymbol{\phi}_0(\boldsymbol{\phi}) + d_T \mathbf{u}) < Q(\boldsymbol{\phi}_0)\right\} \ge 1 - \varepsilon. \quad \text{(B.1)}$$

Thus, with large probability, there is a local maximum in the ball $\{\boldsymbol{\phi}_0 + d_T \mathbf{u} : \|\mathbf{u}\| \le M\}$. This local maximizer satisfies $\|\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0\| = O_p(d_T)$.

Using $p_{\gamma_T}(0) = 0$ and $\sum_{k=1}^{s}(\lambda_k - \lambda_{0k}) = 0$, we have

$$Q(\boldsymbol{\phi}_0(\boldsymbol{\phi}) + d_T\mathbf{u}) - Q(\boldsymbol{\phi}_0) \le \left[L(\boldsymbol{\phi}_0(\boldsymbol{\phi}) + d_T\mathbf{u}) - L(\boldsymbol{\phi}_0)\right] - T\sum_{k=1}^{r}\left[p_{\gamma_T}(\lambda_k) - p_{\gamma_T}(\lambda_{0k})\right]. \quad \text{(B.2)}$$

Let $L^{'}(\boldsymbol{\phi})$ be the gradient vector of L. Applying Taylor's expansion to $L(\boldsymbol{\phi})$ at point $\boldsymbol{\phi}_0$, we have

$$L(\boldsymbol{\phi}_0(\boldsymbol{\phi}) + d_T\mathbf{u}) - L(\boldsymbol{\phi}_0) = d_T L^{'}(\boldsymbol{\phi}_0(\boldsymbol{\phi}))^\top\mathbf{u} - \frac{1}{2}Td_T^2\mathbf{u}^\top\mathbf{I}(\boldsymbol{\phi}_0(\boldsymbol{\phi}))\mathbf{u}(1 + o_p(1)),$$

and

$$\left|\sum_{k=1}^{r}\left[p_{\gamma_T}(\lambda_k) - p_{\gamma_T}(\lambda_{0k})\right]\right| \le T\sqrt{r}b_T d_T\|\mathbf{u}\| + \frac{1}{2}Td_T^2\max_{1\le k\le r} \times \{|p''_{\gamma_T}(\lambda_{0k})|\}\{1 + o(1)\},$$

where r is the number of the components of $\boldsymbol{\phi}_{10}$. Regularity conditions imply that $T^{-1/2}L^{'}(\boldsymbol{\phi}_0(\boldsymbol{\phi})) = O_p(1)$ and $\mathbf{I}(\boldsymbol{\phi}_0(\boldsymbol{\phi}))$ is positive definite by Assumption A3. In addition, by assumption that $\max\{|p''_{\gamma_T}(\lambda_{0k})|\}$ goes to zero, the order comparison of the terms in Equation (B.2) implies that

$$-\frac{1}{2}Td_T^2\mathbf{u}^\top\mathbf{I}(\boldsymbol{\phi}_0(\boldsymbol{\phi}))\mathbf{u}(1 + o_p(1))$$

is the sole leading term in the right side of Equation (B.2). Therefore, for any given $\varepsilon > 0$, by choosing a sufficiently large M, (B.1) holds. This completes the proof of the theorem.

Lemma B.1 Assume the conditions in Theorem 2 hold. If $\liminf_{T\to\infty}\liminf_{\lambda_k\to 0+} p'_{\gamma_T}(\lambda_k)/\gamma_T > 0$ and $\sqrt{T}\gamma_T \to \infty$, for any given $\boldsymbol{\phi}$ such that $\|\boldsymbol{\phi} - \boldsymbol{\phi}_0\| = O_p(T^{-1/2})$, then, with probability tending to 1,

$$Q(\boldsymbol{\phi}_1, 0) \ge Q(\boldsymbol{\phi}_1,\ \lambda_2).$$

Proof. It suffices to show that as $T \to \infty$, for any $\boldsymbol{\phi}_2$ satisfying $\|\boldsymbol{\phi}_2 - \boldsymbol{\phi}_{20}\| = O_p(T^{-1/2})$ and small $\varepsilon_T = B_2T^{-1/2}$ with constant $B_2 > 0$, $\partial Q(\boldsymbol{\phi})/\partial\lambda_j < 0$ for $0 < \lambda_j < \varepsilon_T$, $j = r+1, \ldots, s$. By Taylor expansion, we have

$$\frac{\partial Q(\boldsymbol{\phi})}{\partial\lambda_j} = \frac{\partial L(\boldsymbol{\phi}_0)}{\partial\lambda_j} + \sum_{l=1}^{d_\gamma}\frac{\partial^2 L(\boldsymbol{\phi}_0)}{\partial\lambda_j\partial\phi_l}(\phi_l - \phi_{l0}) + \sum_{l=1}^{d_\gamma}\sum_{k=1}^{d_\gamma}\frac{\partial^3 L(\boldsymbol{\phi}^*)}{\partial\lambda_j\partial\phi_l\partial\phi_k}(\phi_l - \phi_{l0})(\phi_k - \phi_{k0}) - Tp'_{\gamma_T}(\lambda_j) - \delta. \quad \text{(B.3)}$$

Note that by the standard argument,

$$\frac{1}{T}\frac{\partial L(\boldsymbol{\phi}_0)}{\partial\lambda_j} = O_p(T^{-1/2}) \quad \text{and} \quad \frac{1}{T}\frac{\partial^2 L(\boldsymbol{\phi}_0)}{\partial\lambda_j\partial\phi_l} = E\left[\frac{\partial^2 L(\boldsymbol{\phi}_0)}{\partial\lambda_j\partial\phi_l}\right] + o_p(1).$$

By the assumption $\|\boldsymbol{\phi} - \boldsymbol{\phi}_0\| = O_p(T^{-1/2})$, we have

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

When $\liminf_{T\to\infty}\liminf_{\lambda_k\to 0+} p'_{\gamma_T}(\lambda_k)/\gamma_T > 0$ and $\sqrt{T}\gamma_T \to \infty$, we can show that $O_p(T^{1/2})\lambda_k < Tp_{\gamma_T}(\lambda_k)$ in probability in a shrinking neighborhood of 0. This completes the proof. □

In what follows, denote $\mathbf{b} = (0, \ldots, 0, p'_{\gamma_T}(\lambda_{01}), \ldots, p'_{\gamma_T}(\lambda_{0s}), 0, \ldots, 0)^\top_{(d_\alpha+s+\sum_{i=1}^r d_i)\times 1}$ and $\boldsymbol{\Sigma} = \text{diag}\{0, \ldots, 0, p''_{\gamma_T}(\lambda_{01}), \ldots, p''_{\gamma_T}(\lambda_{0s}), 0, \ldots, 0\}_{(d_\alpha+s+\sum_{i=1}^r d_i)\times(d_\alpha+s+\sum_{i=1}^r d_i)}$. Let $\boldsymbol{I}_0 = \boldsymbol{I}(\boldsymbol{\phi}_0) + \boldsymbol{\Sigma}$ and $Z_T = T^{-}\partial L(\boldsymbol{\phi}_0)/\partial\boldsymbol{\phi}$.

Lemma B.2. Under the regularity conditions in Theorem 2, if $b_T = O_p(T^{-\frac{1}{2}})$,

$$\frac{2}{T}(Q(\boldsymbol{\phi}) - Q(\boldsymbol{\phi}_0)) = g_T(\boldsymbol{\phi}) + (Z_T - \mathbf{b})^\top\boldsymbol{I}_0^{-1}(Z_T - \mathbf{b}) + O_p \times \left(\|\boldsymbol{\phi} - \boldsymbol{\phi}_0\|^3\right), \quad \text{(B.4)}$$

where $g_T(\boldsymbol{\phi}) = -[\boldsymbol{I}_0^{-1}(Z_T - \mathbf{b}) - (\boldsymbol{\phi} - \boldsymbol{\phi}_0)]^\top\boldsymbol{I}_0[\boldsymbol{I}_0^{-1}(Z_T - \mathbf{b}) - (\boldsymbol{\phi} - \boldsymbol{\phi}_0)]$. Moreover, if Γ is convex in a neighborhood of $\boldsymbol{\phi}_0$, $\|\hat{\boldsymbol{\phi}} - \widetilde{\boldsymbol{\phi}}\| = o_p(T^{-1/2})$, where $\widetilde{\boldsymbol{\phi}}$ is the maximized value of quadratic form $g_T(\boldsymbol{\phi})$ over Γ, that is, $\widetilde{\boldsymbol{\phi}} = \max_\Gamma g_T(\boldsymbol{\phi} - \boldsymbol{\phi}_0)$.

Proof. It is easy to verify that

$$Q(\boldsymbol{\phi}) - Q(\boldsymbol{\phi}_0) = Tg_T(\boldsymbol{\phi}) + (Z_T - b)^\top\boldsymbol{I}_0^{-1}(Z_T - \mathbf{b}) + O_p(\|\boldsymbol{\phi} - \boldsymbol{\phi}_0\|^3).$$

To show $\sqrt{T}\,\|\hat{\boldsymbol{\phi}} - \widetilde{\boldsymbol{\phi}}\| = o_p(1)$, it is equivalent to showing that $|g_T(\hat{\boldsymbol{\phi}}) - g_T(\widetilde{\boldsymbol{\phi}})| = o_p(T^{-1})$ because $g_T(\boldsymbol{\phi})$ is a quadratic function. Let $R_T(\boldsymbol{\phi})$ denote the last term in $(B.4)$. It is clear that $0 \le \frac{1}{T}\left(Q(\hat{\boldsymbol{\phi}}) - Q(\widetilde{\boldsymbol{\phi}})\right) = g_T(\hat{\boldsymbol{\phi}}) - g_T(\widetilde{\boldsymbol{\phi}}) + R_T(\hat{\boldsymbol{\phi}}) - R_T(\widetilde{\boldsymbol{\phi}})$. It is clear $g_T(\hat{\boldsymbol{\phi}}) - g_T(\widetilde{\boldsymbol{\phi}})$ is negative because $\widetilde{\boldsymbol{\phi}}$ is the maximum of $g_T(\boldsymbol{\phi})$. Then,

$$|g_T(\hat{\boldsymbol{\phi}}) - g_T(\widetilde{\boldsymbol{\phi}})| \le |R_T(\hat{\boldsymbol{\phi}}) - R_T(\widetilde{\boldsymbol{\phi}})|.$$

We know $\|\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0\| = O_p(T^{-1/2})$ and it can be shown that $\|\widetilde{\boldsymbol{\phi}} - \boldsymbol{\phi}_0\| = O_p(T^{-1/2})$ by using the same arguments used in the proof of Theorem 1. Thus, $|R_T(\hat{\boldsymbol{\phi}}) - R_T(\widetilde{\boldsymbol{\phi}})| = O_p(T^{-\frac{3}{2}})$. Hence, $|g_T(\hat{\boldsymbol{\phi}}) - g_T(\widetilde{\boldsymbol{\phi}})| = o_p(T^{-1})$. This completes the proof. □

Lemma B.3. Let $\widetilde{\widetilde{\boldsymbol{\phi}}} = \max_{C_\Gamma} g_T(\boldsymbol{\phi} - \boldsymbol{\phi}_0)$, where C_Γ is a cone. Then, $\|\widetilde{\boldsymbol{\phi}} - \widetilde{\widetilde{\boldsymbol{\phi}}}\| = o_p(T^{-1/2})$ if the assumptions in Theorem 2 are satisfied.

Proof. This lemma can be justified by the square root-T consistency of $\widetilde{\boldsymbol{\phi}}$ and the definition of cone. Define $q_T(W_T, \boldsymbol{\phi}) = [W_T - (\boldsymbol{\phi} - \boldsymbol{\phi}_0)]^\top(\boldsymbol{I}(\boldsymbol{\phi}_0) + \boldsymbol{\Sigma})[W_T - (\boldsymbol{\phi} - \boldsymbol{\phi}_0)]$, where $W_T = (\boldsymbol{I}(\boldsymbol{\phi}_0) + \boldsymbol{\Sigma})^{-1}(Z_T - \mathbf{b})$. According to the definitions of $\widetilde{\widetilde{\boldsymbol{\phi}}}$ and $\widetilde{\boldsymbol{\phi}}$, we have

$$\widetilde{\widetilde{\boldsymbol{\phi}}} = \inf_{\boldsymbol{\phi}\in C_\Gamma} q_T(W_T, \boldsymbol{\phi}) \quad \text{and} \quad \widetilde{\boldsymbol{\phi}} = \inf_{\boldsymbol{\phi}\in\Gamma} q_T(W_T, \boldsymbol{\phi}).$$

Let $\boldsymbol{\phi}^* \in \Gamma^*$ be such that $\inf_{\boldsymbol{\phi}\in\Gamma} q_T(W_T, \boldsymbol{\phi}) = \inf_{\boldsymbol{\phi}^*\in\Gamma^*} q_T(W_T, \boldsymbol{\phi}^*) + o_p(T^{-1/2})$. Thus, by the triangle inequality and the definition of cone, we have

$$\begin{aligned}\|\widetilde{\widetilde{\boldsymbol{\phi}}} - \widetilde{\boldsymbol{\phi}}\| &= \inf_{\boldsymbol{\phi}\in C_\Gamma} q_T(W_T, \boldsymbol{\phi}) - \inf_{\boldsymbol{\phi}\in\Gamma} q_T(W_T, \boldsymbol{\phi})\\ &\le \inf_{\boldsymbol{\phi}^*\in\Gamma^*} q_T(W_T, \boldsymbol{\phi}^*) + \inf_{\boldsymbol{\phi}\in C_\Gamma} q_T(\boldsymbol{\phi}^*, \boldsymbol{\phi}) - \inf_{\boldsymbol{\phi}\in\Gamma} q_T(W_T, \boldsymbol{\phi})\\ &\le o_p(\|\boldsymbol{\phi}^* - \widetilde{\boldsymbol{\phi}}\|) + o_p(\|\widetilde{\boldsymbol{\phi}} - \boldsymbol{\phi}_0\|) + o_P(T^{-1/2}) = o_p(T^{-1/2}).\end{aligned}$$

Similarly, let $\boldsymbol{\phi}^{**} \in C_\Gamma^*$ be such that $\inf_{\boldsymbol{\phi}\in C_\Gamma} q_T(W_T, \boldsymbol{\phi}) = \inf_{\boldsymbol{\phi}^{**}\in C_\Gamma^*} q_T(W_T, \boldsymbol{\phi}^{**}) + o_P(T^{-1/2})$. We have

$$\begin{aligned}\|\widetilde{\widetilde{\boldsymbol{\phi}}} - \widetilde{\boldsymbol{\phi}}\| &= \inf_{\boldsymbol{\phi}\in C_\Gamma} q_T(W_T, \boldsymbol{\phi}) - \inf_{\boldsymbol{\phi}\in\Gamma} q_T(W_T, \boldsymbol{\phi})\\ &\ge \inf_{\boldsymbol{\phi}\in C_\Gamma} q_T(W_T, \boldsymbol{\phi}) - \inf_{\boldsymbol{\phi}^{**}\in C_\Gamma^*} q_T(W_T, \boldsymbol{\phi}^{**}) + \inf_{\boldsymbol{\phi}\in\Gamma} q_T(\boldsymbol{\phi}^{**}, \boldsymbol{\phi})\\ &\ge o_p(\|\boldsymbol{\phi}^{**} - \widetilde{\widetilde{\boldsymbol{\phi}}}\|) + o_p(\|\widetilde{\widetilde{\boldsymbol{\phi}}} - \boldsymbol{\phi}_0\|) + o_P(T^{-1/2}) = o_p(T^{-1/2}).\end{aligned}$$

This completes the proof. □

Proof of Theorem 2. In the following proof, we consider the partition $(\boldsymbol{\phi}_1^\top, \boldsymbol{\lambda}_2^\top)^\top$. By assuming that $(\widehat{\boldsymbol{\phi}}_1^\top, 0^\top)^\top$ is the local maximizer of the penalized log-likelihood function $Q\{(\boldsymbol{\phi}_1, 0)\}$, it suffices to show that as $n \to \infty$,

$$P\left\{Q\{(\boldsymbol{\phi}_1,\ \boldsymbol{\lambda}_2)\} < Q\{(\widehat{\boldsymbol{\phi}}_1,\ 0)\}\right\} \ \to\ 1.$$

First, by the assumption above, we have the following expressions:

$$Q\left\{(\boldsymbol{\phi}_1,\ \pi_2)\right\} - Q\left\{(\widehat{\boldsymbol{\phi}}_1,\ 0)\right\} = Q\left\{(\boldsymbol{\phi}_1,\ \pi_2)\right\} - Q\left\{(\boldsymbol{\phi}_1,\ 0)\right\} + Q\left\{(\boldsymbol{\phi}_1,\ 0)\right\} - Q\left\{(\widehat{\boldsymbol{\phi}}_1,\ 0)\right\} \le Q\left\{(\boldsymbol{\phi}_1,\ \pi_2)\right\} - Q\left\{(\boldsymbol{\phi}_1,\ 0)\right\}.$$

Note that by Lemma 1, the last expression is negative with probability tending to one as T 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 $\sqrt{T}$-consistent estimator, say $\widehat{\boldsymbol{\phi}}_r$, which is the local maximizer of $Q\{\boldsymbol{\phi}_r\}$. Let $Z_{T1} = \frac{1}{T}\frac{\partial L(\boldsymbol{\phi}_{r0})}{\partial \boldsymbol{\phi}_r}$ and $\boldsymbol{I}_r = \boldsymbol{I}_1(\boldsymbol{\phi}_{r0}) + \boldsymbol{\Sigma}_1$, in which $\boldsymbol{I}_1(\boldsymbol{\phi}_r)$ is the Fisher information when all zero effects are removed. Note that

$$\frac{1}{T}\left[\ Q\left\{(\boldsymbol{\phi}_r,\ 0)\right\} - Q\left\{(\boldsymbol{\phi}_{r0},\ 0)\right\}\ \right] = g_T(\boldsymbol{\phi}_r) + R_T(\boldsymbol{\phi}_r) - (Z_{T1} - \mathbf{b}_1)^\top \times \boldsymbol{I}_r^{-1}(Z_{T1} - \mathbf{b}_1),$$

where $g_T(\boldsymbol{\phi}_r) = -[\boldsymbol{I}_r^{-1}(Z_{T1} - \mathbf{b}_1) - (\boldsymbol{\phi}_r - \boldsymbol{\phi}_{r0})]^\top \boldsymbol{I}_r[\boldsymbol{I}_r^{-1}(Z_{T_1} - \mathbf{b}) - (\boldsymbol{\phi}_r - \boldsymbol{\phi}_{r0})]$. Let $\widetilde{\boldsymbol{\phi}}_r = \max_{\Gamma_r} g_T(\boldsymbol{\phi}_r - \boldsymbol{\phi}_{r0})$. It follows from Lemmas 2 and 3 that

$$\left\|\widehat{\boldsymbol{\phi}}_r - \widetilde{\boldsymbol{\phi}}_r\right\| = o_p(T^{-1/2}), \quad \text{and} \quad \left\|\widetilde{\boldsymbol{\phi}}_r - \widetilde{\widetilde{\boldsymbol{\phi}}}_r\right\| = o_p(T^{-1/2}),$$

where $\widetilde{\widetilde{\boldsymbol{\phi}}}_r = \max_{C_{\Gamma_r}} g_T(\boldsymbol{\phi}_r - \boldsymbol{\phi}_{r0})$. Then, combining above two equations, we have

$$\sqrt{T}\left\|\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}_{r0}\right\| \le \sqrt{T}\left(\left\|\widehat{\boldsymbol{\phi}}_r - \widetilde{\boldsymbol{\phi}}_r\right\| + \left\|\widetilde{\boldsymbol{\phi}}_r - \widetilde{\widetilde{\boldsymbol{\phi}}}_r\right\| + \left\|\widetilde{\widetilde{\boldsymbol{\phi}}}_r - \boldsymbol{\phi}_{r0}\right\|\right) = \sqrt{T}\left\|\widetilde{\widetilde{\boldsymbol{\phi}}}_r - \boldsymbol{\phi}_{r0}\right\| + o_p(1).$$

To continue our proof, two different cases have to be considered.

Case I. $\boldsymbol{\phi}_{r0}$ is an interior point of the subset of Γ_r. That is, $C_{\Gamma_r} = \mathbb{R}^q$. Based on the definition of $\widetilde{\widetilde{\boldsymbol{\phi}}}_r$, we have $\widetilde{\widetilde{\boldsymbol{\phi}}}_r - \boldsymbol{\phi}_{r0} = (\boldsymbol{I}_1(\boldsymbol{\phi}_{r0}) + \boldsymbol{\Sigma}_1)^{-1}(Z_{T1} - \mathbf{b}_1)$. Therefore, $\sqrt{T}[\widehat{\boldsymbol{\phi}}_r - \boldsymbol{\phi}_{r0} + (\boldsymbol{I}_1(\boldsymbol{\phi}_{r0}) + \boldsymbol{\Sigma}_1)^{-1}\mathbf{b}_1] = \sqrt{T}(\boldsymbol{I}_1(\boldsymbol{\phi}_{r0}) + \boldsymbol{\Sigma}_1)^{-1}Z_{T1} + o_p(1)$. By the central limit theorem and Slutsky's theorem, we have

$$\sqrt{T}\left[\boldsymbol{I}_1(\boldsymbol{\phi}_{r0}) + \boldsymbol{\Sigma}_1\right](\widetilde{\widetilde{\boldsymbol{\phi}}}_r - \boldsymbol{\phi}_{r0}) \to N(0,\ \boldsymbol{I}_1).$$

Case II. $\boldsymbol{\phi}_{r0}$ is on the boundary of the subset of Γ_r. In other words, $\boldsymbol{\phi}_{r0} = (\lambda_{10}, \theta_{10}^\top, \alpha_0^\top)^\top = (1, \theta_{10}^\top, \alpha_0^\top)^\top$. Therefore, $C_{\Gamma_r} = \mathbb{R}^q = [0, \infty) \times \mathbb{R}^{d_1 + d_\alpha}$, where $q = 1 + d_1 + d_\alpha$. Let $g(\boldsymbol{\phi}_r) = -\left[\boldsymbol{I}_r^{-1}(Z_1 - \mathbf{b}_1) - (\boldsymbol{\phi}_r - \boldsymbol{\phi}_{r0})\right]^\top \boldsymbol{I}_r\left[\boldsymbol{I}_r^{-1}(Z_1 - \mathbf{b}) - (\boldsymbol{\phi}_r - \boldsymbol{\phi}_{r0})\right]$. Since $\boldsymbol{I}_r$ is positive definite, by maximizing the quadratic form $g(\boldsymbol{\phi}_r)$ over C_{Γ_r}, the continuous mapping theorem gives that $\widetilde{\widetilde{\boldsymbol{\phi}}}_r \to \check{\boldsymbol{\phi}}_r$, where the limiting random variable $\check{\boldsymbol{\phi}}_r$ is the maximizer of $g(\boldsymbol{\phi}_r)$ and has the following representation

$$\begin{pmatrix} Z_{11} - c_1 \\ Z_{12} - c_2 \\ \vdots \\ Z_{1q} - c_q \end{pmatrix} I\{Z_{11} > c_1\} + \begin{pmatrix} 0 \\ Z_{12} - c_2 - (\boldsymbol{I}_1^{21}/\boldsymbol{I}_1^{11})(Z_{11} - c_1) \\ \vdots \\ Z_{1q} - c_q - (\boldsymbol{I}_1^{q1}/\boldsymbol{I}_1^{11})(Z_{11} - c_1) \end{pmatrix} I\{Z_{11} < c_1\},$$

where $Z_1 = (Z_{11}, Z_{12}, \ldots, Z_{1q})^\top$ is a random variable with multivariate Gaussian distribution with mean 0 and covariance matrix $\boldsymbol{I}_1(\boldsymbol{\phi}_{r0})$, $\boldsymbol{\phi}_r \in C_{\Gamma_r} - \boldsymbol{\phi}_{r0}$, $\boldsymbol{I}_1^{ij} = \boldsymbol{I}_1^{ij}(\boldsymbol{\phi}_{r0})$ are elements of matrix $\boldsymbol{I}_1^{-1}(\boldsymbol{\phi}_{r0})$ and $\boldsymbol{I}_1(\boldsymbol{\phi}_r)$ is the Fisher information when all zero effects are removed. Therefore, $\widetilde{\boldsymbol{\phi}}_r \to \check{\boldsymbol{\phi}}_r$. This completes the proof. □

[Received April 2012. Revised November 2013.]

REFERENCES

Anderson, T. W., and Darling, D. A. (1952), "Asymptotic Theory of Certain Goodness-of-Fit Criteria Based on Stochastic Processes," *Annals of Mathematical Statistics*, 23, 193–212. [788]

Andrews, D. (1999), "Estimation When a Parameter is on a Boundary," *Econometrica*, 67, 1341–1383. [789]

Ang, A., and Chen, J. (2002), "Asymmetric Correlations of Equity Portfolios," *Review of Financial Studies*, 63, 443–494. [788]

Bouyé, E., Durrleman, V., Nikeghbali, A., Riboulet, G., and Roncalli, T. (2001), "Copulas: An Open Field for Risk Management," Discussion Paper WP01-01, Paris, France: Groupe de Recherche Opérationnelle, Crédit-Lyonnais. [788]

Cai, Z., Fan, J., and Yao, Q. (2000), "Functional-Coefficient Regression Models for Nonlinear Time Series," *Journal of the American Statistical Association*, 95, 941–956. [798]

Cai, Z., Peng, Z., and Ren, Y. (2011), "Time-Varying Portfolio Choice and Predictor Selection," Working Paper 2011-04, Xiamen University, China. [798]

Cai, Z., and Wang, X. (2008), "Selection of Finite Mixture Distribution," Working Paper 08-03, University of North Carolina at Charlotte. [793,794]

——— (2013). "Semiparametric Copula Selection via Penalized Likelihood," Working Paper 13-01, University of North Carolina at Charlotte. [791]

Chen, X., Fan, Y., and Patton, A. (2003), "Simple Tests for Models of Dependence Between Financial Time Series: With Applications to US Equity Returns and Exchange Rates," Mimeo, University of California at San Diego. [788]

Chen, J., and Khalili, A. (2008), "Order Selection in Finite Mixture Models," *Journal of the American Statistical Association*, 103, 1674–1683. [789]

Cherubini, U., Vecchiato, W., and Luciano, E. (2004), *Copula Methods in Finance*, New York: Wiley. [788,790]

Chollete, L., Pena, V., and Lu, C. (2005), "International Diversication: A Copula Approach," *Journal of Banking and Finance*, 35, 403–417. [788]

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood From Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society,* Series B, 39, 1–38. [798]

Denuit, M., and Scaillet, O. (2004), "Nonparametric Tests for Positive Quadrant Dependence," *Journal of Financial Econometrics*, 2, 422–450. [788]

Donoho, D. L., and Johnstone, I. M. (1994), "Ideal Spatial Adaptation by Wavelet Shrinkage," *Biometrika*, 81, 425–455. [794]

Embrechts, P., Lindskog, F., and McNell, A. J. (2003), "Modeling Dependence With Copulas and Applications to Risk Management," in *Handbook of Heavy Tailed Distributions in Finance*, ed., S. Rachev, Amsterdam: Elsevier Science, Chapter 8, pp. 329–384. [788]

Fan, J. (1997), "Comments on 'Wavelets in Statistics: A Review' by A. Antoniadis," *Journal of the Italian Statistical Association*, 6, 131–138. [792]

Fan, J., and Li, R. (2001), "Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties," *Journal of the American Statistical Association*, 96, 1348–1360. [789,792,798]

Fan, J., and Lv, J. (2010), "A Selective Overview of Variable Selection in High Dimensional Feature Space," *Statistica Sinica*, 20, 101–148. [789,792]

Fan, J., and Peng, H. (2004), "Nonconcave Penalized Likelihood With Diverging Number of Parameters," *The Annals of Statistics*, 32, 928–961. [789]

Fermanian, J. D. (2005), "Goodness of Fit Tests for Copulas," *Journal of Multivariate Analysis*, 95, 119–152. [788]

Fermanian, J. D., and Scaillet, O. (2005), "Some Statistical Pitfalls in Copula Modeling for Financial Applications," in *Capital Formation, Governance and Banking*, ed., E. Klein, New York: Nova Science Publishers, pp. 57–72. [791]

Frey, R., and McNeil, A. J. (2003), "Dependent Defaults in Models of Portfolio Credit Risk," *Journal of Risk*, 6, 59–92. [788]

Genest, C., Rémillard, B., and Beaudoin, D. (2009), "Goodness-of-Fit Tests for Copulas: A Review and a Power Study," *Insurance: Mathematics and Economics*, 44, 199–123. [789]

Hamerle, A., and Rösch, D. (2005), "Misspecified Copulas in Credit Risk Models: How Good is Gaussian," *Journal of Risk*, 8, 41–58. [788]

Hu, L. (2006), "Dependence Patterns Across Financial Markets: A Mixed Copula Approach," *Applied Financial Economics*, 16, 717–729. [788,797]

Joe, H. (1997), *Multivariate Models and Dependence Concepts*, London: Chapman & Hall. [789,792]

Jondeau, E., and Rockinger, M. (2006), "The Copula-GARCH Model of Conditional Dependencies: An International Stock Market Application," *Journal of International Money and Finance*, 25, 827–853. [788]

Junker, M., Szimayer, A., and Wagner, N. (2006), "Nonlinear Term Structure Dependence: Copula Functions, Empirics and Risk Implications," *Journal of Banking & Finance*, 30, 1171–1199. [788]

Kendall, M. G. (1938), "A New Measure of Rank Correlation," *Biometrika*, 30, 81–93. [790]

Kole, E., Koedijk, K., and Verbeek, M. (2007), "Selecting Copulas for Risk Management," *Journal of Banking & Finance*, 31, 2405–2423. [788]

Lee, T.-H., and Long, X. (2009), "Copula-Based Multivariate GARCH Model With Uncorrelated Dependent Errors," *Journal of Econometrics*, 150, 207–218. [788,798]

Li, D. (2000), "On Default Correlation: A Copula Function Approach," *Journal of Fixed Income*, 9, 43–54. [788]

Longin, F., and Solnik, B. (2001), "Extreme Correlation of International Equity Markets," *Journal of Finance*, 56, 649–676. [788]

Mashal, R., Naldi, M., and Zeevi, A. (2003), "Comparing the Dependence Structure of Equity and Asset Returns," *Risk*, 16, 83–87. [788]

Nelsen, R. B. (1998), *An Introduction to Copulas*, New York: Springer-Verlag. [789]

Newey, W. K., and West, K. D. (1987), "A Simple, Positive-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708. [794]

Scaillet, O. (2007), "Kernel Based Goodness-of-Fit Tests for Copulas With Fixed Smoothing Parameters," *Journal of Multivariate Analysis*, 98, 533–543. [788]

Sklar, A. (1959), "Fonctions de Répartition *à* n Dimensions et Leures Marges," *Publications de L'Institut de Statistique de L'Université de Paris*, 8, 229–231. [791]

Spearman, C. (1904), "The Proof and Measurement of Association Between Two Things," *The American Journal of Psychology*, 15, 72–101. [790]

Tibshirani, R. J. (1996), "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society,* Series B, 58, 267–288. [789]