Bivariate temporal dependence via mixtures of rotated copulas

Ruyi Pan, Luis E. Nieto-Barajas, Radu Craiu

University of Toronto & ITAM

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

Abstract

Parametric bivariate 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. We propose a mixture of 4-way 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 modeling 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 bivariate 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. The first attempt was done by [15] who 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 the mixture [20].

A survival copula is a 180^o rotation of any 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^o and 270^o, are also possible. For instance, [16] considered all four rotations of a Clayton copula 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 $0^o - 180^o$ and $90^o - 270^o$, 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 $\mathbf{u} = (u_1, \ldots, u_T)$ as $c(\mathbf{u}) = \prod_{t=1}^{T} f(u_t \mid u_{t-1})$ assuming Markov conditional distributions. They further used a bivariate copula to model the transitions, i.e. $f(u_t \mid u_{t-1}) = c_2(u_{t-1}, u_t)$. Specifically for the bivariate copula they use a mixture of rotations of degrees $0^o - 90^o$.

In this paper, we first generalise the concept of bivariate tail dependence to the four corners of the unit square and propose a flexible copula that is able to capture multiple types of tail dependence. Our goal is achieved by mixing the four rotated instances of the Clayton copula, to 0, 90, 180 and 270 degrees. Furthermore, the 4-dimensional mixture weights $\boldsymbol{\pi}_t$ are all different and are allowed to change over time through a moving average type of process of order q and a seasonal cponent 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. We monitor the dependence of several individuals in time, and individuals might not be the same in different times. So our context is a time series of distributions (copulas).

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 the four-way tail dependence and show that the Clayton rotations measure the four types of tail dependence. In Section 4 we define our mixture model for a specific time and define the time dependent mixture weights and association parameters. Section 5 provides prior and posterior distributions to achieve a Bayesian analysis of the model and an illustration of its performance is reported in Section 6. Section 7 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 completed via a unique copula $C : [0,1]^m \to [0,1]$. 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 $F_1, \ldots, F_m$, then $F(x_1, \ldots, x_m) = C\{F_1(x_1), \ldots, F_m(x_m)\}$. Additionally, the copula function can be obtained as $C(u_1, \ldots, u_m) = F\{F_1^{-1}(u_1), \ldots, F_m^{-1}(u_m)\}$, where F_j^{-1} for $j = 1, \ldots, m$ 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

$$\lambda_{UU} = \lim_{\nu \to 0} \mathrm{P}\{X_1 > F_1^{-1}(1-\nu) \mid X_2 > F_2^{-1}(1-\nu)\}$$

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

$$\lambda_{LL} = \lim_{\nu \to 0} \mathrm{P}\{X_1 \leq F_1^{-1}(\nu) \mid X_2 \leq F_2^{-1}(\nu)\}$$

for the lower-down (lower-lower) corner (the sub indexes U and L stand for upper and lower), respectively. However, it is possible that both variables have comovements 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 dependence is defined as

$$\lambda_{UL} = \lim_{\nu \to 0} \mathrm{P}\{X_1 > F_1^{-1}(1-\nu) \mid X_2 \leq F_2^{-1}(\nu)\}$$

for the upper-lower corner, and

$$\lambda_{LU} = \lim_{\nu \to 0} \mathrm{P}\{X_1 \leq F_1^{-1}(\nu) \mid X_2 > F_2^{-1}(1-\nu)\}$$

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

$$\begin{aligned} \lambda_{UU} &= \lim_{\nu \to 0} \frac{2\nu - 1 + C(1-\nu, 1-\nu)}{\nu}, \quad \lambda_{LL} = \lim_{\nu \to 0} \frac{C(\nu, \nu)}{\nu}, \\ \lambda_{UL} &= \lim_{\nu \to 0} \frac{\nu - C(1-\nu, \nu)}{\nu}, \quad \lambda_{LU} = \lim_{\nu \to 0} \frac{\nu - C(\nu, 1-\nu)}{\nu}. \end{aligned} \tag{1}$$

It is 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 $\mathrm{Ga}(\alpha, \beta)$ denote a gamma density with mean α/β, $\mathrm{Be}(\alpha, \beta)$ a beta density with mean $\alpha/(\alpha+\beta)$, $\mathrm{Dir}(\boldsymbol{\alpha})$ a Dirichlet density with parameter vector $\boldsymbol{\alpha}$, and $\mathrm{Mult}(c, \mathbf{p})$ 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. $\mathrm{Ga}(x \mid \alpha, \beta)$, in the gamma case.

3 Four-way tail dependence

In what follows, we illustrate the four-way mixture using the Clayton copula, but the construction is general and can be applied to other copula families.

Let C be a copula in the Clayton family, indexed by parameter θ and 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 lower-lower tail coefficient is $\lambda_{LL} = 2^{-1/\theta}$, and the Kendall's tau association parameter is $\tau = \theta/(2 + \theta)$. Furthermore, the Clayton family is in the class of Archimedean copulas with generator $\varphi(t) = t^{-\theta} - 1$.

To understand the rotations, let us consider the unit square $[0, 1]^2$ and divide it into four quadrants as in Figure 1. To define the 90-degree rotation, we consider the probability in quadrant II, $\mathrm{P}(U_1 > u_1, U_2 \leq u_2) = \mathrm{P}(U_2 \leq u_2) - \mathrm{P}(U_1 \leq u_1, U_2 \leq u_2)$, 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

$$C_{II}(u_1, u_2) = \mathrm{P}(U_1' \leq u_1, U_2 \leq u_2) = u_2 - C(1 - u_1, u_2).$$

To define the 180-degree rotation, we consider the probability in the quadrant III, $\mathrm{P}(U_1 > u_1, U_2 > u_2) = 1 - \mathrm{P}(U_1 \leq u_1) - \mathrm{P}(U_2 \leq u_2) + \mathrm{P}(U_1 \leq u_1, U_2 \leq u_2)$, 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)

$$C_{III}(u_1, u_2) = \mathrm{P}(U_1' \leq u_1, U_2' \leq u_2) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2).$$

Lastly, to define the 270-degree rotation we consider the probability in the quadrant IV, $\mathrm{P}(U_1 \leq u_1, U_2 > u_2) = \mathrm{P}(U_1 \leq u_1) - \mathrm{P}(U_1 \leq u_1, U_2 \leq u_2)$, 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)

$$C_{IV}(u_1, u_2) = \mathrm{P}(U_1 \leq u_1, U_2' \leq u_2) = u_1 - C(u_1, 1 - u_2).$$

For completeness, we denote the original, un-rotated, copula as $C_I(u_1, u_2)$.

It is not difficult to prove that each of the previous four rotated copulas have the same tail dependence coefficients, but in different corners, i.e., lower-lower tail coefficient for copula I, upper-lower tail coefficient for copula II, upper-upper tail coefficient for copula III and lower-upper tail coefficient for copula IV are the same. Using (1), the tail coefficients become

$$\lambda_{LL}^{I} = \lambda_{UL}^{II} = \lambda_{UU}^{III} = \lambda_{LU}^{IV} = \lim_{\nu \to 0} \frac{\left(2\nu^{-\theta} - 1\right)^{-1/\theta}}{\nu} = \lim_{\nu \to 0} (2 - \nu^{\theta})^{-1/\theta} = 2^{-1/\theta}, \tag{2}$$

for $\theta > 0$, and any other tail dependence coefficients are zero for the four rotated copulas.

4 Dynamic Clayton mixtures

Let (U_{1t}, U_{2t}) be a bivariate vector with Unif$(0, 1)$ marginal distributions for each $t = 1, 2, \ldots, T$. The idea is to model the joint distribution between U_{1t} and U_{2t} through a flexible copula C_t which is able to capture any kind of tail dependence as it evolves in time. For that we define the following mixture copula

$$C_t(u_{1t}, u_{2t} \mid \boldsymbol{\pi}_t, \boldsymbol{\theta}_t) = \sum_{k=1}^{4} \pi_{tk} C_k(u_{1t}, u_{2t}, \mid \theta_{tk}), \tag{3}$$

with parameters $\boldsymbol{\pi}_t = (\pi_{t1}, \ldots, \pi_{t4})$ and $\boldsymbol{\theta}_t = (\theta_{t1}, \ldots, \theta_{t4})$, where C_k is a rotated Clayton copula as defined in Section 3, but expressed in arabic numbers instead of roman numbers for simplicity, with a different association parameter $\theta_{tk} > 0$ for each rotated copula $k = 1, \ldots, 4$. Specifically,

$$\begin{aligned}
C_1(u_1, u_2 \mid \theta_1) &= \left(u_1^{-\theta_1} + u_2^{-\theta_1} - 1\right)^{-1/\theta_1}, \\
C_2(u_1, u_2 \mid \theta_2) &= u_2 - \left\{(1 - u_1)^{-\theta_2} + u_2^{-\theta_2} - 1\right\}^{-1/\theta_2}, \\
C_3(u_1, u_2 \mid \theta_3) &= u_1 + u_2 - 1 + \left\{(1 - u_1)^{-\theta_3} + (1 - u_2)^{-\theta_3} - 1\right\}^{-1/\theta_3}, \\
C_4(u_1, u_2 \mid \theta_4) &= u_1 - \left\{u_1^{-\theta_4} + (1 - u_2)^{-\theta_4} - 1\right\}^{-1/\theta_4}.
\end{aligned} \tag{4}$$

Parameters $\pi_{tk} > 0$ are mixture weights such that $\pi_{t1} + \pi_{t2} + \pi_{t3} + \pi_{t4} = 1$ for $t = 1, \ldots, T$.

It is not difficult to derive association coefficients, like the 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 for the mixture copula (3) is

$$\begin{aligned}\tau_t &= 4\mathrm{E}\{C_t(U_{1t}, U_{2t} \mid \boldsymbol{\pi}_t, \boldsymbol{\theta}_t)\} - 1 = 4\sum_{k=1}^{4} \pi_{tk}\mathrm{E}\{C_k(U_{1t}, U_{2t} \mid \boldsymbol{\theta}_t)\} - 1 \\ &= \sum_{k=1}^{4} \pi_{tk}\left[4\mathrm{E}\{C_k(U_{1t}, U_{2t}, \mid \boldsymbol{\theta}_t)\} - 1\right] = \sum_{k=1}^{4} \pi_{tk}\tau_{tk},\end{aligned}$$

where τ_{tk} is the individual Kendall's tau for each of the mixture copula components C_k. Since all mixture components C_k as in (4) are Clayton copulas, individual Kendall's tau coefficient are $\tau_{tk} = \theta_{tk}/(2+\theta_{tk})$ for $k = 1, 3$ and $\tau_{tk} = -\theta_{tk}/(2+\theta_{tk})$ for $k = 2, 4$. Therefore, the Kendall's tau coefficient for the mixture copula (3) becomes

$$\tau_t = \pi_{t1}\frac{\theta_{t1}}{2+\theta_{t1}} - \pi_{t2}\frac{\theta_{t2}}{2+\theta_{t2}} + \pi_{t3}\frac{\theta_{t3}}{2+\theta_{t3}} - \pi_{t4}\frac{\theta_{t4}}{2+\theta_{t4}}. \tag{5}$$

Now, tail dependence coefficients (1) for mixture copula (3) at time t, considering the upper-upper tail, becomes

$$\begin{aligned}\lambda_{UU,t} &= \lim_{\nu\to 0}\frac{2\nu - 1 + C_t(1-\nu, 1-\nu)}{\nu} = \lim_{\nu\to 0}\frac{2\nu - 1 + \sum_{k=1}^{4}\pi_{tk}C_k(1-\nu, 1-\nu)}{\nu} \\ &= \sum_{k=1}^{4}\pi_{tk}\lim_{\nu\to 0}\left\{\frac{2\nu - 1 + C_k(1-\nu, 1-\nu)}{\nu}\right\} = \sum_{k=1}^{4}\pi_{tk}\lambda_{UU,t}^{(k)},\end{aligned}$$

where $\lambda_{UU,t}^{(k)}$ is the individual upper-upper tail dependence coefficient for individual copula C_k at time t. Similar equations are obtained for the other three tail dependence coefficients. Therefore, considering that the only tail dependence coefficients for the mixture components of (3) are those in (2), the overall tail dependence coefficients in the case of the Clayton family are

$$\lambda_{t1} = \pi_{t1}\,2^{-1/\theta_{t1}}, \quad \lambda_{t2} = \pi_{t2}\,2^{-1/\theta_{t2}}, \quad \lambda_{t3} = \pi_{t3}\,2^{-1/\theta_{t3}} \quad \text{and} \quad \lambda_{t4} = \pi_{t4}\,2^{-1/\theta_{t4}} \tag{6}$$

which denote lower-lower, upper-lower, upper-upper and lower-upper tail dependencies, respectively.

In summary, our mixture copula proposal (3) is flexible enough to capture a larger class of dependence associations and the four tail dependencies, according to the copula parameters π_{tk} and θ_{tk} for $k = 1, \ldots, 4$ and $t = 1, \ldots, T$.

5 Bayesian analysis

5.1 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 $\boldsymbol{\pi}_t = (\pi_{t1}, \pi_{t2}, \pi_{t3}, \pi_{t4})$. Since $\sum_{k=1}^{4} \pi_{tk} = 1$ for all t, the natural marginal prior for $\boldsymbol{\pi}_t$ would be a Dirichlet distribution with parameter $a_0\mathbf{p}$, where $\mathbf{p} = (p_1, p_2, p_3, p_4)$ such that $a_0 > 0$, $p_k > 0$ and $\sum_{k=1}^{4} p_k = 1$. To relate a set of Dirichlet random variables, we use ideas from [25], who defined dependence in univariate random variables whose distribution belongs to the exponential family, and define a dynamic prior with temporal dependence as follows.

Let $\boldsymbol{\eta}_t = (\eta_{t1}, \ldots, \eta_{t4}) \in \mathbb{R}^4$ be a latent vector corresponding to each $\boldsymbol{\pi}_t$ and let $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_4)$ be a unique latent vector such that

$$\boldsymbol{\omega} \sim \text{Dir}(a_0\mathbf{p}) \quad \text{and} \quad \boldsymbol{\eta}_t \mid \boldsymbol{\omega} \overset{\text{ind}}{\sim} \text{Mult}(a_t, \boldsymbol{\omega}), \tag{7}$$

with $a_t \in \mathbb{N}$, $\eta_{tk} \in \mathbb{N}$ and $\sum_{k=1}^{4} \eta_{tk} = a_t$. Then, the prior dependence in π_t is modeled through a subset ∂_t of previous latent variables $\{\eta_1, \eta_2, \ldots, \eta_{t-1}\}$

$$\boldsymbol{\pi}_t \mid \boldsymbol{\eta} \overset{\text{ind}}{\sim} \text{Dir}\left(a_0\mathbf{p} + \sum_{j \in \partial_t} \boldsymbol{\eta}_j\right). \tag{8}$$

We denote this construction as $\text{DDir}(a_0, \mathbf{a}, \boldsymbol{\partial})$ with $a_0 > 0$, $\mathbf{a} = (a_1, \ldots, a_T)$ 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

$$\partial_t = \{t, t-1, \ldots, t-q\}; \tag{9}$$

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

$$\partial_t = \{t, t-s, t-2s, \ldots, t-ps\}; \tag{10}$$

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 $\boldsymbol{\pi} = \{\boldsymbol{\pi}_t\} \sim DDir(a_0, \mathbf{a}, \boldsymbol{\partial})$ a sequence of vectors whose probability law is defined by (7) and (8) for $a_0 > 0$, $a_t \in \mathbb{N}$ and subsets $\boldsymbol{\partial}$. Then,*

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

(ii) The correlation between $\pi_{t,k}$ and $\pi_{r,k}$, for $t \neq r$ and $k = 1, \ldots, 4$, does not depend on k and is given by

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

(iii) If $a_t = 0$ for all $t = 1, 2 \ldots$ 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 η_t, $t = 1, 2, \ldots$ are conditionally independent given $\boldsymbol{\omega}$ from $\text{Mult}(a_t, \boldsymbol{\omega})$, and the prior distribution for $\boldsymbol{\omega}$ is $\text{Dir}(a_0\mathbf{p})$, then the posterior distribution for $\boldsymbol{\omega}$ given the η_t's is $\text{Dir}\left(a_0\mathbf{p} + \sum_t \boldsymbol{\eta}_t\right)$. 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 $\boldsymbol{\omega}$ which is $\text{Dir}(a_0\mathbf{p})$.

For (ii) we first note that for a specific k, the distributions for ω_k, η_{tk} and π_{tk} reduce to beta, binomial and beta, respectively. To obtain the correlation we rely on iterative formulae. The covariance is $\text{Cov}(\pi_{t,k}, \pi_{r,k}) = \text{E}\{\text{Cov}(\pi_{t,k}, \pi_{r,k} \mid \boldsymbol{\eta})\} + \text{Cov}\{\text{E}(\pi_{t,k} \mid \boldsymbol{\eta}), \text{E}(\pi_{r,k} \mid \boldsymbol{\eta})\}$, where

the first term is zero due to conditional independence. Then

$$\mathrm{Cov}(\pi_{t,k}, \pi_{r,k}) = \mathrm{Cov}\left\{ \frac{a_0 p_k + \sum_{j\in\partial_t} \eta_{j,k}}{a_0 + \sum_{j\in\partial_r} a_j}, \frac{a_0 p_k + \sum_{j\in\partial_r} \eta_{j,k}}{a_0 + \sum_{j\in\partial_r} a_j} \right\},$$

which after cancelling the additive constants and using the linearity of the covariance becomes

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

After using the iterative formula for a second time we get

$$\mathrm{E}\left[\mathrm{Cov}\left\{ \sum_{j\in\partial_t} \eta_{j,k}, \sum_{j\in\partial_r} \eta_{j,k} \middle| \boldsymbol{\omega} \right\} \right] + \mathrm{Cov}\left\{ \mathrm{E}\left(\sum_{j\in\partial_t} \eta_{j,k} \middle| \boldsymbol{\omega} \right), \mathrm{E}\left(\sum_{j\in\partial_r}^{q} \eta_{j,k} \middle| \boldsymbol{\omega} \right) \right\}. \tag{11}$$

Within each sum we can isolate the common part as $\sum_{j\in\partial_t} \eta_{j,k} = \sum_{j\in\partial_t\cap\partial_r} \eta_{j,k} + \sum_{j\in\partial_t-\partial_r} \eta_{j,k}$ and $\sum_{j\partial_r} \eta_{j,k} = \sum_{j\in\partial_t\cap\partial_r} \eta_{j,k} + \sum_{j\in\partial_r-\partial_t} \eta_{j,k}$, and using covariance properties and conditional independence, (11) becomes

$$\mathrm{E}\left\{ \mathrm{Var}\left(\sum_{j\in\partial_t\cap\partial_r} \eta_{j,k} \middle| \omega_k \right) \right\} + \mathrm{Cov}\left\{ \sum_{j\in\partial_t} a_j\omega_k, \sum_{j\in\partial_r} a_j\omega_k \right\}.$$

The first expected value, after obtaining the conditional variance is $\mathrm{E}\{\sum_{j\in\partial_t\cap\partial_r} a_j\omega_k(1-\omega_k)\} = (\sum_{j\in\partial_t\cap\partial_r} a_j)\mathrm{E}(\omega_k - \omega_k^2)$ with $\mathrm{E}(\omega_k - \omega_k^2) = \mathrm{E}(\omega_k) - \mathrm{E}^2(\omega_k) - \mathrm{Var}(\omega_k) = a_0\mathrm{Var}(\omega_k)$. The second term is $(\sum_{j\in\partial_t} a_j)(\sum_{j\in\partial_r} a_j)\mathrm{Var}(\omega_k)$. In conclusion, we obtain

$$\mathrm{Cov}(\omega_{t,k}, \omega_{r,k}) = \frac{a_0\left(\sum_{j\in\partial_t\cap\partial_r} a_j\right) + \left(\sum_{j\in\partial_t} a_j\right)\left(\sum_{j\in\partial_r} a_j\right)}{\left(a_0 + \sum_{j\in\partial_t} a_j\right)\left(a_0 + \sum_{j\in\partial_r} a_j\right)} \mathrm{Var}(\omega_k).$$

Since ω_k, $\pi_{t,k}$ and $\pi_{r,k}$ 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 $\mathrm{Dir}(a_0\mathbf{p})$. $\diamond$

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 θ_{tk}, so that information is shared across times t for each k. That is,

$$\theta_{tk} \mid \beta_k \overset{\text{ind}}{\sim} \text{Ga}(d_k, \beta_k), \quad \text{and} \quad \beta_k \sim \text{Ga}(e_k, g_k) \tag{12}$$

for $t \geq 1$ and $k = 1, \ldots, 4$.

5.2 Posterior distributions

Let $\mathbf{U}_{ti} = (U_{1ti}, U_{2ti})$ for $i = 1, \ldots, n_t$ a sample of size n_t from model (3) for each $t = 1, \ldots, T$. Let $\mathbf{Z}_{ti}$ be a latent vector that identifies the mixture component which is observation i's pdf, that is, $\mathbf{Z}'_{ti} = (Z_{t1i}, Z_{t2i}, Z_{t3i}, Z_{t4i}) \sim \text{Mult}(1, \boldsymbol{\pi}_t)$. Assuming for the moment that together with $\mathbf{U}_{ti}$ we observe $\mathbf{Z}_{ti}$, then the extended likelihood has the form

$$f(\mathbf{u}, \mathbf{z} \mid \boldsymbol{\pi}, \boldsymbol{\theta}) = \prod_{t=1}^{T} \prod_{i=1}^{n_t} \prod_{k=1}^{4} \left\{ \pi_{tk} f_k(u_{1ti}, u_{2ti} \mid \theta_{tk}) \right\}^{z_{tki}},$$

where

$$\begin{aligned}
f_1(u_1, u_2 \mid \theta_1) &= (\theta_1 + 1)(u_1 u_2)^{-(\theta_1+1)} \left(u_1^{-\theta_1} + u_2^{-\theta_1} - 1 \right)^{-(1/\theta_1+2)}, \\
f_2(u_1, u_2 \mid \theta_2) &= (\theta_2 + 1)\{(1-u_1)u_2\}^{-(\theta_2+1)} \left\{ (1-u_1)^{-\theta_2} + u_2^{-\theta_2} - 1 \right\}^{-(1/\theta_2+2)}, \\
f_3(u_1, u_2 \mid \theta_3) &= (\theta_3 + 1)\{(1-u_1)(1-u_2)\}^{-(\theta_3+1)} \left\{ (1-u_1)^{-\theta_3} + (1-u_2)^{-\theta_3} - 1 \right\}^{-(1/\theta_3+2)}, \\
f_4(u_1, u_2 \mid \theta_4) &= (\theta_4 + 1)\{u_1(1-u_2)\}^{-(\theta_4+1)} \left\{ u_1^{-\theta_4} + (1-u_2)^{-\theta_4} - 1 \right\}^{-(1/\theta_4+2)}.
\end{aligned} \tag{13}$$

The prior distribution for $(\boldsymbol{\pi}, \boldsymbol{\theta})$ is defined by equations (7), (8) and (12). Again, extending the prior to include the latent variables $\boldsymbol{\eta}$ and $\boldsymbol{\omega}$ we get

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

and

$$f(\boldsymbol{\theta}) = \prod_{k=1}^{4} \text{Ga}(\beta_k \mid e_k, g_k) \prod_{t=1}^{T} \text{Ga}(\theta_{tk} \mid d_k, \beta_k)$$

independent of each other.

Posterior distributions are characterised 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 $\boldsymbol{\pi}^* = \{\pi_{tk}^*\}$ and

$$\pi_{tk}^* = \frac{\pi_{tk} f_k(u_{1ti}, u_{2ti} \mid \theta_{tk})}{\sum_{j=1}^4 \pi_{tj} f_j(u_{1ti}, u_{2ti} \mid \theta_{tj})},$$

with f_k is given in (13) for $k = 1, \ldots, 4$.

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

$$\boldsymbol{\pi}_t \mid \text{rest} \sim \text{Dir}\left(a_o \mathbf{p} + \sum_{j \in \partial_t} \boldsymbol{\eta}_j + \sum_{i=1}^{n_t} \mathbf{z}_{ti}\right).$$

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

$$f(\boldsymbol{\eta}_t \mid \text{rest}) \propto \left\{ \prod_{k=1}^4 \frac{\left(\omega_k \prod_{l \in \varrho_t} \pi_{l,k}\right)^{\eta_{tk}}}{\Gamma(\eta_{tk}+1) \prod_{l \in \varrho_t} \Gamma\left(a_0 p_k + \sum_{j \in \partial_l} \eta_{j,k}\right)} \right\} I\left(\sum_{k=1}^4 \eta_{tk} = a_t\right),$$

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

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

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

(e) The posterior conditional for θ_{tk} is

$$f(\theta_{tk} \mid \text{rest}) \propto \theta_{tk}^{d_k - 1} e^{-\beta_k \theta_{tk}} \prod_{i=1}^{n_t} \left\{ f_k(u_{1ti}, u_{2ti} \mid \theta_{tk}) \right\}^{z_{tki}},$$

where f_k is given in (13).

(f) The posterior conditional for β_k is

$$\beta_k \mid \text{rest} \sim \text{Ga}\left(e_k + T d_k \,,\, g_k + \sum_{t=1}^T \theta_{tk}\right).$$

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 4 with a sum restriction, it is easier if we sample from each of the components η_{tk} for $k = 1, 2, 3$ using $f(\eta_{tk} \mid \text{rest}) \propto$

$$\frac{\left\{\omega_k \prod_{l\in\varrho_t} \pi_{l,k} / \left(\omega_4 \prod_{l\in\varrho_t} \pi_{l,4}\right)\right\}^{\eta_{tk}} I\left(\eta_{tk} \in \{0, 1, \ldots, a_t - \sum_{j\neq k}^{3} \eta_{tj}\}\right)}{\Gamma(\eta_{tk}+1) \prod_{l\in\varrho_t} \Gamma\left(a_0 p_k + \sum_{j\in\partial_l} \eta_{j,k}\right) \Gamma(\eta_{t4}+1) \prod_{l\in\varrho_t} \Gamma\left(a_0 p_4 + \sum_{j\in\partial_l} \eta_{j,4}\right)},$$

with $\eta_{t4} = a_t - \sum_{j=1}^{3} \eta_{tj}$. 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 $\theta^*_{tk} \sim \text{Ga}(\kappa, \kappa/\theta^{(r)}_{tk})$ and accept it with probability

$$\alpha(\theta^*_{tk}, \theta^{(r)}_{tk}) = \frac{f(\theta^*_{tk} \mid \text{rest}) \text{Ga}(\theta^{(r)}_{tk} \mid \kappa, \kappa/\theta^*_{tk})}{f(\theta^{(r)}_{tk} \mid \text{rest}) \text{Ga}(\theta^*_{tk} \mid \kappa, \kappa/\theta^{(r)}_{tk})},$$

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 $h \to \infty$ 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 can be found at the GitHub repository `https://github.com/RuyiPan/TD-MRC`.

6 Numerical analyses

6.1 Simulation study

We conduct a comprehensive simulation study to evaluate the performance of the proposed model. The true generative model is set using $\theta_t = (\theta_{t1}, \theta_{t2}, \theta_{t3}, \theta_{t4}) = (5, 3, 4, 3)$

for $t = 1, \ldots, T$, with $T = 20$. The weights of rotated Clayton copulas are chosen to be linearly dependent in time. More specifically, we first set $\boldsymbol{\pi}_1 = (\pi_{11}, \pi_{12}, \pi_{13}, \pi_{14}) = (0.4, 0.25, 0.1, 0.25)$ as the initial values at $t = 1$ and subsequently the weights are constructed using $\pi_{t1} = 0.95\pi_{t-1,1}, \pi_{t1} = 1.05\pi_{t-1,2}, \pi_{t3} = 0.1, \pi_{t4} = 1 - \sum_{i=1}^{3} \pi_{ti}$ for $t = 2, \ldots, 20$. We sampled $n_t = 300$ realizations from this model for each time $t = 1, \ldots, T$ 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 for capturing the temporal dependence, we considered a moving average type of order q, that is, the subsets of lags are defined by (9). We considered different hyper-parameters for the dynamic process: $a_t = 0, 1, 3, 5, 10, 20, 30, 40$ and $q = 0, 1, \ldots, 7$. 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 after 60 batch iterations, and the acceptance rate stabilizes between 0.3 and 0.4 after 60 batches. Therefore we set the burn-in to be equal to $3,000$ iterations. This also confirms that the adaptation method proposed at the end of Section 5 performs well. Computing time is around 50 minutes for each run in 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 π_{tk}, 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 Tables 1. The case of independence across

times for π_{tk} 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 θ_{tk}. This is achieved by considering a very low variance in β_k, as obtained by setting $e_k = d_k = 1000$. This latter model produces inferior gof measures compared to $\mathcal{M}_{a_t,q}$.

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 shows posterior estimates of π_{tk} as time series for $t = 1, \ldots, 20$ in four panels for $k = 1, \ldots, 4$. Posterior estimates follow very closely the path of the true values. Similarly, Figure 4 shows posterior estimates of θ_{tk} as time series in four panels. All the true values lie within the 95% credible intervals. We note that credible intervals are narrower for θ_{t1} (top-left panel) and θ_{t4} (bottom-right panel) as compared to those for θ_{t2} (top-right panel) and θ_{t3} (bottom-left panel). Wider credible intervals are due to smaller weights (less data points) associated to the corresponding mixture components. Specifically, the higher variability for θ_{t3} for larger t is a consequence of the smaller weights for the third component π_{t3}.

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 $\lambda_t = (1/2)\log\left((1+\rho_t)/(1-\rho_t)\right)$ and models the dynamics via an autoregressive process of the form $\lambda_t = \alpha + \beta\lambda_{t-1} + \epsilon_t$ with $\epsilon_t \sim \mathrm{N}(0, \tau)$. We perform a Bayesian analysis for this model with vague prior distributions $\alpha \sim \mathrm{N}(0, 0.01)$, $\beta \sim \mathrm{N}(0, 0.01)$ and $\tau \sim \mathrm{Ga}(0.01, 0.01)$. 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, $LPS(t) = \sum_{i=1}^{n_t} \log f(\mathbf{u}_{ti} \mid \mathbf{u}_{t-1})$, where the predictive distribution is approximated via Monte Carlo as

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

and $(\boldsymbol{\pi}_t^{(r)}, \boldsymbol{\theta}_t^{(r)})$ are obtained from the posterior distribution $f(\boldsymbol{\pi}_t, \boldsymbol{\theta}_t \mid \mathbf{u}_1, \ldots, \mathbf{u}_{t-1})$ 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 $q = 7$.

6.2 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, $t = 1, \ldots, T$ with $T = 36$ for a total time span of three years. The number of stations varies across months, given that some of them have missing values for the whole month. On average, sample sizes range around $n_t \approx 200$ for each t.

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 $\widehat{F}_{X_1}$ and $\widehat{F}_{X_2}$ and apply the inverse transformation $U_{1ti} = \widehat{F}^{-1}_{X_1}(X_{1ti})$ and $U_{2ti} = \widehat{F}^{-1}_{X_2}(X_{2ti})$.

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 (9); seasonal model of order p with annual seasonality $s = 12$ as in (10); 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 $b_k = c_k = 1$ to define the prior distributions. In this real data analysis, we also varied the dependence parameters $a_t = 0, 1, \ldots, 5$, $q = 0, 1, \ldots, 4$ 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. We also used LPML and WAIC gof measures to assess model fit.

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

required to model this particular dataset.

In Figures 6 and 7 we show posterior estimates of the weights π_{tk} and copula coefficients θ_{tk}, 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 (π_{t1}, π_{t3}) and (π_{t2}, π_{t4}). 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, π_{t2} has its peak around September-October (autumn), whereas π_{t4} 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 π_{t3} 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 θ_{tk}. Their posterior estimates are all around slightly below the value of one, with θ_{t2} and θ_{t4} showing more variability along time. Uncertainty in the estimation of θ_{t3} is somehow higher, due to the slightly smaller weights π_{t3} and thus smaller sample size for estimating θ_{t3}. According to θ_{t1}, 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 θ_{t2} and θ_{t4}, 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 $\boldsymbol{\lambda}_t$, given in (6). These are reported in Figure 8. We first note that none of them is larger than 1/2, the only exception being the upper-lower λ_{t2} 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 λ_{t1} and λ_{t3} 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 λ_{2t}

and λ_{4t} 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 λ_{2t} and in the winter for λ_{4t}.

Finally, we show joint density estimates in Figure 9 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: $\boldsymbol{\pi}_{2017-8} = (0.41, 0.52, 0.05, 0.02)$, $\boldsymbol{\theta}_{2017-8} = (0.69, 0.69, 0.59, 0.81)$; $\boldsymbol{\pi}_{2018-8} = (0.07, 0.86, 0.03, 0.04)$, $\boldsymbol{\theta}_{2018-8} = (0.48, 0.31, 0.50, 0.84)$; $\boldsymbol{\pi}_{2019-8} = (0.11, 0.59, 0.21, 0.09)$, $\boldsymbol{\theta}_{2019-8} = (0.57, 0.31, 0.58, 0.59)$. What makes the heat plots to show the 4-way dependence is a combination of the weight π_{tk} and the intensity θ_{tk}.

We compare the fit of $\mathcal{M}_{1,1}$ 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 $t = 1, \ldots, T$ 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 $C_t(u_{1t} \mid u_{2t}, \boldsymbol{\pi}_t, \boldsymbol{\theta}_t) = \sum_{k=1}^{4} \pi_{tk} C_k(u_{2t} \mid u_{1t}, \theta_{tk})$ with $C_k(u_{2t} \mid u_{1t}, \theta_{tk}) = (\partial/\partial u_{1t}) C_k(u_{1t}, u_{2t} \mid \theta_{tk})$ for $k = 1, \ldots, 4$, and obtain the posterior predictive mean $\widehat{u}_{2t} = \mathrm{E}(U_{2t} \mid u_{1t}, \boldsymbol{\pi}_t, \boldsymbol{\theta}_t)$ as point prediction.

We assess model performance by computing the mean square error between the observed u_{2t} and predicted $\widehat{u}_{2t}$ 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 $MSE_t = (1/n_{2t}) \sum_{i=1}^{n_{2t}} (u_{2t} - \widehat{u}_{2t})^2$, 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 $t = s+1, \ldots, T$ and aggregate the scores as

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

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 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 bivariate distribution 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 the study of it for future work.

Additionally, future work will explore extensions of dynamic copulas to higher dimensions. For instance, one can easily see that if we were to capture dependence in all the extreme regions of a 10-dimensional copula, we would need 100 mixture components, each component being a 10-dimensional copula. Although feasible, such an approach is likely impractical as not all extremes are likely to be significant. In order to impose sparsity, we plan to exploit a Dirichlet process mixture prior to reduce the number of components needed to model the data and yet 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 dṕendance empirique et ses proprietś. Un test non paramt́rique d'indṕendance. *Bulletin Royale Belge de l'Acadḿie 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 evaluati9ng 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 Medecine* **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-Argëlles, 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'Universite 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.

b_k, c_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	3	1685	1690	1692	1694	1697	1698	1699	1700
1	5	1688	1694	1698	1699	1700	1700	1702	1701
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	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	3	-3372	-3382	-3386	-3390	-3395	-3397	-3399	-3401
1	5	-3378	-3389	-3396	-3399	-3402	-3401	-3405	-3403
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 $1, \ldots, 19$ for fitting and $t = 20$ for prediction. The other gof measures, LPML and WAIC, were calculated with the fitting data.

	LPML					WAIC				
a_t	$MA(0)$	$MA(1)$	$MA(2)$	$MA(3)$	$MA(4)$	$MA(0)$	$MA(1)$	$MA(2)$	$MA(3)$	$MA(4)$
	$S(0)$									
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
	$S(1)$									
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
	$S(2)$									
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: Real data. LPML and WAIC statistics for different prior choices for $\mathcal{M}_{a_t,q,p}$.

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: 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: Real data. LPS statistic computed by fitting models from time 1 to $t - 1$ and predicting time t.

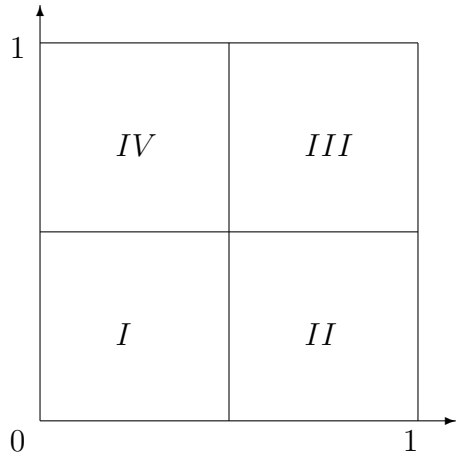

Figure 1: Unit square divided in four quadrants.

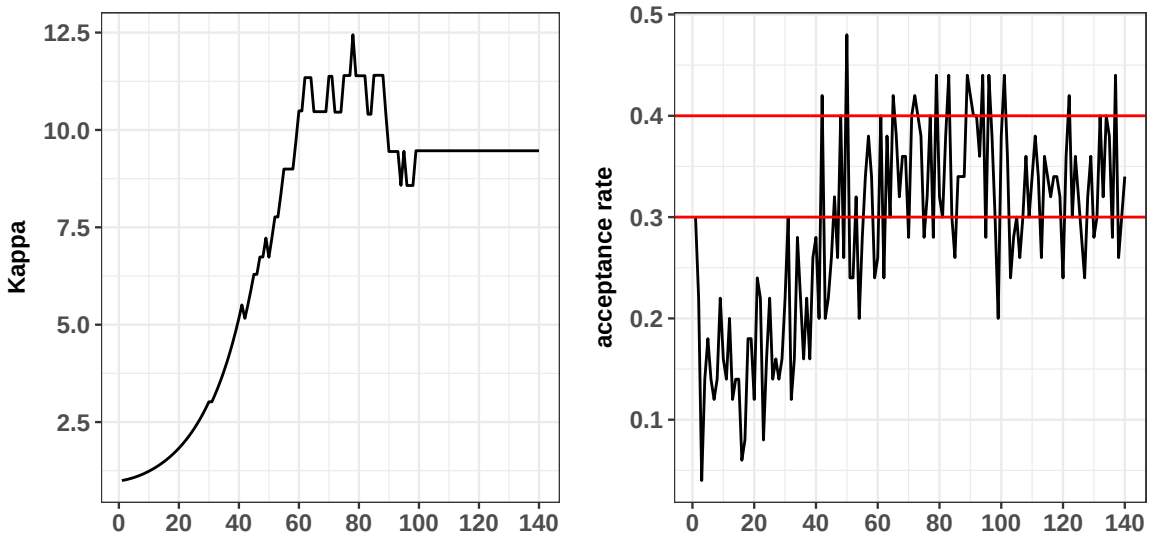

Figure 2: The recorded $\kappa^{(h)}$ and acceptance rate for each batch h. The batch size is 50.

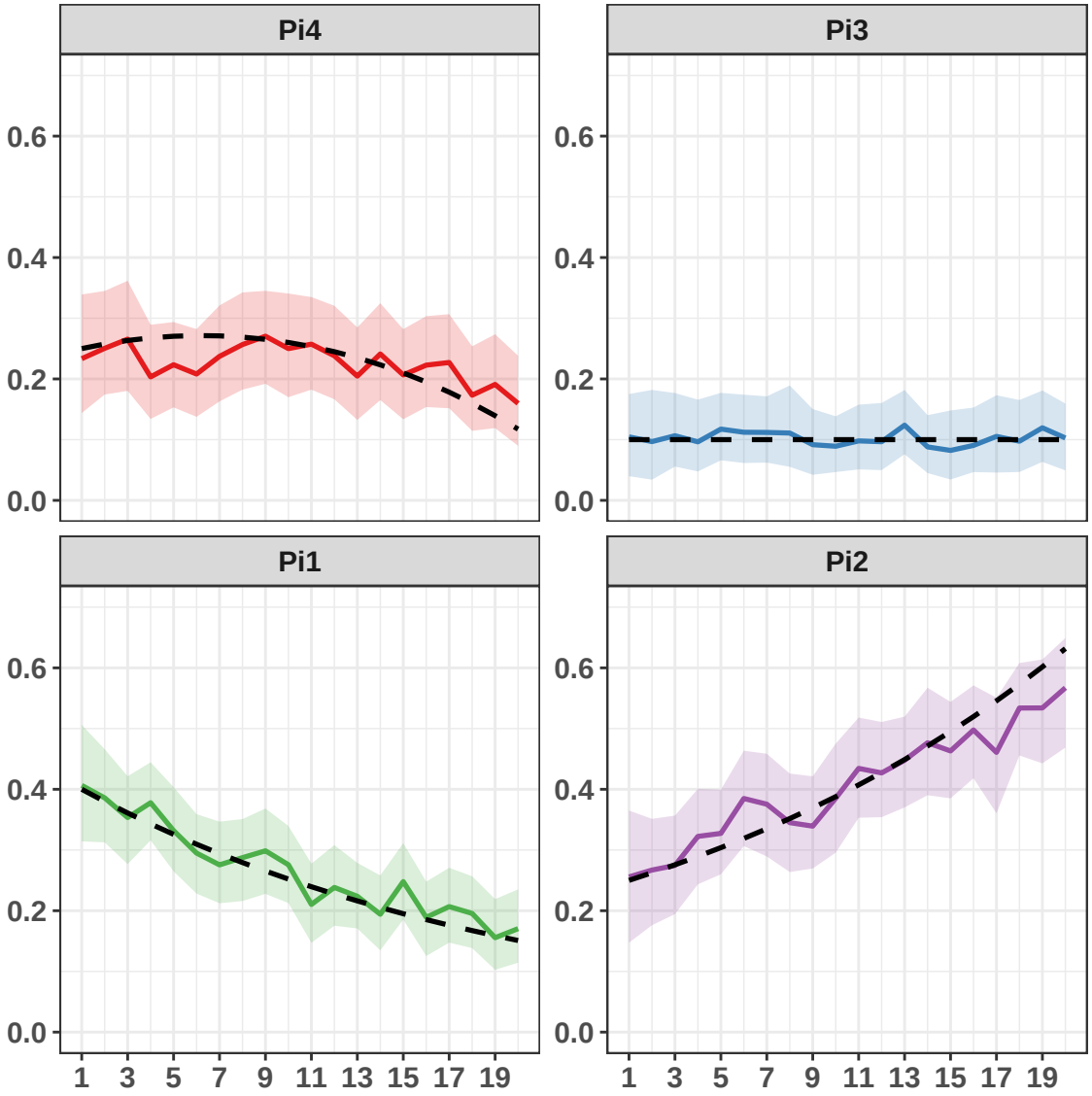

Figure 3: Simulated data. Posterior estimates of $\boldsymbol{\pi}$: posterior mean (solid line) with 95% credible intervals (shadows), together with the true value (dotted black line).

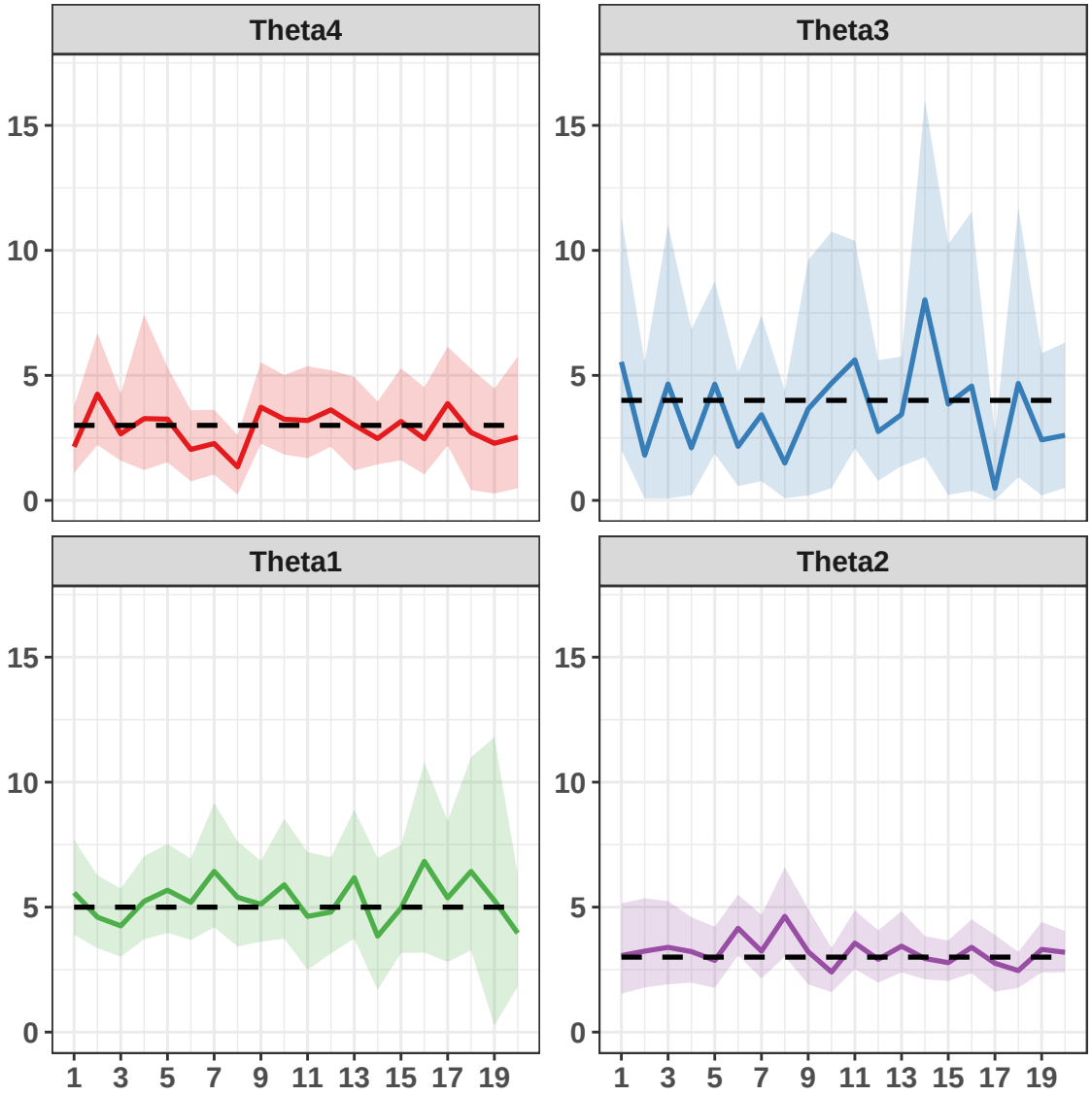

Figure 4: Simulated data. Posterior estimates of $\boldsymbol{\theta}$: posterior mean (solid line) with 95% credible intervals (shadows), together with the true value (dotted black line).

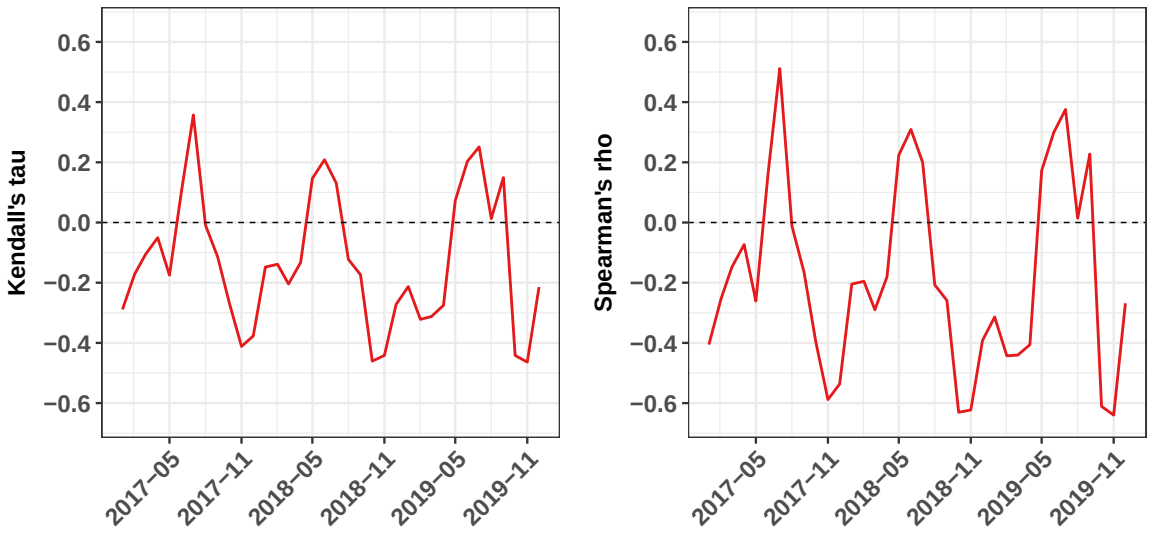

Figure 5: Real dataset. Empirical Kendall's tau with estimated vlaues (left) and Spearman's rho (right).

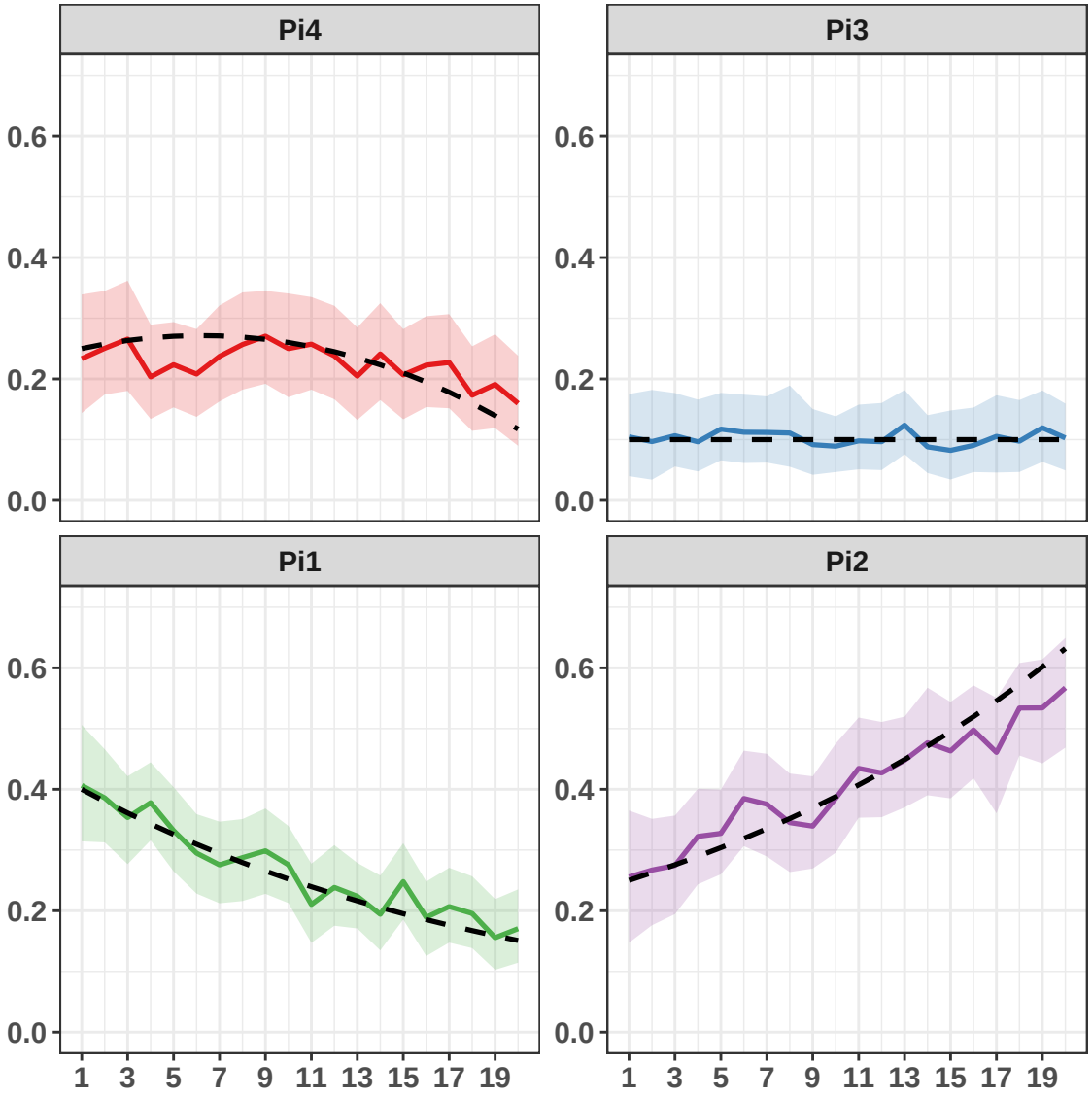

Figure 6: Real dataset. Posterior estimates of $\boldsymbol{\pi}$: posterior mean (solid line) with 95% credible intervals (shadows).

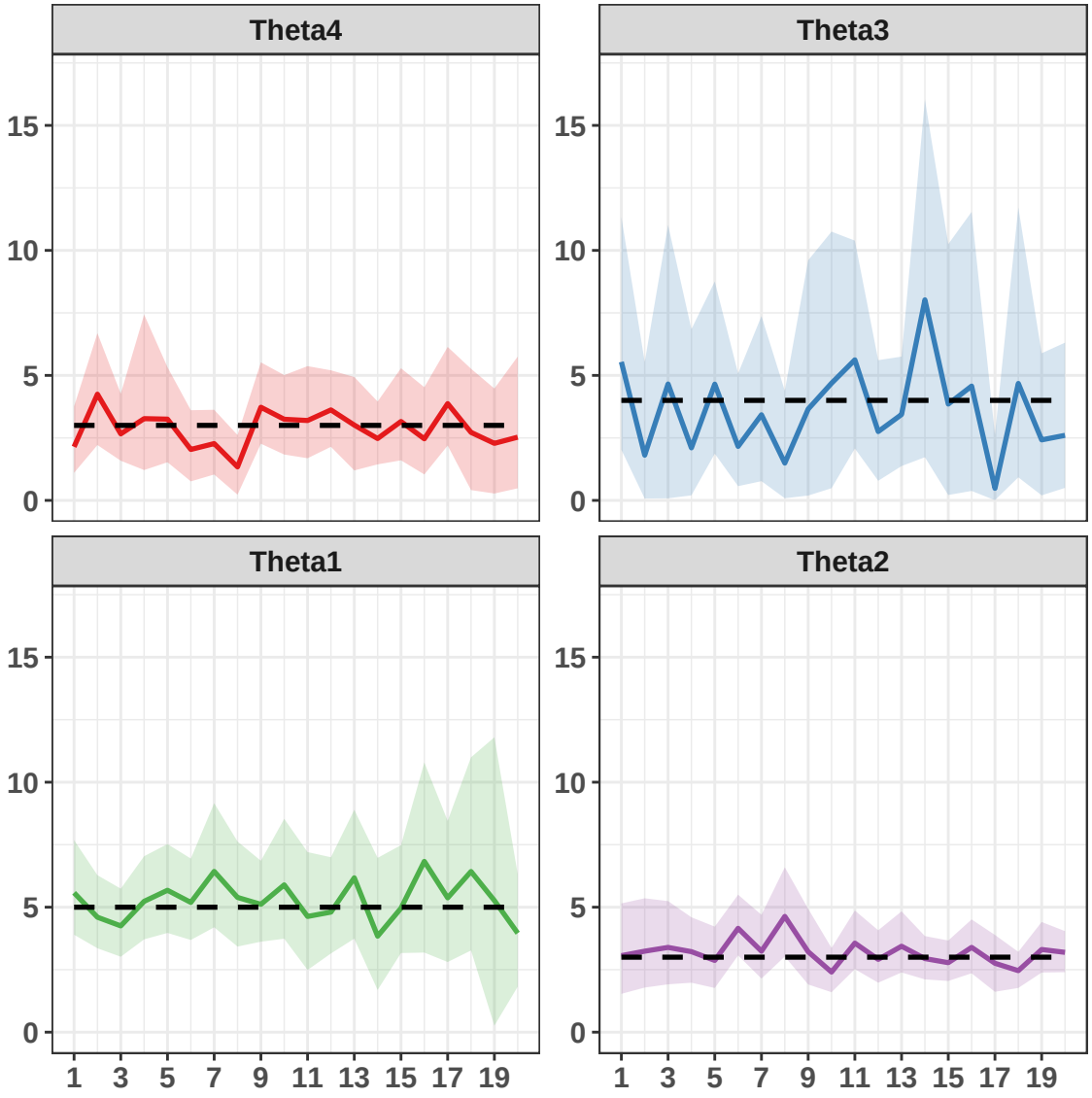

Figure 7: Real dataset. Posterior estimates of $\boldsymbol{\theta}$: posterior mean (solid line) with 95% credible intervals (shadows).

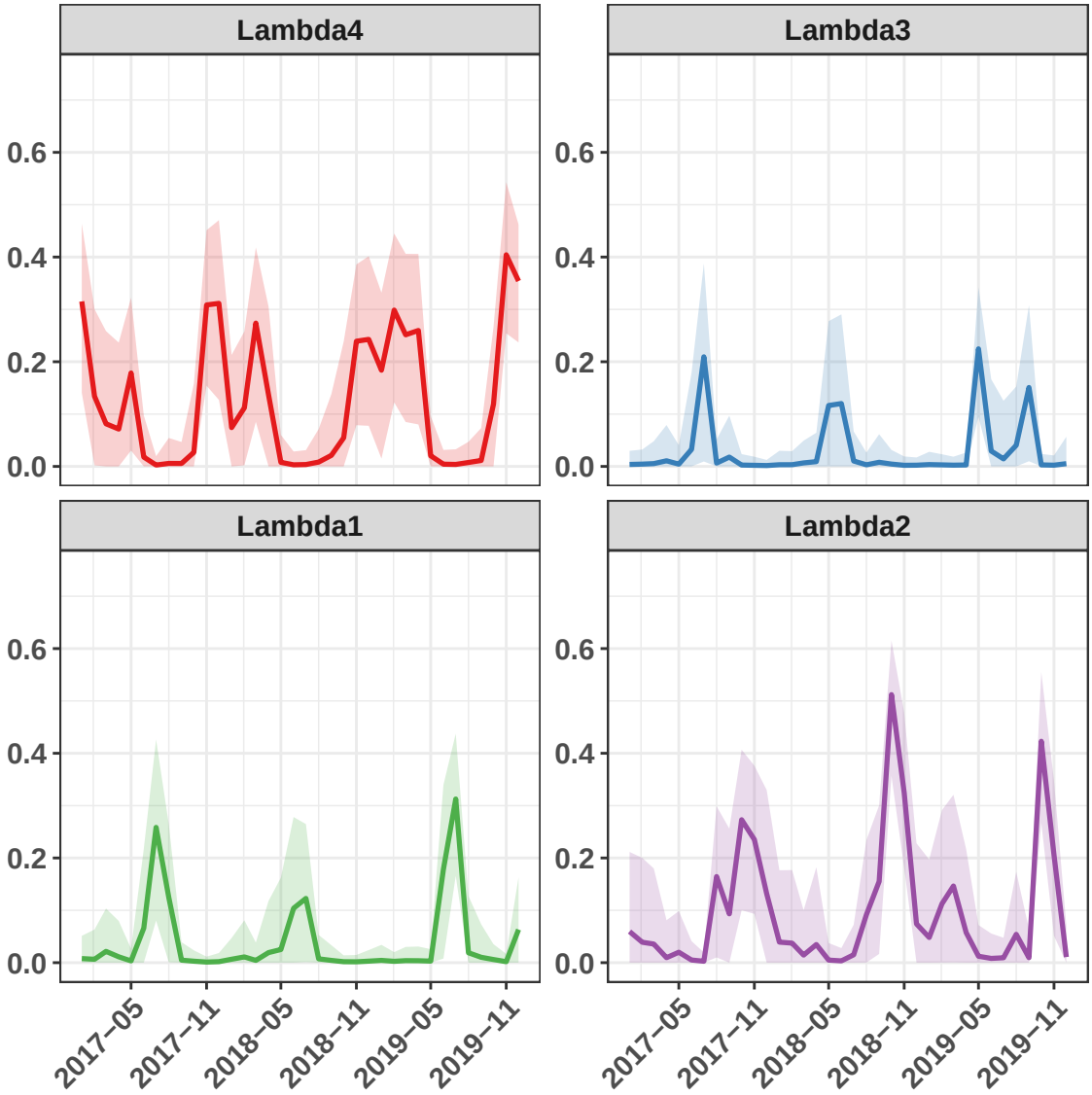

Figure 8: Real dataset. Posterior estimates of $\boldsymbol{\lambda}$: posterior mean (solid line) with 95% credible intervals (shadows).

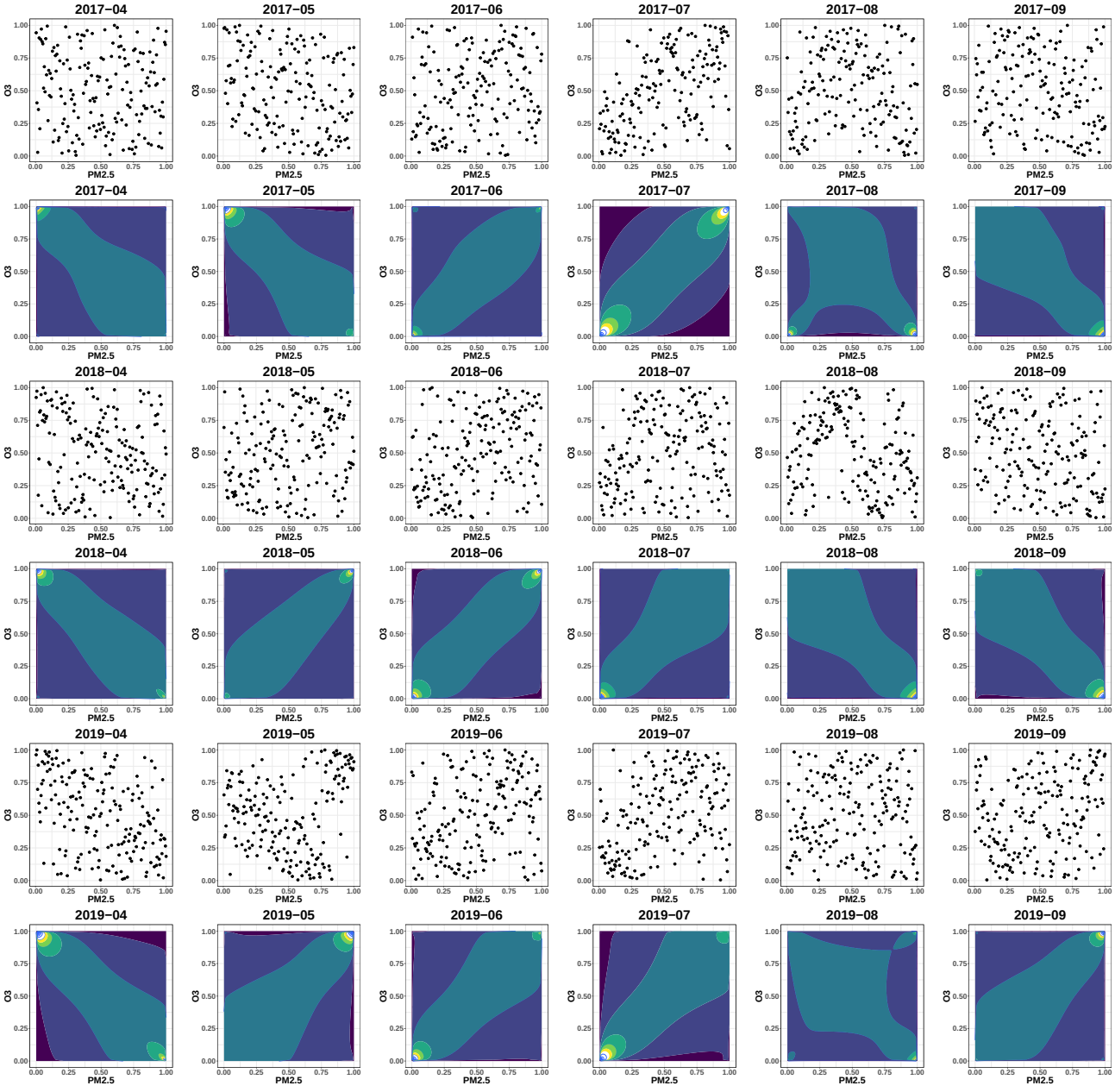

Figure 9: Observed real data and estimated joint density from April to September 2017 (top), 2018 (middle), and 2019 (bottom). Dependence patterns tend to vary between seasons with some months exhibiting transitional regimes.