Trending Mixture Copula Models with Copula Selection*

Bingduo Yang[a], Zongwu Cai[b], Christian M. Hafner[c] and Guannan Liu[c]

[a]Lingnan (University) College, Sun Yat-sen University, Guangzhou, Guangdong 510275, China

[b]Department of Economics, University of Kansas, Lawrence, KS 66045, USA.

[c]Institut de statistique and CORE, Université catholique de Louvain, Voie du Roman Pays, 20, 1348 Louvain-la-Neuve, Belgium.

[d]Wang Yanan Institute for Studies in Economics and School of Economics, Xiamen University, Xiamen, Fujian 361005, China

Abstract: Modeling the joint tails of multiple financial time series has important implications for risk management. Classical models for dependence often encounter a lack of fit in the joint tails, calling for additional flexibility. In this paper we introduce a new nonparametric time-varying mixture copula model, in which both weights and dependence parameters are deterministic functions of time. We propose penalized trending mixture copula models with group smoothly clipped absolute deviation (SCAD) penalty functions to do the estimation and copula selection simultaneously. Monte Carlo simulation results suggest that the shrinkage estimation procedure performs well in selecting and estimating both constant and trending mixture copula models. Using the proposed model and method, we analyze the evolution of the dependence among four international stock markets, and find substantial changes in the levels and patterns of the dependence, in particular around crisis periods.

Keywords: Copula, Time-Varying Copula, Mixture Copula, Copula Selection

*Yang's research was partly supported by NSF of Jiangxi Province (No. 20171BAA208002) and the science and technology project of Jiangxi Province Education Department (No. GJJ170329). Cai's research is partially supported by the National Natural Science Foundation of China grant #71631004 (Key Project).
The corresponding author: Guannan Liu, email: gliuecon@gmail.com.

1 Introduction

Copulas have received considerable attention recently because they offer great flexibility to model multivariate distributions and to characterize nonlinear dependence and tail dependency. A copula function glues various types of marginal distributions, including symmetric, skewed and heavy-tailed distributions, into a multivariate distribution, and by Sklar's theorem (1959) this is always possible. Copula functions are capable of capturing different types of dependence: linear or nonlinear, symmetric or asymmetric, tail dependence or no-tail dependence. These features are of great importance in financial or macroeconomic time series, which lead to many applications, such as volatility clustering (Ning, Xu and Wirjanto, 2015), real-time density forecasting (Smith and Vahey, 2016), non-stationarity (Wollschläger and Schäfer, 2016), and systemic risk (Mensi, Hammoudeh, Shahzad and Shahbaz, 2016), among others.

In the literature, time-varying copulas have been extensively used to model multiple financial time series. For example, Patton (2006) uses a symmetrized Joe-Clayton copula in which the dependence structure follows an autoregressive moving average (ARMA)-type process to capture asymmetric dependence between mark-dollar and yen-dollar exchange rates. Other time-varying copulas include dynamic stochastic copula models (Hafner and Manner, 2012), stochastic copula autoregressive models (Almeida and Czado, 2012), generalized autoregressive score models (Creal, Koopman and Lucas, 2013), and variational mode decomposition methods (Mensi, Hammoudeh, Shahzad and Shahbaz, 2016), among others. For a comprehensive survey of time-varying copulas and their applications in financial time series analysis, see Manner and Reznikova (2012) and Patton (2012a).

Another line of research is using a nonparametric approach to characterize time-varying copulas. Compared to above parametric approaches, the nonparametric approach can increase the flexibility in modelling dynamic dependence among variables. For example, Hafner and Reznikova (2010) assume that the dependence parameters in copulas are deterministic functions of time, while Acar, Craiu and Yao (2011) specify the dependence parameters as functions of explanatory variables. Both papers employ kernel smoothers to estimate the dependence parameters in the context of known conditional marginal distributions. Recently, Fermanian and Lopez (2018) introduce a single-index copula whose parameter is an unknown link function of a univariate index and propose estimates of this link function and the parameters in the index part.

The third line of research is the mixture copula approach. For example, Garcia and Tsafack (2011) study the dependence structure of international equity and bond markets by a regime switching method, which models symmetric dependence in one regime and asymmetric dependence in another regime. Their approach can also be regarded as a mixture copula. Liu, Ji and Fan (2017) propose a time-varying optimal copula (TVOC) model to identify and capture the optimal dependence structure of bivariate time series at every time point. Their modelling procedures consist of two steps, i.e. time-varying modelling using rolling windows and a distribution-free test to identify the optimal copula. Recently, Liu, Long, Zhang and Li (2018) use a model averaging approach to estimate the dependence structure in a mixture copula framework by assuming homogeneity of the weights and dependence parameters.

In this paper we propose a trending mixture copula model which allows both the weights and dependence parameters to be time-varying in a purely nonparametric way. We do not specify any parametric form for the weight and dependence parameters and use a data-driven method for their specification. In this way, we alleviate typical misspecification problems in copulas. Moreover, the proposed model can be considered as an ideal copula model, as described in Patton (2012b), in the sense that it accommodates dependence of either positive or negative sign, captures both symmetric and asymmetric dependence, and allows for the possibility of non-zero tail dependence. The proposed model is different from prior studies which focus exclusively on single copula models using a nonparametric approach (Hafner and Reznikova, 2010; Acar et al., 2011). It also differs from previous mixture copula models that assume homogeneity in either the weights or dependence parameters (Garcia and Tsafack, 2011; Liu et al., 2018). Finally, it generalizes the TVOC model of Liu et al. (2017) which assumes a single copula at each time point.

An important issue is that the range of both copula dependence parameters and their corresponding weights are restricted, e.g. $\theta \in (-1, 1)$ for a Gaussian copula, $\theta \in (0, \infty)$ for a Clayton copula, and the weights satisfying $\lambda_k \in [0, 1]$ and $\sum_k \lambda_k = 1$. To overcome this difficulty in the nonparametric estimation, Abegaz et al. (2012) and Acar et al. (2011) use some known inverse functions to ensure that the copula parameters are properly defined and employ a local polynomial framework to estimate the dependence parameters. However, in the asymptotic properties either the bias or variance depends on the choice of the inverse link functions, see e.g. Theorem 2 of Abegaz et al. (2012) and Corollary 1 of Acar et al. (2011). It is nontrivial to find an optimal inverse link function in a large functional space. In this study, we employ a local constant (Nadaraya-Watson) kernel method without choosing any inverse link function and show that the local constant estimators have the same asymptotic behavior as the local linear estimators at the interior points: both have the same bias and variance terms as well as the same convergence rate.

To reduce the risk of over-fitting and efficiency loss, we propose penalized trending mixture copula models with group smoothly clipped absolute deviation (SCAD) penalty term (Cai, Juhl and Yang, 2015) to do the estimation and copula selection simultaneously. The functional norms of the weight functions are penalized so that we can shrink them to zeros if contributions of corresponding copulas are small. To facilitate the estimation, we propose a nonparametric version of the expectation maximization (EM) algorithm to estimate the weights and dependence parameters in the penalized local copula log-likelihood function. Other important practical issues including goodness-of-fit, the bandwidth and tuning parameter selection, and confidence intervals are also discussed. When doing simulations, we consider mixture copulas with both constant and time-varying weights and dependence parameters. Simulation results show that the proposed method can correctly select the appropriate copulas and accurately estimate the unknown parameters in both cases.

In the empirical section, we employ the proposed model and method to investigate the evolution of the dependence structures among four international stock markets (the United States, the United Kingdom, Hong Kong and South Korea), using 28 years of data on weekly returns from the four economies' equity indexes. We find that all pairs of markets show lower tail dependence but no upper tail dependence, since the Clayton and Frank copulas are always selected while the Gumbel is always filtered out. We also observe that the dependence exhibits quite different levels and patterns during different periods, e.g., in tranquil periods and in crisis periods.

The remainder of paper is organized as follows. Section 2 introduces the proposed trending mixture copula models. In the same section, we introduce penalized trending mixture copula models. Four practical issues are discussed including a nonparametric EM algorithm, goodness-of-fit tests, the bandwidth and tuning parameter selection, and the construction of pointwise confidence intervals. Section 3 reports the Monte Carlo simulation results. Section 4 applies the model and method to examine the evolution of the dependence among four international stock markets. The final section provides some concluding comments, and the mathematical proofs are gathered in the Appendix.

2 Trending mixture copula models

In this section, we model the time-varying mixture copula in a purely nonparametric way so that the dynamics in both the weights and dependence parameters are simultaneously captured. To simplify the presentation we will only consider bivariate copulas although the extension to the multivariate case is straightforward.

First, we rescale time as t_i by $t_i = i/T$ to provide the asymptotic justification for nonparametric smoothing estimators. The underlying assumption is that there will be an increasingly intense sampling of data points that can be used to derive consistent estimation, see e.g. Robinson (1989, 1991). The trending mixture copula model is defined as

$$C(x_{1i}, x_{2i}, \psi, \delta(t_i)) = \sum_{k=1}^{d} \lambda_k(t_i) C_k(F_1(x_{1i}; \psi_1), F_2(x_{2i}; \psi_2), \theta_k(t_i)), \tag{1}$$

where copulas C_k are from different copula families, $\psi = (\psi_1^\intercal, \psi_2^\intercal)^\intercal$ is a parametric vector with ψ_s being the q_s-dimensional component for the marginal distribution $F_s(x_s; \psi_s)$, $s = 1, 2$. Furthermore, $\delta(t_i) = (\theta(t_i)^\intercal, \lambda(t_i)^\intercal)^\intercal$ is a vector of $(p_1 + \cdots + p_d)$-dimensional dependence parameters $\theta(t_i) = (\theta_1(t_i)^\intercal, ..., \theta_d(t_i)^\intercal)^\intercal$ and d-dimensional weights $\lambda(t_i) = (\lambda_1(t_i), ..., \lambda_d(t_i))^\intercal$. For simplicity of presentation, we set $p_1 = \cdots = p_d = 1$. The weight $\lambda_k(t_i)$ controls the contribution of the copula C_k and satisfies both $0 \le \lambda_k(t_i) \le 1$ and $\sum_{k=1}^{d} \lambda_k(t_i) = 1$ for all $t_i \in [0, 1]$. The parameter $\theta_k(t_i)$ represents the level of the dependence corresponding to the copula C_k at time t_i. Since our main focus is on the dependence structure, we assume that the mixture copula parameter vector $\delta(t_i)$ is potentially time-varying, whereas the marginal parameter vector ψ does not depend on t_i. The above mixture copula model implies that the joint cumulative distribution function of a bivariate random vector (x_{1i}, x_{2i}) is given by a linear combination of $C_k(F_1(x_{1i}, \psi_1), F_2(x_{2i}, \psi_2), \theta_k(t_i))$ with time-varying weights $\lambda_k(t_i)$.

When $d = 1$, the above model reduces to the nonparametric dynamic copula model of

Hafner and Reznikova (2010). It also reduces to the model of Acar et al. (2011) when the conditioning variable is a single exogenous variable. Furthermore, if both $\lambda(t_i)$ and $\theta(t_i)$ are constant, the above model is a standard mixture copula model (Cai and Wang, 2014; Liu et al., 2018).

To find estimators for the proposed trending mixture copula model, we take first derivatives of the distribution function with respect to x_{1i} and x_{2i} sequentially, and its density function can be written as

$$c(u_{1i}, u_{2i}, \psi, \delta) = f_1(x_{1i}; \psi_1) f_2(x_{2i}; \psi_2) \sum_{k=1}^{d} \lambda_k(t_i) c_k(u_{1i}, u_{2i}, \psi, \theta_k(t_i)),$$

where $u_{si} = F_s(x_{si}; \psi_s)$, for $s = 1, 2$, $c_k(\cdot)$ is the density function of copula $C_k(\cdot)$, and $f_s(\cdot; \psi_s)$ are the marginal densities. Then the log-likelihood function at time t_i for a sequence $\{x_{1i}, x_{2i}\}_{i=1}^T$ is obtained as

$$L(u_1, u_2, \psi, \delta(t_i)) = L_m(\psi) + L_c(\psi, \delta(t_i)),$$

where $L_m(\psi) = \sum_{i=1}^T \log f_1(x_{1i}, \psi_1) + \sum_{i=1}^T \log f_2(x_{2i}, \psi_2)$ and $L_c(\psi, \delta(t_i)) = \sum_{i=1}^T \ell_i(\psi, \delta(t_i))$ with $\ell_i(\psi, \delta(t_i)) = \log(\sum_{k=1}^d \lambda_k(t_i) c_k(u_{1i}, u_{2i}, \psi, \theta_k(t_i)))$.

In the following we propose a two-step procedure to estimate the parameters. First, we estimate the parameters of the marginal distributions, (ψ_1, ψ_2), by maximizing the marginal log-likelihood:

$$(\hat{\psi}_1, \hat{\psi}_2) = \arg\max_{\psi_1, \psi_2} \sum_{i=1}^T \log f_1(X_{1i}, \psi_1) + \sum_{i=1}^T \log f_2(X_{2i}, \psi_2).$$

This step uses the classical parametric estimation and is referred to as the inference function for margins (IFM) approach by Joe (2000). Under weak regularity conditions, the standard convergence rate of the parametric estimators is $\|\hat{\psi}_s - \psi_s\| = O_p(1/\sqrt{T})$ for $s = 1, 2$, where $\|\cdot\|$ represents the Euclidean norm (L_2-norm).

Second, given the estimators $(\hat{\psi}_1, \hat{\psi}_2)$ obtained from the first step, we calculate the nonparametric estimator $\hat{\delta}(\tau)$ at any grid point $\tau \in (0, 1)$ by maximizing the local copula log-likelihood function as:

$$\hat{\delta}(\tau) = \arg \max_{(\theta(\tau),\lambda(\tau))} \sum_{i=1}^{T} \log(\sum_{k=1}^{d} \lambda_k(\tau) c_k(\hat{u}_{1i}, \hat{u}_{2i}, \hat{\psi}, \theta_k(\tau))) K_h(t_i - \tau), \tag{2}$$

where $\hat{u}_{si} = F_s(x_{si}, \hat{\psi}_s)$, $K_h(\cdot) = K(\cdot/h)/h$ with K being a kernel function and h a bandwidth that tunes the smoothness of the kernel estimator. In our simulation and empirical study, the commonly adopted Epanechnikov kernel function $K(z) = 3/4(1-z^2)I(|z| \leq 1)$ is used, where $I(|z| \leq 1)$ takes the value 1 if $|z| \leq 1$ and 0 otherwise. In the following we make assumptions and derive the asymptotic properties of the proposed estimators under α-mixing conditions.

2.1 Asymptotic properties

The above two-step procedure is motivated by the fact that the parametric estimators $\hat{\psi}$ have little effect on the nonparametric estimators $\hat{\delta}$ in large samples since the convergence rate $\sqrt{T}$ in the parametric part of the model is faster than $\sqrt{Th}$ in the nonparametric component. The detailed proof for this argument can be found in Lemma 1 in Appendix B. In the following, we present the asymptotic properties of the nonparametric estimators without considering the errors from the marginal parametric estimation in the first step. For this purpose, we introduce regularity conditions as below.

A1. The vector of functions $\delta(\tau)$ is continuous, bounded and has second order continuous derivatives on the support $[0, 1]$.

A2. The local copula log-likelihood function $L_c(\psi, \delta(\tau))$ is three times differentiable with respect to δ and twice differentiable with respect to ψ. The marginal log-likelihood function $L_m(\psi)$ is twice differentiable with respect to ψ.

A3. $0 \leq \lambda_k(\tau) \leq 1$ and $\sum_{k=1}^{d} \lambda_k(\tau) = 1$ for all $\tau \in [0, 1]$.

A4. The kernel function $K(z)$ is twice continuously differentiable on the support $[0, 1]$, and its second derivative satisfies a Lipschitz condition. Let $v_0 = \int K^2(z)dz$, $v_2 = \int z^2 K^2(z)dz$ and $\mu_2 = \int z^2 K(z)dz$.

A5. The bandwidth h satisfies that $h \to 0$ and $Th \to \infty$, as $T \to \infty$.

A6. Assume that $\{X_{1i}, X_{2i}\}_{i=1}^{T}$ is a strictly stationary α-mixing sequence. Further, assume that there exists some constant $c > 0$, $E|X_{1i}|^{2(2+c)} < \infty$, $E|X_{2i}|^{2(2+c)} < \infty$ and the mixing coefficient $\alpha(m)$ satisfying $\alpha(m) = O(m^{-\vartheta})$ with $\vartheta = (2+c)(1+c)/c$.

Remarks: The conditions in A1 and A2 are needed for deriving the asymptotic properties of the nonparametric estimators. Moreover, by the conditions in A1, the continuity of $\delta(\tau)$ implies that $\|\hat{\delta}(t_i) - \hat{\delta}(t_{i-1})\| = O_p(1/T)$ which is of much smaller order than the nonparametric convergence rate $T^{-2/5}$. It suggests that we only need to estimate $\hat{\delta}(t_i)$ for $i = 1, \cdots, T$ rather than $\hat{\delta}(\tau)$ for all values $\tau \in (0,1)$. The conditions in A3 are mild conditions for identification, while the conditions in A4 and A5 are commonly employed in nonparametric estimation. The conditions in A6 are the common conditions with weakly dependent data. Most financial models satisfy these conditions, such as ARMA and GARCH models, see e.g. Cai (2007).

Theorem 1 *Let $\{x_{1i}, x_{2i}\}_{i=1}^{T}$ be a strictly stationary α-mixing sequence following the copula model in (1). Assume that $\|\hat{\psi} - \psi\| = O_p(1/\sqrt{T})$, $h \to 0$ and $Th \to \infty$ as $T \to \infty$. For a fixed point $\tau \in (0,1)$, under Conditions A1-A6, we have*

$$\sqrt{Th}(\hat{\delta}(\tau) - \delta(\tau) - h^2 B(\tau)) \to N(0, v_0 \Sigma(\tau)^{-1}\Omega(\tau)\Sigma(\tau)^{-1}),$$

where $\Sigma(\tau) = -E\{\ell_i''(\delta(\tau))|t_i = \tau\}$, $\Omega(\tau) = \sum_{\ell=-\infty}^{\infty} \Gamma_\ell(\tau)$ with $\Gamma_\ell(\tau) = E\{\ell_i'(\delta(\tau))\ell_{i+\ell}'(\delta(\tau))^\intercal | t_i = \tau\}$ and the bias term $h^2 B(\tau) = \frac{h^2}{2}\delta''(\tau)\mu_2$.

Remark 1 *The condition $\|\hat{\psi} - \psi\| = O_p(1/\sqrt{T})$ can be derived from the marginal log-likelihood estimation. From Theorem 1, as expected, the initial estimators $\hat{\psi}$ have little effect on the estimation of $\hat{\delta}(\cdot)$ in large samples. In classical local constant estimation, the bias term is usually written as $h^2 B(\tau) = \frac{h^2}{f(\tau)}\delta'(\tau) f'(\tau)\mu_2 + \frac{h^2}{2}\delta''(\tau)\mu_2$, where $f(\tau)$ is the density at the point τ. However, the first term on the right hand size disappears since $f(\tau) = 1$ and $f'(\tau) = 0$ for all $\tau \in (0,1)$.*

Theorem 1 suggests that the local constant estimators $\hat{\delta}(\tau)$ have the same asymptotic behavior as the local linear estimators at the interior points: both have the same bias and variance terms as well as the same convergence rate $\sqrt{Th}$.

To see whether the large sample properties of the local constant estimators still hold at the boundary, we introduce Theorem 2 as below. For this purpose, we define $v_{0,b} = \int_{-b}^{1} K^2(z)dz$,

$\mu_{0,b} = \int_{-b}^{1} K(z)dz$ and $\mu_{1,b} = \int_{-b}^{1} zK(z)dz$, for $0 < b < 1$. Without loss of generality, we only consider the left boundary point, $\tau = bh$. Similar results hold for a right boundary point $\tau = 1 - bh$.

Theorem 2 *Let $\{x_{1i}, x_{2i}\}_{i=1}^{T}$ be a strictly stationary α-mixing sequence following the copula model in (1). Assume that $\|\hat{\psi} - \psi\| = O_p(1/\sqrt{T})$, $h \to 0$ and $Th \to \infty$ as $T \to \infty$. For a left boundary point $\tau = bh$, under Conditions A1-A6, we have*

$$\sqrt{Th}(\hat{\delta}(bh) - \delta(bh) - hB^*(0+)) \to N(0, \frac{v_{0,b}}{\mu_{0,b}^2}\Sigma(0+)^{-1}\Omega(0+)\Sigma(0+)^{-1}),$$

where the bias term $hB^*(0+) = \frac{h}{\mu_{0,b}}\delta'(0+)\mu_{1,b}$.

Remark 2 *The bias term is of order h for a boundary point $\tau = bh$, which suggests that the local constant estimator suffers from boundary effects.*

2.2 Penalized trending mixture copula models

When many candidate copula families are included in the proposed trending mixture copula model, there is a risk of overfitting and efficiency loss, which motivates us to do the estimation and copula selection simultaneously. For this purpose, we define a $T \times (2d)$ matrix $\delta = (\delta(t_1), \cdots, \delta(t_T))^\intercal = (\theta_{\cdot 1}, \cdots, \theta_{\cdot d}, \lambda_{\cdot 1}, \cdots, \lambda_{\cdot d})$, where $\delta(t_j) = (\theta_1(t_j), \cdots, \theta_d(t_j), \lambda_1(t_j), \cdots, \lambda_d(t_j))^\intercal$ for $j = 1, \cdots, T$, and $\theta_{\cdot k} = (\theta_k(t_1), \cdots, \theta_k(t_T))^\intercal$ and $\lambda_{\cdot k} = (\lambda_k(t_1), \cdots, \lambda_k(t_T))^\intercal$ for $k = 1, \cdots, d$. We follow the idea of the group LASSO (Yuan and Lin, 2006) and propose the following penalized local log-likelihood function as

$$Q^P(\delta) = \sum_{j=1}^{T}\sum_{i=1}^{T} \ell_i(\delta(t_j))K_h(t_i - t_j) - T\sum_{k=1}^{d} P_{\gamma_k}(\|\lambda_{\cdot k}\|) \quad (3)$$

where $\ell_i(\delta(t_j)) \equiv \ell_i(\psi, \delta(t_j))$, $P_{\gamma_k}(\cdot)$ is a penalty function with tuning parameter γ_k and $\|\lambda_{\cdot k}\|$=$(\lambda_k^2(t_1) + \cdots + \lambda_k^2(t_T))^{1/2}$. The norm of $\lambda_{\cdot k}$, i.e. $\|\lambda_{\cdot k}\|$, is penalized so that we can shrink the weight function $\lambda_k(\cdot)$ to zero if the contribution of copula $C_k(\cdot)$ is small. We do not penalize the dependence parameters $\theta_k(\cdot)$ since our main focus is on the copula selection. Clearly, the purpose of using the penalized locally weighted log-likelihood function is to select important copula families.

Various penalty functions have been proposed over the last decades. As pointed out by Fan and Li (2001), a good penalty function should satisfy the following three properties: unbiasedness for the non-zero coefficients, sparsity, and continuity of the resulting estimators to avoid instability in model prediction. Here, we propose to use the SCAD penalty function (Fan and Li, 2001; Cai et al., 2015) that enjoys all three properties, although many other penalty functions are applicable including LASSO (Tibshirani, 1996) and adaptive LASSO (Zou, 2006). The first-order derivative $P'_{\gamma_k}(z)$ of the continuous SCAD penalty function $P_{\gamma_k}(z)$ is given by

$$P'_{\gamma_k}(z) = \gamma_k I(z \leq \gamma_k) + \frac{(\varrho\gamma_k - z)_+}{(\varrho - 1)} I(z > \gamma_k)$$

for some $\varrho > 2$, where $I(\cdot)$ is the indicator function and $(\varrho\gamma_k - z)_+ = \max(\varrho\gamma_k - z, 0)$. For simplicity of presentation, we assume that the tuning parameters γ_k are the same for all $k = 1, \cdots, d$ by taking $\gamma_k = \gamma_T$. We select $\varrho = 3.7$ from a Bayesian risk point of view as suggested by Fan and Li (2001). They note that this choice provides a good practical performance for various model selection problems.

To find the asymptotic properties of the penalized estimator, we assume that the first d_0 functional weights are nonzero and the remaining $d - d_0$ functional weights are zero. That is, $\lambda_0(\tau) = [\lambda_{0a}^{\intercal}(\tau), \lambda_{0b}^{\intercal}(\tau)]^{\intercal}$, where $\lambda_{0a}(\tau) = [\lambda_{01}(\tau), \cdots, \lambda_{0d_0}(\tau)]^{\intercal}$ with $\|\lambda_{0k}\| \neq 0$ for $1 \leq k \leq d_0$ and $\lambda_{0b}(\tau) = [\lambda_{0(d_0+1)}(\tau), \cdots, \lambda_{0d}(\tau)]^{\intercal}$ with $\|\lambda_{0k}\| = 0$ for $d_0 + 1 \leq k \leq d$. Similarly, we let $\theta_0(\tau) = [\theta_{0a}^{\intercal}(\tau), \theta_{0b}^{\intercal}(\tau)]^{\intercal}$ with $\theta_{0a}(\tau) = [\theta_{01}(\tau), \cdots, \theta_{0d_0}(\tau)]^{\intercal}$ and $\theta_{0b}(\tau) = [\theta_{0(d_0+1)}(\tau), \cdots, \theta_{0d}(\tau)]^{\intercal}$, in which $\theta_{0b}(\tau)$ can be arbitrary since the corresponding weights are zeros. Moreover, we define $\delta_0(\tau) = [\theta_0^{\intercal}(\tau), \lambda_0^{\intercal}(\tau)]^{\intercal}$ and $\delta_{0a}(\tau) = [\theta_{0a}^{\intercal}(\tau), \lambda_{0a}^{\intercal}(\tau)]^{\intercal}$, and their corresponding penalized estimators $\hat{\delta}_{\gamma_T}(\tau) = [\hat{\theta}_{\gamma_T}^{\intercal}(\tau), \hat{\lambda}_{\gamma_T}^{\intercal}(\tau)]^{\intercal}$ and $\hat{\delta}_{a,\gamma_T}(\tau) = [\hat{\theta}_{a,\gamma_T}^{\intercal}(\tau), \hat{\lambda}_{a,\gamma_T}^{\intercal}(\tau)]^{\intercal}$, respectively. One can partition $\delta_0(\tau)$ into an identified subset $[\theta_{0a}^{\intercal}(\tau), \lambda_{0a}^{\intercal}(\tau), \lambda_{0b}^{\intercal}(\tau)]^{\intercal}$ and an unidentified subset $\theta_{0b}(\tau)$ in which the former is unique and the latter is a vector of arbitrary fixed points. Further, we include the following additional technical conditions:

(B1) $\lim_{T\to\infty} \inf_{z\to 0^+} P'_{\gamma_T}(z)/\gamma_T > 0$, $h \propto T^{-1/5}$, $T^{-1/2}\gamma_T \to 0$ and $T^{-1/10}\gamma_T \to \infty$, as $T \to \infty$.

The regularity condition (B1) implies that the order of tuning parameter γ_T needs to be greater than $T^{1/10}$ and smaller than $T^{1/2}$, which will be crucial for the consistency result in

Theorem 3 and the oracle property in Theorem 4.

Theorem 3 *Let $\{x_{1i}, x_{2i}\}_{i=1}^T$ be a strictly stationary α-mixing sequence following the copula model in (1). For a fixed point $\tau \in (0,1)$, under Conditions A1-A6 and B1, $h \propto T^{-1/5}$ and $T^{-1/2}\gamma_T \to 0$ as $T \to \infty$, there exists a $\sqrt{Th}$-consistent estimator $\hat{\delta}_{\gamma_T}(\tau)$ that maximizes (3) satisfying $\|\hat{\delta}_{\gamma_T}(\tau) - \delta_0(\tau)\| = O_p(1/\sqrt{Th})$.*

Remark 3 *Theorem 3 shows the consistency for the nonparametric kernel-based estimator $\hat{\delta}_{\gamma_T}(\tau)$ at a given point $\tau \in (0,1)$.*

Theorem 4 *(Oracle Property). Let $\{x_{1i}, x_{2i}\}_{i=1}^T$ be a strictly stationary α-mixing sequence following the copula model in (1). For a fixed point $\tau \in (0,1)$, under Conditions A1-A6 and B1, $h \propto T^{-1/5}$, $T^{-1/2}\gamma_T \to 0$ and $T^{-1/10}\gamma_T \to \infty$ as $T \to \infty$, we have*

(a) Sparsity: $\|\hat{\lambda}_{\cdot k}\| = 0$ for $k = d_0 + 1, \cdots, d$,

(b) Asymptotic normality:

$$\sqrt{Th}(\hat{\delta}_{a,\gamma_T}(\tau) - \delta_{0a}(\tau) - h^2 B_a(\tau)) \to N(0, \upsilon_0 \Sigma_a(\tau)^{-1}\Omega_a(\tau)\Sigma_a(\tau)^{-1}).$$

where $\Sigma_a(\tau) = -E\{\ell_i''(\delta_{0a}(\tau))|t_i = \tau\}$, $\Omega_a(\tau) = E\{\ell_i'(\delta_{0a}(\tau))\ell_i'(\delta_{0a}(\tau))^\intercal|t_i = \tau\}$, and the bias term $h^2 B_a(\tau) = \frac{h^2}{2}\delta_{0a}''(\tau)\mu_2$.

Sparsity is an important statistical property in high-dimensional statistics. By assuming that only a small subset of copula families is important, it can reduce complexity so that it improves interpretability and predictability of the model. The sparsity property from Theorem 4 demonstrates that the penalized trending mixture copula model shrinks superfluous components of the weight vector exactly to zero with probability one as the sample size T goes to infinity.

2.3 Practical issues

A. A nonparametric EM algorithm. Clearly, it will be hard to obtain the nonparametric estimators by maximizing the penalized local log-likelihood copula function in (3) if the number of copulas is large. In this section, we propose a nonparametric version of the expectation maximization (EM) algorithm to estimate the weights and dependence parameters, which dramatically reduces the computational complexity. It iteratively alternates between an expectation step and a maximization step. The E-step updates the weights of each copula with given dependence parameters, and the M-step maximizes the local log-likelihood with respect to the dependence parameters for given copula weights. For details of the EM algorithm and its applications in parametric mixture copula models, see Dempster et al. (1977) and Cai and Wang (2014).

To develop a nonparametric version of the EM algorithm for the proposed model, we add constraints in the penalized objective function. The estimator $\hat{\delta}(t_j)$ at a given iteration step can be obtained by maximizing the criterion function

$$Q^P(\delta(t_j)) = \sum_{i=1}^{T} \ell_i(\delta(t_j))K_h(t_i - t_j) - T\sum_{k=1}^{d} \frac{P'_{\gamma_k}(\|\hat{\lambda}^{(0)}_{\cdot k}\|)}{2\|\hat{\lambda}^{(0)}_{\cdot k}\|}\lambda_k^2(t_j) + \rho_{t_j}(1 - \sum_{k=1}^{d}\lambda_k(t_j)),$$

with $\hat{\lambda}^{(0)}_{\cdot k}$ being the estimates from the previous iteration. At the first iteration, $\hat{\lambda}^{(0)}_{\cdot k}$ is a set of starting values for the weights. We take the first derivative of $Q^P(\delta(t_j))$ with respect to $\lambda_k(t_j)$, and multiply both sides by $\lambda_k(t_j)$, which leads to

$$\sum_{i=1}^{T} \frac{\lambda_k(t_j)c_k(u_{1i}, u_{2i}, \psi, \theta_k(t_i))}{c_c(u_{1i}, u_{2i}, \psi, \delta_k(t_i))}K_h(t_i - t_j) - T\frac{P'_{\gamma_k}(\|\hat{\lambda}^{(0)}_{\cdot k}\|)}{\|\hat{\lambda}^{(0)}_{\cdot k}\|}\lambda_k^2(t_j) - \rho_{t_j}\lambda_k(t_j) = 0. \quad k = 1, \ldots, d,$$

where $c_c(u_{1i}, u_{2i}, \psi, \delta_k(t_i)) = \sum_{k=1}^{d} \lambda_k(t_i)c_k(u_{1i}, u_{2i}, \psi, \theta_k(t_i))$. Taking sums of both sides of the equation over all k's, we obtain

$$\rho_{t_j} = \sum_{i=1}^{T} K_h(t_i - t_j) - T\sum_{k=1}^{d} \frac{P'_{\gamma_k}(\|\hat{\lambda}^{(0)}_{\cdot k}\|)}{\|\hat{\lambda}^{(0)}_{\cdot k}\|}\lambda_k^2(t_j).$$

Next, we introduce the expectation and maximization steps.

Expectation step

Let $\lambda_k^{(0)}(\tau)$ and $\theta_k^{(0)}(\tau)$ be the initial estimators in each iterative step. Given a grid point τ, we update the new weight parameters $\lambda_k^{(1)}(\tau)$ as

$$\lambda_k^{(1)}(\tau) = \Big(\sum_{i=1}^{T} \frac{\lambda_k^{(0)}(\tau)c_k(u_{1i}, u_{2i}, \psi, \theta_k^{(0)}(\tau))}{c_c(u_{1i}, u_{2i}, \psi, \delta_k^{(0)}(\tau))}K_h(t_i - \tau) - TD_k^{(0)}\Big)/\rho_\tau^{(0)}, \qquad \text{for} \quad k = 1, \ldots, d,$$

where $\rho_\tau^{(0)} = \sum_{i=1}^T K_h(t_i - \tau) - T\sum_{k=1}^d D_k^{(0)}$ with $D_k^{(0)} = \frac{P'_{\gamma_k}(\|\hat{\lambda}_{\cdot k}^{(0)}\|)}{\|\hat{\lambda}_{\cdot k}^{(0)}\|}\lambda_k^{(0)2}(\tau)$.

Maximization step

After updating the weight $\lambda_k^{(0)}(\tau)$ with $\lambda_k^{(1)}(\tau)$ from the above E-step, we obtain the dependence estimator $\theta^{(1)}(\tau)$ by maximizing the objective function $Q^P(\delta(\tau))$ with respect to the dependence parameter θ. Note that the penalty and constraint terms of $Q^P(\delta(\tau))$ do not depend on θ, so that it is equivalent to maximize $Q(\delta(\tau)) = \sum_{i=1}^T \ell_i(\delta(\tau))K_h(t_i - \tau)$. We use a one-step Newton-Raphson method:

$$\theta^{(1)}(\tau) = \theta^{(0)}(\tau) - \frac{Q'_\theta(\delta^{(0)}(\tau))}{Q''_\theta(\delta^{(0)}(\tau))},$$

where $Q'_\theta(\delta(\tau))$ and $Q''_\theta(\delta(\tau))$ are the first and second derivatives of $Q_\theta(\delta(\tau))$ with respect to θ, respectively. It may not be easy to find explicit expressions for $Q'_\theta(\delta(\tau))$ and $Q''_\theta(\delta(\tau))$, in which case one can use numerical derivatives as

$$Q'_{\theta_k}(\delta(\tau)) \approx \frac{Q(\delta(\tau)+\epsilon\iota_k) - Q(\delta(\tau)-\epsilon\iota_k)}{2\epsilon} \quad \text{and} \quad Q''_{\theta_k}(\delta(\tau)) \approx \frac{Q'_{\theta_k}(\delta(\tau)+\epsilon\iota_k) - Q'_{\theta_k}(\delta(\tau)-\epsilon\iota_k)}{2\epsilon}$$

where ϵ is a small positive real number and ι_k is a $(2d)$-dimensional vector with the k-th element being one and the others being zero.

B. Goodness-of-fit. In the literature, various goodness-of-fit tests for the specification of parametric copula functions have been proposed (Dobric and Schmid, 2007; Lin and Wu, 2015; Zhang, Okhrin, Zhou and Song, 2016). However, to the best of our knowledge, goodness-of-fit tests for nonparametric copula models have not been studied. To evaluate the performance of the proposed mixture copula model, we use a Rosenblatt probability integral transformation as in Dobric and Schmid (2007). We define the random variable

$$S(u_1, u_2) = [\Phi^{-1}(u_1)]^2 + [\Phi^{-1}(C(u_2|u_1))]^2, \tag{4}$$

where $C(u_2|u_1) = P(U_2 \le u_2 | U_1 = u_1)$ and $\Phi(\cdot)$ is the standard normal cumulative distribution function. Note that $C(u_2|u_1) = \partial C(u_1, u_2)/\partial u_1$, which is available in analytical form for most copulas. We consider the null hypothesis $H_0 : (u_1, u_2)$ follows copula $C(u_1, u_2)$. Under H_0, u_1 and $C(u_2|u_1)$ are i.i.d. and mutually independent $U(0, 1)$ distributed random variables. Thus, H_0 implies that $S(u_1, u_2)$ follows a $\chi^2(2)$ distribution, and we can use a random sample $\{u_{1i}, u_{2i}\}_{i=1}^T$, to test this hypothesis.

We consider three tests including the Kolmogorov-Smirnov (KS) test, the Cramer-von Mises (CM) test, and the Anderson-Darling (AD) test:

$$t_{KS} = \sup_S |F_T(S) - F(S)|,$$

$$t_{CM} = \int_{-\infty}^{\infty} [F_T(S) - F(S)]^2 dF(S),$$

$$t_{AD} = \sup_S \frac{\sqrt{T}|F_T(S) - F(S)|}{\sqrt{F(S)(1 - F(S))}},$$

where $F_T(S)$ is the empirical cumulative distribution function for the random variable S, and $F(S)$ is the cumulative distribution function for the Chi-squared distribution with two degrees of freedom. Standard critical values cannot be used to make inference since the time series data are weakly dependent. Moreover, the parameters are time-varying and the estimation error should not be ignored. To overcome these difficulties, we propose the following bootstrap algorithm to compute p-values of these three test statistics:

1. Generate a sample sequence $\{x_{1,i}^*, x_{2,i}^*\}_{i=1}^T$ from the original data $\{x_{1,i}, x_{2,i}\}_{i=1}^T$ using a stationary bootstrap technique as described in Appendix A.
2. Obtain $\hat{u}_{1i}^*$ and $\hat{u}_{2i}^*$ using the marginal distributions, respectively.
3. Calculate new local constant estimators $\hat{\delta}^*(t_i)$ by equation (3) with paired estimators $\{\hat{u}_{1i}^*, \hat{u}_{2i}^*\}_{i=1}^T$, and obtain $S(\hat{u}_{1i}^*, \hat{u}_{2i}^*)$ by equation (4).
4. Use the values $S(\hat{u}_{1i}^*, \hat{u}_{2i}^*)$ to construct the bootstrap statistics t_{KS}^*, t_{CM}^*, and t_{AD}^*.
5. Repeat Steps 1-4 B times (say, B=1000) and get B values of the statistics t_{KS}^*, t_{CM}^*, and t_{AD}^* respectively.
6. Calculate the values of t_{KS}, t_{CM} and t_{AD} from the original sample $\{u_{1i}, u_{2i}\}_{i=1}^T$ and compute the p-value of the tests based on the relative frequency of the events $\{t_{KS}^* \geq t_{KS}\}$, $\{t_{CM}^* \geq t_{CM}\}$, $\{t_{AD}^* \geq t_{AD}\}$ in the replications of the bootstrap sampling.

C. Bandwidth and tuning parameter selection. The bandwidth h determines the trade-off between the bias and variance of the nonparametric estimators, while the tuning parameter

γ_T adjusts the weight for the penalty term. We need to choose suitable regularization parameters to do the nonparametric estimation and variable selection simultaneously. Various methods for the selection of bandwidths and tuning parameters have been proposed in the variable selection literature, such as cross-validation, AIC- and BIC-type criteria, among others. Due to the time series nature of the sequence $\{x_{1i}, x_{2i}\}_{t=1}^T$, we propose to use a *forward leave-one-out* cross-validation to select both the bandwidth h and tuning parameter γ_T in the penalty term simultaneously.

Define $\hat{\delta}(h, \gamma_T)$ as the nonparametric estimators for the penalized trending mixture copula models in (3) with known bandwidth h and tuning parameter γ_T. For each data point $i_0 + 1 \leq i^* \leq T$, we use the data $\{x_{1i}, x_{2i}, i < i^*\}$ to construct the estimate $\hat{\delta}_{t^*}(h, \gamma_T)$ at the sample point $\{x_{1i^*}, x_{2i^*}\}$, where i_0 is the minimum window size used to estimate $\hat{\delta}_{i_0+1}(h, \gamma_T)$. Under this forward recursive scheme, we obtain the sequential estimators $\{\hat{\delta}_{i^*}(h, \gamma_T)\}_{i^*=i_0+1}^T$. The optimal bandwidth h^* and tuning parameter γ_T^* can be obtained by maximizing the objective function

$$(h^*, \gamma_T^*) = \arg\max_{(h,\gamma_T)} \sum_{i^*=i_0+1}^{T} \{\ell_{i^*}(\hat{\psi}, \hat{\delta}(t_i^*)) | \hat{\delta}_{i^*}(h, \gamma_T)\},$$

and (h^*, γ_T^*) is the *forward leave-one-out* cross-validation estimator in terms of the log-likehood.

D. Confidence intervals. For inference, i.i.d bootstrap approaches are not applicable here since most of the financial/economic data are dependent. Patton (2012a) suggested a block bootstrap to construct the pointwise confidence intervals on copula dependence parameters for serially dependent data although its theoretical properties require formal justification. The intuition behind this method is that, by dividing the data into several blocks, it can preserve the original time series structure within a block. A simple block bootstrap for calculating confidence intercals can be implemented as follows:

1. Generate a sample sequence $\{x_{1,i}^*, x_{2,i}^*\}_{i=1}^T$ from the original data $\{x_{1,i}, x_{2,i}\}_{i=1}^T$ using a stationary bootstrap technique as described in Appendix A.
2. Obtain $\hat{u}_{1i}^*$ and $\hat{u}_{2i}^*$ using the marginal distributions, respectively.
3. Calculate new local constant estimators $\hat{\delta}^*(\tau)$ at the grid point τ by equation (3) with paired estimators $\{\hat{u}_{1i}^*, \hat{u}_{2i}^*\}_{i=1}^T$.

4. Repeat Step 1-3 B times (say, B=1000), and get B values of the estimators $\hat{\delta}^*(\tau)$ as an empirical sample at each grid point τ. Let the $\alpha/2$-th and $(1-\alpha/2)$-th percentiles of the sample sequence $\{\hat{\delta}^*(\tau)\}$ be $q_{\alpha/2}$ and $q_{1-\alpha/2}$, respectively.
5. The empirical $100(1-\alpha)\%$ confidence interval for $\hat{\delta}(\tau)$ is $[q_{\alpha/2}, q_{1-\alpha/2}]$.

3 Numerical studies

This section illustrates the finite-sample performance of the proposed estimation and selection method through a series of simulation studies. We consider the bivariate case where the data are generated by AR(1)-GARCH(1,1) processes:

$$x_{si} = \gamma_s x_{s,i-1} + e_{si}, \qquad\qquad s = 1,2;\ i = 2, ..., T,$$

where $\gamma_1 = 0.1$, $\gamma_2 = 0.05$, $e_{si} = \sigma_{si}\epsilon_{si}$, and ϵ_{si} has a $t(3)$ marginal distribution. The dependence structure between ϵ_{1i} and ϵ_{2i} is governed by a given copula function. Furthermore,

$$\sigma_{si}^2 = \alpha_{s0} + \alpha_{s1} e_{s,i-1}^2 + \beta_{s1}\sigma_{s,i-1}^2,$$

where $\alpha_{10} = 0.0001$, $\alpha_{11} = 0.02$, $\beta_{11} = 0.93$ for the first margin, and $\alpha_{20} = 0.0001$, $\alpha_{21} = 0.03$, $\beta_{21} = 0.92$ for the second margin. Our working mixture copula model consists of three copulas: the Gumbel, Frank and Clayton copulas, which are widely used in empirical studies. The Frank copula shows a symmetric dependence structure, while the Clayton and Gumbel copulas are asymmetric. In particular, the Clayton copula displays strong lower tail dependence, while the Gumbel copula exhibits strong upper tail dependence. Data are generated from mixture copula models consisting of two of the three copulas. That is,

$$(\epsilon_{1i}, \epsilon_{2i}) \sim \sum_{k=1}^{3} \lambda_k(t_i) C_k(F_1(\epsilon_{1i}), F_2(\epsilon_{2i}), \theta_k(t_i)),$$

where one of the three weight parameters (λ_1, λ_2 and λ_3) is zero.

We consider two cases for the weights and dependence parameters. First, the weights and dependence parameters are set to constants. Second, they are time-varying according to some given functions. In each case, we simulate three mixture copulas with two components. The sample size $T = 400$ and 800, and each simulation is repeated 500 times ($M = 500$).

For each sample we calculate the estimated weights and dependence parameters on a grid of 50 equally spaced points $\tau_i = -0.01 + 0.02i$ for $i \in \{1, 2, ..., 50\}$.

3.1 Case I simulations

We first consider the scenario where data are generated from mixture copulas with constant weights and dependence parameters. Let λ_1, λ_2 and λ_3 denote the weights of the Gumbel, Frank and Clayton copulas respectively, and θ_1, θ_2 and θ_3 the corresponding dependence parameters. We consider the following models for the weights and dependence parameters:

- Model 1: $\lambda_1 = 1/2$, $\lambda_2 = 1/2$, $\lambda_3 = 0$, $\theta_1 = 6$, $\theta_2 = 4$;
- Model 2: $\lambda_1 = 1/2$, $\lambda_2 = 0$, $\lambda_3 = 1/2$, $\theta_1 = 6$, $\theta_3 = 5$;
- Model 3: $\lambda_1 = 0$, $\lambda_2 = 1/2$, $\lambda_3 = 1/2$, $\theta_2 = 4$, $\theta_3 = 5$.

We summarize the estimation results for the weights and dependence parameters in the Case I simulations in Tables 1-2, Panel A and Figures 1-3[1]. Table 1, Panel A presents the percentages corresponding to the correctly and incorrectly (in parentheses) selected copulas. From Table 1, Panel A, the proposed method performs very well for selecting appropriate copulas from mixture copula models with constant parameters, although our method is designed for trending mixture copula models. For all three models, the correct component copulas are selected with 100% probability. Moreover, the probability that the incorrect copulas are chosen is small. There is zero probability of selecting the incorrect copulas for the mixture of Gumbel and Frank, and the mixture of Clayton and Frank. For the mixture model consisting of the Gumbel and Clayton copulas, the chance to incorrectly select the Frank copula is also small and decreases with T.

[INSERT TABLE 1 ABOUT HERE]

To examine the performance of the proposed method in estimating the unknown parameters, we calculate the mean square errors (MSEs) of the estimated weights and dependence

[1] We omit the results of the marginal parameters to save space.

parameters for the mixture copula models under Case I simulations. The MSEs are calculated as

$$
\begin{aligned}
MSE(\widehat{\theta}_k) &= \frac{1}{M}\frac{1}{50}\sum_{j=1}^{M}\sum_{i=1}^{50}\left(\widehat{\theta}_{jk}(\tau_i)-\theta_k\right)^2, \text{ for } k=1,2,3,\\
MSE(\widehat{\lambda}_k) &= \frac{1}{M}\frac{1}{50}\sum_{j=1}^{M}\sum_{i=1}^{50}\left(\widehat{\lambda}_{jk}(\tau_i)-\lambda_k\right)^2, \text{ for } k=1,2,3,
\end{aligned}
$$

where 50 is the number of grid points and $M = 500$ is the replication time.

The results are shown in Table 2, Panel A. As expected, the MSEs decrease when the sample size increases for all three models.

[INSERT TABLE 2 ABOUT HERE]

We further evaluate the quality of the estimators graphically. Figures 1-3 respectively display simulation results of the weights and dependence parameters for Models 1-3 under Case I simulations. In each figure, the black solid line denotes true parameters (the weight or dependence parameter), and two curves respectively represent medians (blue) and means (red) of the 500 simulation parameter function estimates at the grid points. The two green dashed lines represent the 5% and 95% percentiles of the parameter estimates at the grid points. To save space, we only present the results for $T = 800$. In all three models, the median and mean curves are close to the true parameter paths, which are constant in this case.

[INSERT FIGURES 1-3 ABOUT HERE]

3.2 Case II simulations

In the Case II simulations, the weights and dependence parameters are dynamic according to the following functions:

- Model 1: $\lambda_1(\tau) = 0.7 - 0.4\sin^2(\frac{\pi}{2}\tau)$, $\lambda_2(\tau) = 1-\lambda_1(\tau)$, $\lambda_3(\tau) = 0$, $\theta_1(\tau) = e^{2\tau} + 3$, $\theta_2(\tau) = 6\tau^2 + 4$;
- Model 2: $\lambda_1(\tau) = 0.7 - 0.4\sin^2(\frac{\pi}{2}\tau)$, $\lambda_2(\tau) = 0$, $\lambda_3(\tau) = 1-\lambda_1(\tau)$, $\theta_1(\tau) = e^{2\tau} + 3$, $\theta_3(\tau) = \ln(1 + \tau T) + 3$;

- Model 3: $\lambda_1(\tau) = 0$, $\lambda_2(\tau) = 0.7 - 0.4\sin^2(\frac{\pi}{2}\tau)$, $\lambda_3(\tau) = 1 - \lambda_2(\tau)$, $\theta_2(\tau) = 6\tau^2 + 4$, $\theta_3(\tau) = \ln(1 + \tau T) + 3$;

where $\lambda_k(\tau)$ and $\theta_k(\tau)$, $k = 1, 2, 3$, respectively represent the weights and dependence parameters of the Gumbel, Frank and Clayton copulas.

Tables 1-2, Panel B and Figures 4-6 show the estimation results for this case. We use Table 1, Panel B to examine whether the proposed method can efficiently select appropriate copulas from different trending mixture copula models. As in Table 1, Panel A, the values without parentheses correspond to the percentages that copulas in the mixture models are selected correctly, and the values within parentheses are the percentages that copulas not in the mixture models are selected incorrectly. For all three trending mixture copula models, the correct copulas are selected in all replications. For the mixture of the Gumbel and Frank, the probability of choosing the incorrect copula (Clayton) is zero. For the other two mixtures, the chance to select incorrect copulas is also small. For example, when $T = 800$, there is only 1.2% to choose Gumbel when the true model is a mixture of Clayton and Frank, and 6% to select Frank when data are generated from a mixture of Gumbel and Clayton. Therefore, Table 1, Panel B demonstrates the good performance of the proposed method in copula selection for mixture copulas with dynamic parameters.

We now use Table 2, Panel B and Figures 4-6 to check whether the proposed method can accurately estimate the unknown parameters under the Case II simulations. Again, we omit the results of the marginal parameters to save space. In Table 2, Panel B, we calculate the MSEs of the estimated weights and dependence parameters for the mixture copulas with dynamic parameters. Similar to the Case I simulations, the MSEs are calculated as

$$
\begin{aligned}
MSE(\widehat{\theta}_k) &= \frac{1}{M}\frac{1}{50}\sum_{j=1}^{M}\sum_{i=1}^{50}\left(\widehat{\theta}_{jk}(\tau_i) - \theta_k(\tau_i)\right)^2, \text{ for } k = 1, 2, 3, \\
MSE(\widehat{\lambda}_k) &= \frac{1}{M}\frac{1}{50}\sum_{j=1}^{M}\sum_{i=1}^{50}\left(\widehat{\lambda}_{jk}(\tau_i) - \lambda_k(\tau_i)\right)^2, \text{ for } k = 1, 2, 3.
\end{aligned}
$$

We note two observations from Table 2, Panel B. First, as the sample size increases from 400 to 800, the MSEs decrease and the estimates become more accurate. Second, compared to the results in Table 2, Panel A, the MSEs in Panel B are larger in most cases. This is not surprising because the models in Case II are trending mixture copulas with dynamic parameters, which are more difficult to estimate than the models in Case I (mixture copulas with constant parameters).

Finally, Figures 4-6 present the estimated and the true parameter paths for different trending mixture copulas models ($T = 800$). Similar to Figures 1-3, we use a black solid line to depict the true parameter function (the weight or dependence parameter), and two curves to respectively represent medians (blue) and means (red) of the 500 simulation parameter function estimates at the grid points. The two green dashed curves represent the 5% and 95% percentiles of the copula estimates at the grid points. From Figures 4-6, one can observe that the median and mean paths are still close to the true parameter functions in all models.

[INSERT FIGURES 4-6 ABOUT HERE]

In sum, the simulation results obtained from the Case I and II simulations demonstrate that the proposed method works reasonably well in selecting and estimating different mixture copula models with both constant and dynamic parameters.

4 An empirical study

In this section, we apply the proposed model and method to analyze the co-movements of returns among international stock markets during different periods. Specifically, we consider weekly returns of the Morgan Stanley Capital International (MSCI) equity indexes of four economies (in U.S. Dollars): the United States (US), the United Kingdom (UK), Hong Kong (HK), and South Korea (KR). These four economies are much affected by the Asian crisis of 1997 and/or the global financial crisis of 2008. By analyzing the evolution of the dependence structures among these four markets, we can examine how these markets are related, for example, in tranquil periods and in crisis periods.

As is well known, correctly understanding and accurately measuring the co-movements across international equity markets play important roles for portfolio allocation and asset pricing. Because financial data are typically non-Gaussian (e.g. Ang and Chen, 2002, Longin and Solnik, 2001), many researchers choose to use copula-based models to measure the dependence structures among international stock markets. For example, Hu (2006), Cai and Wang (2014), and Liu et al. (2018) employ time-invariant mixture copula models and find that international stock markets usually show lower tail dependence, which implies that the markets are more likely to crash together than to boom together. The semiparametric dynamic copula (SDC) model of Hafner and Reznikova (2010) is based on a single copula whose parameter changes over time in a nonparametric way. We extend these approaches by first noting that the dependence structures among international markets are likely to change substantially over time. A time-invariant copula model is incapable of capturing the evolution of the dependence structures. Second, not only the dependence parameter but also the dependence pattern (the weight of each copula) may change over time and hence a single copula model may not be adequate. Therefore, it is of particular interest to use the trending mixture copula model with both time-varying dependence parameters and weights to analyze the co-movements across international stock markets.

4.1 Data

The data we use span the period of over 28 years from January 1990 until July 2018, with a total of 1488 observations for each economy. We first obtain the equity indexes from Datastream and then calculate their log-returns by $r_{i,t} = \log(P_{i,t}) - \log(P_{i,t-1})$, where $P_{i,t}$ is the stock index of the i-th market at time t. We use weekly data instead of daily data to remove the effect of different trading hours for international stock markets (Chollete et al., 2009, and Hafner and Reznikova, 2010). Descriptive statistics are presented in Table 3, Panel A. The United States market exhibits the highest mean and median returns. The Korea market shows the largest volatility of returns. Negative skewness and excess kurtosis are found in all series, suggesting that the observations should be filtered. We also employ the Jarque-Bera test for normality and the test strongly rejects the null hypothesis for all series.

[INSERT TABLE 3 ABOUT HERE]

Table 3, Panel B reports the unconditional correlation coefficients and Kendall's τs (in parentheses) across the four markets. We observe that the US and UK markets display the highest correlation, based on both the correlation coefficient (0.681) and the Kendall's τ (0.451). The US-HK, UK-HK, and HK-KR pairs show similar dependence of moderate size (around 0.5 for the correlation coefficients and around 0.35 for Kendall's τs). The least dependent pairs are US-KR (0.389 for correlation and 0.249 for Kendall's τ) and UK-KR (0.409 for correlation and 0.280 for Kendall's τ).

4.2 The models for the marginal distributions

First of all, we model the marginal distributions of the data. We employ AR(p)-GARCH(1,1) models to capture possible autocorrelation and conditional heteroscedasticity in returns. The Bayesian information criterion (BIC) is used to select the appropriate number of lags p of the AR(p) models. Because the returns show conditional leptokurtosis properties, the error terms are assumed to follow a standardized Student-t distribution. Specifically, we use the following models for the marginal distributions:

$$
\begin{aligned}
X_{it} &= \gamma_{i0} + \sum_{j=1}^{p} \gamma_{ij} X_{i,t-j} + e_{it}, \quad e_{it} = \sigma_{it}\epsilon_{it}, \\
\sigma_{it}^2 &= \alpha_{i0} + \alpha_{i1} e_{i,t-1}^2 + \beta_{i1}\sigma_{i,t-1}^2,
\end{aligned}
$$

where X_{it} denotes the return of the i-th market at time t. The innovations ϵ_{it} are assumed to be independently identically distributed (i.i.d.) with the distribution given by a standardized student-t distribution with ν_i degrees of freedom.

The estimated parameters of the marginal models are presented in Table 4. AR(1) models are selected for the US and UK markets. For Hong Kong and Korea, BIC and Ljung-Box statistics show lack of autocorrelation. The degrees of freedom parameter of the t-distribution is much lower for the United States than for other economies, implying that the US returns have the fattest tails. As shown in Fermanian and Scaillet (2005), correctly specified marginal models are critical for copula selection and estimation. Therefore, after filtering the data, we employ the Ljung-Box (LB) tests for autocorrelation and the Kolmogorov-Smirnov (KS) tests for density specification. The p-values of the tests are displayed in the last three columns of Table 4. All models pass the LB and KS test at the 10% level, showing that the marginal models are well-specified.

[INSERT TABLE 4 ABOUT HERE]

4.3 The models for the copula

We focus on studying the dependence structures of four pairs (US-UK, US-HK, UK-HK, and HK-KR) with higher correlations. Scatter plots (omitted here) of four pairs of standardized returns show violations of elliptical multivariate distributions, because asymmetry and a large number of outliers can be observed in all pairs. Therefore, we employ a mixture copula model including the Clayton, Frank and Gumbel copulas to implement copula selection and estimation. In such a way, we can capture various dependence structures in the data such as a lower or upper tail dependence, or a symmetric but non-elliptical dependence structure.

We first fit the data to a time-invariant mixture copula model to examine the overall dependence structures during the period of 28 years. The Cai and Wang (2014) penalized likelihood method is employed to select and estimate the model. The results are reported in Table 5. We have two findings from Table 5. First, the Gumbel copula is excluded from the mixture model for all pairs of data, implying that no pairs exhibit upper tail dependence. Second, the Clayton copula is selected and the weight and dependence parameters are statistically significant away from zero for all pairs. This indicates that lower tail dependence can be found for all pairs of markets. These two findings are similar to Cai and Wang (2014).

[INSERT TABLE 5 ABOUT HERE]

Although the time-invariant model can tell us that overall the pairs of markets show lower tail dependence, it can neither capture the evolution of the dependence structures, nor distinguish between the dependence structures in tranquil periods and those in crisis periods. Therefore, we next employ the trending mixture copula model proposed in this paper to analyze the dependence structures of the international stock markets. Figures 7-10 respectively present the estimation results and the 90% confidence intervals of all nonzero weights and dependence parameters for the US-UK, US-HK, UK-HK, and HK-KR pairs. In each figure, the path of the estimated parameter (the weight or dependence parameter) is represented by a blue solid line. The two red dashed curves show the 90% confidence intervals of each estimated parameter. The green two-dashed line (horizontal line) is the estimate using the time-invariant mixture copula model. We have several interesting results from these figures.

First, for all pairs of markets, the Clayton and Frank copulas are selected at any time period of the 28 years. The confidence intervals for the weights on Clayton and Frank do not cover zeros, showing that they are always statistically significant. On the other hand, the weight on Gumbel is always zero during the 28 years for all pairs. Therefore, the four pairs of markets show significantly lower tail dependence, but no upper tail dependence from 1990 to 2018. Second, we observe notable fluctuations of both the weights and dependence parameters during the 28-year period for all four pairs of markets, implying the limitation of time-invariant copula models.

For the US-UK pair presented in Figure 7, the weight and dependence parameters of the Clayton copula are both relatively small in the early 1990s. Meanwhile, the dependence parameter of the Frank copula is also small during this period. These findings show that both the lower tail dependence and the overall dependence are weak at the beginning of the 1990s. The dependence parameter of Clayton copula increases sharply after the events of September 11, 2001. At the same time, the weight of Clayton copula also reaches a relatively high value. During the financial crisis of 2008, weight and dependence parameters of the Clayton copula, and the dependence parameter of the Frank copula all increase sharply, reaching their maxima around 2010. This implies that the lower tail and general dependence of the two markets attain high levels in crisis periods.

Turning to the US-HK and UK-HK pairs, the dependence structures display similar evolution paths. Both pairs show relatively weak lower tail dependence and overall dependence during the 1990s. A notable jump in the Clayton parameter took place in 2008 for both pairs due to the financial crisis. An increase in the weight on Clayton can be observed during the same time period.

The last figure exhibits the dependence structure of the HK-KR pair. These two markets are strongly affected by the Asian crisis of 1997. Therefore, we can observe a relatively high level of the Clayton parameter, and a quick increase in the weight on Clayton in 1997. During the periods of the financial crisis of 2008, a significant increase and a remarkable jump in the weight and dependence parameter of the Clayton copula are also detected for this pair of markets.

[INSERT FIGURES 7-10 ABOUT HERE]

Finally, we check the goodness-of-fit of the estimated trending mixture copula model with the KS, CM, and AD tests for correct copula specification. The description of the three tests can be found in Section 2.3. Table 6 reports the bootstrap p-values of the three tests. All models pass these three tests with large p-values.

[INSERT TABLE 6 ABOUT HERE]

5 Conclusion

In this paper, we introduce a trending mixture copula model, in which both the weights and dependence parameters are deterministic functions of time. To reduce the risk of over-fitting and efficiency loss, we propose penalized trending mixture copula models with SCAD penalty term to do the estimation and copula selection simultaneously. Based on α-mixing conditions, asymptotic properties of the penalized and unpenalized estimators have been established. Meanwhile, we propose a nonparametric EM algorithm for computational feasibility to estimate the parameters. Three goodness-of-fit tests have been constructed to assess the appropriateness of the trending mixture model for fitting the data. We study and discuss the bandwidth selection and construction of pointwise confidence intervals. Moreover, we conduct Monte Carlo simulations which demonstrate the good performance of the proposed method in copula selection and estimation for mixture copulas with both constant and time-varying parameters. Furthermore, the proposed methodology has been applied to study the evolution of the dependence among four international stock markets. All pairs of markets present strong dependence at the lower tail that fluctuates significantly over time. All pairs exhibit the highest levels of both the lower tail and overall dependence during the financial crisis of 2008.

References

Abegaz, F., Gijbels, I. and Veraverbeke, N. (2012). Semiparametric estimation of conditional copulas. Journal of Multivariate Analysis, 110, 43-73.

Acar, E.F., Craiu, R.V. and Yao, F. (2011). Dependence calibration in conditional copulas: A nonparametric approach. Biometrics, 67, 445-453.

Almeida, C. and Czado, C. (2012). Efficient Bayesian inference for stochastic time-varying copula models. Computational Statistics & Data Analysis, 56, 1511-1527.

Ang, A. and Chen, J. (2002). Asymmetric correlations of equity portfolios. Review of Financial Studies, 63, 443-494.

Cai, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. Journal of Econometrics, 136, 163-188.

Cai, Z., Juhl, T. and Yang, B. (2015). Functional index coefficient models with variable selection. Journal of Econometrics, 189, 272-284.

Cai, Z. and Wang, X. (2014). Selection of mixed Copula model via penalized likelihood. Journal of the American Statistical Association, 109, 788-801.

Chollete, L., Heinen, A. and Valdesogo, A. (2009). Modeling international financial returns with a multivariate regime-switching copula. Journal of Financial Econometrics, 7, 437-480.

Creal, D., Koopman, S.J. and Lucas, A. (2013). Generalized autoregressive score models. Journal of Applied Econometrics, 28, 777-795.

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.

Dobric, J. and Schmid, F. (2007). A goodness of fit test for copulas based on rosenblatt's transformation. Computational Statistics and Data Analysis, 51, 4633-4642.

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.

Fermanian, J.D. and Lopez, O. (2018). Single-index copulas. Journal of Multivariate Analysis, 165, 27-55.

Fermanian, J.D. and Scaillet, O. (2005). Some statistical pitfalls in copula modeling for financial applications. In: Capital Formation, Governance and Banking. Nova Science Publishers, pp. 57-72.

Garcia, R. and Tsafack, G. (2011). Dependence structure and extreme comovements in international equity and bond markets. Journal of Banking & Finance, 35, 1954-1970.

Hafner, C.M. and Manner, H. (2012). Dynamic stochastic copula models: estimation, inference and applications. Journal of Applied Econometrics, 27, 269-295.

Hafner, C.M. and Reznikova, O. (2010). Efficient estimation of a semiparametric dynamic copula model. Computational Statistics & Data Analysis, 54, 2609-2627.

Hu, L. (2006). Dependence patterns across financial markets: A mixed copula approach. Applied Financial Economics, 16, 717-729.

Joe, H. (2000). Multivariate models and dependence concepts, Series: Monographs on Statistics and Applied Probability Series, Publisher: Taylor and Francis, Inc.

Lin, J. and Wu, X. (2015). Smooth tests of copula specifications. Journal of Business & Economic Statistics, 33, 128-143.

Liu, B.Y., Ji, Q. and Fan, Y. (2017). A new time-varying optimal copula model identifying the dependence across markets. Quantitative Finance, 17, 1-17.

Liu, G., Long, W., Zhang, X. and Li, Q. (2018). Detecting financial data dependence structure by averaging mixture copulas. Econometric Theory (forthcoming).

Longin, F. and Solnik, B. (2001). Extreme correlation of international equity markets, Journal of Finance, 56, 649-676.

Manner, H. and Reznikova, O. (2012). A survey on time-varying copulas: specification, simulation, and application. Econometric Review, 31, 654-687.

Mensi, W., Hammoudeh, S., Shahzad, S.J.H. and Shahbaz, M. (2016). Modeling systemic risk and dependence structure between oil and stock markets using a variational mode decomposition-based copula method. Journal of Banking & Finance, 75, 258-279.

Ning, C., Xu, D. and Wirjanto, T.S. (2015). Is volatility clustering of asset returns asymmetric. Journal of Banking & Finance, 52, 62-76.

Patton, A.J. (2006). Modeling asymmetric exchange rate dependence. International Economic Review, 47, 527-556.

Patton, A.J. (2012a). Copula methods for forecasting multivariate time series, in: Handbook of Economic Forecasting, Vol. 2, Elsevier, Oxford, 2011.

Patton, A.J. (2012b). A review of copula models for economic time series. Journal of Multivariate Analysis, 110, 4-18.

Politis, D.N. and Romano, J.P. (1994). The stationary bootstrap. Journal of the American Statistical Association, 89, 1303-1313.

Robinson, P.M. (1989). Nonparametric estimation of time-varying parameters. In: Hackl, P. (Ed.), Statistical Analysis and Forecasting of Economic Structural Change. Springer, Berlin, pp. 164-253.

Robinson, P.M. (1991). Time-varying nonlinear regression. In: Hackl, P., Westland, A.H. (Eds.), Economic Structure Change Analysis and Forecasting. Springer, Berlin, pp. 179-190.

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

Smith, M.S. and Vahey, S.P. (2016). Asymmetric forecast densities for U.S. macroeconomic variables from a gaussian copula model of cross-sectional and serial dependence. Journal of Business & Economic Statistics, 34, 52-70.

Tibshirani, R.J. (1996). Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society, Ser. B, 58, 267-288.

Wollschläger, M. and Schäfer, R. (2016). Impact of non-stationarity on estimating and modeling empirical copulas of daily stock returns. Journal of Risk, 19, 1-23.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Ser.B, 68, 49-67.

Zhang, S., Okhrin, O. Zhou, Q.M. and Song, P.X.K. (2016). Goodness-of-fit test for specification of semiparametric copula dependence models. Journal of Econometrics, 193, 215-233.

Zou, H. (2006). The adaptive Lasso and its oracle properties. Journal of the American Statistical Association, 101, 1418-1429.

Table 1: Percentages that the corresponding copulas are chosen correctly (incorrectly) for the mixture copula models in Case I (Panel A) and Case II (Panel B) simulations

Model	T	Gumbel	Frank	Clayton
Panel A: Case I				
Gumbel+Frank	400	1.000	1.000	(0.000)
	800	1.000	1.000	(0.000)
Gumbel+Clayton	400	1.000	(0.118)	1.000
	800	1.000	(0.078)	1.000
Clayton+Frank	400	(0.000)	1.000	1.000
	800	(0.000)	1.000	1.000
Panel B: Case II				
Gumbel+Frank	400	1.000	1.000	(0.000)
	800	1.000	1.000	(0.000)
Gumbel+Clayton	400	1.000	(0.142)	1.000
	800	1.000	(0.060)	1.000
Clayton+Frank	400	(0.000)	1.000	1.000
	800	(0.012)	1.000	1.000

NOTE: Values without parentheses are the percentages that copulas in the mixture copulas are chosen correctly. Values with parentheses are the percentages that copulas not in the mixture copulas are chosen incorrectly.

Table 2: Mean squared errors of the estimated weights and dependence parameters for the mixture copula models in Case I (Panel A) and Case II (Panel B) simulations

Model	T	(λ_1, θ_1)	(λ_2, θ_2)	(λ_3, θ_3)
Panel A: Case I				
Gumbel+Frank	400	(0.007, 0.516)	(0.007, 0.857)	
	800	(0.003, 0.240)	(0.003, 0.379)	
Gumbel+Clayton	400	(0.006, 0.574)		(0.008, 0.726)
	800	(0.003, 0.328)		(0.004, 0.365)
Clayton+Frank	400		(0.006, 0.825)	(0.006, 0.655)
	800		(0.001, 0.431)	(0.001, 0.393)
Panel B: Case II				
Gumbel+Frank	400	(0.009, 2.075)	(0.009, 1.502)	
	800	(0.004, 0.627)	(0.004, 0.498)	
Gumbel+Clayton	400	(0.006, 0.918)		(0.008, 2.957)
	800	(0.002, 0.417)		(0.003, 0.981)
Clayton+Frank	400		(0.009, 1.983)	(0.009, 1.767)
	800		(0.004, 0.720)	(0.004, 0.858)

Table 3: Summary statistics and correlations

	US	UK	HK	KR
Panel A: Summary statistics				
Mean (%)	0.140	0.061	0.136	0.061
Median (%)	0.284	0.193	0.277	0.218
Min (%)	-16.75	-15.21	-16.79	-40.25
Max (%)	10.34	11.56	14.03	30.02
Std. Dev.	0.022	0.026	0.032	0.047
Skewness	-0.669	-0.425	-0.482	-0.515
Kurtosis	7.892	6.117	5.813	10.556
JB statistic	1595	647	548	3606
JB p-value	0.000	0.000	0.000	0.000

Panel B: Correlations

	UK	KR	HK
US	0.681 (0.451)	0.389 (0.249)	0.487 (0.330)
UK		0.409 (0.280)	0.515 (0.351)
KR			0.495 (0.350)

NOTE: Panel A presents the summary statistics of weekly index returns for the United States (US), the United Kingdom (UK), Hong Kong (HK) and South Korea (KR). All returns are expressed in U.S. dollars from January, 1990 to July, 2018, which correspond to a sample of 1488 observations. JB statistic and JB p-value refer to Jarque-Bera test of normality. Panel B reports the linear correlation coefficients and the Kendall's τs (Kendall's τs are in parentheses) across the US, UK, HK and KR markets.

Table 4: Estimation results and tests of the marginal distribution models

	AR(p)	GARCH(1,1)			d.o.f	LB		KS
	γ_1	α_0	α_1	β_1	ν	4	16	
	(s.e.)	(s.e.)	(s.e.)	(s.e.)	(s.e.)			
US	-0.118	0.007E-3	0.100	0.890	5.389	0.617	0.643	0.289
	(0.026)	(0.004E-3)	(0.019)	(0.019)	(0.722)			
UK	-0.098	0.023E-3	0.127	0.840	9.034	0.892	0.577	0.383
	(0.027)	(0.009E-3)	(0.028)	(0.036)	(1.878)			
HK		0.020E-3	0.098	0.885	8.203	0.112	0.155	0.527
		(0.008E-3)	(0.020)	(0.023)	(1.482)			
KR		0.045E-3	0.115	0.862	10.707	0.120	0.337	0.634
		(0.015E-3)	(0.020)	(0.023)	(2.340)			

NOTE: The second to sixth columns report parameter estimates of AR(p)-GARCH(1,1) models with Student-t error terms for the index returns of the four markets. Values in parentheses are corresponding standard errors. The seventh and eighth columns report the p-values of the Ljung-Box (LB) tests for autocorrelation of the residuals using 4 and 16 lags, respectively. The last column presents the p-values of the Kolmogorov-Smirnov (KS) tests for evaluating the goodness-of-fit of the marginal distributions.

Table 5: Estimation results of the time-invariant mixture copula models for international markets

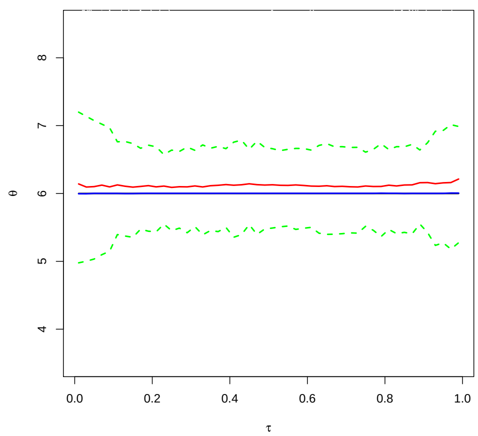
	Markets	Clayton	Gumbel	Frank
λ	US-UK	0.285(0.241,0.329)	0	0.715(0.671,0.759)
	US-HK	0.314(0.268,0.360)	0	0.686(0.640,0.732)
	UK-HK	0.352(0.304,0.399)	0	0.648(0.601,0.696)
	HK-KR	0.208(0.150,0.267)	0	0.792(0.733,0.850)
θ	US-UK	0.836(0.781,0.890)		5.172(4.823,5.521)
	US-HK	0.594(0.536,0.652)		3.761(3.565,3.956)
	UK-HK	0.657(0.585,0.729)		4.148(3.862,4.433)
	HK-KR	0.918(0.848,0.987)		3.208(2.923,3.493)

NOTE: This table presents estimates of the weights (λ) and dependence parameters (θ) of time-invariant mixture copula models using Cai and Wang (2014) penalized likelihood method. Values in parentheses are the 90% confidence interval of the estimates.

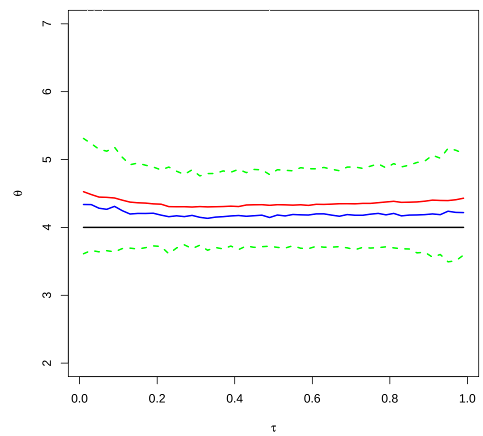
Table 6: Goodness-of-fit tests for the trending-mixture copula models

Markets	KS	CM	AD
US-UK	0.256	0.244	0.250
US-HK	0.670	0.470	0.686
UK-HK	0.468	0.474	0.530
HK-KR	0.362	0.372	0.592

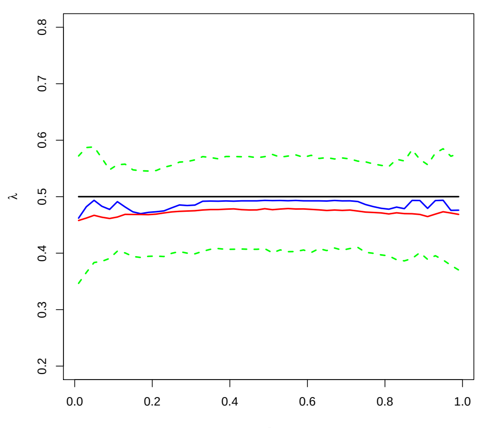
NOTE: This table reports the p-values from three goodness-of-fit tests including the Kolmogorov-Smirnov (KS) test, the Cramer-von Mises (CM) test and the Anderson-Darling (AD) test.



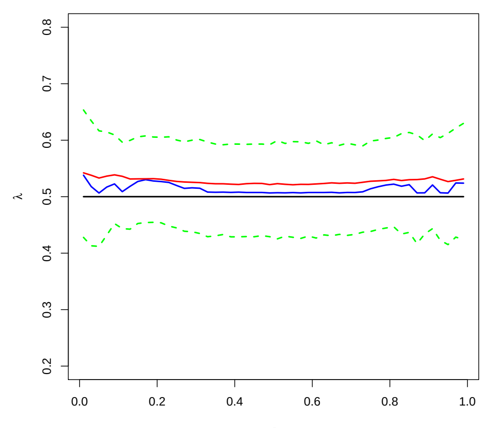
(a) Dependence parameter (Gumbel)



(b) Dependence parameter (Frank)

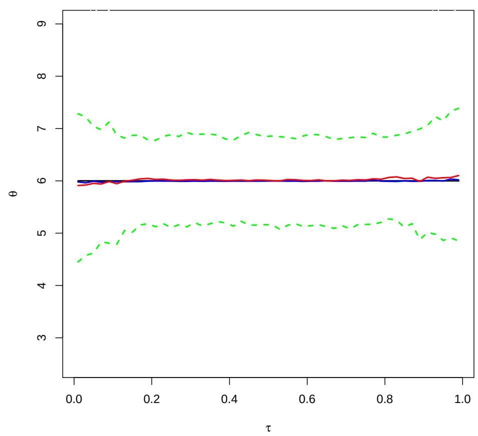


(c) Weight (Gumbel)

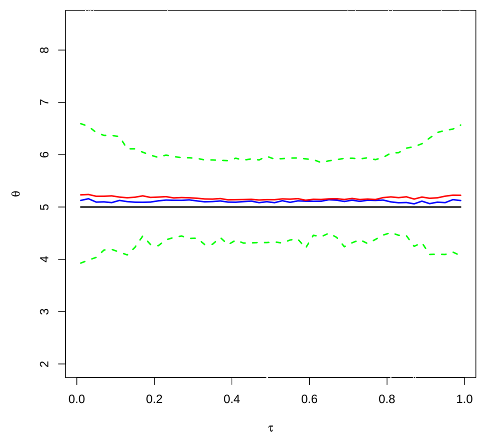


(d) Weight (Frank)

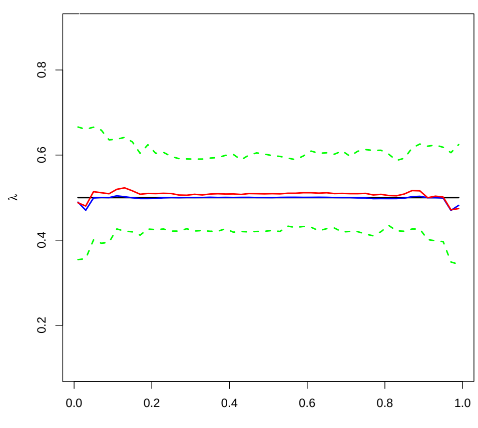
Figure 1: Simulation results of the weights and dependence parameters for Model 1 (500 repeats) in Case I simulations: true values (black solid lines), mean and median estimates (red and blue lines), and 5% and 95% percentile curves (green dashed lines). The sample size is 800.



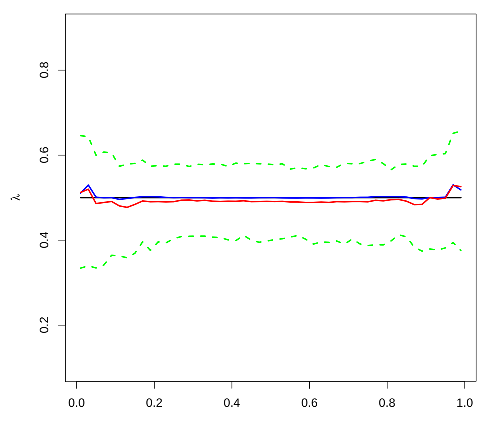
(a) Dependence parameter (Gumbel)



(b) Dependence parameter (Clayton)

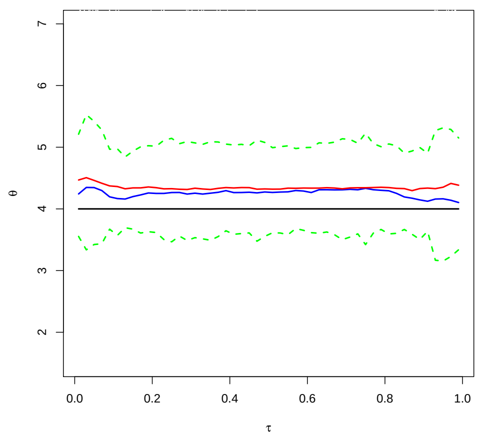


(c) Weight (Gumbel)

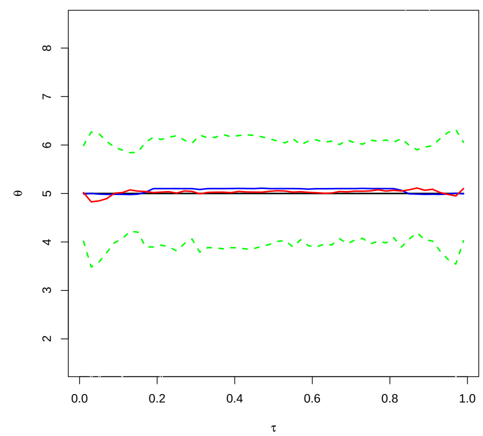


(d) Weight (Clayton)

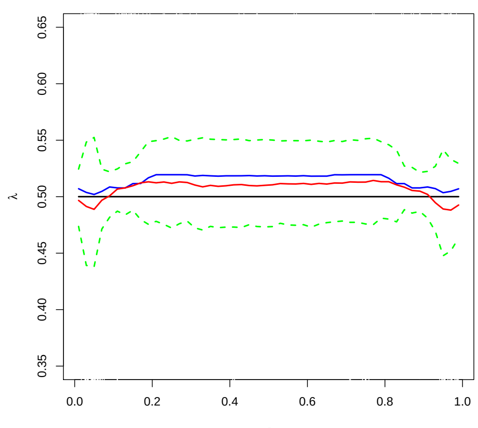
Figure 2: Simulation results of the weights and dependence parameters for Model 2 (500 repeats) in Case I simulations: true values (black solid lines), mean and median estimates (red and blue lines), and 5% and 95% percentile curves (green dashed lines). The sample size is 800.



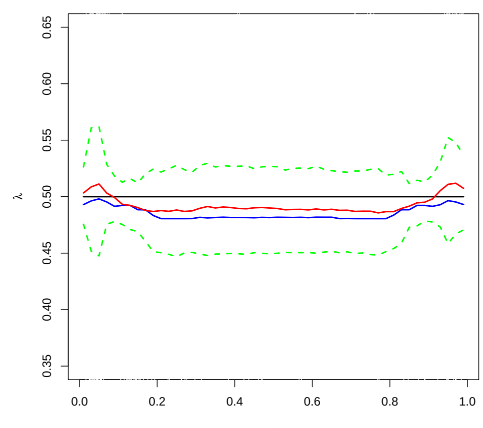
(a) Dependence parameter (Frank)



(b) Dependence parameter (Clayton)

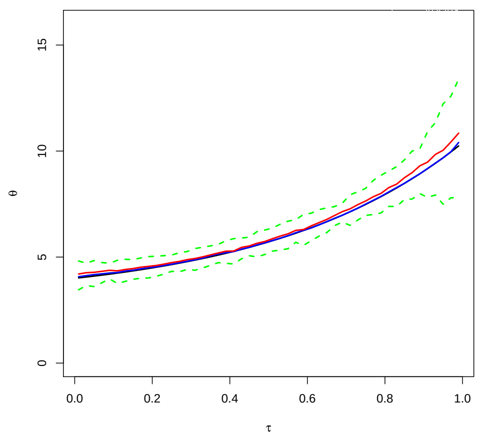


(c) Weight (Frank)

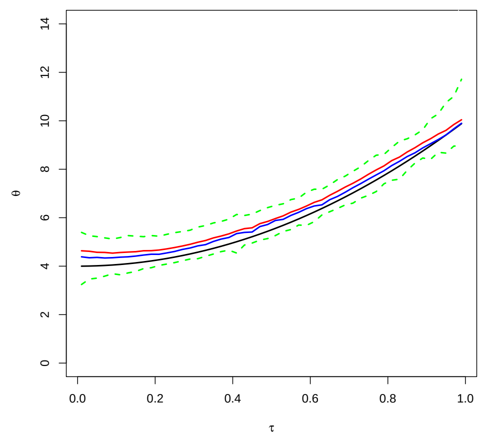


(d) Weight (Clayton)

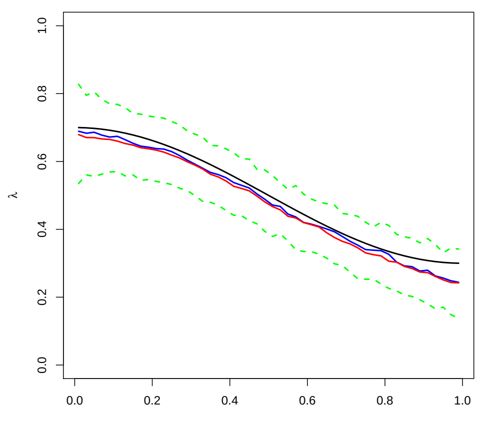
Figure 3: Simulation results of the weights and dependence parameters for Model 3 (500 repeats) in Case I simulations: true values (black solid lines), mean and median estimates (red and blue lines), and 5% and 95% percentile curves (green dashed lines). The sample size is 800.



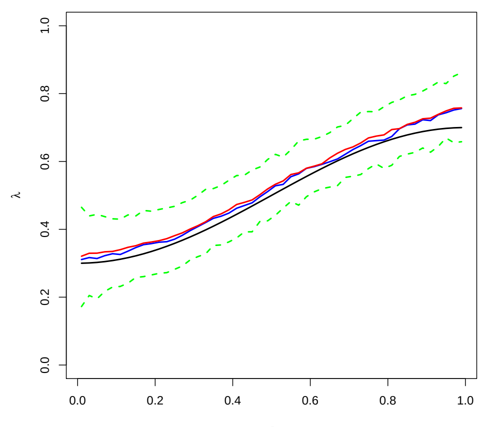
(a) Dependence parameter (Gumbel)



(b) Dependence parameter (Frank)

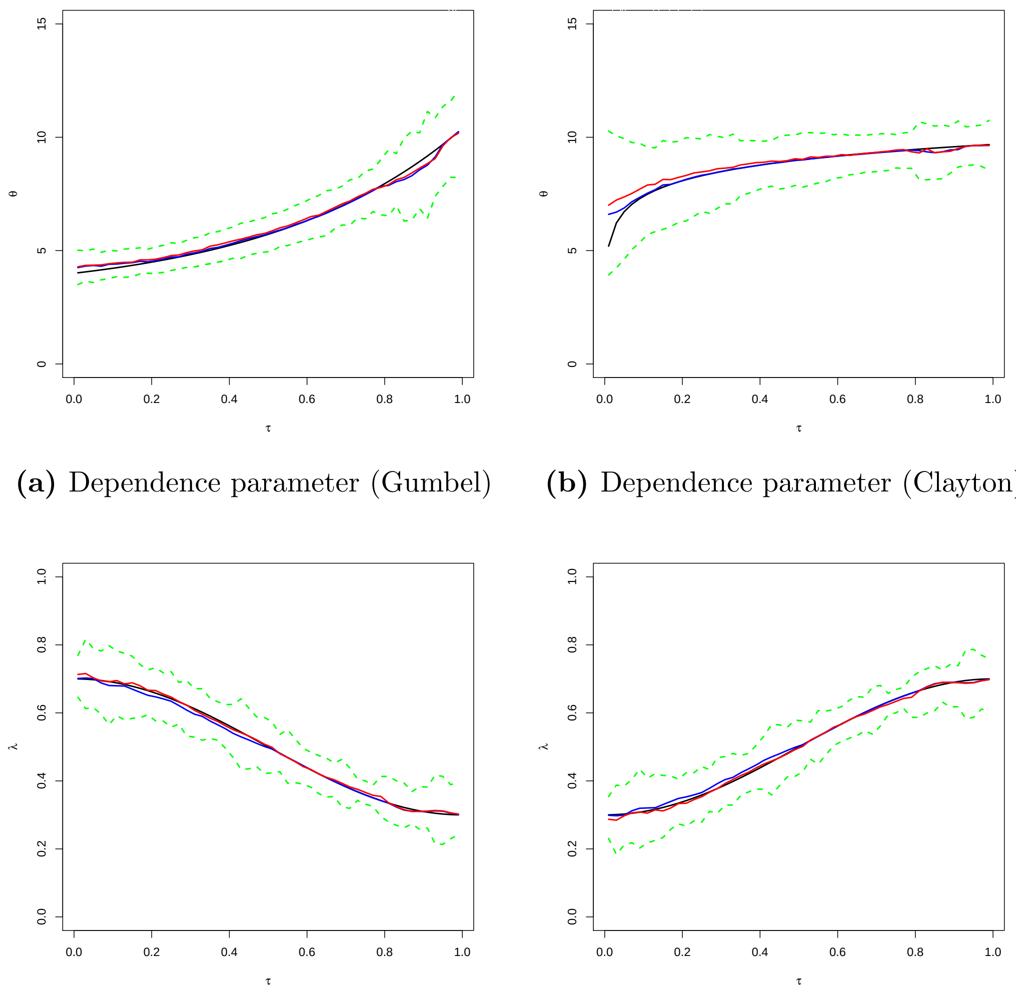


(c) Weight (Gumbel)



(d) Weight (Frank)

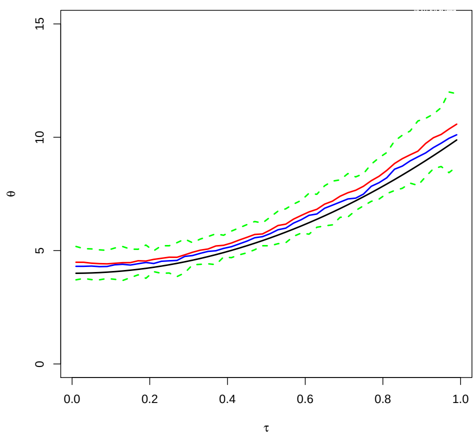
Figure 4: Simulation results of the weights and dependence parameters for Model 1 (500 repeats) in Case II simulations: true values (black solid lines), mean and median estimates (red and blue lines), and 5% and 95% percentile curves (green dashed lines). The sample size is 800.



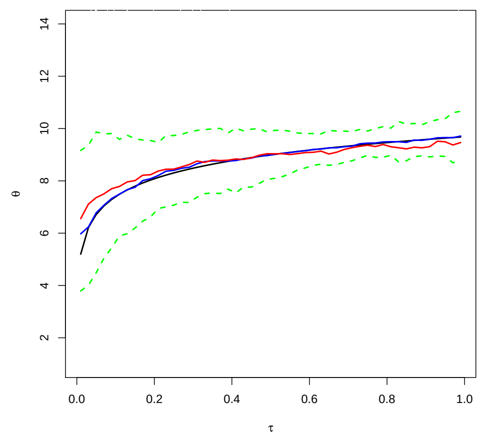
(a) Dependence parameter (Gumbel) **(b)** Dependence parameter (Clayton)

(c) Weight (Gumbel) **(d)** Weight (Clayton)

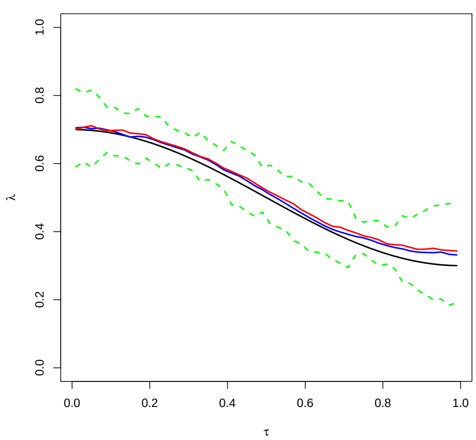
Figure 5: Simulation results of the weights and dependence parameters for Model 2 (500 repeats) in Case II simulations: true values (black solid lines), mean and median estimates (red and blue lines), and 5% and 95% percentile curves (green dashed lines). The sample size is 800.



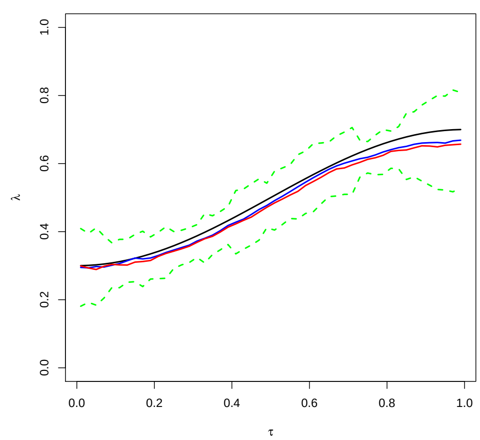
(a) Dependence parameter (Frank)



(b) Dependence parameter (Clayton)

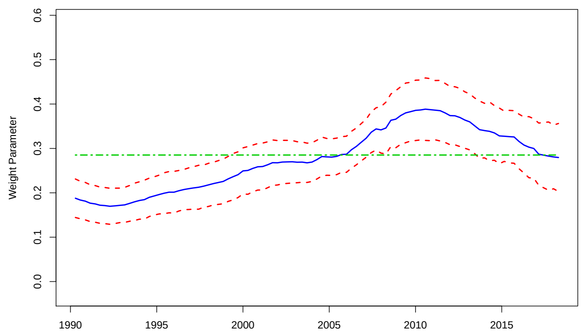


(c) Weight (Frank)

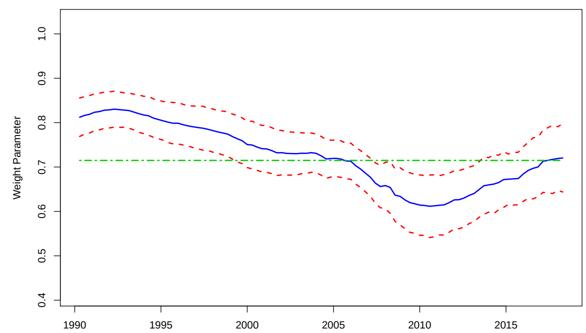


(d) Weight (Clayton)

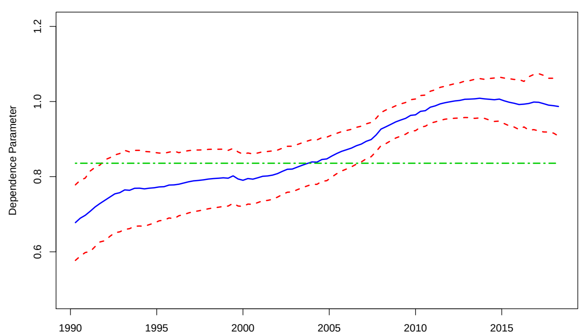
Figure 6: Simulation results of the weights and dependence parameters for Model 3 (500 repeats) in Case II simulations: true values (black solid lines), mean and median estimates (red and blue lines), and 5% and 95% percentile curves (green dashed lines). The sample size is 800.



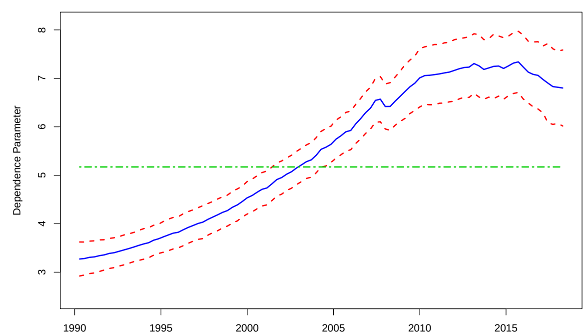
(a) Weight (Clayton)



(b) Weight (Frank)

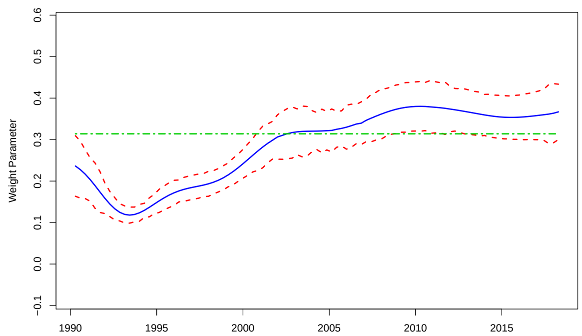


(c) Dependence parameter (Clayton)

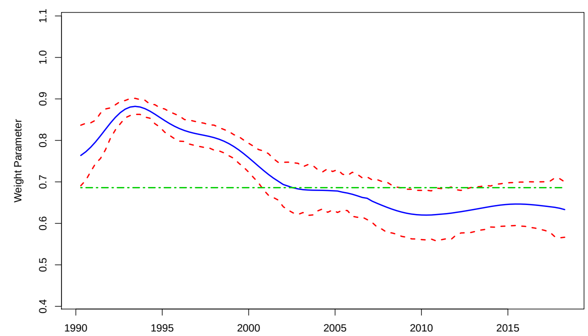


(d) Dependence parameter (Frank)

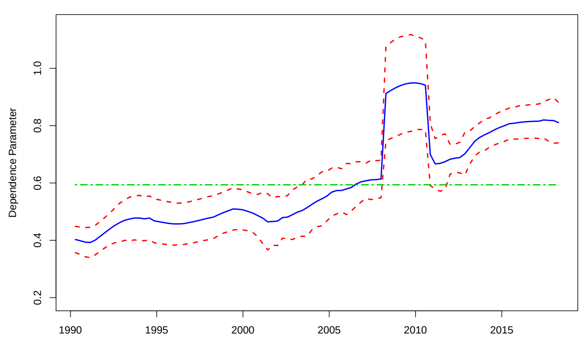
Figure 7: Estimation results of nonzero weights and dependence parameters (blue solid lines) along with the 90% confidence intervals (red dashed curves) for the US-UK pair. The green two-dashed lines (horizontal lines) show the estimates using the time-invariant mixture copula model.



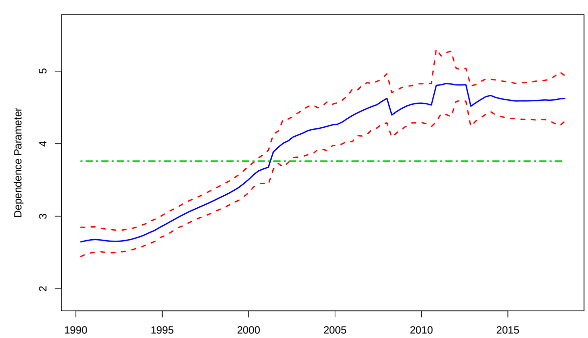
(a) Weight (Clayton)



(b) Weight (Frank)

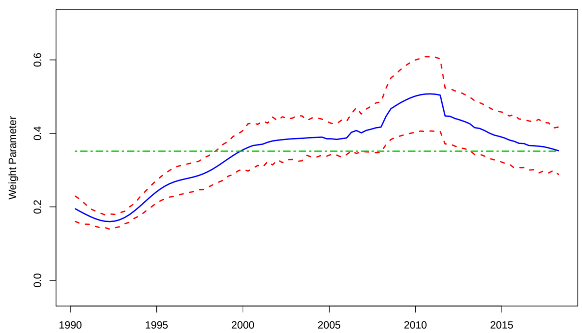


(c) Dependence parameter (Clayton)

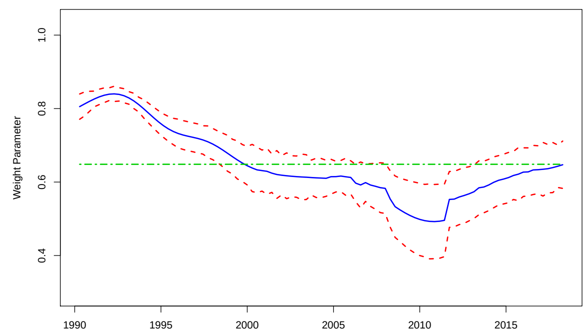


(d) Dependence parameter (Frank)

Figure 8: Estimation results of nonzero weights and dependence parameters (blue solid lines) along with the 90% confidence intervals (red dashed curves) for the US-HK pair. The green two-dashed lines (horizontal lines) show the estimates using the time-invariant mixture copula model.

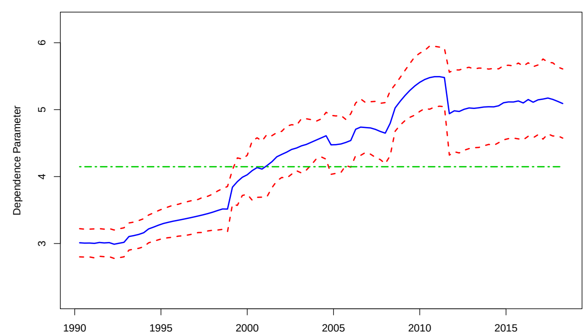


(a) Weight (Clayton)



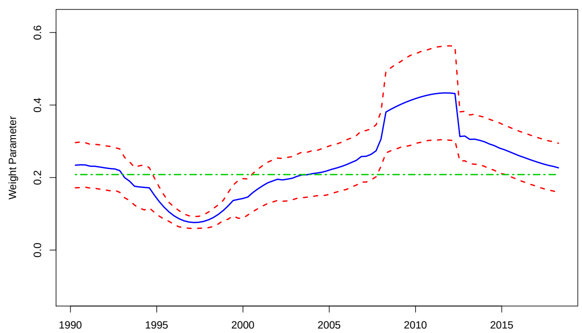
(b) Weight (Frank)

(c) Dependence parameter (Clayton)

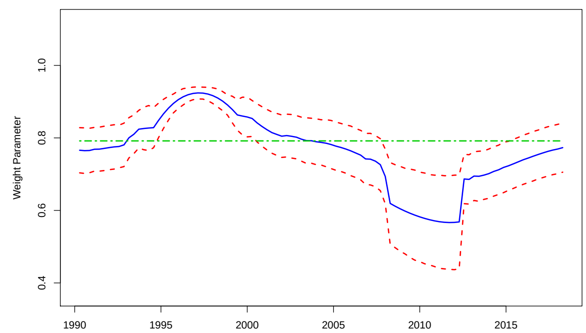


(d) Dependence parameter (Frank)

Figure 9: Estimation results of nonzero weights and dependence parameters (blue solid lines) along with the 90% confidence intervals (red dashed curves) for the UK-HK pair. The green two-dashed lines (horizontal lines) show the estimates using the time-invariant mixture copula model.

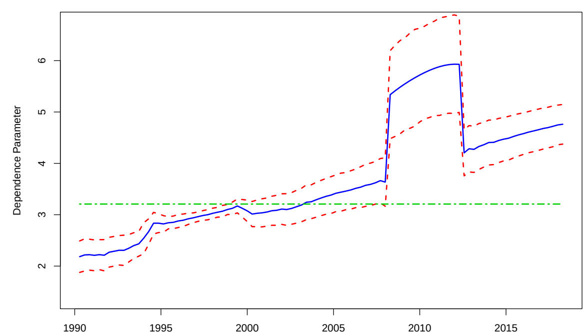


(a) Weight (Clayton)



(b) Weight (Frank)

(c) Dependence parameter (Clayton)



(d) Dependence parameter (Frank)

Figure 10: Estimation results of nonzero weights and dependence parameters (blue solid lines) along with the 90% confidence intervals (red dashed curves) for the HK-KR pair. The green two-dashed lines (horizontal lines) show the estimates using the time-invariant mixture copula model.

Appendix

A: The stationary bootstrap resampling scheme

Suppose that $\{x_{1,i}, x_{2,i}\}_{i=1}^{T}$ is a strictly stationary and weakly dependent time series. Let

$$B_{i,b} = \{(x_{1,i}, x_{2,i}), (x_{1,i+1}, x_{2,i+1}), \cdots, (x_{1,i+b-1}, x_{2,i+b-1})\}$$

be the block consisting of b observations starting from $(x_{1,i}, x_{2,i})$ to $(x_{1,i+b-1}, x_{2,i+b-1})$. In the case $j > T$, $(x_{1,j}, x_{2,j})$ is defined to be $(x_{1,i}, x_{2,i})$, where $i = j(modT)$ and $(x_{1,0}, x_{2,0}) = (x_{1,T}, x_{2,T})$. Let p be a constant such that $p \in [0, 1]$. Independent of $\{x_{1,i}, x_{2,i}\}_{i=1}^{T}$, let $L_1, L_2, \cdots$ be a sequence of i.i.d. random variables having the geometric distribution, i.e.,

$$P\{L_k = m\} = (1-p)^{m-1}p \quad m = 1, 2, \cdots.$$

where $p = T^{-1/3}$. Independent of both $\{x_{1,i}, x_{2,i}\}_{i=1}^{T}$ and L_k, let $I_1, I_2, \cdots$ be a sequence of i.i.d variables which have the discrete uniform distribution on $\{1, \cdots, T\}$.

A pseudo time series $\{x^*_{1,i}, x^*_{2,i}\}_{i=1}^{T}$ is generated in the following way. Sample a sequence of blocks of random length by the prescription $B_{I_1,L_1}, B_{I_2,L_2}, \cdots$, where I_k is generated from a uniform distribution on $\{1, \cdots, T\}$ and L_k is generated from the distribution as defined earlier. The first L_1 observations in the pseudo time series $\{x^*_{1,i}, x^*_{2,i}\}_{i=1}^{T}$ are determined by the first block B_{I_1,L_1} of observations $(x_{1,I_1}, x_{2,I_1}), \cdots, (x_{1,I_1+L_1-1}, x_{2,I_1+L_1-1})$, the next L_2 observations in the pseudo time series are the observations in the second sampled block B_{I_2,L_2}, namely $(x_{1,I_2}, x_{2,I_2}), \cdots, (x_{1,I_2+L_2-1}, x_{2,I_2+L_2-1})$. This process is not stopped until T observations in the pseudo time series have been generated.

By randomly varying the block length, Politis and Romano (1994) show that the pseudo time series $\{x^*_{1,i}, x^*_{2,i}\}_{i=1}^{T}$, conditional on the original data $\{x_{1,i}, x_{2,i}\}_{i=1}^{T}$, is actually stationary. Hence, this resampling method is applicable for stationary and weekly dependent time series.

B: Mathematical proofs

In this Appendix, we prove the main results of Section 2. Let C be a constant and R_m be a generic remainder term of small order, and they may take different values at different places.

Lemma 1 *Assume that the parametric estimators $\hat{\psi}$ and the local constant estimators $\hat{\delta}(\tau)$*

are obtained from the two-step procedure in Section 2 and satisfy $\|\hat{\psi}-\psi\| = O_p(1/\sqrt{T})$ *and* $\|\hat{\delta}(\tau)-\delta(\tau)\| = O_p(1/\sqrt{Th})$. *Define the local log likelihood function as*

$$L_h(\psi,\delta) = \frac{1}{T}\sum_{i=1}^{T} \ell_i(\psi,\delta(\tau))K_h(t_i-\tau)$$

where $\ell_i(\psi,\delta(\tau)) =$*log* $(\sum_{k=1}^{d} \lambda_k(\tau)c_k(F_1(X_{1i},\psi_1), F_2(X_{2i},\psi_2), \theta_k(\tau)))$. *Under conditions A1-A6, we have*

$$L_h(\hat{\psi},\hat{\delta}) - L_h(\psi,\delta) = L_h(\psi,\hat{\delta}) - L_h(\psi,\delta) + R_m.$$

Proof:

Let

$$L_h(\hat{\psi},\hat{\delta}) - L_h(\psi,\delta) = \underbrace{L_h(\hat{\psi},\hat{\delta}) - L_h(\psi,\hat{\delta})}_{I_1} + \underbrace{L_h(\psi,\hat{\delta}) - L_h(\psi,\delta)}_{I_2}$$

By Taylor's expansion and the conditions $\|\hat{\psi}-\psi\| = O_p(1/\sqrt{T})$ and $\|\hat{\delta}-\delta\| = O_p(1/\sqrt{Th})$, the first term on the right hand size can be written as

$$\begin{aligned} I_1 = \quad & L_h(\hat{\psi},\hat{\delta}(\hat{\psi})) - L_h(\psi,\hat{\delta}(\psi)) \\ = \quad & \left[\sqrt{T}\frac{\partial L_h(\psi,\hat{\delta}(\psi))}{\partial\psi}\right]\frac{1}{\sqrt{T}}(\hat{\psi}-\psi)\{1+o_p(1)\} \\ = \quad & \left[\sqrt{T}\frac{\partial L_h(\psi,\delta(\psi))}{\partial\psi}\{1+o_p(1)\}\right]\frac{1}{\sqrt{T}}(\hat{\psi}-\psi)\{1+o_p(1)\}. \end{aligned}$$

The term $\sqrt{T}\partial L_h(\psi,\delta(\psi))/\partial\psi$ is of order $O_p(1)$ since the first derivative for the marginal maximum likelihood $\sqrt{T}\partial L_c(\psi)/\partial\psi$ and the first derivative for the full maximum likelihood $\sqrt{T}\partial L_c(\psi)/\partial\psi + \sqrt{T}\partial L_h(\psi,\delta(\psi))/\partial\psi$ are of order $O_p(1)$. This implies that $L_h(\hat{\psi},\hat{\delta}) - L_h(\psi,\hat{\delta})$ is of order $O_p(1/T)$.

Further, by Taylor's expansion and the condition $\|\hat{\delta}-\delta\| = O_p(1/\sqrt{Th})$, the term

$$I_2 = \quad \left[\sqrt{Th}\frac{\partial L_h(\psi,\delta)}{\partial\delta}\right]\frac{1}{\sqrt{Th}}(\hat{\delta}-\delta)\{1+o_p(1)\}$$

which is of order $O_p(1/(Th))$, dominates I_1. This completes the proof.

$\square$

Lemma 1 suggests that we can derive the asymptotic distribution of $\hat{\delta}$ without considering the errors from parametric estimation. The parametric estimators $\hat{\psi}$ in the marginal parts have little effect on the estimation of $\hat{\delta}$ if the sample size T is large. This result is in line with the fact that the convergence rate of the parametric part of the model is faster than that of the nonparametric component.

Proof of Theorem 1 and Theorem 2:

Using Lemma 1 we can assume that ψ is known for simplicity. Let $\ell_i(\delta(\tau)) = \ell_i(\psi, \delta(\tau))$, $\mathcal{L}(\delta(\tau)) = \frac{1}{T}\sum_{i=1}^{T} \ell_i(\delta(\tau))K_h(t_i - \tau)$, $\mathcal{L}'(\delta(\tau)) = \frac{1}{T}\sum_{i=1}^{T} \ell_i'(\delta(\tau))K_h(t_i - \tau)$ and $\mathcal{L}''(\delta(\tau)) = \frac{1}{T}\sum_{i=1}^{T} \ell_i''(\delta(\tau))K_h(t_i - \tau)$. For a fixed point $\tau \in (0,1)$, the normal equation for the local likelihood-based estimator is given by

$$\mathcal{L}'(\hat{\delta}(\tau)) = 0.$$

By Taylor's expansion, it can be written as

$$\mathcal{L}'(\delta(\tau)) + \mathcal{L}''(\delta(\tau))(\hat{\delta}(\tau) - \delta(\tau)) + o_p(1/\sqrt{Th}) = 0,$$

which leads to

$$\hat{\delta}(\tau) - \delta(\tau) = -[\mathcal{L}''(\delta(\tau))]^{-1}\mathcal{L}'(\delta(\tau)) + o_p(1/\sqrt{Th}).$$

By the moment condition, we have

$$\begin{aligned} 0 &= E\{\ell_i'(\delta(t_i))|t_i = \tau\} \\ &= E\{\ell_i'(\delta(\tau) + r_i)|t_i = \tau\} \\ &= E\{\ell_i'(\delta(\tau))|t_i = \tau\} + r_i E\{\ell_i''(\delta(\tau))|t_i = \tau\} + o_p(r_i), \end{aligned}$$

where $r_i = \delta'(\tau)(t_i - \tau) + \frac{1}{2}\delta''(\tau)(t_i - \tau)^2 + o_p(t_i - \tau)^2$. By construction, we have $E\{\ell_i'(\delta(\tau))|t_i = \tau\} = -r_i E\{\ell_i''(\delta(\tau))|t_i = \tau\} + o_p(r_i)$. Thus

$$\begin{aligned} E\{\mathcal{L}'(\delta(\tau))|t_i = \tau\} &= -\frac{1}{T}\sum_{i=1}^{T} r_i E\{\ell_i''(\delta(\tau))|t_i = \tau\}K_h(t_i - \tau) \\ &= \frac{1}{T}\Sigma(\tau)\sum_{i=1}^{T} r_i K_h(t_i - \tau) \end{aligned}$$

where $\Sigma(\tau) = -E\{\ell_i''(\delta(\tau))|t_i = \tau\}$. Note that

$$
\begin{aligned}
E\{\mathcal{L}''(\delta(\tau))|t_i = \tau\} &= \frac{1}{T}\sum_{i=1}^{T} E\{\ell_i''(\delta(\tau))|t_i = \tau\} k_h(t_i - \tau) \\
&= \begin{cases} -\Sigma(\tau) + o_p(1), & \text{if } \tau \in (0,1); \\ -\mu_{0,b}\Sigma(0+) + o_p(1), & \text{if } \tau = bh. \end{cases}
\end{aligned}
$$

It follows by Taylor's expansion and the Riemann sum approximation of an integral that the bias term of $\hat{\delta}(\tau)$ can be expressed as

$$
\begin{aligned}
& E(\hat{\delta}(\tau)|t_i = \tau) - \delta(\tau) \\
&= -[E(\mathcal{L}''(\delta(\tau))|t_i = \tau)]^{-1} E(\mathcal{L}'(\delta(\tau))|t_i = \tau) \\
&= \frac{1}{T}\sum_{i=1}^{T} \left[\delta'(\tau)(t_i - \tau) + \frac{1}{2}\delta''(\tau)(t_i - \tau)^2\right] K_h(t_i - \tau) + R_m \\
&= \int \delta'(\tau)(t_i - \tau) K_h(t_i - \tau) dt_i + \frac{1}{2}\int \delta''(\tau)(t_i - \tau)^2 K_h(t_i - \tau) dt_i + R_m \\
&= \begin{cases} \frac{h^2}{2}\delta''(\tau)\mu_2 + o_p(h^2), & \text{if } \tau \in (0,1); \\ \frac{h}{\mu_{0,b}}\delta'(0+)\mu_{1,b} + o_p(h), & \text{if } \tau = bh. \end{cases}
\end{aligned}
$$

To find the expression for $Var\{\mathcal{L}'(\delta(\tau))|t_i = \tau\}$, we let $Q_T = \frac{1}{T}\sum_{i=1}^{T} Z_i$, where $Z_i = \ell_i'(\delta(\tau))K_h(t_i - \tau)$. Using the same argument as in Cai (2007) , we can show that

$$
Var\{\mathcal{L}'(\delta(\tau))|t_i = \tau\} = Var(Q_T) = \begin{cases} \frac{1}{Th}v_0\left(\Gamma_0(\tau) + 2\sum_{\ell=1}^{\infty}\Gamma_\ell(\tau)\right) + o_p(\frac{1}{Th}), & \text{if } \tau \in (0,1); \\ \frac{1}{Th}v_{0,b}\left(\Gamma_0(0+) + 2\sum_{\ell=1}^{\infty}\Gamma_\ell(0+)\right) + o_p(\frac{1}{Th}), & \text{if } \tau = bh, \end{cases}
$$

where $\Gamma_\ell(\tau) = E\{\ell_i'(\delta(\tau))\ell_{i+\ell}'(\delta(\tau))^\intercal|t_i = \tau\}$.
The variance term is given by

$$
\begin{aligned}
& Var\{\hat{\delta}(\tau)|t_i = \tau\} \\
&= E\{\mathcal{L}''(\delta(\tau))|t_i = \tau\}^{-1} Var\{\mathcal{L}'(\delta(\tau))|t_i = \tau\} E\{\mathcal{L}''(\delta(\tau))|t_i = \tau\}^{-1} \\
&= \begin{cases} \frac{1}{Th}v_0\Sigma(\tau)^{-1}\Omega(\tau)\Sigma(\tau)^{-1}, & \text{if } \tau \in (0,1); \\ \frac{1}{Th}\frac{v_{0,b}}{\mu_{0,b}^2}\Sigma(0+)^{-1}\Omega(0+)\Sigma(0+)^{-1}, & \text{if } \tau = bh. \end{cases}
\end{aligned}
$$

This completes the proof.

□

Proof of Theorem 3:

Let $v = (v_{jk}) \in R^{T\times(2d)}$ be an arbitrary $T \times (2d)$ matrix with rows $v_{j\cdot}$ and columns $v_{\cdot k}$, i.e. $v = (v_{1\cdot}, v_{2\cdot}, \cdots, v_{T\cdot})^T = (v_{\cdot 1}, v_{\cdot 2}, \cdots, v_{\cdot(2d)})$. Define $v_{\cdot k}^{\lambda} = v_{\cdot(d+k)}$ and set $\|v\| = \sqrt{\sum_{j,k} v_{j,k}^2}$ to be the L_2-norm for the matrix $v = (v_{jk})$. The true value δ_0 is defined in the same way as δ. For any small $\varepsilon > 0$, if we can show that there is a large constant C such that $P\{\inf_{T^{-1}\|v\|^2=C} Q^P(\delta_0 + (Th)^{-1/2}v) < Q^P(\delta_0)\} > 1-\varepsilon$, then the result is established. To this end, define

$$\begin{aligned}
D &\equiv \frac{h}{T}\left[Q^P(\delta_0 + (Th)^{-1/2}v) - Q^P(\delta_0)\right] \\
&= \frac{h}{T}\left[\sum_{j=1}^{T}\sum_{i=1}^{T} \ell_i(\delta_0(t_j) + (Th)^{-1/2}v_{j\cdot})K_h(t_i - t_j) - \sum_{j=1}^{T}\sum_{i=1}^{T} \ell_i(\delta_0(t_j))K_h(t_i - t_j)\right] \\
&\quad - h\sum_{k=1}^{d}\left[P_{\gamma_T}(\|\lambda_{0k} + (Th)^{-1/2}v_{\cdot k}^{\lambda}\|) - P_{\gamma_T}(\|\lambda_{0k}\|)\right] \\
&\doteq D_1 + D_2,
\end{aligned}$$

where

$$\begin{aligned}
D_1 &\equiv \frac{h}{T}\left[\sum_{j=1}^{T}\sum_{i=1}^{T} \ell_i(\delta_0(t_j) + (Th)^{-1/2}v_{j\cdot})K_h(t_i - t_j) - \sum_{j=1}^{T}\sum_{i=1}^{T} \ell_i(\delta_0(t_j))K_h(t_i - t_j)\right] \\
&= \frac{h}{T}\sum_{j=1}^{T}\sum_{i=1}^{T}\left[\ell_i(\delta_0(t_j) + (Th)^{-1/2}v_{j\cdot}) - \ell_i(\delta_0(t_j))\right]K_h(t_i - t_j) \\
&= \frac{h}{T}\sum_{j=1}^{T}\sum_{i=1}^{T}\left[(Th)^{-1/2}v_{j\cdot}^{\intercal}\ell_i'(\delta_0(t_j)) + (2Th)^{-1}v_{j\cdot}^{\intercal}\ell_i''(\delta_0(t_j))v_{j\cdot} + o_p(1/(Th))\right]K_h(t_i - t_j) \\
&= \frac{1}{T}\sum_{j=1}^{T} v_{j\cdot}^{\intercal}e_j - \frac{1}{2T}\sum_{j=1}^{T} v_{j\cdot}^{\intercal}\{\Sigma(\delta_0(t_j)) + o_p(1)\}v_{j\cdot} + o_p(1)
\end{aligned}$$

where $e_j = h^{1/2}T^{-1/2}\sum_{i=1}^{T} \ell_i'(\delta_0(t_j))K_h(t_i - t_j)$ and $\Sigma(\delta_0(\tau)) = -E(\ell_i''(\delta_0(\tau))|t_i = \tau)$.

By the Cauchy-Schwarz inequality,

$$\begin{aligned} D_1 \quad & \leq \frac{1}{T}\sum_{j=1}^{T}\|v_{j\cdot}\|\|e_j\| - \frac{1}{2T}\sum_{j=1}^{T}\lambda_{t_j}^{min}\|v_{j\cdot}\|^2 + o_p(1) \\ & \leq \sqrt{\|v\|^2/T}\sqrt{\|e\|^2/T} - \frac{\lambda_{min}}{2}\|v\|^2/T + o_p(1) \\ & = \sqrt{C}\sqrt{\|e\|^2/T} - \frac{C\lambda_{min}}{2} + o_p(1) \end{aligned}$$

where $\lambda_{t_j}^{min}$ is the smallest eigenvalue of $\Sigma(\delta_0(t_j))$ and λ_{min} is the minimal value of the sequence $\{\lambda_{t_j}^{min}\}_{j=1}^{T}$. By standard nonparametric arguments, we can show that $E(e_j|t_i = t_j)$ is of order $O_p(1)$. Along with the law of large numbers and equation (a.1), we have $E(e_j^2|t_i = t_j) = Var(e_j|t_i = t_j) + (E(e_j|t_i = t_j))^2 = \Omega(t_j)v_0 + (E(e_j|t_i = t_j))^2 + R_m$ which is of order $O_p(1)$ and $\|e\|^2/T = E(e_j^2) + R_m = E(E(e_j^2|t_i = t_j)) + R_m$ which is of order $O_p(1)$.

By the fact that $\|\lambda_{0k}\| = 0$ for $k = d_0 + 1, \cdots, d$, Taylor's expansion and the triangle inequality, we have

$$\begin{aligned} D_2 \quad & \equiv -h\sum_{k=1}^{d}\Big(P_{\gamma_T}(\|\lambda_{0k} + (Th)^{-1/2}v_{\cdot k}^{\lambda}\|) - P_{\gamma_T}(\|\lambda_{0k}\|)\Big) \\ & \leq -h\sum_{k=1}^{d_0}\Big(P_{\gamma_T}(\|\lambda_{0k} + (Th)^{-1/2}v_{\cdot k}^{\lambda}\|) - P_{\gamma_T}(\|\lambda_{0k}\|)\Big) \\ & \leq -h\sum_{k=1}^{d_0}P_{\gamma_T}'(\|\lambda_{0k}\|)\big(\|\lambda_{0k} + (Th)^{-1/2}v_{\cdot k}^{\lambda}\| - \|\lambda_{0k}\|\big) + R_m \\ & \leq h^{1/2}T^{-1/2}\sum_{k=1}^{d_0}P_{\gamma_T}'(\|\lambda_{0k}\|)\|v_{\cdot k}^{\lambda}\| + R_m \\ & \doteq D_{21}. \end{aligned}$$

Define $a_T = \max\{P_{\gamma_T}'(\|\lambda_{0k}\|) : \|\lambda_{0k}\| \neq 0\}$, by Cauchy-Schwarz inequality, we have

$$\begin{aligned} D_{21} \quad & \leq h^{1/2}a_T\sqrt{d_0}\left[T^{-1}\sum_{k=1}^{d_0}\|v_{\cdot k}^{\lambda}\|^2\right]^{1/2} + R_m \\ & \leq h^{1/2}a_T\sqrt{d_0}\left[T^{-1}\sum_{k=1}^{d}\|v_{\cdot k}\|^2\right]^{1/2} + R_m \\ & \leq h^{1/2}a_T\sqrt{d_0}\sqrt{C} + R_m. \end{aligned}$$

It follows by the Riemann sum approximation of an integral and the condition (A1) that $\|\lambda_{0k}\|^2/T = \int_0^1 \lambda_k^2(\tau)d\tau + o_p(1)$ is a bounded constant. By choosing the SCAD penalty function, as $T \to \infty$, the condition $T^{-1/2}\gamma_T \to 0$ in (B1) implies $\gamma_T < \|\lambda_{0k}\|$. Therefore, $a_T \to 0$ and $D_2 = 0$. By choosing a sufficient large C, the second term in D_1 dominates other terms, which implies $D < 0$. This completes the proof.

$\square$

Proof of Theorem 4:

(a) Firstly, we show the sparsity $\|\hat{\lambda}_{\cdot k}\| = 0$ for all $k = d_0 + 1, \cdots, d$. We assume that $\|\hat{\lambda}_{\cdot k}\| \neq 0$ and there exists a $\sqrt{Th}$-consistent penalized estimator $\hat{\delta}$ such that

$$\frac{\partial Q^P(\hat{\delta})}{\partial \lambda_{\cdot k}} = J_1 + J_2 = 0$$

where $J_1 = (J_{11}, \cdots, J_{1T})^T$ with $J_{1j} = \sum_{i=1}^T \frac{\partial \ell_i(\hat{\delta}(t_j))}{\partial \lambda_k(t_j)} k_h(t_i - t_j)$ and $J_2 = -TP'_{\gamma_T}(\|\hat{\lambda}_{\cdot k}\|)\frac{\hat{\lambda}_{\cdot k}}{\|\hat{\lambda}_{\cdot k}\|}$.

By the law of large numbers, we have

$$\|J_1\| = \sqrt{J_{11}^2 + J_{12}^2 + \cdots + J_{1T}^2} = \sqrt{T}\sqrt{EJ_{1j}^2(1 + o_p(1))}.$$

By the result $\|\hat{\delta}_{\gamma_T}(\tau) - \delta_0(\tau)\| = O_p(1/\sqrt{Th})$ in Theorem 3, similar to the proof for $E(e_j^2|t_i = t_j)$ which is of order $O_p(1)$, we can show $Var(J_{1j}|t_i = t_j) = O_p(T/h)$. Moreover, by Taylor's expansion, we have

$$\begin{aligned}
J_{1j} = \quad & \sum_{i=1}^T \frac{\partial \ell_i(\hat{\delta}(t_j))}{\partial \lambda_k(t_j)} K_h(t_i - t_j) \\
= \quad & \sum_{i=1}^T \frac{\partial \ell_i(\delta_0(t_j))}{\partial \lambda_k(t_j)} K_h(t_i - t_j) \\
& + \sum_{i=1}^T \left[\sum_{m=1}^{2d} \frac{\partial^2 \ell_i(\delta_0(t_j))}{\partial \lambda_k(t_j) \partial \delta_m(t_j)} (\hat{\delta}_m(t_j) - \delta_{0m}(t_j)) \right] K_h(t_i - t_j) + R_m \\
\doteq \quad & A_1 + A_2 + R_m.
\end{aligned}$$

By standard nonparametric arguments, we can show that both A_1 and A_2 are of order $O_p(\sqrt{T/h})$, which suggests that $(E(J_{1j}|t_i = t_j))^2$ is of order $O_p(T/h)$. It follows that $E(J_{1j}^2|t_i = t_j) = Var(J_{1j}|t_i = t_j) + (E(J_{1j}|t_i = t_j))^2$ is of order $O_p(T/h)$ and $\|J_1\| = \sqrt{T}\{EJ_{1j}^2(1 + o_p(1))\}^{1/2} = \sqrt{T}\{E(E(J_{1j}^2|t_i = t_j))(1 + o_p(1))\}^{1/2}$ is of order $O_p(Th^{-1/2})$. By the condition that $P'_{\gamma_T}(\|\hat{\lambda}_{\cdot k}\|)/\gamma_T > 0$ and $\sqrt{h}\gamma_T \to \infty$, we can conclude that

$$\|J_2\| = TP'_{\gamma_T}(\|\hat{\lambda}_{\cdot k}\|) = \frac{P'_{\gamma_T}(\|\hat{\lambda}_{\cdot k}\|)}{\gamma_T}(\sqrt{h}\gamma_T)(Th^{-1/2})$$

dominates $\|J_1\|$ as $T \to \infty$. Therefore, the assumption $\|\hat{\lambda}_{\cdot k}\| \neq 0$ does not hold and we conclude $\|\hat{\lambda}_{\cdot k}\| = 0$.

(b) Secondly, we show the asymptotic normality.

By the sparsity $\|\hat{\lambda}_{\cdot k}\| = 0$ for all $k = d_0 + 1, \cdots, d$, we rewrite equation (3) as

$$Q^P(\delta_a) = \sum_{j=1}^{T}\sum_{i=1}^{T} \ell_i(\delta_a(t_j))K_h(t_i - t_j) - T\sum_{k=1}^{d_0} P_{\gamma_k}(\|\lambda_{\cdot k}\|),$$

where δ_a is a $T \times (2d_0)$ matrix as $\delta_a = (\delta_a(t_1), \cdots, \delta_a(t_T))^\intercal = (\theta_{\cdot 1}, \cdots, \theta_{\cdot d_0}, \lambda_{\cdot 1}, \cdots, \lambda_{\cdot d_0})$.

Taking the first derivative of the above equation with respect to $\delta_a(t_j)$, we have

$$\sum_{i=1}^{T} \ell'_i(\hat{\delta}_{a,\gamma_T}(t_j))K_h(t_i - t_j) - T\boldsymbol{P}' = 0.$$

where $\boldsymbol{P}' = \big(0, \cdots, 0, P'_{\gamma_1}(\|\hat{\lambda}_{\cdot 1}\|)\frac{\hat{\lambda}_1(t_j)}{\|\hat{\lambda}_{\cdot 1}\|}, \cdots, P'_{\gamma_{d_0}}(\|\hat{\lambda}_{\cdot d_0}\|)\frac{\hat{\lambda}_{d_0}(t_j)}{\|\hat{\lambda}_{\cdot d_0}\|}\big)$.

By the result $\|\hat{\delta}_{\gamma_T}(\tau) - \delta_0(\tau)\| = O_p(1/\sqrt{Th})$ in Theorem 3, we have $\|\hat{\lambda}_k(\tau) - \lambda_{0k}(\tau)\| = O_p(1/\sqrt{Th})$ which leads to $\|\hat{\lambda}_k\|^2/T = \|\lambda_{0k}\|^2/T + o_p(1) = \int_0^1 \lambda_k^2(\tau)d\tau + o_p(1)$. By choosing the SCAD penalty function, as $T \to \infty$, the condition $T^{-1/2}\gamma_T \to 0$ in (B1) implies $\gamma_T < \|\hat{\lambda}_{0k}\|$. Therefore, as $T \to \infty$, $a_T \to 0$ and $\boldsymbol{P}' = 0$. The asymptotic normality results directly follows Theorem 1.

This completes the proof.

$\square$