Time-Varying Mixture Copula Models with Copula Selection

Bingduo Yang[1], Zongwu Cai[2], Christian M. Hafner[3] and Guannan Liu[4]

[1]*School of Finance, Guangdong University of Finance and Economics*

[2]*Department of Economics, University of Kansas*

[3]*Louvain Institute of Data Analysis and Modeling, Université catholique de Louvain*

[4]*School of Economics and WISE, Xiamen University*

Supplementary Material

This supplement provides the stationary bootstrap resampling scheme, the regularity conditions and asymptotic properties for unpenalized estimators, the mathematical proofs, the procedures of doing the goodness-of-fit, and some additional figures.

S1 The stationary bootstrap resampling scheme

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

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

be the block consisting of b observations starting from $(x_{1,i}, \cdots, x_{N,i})$ to $(x_{1,i+b-1}, \cdots, x_{N,i+b-1})$. In the case $j > T$, $(x_{1,j}, \cdots, x_{N,j})$ is defined to be $(x_{1,i}, \cdots, x_{N,i})$, where $i = j(modT)$ and $(x_{1,0}, \cdots, x_{N,0}) = (x_{1,T}, \cdots, x_{N,T})$. Let p be a constant such that $p \in [0, 1]$. Independent of $\{x_{1,i}, \cdots, x_{N,i}\}_{i=1}^{T}$, 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 $\{x_{1,i}, \cdots, x_{N,i}\}_{i=1}^{T}$ 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 $\{x^*_{1,i}, \cdots, x^*_{N,i}\}_{i=1}^{T}$ 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 $\{x^*_{1,i}, \cdots, x^*_{N,i}\}_{i=1}^{T}$ are determined by the first block B_{I_1,L_1} of observations $(x_{1,I_1}, \cdots, x_{N,I_1}), \cdots, (x_{1,I_1+L_1-1}, \cdots, x_{N,I_1+L_1-1})$, the next L_2 observations in the pseudo time series are the observations in the second sampled block B_{I_2,L_2}, namely $(x_{1,I_2}, \cdots, x_{N,I_2}), \cdots, (x_{1,I_2+L_2-1}, \cdots, x_{N,I_2+L_2-1})$. 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 $\{x^*_{1,i}, \cdots, x^*_{N,i}\}_{i=1}^T$, conditional on the original data $\{x_{1,i}, \cdots, x_{N,i}\}_{i=1}^T$, is actually stationary. Hence, this resampling method is applicable for stationary and weekly dependent time series.

S2 Asymptotic properties for unpenalized estimators

The three-step procedure in Section 2.1 suggests that the marginal distribution estimators $\hat{\psi}_1$, $\hat{\psi}_2$, and $\hat{F}_s$ have little effect on the nonparametric estimators $\hat{\delta}$ in large samples since the convergence rate of the marginal distribution estimators $\hat{F}_s$, $\sqrt{T}$, is faster than $\sqrt{Th}$ in the nonparametric component. In the following, we present the asymptotic properties of the nonparametric estimators without considering the errors from the marginal estimation in the first two steps. For this purpose, we introduce regularity conditions as below.

A1. The estimators of unknown marginal distribution functions satisfy $\sqrt{T} sup_{y_s} |\hat{F}_s(y_s) - F_s(y_s)| = O_p(1)$.

A2. The vector of functions $\delta(\tau)$ is continuous, bounded and has second order continuous derivatives on the support $[0, 1]$. The function $\ell(u_i, \delta(\tau))$ is three times differentiable with respect to δ and twice differentiable with respect to u_i.

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

A6. The bandwidth h satisfies that $h \to 0$ and $Th^{1+4/c} \to \infty$, as $T \to \infty$.

Remark 1. The condition in A1 directly follows Lemma A.1 in Chen and Fan (2006b). The conditions in A2 are for deriving the asymptotic properties of the nonparametric estimators. Moreover, by the conditions in A2, 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 for kernel function in A4 are commonly employed in nonparametric estimation. The conditions in A5 are α-mixing conditions for weakly dependent data. Most financial models satisfy these conditions, such as ARMA and GARCH models, see Carrasco and Chen (2002). When $c > 1$, the optimal bandwidth $h = O(T^{-1/5})$ satisfies Condition A6.

Theorem A.1: *Let $\{X_{1i}, \cdots, X_{Ni}\}_{i=1}^T$ be a strictly stationary α-mixing sequence following the proposed models (2.1)-(2.2) in the main text. For a fixed point $\tau \in$*

$(0,1)$, *under Conditions A1-A6, we have*

$$\sqrt{Th}(\hat{\delta}(\tau) - \delta_0(\tau) - h^2 B(\tau)) \to N(0, v_0 \Sigma(\tau)^{-1} \Omega(\tau) \Sigma(\tau)^{-1}),$$

where $\Sigma(\tau) = -E\{\ell''(u_i, \delta_0(\tau))|t_i = \tau\}$, $\Omega(\tau) = \sum_{s=-\infty}^{\infty} \Gamma_s(\tau)$ *with* $\Gamma_s(\tau) = E\{\ell'(u_i, \delta_0(\tau))\ell'(u_{i+s}, \delta_0(\tau))^\intercal|t_i = \tau\}$ *and the bias term* $h^2 B(\tau) = \frac{h^2}{2}\delta_0''(\tau)\mu_2$*.* $\ell'(u_i, \delta_0(\tau))$ *and* $\ell''(u_i, \delta_0(\tau))$ *respectively denote the first and second derivatives of* $\ell(u_i, \delta_0(\tau))$ *with respect to* $\delta_0(\tau)$*.*

Remark 2. In classical local constant estimation, the bias term is usually written as $h^2 B(\tau) = \frac{h^2}{f(\tau)}\delta_0'(\tau) f'(\tau)\mu_2 + \frac{h^2}{2}\delta_0''(\tau)\mu_2$, 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 A.1 suggests that the local constant estimator $\hat{\delta}(\tau)$ has the same asymptotic behavior as the local linear estimator 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 A.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 A.2: *Let $\{X_{1i}, \cdots, X_{Ni}\}_{i=1}^{T}$ be a strictly stationary α-mixing sequence following the proposed models (2.1)-(2.2) in the main text. For a left boundary point $\tau = bh$, under Conditions A1-A6, we have*

$$\sqrt{Th}(\hat{\delta}(bh) - \delta_0(bh) - hB^*(0+)) \to N(0, \frac{v_{0,b}}{\mu_{0,b}^2}\Sigma(0+)^{-1}\Omega(0+)\Sigma(0+)^{-1}),$$

where the bias term $hB^*(0+) = \frac{h}{\mu_{0,b}}\delta_0'(0+)\mu_{1,b}$.

Remark 3. The bias term is of order h for a boundary point $\tau = bh$, which suggests that the local constant estimator suffers from boundary effects.

S3 Mathematical proofs

In this section, we first show the proof of Theorem A.1 and Theorem A.2 for unpenalized estimators. These two theorems are critical for the proof of sparsity and asymptotic normality in Theorem 2 for penalized estimators. The proof of Theorem 1 is to show that the penalized estimators employ the $\sqrt{Th}$ convergence rate.

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.

Proof of Theorem A.1 and Theorem A.2:

First, we define

$$A_T = \frac{h}{\sqrt{Th}}\sum_{i=1}^{T} \ell'(u_i, \delta_0(\tau))K_h(t_i - \tau)$$

and

$$B_T = \frac{1}{T}\sum_{i=1}^{T} \ell''(u_i, \delta_0(\tau))K_h(t_i - \tau),$$

where $\ell'(u_i, \delta_0(\tau))$ and $\ell''(u_i, \delta_0(\tau))$ respectively denote the first and second derivatives of $\ell(u_i, \delta_0(\tau))$ with respect to $\delta_0(\tau)$.

To minimize the objective function $\sum_{i=1}^{T} \ell(u_i, \hat{\delta}(\tau))K_h(t_i - \tau)$ at point τ, it is equivalent to minimize $h\{\sum_{i=1}^{T} \ell(u_i, \hat{\delta}(\tau))K_h(t_i - \tau) - \sum_{i=1}^{T} \ell(u_i, \delta_0(\tau))K_h(t_i - \tau)\}$, which can be written as

$$\begin{aligned}
& h\{\sum_{i=1}^{T} \ell(u_i, \hat{\delta}(\tau))K_h(t_i - \tau) - \sum_{i=1}^{T} \ell(u_i, \delta_0(\tau))K_h(t_i - \tau)\} \\
= \;& h\sum_{i=1}^{T} (\hat{\delta}(\tau) - \delta_0(\tau))^\intercal \ell'(u_i, \delta_0(\tau))K_h(t_i - \tau) \\
& + \frac{1}{2}h\sum_{i=1}^{T} (\hat{\delta}(\tau) - \delta_0(\tau))^\intercal \ell''(u_i, \delta_0(\tau))(\hat{\delta}(\tau) - \delta_0(\tau))K_h(t_i - \tau) + o_p(1) \\
= \;& \sqrt{Th}(\hat{\delta}(\tau) - \delta_0(\tau))^\intercal A_T + \frac{1}{2}\sqrt{Th}(\hat{\delta}(\tau) - \delta_0(\tau))^\intercal B_T \sqrt{Th}(\hat{\delta}(\tau) - \delta_0(\tau)) + o_p(1).
\end{aligned}$$

After taking the first derivative with respect to $\sqrt{Th}(\hat{\delta}(\tau) - \delta_0(\tau))$, we obtain

$$\sqrt{Th}(\hat{\delta}(\tau) - \delta_0(\tau)) = -B_T^{-1}A_T + o_p(1).$$

By the moment condition, we have

$$
\begin{aligned}
0 &= E\{\ell'(u_i, \delta_0(t_i))|t_i = \tau\} \\
&= E\{\ell'(u_i, \delta_0(\tau) + r_i)|t_i = \tau\} \\
&= E\{\ell'(u_i, \delta_0(\tau))|t_i = \tau\} + r_i E\{\ell''(u_i, \delta_0(\tau))|t_i = \tau\} + o_p(r_i),
\end{aligned}
$$

where $r_i = \delta_0'(\tau)(t_i - \tau) + \frac{1}{2}\delta_0''(\tau)(t_i - \tau)^2 + o_p(t_i - \tau)^2$. By construction, we have $E\{\ell'(u_i, \delta_0(\tau))|t_i = \tau\} = r_i\Sigma(\tau) + o_p(r_i)$, where $\Sigma(\tau) = -E\{\ell''(u_i, \delta_0(\tau))|t_i = \tau\}$. Thus,

$$
\begin{aligned}
E\{A_T|t_i = \tau\} &= \frac{h}{\sqrt{Th}} \sum_{i=1}^{T} E\{\ell'(u_i, \delta_0(\tau))|t_i = \tau\} K_h(t_i - \tau) \\
&= \frac{h}{\sqrt{Th}} \Sigma(\tau) \sum_{i=1}^{T} r_i K_h(t_i - \tau).
\end{aligned}
$$

Note that

$$
\begin{aligned}
E(-B_T|t_i = \tau) &= E(\ell''(u_i, \delta_0(\tau)) K_h(t_i - \tau)) \\
&= \begin{cases} -\Sigma(\tau) + o_p(1), & \text{if } \tau \in (0, 1); \\ -\mu_{0,b}\Sigma(0+) + o_p(1), & \text{if } \tau = bh. \end{cases}
\end{aligned}
$$

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

$$
\begin{aligned}
& E(\hat{\delta}(\tau)|t_i=\tau)-\delta_0(\tau) \\
&= -\frac{1}{\sqrt{Th}}[E(B_T|t_i=\tau)]^{-1}E(A_T|t_i=\tau) \\
&= \frac{1}{T}\sum_{i=1}^{T}\left[\delta_0'(\tau)(t_i-\tau)+\frac{1}{2}\delta_0''(\tau)(t_i-\tau)^2\right]K_h(t_i-\tau)+R_m \\
&= \int \delta_0'(\tau)(t_i-\tau)K_h(t_i-\tau)dt_i+\frac{1}{2}\int \delta_0''(\tau)(t_i-\tau)^2K_h(t_i-\tau)dt_i+R_m \\
&= \begin{cases} \frac{h^2}{2}\delta_0''(\tau)\mu_2+o_p(h^2), & \text{if } \tau\in(0,1); \\ \frac{h}{\mu_{0,b}}\delta_0'(0+)\mu_{1,b}+o_p(h), & \text{if } \tau=bh. \end{cases}
\end{aligned}
$$

To find the expression for $Var\{A_T|t_i=\tau\}$, we let $Q_T=\frac{1}{T}\sum_{i=1}^{T}Z_i$, where $Z_i=\ell'(u_i,\delta_0(\tau))K_h(t_i-\tau)$. Using the same argument as in Lemma 3 of Cai (2007), we can show that

$$
Var(Q_T)=\begin{cases} \frac{1}{Th}v_0\left(\Gamma_0(\tau)+2\sum_{s=1}^{\infty}\Gamma_s(\tau)\right)+o_p(\frac{1}{Th}), & \text{if } \tau\in(0,1); \\ \frac{1}{Th}v_{0,b}\left(\Gamma_0(0+)+2\sum_{s=1}^{\infty}\Gamma_s(0+)\right)+o_p(\frac{1}{Th}), & \text{if } \tau=bh, \end{cases}
$$

where $\Gamma_s(\tau) = E\{\ell'(u_i, \delta_0(\tau))\ell'(u_{i+s}, \delta_0(\tau))^{\intercal}|t_i = \tau\}$. Therefore,

$$Var(A_T) = \begin{cases} v_0\left(\Gamma_0(\tau) + 2\sum_{s=1}^{\infty}\Gamma_s(\tau)\right) + o_p(1), & \text{if } \tau \in (0,1); \\ v_{0,b}\left(\Gamma_0(0+) + 2\sum_{s=1}^{\infty}\Gamma_s(0+)\right) + o_p(1), & \text{if } \tau = bh, \end{cases}$$

It follows that the variance term is given by

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

To establish the asymptotic normality for $\hat{\delta}(\tau)$, we use the score function with $Z_i = \ell'(u_i, \delta_0(\tau))K_h(t_i - \tau)$ and present the results by the Doobs small-block and large-block technique, which is similar to the proofs in Cai (2007) and pages 251-255 of Fan and Gijbels (1996). Details are omitted here. This completes the proof.

$\square$

Proof of Theorem 1:

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 $v^{\lambda}_{\cdot k} = v_{\cdot (d+k)}$ and set

$\|v\| = \sqrt{\sum_{j,k} v_{j,k}^2}$ to be the L_2-norm for the matrix $v = (v_{jk})$. 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 $D \equiv \frac{h}{T}\left[Q^P(\delta_0 + (Th)^{-1/2}v) - Q^P(\delta_0)\right]$. Following the proof of Theorem 1 in Cai and Wang (2014), we use the the facts that $\sum_{k=1}^d (\lambda_{0k}(t_j) + (Th)^{-1/2}v_{jk}^{\lambda}) - \sum_{k=1}^d \lambda_{0k}(t_j) = 0$ and $\|\lambda_{\cdot 0k}\| = 0$ for $k = d_0 + 1, \cdots, d$, and have

$$
\begin{aligned}
D \quad & \leq \frac{h}{T}\left[\sum_{j=1}^T\sum_{i=1}^T \ell(u_i, \delta_0(t_j) + (Th)^{-1/2}v_{j\cdot})K_h(t_i - t_j) - \sum_{j=1}^T\sum_{i=1}^T \ell(u_i, \delta_0(t_j))K_h(t_i - t_j)\right] \\
& -h\sum_{k=1}^{d_0}\left[P_{\gamma_T}(\|\lambda_{\cdot 0k} + (Th)^{-1/2}v_{\cdot k}^{\lambda}\|) - P_{\gamma_T}(\|\lambda_{\cdot 0k}\|)\right] \\
& \doteq D_1 + D_2,
\end{aligned}
$$

where

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

with $e_j = h^{1/2}T^{-1/2}\sum_{i=1}^{T}\ell'(u_i, \delta_0(t_j))K_h(t_i - t_j)$. By the Cauchy-Schwarz inequality,

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

where $\lambda_{t_j}^{min}$ is the smallest eigenvalue of $\Sigma(t_j)$ 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 $O_p(1)$. It follows by the law of large numbers that $E(e_j^2|t_i = t_j) = Var(e_j|t_i = t_j) + (E(e_j|t_i = t_j))^2 = \Omega(t_j)v_0 + (E(e_j|t_i = t_j))^2 + R_m$ 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 Taylor's expansion and the triangle inequality, we have

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

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

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

It follows by the Riemann sum approximation of an integral and condition (A2) that $\|\lambda_{\cdot 0k}\|^2/T = \int_0^1 \lambda_k^2(\tau) d\tau + o_p(1)$ is a bounded constant. By choosing the SCAD penalty function, as $T \to \infty$, the condition $T^{-1/10}\gamma_T \to 0$ in (B1) implies $\gamma_T <<$ $\|\lambda_{\cdot 0k}\|$. Therefore, $a_T \to 0$ and $D_2 \to 0$. By choosing a sufficient large C, the second term in D_1 dominates other terms. This completes the proof.

□

Proof of Theorem 2:

(a) Firstly, we show the sparsity $\|\hat{\lambda}_{\cdot k}\| = 0$ for all $k = d_0 + 1, \cdots, d$. We follow the proof of Lemma B.1 and Theorem 2 (a) in Cai and Wang (2014), and assume that $\|\hat{\lambda}_{\cdot k}\| \neq 0$ and there exists a $\sqrt{Th}$-consistent penalized estimator $\hat{\delta}_{\gamma_T}$ such that

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

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

and $\rho = (\rho_{t_1}, \cdots, \rho_{t_T})^\intercal$.

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 1, 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

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

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 conditions in (B1), we have $\sqrt{h}\gamma_T \to 0$ and $P'_{\gamma_T}(\|\hat{\lambda}_{\cdot k}\|)/\gamma_T > 0$

for $k = d_0 + 1, \cdots, d$, thus

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

is dominated by $\|J_1\|$ as $T \to \infty$. That is, $\mathrm{P}(\|J_2\| < \|J_1\|) \to 1$ as $T \to \infty$. Further, $\|\rho\|$ is dominated by $\|J_1\|$. 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 (2.4) in the main text as

$$Q^P(\delta_a) = \sum_{j=1}^{T}\sum_{i=1}^{T}\ell(u_i, \delta_a(t_j))K_h(t_i - t_j) - T\sum_{k=1}^{d_0}P_{\gamma_k}(\|\lambda_{\cdot k}\|) + \sum_{j=1}^{T}\rho_{t_j}\left(1 - \sum_{k=1}^{d_0}\lambda_k\left(t_j\right)\right),$$

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

We follow Fan and Li (2001) and Cai et al. (2015) and approximate the above equation by

$$Q^P(\delta_a) = \sum_{j=1}^{T}\left[\sum_{i=1}^{T}\ell(u_i, \delta_a(t_j))K_h(t_i - t_j) - T\sum_{k=1}^{d_0}\frac{P'_{\gamma_k}(\|\lambda_{\cdot k}\|)}{2\|\lambda_{\cdot k}\|}\lambda_k^2(t_j) + \rho_{t_j}\left(1 - \sum_{k=1}^{d_0}\lambda_k\left(t_j\right)\right)\right]$$
$$+ \{\text{terms unrelated to } \delta_a\}.$$

To find the minimizer $\hat{\delta}_{a,\gamma_T}(t_j)$ at point t_j, we minimize the objective function

$$Q^P(\hat{\delta}_{a,\gamma_T}(t_j)) = \sum_{i=1}^{T} \ell(u_i, \hat{\delta}_{a,\gamma_T}(t_j)) K_h(t_i - t_j) - \frac{T}{2} \sum_{k=1}^{d_0} \frac{P'_{\gamma_k}(\|\hat{\lambda}_{\cdot k}\|)}{\|\hat{\lambda}_{\cdot k}\|} \hat{\lambda}_k^2(t_j) + \rho_{t_j} \left(1 - \sum_{k=1}^{d_0} \hat{\lambda}_k(t_j)\right),$$

and it is equivalent to minimize $Q^P(\hat{\delta}_{a,\gamma_T}(t_j)) - Q^P(\delta_{0a}(t_j))$. Following the proof of Lemma B.2 and Theorem 2 (b) in Cai and Wang (2014), we use the fact $\sum_{k=1}^{d_0}(\hat{\lambda}_k(t_j) - \lambda_{0k}(t_j)) = 0$ and have

$$\begin{aligned} Q^P(\hat{\delta}_{a,\gamma_T}(t_j)) - Q^P(\delta_{0a}(t_j)) &= \sum_{i=1}^{T} \ell(u_i, \hat{\delta}_{a,\gamma_T}(t_j)) K_h(t_i - t_j) - \sum_{i=1}^{T} \ell(u_i, \delta_{0a}(t_j)) K_h(t_i - t_j) \\ &\quad - \frac{T}{2} \sum_{k=1}^{d_0} \frac{P'_{\gamma_k}(\|\hat{\lambda}_{\cdot k}\|)}{\|\hat{\lambda}_{\cdot k}\|} \hat{\lambda}_k^2(t_j) + \frac{T}{2} \sum_{k=1}^{d_0} \frac{P'_{\gamma_k}(\|\lambda_{\cdot 0k}\|)}{\|\lambda_{\cdot 0k}\|} \lambda_{0k}^2(t_j). \end{aligned}$$

By the result $\|\hat{\delta}_{\gamma_T}(\tau) - \delta_0(\tau)\| = O_p(1/\sqrt{Th})$ in Theorem 1, we have $\|\hat{\lambda}_{\cdot k}\|^2/T = \|\lambda_{\cdot 0k}\|^2/T + o_p(1) = \int_0^1 \lambda_k^2(\tau) d\tau + o_p(1)$. By choosing the SCAD penalty function, as $T \to \infty$, the condition $T^{-1/10}\gamma_T \to 0$ in (B1) implies $\mathrm{P}(\varrho\gamma_T < \|\hat{\lambda}_{\cdot k}\|) \to 1$ and $\mathrm{P}(\varrho\gamma_T < \|\lambda_{\cdot 0k}\|) \to 1$. Therefore, as $T \to \infty$, $P'_{\gamma_k}(\|\hat{\lambda}_{\cdot k}\|) = 0$ and $P'_{\gamma_k}(\|\lambda_{\cdot 0k}\|) = 0$. The asymptotic normality results directly follow Theorem A.1. This completes the proof.

□

S4 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). However, to the best of our knowledge, goodness-of-fit tests for semiparametric time-varying copula models have not been studied. To evaluate the performance of the estimated time-varying mixture copula model, we use a Rosenblatt probability integral transformation as in Dobric and Schmid (2007). We define the random variable

$$S(u_1, u_2) = [\Phi^{-1}(u_1)]^2 + [\Phi^{-1}(C(u_2|u_1))]^2, \tag{S4.1}$$

where $C(u_2|u_1) = P(U_2 \leq u_2|U_1 = u_1)$ 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:

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

and

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

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:

i. 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 above (Section S1);

ii. Obtain $\hat{u}^*_{1i}$ and $\hat{u}^*_{2i}$ by Steps 1-2 in Section 2.1 of the main text;

iii. Calculate new local constant estimators $\hat{\delta}^*(t_i)$ by the proposed method with paired estimators $\{\hat{u}^*_{1i}, \hat{u}^*_{2i}\}_{i=1}^T$, and obtain $S(\hat{u}^*_{1i}, \hat{u}^*_{2i})$ by equation (S4.1);

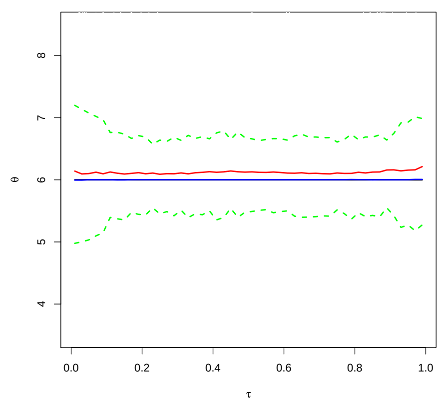
iv. Use the values $S(\hat{u}^*_{1i}, \hat{u}^*_{2i})$ to construct the bootstrap statistics t^*_{KS}, t^*_{CM}, and t^*_{AD};

v. Repeat Steps i $\sim$ iv M times (say, M=1000) and obtain M values of the statistics t^*_{KS}, t^*_{CM}, and t^*_{AD} respectively; and

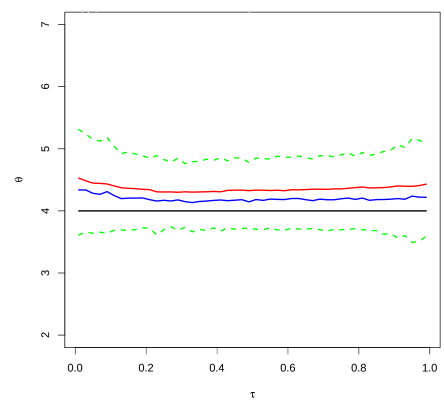
vi. Calculate the values of t_{KS}, t_{CM} and t_{AD} from the original sample $\{\hat{u}_{1i}, \hat{u}_{2i}\}_{i=1}^T$ and compute the p-values of the tests based on the relative frequency of the events $\{t^*_{KS} \geq t_{KS}\}$, $\{t^*_{CM} \geq t_{CM}\}$, and $\{t^*_{AD} \geq t_{AD}\}$ in the M repetitions.

S5 Additional figures

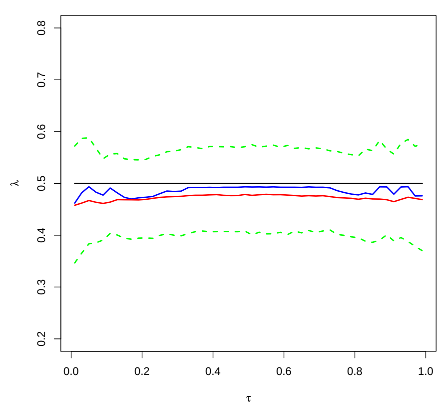
Figures S1-S3 respectively display simulation results of the weights and dependence parameters for Models 1-3 under Case I simulations. Figures S4-S6 respectively display simulation results of the weights and dependence parameters for Models 1-3 under Case II simulations. Figure S7 shows simulation results of the weights and dependence parameters when considering five candidate copulas (Gumbel, Frank, Clayton, rotated Gumbel, and rotated Clayton). 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 1000 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. The sample size is 800.



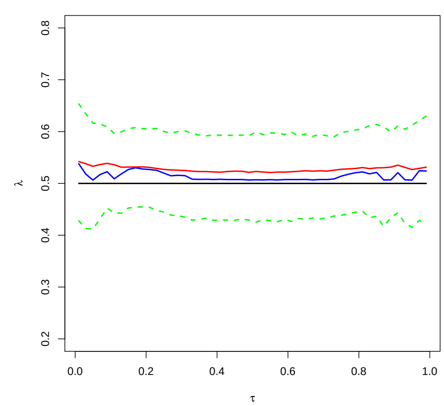
(a) Dependence parameter (Gumbel)



(b) Dependence parameter (Frank)

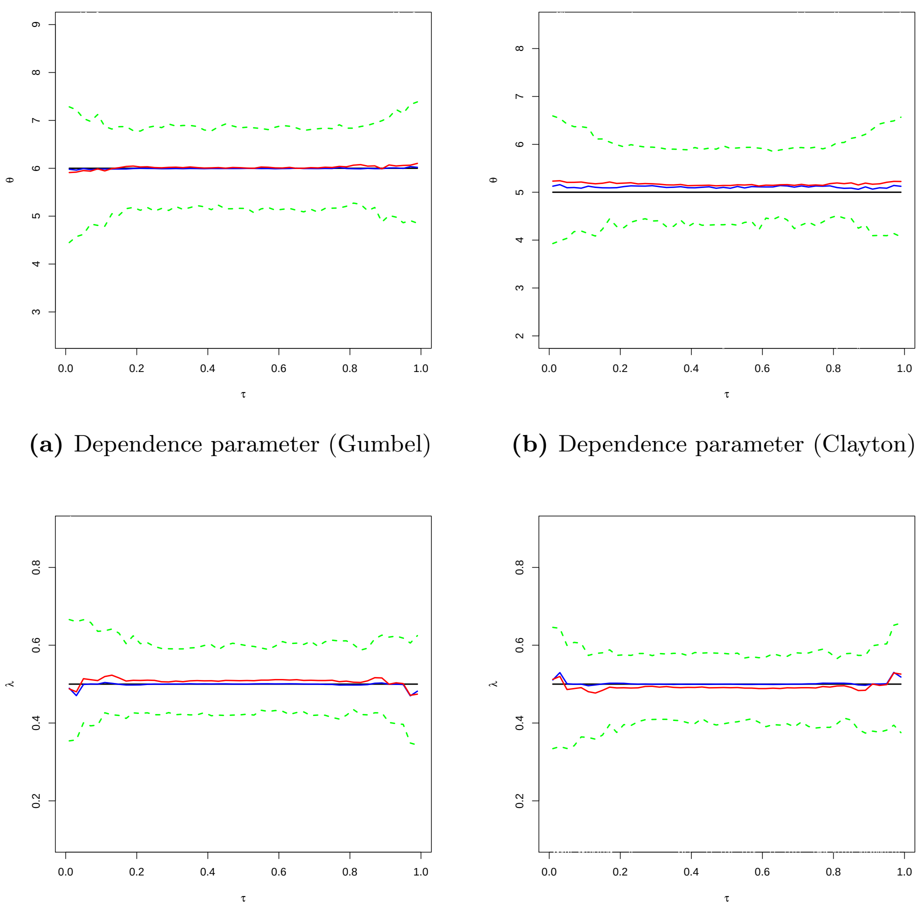


(c) Weight (Gumbel)



(d) Weight (Frank)

Figure S1: Simulation results of the weights and dependence parameters for Model 1 (1000 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.



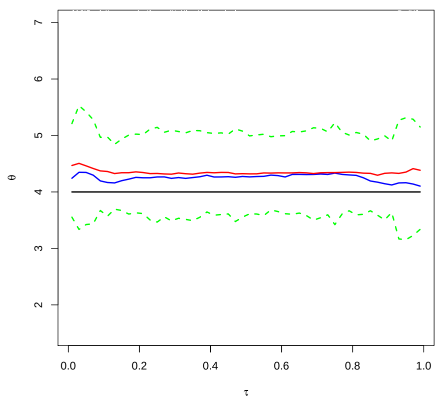
(a) Dependence parameter (Gumbel)

(b) Dependence parameter (Clayton)

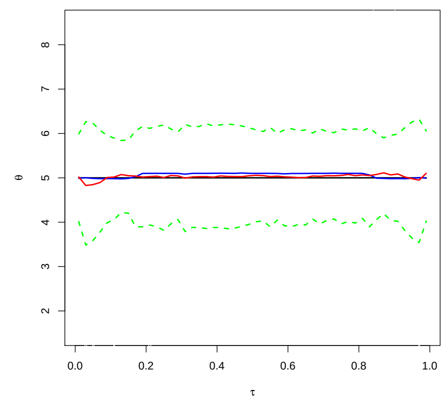
(c) Weight (Gumbel)

(d) Weight (Clayton)

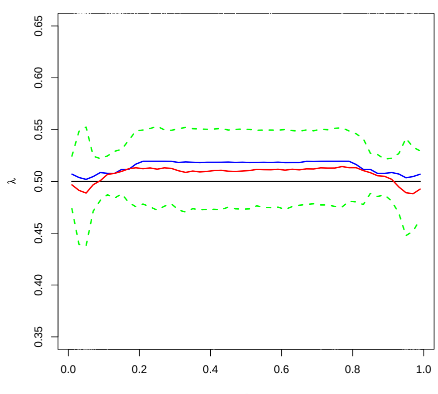
Figure S2: Simulation results of the weights and dependence parameters for Model 2 (1000 repeats) in Case I simulations: true values (black solid lines), mean and median estimates (red and blue lines), and 5% and 95% percentile curves (green dashed lines). The sample size is 800.



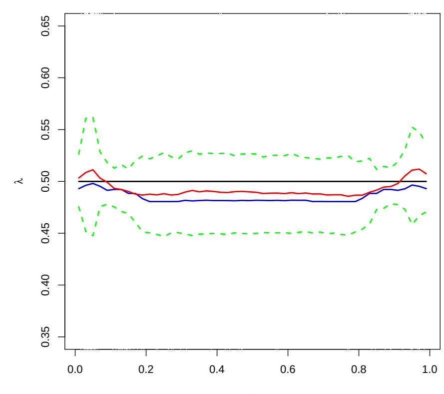
(a) Dependence parameter (Frank)



(b) Dependence parameter (Clayton)

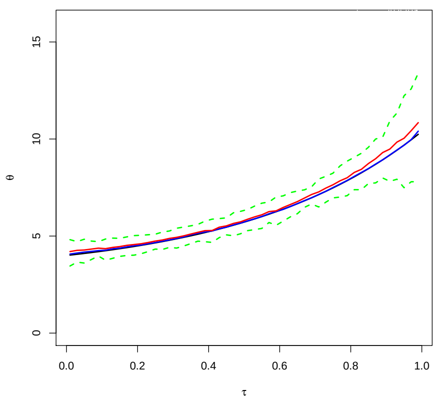


(c) Weight (Frank)

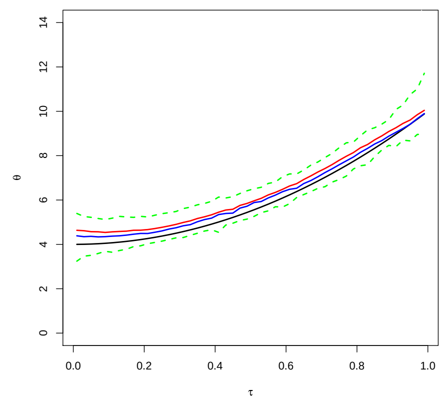


(d) Weight (Clayton)

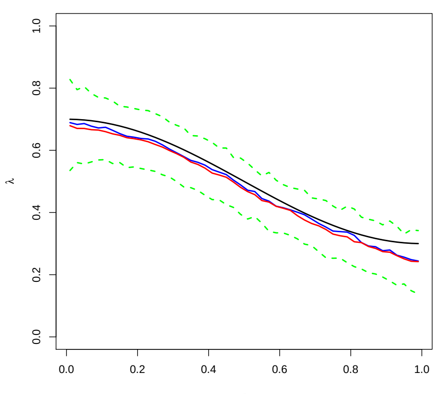
Figure S3: Simulation results of the weights and dependence parameters for Model 3 (1000 repeats) in Case I simulations: true values (black solid lines), mean and median estimates (red and blue lines), and 5% and 95% percentile curves (green dashed lines). The sample size is 800.



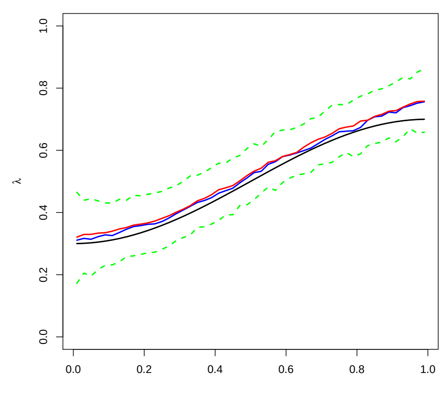
(a) Dependence parameter (Gumbel)



(b) Dependence parameter (Frank)

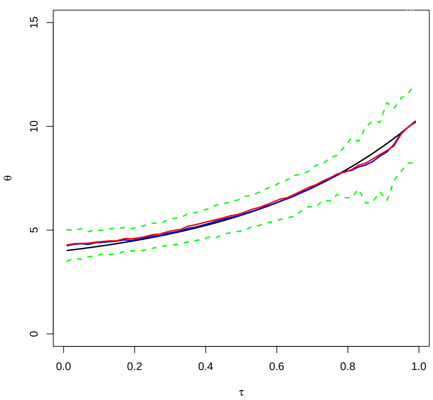


(c) Weight (Gumbel)

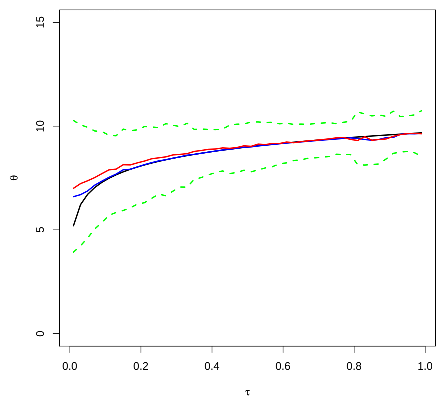


(d) Weight (Frank)

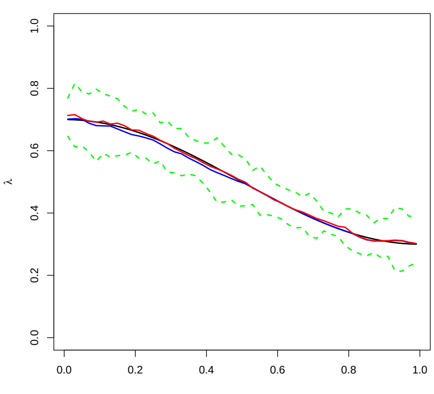
Figure S4: Simulation results of the weights and dependence parameters for Model 1 (1000 repeats) in Case II simulations: true values (black solid lines), mean and median estimates (red and blue lines), and 5% and 95% percentile curves (green dashed lines). The sample size is 800.



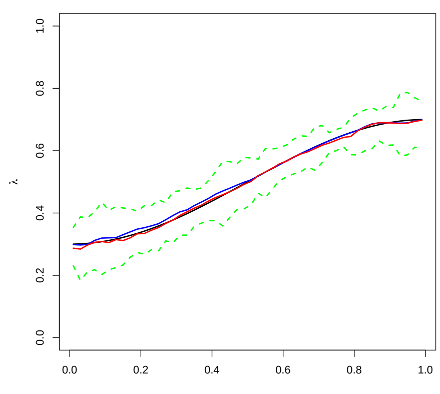
(a) Dependence parameter (Gumbel)



(b) Dependence parameter (Clayton)

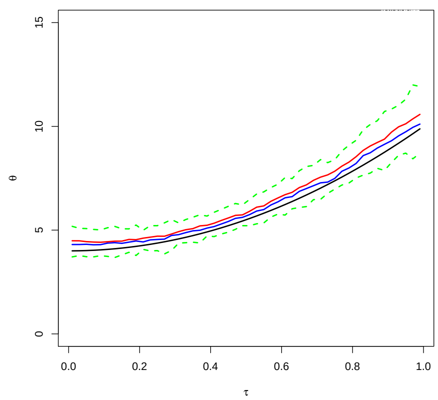


(c) Weight (Gumbel)

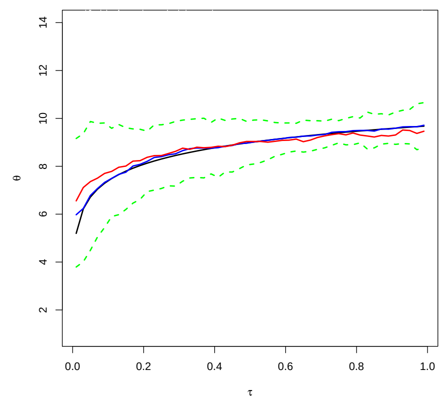


(d) Weight (Clayton)

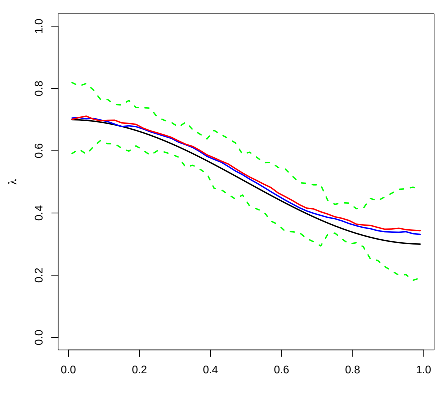
Figure S5: Simulation results of the weights and dependence parameters for Model 2 (1000 repeats) in Case II simulations: true values (black solid lines), mean and median estimates (red and blue lines), and 5% and 95% percentile curves (green dashed lines). The sample size is 800.



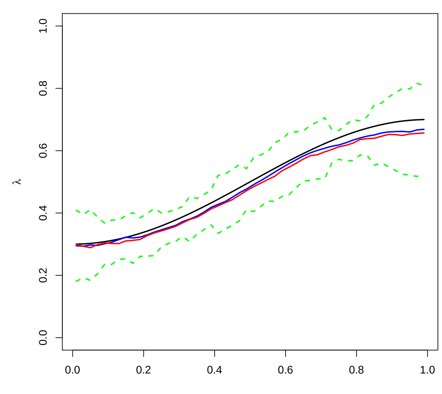
(a) Dependence parameter (Frank)



(b) Dependence parameter (Clayton)

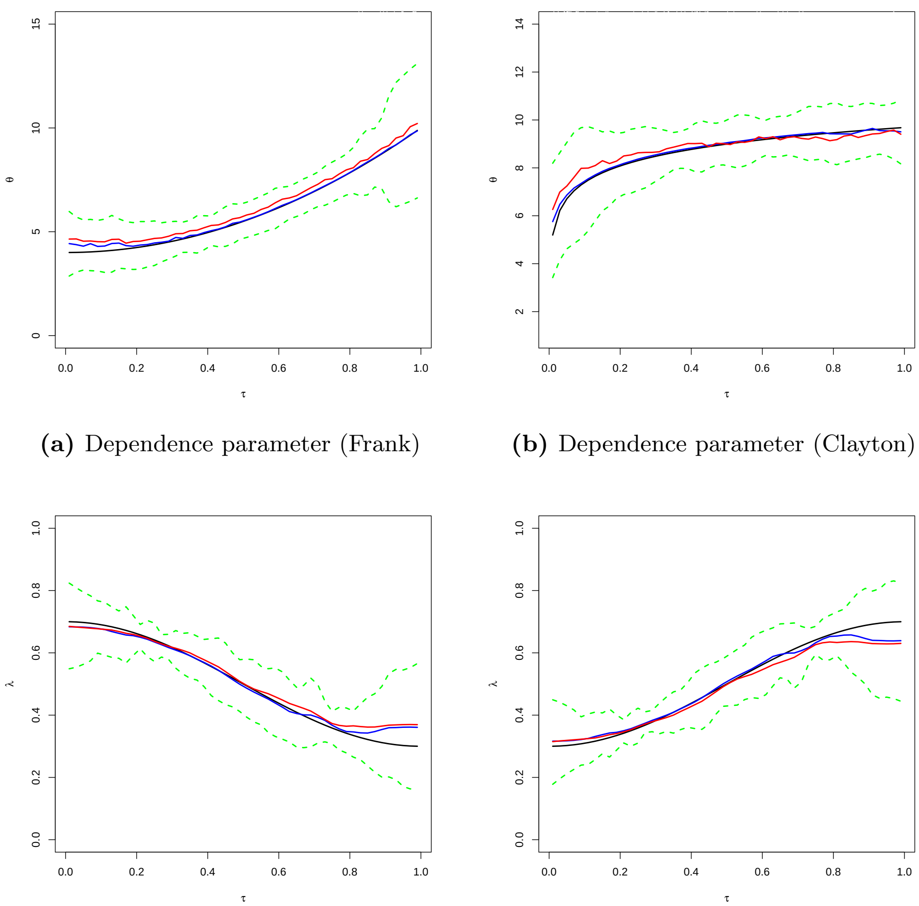


(c) Weight (Frank)



(d) Weight (Clayton)

Figure S6: Simulation results of the weights and dependence parameters for Model 3 (1000 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 S7: Simulation results of the weights and dependence parameters when considering five candidate copulas (Gumbel, Frank, Clayton, rotated Gumbel, and rotated Clayton): true values (black solid lines), mean and median estimates (red and blue lines), and 5% and 95% percentile curves (green dashed lines). The true model is a mixture copula of Clayton and Frank: $\lambda_1(\tau) = 0$, $\lambda_2(\tau) = 0.7 - 0.4\sin^2(\frac{\pi}{2}\tau)$, $\lambda_3(\tau) = 1-\lambda_2(\tau)$, $\lambda_4(\tau) = 0$, $\lambda_5(\tau) = 0$, $\theta_2(\tau) = 6\tau^2 + 4$, $\theta_3(\tau) = \ln(1 + \tau T) + 3$. The sample size is 800.

References

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.

Carrasco, M. and Chen, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. Econometric Theory, 18, 17-39.

Chen, X. and Fan, Y. (2006a). Estimation of copula-based semiparametric time series models. Journal of Econometrics, 130, 307-335.

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

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 Gijbels, I. (1996). Local polynomial modelling and its applications, Chapman & Hall, UK.

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.

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