

Multivariate temporal dependence via mixtures of rotated copulas

RUYI PAN, LUIS E. NIETO-BARAJAS, RADU V. CRAIU

University of Toronto & ITAM

`ruyi.pan@mail.utoronto.ca, luis.nieto@itam.mx & radu.craiu@utoronto.ca`

Abstract

Parametric copula families have been known to flexibly capture various dependence patterns, e.g., either positive or negative dependence in either the lower or upper tails of bivariate distributions. In this paper, our objective is to construct a model that is adaptable enough to capture several of these features simultaneously in m dimensions. We propose a mixture of 2^m rotations of a parametric copula that can achieve this goal. We illustrate the construction using the Clayton family but the concept is general and can be applied to other families. In order to include dynamic dependence regimes, the approach is extended to a time-dependent sequence of mixture copulas in which the mixture probabilities are allowed to evolve in time via a moving average and seasonal types of relationship. The properties of the proposed model and its performance are examined using simulated and real data sets.

Keywords: Bayesian inference, dynamic dependence models, moving average process, seasonal model, time varying copulas.

1 Introduction

Copulas have emerged in recent years as viable tools for modelling dependence in non-standard situations in which the usual “suspects” such as multivariate Gaussian, Student or Wishart distributions, are not appropriate. Besides being an important tool for methodological development and having considerable potential for applications, copulas have gained popularity due to several features that are desirable to a statistician. Allowing the separation

of modeling effort for the marginal models and the dependence structure continues to rank high, but so is the flexibility it exhibits in capturing dependence patterns using parametric families, especially for bivariate data. In higher dimensions, this flexibility is expressed through the use of C- or D- vine copulas that make efficient use of bivariate conditional copulas to flexibly model multivariate ones [5].

In the analysis of extreme value data, it is often desirable to measure the tail dependence in a vector. Some copulas are able to capture tail dependence for instance, the Clayton/Gumbel copula with positive θ parameter exhibits upper/lower tail dependence [22]. However, while one can identify copula families able to capture a bivariate distribution's various patterns of lower or upper tail dependence, be they positive or negative, there is interest for developing more flexible parametric families that can capture several such patterns simultaneously in higher dimensions. Most attempts have been done in two dimensions. [15] proposed the Joe-Clayton Archimedean copula, which is able to model lower and upper tail dependence. Later, [14] proposed the use of a three-component mixture of Gaussian, Gumbel and survival Gumbel copulas that allows for no, lower or upper tail dependence. Alternatively, the survival Gumbel copula was replaced by the Clayton copula in [20].

A bivariate survival copula is a 180° rotation of a bivariate copula. The advantage of such rotation is that, tail properties are reflected with respect to the $v = 1 - u$ line in the unit square. However, other degrees of rotations, like 90° and 270° , are also possible. For instance, [16] considered all four rotations of a Clayton copula in two dimensions and developed model selection criteria for selecting the correct one. On the other hand, [26] proposed a jointly symmetric copula with an equally weighted mixture of the four way rotations of a copula with the same parameter. In a similar fashion, [29] proposed two rotation mixtures of $0^\circ - 180^\circ$ and $90^\circ - 270^\circ$, both with the same dependence parameter, in order to tackle both serial and cross sectional dependence.

In time series analysis, copulas have been used to capture serial dependence. For example,

[11] proposed a dynamic copula model in which copula parameters follow an autoregressive process, and [21] factored the joint density of a unit vector $\mathbf{u} = (u_1, \dots, u_T)$ as $c(\mathbf{u}) = \prod_{t=1}^T f(u_t | u_{t-1})$ assuming Markov conditional distributions. They further used a bivariate copula to model the transitions, i.e. $f(u_t | u_{t-1}) = c_2(u_{t-1}, u_t)$. Specifically, in the bivariate copula they use a mixture of rotations of degrees $0^\circ - 90^\circ$.

In this paper, we first generalise the concept of bivariate tail dependence to any corner of the unit m -dimensional hypercube and propose a flexible copula that is able to capture multiple types of tail dependence. Our goal is achieved by mixing all 2^m rotated versions of the multivariate copula. Furthermore, the 2^m -dimensional mixture weights $\boldsymbol{\pi}_t$ are all different and are allowed to change over time through a moving average type process of order q and a seasonal component of order p that maintains the marginal distribution invariant over time. For each of the mixture copula components, dependence parameters are allowed to vary over time and are, a priori, assumed to be exchangeable. Our context is not a traditional time series problem in the sense that we do not follow a single individual through time, but rather we monitor the dependence of several individuals in time.

The rest of the paper is organized as follows. In Section 2, we provide the motivation of the paper and the required notation. Section 3 contains a discussion about tail dependence. In Section 5, we define our mixture model for a specific time and define the time dependent mixture weights and association parameters. Section 6 provides the prior and posterior distributions that are required to conduct a Bayesian analysis of the model. An illustration of the model’s performance is reported in Section 7. Section 8 contains conclusions and directions for future work.

2 Motivation and Notation

The emergence of copulas as important tools for modeling dependence has its origins in Sklar’s paper [28] which demonstrated that the link between any continuous multivariate

distribution and its marginals can be achieved via a unique copula $C : [0, 1]^m \rightarrow [0, 1]$. The latter is a multivariate distribution with uniform marginals on the interval $[0, 1]$. Specifically, if F is a multivariate cumulative distribution function (CDF) with marginal CDFs F_1, \dots, F_m , then $F(x_1, \dots, x_m) = C\{F_1(x_1), \dots, F_m(x_m)\}$. Additionally, the copula function can be obtained as $C(u_1, \dots, u_m) = F\{F_1^{-1}(u_1), \dots, F_m^{-1}(u_m)\}$, where F_j^{-1} for $j = 1, \dots, m$ are the marginal inverse CDFs or quantile functions.

There is a large body of literature devoted to identifying parametric copula families that are able to capture various dependence patterns in the tails [15]. For instance, in the analysis of extreme value theory, an important concept is that of dependence in the upper-right or lower-down quadrants of a joint bivariate distribution. This is quantified by the so-called upper and lower tail dependence coefficients [7, 15].

Let (X_1, X_2) be a bivariate vector with marginal CDFs, F_1 and, respectively, F_2 , such that the joint CDF is given in terms of the copula C as $F(x_1, x_2) = C(F_1(x_1), F_2(x_2))$. Tail dependence coefficients are defined as limits of the conditional probabilities that both variables are above an upper quantile of order $1 - \nu$, or both variables are below a lower quantile of order ν , as ν approaches zero. We denote

$$\lambda_{11} = \lim_{\nu \rightarrow 0} P\{X_1 > F_1^{-1}(1 - \nu) \mid X_2 > F_2^{-1}(1 - \nu)\} = \lim_{\nu \rightarrow 0} P(U_1 > 1 - \nu \mid U_2 > 1 - \nu)$$

for the upper-right (upper-upper) corner, and

$$\lambda_{00} = \lim_{\nu \rightarrow 0} P\{X_1 \leq F_1^{-1}(\nu) \mid X_2 \leq F_2^{-1}(\nu)\} = \lim_{\nu \rightarrow 0} P(U_1 \leq \nu \mid U_2 \leq \nu)$$

for the lower-down (lower-lower) corner. The binary sub-indexes 0 and 1 stand for lower and upper, respectively. However, it is possible that both variables have co-movements in the opposite tails, that is, one variable has values in the upper quantile and the other in the lower quantile, or conversely. In this case the opposite tail dependencies are defined as

$$\lambda_{10} = \lim_{\nu \rightarrow 0} P\{X_1 > F_1^{-1}(1 - \nu) \mid X_2 \leq F_2^{-1}(\nu)\} = \lim_{\nu \rightarrow 0} P(U_1 > 1 - \nu \mid U_2 \leq \nu)$$

for the upper-lower corner, and

$$\lambda_{01} = \lim_{\nu \rightarrow 0} P\{X_1 \leq F_1^{-1}(\nu) \mid X_2 > F_2^{-1}(1 - \nu)\} = \lim_{\nu \rightarrow 0} P(U_1 \leq \nu \mid U_2 > 1 - \nu)$$

for the lower-upper corner.

These four tail dependence coefficients can be written entirely in terms of the copula. It is straightforward to show that

$$\begin{aligned} \lambda_{11} &= \lim_{\nu \rightarrow 0} \frac{2\nu - 1 + C(1 - \nu, 1 - \nu)}{\nu}, & \lambda_{00} &= \lim_{\nu \rightarrow 0} \frac{C(\nu, \nu)}{\nu}, \\ \lambda_{10} &= \lim_{\nu \rightarrow 0} \frac{\nu - C(1 - \nu, \nu)}{\nu}, & \lambda_{01} &= \lim_{\nu \rightarrow 0} \frac{\nu - C(\nu, 1 - \nu)}{\nu}. \end{aligned} \quad (1)$$

It is also well known [e.g. 33] that the Clayton copula exhibits lower-lower tail dependence, whereas the Gumbel copula has upper-upper tail dependence. One way of defining copulas with the four types of tail dependence (1) is by means of rotation as in [16]. It is easy to see from (1) that for most copulas, the four tail dependence coefficients will be different. In the next section, we develop a mixture of copulas that allows identical tail dependence coefficients.

Before we proceed, let us introduce some notation. Let $\text{Ga}(\alpha, \beta)$ denote a gamma density with mean α/β , $\text{Be}(\alpha, \beta)$ a beta density with mean $\alpha/(\alpha + \beta)$, $\text{N}(\mu, \tau)$ a normal density with mean μ and precision τ , $\text{Dir}(\boldsymbol{\alpha})$ a Dirichlet density with parameter vector $\boldsymbol{\alpha}$, and $\text{Mult}(c, \mathbf{p})$ a multinomial density with total trials c and probability vector \mathbf{p} . The density evaluated at a specific point x , will be denoted using the notation for the density, e.g. $\text{Ga}(x \mid \alpha, \beta)$, in the gamma case.

3 Bivariate rotations

Before we introduce the general case, consider for illustration the case $m = 2$ in which the unit square $[0, 1]^2$ is divided into four quadrants as in Figure 1. To define the 90-degree rotation, we consider the probability in quadrant II, $P(U_1 > u_1, U_2 \leq u_2) = P(U_2 \leq u_2) - P(U_1 \leq$

$u_1, U_2 \leq u_2$), which in terms of the copula becomes $u_2 - C(u_1, u_2)$. Finally by making the transformation $U'_1 = 1 - U_1$, we maintain the marginal uniformity in U'_1 and can obtain a new CDF (copula) of the form

$$C_{10}(u_1, u_2) = P(U'_1 \leq u_1, U_2 \leq u_2) = u_2 - C(1 - u_1, u_2).$$

To define the 180° rotation, we consider the following probability in quadrant III, $P(U_1 > u_1, U_2 > u_2) = 1 - P(U_1 \leq u_1) - P(U_2 \leq u_2) + P(U_1 \leq u_1, U_2 \leq u_2)$, which in terms of the copula becomes $1 - u_1 - u_2 + C(u_1, u_2)$. Again, making the transformation $U'_1 = 1 - U_1$ and $U'_2 = 1 - U_2$ we get a new CDF (copula)

$$C_{11}(u_1, u_2) = P(U'_1 \leq u_1, U'_2 \leq u_2) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2).$$

Lastly, to define the 270 -degree rotation we consider the probability, $P(U_1 \leq u_1, U_2 > u_2) = P(U_1 \leq u_1) - P(U_1 \leq u_1, U_2 \leq u_2)$, which in terms of the copula can be written as $u_1 - C(u_1, u_2)$. Making the transformation $U'_2 = 1 - U_2$, we obtain the new CDF (copula)

$$C_{01}(u_1, u_2) = P(U_1 \leq u_1, U'_2 \leq u_2) = u_1 - C(u_1, 1 - u_2).$$

For completeness, we denote the original, un-rotated, copula as $C_{00}(u_1, u_2)$.

If a particular copula has lower-lower or upper-upper tail dependence, each of the four rotated versions of the copula will have the same tail dependence, but in different corners. To illustrate, consider the Clayton copula defined as $C(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}$ for $\theta \geq -1$. If $\theta = 0$ Clayton copula reduces to the independence copula, and for $\theta > 0$ the Kendall's tau association coefficient is $\tau = \theta/(2+\theta)$ and there is a lower-lower tail dependence with coefficient $\lambda_{00} = 2^{-1/\theta}$.

The four rotations of the Clayton copula are as follows.

(i) 0-degree rotation:

$$C_{00}(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}$$

(ii) 90- and 270-degrees rotations:

$$C_{10}(u_1, u_2) = u_2 - \{(1 - u_1)^{-\theta} + u_2^{-\theta} - 1\}^{-1/\theta}$$

$$C_{01}(u_1, u_2) = u_1 - \{u_1^{-\theta} + (1 - u_2)^{-\theta} - 1\}^{-1/\theta}$$

(iii) 180-degree rotation:

$$C_{11}(u_1, u_2) = u_1 + u_2 - 1 + \{(1 - u_1)^{-\theta} + (1 - u_2)^{-\theta} - 1\}^{-1/\theta}.$$

Using (i)–(iii) and (1), it is not difficult to prove that the tail dependence coefficients, when the tail (corner) corresponds to the rotation, are the same $\lambda_{00}(C_{00}) = \lambda_{10}(C_{10}) = \lambda_{01}(C_{01}) = \lambda_{11}(C_{11})$ and are given by

$$\lambda_{jk}(C_{jk}) = \lim_{\nu \rightarrow 0} \frac{(2\nu^{-\theta} - 1)^{-1/\theta}}{\nu} = \lim_{\nu \rightarrow 0} (2 - \nu^\theta)^{-1/\theta} = 2^{-1/\theta}, \quad (2)$$

for $j, k \in \{0, 1\}$ and for $\theta > 0$. If the tail coefficient does not correspond to a rotation then it has the value zero.

4 Multivariate extensions

Let $\mathbf{U} = (U_1, \dots, U_m)$ be an m -multivariate random vector with support on the unit hypercube $[0, 1]^m$, for $m \geq 2$. Let $C^{(m)}$ be an m -dimensional symmetric copula. Most dependence measures involve two random variables, like the Pearson and Spearman correlation coefficients and Kendall's tau. Additionally, as far as we are aware, the tail dependence is only defined for a pair of random variables. Since this last measure is defined in terms of a conditional probability, it can be generalised to more than two random variables in different ways.

We proposed a generalised tail (corner) dependence multivariate coefficient as follows. In a m -dimensional copula there are 2^m corners that can be identified by a binary sequence $\mathbf{j} \in \mathbb{J} = \{0, 1\}^m$ such that a value of 0/1 at position l denotes the lower/upper corner of

variable l for $l = 1, \dots, m$. Let us also denote the intervals $B_0(\nu) = [0, \nu]$ and $B_1(\nu) = (\nu, 1]$.

Then the tail dependence coefficient for variable l at corner $\mathbf{j} = (j_1, \dots, j_m)$ is defined as

$$\lambda_{\mathbf{j}}^l = \lim_{\nu \rightarrow 0} \mathbb{P} [U_l \in B_{j_l}(\nu) \mid \cap_{k \neq l}^m \{U_k \in B_{j_k}(\nu)\}], \quad (3)$$

for $l = 1, \dots, m$. There will be a total of $\#\mathbb{J} = 2^m$ tail dependence multivariate coefficients.

It is straightforward to see that each of these coefficients can be expressed in terms of the copula $C^{(m)}$. Moreover, if the copula is symmetric, then $\lambda_{\mathbf{j}}^l$ is the same for any $l = 1, \dots, m$.

For example, if $m = 3$ there are $2^3 = 8$ corners. If $\mathbf{j} = (0, 0, 1)$ and $l = 2$ then $\lambda_{001}^2 = \lim_{\nu \rightarrow 0} \mathbb{P}(U_2 \leq \nu \mid U_1 \leq \nu, U_3 > 1 - \nu)$. By using the definition of conditional probability we can re-express this tail dependence coefficient as $\lambda_{001}^2 = \lim_{\nu \rightarrow 0} \{C^{(2)}(\nu, \nu) - C^{(3)}(\nu, \nu, \nu)\} / \{C^{(1)}(\nu) - C^{(2)}(\nu)\}$ with $C^{(1)}(\nu) = \nu$.

As in the bivariate case, it is possible to rotate an m -variate copula to each of the 2^m corners. Let \mathbb{J}_k the subset of \mathbb{J} with k 1's, i.e., if $\mathbf{j} \in \mathbb{J}_k$ then $\sum_{l=1}^m j_l = k$, for $k = 0, \dots, m$. Moreover, we can partition the set \mathbb{J} into disjoint subsets \mathbb{J}_k such that, $\mathbb{J} = \cup_{k=0}^m \mathbb{J}_k$ with $\mathbb{J}_k \cap \mathbb{J}_j = \emptyset$ for $k \neq j$. The number of elements in each subset is $\#\mathbb{J}_k = \binom{m}{k}$, and $2^m = \sum_{k=0}^m \binom{m}{k}$. Let also define $\mathbf{u}^{\mathbf{j}} = (u_1^{1-j_1}(1-u_1)^{j_1}, \dots, u_m^{1-j_m}(1-u_m)^{j_m})$.

For $k = 0$ and $\mathbf{j} \in \mathbb{J}_0$, then

$$C_{\mathbf{j}}(\mathbf{u}) = C^{(m)}(\mathbf{u}^{\mathbf{j}}),$$

where $C^{(m)}$ is the original copula without rotation. For $k = 1$ and $\mathbf{j} \in \mathbb{J}_1$ and if the only index 1 is situated at the l^{th} position, then

$$C_{\mathbf{j}}(\mathbf{u}) = C^{(m-1)}(\mathbf{u}_{(-l)}^{\mathbf{j}}) - C^{(m)}(\mathbf{u}^{\mathbf{j}}).$$

For $k = 2$ and $\mathbf{j} \in \mathbb{J}_2$ and if the two indexes 1 are situated at positions (l_1, l_2) , then

$$C_{\mathbf{j}}(\mathbf{u}) = C^{(m-2)}(\mathbf{u}_{(-l_1, l_2)}^{\mathbf{j}}) - \sum_{i=1}^2 C^{(m-1)}(\mathbf{u}_{(-l_i)}^{\mathbf{j}}) + C^{(m)}(\mathbf{u}^{\mathbf{j}}).$$

For $k = 3$ and $\mathbf{j} \in \mathbb{J}_3$ and if the three indexes 1 are situated at positions (l_1, l_2, l_3) , then

$$C_{\mathbf{j}}(\mathbf{u}) = C^{(m-3)}(\mathbf{u}_{(-l_1, l_2, l_3)}^{\mathbf{j}}) - \sum_{i < j}^3 C^{(m-2)}(\mathbf{u}_{(-l_i, l_j)}^{\mathbf{j}}) + \sum_{i=1}^3 C^{(m-1)}(\mathbf{u}_{(-l_i)}^{\mathbf{j}}) - C^m(\mathbf{u}^{\mathbf{j}}).$$

We carry on until $k = m$ and $\mathbf{j} \in \mathbb{J}_m$, so all indexes in \mathbf{j} are 1, then

$$\begin{aligned} C_{\mathbf{j}}(\mathbf{u}) = & 1 - \sum_{i=1}^m C^{(1)}(1 - u_i) + \sum_{i < j}^m C^{(2)}(1 - u_i, 1 - u_j) - \sum_{i < j < k}^m C^{(3)}(1 - u_i, 1 - u_j, 1 - u_k) \\ & + \cdots + (-1)^m C^{(m)}(1 - u_1, 1 - u_2, \dots, 1 - u_m). \end{aligned}$$

If a copula exhibits multivariate tail dependence in a particular corner, each of the rotated versions of the copula will also have tail dependence in one corner. To illustrate, we also consider the Clayton family whose multivariate version is defined as

$$C^{(m)}(u_1, \dots, u_m) = \left\{ \sum_{j=1}^m (u_j^{-\theta} - 1) + 1 \right\}^{-1/\theta} \quad (4)$$

for $\theta \geq 0$.

Specifically, we concentrate in the case $m = 3$. The number of possible rotations is $2^3 = 8$. The four subsets of rotated copulas generated by \mathbb{J}_k and $k = 0, 1, 2, 3$ are:

(iv) For $k = 0$ we have one element

$$C_{000}(\mathbf{u}) = C^{(3)}(\mathbf{u}) = (u_1^{-\theta} + u_2^{-\theta} + u_3^{-\theta} - 2)^{-1/\theta}$$

(v) For $k = 1$ we have three elements

$$\begin{aligned} C_{100}(\mathbf{u}) &= (u_2^{-\theta} + u_3^{-\theta} - 1)^{-1/\theta} - \{(1 - u_1)^{-\theta} + u_2^{-\theta} + u_3^{-\theta} - 2\}^{-1/\theta} \\ C_{010}(\mathbf{u}) &= (u_1^{-\theta} + u_3^{-\theta} - 1)^{-1/\theta} - \{u_1^{-\theta} + (1 - u_2)^{-\theta} + u_3^{-\theta} - 2\}^{-1/\theta} \\ C_{001}(\mathbf{u}) &= (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta} - \{u_1^{-\theta} + u_2^{-\theta} + (1 - u_3)^{-\theta} - 2\}^{-1/\theta} \end{aligned}$$

(vi) For $k = 2$ we have three elements

$$\begin{aligned}
C_{110}(\mathbf{u}) &= u_3 - \{(1 - u_2)^{-\theta} + u_3^{-\theta} - 1\}^{-1/\theta} - \{(1 - u_1)^{-\theta} + u_3^{-\theta} - 1\}^{-1/\theta} \\
&\quad + \{(1 - u_1)^{-\theta} + (1 - u_2)^{-\theta} + u_3^{-\theta} - 2\}^{-1/\theta} \\
C_{101}(\mathbf{u}) &= u_2 - \{(1 - u_1)^{-\theta} + u_2^{-\theta} - 1\}^{-1/\theta} - \{u_2^{-\theta} + (1 - u_3)^{-\theta} - 1\}^{-1/\theta} \\
&\quad + \{(1 - u_1)^{-\theta} + u_2^{-\theta} + (1 - u_3)^{-\theta} - 2\}^{-1/\theta} \\
C_{011}(\mathbf{u}) &= u_1 - \{u_1^{-\theta} + (1 - u_2)^{-\theta} - 1\}^{-1/\theta} - \{u_1^{-\theta} + (1 - u_3)^{-\theta} - 1\}^{-1/\theta} \\
&\quad + \{u_1^{-\theta} + (1 - u_2)^{-\theta} + (1 - u_3)^{-\theta} - 2\}^{-1/\theta}
\end{aligned}$$

(vii) For $k = 3$ we have one element

$$\begin{aligned}
C_{111}(\mathbf{u}) &= 1 - (1 - u_1) - (1 - u_2) - (1 - u_3) + \{(1 - u_1)^{-\theta} + (1 - u_2)^{-\theta} - 1\}^{-1/\theta} \\
&\quad + \{(1 - u_1)^{-\theta} + (1 - u_3)^{-\theta} - 1\}^{-1/\theta} + \{(1 - u_2)^{-\theta} + (1 - u_3)^{-\theta} - 1\}^{-1/\theta} \\
&\quad - \{(1 - u_1)^{-\theta} + (1 - u_2)^{-\theta} + (1 - u_3)^{-\theta} - 2\}^{-1/\theta}
\end{aligned}$$

Similarly to the bivariate case, if the tail coefficient corresponds to a rotation, then we get a coefficient different from zero. Specifically for the Clayton family with $\theta > 0$, using (iv)–(vii) and (3), the tail coefficients are given by

$$\lambda_{j_1, j_2, j_3}(C_{j_1, j_2, j_3}) = \lim_{\nu \rightarrow 0} \frac{(3\nu^{-\theta} - 2)^{-1/\theta}}{(2\nu^{-\theta} - 1)^{-1/\theta}} = \lim_{\nu \rightarrow 0} \left(\frac{3 - 2\nu^\theta}{2 - \nu^\theta} \right)^{-1/\theta} = \left(\frac{3}{2} \right)^{-1/\theta}, \quad (5)$$

for $j_l \in \{0, 1\}$, $l = 1, 2, 3$. Note that the superindex in λ has been removed because these tail coefficients are the same for any l due to the symmetry in the copula. If the tail (corner) does not correspond to a rotation, the tail coefficient is zero.

5 Dynamic mixtures

Let $\mathbf{U}_t = (U_{t1}, U_{t2}, \dots, U_{tm})$ be an m -variate vector with $\text{Unif}(0, 1)$ marginal distributions for each $t = 1, 2, \dots, T$. The aim in this section is to model the joint distribution of \mathbf{U}_t through

a flexible copula C_t which is able to capture any kind of tail dependence as it evolves in time. For that, we use all 2^m rotations, each one with different parameter $\theta_{t,\mathbf{j}}$ for $\mathbf{j} \in \mathbb{J}$, and define the following mixture copula

$$C_t(\mathbf{u}_t \mid \boldsymbol{\pi}_t, \boldsymbol{\theta}_t) = \sum_{\mathbf{j} \in \mathbb{J}} \pi_{t,\mathbf{j}} C_{\mathbf{j}}(\mathbf{u}_t \mid \theta_{t,\mathbf{j}}), \quad (6)$$

with 2^m -dimensional parameters $\boldsymbol{\pi}_t = (\pi_{t,00\dots 0}, \pi_{t,10\dots 0}, \dots, \pi_{t,11\dots 1})$,

$\boldsymbol{\theta}_t = (\theta_{t,00\dots 0}, \theta_{t,10\dots 0}, \dots, \theta_{t,11\dots 1})$, and $C_{\mathbf{j}}$ is a rotated copula as defined in Section 4. Parameters $\pi_{t,\mathbf{j}} > 0$ are mixture weights such that $\sum_{\mathbf{j} \in \mathbb{J}} \pi_{t,\mathbf{j}} = 1$, whereas $\theta_{t,\mathbf{j}}$ are copula parameters whose parameter space depends on the specific copula chosen, for $t = 1, \dots, T$.

It is not difficult to derive association coefficients, like Kendall's tau, and tail dependence coefficients for a mixture copula in terms of the corresponding coefficients for the individual copulas. In particular, the Kendall's tau between any pair of variables, say $(U_{t,i}, U_{t,k})$, $i \neq k = 1, \dots, m$, for the mixture copula (6) is

$$\begin{aligned} \tau_t &= 4\mathbb{E}\{C_t^{(2)}(U_{t,i}, U_{t,k} \mid \boldsymbol{\pi}_t, \boldsymbol{\theta}_t)\} - 1 = 4 \sum_{\mathbf{j} \in \mathbb{J}} \pi_{t,\mathbf{j}} \mathbb{E}\{C_{\mathbf{j}}^{(2)}(U_{t,i}, U_{t,k} \mid \boldsymbol{\theta}_t)\} - 1 \\ &= \sum_{\mathbf{j} \in \mathbb{J}} \pi_{t,\mathbf{j}} \left[4\mathbb{E}\{C_{\mathbf{j}}^{(2)}(U_{t,i}, U_{t,k} \mid \boldsymbol{\theta}_t)\} - 1 \right] = \sum_{\mathbf{j} \in \mathbb{J}} \pi_{t,\mathbf{j}} \tau_{t,\mathbf{j}}, \end{aligned}$$

where $\tau_{t,\mathbf{j}}$ is the individual Kendall's tau, between $(U_{t,i}, U_{t,k})$, for each of the mixture copula components $C_{\mathbf{j}}$. In particular, for the Clayton family (4), all rotated mixture components $C_{\mathbf{j}}$, as in (i)–(vii), share the same properties of the Clayton family, therefore their Kendall's tau coefficients are $\tau_{t,\mathbf{j}} = \theta_{t,\mathbf{j}}/(2 + \theta_{t,\mathbf{j}})$ if $\sum_{l=1}^m j_l$ is an even number, and $\tau_{t,\mathbf{j}} = -\theta_{t,\mathbf{j}}/(2 + \theta_{t,\mathbf{j}})$ if $\sum_{l=1}^m j_l$ is an odd number. Therefore, the Kendall's tau coefficient for the mixture copula (6) of Clayton components becomes

$$\tau_t = \sum_{\mathbf{j} \in \mathbb{J}} (-1)^{\sum_{l=1}^m j_l} \pi_{t,\mathbf{j}} \left(\frac{\theta_{t,\mathbf{j}}}{2 + \theta_{t,\mathbf{j}}} \right). \quad (7)$$

Similarly to Kendall's tau, it is not difficult to prove that multivariate tail dependence coefficients (5) for mixture copula (6) at time t , become the mixture of components-wise tail

dependence coefficients, that is

$$\lambda_{t,\mathbf{k}}(C_t) = \sum_{\mathbf{j} \in \mathbb{J}} \pi_{t,\mathbf{j}} \lambda_{t,\mathbf{k}}(C_{\mathbf{j}}), \quad (8)$$

for any $\mathbf{k} \in \mathbb{J}$ and with $\lambda_{t,\mathbf{k}}(C_{\mathbf{j}})$ the \mathbf{k} -tail dependence coefficient for rotated copula $C_{\mathbf{j}}$. For the Clayton family case, $\lambda_{t,\mathbf{k}}(C_{\mathbf{j}}) > 0$ only for $\mathbf{k} = \mathbf{j}$. Therefore, when $m = 2$ and using (2), we obtain that the tail dependence coefficients simplify to $\lambda_{t,j,k}(C_t) = \pi_{t,j,k}(2)^{-1/\theta_{t,j,k}}$. Now, for $m = 3$ we use (5) and obtain that the tail dependence coefficients are $\lambda_{t,j_1,j_2,j_3}(C_t) = \pi_{t,j_1,j_2,j_3}(3/2)^{-1/\theta_{t,j_1,j_2,j_3}}$.

In summary, our mixture copula proposal (6) is flexible enough to capture a larger class of dependence associations and all 2^m tail dependencies, in terms of the copula parameters $\pi_{t,\mathbf{j}}$ and $\theta_{t,\mathbf{j}}$ for $\mathbf{j} \in \mathbb{J}$ and $t = 1, \dots, T$.

6 Bayesian analysis

6.1 Prior distributions

To allow for temporal dependence in the parameter estimation, we propose a prior dynamic process for $\boldsymbol{\pi} = \{\boldsymbol{\pi}_t\}_{1 \leq t \leq T}$, where $\boldsymbol{\pi}_t = \{\pi_{t,\mathbf{j}}, \mathbf{j} \in \mathbb{J}\}$. Since $\sum_{\mathbf{j} \in \mathbb{J}} \pi_{t,\mathbf{j}} = 1$ for all t , the natural marginal prior for $\boldsymbol{\pi}_t$ would be a Dirichlet distribution with parameter $a_0 \mathbf{p}$, where $\mathbf{p} = \{p_{\mathbf{j}}, \mathbf{j} \in \mathbb{J}\}$ such that $a_0 > 0$, $p_{\mathbf{j}} > 0$ and $\sum_{\mathbf{j} \in \mathbb{J}} p_{\mathbf{j}} = 1$. To relate a set of Dirichlet random variables, we use ideas from [25], who defined dependence in univariate random variables whose distributions belong to the exponential family, and define a dynamic prior with temporal dependence as follows.

Let $\boldsymbol{\eta}_t = \{\eta_{t,\mathbf{j}}, \mathbf{j} \in \mathbb{J}\} \in \mathbb{R}^{2^m}$ be a latent vector corresponding to each $\boldsymbol{\pi}_t$ and let $\boldsymbol{\omega} = \{\omega_{\mathbf{j}}, \mathbf{j} \in \mathbb{J}\}$ be a unique latent vector such that

$$\boldsymbol{\omega} \sim \text{Dir}(a_0 \mathbf{p}) \quad \text{and} \quad \boldsymbol{\eta}_t \mid \boldsymbol{\omega} \stackrel{\text{ind}}{\sim} \text{Mult}(a_t, \boldsymbol{\omega}), \quad (9)$$

with $a_t \in \mathbb{N}$, $\eta_{tk} \in \mathbb{N}$ and $\sum_{\mathbf{j} \in \mathbb{J}} \eta_{t,\mathbf{j}} = a_t$. Then, the prior dependence in $\boldsymbol{\pi}_t$ is modeled

through a subset ∂_t of previous latent variables $\{\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \dots, \boldsymbol{\eta}_{t-1}\}$

$$\boldsymbol{\pi}_t \mid \boldsymbol{\eta} \stackrel{\text{ind}}{\sim} \text{Dir} \left(a_0 \mathbf{p} + \sum_{k \in \partial_t} \boldsymbol{\eta}_k \right). \quad (10)$$

We denote this construction as $\text{DDir}(a_0, \mathbf{a}, \boldsymbol{\partial})$ with $a_0 > 0$, $\mathbf{a} = (a_1, \dots, a_T)$ and subsets $\boldsymbol{\partial} = \{\partial_t\}$ of lags. Different temporal dependencies can be induced by an appropriate selection of subsets of lags. For instance: *moving average* type of order q can be induced by defining

$$\partial_t = \{t, t-1, \dots, t-q\}; \quad (11)$$

seasonal dependence of order p can be induced by defining, for seasonality s ,

$$\partial_t = \{t, t-s, t-2s, \dots, t-ps\}; \quad (12)$$

or a combination of moving average of order q and seasonal dependence of order p . In general, the only requirement is that $t \in \partial_t$.

Properties of this prior are given in the following proposition.

Proposition 1 *Let $\boldsymbol{\pi} = \{\boldsymbol{\pi}_t\} \sim \text{DDir}(a_0, \mathbf{a}, \boldsymbol{\partial})$ a sequence of vectors whose probability law is defined by (9) and (10) for $a_0 > 0$, $a_t \in \mathbb{N}$ and subsets $\boldsymbol{\partial}$. Then,*

(i) *The marginal distribution for each $\boldsymbol{\pi}_t$ is $\text{Dir}(a_0 \mathbf{p})$,*

(ii) *The correlation between $\pi_{t,\mathbf{j}}$ and $\pi_{r,\mathbf{j}}$, for $t \neq r$ and $\mathbf{j} \in \mathbb{J}$, does not depend on the specific \mathbf{j} and is given by*

$$\text{Corr}(\pi_{t,\mathbf{j}}, \pi_{r,\mathbf{j}}) = \frac{a_0 \left(\sum_{k \in \partial_t \cap \partial_r} a_k \right) + \left(\sum_{k \in \partial_t} a_k \right) \left(\sum_{k \in \partial_r} a_k \right)}{\left(a_0 + \sum_{k \in \partial_t} a_k \right) \left(a_0 + \sum_{k \in \partial_r} a_k \right)}$$

(iii) *If $a_t = 0$ for all $t = 1, 2, \dots$ then the $\boldsymbol{\pi}_t$'s become independent.*

Proof

For (i) we rely on conjugacy properties of the Dirichlet multinomial Bayesian updating [3]. This states that if $\boldsymbol{\eta}_t$, $t = 1, 2, \dots$ are conditionally independent given $\boldsymbol{\omega}$ in (9), whose prior

is $\boldsymbol{\omega} \sim \text{Dir}(a_0 \mathbf{p})$, then the posterior distribution for $\boldsymbol{\omega}$ given the $\boldsymbol{\eta}_t$'s is $\text{Dir}(a_0 \mathbf{p} + \sum_t \boldsymbol{\eta}_t)$. Replacing $\boldsymbol{\omega}$ in the posterior by $\boldsymbol{\pi}_t$ we obtain that the marginal distribution for $\boldsymbol{\pi}_t$ is the same as the prior for $\boldsymbol{\omega}$.

For (ii) we first note that for a specific \mathbf{j} , the distributions for $\omega_{\mathbf{j}}$, $\eta_{t,\mathbf{j}}$ and $\pi_{t,\mathbf{j}}$ reduce to beta, binomial and beta, respectively. To obtain the correlation we rely on iterative formulae. The covariance is $\text{Cov}(\pi_{t,\mathbf{j}}, \pi_{r,\mathbf{j}}) = \text{E}\{\text{Cov}(\pi_{t,\mathbf{j}}, \pi_{r,\mathbf{j}} \mid \boldsymbol{\eta})\} + \text{Cov}\{\text{E}(\pi_{t,\mathbf{j}} \mid \boldsymbol{\eta}), \text{E}(\pi_{r,\mathbf{j}} \mid \boldsymbol{\eta})\}$, where the first term is zero due to conditional independence. Then

$$\text{Cov}(\pi_{t,\mathbf{j}}, \pi_{r,\mathbf{j}}) = \text{Cov} \left\{ \frac{a_0 p_{\mathbf{j}} + \sum_{k \in \partial_t} \eta_{k,\mathbf{j}}}{a_0 + \sum_{k \in \partial_t} a_k}, \frac{a_0 p_{\mathbf{j}} + \sum_{k \in \partial_r} \eta_{k,\mathbf{j}}}{a_0 + \sum_{k \in \partial_r} a_k} \right\},$$

which, after canceling the additive constants and using the linearity of the covariance, becomes

$$\text{Cov}(\pi_{t,\mathbf{j}}, \pi_{r,\mathbf{j}}) = \frac{1}{(a_0 + \sum_{k \in \partial_t} a_k)(a_0 + \sum_{k \in \partial_r} a_k)} \text{Cov} \left\{ \sum_{k \in \partial_t} \eta_{k,\mathbf{j}}, \sum_{k \in \partial_r} \eta_{k,\mathbf{j}} \right\}.$$

After using the iterative formula for a second time, we get

$$\text{E} \left[\text{Cov} \left\{ \sum_{k \in \partial_t} \eta_{k,\mathbf{j}}, \sum_{k \in \partial_r} \eta_{k,\mathbf{j}} \mid \boldsymbol{\omega} \right\} \right] + \text{Cov} \left\{ \text{E} \left(\sum_{k \in \partial_t} \eta_{k,\mathbf{j}} \mid \boldsymbol{\omega} \right), \text{E} \left(\sum_{k \in \partial_r} \eta_{k,\mathbf{j}} \mid \boldsymbol{\omega} \right) \right\}. \quad (13)$$

Within each sum we can isolate the common part as $\sum_{k \in \partial_t} \eta_{k,\mathbf{j}} = \sum_{k \in \partial_t \cap \partial_r} \eta_{k,\mathbf{j}} + \sum_{k \in \partial_t - \partial_r} \eta_{k,\mathbf{j}}$ and $\sum_{k \in \partial_r} \eta_{k,\mathbf{j}} = \sum_{k \in \partial_t \cap \partial_r} \eta_{k,\mathbf{j}} + \sum_{k \in \partial_r - \partial_t} \eta_{k,\mathbf{j}}$, and using covariance properties and conditional independence, (13) becomes

$$\text{E} \left\{ \text{Var} \left(\sum_{k \in \partial_t \cap \partial_r} \eta_{k,\mathbf{j}} \mid \omega_{\mathbf{j}} \right) \right\} + \text{Cov} \left\{ \sum_{k \in \partial_t} a_k \omega_{\mathbf{j}}, \sum_{k \in \partial_r} a_k \omega_{\mathbf{j}} \right\}.$$

The first expected value, after obtaining the conditional variance is $\text{E}\{\sum_{k \in \partial_t \cap \partial_r} a_k \omega_{\mathbf{j}}(1 - \omega_{\mathbf{j}})\} = (\sum_{k \in \partial_t \cap \partial_r} a_k) \text{E}(\omega_{\mathbf{j}} - \omega_{\mathbf{j}}^2)$ with $\text{E}(\omega_{\mathbf{j}} - \omega_{\mathbf{j}}^2) = \text{E}(\omega_{\mathbf{j}}) - \text{E}^2(\omega_{\mathbf{j}}) - \text{Var}(\omega_{\mathbf{j}}) = a_0 \text{Var}(\omega_{\mathbf{j}})$.

The second term is $(\sum_{k \in \partial_t} a_k)(\sum_{k \in \partial_r} a_k) \text{Var}(\omega_{\mathbf{j}})$. In conclusion, we obtain

$$\text{Cov}(\omega_{t,\mathbf{j}}, \omega_{r,\mathbf{j}}) = \frac{a_0 (\sum_{k \in \partial_t \cap \partial_r} a_k) + (\sum_{k \in \partial_t} a_k)(\sum_{k \in \partial_r} a_k)}{(a_0 + \sum_{k \in \partial_t} a_k)(a_0 + \sum_{k \in \partial_r} a_k)} \text{Var}(\omega_{\mathbf{j}}).$$

Since $\omega_{\mathbf{j}}$, $\pi_{t,\mathbf{j}}$ and $\pi_{r,\mathbf{j}}$ all have the same beta marginal distribution, (ii) is demonstrated.

For (iii) we note that $a_t = 0$ for all t implies that $\boldsymbol{\eta}_t = 0$ with probability one so the dependence disappears and $\boldsymbol{\pi}_t$ become independent with marginal distribution $\text{Dir}(a_0 \mathbf{p})$. \diamond

The strength of dependence in the prior for $\boldsymbol{\pi}$ depends on the model parameters a_0 , \mathbf{a} and subsets $\boldsymbol{\partial}$. Larger values of any of the first two induce stronger dependence. More shared elements in ∂_t and ∂_r also indicate stronger dependence. However, if the intersection between sets ∂_t and ∂_r is empty, the correlation is still positive.

Prior distributions are completed by assigning hierarchical gamma distributions for each $\theta_{t,\mathbf{j}}$, so that information is shared across times t for each \mathbf{j} . That is,

$$\theta_{t,\mathbf{j}} \mid \beta_{\mathbf{j}} \stackrel{\text{ind}}{\sim} \text{Ga}(d_{\mathbf{j}}, \beta_{\mathbf{j}}), \text{ and } \beta_{\mathbf{j}} \sim \text{Ga}(e_{\mathbf{j}}, g_{\mathbf{j}}) \quad (14)$$

for $t \geq 1$ and $\mathbf{j} \in \mathbb{J}$.

6.2 Posterior distributions

Let $\mathbf{U}_{t,i} = (U_{1,t,i}, \dots, U_{m,t,i})$ for $i = 1, \dots, n_t$ a sample of size n_t from model (6) for each $t = 1, \dots, T$. Let $\mathbf{Z}_{t,i}$ be a latent vector that identifies the mixture component from where observation i is coming from, that is, $\mathbf{Z}_{t,i} = \{Z_{\mathbf{j},t,i}, \mathbf{j} \in \mathbb{J}\} \sim \text{Mult}(1, \boldsymbol{\pi}_t)$. Assuming for the moment that together with $\mathbf{U}_{t,i}$ we observe $\mathbf{Z}_{t,i}$, then the extended likelihood has the form

$$f(\mathbf{u}, \mathbf{z} \mid \boldsymbol{\pi}, \boldsymbol{\theta}) = \prod_{t=1}^T \prod_{i=1}^{n_t} \prod_{\mathbf{j} \in \mathbb{J}} \{\pi_{t,\mathbf{j}} f_{\mathbf{j}}(u_{1,t,i}, \dots, u_{m,t,i} \mid \theta_{t,\mathbf{j}})\}^{z_{\mathbf{j},t,i}},$$

where $f_{\mathbf{j}}(\mathbf{u}_{t,i} \mid \theta_{t,\mathbf{j}})$ are the density functions associated to each of the copula rotations $C_{\mathbf{j}}$ for $\mathbf{j} \in \mathbb{J}$.

In particular, for the Clayton family (4), the corresponding density becomes

$$f_C(\mathbf{u}) = \left\{ \sum_{i=1}^m u_i^{-\theta} - m - 1 \right\}^{-1/\theta-m} \prod_{i=1}^m \{1 + (i-1)\theta\} u_i^{-\theta-1}.$$

For any rotated version we just evaluate at $\tilde{\mathbf{u}} = (\tilde{u}_1, \tilde{u}_2, \dots, \tilde{u}_m)$ where \tilde{u}_i can be u_i or $1 - u_i$.

The prior distribution for $(\boldsymbol{\pi}, \boldsymbol{\theta})$ is defined by equations (9), (10) and (14). Again, extending the prior to include the latent variables $\boldsymbol{\eta}$ and $\boldsymbol{\omega}$ we get

$$f(\boldsymbol{\pi}, \boldsymbol{\eta}, \boldsymbol{\omega}) = \text{Dir}(\boldsymbol{\omega} \mid a_0 \mathbf{p}) \prod_{t=1}^T \text{Dir} \left(\boldsymbol{\pi}_t \left| a_0 \mathbf{p} + \sum_{j \in \partial_t} \boldsymbol{\eta}_j \right. \right) \text{Mult}(\boldsymbol{\eta}_t \mid a_t, \boldsymbol{\omega})$$

and

$$f(\boldsymbol{\theta}) = \prod_{\mathbf{j} \in \mathbb{J}} \text{Ga}(\beta_{\mathbf{j}} \mid e_{\mathbf{j}}, g_{\mathbf{j}}) \prod_{t=1}^T \text{Ga}(\theta_{t,\mathbf{j}} \mid d_{\mathbf{j}}, \beta_{\mathbf{j}}),$$

independent of each other.

Posterior distributions are characterized through their full conditional distributions. These include actual parameters as well as latent variables and are given as follows.

(a) The posterior conditional for \mathbf{Z}_{ti} is

$$\mathbf{Z}_{ti} \mid \text{rest} \sim \text{Mult}(1, \boldsymbol{\pi}_t^*),$$

where $\boldsymbol{\pi}^* = \{\pi_{t,\mathbf{j}}^*\}$ and

$$\pi_{t,\mathbf{j}}^* = \frac{\pi_{t,\mathbf{j}} f_{\mathbf{j}}(\mathbf{u}_{t,i} \mid \theta_{t,\mathbf{j}})}{\sum_{\mathbf{k} \in \mathbb{J}} \pi_{t,\mathbf{k}} f_{\mathbf{k}}(\mathbf{u}_{t,i} \mid \theta_{t,\mathbf{k}})}.$$

(b) The posterior conditional for $\boldsymbol{\pi}_t$ is

$$\boldsymbol{\pi}_t \mid \text{rest} \sim \text{Dir} \left(a_0 \mathbf{p} + \sum_{k \in \partial_t} \boldsymbol{\eta}_k + \sum_{i=1}^{n_t} \mathbf{z}_{t,i} \right).$$

(c) The posterior conditional for $\boldsymbol{\eta}_t$ is

$$f(\boldsymbol{\eta}_t \mid \text{rest}) \propto \left\{ \prod_{\mathbf{j} \in \mathbb{J}} \frac{\left(\omega_{\mathbf{j}} \prod_{l \in \varrho_t} \pi_{l,\mathbf{j}} \right)^{\eta_{t,\mathbf{j}}}}{\Gamma(\eta_{t,\mathbf{j}} + 1) \prod_{l \in \varrho_t} \Gamma(a_0 p_{\mathbf{j}} + \sum_{k \in \partial_l} \eta_{k,\mathbf{j}})} \right\} I \left(\sum_{\mathbf{j} \in \mathbb{J}} \eta_{t,\mathbf{j}} = a_t \right),$$

where $\varrho_t = \{l : t \in \partial_l\}$ is the set of inverse subsets.

(d) The posterior conditional for $\boldsymbol{\omega}$ is

$$f(\boldsymbol{\omega} \mid \text{rest}) = \text{Dir} \left(\boldsymbol{\omega} \left| c_0 \mathbf{p} + \sum_{t=1}^T \boldsymbol{\eta}_t \right. \right).$$

(e) The posterior conditional for $\theta_{t,\mathbf{j}}$ is

$$f(\theta_{t,\mathbf{j}} \mid \text{rest}) \propto \theta_{t,\mathbf{j}}^{d_{\mathbf{j}}-1} e^{-\beta_{\mathbf{j}} \theta_{t,\mathbf{j}}} \prod_{i=1}^{n_t} \{f_{\mathbf{j}}(\mathbf{u}_{t,i} \mid \theta_{t,\mathbf{j}})\}^{z_{t,\mathbf{j},i}}.$$

(f) The posterior conditional for $\beta_{\mathbf{j}}$ is

$$\beta_{\mathbf{j}} \mid \text{rest} \sim \text{Ga} \left(e_{\mathbf{j}} + T d_{\mathbf{j}}, g_{\mathbf{j}} + \sum_{t=1}^T \theta_{t,\mathbf{j}} \right).$$

Posterior inference will rely on the implementation of a Gibbs sampler [30] based on the previous posterior conditional distributions. Sampling from (a), (b), (d) and (f) is straightforward since they are of standard form. To sample from (c), since $\boldsymbol{\eta}_t$ is a vector of dimension 2^m with a sum restriction, it is easier if we sample from each of the components $\eta_{t,\mathbf{j}}$ for $\mathbf{j} \in \mathbb{J}_{(-1)} = \mathbb{J} - \mathbf{1}$, with $\mathbf{1} = 11 \cdots 1$ the vector of all 1's, using $f(\eta_{t,\mathbf{j}} \mid \text{rest}) \propto$

$$\frac{\left\{ \omega_{\mathbf{j}} \prod_{l \in \varrho_t} \pi_{l,\mathbf{j}} / \left(\omega_{\mathbf{1}} \prod_{l \in \varrho_t} \pi_{l,\mathbf{1}} \right) \right\}^{\eta_{t,\mathbf{j}}} I \left(\eta_{t,\mathbf{j}} \in \{0, 1, \dots, a_t - \sum_{\mathbf{k} \in \mathbb{J}_{(-1)}} \eta_{t,\mathbf{k}}\} \right)}{\Gamma(\eta_{t,\mathbf{j}} + 1) \prod_{l \in \varrho_t} \Gamma(a_0 p_{\mathbf{j}} + \sum_{k \in \partial_l} \eta_{k,\mathbf{j}}) \Gamma(\eta_{t,\mathbf{1}} + 1) \prod_{l \in \varrho_t} \Gamma(a_0 p_{\mathbf{1}} + \sum_{k \in \partial_l} \eta_{k,\mathbf{1}})},$$

with $\eta_{t,\mathbf{1}} = a_t - \sum_{\mathbf{j} \in \mathbb{J}_{(-1)}} \eta_{t,\mathbf{j}}$. Sampling from (e) will require a Metropolis-Hastings step [32]. We suggest to use an adaptive random walk proposal defined as follows. At iteration $(r+1)$ sample $\theta_{t,\mathbf{j}}^* \sim \text{Ga}(\kappa, \kappa / \theta_{t,\mathbf{j}}^{(r)})$ and accept it with probability

$$\alpha(\theta_{t,\mathbf{j}}^*, \theta_{t,\mathbf{j}}^{(r)}) = \frac{f(\theta_{t,\mathbf{j}}^* \mid \text{rest}) \text{Ga}(\theta_{t,\mathbf{j}}^{(r)} \mid \kappa, \kappa / \theta_{t,\mathbf{j}}^*)}{f(\theta_{t,\mathbf{j}}^{(r)} \mid \text{rest}) \text{Ga}(\theta_{t,\mathbf{j}}^* \mid \kappa, \kappa / \theta_{t,\mathbf{j}}^{(r)})},$$

where α is truncated to the interval $[0, 1]$ and κ is a tuning parameter that controls the acceptance rate. We adapt κ following the method of [23]. The adaptation method uses batches of 50 iterations and for every batch h we compute the acceptance rate $AR^{(h)}$ and increase $\kappa^{(h+1)} = \kappa^{(h)} 1.01^{\sqrt{h}}$ if $AR^{(h)} < 0.3$ and decrease $\kappa^{(h+1)} = \kappa^{(h)} 1.01^{-\sqrt{h}}$ if $AR^{(h)} > 0.4$, with $\kappa^{(1)} = 1$ as starting value. This adaptation scheme satisfies diminishing adaptation as $h \rightarrow \infty$ and in the applications we restrict the parameters to a compact thus ensuring that the sampler is valid [27].

Implementation code for our dynamic Clayton mixture model, for dimensions $m = 2$ and $m = 3$, can be found at the GitHub repository <https://github.com/RuyiPan/TD-MRC>.

7 Numerical analyses

7.1 Simulation study

We conduct a comprehensive simulation study with $m = 2$ to evaluate the performance of the proposed model. The true generative model is set using $\theta_t = (\theta_{t,00}, \theta_{t,10}, \theta_{t,01}, \theta_{t,11}) = (5, 3, 4, 3)$ for $t = 1, \dots, T$, with $T = 20$. The weights of rotated Clayton copulas are chosen to be linearly dependent in time. More specifically, we first set $\boldsymbol{\pi}_1 = (\pi_{1,00}, \pi_{1,10}, \pi_{1,01}, \pi_{1,11}) = (0.4, 0.25, 0.25, 0.1)$ as the initial values at $t = 1$ and subsequently the weights are constructed using $\pi_{t,00} = 0.95\pi_{t-1,00}, \pi_{t,10} = 1.05\pi_{t-1,10}, \pi_{t,11} = 0.1, \pi_{t,01} = 1 - \pi_{t,00} - \pi_{t,10} - \pi_{t,11}$ for $t = 2, \dots, 20$. We sampled $n_t = 300$ realizations from this model for each time $t = 1, \dots, T$ as the simulated data.

For the prior distributions, we set hyper-parameters $a_0 = 1$, and $e_k = g_k = 1$. To evaluate the performance of the model in capturing the temporal dependence, we considered a moving average type of order q , that is, the subsets of lags are defined by (11). We considered different hyper-parameters for the dynamic process: $a_t = 0, 1, 10, 20, 30, 40$ and $q = 0, 1, \dots, 7$. The model with a_t and q is denoted as $\mathcal{M}_{a_t, q}$. We ran the MCMC for 7,000 iterations. To determine the burn-in, we monitor the adaptive κ parameter and the acceptance rate for each batch. These are included in Figure 2 where we observe that the κ becomes stable and the acceptance rate stabilizes between 0.3 and 0.4 after 60 batches. Therefore we set the burn-in to be equal to $3,000 = 60 \times 50$ iterations. This also confirms that the adaptation method proposed at the end of Section 6 performs well. Computing time is around 50 minutes for each run on an Intel Core i9 processor at 2.3 GHz with 32 GB of RAM.

To assess model performance, we computed two goodness of fit (gof) measures, the logarithm of the pseudo marginal likelihood (LPML) [8] and Watanabe–Akaike Information Criterion (WAIC) [34]. Table 1 shows these values. In general, the two gof measures concur

in determining the best model for each value of q . Briefly put, for smaller values of a_t , better fitting is achieved for larger orders of dependence q in the $\boldsymbol{\pi}$, whereas for larger values of a_t , smaller orders of dependence q produce better fit. Overall, the best model is, $\mathcal{M}_{30,7}$, obtained with $a_t = 30$ and $q = 7$.

Two more comparisons are also included in Table 1. The case of independence across times for $\pi_{t,\mathbf{j}}$ is obtained when $a_t = 0$, regardless of the value of q . Goodness of fit statistics show that the independence fitting is not the worst, but is definitely underperforming in comparison to the other dependence models. Additionally, we also consider the model that assumes independence in the $\theta_{t,\mathbf{j}}$. This is achieved by considering a very low variance in $\beta_{\mathbf{j}}$, as obtained by setting $e_{\mathbf{j}} = g_{\mathbf{j}} = 1000$. This latter model produces inferior gof measures as compared to the hierarchical priors.

To assess in more detail our model's performance, we compare posterior estimates of $\boldsymbol{\pi}$ and $\boldsymbol{\theta}$ with the true values. We use the best fitting model and use posterior means as point estimates, together with quantiles 2.5% and 97.5% to produce 95% credible intervals. Figure 3 (left panel) shows posterior estimates of $\pi_{t,\mathbf{j}}$ as time series for $t = 1, \dots, 20$ in four panels for $\mathbf{j} \in \{00, 10, 01, 11\}$. Posterior estimates follow very closely the path of the true values. Similarly, Figure 3 (right panel) shows posterior estimates of $\theta_{t,\mathbf{j}}$ as time series in four panels. All the true values lie within the 95% credible intervals. We note that credible intervals for $\theta_{t,00}$ are narrower at the beginning and become wider toward the end (bottom-left panel) whereas the credible intervals for $\theta_{t,10}$ are wider at the beginning and narrow toward the end (bottom-right panel). The credible intervals for $\theta_{t,11}$ remain wide overall (top-right panel), while those for $\theta_{t,01}$ are narrow (top-left panel) in general. Wider credible intervals are due to smaller weights (less data points) associated to the corresponding mixture components. Specifically, the higher variability for $\theta_{t,11}$ for larger t is a consequence of the smaller weights for the third component $\pi_{t,11}$.

We compare the best fit produced by $\mathcal{M}_{a_t,q}$ with the dynamic copula model of Hafner

& Manner [11]. This latter model assumes a Gaussian copula with time-varying association parameter ρ_t . It relies on a Fisher transformation (inverse hyperbolic tangent) of the association parameter as $\lambda_t = (1/2) \log((1 + \rho_t)/(1 - \rho_t))$ and models the dynamics via an autoregressive process of the form $\lambda_t = \alpha + \beta \lambda_{t-1} + \epsilon_t$ with $\epsilon_t \sim N(0, \tau)$. We perform a Bayesian analysis for this model with vague prior distributions $\alpha \sim N(0, 0.01)$, $\beta \sim N(0, 0.01)$ and $\tau \sim Ga(0.01, 0.01)$. We will refer to this model as dynamic Gaussian.

As a second competitor, we consider a simple Clayton copula, which is the result of assigning fixed weight one to the mixture model and keeping the exchangeable prior on the association copula time varying parameters.

To compare, we compute the log predictive scores (LPS) defined in [9]. We fit models with data up to time $t-1$ and compute the LPS for time t , that is, $LPS(t) = \sum_{i=1}^{n_t} \log f(\mathbf{u}_{t,i} | \mathbf{u}_{t-1})$, where the predictive distribution is approximated via Monte Carlo as

$$f(\mathbf{u}_{t,i} | \mathbf{u}_{t-1}) = (1/R) \sum_{r=1}^R f(\mathbf{u}_{t,i} | \boldsymbol{\pi}_t^{(r)}, \boldsymbol{\theta}_t^{(r)})$$

and $(\boldsymbol{\pi}_t^{(r)}, \boldsymbol{\theta}_t^{(r)})$ are obtained from the posterior distribution $f(\boldsymbol{\pi}_t, \boldsymbol{\theta}_t | \mathbf{u}_1, \dots, \mathbf{u}_{t-1})$ for a total of R iterations.

The LPS measures at $t = 20$ as well as the LPML and WAIC for the different models are reported in Table 2. The performance of the simple Clayton copula is the worst, as expected. The dynamic Gaussian improves a little the goodness of fit measures, but is far behind the dynamic Clayton mixtures. When removing the n_{20} data points from the fitting, the best performance is achieved when taking $a_t = 20$ and $q = 7$.

7.2 Bivariate real data analysis

We also assess how well our model can capture the relationship between variables in a real life application where data is generated from some unknown distribution, rather than directly from a mixture copula.

We used the Environment and Climate Change Canada (ECCC) data catalogue from the government of Canada. The ECCC managed the National Air Pollution Surveillance Program (NAPS), which began in 1969 and is now comprised of nearly 260 stations in 150 rural and urban communities reporting to the Canada-wide air quality database (for more details about the dataset, please visit <https://data-donnees.az.ec.gc.ca/data/air/monitor>).

Specifically, we selected ozone (O_3) and particulate matter with diameters 2.5 and smaller ($PM_{2.5}$) as the bivariate data. We used the hourly data from the years 2017 to 2019. Due to a large number of missing values, we took averages across hours and across days to produce monthly data for each station, $t = 1, \dots, T$ with $T = 36$ for a total time span of three years. The number of stations varies across months, specifically we have $n_t = 177$ for the months in 2017, $n_t = 187$ for 2018 and $n_t = 189$ for 2019.

Since our focus is not on the marginal distributions, but on the association between these two variables, we applied the modified rank transformation [6] to produce data in the interval $[0, 1]$. Specifically, if observed data is denoted by (X_{1ti}, X_{2ti}) , we compute the empirical cdf's, independently for each variable, say \hat{F}_{X_1} and \hat{F}_{X_2} and apply the inverse transformation $U_{1ti} = \hat{F}_{X_1}^{-1}(X_{1ti})$ and $U_{2ti} = \hat{F}_{X_2}^{-1}(X_{2ti})$.

To explore the data, we computed empirical Kendall's tau and Spearman's rho association coefficients per month. These are shown in Figure 2. In both cases the dependence is cyclical around zero, reaching its maximum in June/July and its minimum in October/November. This suggests that a seasonal specification of our model seems to be a good candidate to capture these cycles. For completeness we considered three types of models: moving average type of order q as in (11); seasonal model of order p with annual seasonality $s = 12$ as in (12); and a combination of moving average and seasonal. We denote the model $\mathcal{M}_{a_t, q, p}$ where the indexes describe chosen values for a_t , q and p .

Similar to the simulation study, we set the parameters $a_0 = 1$ and $e_k = g_k = 1$ to define the prior distributions. In this real data analysis, we also varied the dependence parameters

$a_t = 0, 1, \dots, 5$, $q = 0, 1, \dots, 4$ and $p = 0, 1, 2$ to assess the performance of the model under different strengths of temporal dependence. We ran the MCMC for 10,000 iterations and set the burn-in equal to 5,000. Computing time is around 60 minutes for each run in an Intel Core i9 processor at 2.3 GHz with 32 GB of RAM.

Table 3 shows the LPML and WAIC values for the different prior specifications. The two gof measures agree that the preferred model is $\mathcal{M}_{1,1,2}$, i.e. when $a_t = 1$, $q = 1$ and $p = 2$. In summary, a combination of moving average and seasonality are required to model this particular dataset.

In Figure 5 we show posterior estimates of the weights $\boldsymbol{\pi}$ (left panel) and copula coefficients $\boldsymbol{\theta}$ (right panel), respectively. The cyclical monthly dependence is captured by the weights. Since the first and third components of the mixture induce positive dependence, and second and fourth induce negative dependence, there is an opposite behaviour between the pairs $(\pi_{t,00}, \pi_{t,11})$ and $(\pi_{t,10}, \pi_{t,01})$. The former reaches its peak in the summer and the latter in the autumn-winter, however the peaks within the second pairs do not occur exactly at the same months, $\pi_{t,10}$ has its peak around September-October (autumn), whereas $\pi_{t,01}$ has its peak around December-January (winter). Among the four components, component 3 is the one with slightly smaller weights, apart from the summer of the year 2019 where $\pi_{t,11}$ has two peaks in May and September of around 0.75 and 0.5, respectively. Therefore, our mixture model is able to capture the seasonal dynamics in the data.

The strength of the association between the pair $(O_3, PM_{2.5})$ is determined by parameters $\theta_{t,j}$. Their posterior estimates are all around slightly below the value of one, with $\theta_{t,10}$ and $\theta_{t,01}$ showing more variability along time. Uncertainty in the estimation of $\theta_{t,11}$ is somehow higher, due to the slightly smaller weights $\pi_{t,11}$ and thus smaller sample size for estimating $\theta_{t,11}$. According to $\theta_{t,00}$, positive dependence is particularly higher in the summer of the years 2017 and 2019 with a lower-lower tail dependence. On the other hand, looking at $\theta_{t,10}$ and $\theta_{t,01}$, negative dependence is high in the winter of the three years.

We can further assess the tail dependence in the four corners of the unit square by computing the tail dependence coefficients λ_t , given in (8). These are reported in Figure 6. We first note that none of them is larger than 1/2, the only exception being the upper-lower $\lambda_{t,10}$ in October of 2018, where perhaps O_3 was extremely high and $PM_{2.5}$ was extremely low in that month. The lower-lower and upper-upper tail parameters $\lambda_{t,00}$ and $\lambda_{t,11}$ show very similar behaviour, they are most of the time close to zero, both with moderate peaks in July of the three years. On the other hand, the upper-lower and lower-upper tail parameters $\lambda_{t,10}$ and $\lambda_{t,01}$ do not behave exactly in the same way. These have wider peaks than the previous tail coefficients with moderate tail dependencies in the autumn for $\lambda_{t,10}$ and in the winter for $\lambda_{t,01}$.

Finally, we show joint density estimates in Figure 7 as heat plots, together with scatter plots of the data for each month t . We particularly concentrate on the months where the dependence changes from negative to positive. This transition is consistent along the three years of study for the months of June and July. It is interesting to see that August is a transition month, where in 2017 and 2019 the dependence is 4-way, i.e., the four mixture components of our model are present, in fact the estimated weights and coefficients are: $\boldsymbol{\pi}_{2017-8} = (0.41, 0.52, 0.02, 0.05)$, $\boldsymbol{\theta}_{2017-8} = (0.69, 0.69, 0.81, 0.59)$; $\boldsymbol{\pi}_{2018-8} = (0.07, 0.86, 0.04, 0.03)$, $\boldsymbol{\theta}_{2018-8} = (0.48, 0.31, 0.84, 0.50)$; $\boldsymbol{\pi}_{2019-8} = (0.11, 0.59, 0.09, 0.21)$, $\boldsymbol{\theta}_{2019-8} = (0.57, 0.31, 0.59, 0.58)$. What makes the heat plots to show the 4-way dependence is a combination of the weight $\pi_{t,j}$ and the intensity $\theta_{t,j}$.

We compare the fit of $\mathcal{M}_{1,1,2}$ with the dynamic Gaussian and simple Clayton copula and carry out two validation studies. In the first validation study we partition the dataset into two sets, fitting and testing. For each month $t = 1, \dots, T$ we took $n_{1t} = 140$ observations for fitting and $n_{2t} = n_t - 140$ for testing. We estimate the model parameters with the fitting set and use the testing set to predict O_3 conditional on the observed value of $PM_{2.5}$. For this we use the conditional copula $C_t(u_{1,t} \mid u_{2,t}, \boldsymbol{\pi}_t, \boldsymbol{\theta}_t) = \sum_j \pi_{t,j} C_k(u_{2,t} \mid u_{1,t}, \theta_{t,j})$

with $C_{\mathbf{j}}(u_{2,t} \mid u_{1,t}, \theta_{t,\mathbf{j}}) = (\partial/\partial u_{1,t})C_{\mathbf{j}}(u_{1,t}, u_{2,t} \mid \theta_{t,\mathbf{j}})$ for $\mathbf{j} \in \{00, 10, 01, 11\}$, and obtain the posterior predictive mean $\widehat{u}_{2,t} = \mathbb{E}(U_{2,t} \mid u_{1,t}, \boldsymbol{\pi}_t, \boldsymbol{\theta}_t)$ as point prediction.

We assess model performance by computing the mean square error between the observed $u_{2,t}$ and predicted $\widehat{u}_{2,t}$ for O_3 in the test set, i.e. $MSE = \left(\sum_{t=1}^T n_{2t} MSE_t \right) / \left(\sum_{t=1}^T n_{2t} \right)$, where $MSE_t = (1/n_{2t}) \sum_{i=1}^{n_{2t}} (u_{2,t} - \widehat{u}_{2,t})^2$, as well as the LPML and WAIC goodness of fit measures for the fitting sets. Results from the first validation study are included in Table 4. All four gof statistics for the dynamic Clayton mixture model are better than those obtained with the dynamic Gaussian and simple Clayton copula models, confirming that our model is preferred for the analysis.

The second validation study consists in comparing the log predictive scores (LPS) defined in [9]. We fit models with data up to time $t - 1$ and compute the LPS for time t . We repeat this for times $t = s + 1, \dots, T$ and aggregate the scores as

$$LPS = \sum_{t=s+1}^T \sum_{i=1}^{n_t} \log f(\mathbf{u}_{t,i} \mid \mathbf{u}_{t-1}).$$

In particular we took $s = 30$. The values of LPS are reported in Table 5. Here we observe that the simple Clayton has the worst predictive scores. For times $t = 31, 32, 33, 35$ the dynamic Gaussian copula obtains a better predictive score, and for times $t = 34, 36$ the dynamic Clayton mixtures has better performance. However, when aggregating the predictive scores for the six predicted times, our proposed model achieves the best performance.

7.3 Multivariate real data analysis

Using the same Canadian pollution repository ECCC, we selected $PM_{2.5}$, nitrogen dioxide (NO_2) and sulfur dioxide (SO_2) to perform a multivariate analysis with $m = 3$ variables. Again, hourly data were averaged to produce monthly data for years 2017 to 2019. In total we have $T = 36$ months and $n_t = 91$ stations for each month in 2017, $n_t = 104$ for 2018, and $n_t = 100$ for 2019. We applied the modified rank transformation to produce data in the interval $[0, 1]$.

We first performed an exploratory analysis and computed pairwise empirical Kendall's tau coefficients for the three variables. These are reported in Figure 8. Dependencies are not seasonal as in the bivariate case and the three coefficients fluctuate between negative and positive values. Largest positive dependence occurs between $PM_{2.5}$ and NO_2 , and lowest negative dependence between NO_2 and CO_2 .

Prior specifications, dependence search and MCMC settings were the same as in the bivariate real data analysis. Table 6 reports the LPML and WAIC gof statistics. Both measures agree on the best model which is $\mathcal{M}_{4,4,2}$, that is, a dynamic total mass $a_t = 4$, a moving average order of $q = 4$ and a seasonal order of $p = 2$. Although the seasonality was not clear in the exploratory analysis, the best model is using information from previous two years.

Posterior inference for parameters $\boldsymbol{\pi}$ and $\boldsymbol{\theta}$ are shown in Figure 9 obtained with the preferred model. We observe that component 000 (no rotation) dominates the mixture with the highest weight π_{000} of around 75%, followed by component 001 with the remaining 25%. The rest of the components have estimated weights close to zero, with component 111 having sometimes a small positive weight in the winter 2017-2018 and in spring of 2019. Copula coefficients were also estimated. Point estimates are all different than zero for all mixture components, however credible intervals are a lot narrower for those components with high weight. Largest $\boldsymbol{\theta}$ values correspond to mixture 110, but the uncertainty is too high (wide credible intervals) to be trusted.

Joint density estimates are presented as pairwise heat maps in Figure 10. We only show the last trimester of 2017 and each row corresponds to a different pair, as follows: $(PM_{2.5}, NO_2)$ in the top row, $(PM_{2.5}, SO_2)$ in the middle row, and (NO_2, SO_2) in the bottom. In all bivariate density estimates we see the presence of tail dependence in the 00 (lower-lower) corner, whereas in the second and third row we see a 01 (lower-upper) tail dependence. This corresponds to SO_2 which is plotted in the y-axis and has a negative dependence with

the other two variables $PM_{2.5}$ and CO_2 . This combination of tail dependencies makes the heatmaps show a spider effect towards the bottom-left and upper-left corners in September and November of 2017. Moreover, in October of 2017, for the pairs $(PM_{2.5}, SO_2)$ and (NO_2, SO_2) , a combination of three trends can be identified with three tail dependencies in the 00, 01 and 11 corners.

Certainly, this non standard dependence among these three pollution variables could not be captured by a single copula model.

8 Concluding remarks and future work

We extend a copula's versatility in capturing dependence patterns using mixtures of copulas with a dynamic component in the weights. The idea is illustrated using Clayton copulas, but it can be applied to any other families. The motivation is given by problems where different extreme regions of the pairwise bivariate distributions exhibit patterns that cannot be captured by a single copula.

In situations in which the dependence varies in time, we propose a dynamic mixture of copulas model in which the mixture weights and the parameters of the copulas involved in the mixture are modelled either through a moving average or a seasonal dynamic. The resulting increase in modelling flexibility is illustrated by all our numerical experiments, be they synthetic or real.

Dependence in our dynamic Dirichlet prior on the weights is controlled by the triplet (a_t, q, p) . For the analyses considered here we have kept a_t to be constant across time, to make our prior easy to define. However, this parameter can certainly be chosen to be different across time, providing additional flexibility in the model. We have left this substantial generalization for future work.

Additionally, special care has to be put for the analysis of multivariate data for large dimension m , since the number of components in the mixture increases exponentially at a

rate 2^m . In practice, not all extremes are likely to be significant. In order to impose sparsity, we plan to exploit a sparse prior designed to reduce the number of components needed to model the data and still maintain the added flexibility demonstrated in this work. An added question of interest is the identification of lower dimensional manifolds where a mixture of lower-dimensional copulas can be used to capture the dependence in the data.

Acknowledgements

This work was supported by *Asociación Mexicana de Cultura, A.C.* while the second author was visiting the Department of Statistical Sciences at the University of Toronto and by an NSERC of Canada discovery grant of the third author. We are also grateful to Jun Young Park and Patrick Brown for their guidance on potential applications.

References

- [1] Aas, K., Czado, C., Frigessi, A. and Bakken, H.. Pair-copula constructions of multiple dependence. *Insurance: Mathematics & Economics* **44**, 182–198.
- [2] Bedford, T., Cooke, R.M. (2002). Vines—a new graphical model for dependent random variables. *Annals of Statistics* **30**, 1031–1068.
- [3] Bernardo, J.M. and Smith, A.M.F. (2000). *Bayesian theory*. Wiley, London.
- [4] Craiu, V.R. and Sabeti, A. (2012). In mixed company: Bayesian inference for bivariate conditional copula models with discrete and continuous outcomes. *Journal of Multivariate Analysis* **110**, 106–120.
- [5] Czado, Claudia and Nagler, Thomas (2022). Vine copula based modeling. *Annual Review of Statistics and Its Application* **9**, 453–477.

[6] Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance. *Bulletin Royale Belge de l'Académie des Sciences* **65**, 274–292.

[7] Frahm, G., Junker, M. and Schmidt, R. (2005). Estimating the tail-dependence coefficient: Properties and pitfalls. *Insurance: Mathematics and Economics* **37**, 80–100.

[8] Geisser, S. and Eddy, W.F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association* **74**, 153–160.

[9] Geweke, J. and Amisano, G. (2010). Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting* **26**, 216–230.

[10] González-Barrios, J.M. and Hoyos-Argüelles, R. (2018). Distributions associated to the counting techniques of the d-sample copula of order m and weak convergence of the sample process. *Communications in Statistics - Simulation and Computing* **49**, 2505–2532.

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

[12] Hasler, C., Craiu, R.V. and Rivest, L.P. (2018). Vine copulas for imputation of monotone non-response. *International Statistical review* **86**, 448–511.

[13] Hoyos-Argüelles, R. and Nieto-Barajas, L.E. (2020). A Bayesian semiparametric archimedean copula. *Journal of Statistical Planning and Inference* **206**, 298–311.

[14] Hu, L. (2006). Dependence patterns across financial markets: A mixed copula approach. *Applied Financial Economics* **16**, 717–729.

[15] Joe, H. (1997). *Multivariate models and dependence concepts*. Chapman and Hall, London.

[16] Klein, N., Kneib, T., Marra, G. and Radice, R. (2020). Bayesian mixed binary-continuous copula regression with an application to childhood under nutrition. In *Flexible Bayesian regression modeling*. Y.Fan, D.Nott, M.S.Smith, J.L. Dortet-Bernadet (eds.) 121–152.

[17] Kruskal, W.H. (1958). Ordinal measures of association. *Journal of the American Statistical Association* **53**, 814–861.

[18] Lambert, P. and Vandenhende, F. (2002). A copula-based model for multivariate non-normal longitudinal data: analysis of a dose titration safety study on a new antidepressant. *Statistics in Medicine* **21**, 3197–3217.

[19] Levi, E. and Craiu, R.V. (2018). Bayesian inference for conditional copulas using Gaussian process single index models. *Computational Statistics and Data Analysis* **122**, 115–134.

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

[21] Loaiza-Maya, R., Smith, M.S. and Maneesoonthorn, W. (2018). Time series copulas for heteroskedastic data. *Journal of Applied Econometrics* **33**, 332–354.

[22] Nelsen, R.B. (2006). *An introduction to copulas*. Springer, New York.

[23] Nieto-Barajas, L.E. (2024). Multivariate and regression models for directional data based on projected Pólya trees. *Statistics and Computing*. To appear.

[24] Nieto-Barajas, L.E. and Hoyos-Argélles, R. (2023). Generalised Bayesian sample copula of order m. *Computational Statistics*. To appear.

[25] Nieto-Barajas, L.E. and Gutiérrez-Peña, E. (2022). General dependence structures for some models based on exponential families with quadratic variance functions. *TEST* **31**, 699–716.

[26] Oh, D.H. and Patton, A.J. (2016). High-dimensional copula-based distributions with mixed frequency data. *Journal of Econometrics* **193**, 349–366.

[27] Roberts, G.O. and Rosenthal, J.S. (2007). Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms. *Journal of Applied Probability*, **44**, 458–475.

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

[29] Smith, M.S. (2015). Copula modelling of dependence in multivariate time series. *International Journal of Forecasting* **31**, 815–833.

[30] Smith, A. and Roberts, G. (1993). Bayesian computations via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B* **55**, 3–23.

[31] Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B* **64**, 583–639.

[32] Tierney, L. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics* **22**, 1701–1722.

[33] Tjøstheim, D., Otneim, H. and Støve, B. (2022). Local Gaussian correlation and the copula. In *Statistical Modeling Using Local Gaussian Approximation*. D.Tjøstheim, H.Otneim, B.Støve (eds.) Academic Press, 135–159.

[34] Watanabe, S. and Opper, M. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research* **11**, 3571–3594.

e_k, g_k	a_t	$q = 0$	$q = 1$	$q = 2$	$q = 3$	$q = 4$	$q = 5$	$q = 6$	$q = 7$
LPML									
1	0	1685							
1	1	1686	1683	1684	1688	1687	1692	1691	1696
1	10	1693	1698	1702	1702	1706	1705	1707	1706
1	20	1696	1702	1705	1705	1707	1706	1707	1706
1	30	1699	1704	1706	1708	1705	1708	1706	1708
1	40	1702	1705	1704	1708	1706	1705	1705	
1000	30	1682	1688	1691	1693	1693	1691	1693	1692
WAIC									
1	0	-3372							
1	1	-3374	-3368	-3370	-3378	-3375	-3384	-3384	-3393
1	10	-3388	-3397	-3403	-3405	-3413	-3410	-3413	-3411
1	20	-3393	-3405	-3409	-3411	-3415	-3413	-3414	-3412
1	30	-3398	-3409	-3411	-3416	-3410	-3415	-3413	-3416
1	40	-3403	-3410	-3409	-3416	-3412	-3410	-3409	-3411
1000	30	-3364	-3376	-3382	-3386	-3386	-3382	-3385	-3385

Table 1: Simulated data. LPML and WAIC gof values of different $\mathcal{M}_{a_t, q}$ models. Best values are bolded.

Model / Measure	LPS	LPML	WAIC
$\mathcal{M}_{2,2}$	73.2	1603	-3209
$\mathcal{M}_{20,7}$	74.9	1625	-3250
$\mathcal{M}_{30,7}$	73.1	1624	-3247
D.Gaussian	25.4	146	-292
S.Clayton	-12.3	9	-18

Table 2: Simulated data. LPS statistic using times 1, ..., 19 for fitting and $t = 20$ for prediction. The other gof measures, LPML and WAIC, were calculated with the fitting data.

a_t	LPML					WAIC				
	$MA(0)$	$MA(1)$	$MA(2)$	$MA(3)$	$MA(4)$	$MA(0)$	$MA(1)$	$MA(2)$	$MA(3)$	$MA(4)$
	$S(0)$					$S(1)$				
0	435					-871				
1	433	438	436	434	434	-867	-875	-873	-868	-867
2	433	433	433	426	422	-866	-867	-866	-852	-843
3	432	432	428	421	416	-864	-863	-857	-843	-832
4	432	427	421	417	410	-863	-855	-841	-834	-820
5	429	424	417	409	403	-858	-849	-834	-817	-805
	$S(2)$					$S(1)$				
0	435					-871				
1	440	441	439	435	433	-880	-881	-878	-869	-866
2	439	435	433	427	422	-877	-871	-865	-854	-843
3	434	432	427	420	414	-868	-865	-854	-839	-829
4	432	428	422	414	408	-863	-857	-844	-829	-815
5	427	424	416	408	401	-854	-848	-832	-815	-802
	$S(2)$					$S(1)$				
0	435					-871				
1	440	442	438	433	430	-879	-883	-876	-867	-859
2	437	433	433	427	420	-874	-865	-865	-854	-840
3	434	432	427	419	413	-867	-864	-853	-838	-826
4	430	425	424	412	407	-861	-850	-847	-824	-813
5	426	422	418	408	397	-852	-844	-836	-817	-794

Table 3: Bivariate real data. LPML and WAIC statistics for different prior choices for $\mathcal{M}_{at,q,p}$.

Model / Measure	MSE	LPML	WAIC
$\mathcal{M}_{1,1,2}$	0.0751	397	-794
D.Gaussian	0.0763	383	-767
S.Clayton	0.0826	39	-78

Table 4: Bivariate real data. Goodness of fit measures in first validation study.

Model	$t = 31$	$t = 32$	$t = 33$	$t = 34$	$t = 35$	$t = 36$	Total
$\mathcal{M}_{1,1,0}$	-10.5	-3.1	-9.1	17.8	21.6	11.4	28.2
$\mathcal{M}_{1,2,0}$	-3.2	-4.7	-12.1	21.8	25.3	12.2	39.3
$\mathcal{M}_{1,1,1}$	-12.4	-3.8	-10.9	16.1	24.0	11.9	24.9
$\mathcal{M}_{1,2,1}$	-5.9	-2.7	-11.6	18.5	26.8	11.6	36.7
$\mathcal{M}_{1,2,2}$	-8.2	-3.1	-10.0	15.4	26.8	12.3	33.2
D.Gaussian	5.2	-1.8	-1.3	-7.1	37.5	6.1	38.8
S.Clayton	-79.5	-170.5	-153.6	-258.2	-246.1	-199.8	-1107.6

Table 5: Bivariate real data. LPS statistic computed by fitting models from time 1 to $t - 1$ and predicting time t .

a_t	LPML					WAIC				
	$MA(0)$	$MA(1)$	$MA(2)$	$MA(3)$	$MA(4)$	$MA(0)$	$MA(1)$	$MA(2)$	$MA(3)$	$MA(4)$
	$S(0)$					$S(1)$				
0	271					-542				
1	286	289	293	294	294	-573	-579	-589	-588	-589
2	281	278	278	278	276	-562	-557	-557	-556	-553
3	287	291	289	292	291	-574	-582	-579	-586	-583
4	290	299	293	297	297	-581	-601	-587	-594	-595
5	282	282	281	282	282	-565	-565	-563	-565	-565
$S(2)$										
0	271					-542				
1	288	288	291	295	287	-576	-577	-583	-590	-575
2	279	279	278	278	276	-559	-558	-557	-557	-553
3	289	289	292	292	294	-579	-579	-584	-586	-589
4	292	294	298	297	293	-586	-591	-597	-594	-587
5	283	282	283	283	283	-568	-565	-567	-566	-567

Table 6: Multivariate real data. LPML and WAIC statistics for different prior choices for $\mathcal{M}_{a_t, q, p}$.

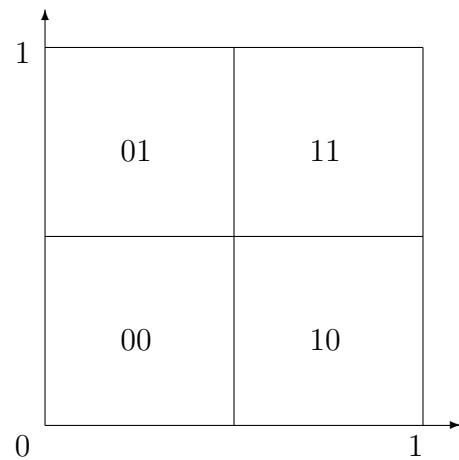


Figure 1: Unit square divided in four quadrants.

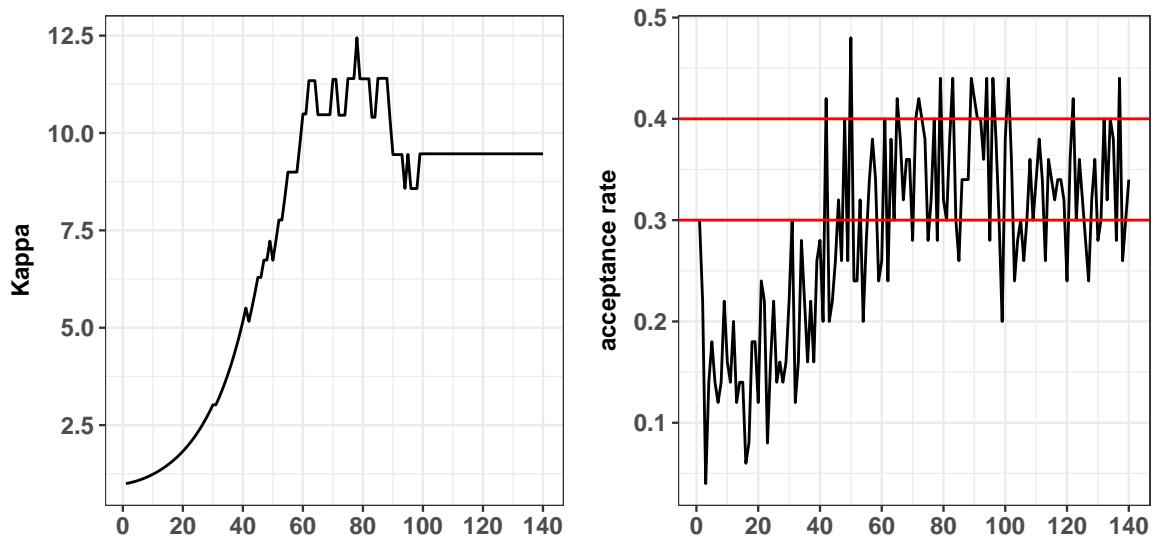


Figure 2: The recorded $\kappa^{(h)}$ and acceptance rate for each batch h . The batch size is 50.

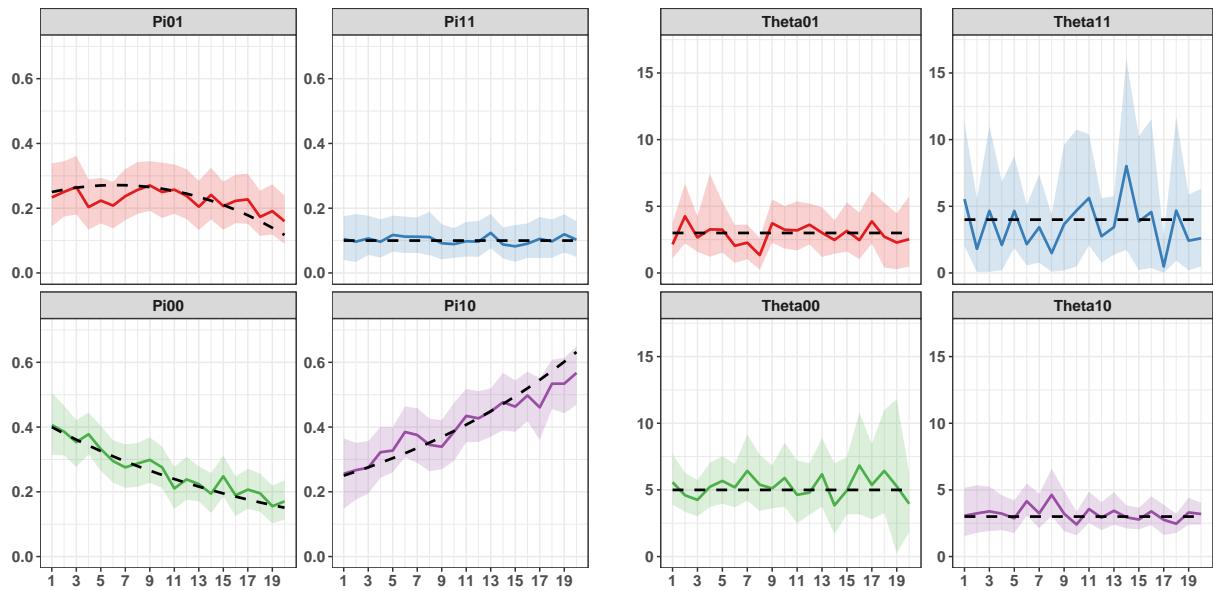


Figure 3: Simulated data. Posterior estimates of π (left) and θ (right): posterior mean (solid line) with 95% credible intervals (shadows), together with the true value (dotted black line).

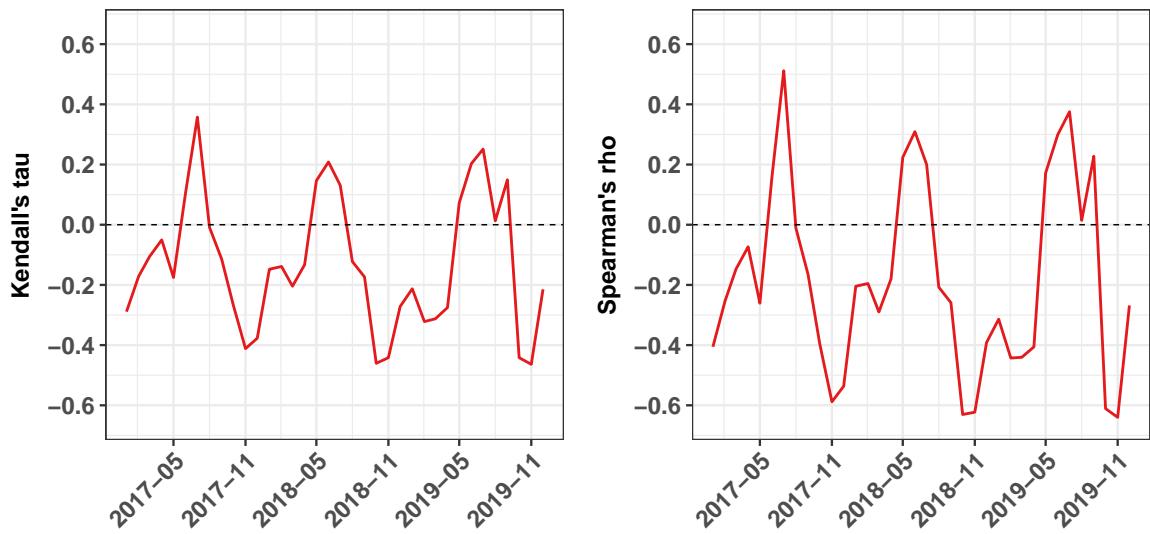


Figure 4: Bivariate real dataset. Empirical Kendall's tau (left) and Spearman's rho (right).

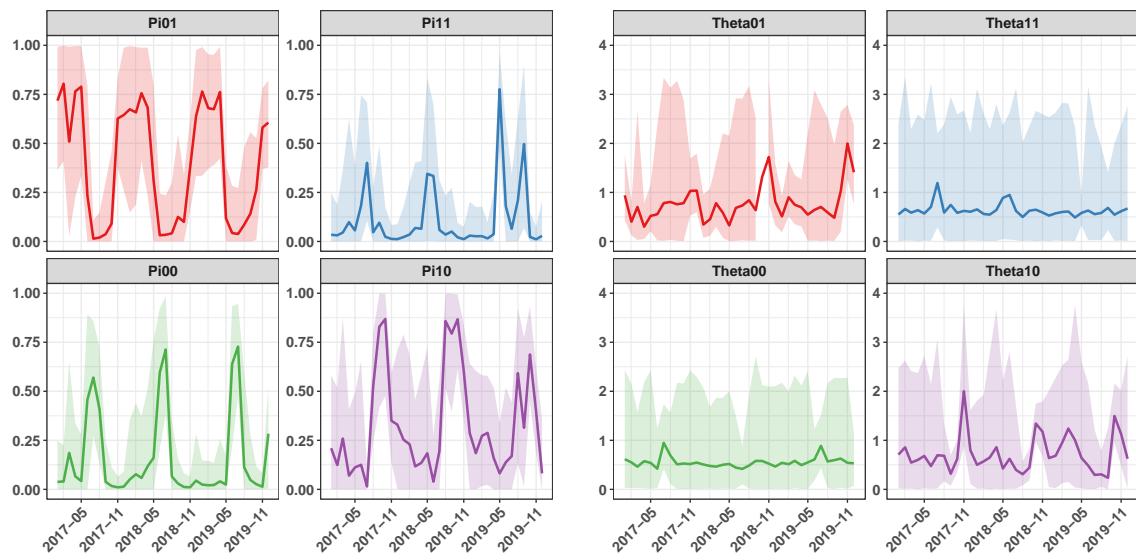


Figure 5: Bivariate real dataset. Posterior estimates of π (left) and θ (right): posterior mean (solid line) with 95% credible intervals (shadows).

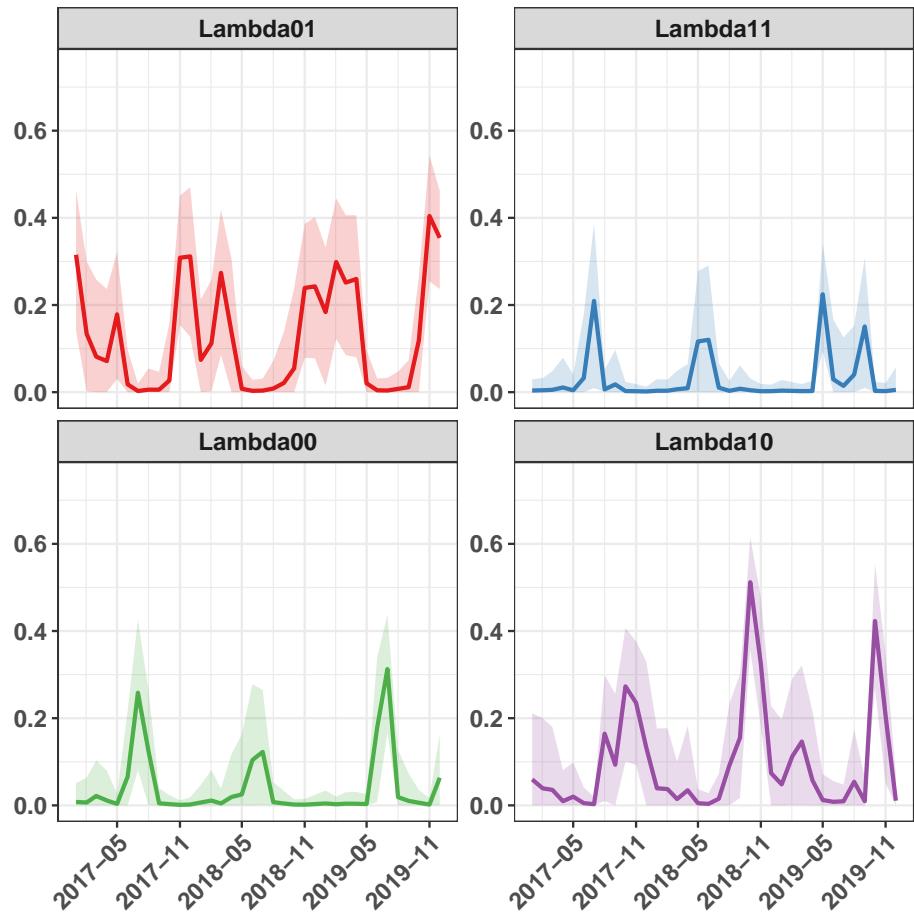


Figure 6: Bivariate real dataset. Posterior estimates of λ : posterior mean (solid line) with 95% credible intervals (shadows).

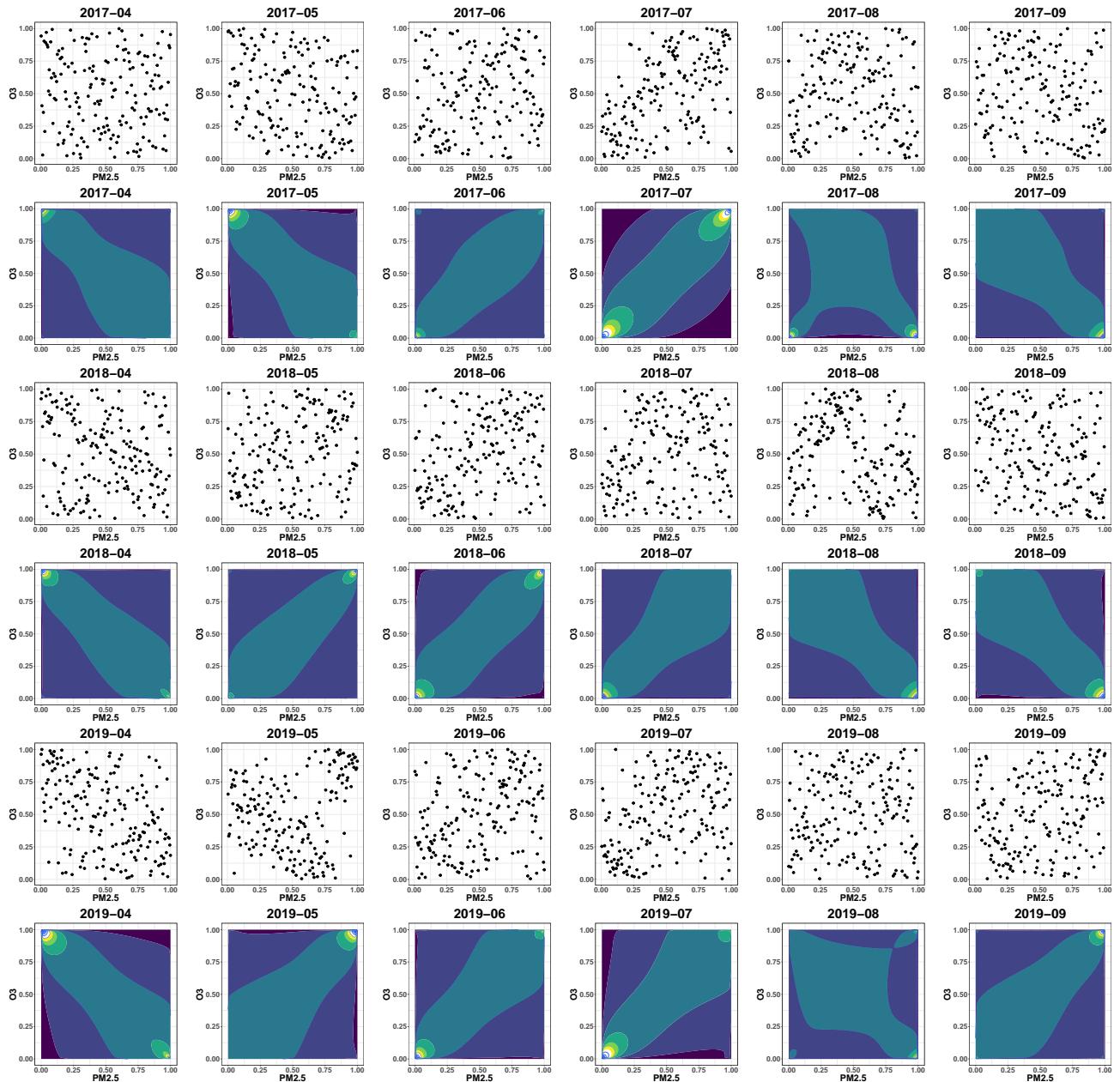


Figure 7: Observed real data and estimated joint density from April to September 2017 (top), 2018 (middle), and 2019 (bottom). Dependence patterns tend to vary between seasons with some months exhibiting transitional regimes.

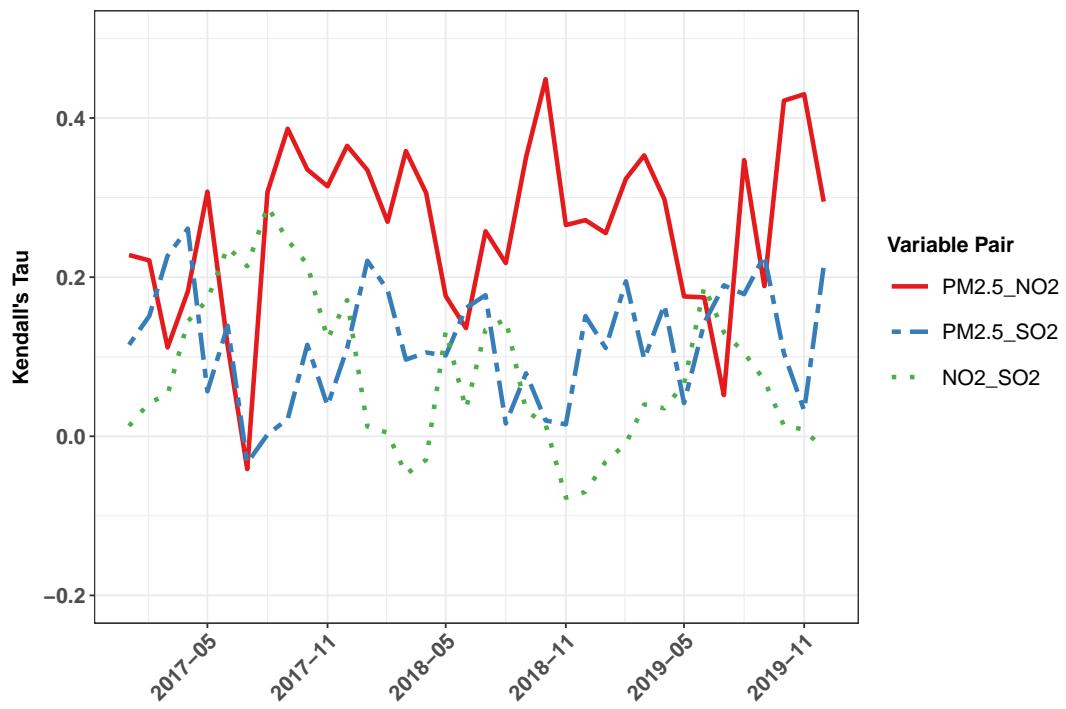


Figure 8: Multivariate real dataset. Pairwise empirical Kendall's tau.

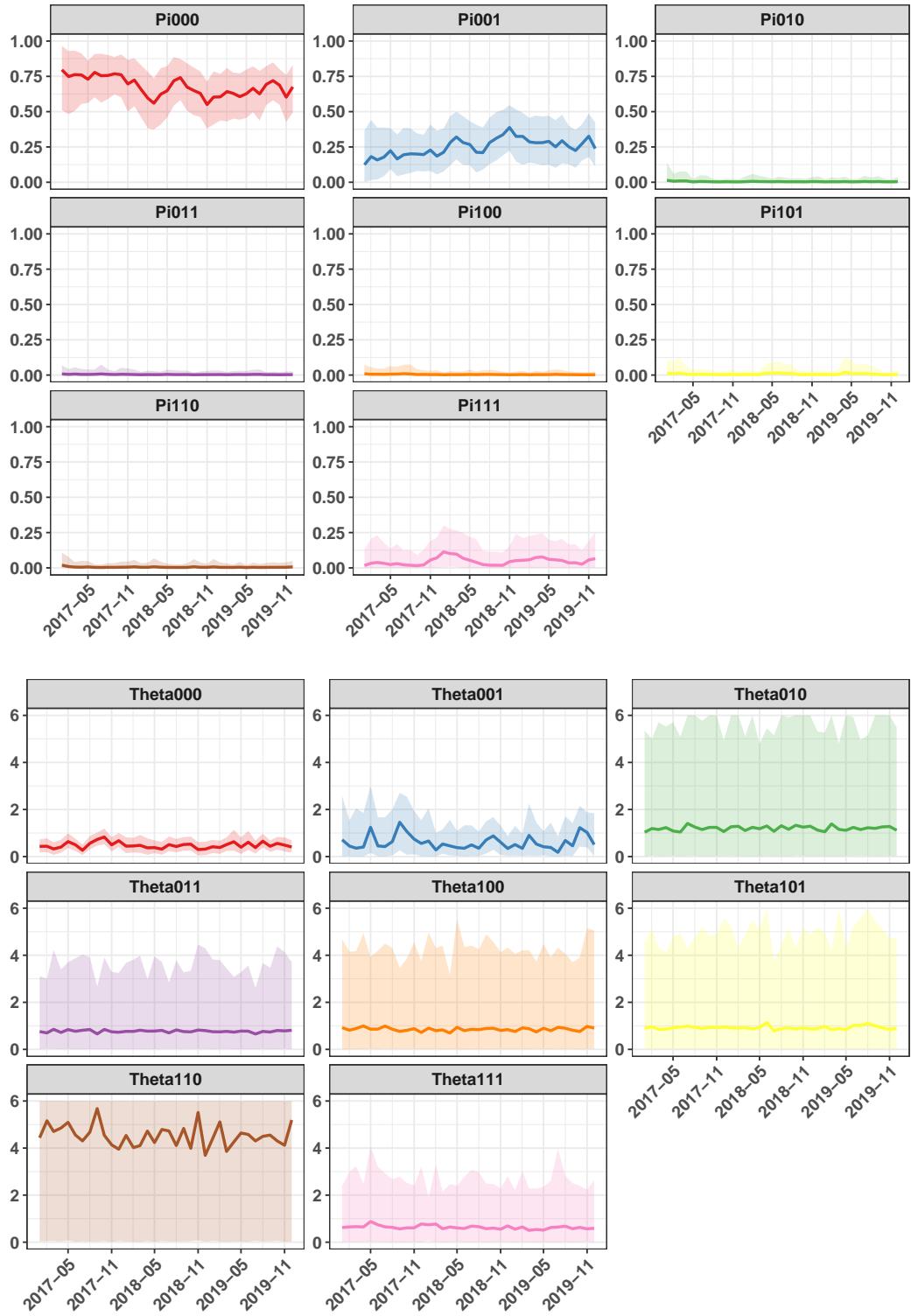


Figure 9: Multivariate real dataset. Posterior estimates of π (top) and θ (down): posterior mean (solid line) with 95% credible intervals (shadows).

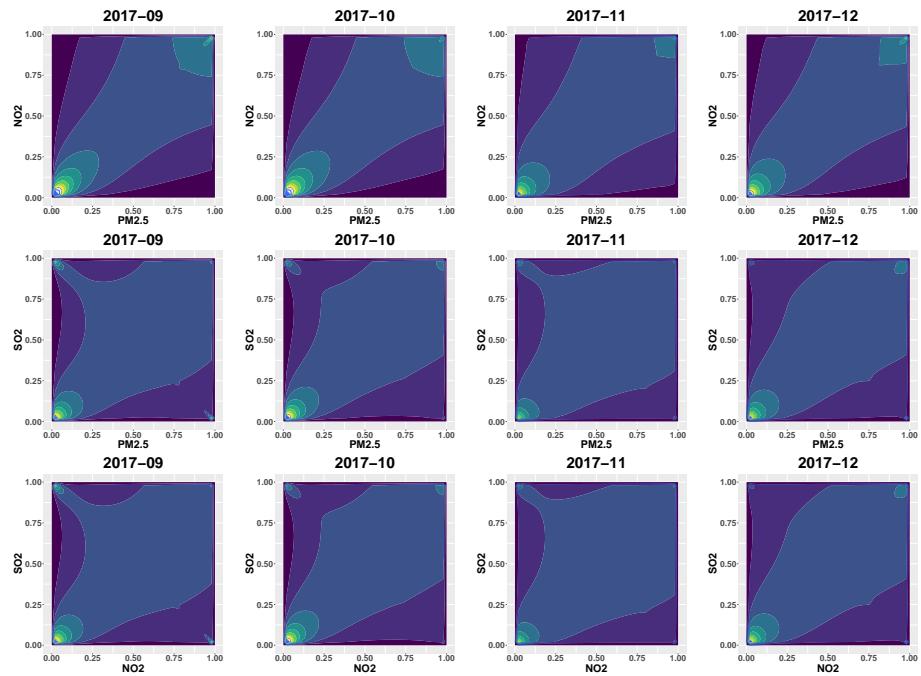


Figure 10: Multivariate real data. Estimated joint density for months September to December of 2017. $(PM_{2.5}, NO_2)$ (top row), $(PM_{2.5}, SO_2)$ (middle row) and (NO_2, SO_2) (bottom row).