

Skew-t Copula-Based Semiparametric Markov Chains

Leming QU

Boise State University, Boise, ID, USA

Abstract Without specifying the structure of a time series, we model the distribution of a multivariate Markov process in discrete time by the corresponding multivariate Markov family and the one-dimensional flows of marginal distributions. Such models tackle simultaneously temporal dependence and contemporaneous dependence between time series. A specific parametric form of stationary copula, namely skew-t copula, is assumed. Skew-t copulas are capable of modeling asymmetry, skewness, and heavy tails. An empirical study with unfiltered daily returns for three stock indices shows that the skew-t copula Markov model provides a better fit than the skew-Normal copula Markov or t-copula Markov model, and the skew-t copula model without Markov property.

Keywords multivariate Markov chain process; copula; skew-t distribution

MR(2010) Subject Classification 62M05; 62M10

1. Introduction

We study the estimation of a multivariate Markov process $\mathbf{X} \in \mathbb{R}^p$ in continuous state space based on the underlying Markov family of copulas, extending in particular the work of [1–3].

Considering $(\mathbf{X}_1, \dots, \mathbf{X}_n)$ - a p -variate Markov process \mathbf{X} observed at discrete time $t = 1, \dots, n$, we assume that the process is stationary, i.e., the distribution of $(\mathbf{X}_{t-1}, \mathbf{X}_t)$ does not depend on t . In addition we assume that (1) the dependence function of this $2p$ dimensional distribution is a skew-t copula, and (2) each of the marginal distributions of \mathbf{X}_t is independent of time and can be estimated by a nonparametric method before the copula parameter estimation stage.

General frameworks for modeling the dependence in a p -dimensional Markov Chain through copulas were presented in [1–3]. In these frameworks, a copula model is assumed to be a member of some specific parametric families determined by a parameter vector, such as meta-elliptical copulas [4] or skew-elliptical copulas [5].

Abegaz & Naik-Nimbalkar [1] assumes parametric models for the marginal distributions and presents two stage parametric pseudo-maximum likelihood estimation (2SPMLE) procedure for copula-based stationary Markov type models. In the first stage the marginal parameters are estimated by maximizing the marginal likelihood function, and in the second stage a pseudo-maximum likelihood estimation of the copula parameter is obtained after fixing the marginal

parameters at the values obtained from the first stage. The consistency and asymptotic normality of the marginal and copula parameter estimators from the 2SPMLE procedure is established.

For a bivariate stationary first-order Markov chains, Yi & Liao [2] assumes parametric models for the marginal distributions, a parametric copula model for temporal dependence in each univariate time series, and a parametric copula model for contemporaneous dependence between the two time series. A three stage parametric pseudo-maximum likelihood estimation (3SPMLE) procedure is presented for copula-based stationary Markov type models. In the first stage the marginal parameters are estimated by maximizing the marginal likelihood function. In the second stage a pseudo-maximum likelihood estimation of the temporal copula parameters is obtained after fixing the marginal parameters at the values obtained from the first stage. In the third stage, a pseudo-maximum likelihood estimation of the contemporaneous copula parameters is obtained after fixing the marginal parameters at the values obtained from the first stage and the temporal copula parameters obtained from the second stage. The consistency and asymptotic normality of the marginal, the temporal, and the contemporaneous copula parameter estimators from the 3SPMLE procedure is established. The 2SPMLE or 3SPMLE is a kind of estimators obtained by maximization by parts proposed by Song et. al. in [6]. The maximization by parts estimation method has recently been applied to integer-valued GARCH time series model [7].

Remillard et. al. [3] estimates a marginal distribution nonparametrically by its empirical CDF before the copula parameter estimation stage. Then the copula parameter is estimated by the maximum pseudo likelihood method. Under the Markovian models with meta-elliptic copulas, the conditional copulas are shown to be again meta-elliptic copulas. Under the Markovian models with Archimedean copulas, the conditional copulas are shown to be again Archimedean copulas. The consistency and asymptotic normality of the copula parameter estimators from this two stage semiparametric pseudo-maximum likelihood estimation procedure is established.

The main contribution of this article is to extend the skew-t copula to stationary Markov processes. The transition density of a skew-t copula Markov chain is derived, which is essential to the random number generation from the underlying stochastic processes and the calculation of the pseudo likelihood function. The computational difficulties for parameter estimation arising out of the model specification are addressed.

The rest of this paper is organized as follows. The skew-t copula model is defined in Section 2. The skew-t copula Markov model is presented in Section 3. Section 4 describes the computational difficulties for estimating the parameters of the skew-t copula Markov model by MLE and shows how to overcome them. Section 5 confirms by Monte-Carlo simulation that the nonlinear constrained MLE implemented by interior point algorithm works well. Section 6 applies the model to unfiltered daily returns of three stock indices: Nikkei225, S&P500 and DAX. Finally, Section 7 concludes the paper.

2. Skew-t copulas

Yoshihara [8] demonstrated that skew-t copulas provided good fit to data with heavy tail de-

pendence and tail asymmetry, and discussed their maximum likelihood estimation by overcoming the inherent numerical difficulties in the underlying optimization problem. The copula model in [8] only took the contemporaneous dependence between multiple time series into account. Our work here extends the skew-t copulas to Markov model - taking into account not only the contemporaneous dependence but also the temporal dependence between multiple time series.

The details of one of the three types of skew-t copulas - namely, the Azzalini and Capitanio (AC) [9] skew-t copula - are given in [8]. For completeness, we repeat some basic facts of an AC skew-t copula here.

The AC skew-t copula represents the dependence structure implicit in a d -variate AC skew-t distribution with the location vector, $\boldsymbol{\xi}^T = (\xi_1, \dots, \xi_d) = (0, \dots, 0) = \mathbf{0}_d^T$ and the scale vector, $\boldsymbol{\sigma}^T = (\sigma_1, \dots, \sigma_d) = (1, \dots, 1)$. AC skew-t distribution is a special case of an extended skew-t distribution, where the extension parameter τ in equation (1) of [10] is 0. The conditional distribution of extended skew-t random vectors is derived in [10]. The skew-t copula-based Markov model in this paper needs to utilize the conditional distribution of skew-t random vectors - for this reason, the notation and formulas for the AC skew-t distribution in this paper follow those in [10].

Definition 2.1 (Skew-t) *The d -dimensional random vector \mathbf{Y} of a skew-t distribution, denoted by $ST_d(\boldsymbol{\xi}, \Omega, \boldsymbol{\lambda}, \nu)$, has the following joint density function at $\mathbf{y} \in \mathbb{R}^d$:*

$$g(\mathbf{y}; \boldsymbol{\xi}, \Omega, \boldsymbol{\lambda}, \nu) = 2t_d(\mathbf{y}; \boldsymbol{\xi}, \Omega, \nu)T_1(\boldsymbol{\lambda}^T \mathbf{y} \sqrt{\frac{\nu + d}{\nu + Q(\mathbf{z})}}; \nu + d), \tag{2.1}$$

where $\mathbf{z} = \boldsymbol{\omega}^{-1}(\mathbf{y} - \boldsymbol{\xi})$, $Q(\mathbf{z}) = \mathbf{z}^T \bar{\Omega}^{-1} \mathbf{z}$, $\boldsymbol{\lambda} \in \mathbb{R}^d$ is the shape parameter,

$$t_d(\mathbf{y}; \boldsymbol{\xi}, \Omega, \nu) = \frac{\Gamma((\nu + d)/2)}{(\pi\nu)^{d/2} \Gamma(\nu/2) |\Omega|^{1/2}} \left(1 + \frac{Q(\mathbf{z})}{\nu}\right)^{-(\nu+d)/2}$$

denotes the usual d -dimensional Student- t density with location $\boldsymbol{\xi} \in \mathbb{R}^d$, positive definite $d \times d$ dispersion matrix Ω , with $d \times d$ scale matrix $\boldsymbol{\omega} = \text{diag}(\Omega)^{1/2}$ and correlation matrices $\bar{\Omega} = \boldsymbol{\omega}^{-1} \Omega \boldsymbol{\omega}^{-1}$, and the degrees of freedom $\nu > 0$. The $T_1(y; \nu)$ is the univariate standard Student- t cumulative distribution function with degrees of freedom $\nu > 0$. The d dimensional skew- t distribution function is denoted as $ST_d(\mathbf{y}; \boldsymbol{\xi}, \Omega, \boldsymbol{\lambda}, \nu)$.

For a Markov time series, we will model the joint distribution of the observation random vector at discrete time $(t - 1)$ and the random vector at discrete time t by a skew-t copula. The marginal distribution of the observation random vector at the start of the Markov time series $t = 1$, and the conditional distribution of the observation random vector at time t given the random vector at time $(t - 1)$ are needed to compute the likelihood function of the time series. An extended skew-t distribution (EST) introduced in [10] generalizes a skew-t distribution by adding the scalar extension parameter τ . A d -dimensional EST-with the location $\boldsymbol{\xi}$, the dispersion matrix Ω , the shape parameter $\boldsymbol{\lambda}$, the degrees of freedom ν , the extension parameter τ - is denoted as $EST_d(\boldsymbol{\xi}, \Omega, \boldsymbol{\lambda}, \nu, \tau)$. Its distribution theory is presented in [10]. By setting the extension parameter τ in the EST to 0 in [10, Propositions 3 and 4], we obtain the marginal and

conditional distribution of skew-t respectively as the following.

Proposition 2.2 (Marginal distribution of skew-t) *Let $\mathbf{Y} \sim ST_d(\boldsymbol{\xi}, \Omega, \boldsymbol{\lambda}, \nu)$. Consider the partition $\mathbf{Y}^T = (\mathbf{Y}_1^T, \mathbf{Y}_2^T)$ with $\dim(\mathbf{Y}_1) = d_1$, $\dim(\mathbf{y}_2) = d_2$, $d_1 + d_2 = d$, and the corresponding partition of the parameters $(\boldsymbol{\xi}, \Omega, \boldsymbol{\lambda})$. Then*

$$Y_1 \sim ST_{d_1}(\boldsymbol{\xi}_1, \Omega_{11}, \boldsymbol{\lambda}_{1(2)}, \nu),$$

where

$$\boldsymbol{\lambda}_{1(2)} = \frac{\boldsymbol{\lambda}_1 + \bar{\Omega}_{11}^{-1} \bar{\Omega}_{12} \boldsymbol{\lambda}_2}{\sqrt{1 + \boldsymbol{\lambda}_2^T \bar{\Omega}_{22 \cdot 1} \boldsymbol{\lambda}_2}}, \quad \bar{\Omega}_{22 \cdot 1} = \bar{\Omega}_{22} - \bar{\Omega}_{21} \bar{\Omega}_{11}^{-1} \bar{\Omega}_{12}.$$

Proposition 2.3 (Conditional distribution of skew-t) *Let $\mathbf{Y} \sim ST_d(\boldsymbol{\xi}, \Omega, \boldsymbol{\lambda}, \nu)$. Consider the partition $\mathbf{Y}^T = (\mathbf{Y}_1^T, \mathbf{Y}_2^T)$ with $\dim(\mathbf{Y}_1) = d_1$, $\dim(\mathbf{y}_2) = d_2$, $d_1 + d_2 = d$, and the corresponding partition of the parameters $(\boldsymbol{\xi}, \Omega, \boldsymbol{\lambda})$. Then*

$$(\mathbf{Y}_2 | \mathbf{Y}_1 = \mathbf{y}_1) \sim EST_{d_2}(\boldsymbol{\xi}_{2 \cdot 1}, \alpha_1 \Omega_{22 \cdot 1}, \boldsymbol{\lambda}_{2 \cdot 1}, \nu_{2 \cdot 1}, \tau_{2 \cdot 1}^*),$$

where

$$\begin{aligned} \boldsymbol{\xi}_{2 \cdot 1} &= \boldsymbol{\xi}_2 + \Omega_{21} \Omega_{11}^{-1} (\mathbf{y}_1 - \boldsymbol{\xi}_1), \quad Q_1(\mathbf{z}_1) = \mathbf{z}_1^T \bar{\Omega}_{11}^{-1} \mathbf{z}_1, \\ \mathbf{z}_1 &= \boldsymbol{\omega}_1^{-1} (\mathbf{y}_1 - \boldsymbol{\xi}_1), \quad \Omega_{22 \cdot 1} = \Omega_{22} - \Omega_{21} \Omega_{11}^{-1} \Omega_{12}, \\ \boldsymbol{\lambda}_{2 \cdot 1} &= \boldsymbol{\omega}_{2 \cdot 1} \boldsymbol{\omega}_2^{-1} \boldsymbol{\lambda}_2, \quad \boldsymbol{\omega}_2 = \text{diag}(\Omega_{22})^{1/2}, \\ \boldsymbol{\omega}_{2 \cdot 1} &= \text{diag}(\Omega_{22 \cdot 1})^{1/2}, \quad \nu_{2 \cdot 1} = \nu + d_1, \\ \tau_{2 \cdot 1} &= (\boldsymbol{\lambda}_2^T \bar{\Omega}_{21} \bar{\Omega}_{11}^{-1} + \boldsymbol{\lambda}_1^T) \mathbf{z}_1, \quad \tau_{2 \cdot 1}^* = \tau_{2 \cdot 1} / \sqrt{\alpha_1}, \\ \alpha_1 &= (\nu + Q_1(\mathbf{z}_1)) / (\nu + d_1). \end{aligned}$$

Unlike the marginal distribution of Y_2 , Arellano-Valle & Genton [10, section 2.2] note that the conditional distribution of a skew-t distribution is no longer a skew-t distribution - rather it is an extended skew-t distribution with an extension parameter $\tau_{2 \cdot 1}^* \neq 0$ in general.

ST distribution can be expressed as a scale mixture of the skew-normal distribution:

$$Y = \boldsymbol{\xi} + \boldsymbol{\omega} V^{-1/2} \tilde{\mathbf{Z}},$$

where $V \sim \Gamma(\nu/2, \nu/2)$ and is independent of the d -variate skew-Normal random vector $\tilde{\mathbf{Z}}$ which is constructed as

$$\tilde{\mathbf{Z}} = \begin{cases} \mathbf{Z}, & \text{if } Z_0 \geq 0, \\ -\mathbf{Z}, & \text{if } Z_0 < 0. \end{cases}$$

The $(1 + d)$ -dimensional random vector $(Z_0, \mathbf{Z}^T)^T$ has the $(1 + d)$ -variate Normal distribution $N_{1+d}(0, R)$. The extended correlation matrix R is defined by

$$R = \begin{bmatrix} 1 & \boldsymbol{\delta}^T \\ \boldsymbol{\delta} & \bar{\Omega} \end{bmatrix},$$

using the original correlation matrix $\bar{\Omega}$ and the skewness vector $\boldsymbol{\delta} = (\delta_1, \delta_2, \dots, \delta_d)^T$. The shape

parameter $\boldsymbol{\lambda}$ appearing in the ST density (2.1) is given as

$$\boldsymbol{\lambda} = \frac{\bar{\Omega}^{-1}\boldsymbol{\delta}}{\sqrt{1 - \boldsymbol{\delta}^T \bar{\Omega}^{-1}\boldsymbol{\delta}}}$$

using the skewness vector $\boldsymbol{\delta}$ and the original correlation matrix $\bar{\Omega}$. The vector $\boldsymbol{\delta}$ - and consequently $\boldsymbol{\lambda}$ - are constrained so that R is positive semi-definite.

As Azzalini & Capitanio [9, Section 4.2.3], Capitanio et al. [11, Eq. (10)] and Joe [12, p. 40] indicate, the i th univariate marginal distribution of $ST_d(\boldsymbol{\xi}, \Omega, \boldsymbol{\lambda}, \nu)$ is $ST_1(\xi_i, \sigma_i^2, \zeta_i, \nu)$ with density

$$g_1(y_i; \xi_i, \sigma_i^2, \zeta_i, \nu) = 2t_1(y_i; \xi_i, \sigma_i^2, \nu)T_1\left(\zeta_i y_i \sqrt{\frac{\nu + 1}{\nu + z_i^2}}; \nu + 1\right), \quad y_i \in \mathbb{R},$$

where $z_i = (y_i - \xi_i)/\sigma_i$, $t_1(\cdot; \xi_i, \sigma_i^2, \nu)$ is the univariate Student-t density with location ξ_i , scale σ_i , degrees of freedom ν , and ζ_i is defined as

$$\zeta_i = \delta_i / \sqrt{1 - \delta_i^2},$$

using the original skewness parameter δ_i .

Hence, applying Sklar’s theorem, we obtain the skew-t copula distribution function at $\mathbf{u} = (u_1, \dots, u_d)^T \in [0, 1]^d$:

$$C(\mathbf{u}; \bar{\Omega}, \boldsymbol{\lambda}, \nu) = ST_d(\mathbf{y}; \mathbf{0}_d, \bar{\Omega}, \boldsymbol{\lambda}, \nu).$$

where $\mathbf{y} = (y_1, \dots, y_d)^T$ with its element defined by

$$y_i = ST_1^{-1}(u_i; 0, 1, \zeta_i, \nu). \tag{2.2}$$

In words, y_i is the inverse probability integral transform of u_i , i.e., the $(100u_i)$ -th quantile - or equivalently, u_i is the probability integral transform of y_i - for the i th marginal distribution of $ST_d(\mathbf{0}_d, \bar{\Omega}, \boldsymbol{\lambda}, \nu)$.

The density function of this skew-t copula is given as

$$c(\mathbf{u}; \bar{\Omega}, \boldsymbol{\lambda}, \nu) = \frac{g(\mathbf{y}; \mathbf{0}_d, \bar{\Omega}, \boldsymbol{\lambda}, \nu)}{\prod_{i=1}^d g_1(y_i; 0, 1, \zeta_i, \nu)}. \tag{2.3}$$

3. Skew-t copulas for Markov models

We assume that the dependence structure of a p -variate time series with Markov property follows a skew-t copula.

Denoting by \mathbf{F} the transformation

$$\mathbf{x}_t = (x_{t1}, \dots, x_{tp})^T \mapsto \mathbf{F}(\mathbf{x}_t) = (F_1(x_{t1}), \dots, F_p(x_{tp}))^T = (u_{t1}, \dots, u_{tp})^T = \mathbf{u}_t,$$

one can then define $\mathbf{U}_t = \mathbf{F}(\mathbf{X}_t)$. Assuming a skew-t copula model:

$$(\mathbf{X}_{t-1}, \mathbf{X}_t) \sim H(\mathbf{x}_{t-1}, \mathbf{x}_t) = C(\mathbf{u}_{t-1}, \mathbf{u}_t; \tilde{\Omega}, \tilde{\boldsymbol{\lambda}}, \nu)$$

where

$$\tilde{\Omega} = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}, \text{ and } \tilde{\boldsymbol{\lambda}} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} \tag{3.1}$$

with $\Omega_{11} = \Omega_{22}$, $\boldsymbol{\delta}_1 = \boldsymbol{\delta}_2$, where $\tilde{\boldsymbol{\delta}} = (\boldsymbol{\delta}_1^T, \boldsymbol{\delta}_2^T)^T$ is the skewness vector corresponding to the shape parameter $\tilde{\boldsymbol{\lambda}}$. We note that $\Omega_{11} = \Omega_{22}$ encodes the contemporaneous correlation. The $\Omega_{12} = \Omega_{21}^T$ encodes the lag-1 cross-correlation [13, Section 8.1] - a specific temporal correlation - between time series. Under this joint skew-t copula model of $(\mathbf{X}_{t-1}, \mathbf{X}_t)$, the copula of \mathbf{X}_{t-1} must be the same as the copula of \mathbf{X}_t , which is another skew-t copula specified as:

$$C(\mathbf{u}, \mathbf{1}_p; \tilde{\Omega}, \tilde{\boldsymbol{\lambda}}, \nu) = C(\mathbf{1}_p, \mathbf{u}; \tilde{\Omega}, \tilde{\boldsymbol{\lambda}}, \nu) = C(\mathbf{u}; \Omega_{11}, \boldsymbol{\lambda}_{1(2)}, \nu)$$

where

$$\boldsymbol{\lambda}_{1(2)} = \frac{\boldsymbol{\lambda}_1 + \Omega_{11}^{-1} \Omega_{12} \boldsymbol{\lambda}_2}{\sqrt{1 + \boldsymbol{\lambda}_2^T \Omega_{22 \cdot 1} \boldsymbol{\lambda}_2}}, \quad \Omega_{22 \cdot 1} = \Omega_{22} - \Omega_{21} \Omega_{11}^{-1} \Omega_{12}$$

according to Proposition 2.2 (Marginal distribution of skew-t).

Now under the AC skew-t copula Markov model

$$(\mathbf{U}_{t-1}, \mathbf{U}_t) \sim C(\mathbf{u}_{t-1}, \mathbf{u}_t; \tilde{\Omega}, \tilde{\boldsymbol{\lambda}}, \nu), \tag{3.2}$$

we derive that

$$\mathbf{U}_t \sim C(\mathbf{u}_t; \Omega_{11}, \boldsymbol{\lambda}_{1(2)}, \nu).$$

To simulate observations for the Markov process \mathbf{U}_t , one needs to compute the conditional distribution of \mathbf{U}_t given \mathbf{U}_{t-1} , which is readily available as following based on the Proposition 2.3 (Conditional distribution of skew-t):

Proposition 3.1 (Transition density of the Markov Chain) *The conditional density of \mathbf{U}_t given $\mathbf{U}_{t-1} = \mathbf{u}_{t-1}$ is:*

$$\begin{aligned} c(\mathbf{u}_t | \mathbf{u}_{t-1}; \tilde{\Omega}, \tilde{\boldsymbol{\lambda}}, \nu) &= \frac{c(\mathbf{u}_{t-1}, \mathbf{u}_t; \tilde{\Omega}, \tilde{\boldsymbol{\lambda}}, \nu)}{c(\mathbf{u}_{t-1}; \Omega_{11}, \boldsymbol{\lambda}_{1(2)}, \nu)} \\ &= \frac{g(\mathbf{y}_{t-1}, \mathbf{y}_t; \tilde{\Omega}, \tilde{\boldsymbol{\lambda}}, \nu)}{g(\mathbf{y}_{t-1}; \Omega_{11}, \boldsymbol{\lambda}_{1(2)}, \nu) \prod_{i=1}^p g_1(y_{it}; 0, 1, \tilde{\zeta}_i, \nu)} \\ &= \frac{g_{EST}(\mathbf{y}_t; \boldsymbol{\xi}_{2 \cdot 1}^{(t)}, \alpha_t \Omega_{22 \cdot 1}, \boldsymbol{\lambda}_{2 \cdot 1}, \nu_{2 \cdot 1}, \tau_{2 \cdot 1}^{*(t)})}{\prod_{i=1}^p g_1(y_{it}; 0, 1, \tilde{\zeta}_i, \nu)}, \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{\xi}_{2 \cdot 1}^{(t)} &= \Omega_{21} \Omega_{11}^{-1} \mathbf{y}_{t-1}, \quad \Omega_{22 \cdot 1} = \Omega_{22} - \Omega_{21} \Omega_{11}^{-1} \Omega_{12}, \\ \boldsymbol{\lambda}_{2 \cdot 1} &= \omega_{2 \cdot 1} \boldsymbol{\lambda}_2, \quad \omega_{2 \cdot 1} = \text{diag}(\Omega_{22 \cdot 1})^{1/2}, \\ \nu_{2 \cdot 1} &= \nu + p, \quad \tau_{2 \cdot 1}^{(t)} = (\boldsymbol{\lambda}_2^T \Omega_{21} \Omega_{11}^{-1} + \boldsymbol{\lambda}_1^T) \mathbf{y}_{t-1}, \\ \tau_{2 \cdot 1}^{*(t)} &= \tau_{2 \cdot 1}^{(t)} / \sqrt{\alpha_t}, \quad \alpha_t = (\nu + \mathbf{y}_{t-1}^T \Omega_{11}^{-1} \mathbf{y}_{t-1}) / (\nu + p), \\ \mathbf{u}_t &= \mathbf{F}_{ST}(\mathbf{y}_t, \mathbf{0}_p, \Omega_{11}, \boldsymbol{\lambda}_{1(2)}, \nu), \end{aligned} \tag{3.3}$$

and the i th element of $\mathbf{F}_{ST}(\mathbf{y}_t, \mathbf{0}_p, \Omega_{11}, \boldsymbol{\lambda}_{1(2)}, \nu)$ is the probability integral transform of y_{it} - where the transform function is the i th marginal distribution of $ST_p(\mathbf{0}_p, \Omega_{11}, \boldsymbol{\lambda}_{1(2)}, \nu)$. The $\tilde{\zeta}_i$ dependent on Ω_{11} , $\boldsymbol{\lambda}_{1(2)}$ is the shape parameter of this i -th marginal distribution.

Using the distribution for \mathbf{U}_t and Proposition 3.1, one can now propose an algorithm to simulate a Markovian time series with a joint skew-t copula.

Algorithm. To generate a Markov chain $\mathbf{U}_1, \dots, \mathbf{U}_n$ with stationary distribution $C(\Omega_{11}, \boldsymbol{\lambda}_{1(2)}, \nu)$ and joint distribution $(\mathbf{U}_{t-1}, \mathbf{U}_t) \sim C(\tilde{\Omega}, \tilde{\boldsymbol{\lambda}}, \nu)$, proceed as follows:

- (1) Generate $\mathbf{Y}_1 \sim ST_p(\mathbf{0}_p, \Omega_{11}, \boldsymbol{\lambda}_{1(2)}, \nu)$, set $\mathbf{U}_1 = \mathbf{F}_{ST}(\mathbf{Y}_1; \mathbf{0}_p, \Omega_{11}, \boldsymbol{\lambda}_{1(2)}, \nu)$.
- (2) For $t = 2, \dots, n$,
 - (a) Update $\tau_{2,1}^{*(t)}$ according to Eq. (3.3)
 - (b) Generate \mathbf{Y}_t according to $(\mathbf{Y}_t | \mathbf{Y}_{t-1} = \mathbf{y}_{t-1}) \sim EST_p(\boldsymbol{\xi}_{2,1}^{(t)}, \alpha_t \Omega_{22,1}, \boldsymbol{\lambda}_{2,1}, \nu_{2,1}, \tau_{2,1}^{*(t)})$. Set $\mathbf{U}_t = \mathbf{F}_{ST}(\mathbf{Y}_t; \mathbf{0}_p, \Omega_{11}, \boldsymbol{\lambda}_{1(2)}, \nu)$.

4. Estimation by the maximum pseudo likelihood method

Given a time series of p -dimensional vectors $\mathbf{x}_t = (x_{t1}, \dots, x_{tp})^T$ with $t = 1, \dots, n$ for which the copula associated with $(\mathbf{X}_{t-1}, \mathbf{X}_t)$ is $C(\mathbf{u}_{t-1}, \mathbf{u}_t; \tilde{\Omega}, \tilde{\boldsymbol{\lambda}}, \nu)$, we wish to estimate model parameters $(\Omega_{11}, \Omega_{12}, \boldsymbol{\delta}_1, \nu)$ without assuming any parametric models for the marginal distributions.

We take the usual approach to estimate a marginal distribution $F_i(x_i)$ by an appropriate parametric estimator, or a nonparametric estimator - such as a kernel estimator [14], a local linear estimator [15], or the classical and simple empirical cumulative distributional function (ECDF). In the case of ECDF, the $u_{it} = F_i(x_{it})$ is estimated by the normalized rank of x_{it} among x_{i1}, \dots, x_{in} , that is $\hat{\mathbf{u}}_t = \text{rank}(\mathbf{x}_t)/(n + 1)$.

As an extension of the maximum pseudo likelihood method [16] to the Markovian case, the maximum pseudo likelihood estimator of $(\tilde{\Omega}, \tilde{\boldsymbol{\delta}}, \nu)$ is defined by the maximizer of the pseudo log-likelihood function:

$$l(\tilde{\Omega}, \tilde{\boldsymbol{\delta}}, \nu; \hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_n) = \log c(\hat{\mathbf{u}}_1; \Omega_{11}, \boldsymbol{\lambda}_{1(2)}, \nu) + \sum_{t=2}^n \log c(\hat{\mathbf{u}}_t | \hat{\mathbf{u}}_{t-1}; \tilde{\Omega}, \tilde{\boldsymbol{\lambda}}, \nu). \tag{4.1}$$

When maximizing this pseudo log-likelihood function, we face the same two problems as stated in [8, Section 3.1]. First, computing the log-likelihood function given in Eq. (4.1) requires computing univariate ST quantile functions, as shown in Eq. (2.2). The brute force computation of the $(100\hat{u}_{it})$ th quantile of the i -th marginal distribution of $ST_p(\mathbf{0}_p, \Omega_{11}, \boldsymbol{\lambda}_{1(2)}, \nu)$ is to carry out the inverse probability integral transform for each \hat{u}_{it} with $i = 1, \dots, p$ and $t = 1, \dots, n$, resulting in $p \times n$ inverse probability integral transforms in total. If implemented without resorting to a fast algorithm such as a monotone interpolator as in [8], each of these $p \times n$ inverse probability integral transforms needs to numerically solve a root finding problem involving integral equation - which is time consuming. The second problem is that the correlation matrix $\tilde{\Omega}$ in Eq. (3.1) should be positive semi-definite and numerical optimization which directly enforces such constraints can be complicated.

To overcome the first problem - computational burden in finding quantiles, Azzalini and Capitanio [9] used a monotone interpolator implemented by a piecewise cubic Hermite interpolating polynomial. Note that the i th variate of the pseudo sample has the values of $(\hat{u}_{i1}, \dots, \hat{u}_{in})$. Let

$$p_{i1} = \min_{t=1, \dots, n} \hat{u}_{it} \text{ and } p_{in} = \max_{t=1, \dots, n} \hat{u}_{it}.$$

For the given set of p -variate ST parameters $(\Omega_{11}, \boldsymbol{\lambda}_{1(2)}, \nu)$, for its i -th marginal distribution

$ST_1(0, 1, \tilde{\zeta}_i, \nu)$, we obtain the $(100p_{i1})$ -th and $(100p_{im})$ -th quantiles

$$\tilde{y}_{i1} = ST_1^{-1}(p_{i1}; 0, 1, \tilde{\zeta}_i, \nu) \text{ and } \tilde{y}_{im} = ST_1^{-1}(p_{im}; 0, 1, \tilde{\zeta}_i, \nu)$$

by numerical root finding algorithm such as Newton-Raphson iterations to solve for y for given $p = p_{i1}$ and $p = p_{im}$ in the problem $ST_1(y; 0, 1, \tilde{\zeta}_i, \nu) - p = 0$, respectively. We then choose the interpolating quantiles \tilde{y}_{ik} 's by m equally spaced points in $[\tilde{y}_{i1}, \tilde{y}_{im}]$:

$$\tilde{y}_{ik} = \tilde{y}_{i1} + \frac{k-1}{m-1}(\tilde{y}_{im} - \tilde{y}_{i1}), \quad k = 1, \dots, m,$$

and calculate their corresponding probability integral transform

$$\tilde{p}_{ik} = ST_1(\tilde{y}_{ik}; 0, 1, \tilde{\zeta}_i, \nu), \quad k = 1, \dots, m.$$

Numerical experiments show that the monotone interpolator computes the quantiles efficiently and accurately.

To overcome the second problem - complexity in ensuring positive semi-definiteness of the correlation matrix, Yoshiba [8] used the hyperspherical reparameterization of its Cholesky factor [17–20]. For a normal or a multivariate Student-t distribution model, the consistency and asymptotic normality of the maximum likelihood estimators of the hyperspherical coordinates or angles for a structured correlation matrix were established in [20]. This reparameterization method is adopted here for the AC skew-t Markov model parameter estimation.

4.1. A monotone interpolator for fast computing of univariate ST quantiles

The pseudo log-likelihood function in terms of Eq. (4.1) can be expressed as

$$\begin{aligned} & l(\tilde{\Omega}, \tilde{\delta}, \nu; \hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_n) \\ &= \sum_{t=2}^n \log \frac{g(\mathbf{y}_{t-1}, \mathbf{y}_t; \tilde{\Omega}, \tilde{\lambda}, \nu)}{g(\mathbf{y}_{t-1}; \Omega_{11}, \boldsymbol{\lambda}_{1(2)}, \nu) \prod_{i=1}^p g_1(y_{it}; 0, 1, \tilde{\zeta}_i, \nu)} \\ &= \sum_{t=2}^n \left\{ \log g(\mathbf{y}_{t-1}, \mathbf{y}_t; \tilde{\Omega}, \tilde{\lambda}, \nu) - \log g(\mathbf{y}_{t-1}; \Omega_{11}, \boldsymbol{\lambda}_{1(2)}, \nu) - \sum_{i=1}^p \log g_1(y_{it}; 0, 1, \tilde{\zeta}_i, \nu) \right\}. \end{aligned}$$

For this formulation, only the set of quantiles $\{\mathbf{y}_t, t = 1, \dots, n\}$ of the univariate marginal distributions of $ST_p(\Omega_{11}, \boldsymbol{\lambda}_{1(2)}, \nu)$ are needed, for which the monotone interpolator can be used to compute them.

4.2. Hyperspherical reparameterization of the Cholesky factor of the extended correlation matrix

Let \tilde{R} denote the extended correlation matrix of the AC skew-t Markov model (3.2):

$$\tilde{R} = \begin{bmatrix} 1 & \tilde{\delta}^T \\ \tilde{\delta} & \tilde{\Omega} \end{bmatrix}.$$

Its symmetry and positive semi-definiteness allow a Cholesky decomposition: $\tilde{R} = LL^T$, where

L is a lower triangular matrix with all elements in $[-1, 1]$, given as

$$L = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ l_{21} & l_{22} & 0 & \dots & 0 \\ l_{31} & l_{32} & l_{33} & \dots & 0 \\ l_{41} & l_{42} & l_{43} & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ l_{2p+1,1} & l_{2p+1,2} & l_{2p+1,3} & \dots & l_{2p+1,2p+1} \end{bmatrix}.$$

The hyperspherical reparameterization of L in matrix form is

$$L = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ \cos \theta_{21} & \sin \theta_{21} & 0 & \dots & 0 \\ \cos \theta_{31} & \cos \theta_{32} \sin \theta_{31} & \sin \theta_{32} \sin \theta_{31} & \dots & 0 \\ \cos \theta_{41} & \cos \theta_{42} \sin \theta_{41} & \cos \theta_{43} \sin \theta_{42} \sin \theta_{41} & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \cos \theta_{2p+1,1} & \cos \theta_{2p+1,2} \sin \theta_{2p+1,2} & \cos \theta_{2p+1,3} \sin \theta_{2p+1,2} \sin \theta_{2p+1,1} & \dots & \prod_{k=1}^{2p} \sin \theta_{2p+1,k} \end{bmatrix}.$$

The angles θ_{ij} measured in radians for $i > j$ are required to be restricted to the range $(0, \pi)$ so that the \tilde{R} has positive diagonal entries and hence its Cholesky factor L is unique [19].

4.3. Constrained maximum pseudo log-likelihood estimator

The AC Skew-t Markov model imposes the constraints $\Omega_{11} = \Omega_{22}$, $\delta_1 = \delta_2$ on the model parameters $(\tilde{\Omega}, \tilde{\delta}, \nu)$ due to the Markov property. But to ensure the positive semi-definiteness of the extended correlation matrix \tilde{R} , the hyperspherical reparameterization of its Cholesky factor L is utilized. The original model parameter $\tilde{\Omega}$ and $\tilde{\delta}$ are nonlinear functions of the hyperspherical coordinates $\theta = (\theta_{ij}, 1 \leq j < i \leq (2p + 1))$. Therefore the constraints $\Omega_{11} = \Omega_{22}$, $\delta_1 = \delta_2$ are nonlinear in terms of θ .

The maximum pseudo log-likelihood estimator $\hat{\theta}$ and $\hat{\nu}$ in constrained parameter spaces maximizes the pseudo log-likelihood

$$l(\theta, \nu; \hat{\mathbf{u}}_1, \dots, \hat{\mathbf{u}}_n) = \sum_{t=2}^n \left\{ \log g(\mathbf{y}_{t-1}, \mathbf{y}_t; \tilde{\Omega}, \tilde{\lambda}, \nu) - \log g(\mathbf{y}_{t-1}; \Omega_{11}, \lambda_{1(2)}, \nu) - \sum_{i=1}^p \log g_1(y_{it}; 0, 1, \tilde{\zeta}_i, \nu) \right\}$$

Subject to

$$0 < \theta_{ij} < \pi, \quad \text{for } 1 \leq j < i \leq (2p + 1),$$

$$\Omega_{11} = \Omega_{22}, \quad \delta_1 = \delta_2, \quad \nu > 0.$$

This nonlinear constrained optimization problem can be solved by the interior point algorithm [21]. MATLAB optimization toolbox's `fmincon()` function with interior point algorithm option [22] is called to solve the above problem.

5. Monte-Carlo simulation

The benchmark parameters in a trivariate AC skew-t copula Markov Model for Monte-Carlo simulation listed in Table 1 is based on the estimated parameters for the empirical data analysis in the next section. The Nikkei225, S&P500 and DAX daily return data $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ from 1 April 2010 to 31 March 2015 are used to estimate the pseudo observations $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$. The pseudo observations $\{u_{j1}, \dots, u_{jn}\}$ ($j = 1, 2, 3$) are estimated by the normalized ranks of $\{x_{j1}, \dots, x_{jn}\}$.

Σ_{11}	Σ_{12}	δ	ν
$\begin{bmatrix} 1 & 0.48 & 0.40 \\ 0.48 & 1 & 0.62 \\ 0.40 & 0.62 & 1 \end{bmatrix}$	$\begin{bmatrix} -0.04 & -0.04 & 0.01 \\ 0.14 & -0.09 & -0.04 \\ 0.28 & 0.16 & 0.01 \end{bmatrix}$	$\begin{bmatrix} -0.21 \\ 0.03 \\ -0.15 \end{bmatrix}$	4.34

Table 1 Benchmark parameters in a trivariate AC skew-t copula Markov model

n	RMSE of the lower triangular matrix of $\hat{\Sigma}_{11}$	RMSE of $\hat{\Sigma}_{12}$	RMSE of $\hat{\delta}$	RMSE of $\hat{\nu}$
500	$\begin{bmatrix} 0.0392 \\ 0.0433 & 0.0309 \end{bmatrix}$	$\begin{bmatrix} 0.0598 & 0.0501 & 0.0565 \\ 0.0478 & 0.0506 & 0.0516 \\ 0.0457 & 0.0490 & 0.0580 \end{bmatrix}$	$\begin{bmatrix} 0.1435 \\ 0.1365 \\ 0.1427 \end{bmatrix}$	0.4638
2000	$\begin{bmatrix} 0.0198 \\ 0.0225 & 0.0151 \end{bmatrix}$	$\begin{bmatrix} 0.0296 & 0.0250 & 0.0279 \\ 0.0253 & 0.0241 & 0.0258 \\ 0.0251 & 0.0244 & 0.0292 \end{bmatrix}$	$\begin{bmatrix} 0.0696 \\ 0.0704 \\ 0.0731 \end{bmatrix}$	0.2332

Table 2 Root mean squared errors of the parameter estimates displayed in the shape of their corresponding parameters for the AC skew-t copula Markov model

Two different sample sizes $n = 500, 2000$ are considered. In one replication of the simulation, for the benchmark parameters at the given sample size, we generate the pseudo sample $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ using the Algorithm discussed in Section 3. The initial value for $\tilde{\Omega}$ is the identity matrix. The initial value for $\tilde{\delta}$ is the zero vector. The initial value for ν is 8. For each sample size setting, we replicate the experiment 500 times. The quality of a copula parameter estimate, say $\hat{\nu}$, of the true copula parameter, say ν , is measured by the root mean squared error:

$$RMSE(\hat{\nu}) = \sqrt{\frac{1}{500} \sum_{i=1}^{500} (\hat{\nu}^{(i)} - \nu)^2},$$

where $\hat{\nu}^{(i)}$ is the parameter estimate for the i th replication of the simulation. In Table 2, the RMSE of the parameter estimates decreases as the sample size n increases - which is expected. The RMSE of the skewness parameter estimator $\hat{\delta}$ are larger than those of the correlation parameter estimators $\hat{\rho}_{ij}$, especially for $n = 500$, but they follow the general pattern of decreasing with increasing sample size.

6. Empirical data analysis

We apply the proposed method to estimate the trivariate AC skew-t copula parameter for Nikkei225, S&P500 and DAX daily return data under Markov Model. For unfiltered returns we compare the fits of the AC skew-t copula, skew-Normal copula [23] and t-copula under Markov Model with the AC skew-t copula under iid assumption.

Table 3 shows the Log-likelihood, AIC and BIC of the 4 models for unfiltered five-year daily returns from 1 April 2005 to 31 March 2015 (sample size $n = 1188$). The corresponding Table 4 is for unfiltered 10-year daily returns from 1 April 2005 to 31 March 2015 ($n = 2367$). In both Tables 3 and 4, the AC skew-t copula Markov model attains the lowest AIC or BIC values among the 4 copula models, thus is the best model among them in terms of AIC or BIC.

	Skew-t Markov	Skew-Normal Markov	Student-t Markov	Skew-t iid
Log-likelihood	696.5	595.5	683.6	526.9
AIC	-1361.0	-1161.0	-1341.2	-1043.8
BIC	-1279.7	-1084.8	-1275.2	-1018.4

Table 3 Copula Model fit comparison for daily return data from 1 April 2010 to 31 March 2015

	Skew-t Markov	Skew-Normal Markov	Student-t Markov	Skew-t iid
Log-likelihood	1626.7	1203.1	1596.8	1108.8
AIC	-3221.3	-2376.2	-3167.7	-2207.6
BIC	-3129.0	-2289.6	-3092.7	-2178.7

Table 4 Copula Model fit comparison for daily return data from 1 April 2005 to 31 March 2015

7. Conclusion

This paper extends the skew-t copula model in [8] to a multivariate Markov process in continuous state space [3]. In the empirical study with unfiltered daily returns for the three stock indices - Nikkei225, S&P500 and DAX - we show that the AC skew-t copula Markov model provides a better fit than the skew-Normal copula Markov or t-copula Markov model, and AC skew-t copula model without Markov property.

Recently, D'Amico and Petroni [24] applied copula based multivariate semi-Markov models to high-frequency finance. In particular, they used the Gumbel copula to preserve the cross-correlation between time series. It would be interesting to apply the skew-t copula Markov model proposed in this article to high-frequency finance.

Acknowledgment I thank the anonymous reviewers for constructive comments which have helped me to improve the paper.

For the Monte-Carlo simulation, I would like to acknowledge high-performance computing support of the R2 computer cluster [25] provided by Boise State University's Research Computing Department.

References

- [1] F. ABEGAZ, U. NAIK-NIMBALKAR. *Modeling statistical dependence of Markov chains via copula models*. J. Stat. Plann. Inference, 2008, **138**(4): 1131–1146.
- [2] Wende YI, S. S. LIAO. *Statistical properties of parametric estimators for Markov chain vectors based on copula models*. J. Stat. Plann. Inference, 2010, **140**(6): 1465–1480.
- [3] B. RÉMILLARD, N. PAPAGEORGIOU, F. SOUSTRA. *Copula-based semiparametric models for multivariate time series*. J. Multivar. Anal., 2012, **110**: 30–42.
- [4] Hongbin FANG, Kaitai FANG, S. KOTZ. *The meta-elliptical distributions with given marginals*. J. Multivar. Anal., 2002, **82**(1): 1–16.
- [5] H. JOE, Haijun LI. *Tail densities of skew-elliptical distributions*. J. Multivar. Anal., 2019, **171**: 421–435.
- [6] P. X. K. SONG, Yanqin FAN, J. D. KALBFLEISCH. *Maximization by parts in likelihood inference*. J. Amer. Stat. Assoc., 2005, **100**(472): 1145–1158.
- [7] Yan CUI, Qi LI, Fukang ZHU. *Flexible bivariate Poission integer-valued GARCH model*. Ann. of the Inst. Stat. Math., 2019, 1–29.
- [8] T. YOSHIBA. *Maximum likelihood esimaion of skew-t copulas with its applications to stock returns*. J. Stat. Comput. Simu., 2018, **88**(13): 2489–2506.
- [9] A. AZZALINI, A. CAPITANIO. *Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution*. J. R. Stat. Soc. Ser. B Stat. Methodol., 2003, **65**(2): 367–389.
- [10] R. B. ARELLANO-VALLE, M. G. GENTON. *Multivariate extended skew-t distributions and related families*. Metron., 2010, **68**(3): 201–234.
- [11] A. CAPITANIO, A. AZZALINI, E. STANGHELLINI. *Graphical models for skew-normal variates*. Scand. J. Stat., 2003, **30**(1): 129–144.
- [12] H. JOE. *Discussion of: “Copulas: tales and facts” by T. Mikosch*. Extremes, 2006, **9**(1): 37–41.
- [13] R. S. TSAY. *Analysis of Financial Time Series. Third Edition*. Wiley & Sons, Inc., Hoboken, NJ, 2010.
- [14] C. TENREIRO. *A new class of boundary kernels for distribution function estimation*. Comm. Stat. Theory Methods, 2018, **47**(21): 5319–5332.
- [15] M. Y. CHENG, Liang PENG. *Regression modeling for nonparametric estimation of distribution and quantile functions*. Stat. Sinica, 2002, **12**(4): 1043–1060.
- [16] C. GENEST, K. GHOUDI, L.-P. RIVEST. *A semiparametric estimation procedure of dependence parameters in multivariate families of distributions*. Biometrika, 1995, **82**(3): 543–552.
- [17] J. C. PINHEIRO, D. M. BATES. *Unconstrained parametrizaions for variance-coariance matrices*. Stat. Comput., 1996, **6**(3): 289–296.
- [18] F. RAPISARDA, D. BRIGO, F. MERCURIO. *Parameterizing correlations: a geometric interpretation*. IMA J. Manag. Math., 2007, **18**(1): 55–73.
- [19] M. POURAHMADI, Xiao WANG. *Distribution of random correlation matrices: hyperspherical parameterization of the Cholesky factor*. Stat. Probab. Lett., 2015, **106**: 5–12.
- [20] R. S. TSAY, M. POURAHMADI. *Modelling structured correlation matrices*. Biometrika, 2017, **104**(1): 237–242.
- [21] M. H. WRIGHT. *Interior Methods for Constrained Optimization*. Cambridge Univ. Press, Cambridge, 1992.
- [22] MATLAB. R2019b, fmincon function documentation. Natick, Massachusetts: The MathWorks Inc.; 2019. Available from: <https://www.mathworks.com/help/optim/ug/fmincon.html>.
- [23] S. KIM, D. KIM. *Directional Dependence Analysis Using Skew-Normal Copula-Based Regression*. Book Chapter In: Statistics and Causality: Methods for Applied Empirical Research, eds. by Wiedermann, W., von Eye, A. Wiley, Hoboken, NJ, 2016.
- [24] G. D'AMICO, F. PETRONI. *Copula based multivariate semi-Markov models with applications in high-frequency finance*. European J. Oper. Res., 2018, **267**(2): 765–777.
- [25] K. SHANNON. bsurc/r2-doi: 1.0.17. Zenodo; 2018. Available from: <https://doi.org/10.5281/zenodo.1646597>.