Statistica Sinica **32** (2022), 1049-1077
doi:https://doi.org/10.5705/ss.202020.0005

TIME-VARYING MIXTURE COPULA MODELS WITH COPULA SELECTION

Bingduo Yang, Zongwu Cai, Christian M. Hafner and Guannan Liu

Guangdong University of Finance and Economics, University of Kansas, Université catholique de Louvain and Xiamen University

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

Key words and phrases: Copula selection, EM algorithm, mixture copula, SCAD, time-varying distribution.

1. Introduction

Copulas have received considerable attention because they offer great flexibility when modeling multivariate distributions and when characterizing 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 (1959) theorem, this is always possible. The variety of dependence patterns is important for financial and macroeconomic time series, leading to applications such as volatility clustering (Ning, Xu and Wirjanto (2015)), real-time density forecasting (Smith and Vahey (2016)), stock returns modeling under nonstationarity (Wollschläger and Schäfer (2016)), systemic risk (Mensi et al. (2017)), and so on.

Among the applications of copula models, studying the co-movements of

Corresponding author: Guannan Liu, School of Economics and WISE, Xiamen University, Xiamen, China. E-mail: gliuecon@gmail.com.

returns across international equity markets is one of the most popular topics. For example, Hu (2006), Cai and Wang (2014), and Liu et al. (2019) employ time-invariant mixture copula models and find that international equity markets usually show lower-tail dependence, which implies that the markets are more likely to crash together than they are to boom together. However, the dependence structures between international equity markets are likely to change substantially over time as international financial markets change from time to time. In this case, a time-invariant copula model is incapable of capturing the evolution of the dependence structures. Therefore, a time-varying copula model is needed to solve this problem.

In the literature, time-varying or dynamic copulas have been used extensively 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 the asymmetric dependence between mark–dollar and yen–dollar exchange rates. Hafner and Reznikova (2010) propose a semiparametric dynamic copula (SDC) model, in which the copula parameter changes over time in a nonparametric way. Other dynamic 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)), variational mode decomposition methods (Mensi et al. (2017)), single-index copula models (Fermanian and Lopez (2018)), and semiparametric copula models under nonstationarity (Nasri, Rémillard and Bouezmarni (2019)), among others. For a comprehensive survey of dynamic copulas and their applications in financial time series analysis, see the survey paper by Patton (2012a).

Although many time-varying copula models have been proposed in the literature, most rely on a single copula instead of a mixture copula. Thus, although the copula dependence parameters that reflect the levels of dependence can change over time, the copula function that represents the pattern of dependence is still time-invariant. However, international equity markets may exhibit different dependence patterns during different periods, e.g., in tranquil periods and in crisis periods, and hence a single copula model is inadequate. Therefore, we contribute to the literature by proposing a time-varying mixture copula model with both time-varying dependence parameters and weights (or coefficients of the component copulas in a mixture copula model) to analyze co-movements across international equity markets. Indeed, the weights in a mixture copula model summarize the dependent patterns or shapes, and their magnitudes signify the importance of the corresponding copulas. By allowing the weights to be dynamic, we can recognize different dependence patterns during various periods. Furthermore, we do not specify a parametric form for the weights and dependence parameters in the time-varying mixture copula model, using a data-driven method for their specification. In this way, we alleviate typical misspecification problems in copulas. The proposed model can be considered 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 nonzero tail dependence. The proposed model differs from those of prior studies, which focus exclusively on single copula models (Hafner and Reznikova (2010); Acar, Craiu and Yao (2011)). It also differs from previous mixture copula models that assume that either the weights or the dependence parameters do not change over time (Garcia and Tsafack (2011); Liu et al. (2019)). Finally, it generalizes the time-varying optimal copula model of Liu, Ji and Fan (2017), which assumes a single copula at each time point.

An important issue is that the ranges of both the copula dependence parameters and their corresponding weights are restricted, for example, $\theta \in (-1, 1)$ for a Gaussian copula, $\theta \in (0, \infty)$ for a Clayton copula, and the weights satisfy $\lambda_k \in [0, 1]$ and $\sum_k \lambda_k = 1$. To overcome this difficulty in the nonparametric estimation, Abegaz, Gijbels and Veraverbeke (2012) and Acar, Craiu and Yao (2011) use 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, both the bias and the variance depend on the choice of the inverse link function, see, for example, Theorem 2 of Abegaz, Gijbels and Veraverbeke (2012) and Corollary 1 of Acar, Craiu and Yao (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 an 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 a penalized time-varying mixture copula model with a group smoothly clipped absolute deviation (SCAD) penalty term, as in Fan and Li (2001), to perform the estimation and copula selection simultaneously. The functional norms of the weight functions are penalized so that we can shrink them to zero if the contributions of the corresponding copulas are small. To facilitate the estimation, we propose a semiparametric version of the expectation maximization (EM) algorithm to estimate the weights and dependence parameters in the penalized local copula log-likelihood function. Furthermore, we discuss other important practical issues, including the bandwidth and tuning parameter selection and confidence intervals. In a simulation study, we consider mixture copulas with constant and time-varying weights and dependence parameters. The results show that the proposed method can correctly select the appropriate copulas and accurately estimate the unknown parameters in both cases.

In the application, we employ the proposed model and method to investigate the evolution of the dependence structures between four international stock markets (the United States, the United Kingdom, Hong Kong, and South Korea), using 28 years of weekly returns on the main equity indices. Interestingly, according to our results, all pairs of markets show lower-tail dependence, but no upper-tail dependence, because 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. Detailed results for this empirical example can be found in Section 4.

The remainder of the paper is organized as follows. Section 2 introduces the proposed time-varying mixture copula models. In the same section, we introduce penalized time-varying mixture copula models. Three practical issues are discussed, including a semiparametric EM algorithm, the bandwidth and tuning parameter selection, and the construction of pointwise confidence intervals using the bootstrap method. Section 3 reports the Monte Carlo simulation results. Section 4 applies the model and method to examine the evolution of the dependence between four international stock markets. The final section provides some concluding comments. The mathematical proofs are gathered in the online Supplementary Material.

2. Time-Varying Mixture Copula Models

2.1. Model setup

In this section, we model the time-varying mixture copula in a semiparametric way so that the dynamics in both the weights and the dependence parameters are captured simultaneously. Let $\{\mathbf{X}_i\}_{i=1}^T = \{X_{1i}, X_{2i}, \ldots, X_{Ni}\}_{i=1}^T$ be an N-dimensional time series sequence, and $\mathbf{Z}_i$ be a vector of predetermined or exogenous variables. Denote by $\mathcal{F}_{i-1}$ the σ-field generated by $\{\mathbf{X}_{i-1}, \mathbf{X}_{i-2}, \ldots; \mathbf{Z}_i, \mathbf{Z}_{i-1}, \ldots\}$. We assume that $\mathbf{X}_i$ follows

$$\mathbf{X}_i = \mu_i(\psi_{01}) + \Sigma_i(\psi_{02})^{1/2}\epsilon_i, \quad i = 1, \ldots, T, \tag{2.1}$$

where $\mu_i(\psi_{01}) = E\{\mathbf{X}_i|\mathcal{F}_{i-1}\}$ with ψ_{01} being the parameters for the conditional mean, $\Sigma_i(\psi_{02}) = \text{diag}(\sigma_{1i}^2(\psi_{02}), \ldots, \sigma_{Ni}^2(\psi_{02}))$ with ψ_{02} being the parameters for the conditional variance, and $\sigma_{si}^2(\psi_{02}) = E\{(X_{si} - \mu_{si}(\psi_{01}))^2|\mathcal{F}_{i-1}\}$, for $s = 1, \ldots, N$. Furthermore, ψ_{01} and ψ_{02} are unknown parameters with fixed dimensions.

We further assume that the standardized innovations $\{\epsilon_i \equiv (\epsilon_{1i}, \ldots, \epsilon_{Ni})^\intercal\}$ are independent of $\mathcal{F}_{i-1}$. For each $s \in \{1, \ldots, N\}$, $\{\epsilon_{si}\}_{i=1}^T$ have a zero mean and one unit standard deviation. The settings specified here cover many commonly used specifications, such as ARCH, GARCH, vector autoregressions (VAR), and so on (see Chen and Fan (2006) for a detailed discussion).

For example, for $s = 1, \ldots, N$, X_{si} may follow an AR-GARCH model with exogenous variables, as follows:

$$\begin{aligned} X_{si} &= \varphi_{s0} + \varphi_{s1}X_{s,i-1} + \cdots + \varphi_{sp}X_{s,i-p} + \phi_s Z_{si} + \sigma_{si}\epsilon_{si}, \\ \sigma_{si}^2 &= \alpha_{s0} + \alpha_{s1}(X_{s,i-1} - \varphi_{s0} - \varphi_{s1}X_{s,i-2} - \cdots - \varphi_{sp}X_{s,i-p-1} - \phi_s Z_{s,i-1})^2 \\ &\quad + \beta_{s1}\sigma_{s,i-1}^2, \end{aligned}$$

where the parameters satisfy $\alpha_{s0} > 0, \alpha_{s1} > 0, \beta_{s1} > 0$, and $(\alpha_{s1} + \beta_{s1}) < 1$. More examples of the specifications of the conditional mean and conditional variance in (2.1) can be found in Chen and Fan (2006). Note that, for simplicity, we consider only AR models (without using exogenous variables $\mathbf{Z}_i$) for the conditional mean in the simulation and application sections.

The goal of this study is to estimate the joint distribution of $\{\epsilon_i \equiv (\epsilon_{1i}, \ldots, \epsilon_{Ni})^\intercal\}$ based on a time-varying mixture copula model. Theoretically, the time-varying mixture copula model can be written as a linear combination of infinite single copula terms, as follows:

$$C(F_1(y_1), \ldots, F_N(y_N); \delta(t_i)) = \sum_{k=1}^{\infty} \lambda_k(t_i) C_k(F_1(y_1), \ldots, F_N(y_N); \theta_k(t_i)),$$

where $\{C_k(\cdot;\cdot)\}_{k=1}^\infty$ is a set of candidate copulas, $y_s, s = 1, \ldots, N$, denote the realizations of innovations, and $F_s(\cdot)$ are the marginal distribution functions. We rescale time t_i as $t_i = i/T$ to provide the asymptotic justification for the nonparametric smoothing estimators. The underlying assumption is that there is an increasingly intense sampling of data points that can be used to derive a consistent estimation, see, for example, Robinson (1989) and Cai (2007). Here $\{C_k(\cdot;\cdot)\}_{k=1}^\infty$ can be regarded as known basis copula functions, so that $C(F_1(y_1), \ldots, F_N(y_N);$

$\delta(t_i))$ can be regarded as a series expansion based on the basis copula functions $\{C_k(\cdot;\cdot)\}_{k=1}^{\infty}$. In real applications, we use a finite number of d single copulas to approximate the true one

$$C(F_1(y_1),\ldots,F_N(y_N);\delta(t_i)) = \sum_{k=1}^{d} \lambda_k(t_i) C_k(F_1(y_1),\ldots,F_N(y_N);\theta_k(t_i)), \quad (2.2)$$

where $\{C_1(\cdot;\cdot),\ldots,C_d(\cdot;\cdot)\}$ is a set of candidate copulas. 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 $\theta(t_i) = (\theta_1(t_i)^{\intercal},\ldots,\theta_d(t_i)^{\intercal})^{\intercal}$ and d-dimensional weights $\lambda(t_i) = (\lambda_1(t_i),\ldots,\lambda_d(t_i))^{\intercal}$. 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. The above mixture copula model implies that the joint cumulative distribution function of a random vector $(\epsilon_{1i},\ldots,\epsilon_{Ni})$ is given by a linear combination of $C_k(F_1(\cdot),\ldots,F_N(\cdot);\theta_k(t_i))$ with time-varying weights $\lambda_k(t_i)$.

When using (2.2) to approximate the true model, we may have a misspecification problem because some true single copulas might be excluded. To avoid this problem, we can first consider a large candidate copula set, and then employ the copula model selection procedures discussed in Section 2.2 to filter out the "insignificant" component copulas. Furthermore, even if some true single copulas are not included in this approximation so that the model is misspecified, we can still estimate and select the closest mixture copula model using the model selection criterion described in Section 2.2 (see Cai and Wang (2014)). Therefore, the model in (2.2) is flexible enough to capture a true copula in real applications.

Remark 1. In models (2.1)–(2.2), one can allow ψ_{01}, ψ_{02}, and $F_1(\cdot),\ldots,F_N(\cdot)$ to depend on time. However, for simplicity, here we assume that the marginals in (2.2) and ψ_{01} and ψ_{02} in (2.1) do not depend on time, because our main focus is on time-varying weights and dependence parameters in a mixture copula and its copula selection. It would be an interesting research topic to investigate the case with time-varying ψ_{01}, ψ_{02}, and marginals $F_1(\cdot),\ldots,F_N(\cdot)$.

Next, to discuss the identification issue of the model in (2.2), similarly to the time-invariant model proposed in Cai and Wang (2014) (see Definition 1 in Cai and Wang (2014)), we let $\mathbf{u} = (u_1,\ldots,u_N)^{\intercal}$ with $u_s = F_s(y_s)$, for $s = 1,\ldots,N$. Then we define two time-varying mixture copulas $C(\mathbf{u};\delta(t_i)) = \sum_{k=1}^{d} \lambda_k(t_i) C_k(\mathbf{u};\theta_k(t_i))$ and $C^*(\mathbf{u};\delta^*(t_i)) = \sum_{k=1}^{d^*} \lambda_k^*(t_i) C_k^*(\mathbf{u};\theta_k^*(t_i))$ as identified, that is, $C(\mathbf{u};\delta(t_i)) \equiv C^*(\mathbf{u};\delta^*(t_i))$, if and only if $d = d^*$ and we can order the summations such that $\lambda_k(t_i) = \lambda_k^*(t_i)$ and $C_k(\mathbf{u}; \theta_k(t_i)) = C_k^*(\mathbf{u}; \theta_k^*(t_i))$ for all possible values of $\mathbf{u}$, $k = 1, \ldots, d$ and $t_i \in [0, 1]$. Without loss of generality, we assume that the time-varying mixture copula model under investigation is identified.

In the following, we propose a three-step estimation procedure.

Step 1 : We estimate ψ_{01} and ψ_{02} in model (2.1) by specifying conditional mean models for $\mu(\psi_{01})$, such as the ARMA model, and conditional volatility models for $\Sigma(\psi_{02})$, such as the GARCH model. For further information on the various conditional mean models and conditional volatility models, as well as their corresponding estimation methods, see Chapters 2 and 3 in Tsay (2010).

Step 2 : After obtaining the estimates $\hat{\psi}_1$ and $\hat{\psi}_2$ from the first step, we calculate

$$\hat{\epsilon}_i = \Sigma_i^{-1/2}(\hat{\psi}_2)(\mathbf{X}_i - \mu_i(\hat{\psi}_1)).$$

Then the marginal distribution functions can be estimated using the rescaled empirical distribution of the residuals

$$\hat{F}_s(y_s) = \frac{1}{T+1} \sum_{j=1}^{T} I(\hat{\epsilon}_{sj} \leq y_s), \quad s = 1, \ldots, N,$$

with $I(\cdot)$ being an indicator function. For $i = 1, \ldots, T$, denote $\hat{u}_i = (\hat{u}_{1i}, \ldots, \hat{u}_{Ni})^\intercal$, with $\hat{u}_{si} = \hat{F}_s(\hat{\epsilon}_{si})$.

Step 3 : Given the estimators $\hat{u}_i$ from the second step, we employ the profile likelihood method to calculate the nonparametric estimator $\hat{\delta}(\tau) = (\hat{\theta}(\tau)^\intercal, \hat{\lambda}(\tau)^\intercal)^\intercal$ at a given point $\tau \in (0, 1)$ by maximizing the local copula log-likelihood function

$$\hat{\delta}(\tau) = \underset{(\theta(\tau), \lambda(\tau))}{\operatorname{argmax}} \sum_{i=1}^{T} \log \left(\sum_{k=1}^{d} \lambda_k(\tau) c_k(\hat{u}_i, \theta_k(\tau)) \right) K_h(t_i - \tau), \tag{2.3}$$

where $c_k(\cdot)$ is the density function of copula $C_k(\cdot)$, $K_h(\cdot) = K(\cdot/h)/h$, with $K(\cdot)$ being a kernel function and h a bandwidth that tunes the smoothness of the kernel estimator. In our simulation and empirical study, we use the commonly adopted Epanechnikov kernel function $K(z) = 3/4(1-z^2)I(|z| \leq 1)$.

Remark 2. Lemma A.1 in Chen and Fan (2006) shows that $\sup_{y_s} |\hat{F}_s(y_s) - F_s(y_s)| = O_p(T^{-1/2})$ holds for the rescaled empirical distribution function. This implies that the estimator $\hat{F}_s(\cdot)$ has a $\sqrt{T}$ convergence rate, such that it has little effect on the nonparametric estimators $\hat{\delta}(\tau)$ in large samples.

In the following, we present the copula selection procedure and its asymptotic properties. The regularity conditions and asymptotic properties for unpenalized estimators are given in the online Supplementary Material.

2.2. Penalized time-varying mixture copula models

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

$$Q^P(\delta) = \sum_{j=1}^{T}\sum_{i=1}^{T} \ell(\hat{u}_i, \delta(t_j)) K_h(t_i - t_j) - T\sum_{k=1}^{d} P_{\gamma_k}(\|\lambda_{\cdot k}\|) + \sum_{j=1}^{T} \rho_{t_j}\left(1 - \sum_{k=1}^{d} \lambda_k\left(t_j\right)\right), \tag{2.4}$$

where $\ell(\hat{u}_i, \delta(t_j)) = \log(\sum_{k=1}^{d} \lambda_k(t_j) c_k(\hat{u}_i, \theta_k(t_j)))$, $\hat{u}_i = (\hat{u}_{1i}, \ldots, \hat{u}_{Ni})^{\intercal}$ are obtained from Steps 1 and 2 as in Section 2.1, $P_{\gamma_k}(\cdot)$ is a penalty function with tuning parameter γ_k, $\|\lambda_{\cdot k}\| = (\lambda_k^2(t_1) + \cdots + \lambda_k^2(t_T))^{1/2}$, and ρ_{t_j} is a Lagrange multiplier for the constraint $\sum_{k=1}^{d} \lambda_k\left(t_j\right) = 1$. The norm of $\lambda_{\cdot k}$, that is, $\|\lambda_{\cdot k}\|$, 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 $\theta_k(\cdot)$ because 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 in recent decades. As pointed out by Fan and Li (2001), a good penalty function should satisfy the following three properties: unbiasedness for the nonzero coefficients, sparsity, and continuity of the resulting estimators to avoid instability in model prediction. Here, we use the SCAD penalty function proposed by Fan and Li (2001), which enjoys all three properties, although many other penalty functions are applicable, including the classical LASSO of Tibshirani (1996) and the adaptive LASSO of Zou (2006). The first-order derivative $P'_{\gamma_k}(z)$ of the continuous SCAD penalty function $P_{\gamma_k}(z)$

is given by

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

for some $\varrho > 2$, where $(\varrho\gamma_k - z)_+ = \max(\varrho\gamma_k - z, 0)$. For simplicity of presentation, we assume that the tuning parameters γ_k are the same for all $k = 1, \ldots, d$ 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), who state that this choice provides 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 $\lambda_{0a}(\tau) = [\lambda_{01}(\tau), \ldots, \lambda_{0d_0}(\tau)]^{\intercal}$ with $\|\lambda_{\cdot 0k}\| \neq 0$ for $1 \le k \le d_0$ and $\lambda_{0b}(\tau) = [\lambda_{0(d_0+1)}(\tau), \ldots, \lambda_{0d}(\tau)]^{\intercal}$ with $\|\lambda_{\cdot 0k}\| = 0$ for $d_0 + 1 \le k \le d$. Similarly, we let $\theta_0(\tau) = [\theta_{0a}^{\intercal}(\tau), \theta_{0b}^{\intercal}(\tau)]^{\intercal}$ with $\theta_{0a}(\tau) = [\theta_{01}(\tau), \ldots, \theta_{0d_0}(\tau)]^{\intercal}$ and $\theta_{0b}(\tau) = [\theta_{0(d_0+1)}(\tau), \ldots, \theta_{0d}(\tau)]^{\intercal}$, in which $\theta_{0b}(\tau)$ can be arbitrary because 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. Furthermore, we include the following additional technical conditions:

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

The condition $\lim_{T\to\infty} \inf_{z\to 0^+} P'_{\gamma_T}(z)/\gamma_T > 0$ can be found in Lemma 1 of Fan and Li (2001). The last condition in B1 implies that the order of the tuning parameter γ_T needs to be smaller than $T^{1/10}$, which is crucial for the consistency result in Theorem 1 and the oracle property in Theorem 2.

Theorem 1. *Let $\{X_{1i}, \ldots, X_{Ni}\}_{i=1}^{T}$ be a strictly stationary α-mixing sequence following the proposed models* (2.1)–(2.2). *For a fixed point $\tau \in (0, 1)$, under Conditions* A1–A6 *in the online Supplementary Material and* B1, *there exists a $\sqrt{Th}$-consistent estimator $\hat{\delta}_{\gamma_T}(\tau)$ that maximizes* (2.4) *satisfying $\|\hat{\delta}_{\gamma_T}(\tau) - \delta_0(\tau)\| = O_p(1/\sqrt{Th})$.*

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

Theorem 2. *(Oracle Property). Let $\{X_{1i}, \ldots, X_{Ni}\}_{i=1}^{T}$ be a strictly stationary α-mixing sequence following the proposed models* (2.1)–(2.2) *. For a fixed point*

$\tau \in (0,1)$, *under Conditions* A1–A6 *in the online Supplementary Material and* B1, *we have*

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

(b) *Asymptotic normality:*

$$\sqrt{Th}(\hat{\delta}_{a,\gamma_T}(\tau) - \delta_{0a}(\tau) - h^2 B_a(\tau)) \to N(0, v_0\Sigma_a(\tau)^{-1}\Omega_a(\tau)\Sigma_a(\tau)^{-1}),$$

where $v_0 = \int K^2(z)dz$, $\Sigma_a(\tau) = -E\{\ell''(u_i,\delta_{0a}(\tau))|t_i = \tau\}$, $\Omega_a(\tau) = \sum_{s=-\infty}^{\infty} \Gamma_{a,s}(\tau)$ *with* $\Gamma_{a,s}(\tau) = E\{\ell'(u_i,\delta_{0a}(\tau))\ell'(u_{i+s},\delta_{0a}(\tau))^{\intercal}|t_i = \tau\}$, *and the bias term* $h^2B_a(\tau) = (h^2/2)\delta''_{0a}(\tau)\mu_2$ *with* $\mu_2 = \int z^2K(z)dz$. *Here,* $\ell'(u_i,\delta_{0a}(\tau))$ *and* $\ell''(u_i,\delta_{0a}(\tau))$ *denote the first and second derivatives, respectively, of* $\ell(u_i,\delta_{0a}(\tau))$ *with respect to* $\delta_{0a}(\tau)$.

Sparsity is an important statistical property in high-dimensional statistics. By assuming that only a small subset of copula families is important, we can reduce the complexity to improve the interpretability and predictability of the model. The sparsity property from Theorem 2 demonstrates that the penalized time-varying 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 semiparametric EM algorithm. One possible way to optimize the penalized local log-likelihood copula function in (2.4) is to use a coordinate descent approach. However, it is usually not easy to obtain the explicit forms of the first and second derivatives of the object function, especially when the number of copulas is large and there exist constraints on the weights and dependence parameters. As stated in Cai and Wang (2014), the EM algorithm is one of the most popular algorithms for finding the maximum likelihood estimation of a finite mixture model. Hence, in this section, we propose a semiparametric version of the EM algorithm to estimate the weights and dependence parameters, which dramatically reduces the computational complexity. The algorithm iteratively alternates between an expectation step (E-step) and a maximization step (M-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 Cai and Wang (2014).

To develop a semiparametric version of the EM algorithm for the proposed model, we follow Fan and Li (2001) and Cai, Juhl and Yang (2015), and approximate equation (2.4) by

$$Q^P(\delta) \approx \sum_{j=1}^{T}\left[\sum_{i=1}^{T}\ell(\hat{u}_i,\delta(t_j))K_h(t_i-t_j) - T\sum_{k=1}^{d}\frac{P'_{\gamma_k}(\|\hat{\lambda}_{\cdot k}^{(0)}\|)}{2\|\hat{\lambda}_{\cdot k}^{(0)}\|}\lambda_k^2(t_j) + \rho_{t_j}\left(1-\sum_{k=1}^{d}\lambda_k\left(t_j\right)\right)\right] + \{\text{terms unrelated to } \delta\},$$

where $\hat{\lambda}_{\cdot k}^{(0)}$ are the estimates from the previous iteration. At the first iteration, $\hat{\lambda}_{\cdot k}^{(0)}$ denotes a set of starting values for the weights.

Then, the estimator $\hat{\delta}(t_j)$ at a given iteration step can be obtained by maximizing the criterion function

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

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

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

where $c_c(\hat{u}_{1i},\ldots,\hat{u}_{Ni},,\delta_k(t_i)) = \sum_{k=1}^{d}\lambda_k(t_i)c_k(\hat{u}_{1i},\ldots,\hat{u}_{Ni},\theta_k(t_i))$.

We next 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

$$\lambda_k^{(1)}(\tau) = \left(\sum_{i=1}^{T}\frac{\lambda_k^{(0)}(\tau)c_k(\hat{u}_{1i},\ldots,\hat{u}_{Ni},\theta_k^{(0)}(\tau))}{c_c(\hat{u}_{1i},\ldots,\hat{u}_{Ni},\delta_k^{(0)}(\tau))}K_h(t_i-\tau) - TD_k^{(0)}\right)\Big/$$

$$\left(\sum_{i=1}^{T} K_h(t_i - \tau) - T\sum_{k=1}^{d} D_k^{(0)}\right),$$

for $k = 1, \ldots, d$, where $D_k^{(0)} = (P'_{\gamma_k}(\|\hat{\lambda}_{\cdot k}^{(0)}\|)/\|\hat{\lambda}_{\cdot k}^{(0)}\|)\lambda_k^{(0)2}(\tau)$.

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 maximizing $Q(\delta(\tau)) = \sum_{i=1}^{T} \ell(\hat{u}_i, \delta(\tau))K_h(t_i - \tau)$. We use the 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, respectively, of $Q_\theta(\delta(\tau))$ with respect to θ. 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

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

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

B. Bandwidth and tuning parameter selection. The bandwidth h determines the trade-off between the bias and the variance of the nonparametric estimators, whereas the tuning parameter γ_T adjusts the weight for the penalty term. We need to choose suitable regularization parameters to perform 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, including cross-validation and AIC- and BIC-type criteria, among others. Owing to the time series nature of the sequence $\{X_{1i}, \ldots, X_{Ni}\}_{t=1}^{T}$, we propose using *forward leave-one-out* cross-validation to select the bandwidth h and the tuning parameter γ_T in the penalty term simultaneously.

Define $\hat{\delta}(h, \gamma_T)$ as the nonparametric estimators for the penalized time-varying mixture copula models in (2.4) with a known bandwidth h and the tuning parameter γ_T. For each data point $i_0+1 \le i^* \le T$, we use the data $\{X_{1i}, \ldots, X_{Ni}, i < i^*\}$

to construct the estimate $\hat{\delta}_{t^*}(h, \gamma_T)$ at the sample point $\{x_{1i^*}, \ldots, x_{Ni^*}\}$, 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

$$(h^*, \gamma_T^*) = \underset{(h,\gamma_T)}{\operatorname{argmax}} \sum_{i^*=i_0+1}^{T} \{\ell(\hat{u}_{i^*}, \hat{\delta}(t_i^*)) | \hat{\delta}_{i^*}(h, \gamma_T)\}, \tag{2.5}$$

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

C. Confidence intervals. For inference, independent and identically distributed (i.i.d.) bootstrap approaches are not applicable here, because most of the financial/economic data are dependent. Patton (2012a) suggests 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, we can preserve the original time series structure within a block. A simple block bootstrap for calculating confidence intervals can be implemented as follows:

i. Generate a sample sequence $\{x_{1,i}^*, \ldots, x_{N,i}^*\}_{i=1}^{T}$ from the original data $\{x_{1,i}, \ldots, x_{N,i}\}_{i=1}^{T}$ using a stationary bootstrap technique, as described in the online Supplementary Material.

ii. Obtain $\hat{u}_{1i}^*, \ldots, \hat{u}_{Ni}^*$ by Steps 1–2 described in Section 2.1.

iii. Calculate new local constant estimators $\hat{\delta}^*(\tau)$ at the grid point τ using equation (2.4) with estimators $\{\hat{u}_{1i}^*, \ldots, \hat{u}_{Ni}^*\}_{i=1}^{T}$.

iv. Repeat Steps i–iii M times (say, M=1,000), and get M 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.

v. The empirical $100(1-\alpha)\%$ confidence interval for $\hat{\delta}(\tau)$ is $[q_{\alpha/2}, q_{1-\alpha/2}]$.

3. Monte Carlo Simulation Studies

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

$$x_{si} = \varphi_s x_{s,i-1} + e_{si}, \quad s = 1, 2; \ i = 2, \ldots, T,$$

where $\varphi_1 = 0.1$, $\varphi_2 = 0.05$, $e_{si} = \sigma_{si}\epsilon_{si}$, ϵ_{si} has a standard normal marginal distribution, and

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

where $\alpha_{10} = 0.0001$, $\alpha_{11} = 0.02$, and $\beta_{11} = 0.93$ for the first margin, and $\alpha_{20} = 0.0001$, $\alpha_{21} = 0.03$, and $\beta_{21} = 0.92$ for the second margin. Our working mixture copula model consists of three copulas that are widely used in empirical studies: the Gumbel, Frank, and Clayton copulas. The Frank copula shows a symmetric dependence structure, whereas the Clayton and Gumbel copulas are asymmetric. In particular, the Clayton copula displays strong lower-tail dependence, whereas the Gumbel copula exhibits strong upper-tail dependence. The dependence structure between ϵ_{1i} and ϵ_{2i} is governed by a time-varying mixture copula model

$$(\epsilon_{1i}, \epsilon_{2i}) \sim \sum_{k=1}^{3} \lambda_k(t_i) C_k(u_1, u_2; \theta_k(t_i)),$$

where one of the three weight parameters (λ_1, λ_2, or 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 1,000 times ($M = 1{,}000$). For each sample, we calculate the estimated weights and dependence parameters on a grid of 50 equally spaced points $\tau_j = -0.01 + 0.02j$, for $j \in \{1, 2, \ldots, 50\}$.

3.1. Case-I simulations

We first consider a scenario in which the 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 denote 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: $\lambda_1 = 0$, $\lambda_2 = 1/2$, $\lambda_3 = 1/2$, $\theta_2 = 4$, $\theta_3 = 5$.

Table 1. Percentages of correctly (incorrectly) chosen copulas for models in Case-I (Panel A) and Case-II (Panel B) simulations, when the working mixture copula model is time-varying.

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	(0.115)	1.000
	800	1.000	(0.073)	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	(0.143)	1.000
	800	1.000	(0.058)	1.000
Clayton+Frank	400	(0.011)	1.000	1.000
	800	(0.000)	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.

We summarize the estimation results for the weights and dependence parameters in the Case-I simulations in Tables 1–2, Panel A. Table 1, Panel A, presents the percentages corresponding to the correctly and incorrectly (in parentheses) selected copulas. The results show that the proposed method performs very well in selecting appropriate copulas from mixture copula models with constant parameters, even though our method is designed for time-varying 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 of incorrectly selecting the Frank copula is small and decreases with T.

To examine the performance of the proposed method in estimating the unknown parameters, we calculate the mean squared errors (MSEs) of the estimated weights and dependence parameters for the mixture copula models under the Case-I simulations. The MSEs are calculated as

Table 2. MSEs of the estimated weights and dependence parameters for models in Case-I (Panel A) and Case-II (Panel B) simulations, when the working mixture copula model is time-varying.

Model	T	(λ_1, θ_1)	(λ_2, θ_2)	(λ_3, θ_3)
Panel A: Case I				
Gumbel+Frank	400	(0.007, 0.513)	(0.007, 0.858)	
	800	(0.003, 0.239)	(0.003, 0.380)	
Gumbel+Clayton	400	(0.006, 0.570)		(0.008, 0.727)
	800	(0.003, 0.326)		(0.004, 0.365)
Clayton+Frank	400		(0.006, 0.829)	(0.006, 0.653)
	800		(0.001, 0.426)	(0.001, 0.395)
Panel B: Case II				
Gumbel+Frank	400	(0.009, 2.085)	(0.009, 1.509)	
	800	(0.004, 0.630)	(0.004, 0.497)	
Gumbel+Clayton	400	(0.006, 0.920)		(0.008, 2.971)
	800	(0.002, 0.416)		(0.003, 0.985)
Clayton+Frank	400		(0.009, 1.978)	(0.009, 1.763)
	800		(0.004, 0.716)	(0.004, 0.855)

$$MSE(\widehat{\theta}_k) = \frac{1}{M}\frac{1}{50}\sum_{m=1}^{M}\sum_{j=1}^{50}\left(\widehat{\theta}_{mk}(\tau_j) - \theta_k\right)^2, \text{ and}$$

$$MSE(\widehat{\lambda}_k) = \frac{1}{M}\frac{1}{50}\sum_{m=1}^{M}\sum_{j=1}^{50}\left(\widehat{\lambda}_{mk}(\tau_j) - \lambda_k\right)^2, \text{ for } k = 1, 2, 3,$$

where 50 is the number of grid points and $M = 1{,}000$ 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.

In the online Supplementary Material, we further evaluate the quality of the estimators graphically. Figures S1–S3 in the Supplementary Material show the simulation results of the weights and dependence parameters for Models 1–3, respectively, under Case-I simulations. In each figure, the black solid line denotes the true parameters (the weight or dependence parameter), and the two curves represent the medians (blue) and the means (red), respectively, of the 1,000 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 present only 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.

3.2. Case-II simulations

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

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

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

Tables 1–2, Panel B, 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 time-varying 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 time-varying 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 of selecting incorrect copulas is small. For example, when $T = 800$, there is only a 5.8% probability of selecting Frank when the 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, to check whether the proposed method can accurately estimate the unknown parameters under the Case-II simulations. Again, we omit the results for 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. Similarly to the Case-I simulations, the MSEs are calculated as $MSE(\widehat{\theta}_k) = (1/M)(1/50)\sum_{m=1}^{M}\sum_{j=1}^{50}(\widehat{\theta}_{mk}(\tau_j) - \theta_k(\tau_j))^2$ and $MSE(\widehat{\lambda}_k) = (1/M)(1/50)\sum_{m=1}^{M}\sum_{j=1}^{50}(\widehat{\lambda}_{mk}(\tau_j) - \lambda_k(\tau_j))^2$, for $k = 1, 2, 3$. 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 with the results in Table 2, Panel A, the MSEs in

Panel B are larger in most cases. This is not surprising, because the true models in Case II are time-varying mixture copulas with dynamic parameters, which are more difficult to estimate than the true models in Case I (mixture copulas with constant parameters).

Figures S4–S6 in the online Supplementary Material present the estimated and true parameter paths for different time-varying mixture copulas models ($T = 800$). We observe from Figures S4–S6 that the median and mean paths are still close to the true parameter functions in all models.

We next employ time-invariant mixture copula models for comparision (we thank one anonymous reviewer for suggesting this.). That is, the working mixture copula model still consists of the Gumbel, Frank, and Clayton copulas, but now the weights and dependence parameters are assumed to be constants. We use the penalized likelihood method of Cai and Wang (2014) for the estimation; the MSEs of the estimates are presented in Table 3 (Panel A for Case I, and Panel B for Case II). Comparing the results in Table 3 with the results in Table 2, we observe that when the true models are time-invariant (Case I), employing time-invariant working mixture copula models may indeed gain some efficiency, but the differences between the two types of working models are minor. However, when the true models are dynamic (Case II), using time-invariant mixture copula models yields much worse estimation results than those of the time-varying working mixture copula models.

Finally, to check the robustness of the proposed method, we consider two additional candidate copulas: the rotated Gumbel, and the rotated Clayton; both capture negative dependence. That is, our working mixture copula model now consists of five copulas: the Gumbel, Frank, Clayton, rotated Gumbel, and rotated Clayton copulas. The true model is still a mixture of two copulas. Therefore, three of the five weight parameters are equal to zero in this case. Denote λ_4 and λ_5 as the weights of the rotated Gumbel and the rotated Clayton, respectively. To save space, we consider only Model 3 of Case II (Clayton+Frank) as the true mixture copula: $\lambda_1(\tau) = 0$, $\lambda_2(\tau) = 0.7 - 0.4\sin^2((\pi/2)\tau)$, $\lambda_3(\tau) = 1 - \lambda_2(\tau)$, $\lambda_4(\tau) = 0$, $\lambda_5(\tau) = 0$, $\theta_2(\tau) = 6\tau^2 + 4$, $\theta_3(\tau) = \ln(1 + \tau T) + 3$.

Table 4 presents the selection and estimation performance of the proposed method when employing a larger candidate copula set. The results in Panel A of Table 4 once again support the finding that the proposed method can accurately select the appropriate copulas. Even for $T = 400$, the probability of correctly selecting Clayton and Frank is very close or equal to 100%. Furthermore, the percentage that redundant copulas (Gumbel, rotated Gumbel, and rotated Clayton) are selected is zero. In Panel B, we report the MSEs of the estimated weights and

Table 3. MSEs of the estimated weights and dependence parameters for models in Case-I (Panel A) and Case-II (Panel B) simulations, when the working mixture copula model is time-invariant.

Model	T	(λ_1, θ_1)	(λ_2, θ_2)	(λ_3, θ_3)
Panel A: Case I				
Gumbel+Frank	400	(0.005, 0.313)	(0.005, 0.596)	
	800	(0.002, 0.161)	(0.002, 0.279)	
Gumbel+Clayton	400	(0.005, 0.405)		(0.006, 0.603)
	800	(0.003, 0.183)		(0.003, 0.290)
Clayton+Frank	400		(0.004, 0.713)	(0.004, 0.551)
	800		(0.001, 0.325)	(0.001, 0.272)
Panel B: Case II				
Gumbel+Frank	400	(0.052, 10.664)	(0.047, 9.023)	
	800	(0.029, 5.026)	(0.026, 4.357)	
Gumbel+Clayton	400	(0.038, 8.428)		(0.055, 20.799)
	800	(0.016, 3.343)		(0.023, 8.098)
Clayton+Frank	400		(0.062, 11.370)	(0.070, 12.385)
	800		(0.027, 5.839)	(0.031, 7.648)

Table 4. Selection and estimation performance of the proposed method when considering five candidate copulas: Gumbel, Frank, Clayton, rotated Gumbel, and rotated Clayton.

Panel A: Selection	Gumbel	Frank	Clayton	Rotated Gumbel	Rotated Clayton
$T = 400$	(0.000)	1.000	0.986	(0.000)	(0.000)
$T = 800$	(0.000)	1.000	1.000	(0.000)	(0.000)
Panel B: Estimation	λ_2	θ_2	λ_3	θ_3	
$T = 400$	0.014	2.873	0.014	3.115	
$T = 800$	0.006	1.066	0.006	1.287	

NOTE: The true model is a mixture copula of Clayton and Frank: $\lambda_1(\tau) = 0$, $\lambda_2(\tau) = 0.7 - 0.4\sin^2((\pi/2)\tau)$, $\lambda_3(\tau) = 1 - \lambda_2(\tau)$, $\lambda_4(\tau) = 0$, $\lambda_5(\tau) = 0$, $\theta_2(\tau) = 6\tau^2 + 4$, $\theta_3(\tau) = \ln(1 + \tau T) + 3$.

dependence parameters for Frank (λ_2 and θ_2) and Clayton (λ_3 and θ_3). Compared with the estimation results in Table 2, Panel B, we find that increasing the number of the candidate copulas from three to five only moderately decreases the accuracy of the estimation. Figure S7 in the online Supplementary Material provides additional graphical evidence of the performance of the proposed method in estimating unknown parameters in this case.

4. An Empirical Study

In this section, we apply the proposed model and method to analyze the co-movements of returns between international stock markets during different periods. Specifically, we consider the weekly returns of the Morgan Stanley Capital International (MSCI) equity indices 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 were severely affected by the Asian crisis of 1997 and/or the global financial crisis of 2008. By analyzing the evolution of the dependence structures between these four markets, we can examine how these markets are related, for example, in tranquil periods and in crisis periods.

4.1. Data

The data we use span the period of 28 years from January 1990 until July 2018, with a total of 1,488 observations for each economy. We first obtain the equity indices from Datastream, and then calculate their log-returns as $r_{s,i} = \log(P_{s,i}) - \log(P_{s,i-1})$, where $P_{s,i}$ is the stock index of the ith market at time i. We use weekly data instead of daily data in order to remove the effect of different trading hours for international stock markets (Chollete, Heinen and Valdesogo (2009); Hafner and Reznikova (2010)). Descriptive statistics are presented in Table 5, Panel A. The US market exhibits the highest mean and median returns. The Korean market shows the largest volatility of returns. We employ the Jarque–Bera test for normality, strongly rejecting the null hypothesis for all series.

Table 5, 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 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, we model the marginal distributions of the data. We employ AR(p)-GARCH(1,1) models, a special case of model (2.1) presented in Section 2, to capture possible autocorrelation and conditional heteroscedasticity in the returns. The Bayesian information criterion (BIC) is used to select the appropriate number of lags p for the AR(p) models. Specifically, we use the following models for the

Table 5. 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 the weekly index returns for the United States (US), the United Kingdom (UK), Hong Kong (HK), and South Korea (KR). All returns are expressed in US dollars from January 1990 to July 2018, which correspond to a sample of 1,488 observations. JB statistic and JB p-value refer to the Jarque–Bera test for normality. Panel B reports the linear correlation coefficients and Kendall's τs (in parentheses) across the US, UK, HK, and KR markets.

marginal distributions:

$$X_{si} = \varphi_{s0} + \sum_{k=1}^{p} \varphi_{sk} X_{s,i-k} + e_{si}, \quad e_{si} = \sigma_{si}\epsilon_{si},$$
$$\sigma_{si}^2 = \alpha_{s0} + \alpha_{s1} e_{s,i-1}^2 + \beta_{s1}\sigma_{s,i-1}^2,$$

where X_{si} denotes the return of the sth market at time i. The innovations ϵ_{si} are assumed to be i.i.d for $i = 1, \ldots, T$ with a fixed s, and the distribution is estimated using the rescaled empirical distribution of the residuals. Model diagnostics such as portmanteau-type tests for the mean and the variance confirm our model specifications. To economize on space, they are not reported here, but are available upon request.

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) that have relatively large correlation coefficients. Scatter plots (omitted here) of the four pairs of standardized returns show vi-

Table 6. Estimation results and tests of the marginal distribution models.

	AR(p)	GARCH(1,1)			LB	
	φ_1	α_0	α_1	β_1	4	16
	(s.e.)	(s.e.)	(s.e.)	(s.e.)		
US	-0.118	0.007E-3	0.100	0.890	0.617	0.643
	(0.026)	(0.004E-3)	(0.019)	(0.019)		
UK	-0.098	0.023E-3	0.127	0.840	0.892	0.577
	(0.027)	(0.009E-3)	(0.028)	(0.036)		
HK		0.020E-3	0.098	0.885	0.112	0.155
		(0.008E-3)	(0.020)	(0.023)		
KR		0.045E-3	0.115	0.862	0.120	0.337
		(0.015E-3)	(0.020)	(0.023)		

NOTE: The second to fifth columns report the parameter estimates of the AR(p)-GARCH(1,1) models. Values in parentheses are corresponding standard errors. The sixth and seventh columns report the p-values of the Ljung–Box (LB) tests for autocorrelation of the residuals, using 4 and 16 lags, respectively.

olations 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 of the Clayton, Frank, and Gumbel copulas to implement the copula selection and estimation. This enables us to capture various dependence structures in the data, such as 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 penalized likelihood method of Cai and Wang (2014) is employed to select and estimate the model. The results are reported in Table 7. We have two findings from Table 7. 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 those in Cai and Wang (2014).

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 proposed time-varying mixture copula model to analyze the dependence structures of the international stock markets. Figures 1–4 present the estimation results and the

Table 7. Estimation results of the time-invariant mixture copula models for international markets.

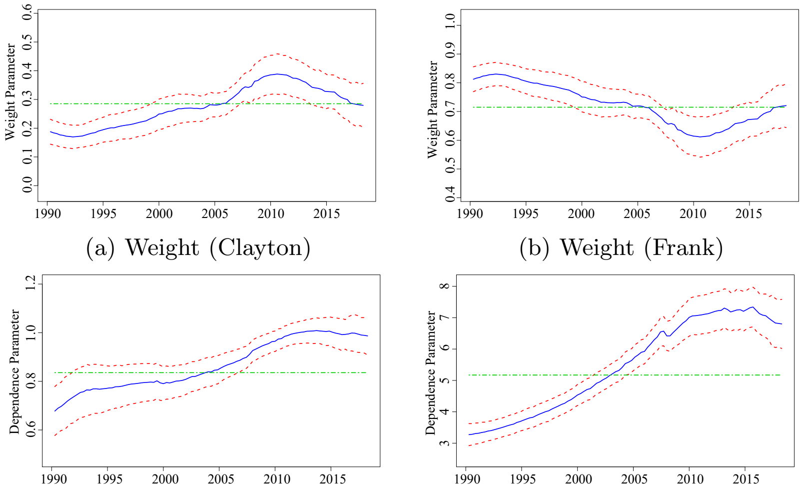
	Markets	Clayton	Gumbel	Frank
λ	US-UK	0.285(0.241,0.329)	0	0.715(0.671,0.759)
	US-HK	0.314(0.268,0.360)	0	0.686(0.640,0.732)
	UK-HK	0.352(0.304,0.399)	0	0.648(0.601,0.696)
	HK-KR	0.208(0.150,0.267)	0	0.792(0.733,0.850)
θ	US-UK	0.836(0.781,0.890)		5.172(4.823,5.521)
	US-HK	0.594(0.536,0.652)		3.761(3.565,3.956)
	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 the penalized likelihood method of Cai and Wang (2014). Values in parentheses are the 90% confidence intervals of the estimates.

90% confidence intervals of all nonzero weights and dependence parameters for the US–UK, US–HK, UK–HK, and HK–KR pairs, respectively. In each figure, the path of the estimated parameter (the weight or dependence parameter) is represented by a solid line. The two dashed curves show the 90% confidence intervals of each estimated parameter. The two-dashed line (horizontal line) is the estimate using the time-invariant mixture copula model. We find several interesting results from these figures.

First, for all pairs of markets, the Clayton and Frank copulas are selected for any period within 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 in both the weights and the 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 1, the weight and dependence parameter of the Clayton copula are both relatively small in the early 1990s. Moreover, 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 increases sharply after the events of September 11, 2001. At the same time, the weight of Clayton also reaches a relatively high value. During the financial crisis of 2008, the weight and dependence parameter of the Clayton



(a) Weight (Clayton) (b) Weight (Frank)

(c) Dependence parameter (Clayton) (d) Dependence parameter (Frank)

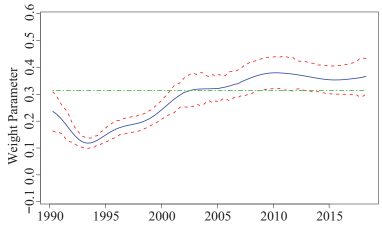
Figure 1. Estimation results of nonzero weights and dependence parameters (solid lines), along with the 90% confidence intervals (dashed curves) for the US–UK pair. The two-dashed lines (horizontal lines) show the estimates using the time-invariant mixture copula model.

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 attains high levels in crisis periods.

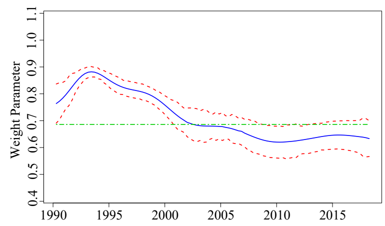
With regard to the US–HK and UK–HK pairs, the dependence structures display similar evolution paths (see Figures 2–3). Both pairs show relatively weak lower-tail dependence and overall dependence during the 1990s. A notable jump in the Clayton parameter took place in 2008 for both pairs, owing to the financial crisis. An increase in the weight on Clayton can be observed during the same period.

The last figure (Figure 4) shows the dependence structure of the HK–KR pair. These two markets were strongly affected by the Asian crisis of 1997. Therefore, we 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.

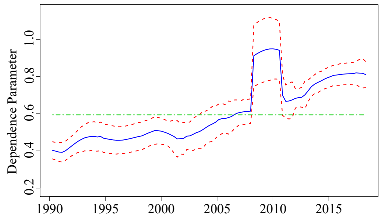
Finally, we check the goodness-of-fit of the estimated time-varying mixture copula model using the Kolmogorov–Smirnov (KS) test, Cramer–von Mises (CM) test, and Anderson–Darling (AD) test for correct copula specification. The goodness-of-fit procedures are available in the Supplementary Material. Table



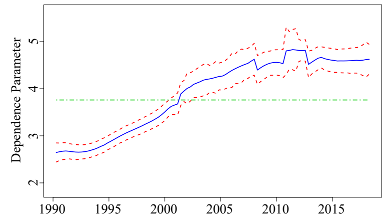
(a) Weight (Clayton)



(b) Weight (Frank)

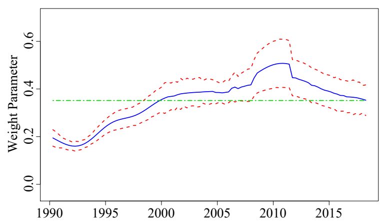


(c) Dependence parameter (Clayton)

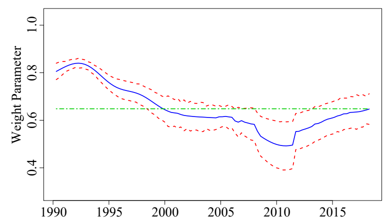


(d) Dependence parameter (Frank)

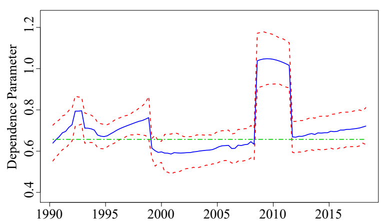
Figure 2. Estimation results of nonzero weights and dependence parameters (solid lines), along with the 90% confidence intervals (dashed curves) for the US–HK pair. The two-dashed lines (horizontal lines) show the estimates using the time-invariant mixture copula model.



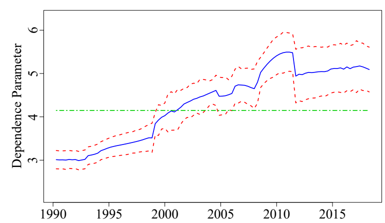
(a) Weight (Clayton)



(b) Weight (Frank)

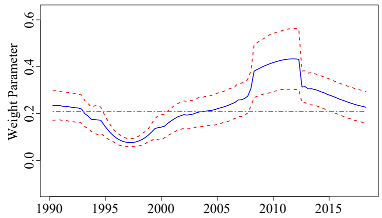


(c) Dependence parameter (Clayton)

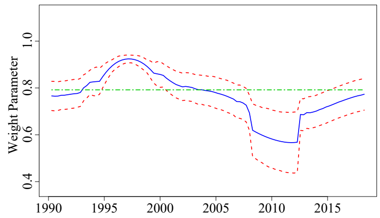


(d) Dependence parameter (Frank)

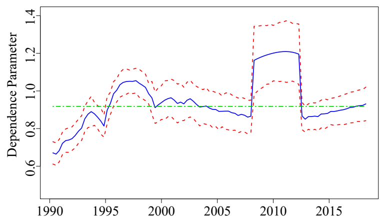
Figure 3. Estimation results of nonzero weights and dependence parameters (solid lines), along with the 90% confidence intervals (dashed curves) for the UK–HK pair. The two-dashed lines (horizontal lines) show the estimates using the time-invariant mixture copula model.



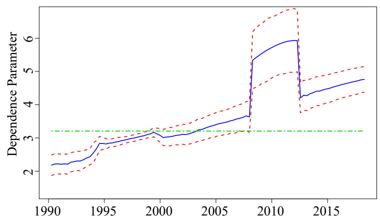
(a) Weight (Clayton)



(b) Weight (Frank)



(c) Dependence parameter (Clayton)



(d) Dependence parameter (Frank)

Figure 4. Estimation results of nonzero weights and dependence parameters (solid lines), along with the 90% confidence intervals (dashed curves) for the HK–KR pair. The two-dashed lines (horizontal lines) show the estimates using the time-invariant mixture copula model.

Table 8. Goodness-of-fit tests for the time-varying 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: the Kolmogorov–Smirnov (KS) test, the Cramer–von Mises (CM) test, and the Anderson–Darling (AD) test.

8 reports the bootstrap p-values of the three tests. All models pass the three tests with large p-values.

5. Conclusion

We have proposed a time-varying mixture copula model, in which both the weights and the dependence parameters are deterministic functions of time. To reduce the risk of over-fitting and efficiency loss, we propose a penalized time-varying mixture copula model with the SCAD penalty term in order to perform the estimation and the copula selection simultaneously. Based on α-mixing conditions, we establish the large sample properties of the penalized and unpenalized estimators. We also study and discuss a semiparametric EM algorithm, the bandwidth selection, and the construction of pointwise confidence intervals. Monte Carlo simulations demonstrate the good performance of the proposed method in copula selection and estimation for time-varying mixture copulas. The proposed methodology is applied to study the evolution of the dependence between four international stock markets. All pairs of markets present strong dependence at the lower tail that fluctuates significantly over time. Furthermore, all pairs exhibit the highest levels of both lower-tail and overall dependence during the financial crisis of 2008.

Future research should seek to employ the proposed model and method in other fields in finance and economics, such as exchange rates, bonds, and crude oil. Moreover, future works should study the higher-dimensional dependence structures between financial markets by using the proposed model and method.

Supplementary Material

The online Supplementary Material provides the stationary bootstrap resampling scheme, the regularity conditions and asymptotic properties for unpenalized estimators, mathematical proofs, the procedures for the goodness-of-fit, and some additional figures.

Acknowledgments

The authors thank the Co-Editor, Professor Yazhen Wang, and the referees for their helpful suggestions for improvements to this paper. Yang's research is partially supported by NSFC(72173140), Humanity and Social Science Youth Foundation of Ministry of Education of China (19YJC790166), the Fundamental Research Funds for the Central Universities (19wkpy61), the Major Program of the National Social Science Foundation of China (17ZDA073) and the Major Program of NSFC (71991474). Cai's research is partially supported by the National Natural Science Foundation of China (Grant Nos. 71631004 and 72033008). Hafner's research is supported by research grant ARC 18-23/089 of the Belgian Federal Science Policy Office. Liu's research is supported by the National Natural Science Foundation of China (Grant No. 71803160). The corresponding author is Guannan Liu, MOE Key Laboratory of Econometrics, School of Economics and Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, China; email: gliuecon@gmail.com.

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.

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

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

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

Chen, X. and Fan, Y. (2006). Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *Journal of Econometrics* **135**, 125–154.

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 with applications. *Journal of Applied Econometrics* **28**, 777–795.

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.

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.

Liu, B. Y., Ji, Q. and Fan, Y. (2017). A new time-varying optimal copula model identifying the dependence across markets. *Quantitative Finance* **17**, 437–453.

Liu, G., Long, W., Zhang, X. and Li, Q. (2019). Detecting financial data dependence structure by averaging mixture copulas. *Econometric Theory* **35**, 777–815.

Mensi, W., Hammoudeh, S., Shahzad, S. J. H. and Shahbaz, M. (2017). 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.

Nasri, B. R., Rémillard, B. N. and Bouezmarni, T. (2019). Semi-parametric copula-based models under non-stationarity. *Journal of Multivariate Analysis* **173**, 347–365.

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.

Robinson, P. M. (1989). Nonparametric estimation of time-varying parameters. In *Statistical Analysis and Forecasting of Economic Structural Change* (Edited by P. Hackl), 253–264. Springer, Berlin.

Sklar, A. (1959). Fonctions de répartition *à* n dimensions et leurs marges. In *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**, 416–434.

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

Tsay, R. (2010). *Analysis of Financial Time Series*. 3rd Edition. John Wiley & Sons, Inc., Hoboken, New Jersey.

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, Series B (Statistical Methodology)* **68**, 49–67.

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

Bingduo Yang

School of Finance, Guangdong University of Finance and Economics, Guangzhou, 510320, China.

E-mail: bdyang2006@sina.com

Zongwu Cai

Department of Economics, University of Kansas, Lawrence, KS 66045, USA.

E-mail: caiz@ku.edu

Christian M. Hafner

Louvain Institute of Data Analysis and Modeling, Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium.

E-mail: christian.hafner@uclouvain.be

Guannan Liu

MOE Key Laboratory of Econometrics, School of Economics and Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, 361005, China.

E-mail: gliuecon@gmail.com

(Received January 2020; accepted October 2020)