

# Variational Inference for high dimensional structured factor copulas

Hoang Nguyen<sup>(a)</sup>, M. Concepción Ausín<sup>(a,b)</sup> and Pedro Galeano<sup>(a,b)</sup>

<sup>(a)</sup> Department of Statistics - Universidad Carlos III de Madrid, Spain

<sup>(b)</sup> UC3M-BS Institute of Financial Big Data (IFiBiD)

October 17, 2018

## Abstract

Factor copula models have been recently proposed for describing the joint distribution of a large number of variables in terms of a few common latent factors. In this paper, we employ a Bayesian procedure to make fast inferences for multi-factor and structured factor copulas. To deal with the high dimensional structure, we apply a variational inference (VI) algorithm to estimate different specifications of factor copula models. Compared to the Markov chain Monte Carlo (MCMC) approach, the variational approximation is much faster and could handle a sizeable problem in a few seconds. Another issue of factor copula models is that the bivariate copula functions connecting the variables are unknown in high dimensions. We derive an automatic procedure to recover the hidden dependence structure. By taking advantage of the posterior modes of the latent variables, we select the bivariate copula functions based on minimizing the Bayesian information criterion (BIC). The simulation studies in different contexts show that the procedure of bivariate copula selection could be very accurate in comparison to the true generated copula model. We illustrate our proposed procedure with two high dimensional real data sets.

**Keywords:** Factor copula; Model selection; Variational inference.

# 1 Introduction

The general class of factor copula models proposed by Krupskii and Joe (2013) explains the dependence structure of high dimensional variables in terms of a few latent variables. Due to the flexibility of copula functions, factor copulas can capture well the correlation along with the tail co-movement in extreme events. Since the Gaussian factor copula model by Hull and White (2004), factor copulas have been extended to fit with different characteristics of data, for example, spatial dependence of temperatures in Krupskii et al. (2016), spatio-temporal dependence of hourly wind data in Krupskii and Genton (2017), mortality dependence of multiple populations in Chen et al. (2015), behavior dependence of item response in Nikoloulopoulos and Joe (2015), extreme dependence of river flows in Lee and Joe (2017), financial time series dependence in Krupskii and Joe (2013), Creal and Tsay (2015), Oh and Patton (2017a), and Nguyen et al. (2017), among others. In comparison to the truncated vine copulas proposed by Brechmann et al. (2012), the factor copula model is not only suitable with the empirical data but also provides parsimonious and interpretable economic meanings.

In this paper, we propose a Bayesian procedure to make inferences for multi-factor and structured factor copulas. To deal with the high dimensional structure, we employ a variational inference (VI) algorithm to estimate different specifications of factor copula models. VI aims to approximate the joint posterior distribution of model parameters by a simpler distribution, usually an exponential family. Blei et al. (2017) review the coordinate ascent variational method for conditionally conjugate models. However, due to the model complexity, it is difficult to come up with a conjugate specification. Instead, we apply the black-box VI based on the reparameterization of parameters proposed by Kucukelbir et al. (2017). Compared to the MCMC approach, the variational approximation is much faster and could handle a sizeable problem in a few seconds. Another issue of factor copula models is that the bivariate copula functions connecting the variables are unknown in high dimensions. We derive an automatic procedure to recover the hidden dependence structure. By taking advantage of the posterior modes of the latent variables, we select the bivariate copula functions based on minimizing the Bayesian information criterion (BIC). All the parameters are estimated jointly hence we take into account the uncertainty of latent variables as well as copula parameters in each tree layer.

There are mainly two approaches to set up factor copula models. Krupskii and Joe (2013, 2015a) proposed pair copula construction-based factor models while Creal and Tsay (2015) and Oh and Patton (2017b) extended the classical factor analysis by inverting the dependence structure from latent elliptical or skew-elliptical distributions to the constrained copula domain. The advantage of the latter approach is that we could incorporate the class of dynamic factor models proposed in the literature of time series analysis with arbitrary marginal distributions. However, the choice of copula functions is limited to some extensions of elliptical distributions such as the Student- $t$  and the skew Student- $t$  distributions. In this paper, we follow Krupskii and Joe (2015a) to construct the dependence structure via bivariate linking copulas between copula variables and the latent factors. The model helps to capture both the tail asymmetry and the tail dependence using different bivariate copula functions.

Estimation of factor copula models in high dimensions is a challenging problem. Krupskii and Joe (2013, 2015a) use maximum likelihood estimation to estimate the copula parameters. They integrate the augmented likelihood over the latent factor variable space and approximate the integral via the Gauss-Legendre quadrature. The estimated variance is obtained as the inverse of the observed Fisher information matrix at the global maximum. Alternatively, from the Bayesian point of view, Murray et al. (2013) use the parameter-expanded Gibbs sampling to estimate the multi-factor Gaussian model. Schamberger et al. (2017) use an adaptive rejection Metropolis sampling to estimate the one-factor copula model, while Tan et al. (2018) use reversible jump MCMC to select the factor copula links during the sampling process. Among those, only the Tan et al. (2018) approach allows for model uncertainty by selecting or averaging over different factor copula specifications. Except for the Gaussian copulas in Murray et al. (2013), the mentioned Bayesian inference strategies are so computationally expensive that it is difficult to extend the algorithm to the general multi-factor copula models.

VI has been applied to different problems in probabilistic latent variable models, see Zhang et al. (2018). VI aims to approximate the model posterior distribution by a simpler distribution such that their distance in term of the Kullback-Leibler divergence is minimized. Therefore, instead of sampling from the posterior distribution, VI turns the inference problem into an optimization problem. We gain the speed of computation in exchange for the simplified posterior distribution. Several VI algorithms have been proposed for a general probabilistic model, see Ranganath et al.

(2014), Ruiz et al. (2016) Kucukelbir et al. (2017), among others. In this paper, we employ VI based on the reparameterization of parameters in Kucukelbir et al. (2017). The central idea is to use several transformations of the factor copula parameters from the constrained space into the real coordinate space and approximate the transformed posterior distribution by a product of univariate Gaussian distributions. Then, VI uses the noisy gradient computed from Monte Carlo simulations to optimize the variational objective. We compare a posteriori estimates from the VI and the MCMC inference. The posterior modes of VI estimation are similar to that of MCMC samples while the standard deviations are only underestimated in the bi-factor copula model. The underestimated results are acceptable which aligns with other findings in the literature of VI, see Blei et al. (2017). However, we greatly benefit from the speed of calculation that VI could be at least 100 times faster. VI has also been used to estimate vine copulas with discrete margins, see Loaiza-Maya and Smith (2018).

We also propose a model selection approach to recover the hidden structure in factor copula models. Starting with arbitrary bivariate links, we obtain the posterior modes of the latent factors. Then, we seek for the best bivariate copula functions between the observable variables and the latent variables assuming that the values of latent variables are fixed at their posterior modes. We assign the new dependence structure to the data and reestimate the latent variables until convergence. Simulation in different context show that the procedures of bivariate copula selection could be at least 80% accurate. We have developed the R package `vifcop` to make it easier to estimate the factor copula models.

We illustrate the proposed methodology with two high dimensional real data sets. The first one considers the daily temperature time series at 479 stations in Germany, while the second one analyzes the stock return dependence of 218 European companies. In general, the structured factor copula model can capture quite well the dependence structure of high dimensional data. The first common factor reveals most of the tail dependence among the variables with a large proportion of bivariate Student- $t$  copulas. The latent factors in higher layers would correct for the remain dependence with a combination of Gaussian, Frank, and Gumbel copulas.

The rest of the paper is organized as follows. Section 2 introduces the structured factor copula models. Section 3 presents the VI algorithm for high dimensional structured factor copula models. Section 4 shows some numerical simulations of the proposed factor copula models. Section 5

illustrates the applications with real data. Finally, conclusions are reached in Section 6.

## 2 Model specification

Let  $X = (X_1, \dots, X_d)'$  be the  $d$ -dimensional continuous random variable that we want to model their joint dependence structure and let  $F_1(x_1), \dots, F_d(x_d)$  be their marginal cumulative distribution functions (cdf). Sklar (1959) considers a copula as a joint cdf function defined in the unit hypercube  $[0, 1]^d$  with uniformly univariate margins,  $C(u_1, \dots, u_d) = F(x_1, \dots, x_d)$ , where  $u_i = F_i(x_i)$ , for  $i = 1, \dots, d$ . It is well-known that the variable  $U_i = F_i(X_i) \sim \mathbb{U}(0, 1)$ , for  $i = 1, \dots, d$ . Hence, copula helps to separate the marginal distributions from the dependence structure. Several examples of bivariate copula functions are shown in Table 1. However, constructing high dimensional copula functions is difficult. Instead, Bedford and Cooke (2001, 2002), and Aas et al. (2009), among others, decompose the copula density into a sequence of bivariate copulas and conditional bivariate copulas. Therefore, the dependence structure could be considered as a hierarchical vine where the dependence among variables is driven by bivariate linkages. As the number of variables increases, the number of possible trees as well as the number of vine copula parameters becomes explosive which results in a truncated vine model, see Brechmann et al. (2012).

On the other hand, Krupskii and Joe (2013) and Joe (2014), for tackling the curse of dimensionality, consider several latent variables at the root and describe the dependence structure through bivariate links between copula data and the latent variables, see Figure 1. In the next section, we focus on the one factor and structured factor copulas proposed by Krupskii and Joe (2013, 2015a) to model the dependence of variables in high dimensions.

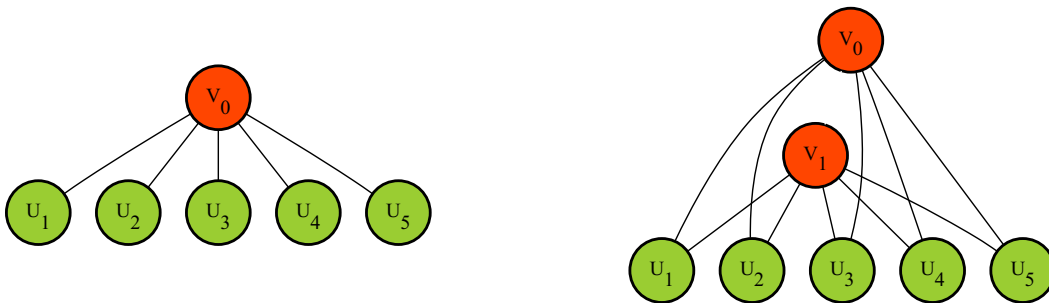


Figure 1: One-factor and two-factor copula models (Krupskii and Joe (2013))

Table 1: Bivariate copula families and their characteristics

Copula	Notation	Copula distribution function	Prior	Range	Kendall's $\tau$
Gaussian	$Gp$	$C_{Gp}(u, v; \theta) = \Phi_2(\Phi^{-1}(u), \Phi^{-1}(v); \theta)$	$\pi_{Gp}(\theta) = \pi_{Gn}(\theta) = \frac{2}{\pi} \frac{1}{\sqrt{1-\theta^2}}$	$\theta \in (0, 1)$	$\frac{2}{\pi} \arcsin(\theta)$
	$Gn$	$C_{Gn}(u, v; \theta) = \Phi_2(\Phi^{-1}(u), \Phi^{-1}(v); \theta)$		$\theta \in (-1, 0)$	
Student- $t$	$Tp$	$C_{Tp}(u, v; \theta, \nu) = T_2(T_\nu^{-1}(u), T_\nu^{-1}(v); \theta, \nu)$	$\pi_{Tp}(\theta) = \pi_{Tn}(\theta) = \frac{2}{\pi} \frac{1}{\sqrt{1-\theta^2}}$	$\theta \in (0, 1), \nu \in (2, 30)$	$\frac{2}{\pi} \arcsin(\theta)$
	$Tn$	$C_{Tn}(u, v; \theta, \nu) = T_2(T_\nu^{-1}(u), T_\nu^{-1}(v); \theta, \nu)$	$\pi_T(\nu) = \text{Gamma}(\nu; 1, 0.1)$	$\theta \in (-1, 0), \nu \in (2, 30)$	
Clayton	$C$	$C_C(u, v; \theta) = (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}}$	$\pi_C(\theta) = \pi_{C180}(\theta) = \frac{2}{(\theta+2)^2}$	$\theta \in (0, \infty)$	$\frac{\theta}{\theta+2}$
	$C_{180}$	$C_{C180}(u, v; \theta) = 1 - u - v + C_C(1 - u, 1 - v; \theta)$			
	$C_{90}$	$C_{C90}(u, v; \theta) = v - C_C(1 - u, v; -\theta)$	$\pi_{C90}(\theta) = \pi_{C270}(\theta) = \frac{2}{(\theta-2)^2}$	$\theta \in (-\infty, 0)$	$\frac{\theta}{\theta-2}$
	$C_{270}$	$C_{C270}(u, v; \theta) = u - C_C(u, 1 - v; -\theta)$			
Gumbel	$G$	$C_G(u, v; \theta) = \exp \left[ - \{ (-\log u)^\theta + (-\log v)^\theta \}^{1/\theta} \right]$	$\pi_G(\theta) = \frac{1}{\theta^2}$	$\theta \in [1, \infty)$	$1 - \frac{1}{\theta}$
	$G_{180}$	$C_{G180}(u, v; \theta) = 1 - u - v + C_G(1 - u, 1 - v; \theta)$			
	$G_{90}$	$C_{G90}(u, v; \theta) = u - C_G(1 - u, v; -\theta)$		$\theta \in (-\infty, -1]$	$-1 - \frac{1}{\theta}$
	$G_{270}$	$C_{G270}(u, v; \theta) = v - C_G(u, 1 - v; -\theta)$			
Frank	$F_p$	$C_{Fp}(u, v; \theta) = -\frac{1}{\theta} \log \left( 1 - \frac{(1-e^{-\theta u})(1-e^{-\theta v})}{1-e^{-\theta}} \right)$	$\pi_F(\theta) = \frac{4}{\theta^2} (1 - B(\theta) + 2D_1(\theta))$	$\theta \in (0, \infty)$	$1 - \frac{4}{\theta} (1 - D_1(\theta))$
	$F_n$	$C_{Fn}(u, v; \theta) = -\frac{1}{\theta} \log \left( 1 - \frac{(1-e^{-\theta u})(1-e^{-\theta v})}{1-e^{-\theta}} \right)$	$\approx \text{Cauchy}(\theta; 0, 6)$	$\theta \in (-\infty, 0)$	
Joe	$J$	$C_J(u, v; \theta) = 1 - \{ (1-u)^\theta + (1-v)^\theta - (1-u)^\theta (1-v)^\theta \}^{1/\theta}$	$\pi_J(\theta) = \sum_{k=1}^{\infty} \frac{8(\theta(k-1)+2-1/k)}{(\theta k+2)^2(\theta(k-1)+2)^2}$	$\theta \in [1, \infty)$	$1 - 4 \sum_{k=1}^{\infty} \frac{1}{k(\theta k+2)(\theta(k-1)+2)}$
	$J_{180}$	$C_{J180}(u, v; \theta) = 1 - u - v + C_J(1 - u, 1 - v; \theta)$	$\approx \frac{2}{(\theta+2)^2}$		
	$J_{90}$	$C_{J90}(u, v; \theta) = v - C_J(1 - u, v; -\theta)$	$\pi_J(\theta) = \sum_{k=1}^{\infty} \frac{8(-\theta(k-1)+2-1/k)}{(\theta k-2)^2(\theta(k-1)-2)^2}$	$\theta \in (-\infty, -1]$	$1 - 4 \sum_{k=1}^{\infty} \frac{1}{k(\theta k-2)(\theta(k-1)-2)}$
	$J_{270}$	$C_{J270}(u, v; \theta) = u - C_J(u, 1 - v; -\theta)$	$\approx \frac{2}{(\theta-2)^2}$		
Independence	$I$	$C_I(u, v) = uv$	-	-	0

$D_1(\theta) = \frac{1}{\theta} \int_0^\theta \frac{\theta}{\exp(\theta)-1}$  denotes the Debye function of order one.  $B(\theta) = \frac{\theta}{\exp(\theta)-1}$  denotes the Bernoulli function.

The table shows some common bivariate copula functions as well as their characteristics such as parameter ranges, and Kendall's  $\tau$  correlation. We divide the symmetric copula functions into positive and negative Kendall's  $\tau$  correlation copulas to prevent the identification issue of the factor copula models.

## 2.1 One-factor copula models

In the one-factor copula model proposed by Krupskii and Joe (2013), the dependence structure is characterized through  $d$  bivariate copulas between the variable  $U_i$  and a latent variable  $V_0$ , where  $V_0 \sim \mathbb{U}(0, 1)$  and  $i = 1, \dots, d$ . Let us denote each bivariate copula density by  $c_{U_i, V_0}(u_i, v_0; \theta_{0i})$ , each bivariate copula function by  $C_{U_i, V_0}(u_i, v_0; \theta_{0i})$ , and each conditional distribution function of  $U_i$  given  $V_0$  by  $F_{U_i|V_0}(u_i|v_0; \theta_{0i})$ , where  $\theta_{0i}$  denotes the bivariate copula parameters. Then assuming that the variables  $U_1, \dots, U_d$  are conditionally independent given the latent variable  $V_0$ , as shown in Krupskii and Joe (2013), we have

$$F_{U_i|V_0}(u_i|v_0; \theta_{0i}) = \frac{\partial F_{U_i, V_0}(u_i, v_0; \theta_{0i})}{\partial v_0} = \frac{\partial C_{U_i, V_0}(u_i, v_0; \theta_{0i})}{\partial v_0}.$$

where  $F_{U_i, V_0}$  is the joint cdf of  $U_i$  and  $V_0$ . The conditional copula density is the following,

$$\begin{aligned} p(u_1, \dots, u_d|v_0; \theta) &= \frac{\partial F(u_1, \dots, u_d|v_0; \theta)}{\partial u_1 \dots \partial u_d} = \prod_{i=1}^d \frac{\partial F_{U_i|V_0}(u_i|v_0; \theta)}{\partial u_i} = \prod_{i=1}^d \frac{\partial C_{U_i, V_0}(u_i, v_0; \theta)}{\partial u_i \partial v_0} \\ &= \prod_{i=1}^d c_{U_i, V_0}(u_i, v_0| \theta_{0i}), \end{aligned} \tag{1}$$

where  $\theta = \{\theta_{01}, \dots, \theta_{0d}\}$  is the set of copula parameters. Note that we have the conditional density  $p(u_1, \dots, u_d|v_0; \theta) = c(u_1, \dots, u_d|v_0; \theta)$  due to the uniform marginal of  $U_i$ , for  $i = 1, \dots, d$ . Krupskii and Joe (2013) calculate the unconditional copula density by integrating over the latent space of  $V_0$  and use maximum likelihood to estimate the parameter  $\theta$  through

$$p(u_1, \dots, u_d; \theta) = \int_0^1 \prod_{i=1}^d c_{U_i, V_0}(u_i, v_0| \theta_{0i}) dv_0.$$

In order to account for more latent variables, Krupskii and Joe (2015a) extend the one-factor copula to the structured factor copulas by adding a hierarchical dependence structure for latent variables, in the case of nested factor copulas or by using the latent variables to capture the conditional dependence in higher tree layers in the case of bi-factor copulas.

## 2.2 Nested factor copula models

Krupskii and Joe (2015a) propose a nested factor copula model by dividing  $d$  variables into  $G$  groups, where  $d_g$  is the number of variables in group  $g$ , for  $g = 1, \dots, G$ , such that  $\sum_{g=1}^G d_g = d$ . In each group, the dependence structure is characterized through  $d_g$  bivariate copulas between the variable  $U_{i_g}$  and the group latent variable  $V_g$ , where  $V_g \sim \mathbb{U}(0, 1)$  and  $i_g = 1, \dots, d_g$ . The dependence among groups is determined through  $G$  bivariate copulas between the group latent variable  $V_g$  and the common latent variable  $V_0$ , where  $V_0 \sim \mathbb{U}(0, 1)$ , see Figure 2. So, the joint dependence is modelled through  $G + 1$  latent variables  $V = \{V_0, V_1, \dots, V_G\}'$  and  $d + G$  bivariate links. Two variables  $U_{i_g}$  and  $U_{j_g}$  in the same group are conditionally independent given the latent group factor  $V_g$  and they are also conditionally independent from the other group factor  $V_{g'}$  for  $g \neq g', i_g \neq j_g$ . Krupskii and Joe (2015a) consider the nested factor copula model as a hierarchical dependence from a common root variable, hence also as an extension of the one-factor copula model. Let  $c_{U_{i_g}, V_g}(u_{i_g}, v_g | \theta_{gi_g})$  be the bivariate copula density of  $U_{i_g}$  and  $V_g$  and let  $c_{V_g, V_0}(v_g, v_0 | \theta_{0g})$  be the bivariate copula density of  $V_g$  and  $V_0$ , for  $g = 1, \dots, G$ , and  $i_g = 1, \dots, d_g$ . Then, the conditional density function for the nested factor copulas is the following,

$$p(u_1, \dots, u_d, v_1, \dots, v_g | v_0; \theta) = p(u_1, \dots, u_d | v_1, \dots, v_g; \theta) p(v_1, \dots, v_g | v_0; \theta), \quad (2)$$

where  $\theta = \{\theta_{01}, \dots, \theta_{0G}, \theta_{11}, \dots, \theta_{Gd_g}\}$  is the set of copula parameters, and from Eq. (1), the conditional density for each layer is,

$$\begin{aligned} p(u_1, \dots, u_d | v_1, \dots, v_g; \theta) &= \prod_{g=1}^G \prod_{i_g=1}^{d_g} c_{U_{i_g}, V_g}(u_{i_g}, v_g | \theta_{gi_g}), \\ \text{and } p(v_1, \dots, v_g | v_0; \theta) &= \prod_{g=1}^G c_{V_g, V_0}(v_g, v_0 | \theta_{0g}). \end{aligned}$$

## 2.3 Bi-factor copula model

Similar to the nested factor copula model, the bi-factor copula model also splits  $d$  variables into  $G$  groups. However, the dependence of variables is modeled by a sequence of bivariate links between the variables and the latent factors, see Figure 3. In fact, bi-factor models extend one-factor



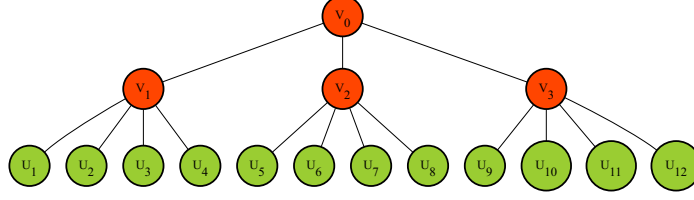


Figure 2: Nested factor copulas with  $d = 12$  and  $G = 3$  (Krupskii and Joe (2015a))

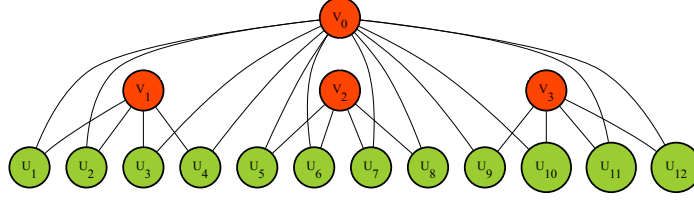


Figure 3: Bi-factor copulas with  $d = 12$  and  $G = 3$  (Krupskii and Joe (2015a))

copulas by first assuming a one-factor copula in the first tree layer. Then, in the second tree, the relationship among variables in group  $g$  is expressed through conditional bivariate copulas between the group latent variable  $V_g$  and its member variable  $U_{i_g}$  conditional on the common factor  $V_0$ , for  $i_g = 1, \dots, d_g$ . Krupskii and Joe (2015a) assume that  $V_0, \dots, V_G$  are independent and identically  $\mathbb{U}(0, 1)$  distributed for identifiability. Therefore, variables in the same group  $g$  are conditionally independent given  $V_0$  and  $V_g$ , while variables in different groups are conditionally independent given  $V_0$  only. The two-factor copula model in Krupskii and Joe (2013) is a special case of bi-factor copula model where  $G = 1$ . In order to construct the conditional bi-factor copula density, we follow the pair-copula decomposition method in Aas et al. (2009). Let  $c_{U_{i_g}, V_0}(u_{i_g}, v_0 | \theta_{0i_g})$  be the bivariate copula density of  $U_{i_g}$  and  $V_0$ , and let  $F_{U_{i_g}|V_0}(u_{i_g} | v_0, \theta_{0i_g})$  be the conditional distribution of  $U_{i_g}$  given  $V_0$ , for  $g = 1, \dots, G$ , and  $i_g = 1, \dots, d_g$ . It is straightforward that the conditional variable  $U_{i_g|v_0} = F(U_{i_g} | V_0, \theta_{0i_g}) \sim \mathbb{U}(0, 1)$ . Additionally, let  $c_{U_{i_g}, V_g|V_0}(u_{i_g|v_0}, v_g | \theta_{gi_g})$  be the conditional bivariate copula density of  $U_{i_g|v_0}$  and  $V_g$  given  $V_0$ . Therefore, the bi-factor copula density function is given by,

$$\begin{aligned}
 p(u_1, \dots, u_d | v_0, \dots, v_G; \theta) &= \prod_{g=1}^G \prod_{i_g=1}^{d_g} p(u_{i_g} | v_0, v_g; \theta) \\
 &= \prod_{g=1}^G \prod_{i_g=1}^{d_g} c_{U_{i_g}, V_g|V_0}(u_{i_g|v_0}, v_g | \theta_{gi_g}) c_{U_{i_g}, V_0}(u_{i_g}, v_0 | \theta_{0i_g}),
 \end{aligned} \tag{3}$$

where  $\theta = \{\theta_{01}, \dots, \theta_{0d}, \theta_{11}, \dots, \theta_{Gd_G}\}$  is the set of copula parameters.

The bi-factor copula model requires  $2d$  bivariate copula links and the computational expensive transformation  $u_{ig|v_0} = F_{U_{ig}|V_0}(u_{ig}|v_0, \theta_{0ig})$  to obtain the transformed variables in the second tree layer. On the other hand, the nested factor copula model only requires  $d + G$  bivariate links. Also, there is no need to obtain the conditional variables. According to the simplified assumption for vine copulas in Haff et al. (2010), we also assume that the bivariate copula density  $c_{U_{ig}, V_g|V_0}(u_{ig}|v_0, v_g)$  does not change with different values of the conditional variable  $V_0$ .

## 2.4 Discussion

The structured factor copula models deliver meaningful interpretations on the dependence structure of observable variables. For instance, in the stock return example that will be shown in Section 5.2, it is expected that the companies operating in the same country are more dependent and they are all affected not only by a common economic latent factor but also by the country latent factors. The structured factor model could also handle hundreds of variables and capture well different behaviours in the upper and lower tails. In general, the hierarchical structure and group members can be specified based on the prior knowledge as the assumption of hierarchical models. However, the bivariate copula links are unknown. Furthermore, we face several challenges when estimating the factor copula models. Firstly, the choice of bivariate copula links in the one-factor model is arbitrary but we can have two global optimal solutions when estimating a model with symmetric copula families. For example, if we simulate from an one-factor Gaussian copula with all positive correlation parameters, depending on the initial point, it could happen that all the estimated correlations are negative and the estimated latent variables are equal to the survival function of the true latent. In order to prevent this phenomenon, we divide the copula families into positive dependence copulas and negative dependence copulas. For instance, a positive Gaussian copula has  $\theta \in [0, 1]$  and a positive Frank copula has  $\theta \in (0, \infty)$  are in the first group, while a negative Gaussian copula  $\theta \in [-1, 0)$  and a negative Frank copula  $\theta \in (-\infty, 0)$  are in the second group. Secondly, in a higher latent space, without restrictions, Krupskii and Joe (2013) report a circumstance for the two-factor copula model that the dependence of the latent variable in the higher layer is stronger than the first one. They implied a zero correlation for one bivariate link in the second layer to make the model identifiable. Nikoloulopoulos and Joe (2015) reported that the two-factor Student- $t$  copula

model is nearly non-identifiable because the likelihood is quite flat. Empirically, when the first tree layer has captured most of the tail dependence, the second tree only reveals a small remaining effect. Hence, we impose not to use the Student- $t$  copula in the higher latent bivariate links and consider other tail dependence families such as the Clayton, Gumbel, and Joe copulas. Thirdly, we also face a situation that the bivariate copula density of Clayton and Survival Joe are very much alike especially when the Kendall’s  $\tau$  correlation is high. This results in a scenario that different factor models containing Clayton and Joe copulas can have similar values of the AIC and/or BIC. Thus, they can be used interchangeable. This phenomenon also happens between the Gaussian and Student copulas where the degree of freedom is high. Finally, Krupskii and Joe (2013) suggest using the bivariate normal score plot or tail-weighted measures of dependence proposed by Krupskii and Joe (2015b) to identify the unknown bivariate links. Because the method requires analyzing each pair of observables, hence becomes infeasible in high dimensions, we propose a procedure in the next section for selecting the best couple links.

### 3 Bayesian approach

In this section, we apply a Bayesian approach to make inferences on both latent variables and bivariate copula parameters. Due to the time demanding cost of MCMC samplers, we employ the variational inference proposed by Kucukelbir et al. (2017) based on the reparameterization of the model parameters. VI looks for a simple proposal distribution that is closest to the posterior in term of minimizing Kullback-Leibler divergence between them. Firstly, we address how the VI approach makes inferences on the factor copula parameters and the latent variables with known bivariate copula links. Due to several restrictions on the parameters, we transform the constraint space of the copula parameters to the real coordinate space and approximate the transformed posterior distribution in the real domain with a product of Gaussian univariate distributions. Then, we apply the stochastic optimization to minimize the Kullback-Leibler divergence from the transformed posterior distribution to the factorized Gaussian distribution. The unbiased noisy gradient used in optimization is calculated using Monte Carlo samples from the proposal distribution. In the case of unknown copula links, we derive an automatic procedure to select the most appropriate bivariate copula functions. Starting with a random initial structure, we obtain the posterior modes of the

proposal distribution of the latent variables. Then, we reassess the agreement of bivariate copulas between the copula data and the latent variables assuming that the latent variables are fixed at the posterior modes. We choose the bivariate copula function that minimizes the BIC for each link in the factor model, that leads to a smaller BIC value of the factor copula model. The procedure is repeated until the bivariate copula functions remain unchanged.

### 3.1 Prior distributions

For all structured factor copula models, we assume  $\mathbb{U}(0, 1)$  prior distributions for the latent variables  $v = (v_0, \dots, v_G)$ . To solve for different restrictions in the bivariate copula parameters, we impose a vague but proper prior distribution for the Kendall's tau correlation such that for positive bivariate copula  $\tau \sim \mathbb{U}(0, 1)$  and for negative bivariate copula  $\tau \sim \mathbb{U}(-1, 0)$ . Then, we calculate the induced prior distributions for copula parameters using the transformation of the probability density functions  $\pi(\theta) = \pi(\tau) \left| \frac{\partial \tau}{\partial \theta} \right|$ . Table 1 shows the induced prior distributions for the copula parameters  $\theta$ . For the Student- $t$  copula, we let the degree of freedom  $\nu \sim \text{Gamma}(1, 0.1)$ . In order to simplify the notations, we consider both latent variables and bivariate copula parameters as the parameters of factor models as  $\Theta = \{v, \theta\}$ . Then, the prior density is  $\pi(\Theta) = \pi(v)\pi(\theta)$ .

### 3.2 Posterior distributions

We obtain the copula data  $u = (u_{t1}, \dots, u_{td})'$  after fitting the appropriate marginal cdf functions to the original data,  $u_{ti} = F_i(x_{ti})$ , for  $t = 1, \dots, T$ , and  $i = 1, \dots, d$ . Assuming that we have specified a factor copula structure together with bivariate links, we are interested in making inferences on the parameters  $\Theta$  of the structured factor model. The augmented likelihood is,

$$p(u|\Theta) = \prod_{t=1}^T p(u_{t1}, \dots, u_{td}|v_t; \theta).$$

Then, the joint posterior density up to a normalized constant is

$$\pi(\Theta|u) \propto p(u, \Theta) = \prod_{t=1}^T p(u_{t1}, \dots, u_{td}|v_t; \theta) \prod_{t=1}^T \pi(v_t) \pi(\theta). \quad (4)$$

Substituting Eq. (1) into Eq. (4) leads to the posterior distribution of  $\Theta$  for the one-factor copula model,

$$p(\Theta|u) \propto \prod_{i=1}^d \left[ \prod_{t=1}^T c_{U_i, V_0}(u_{ti}, v_{t0} | \theta_{0i}) \pi(\theta_{0i}) \right].$$

Similarly, the posterior distribution of  $\Theta$  for the nested factor copula is given by,

$$p(\Theta|u) \propto \left[ \prod_{g=1}^G \prod_{i_g=1}^{d_g} \left[ \prod_{t=1}^T c_{U_{i_g}, V_g}(u_{ti_g}, v_{tg} | \theta_{gi_g}) \pi(\theta_{gi_g}) \right] \right] \left[ \prod_{g=1}^G \left[ \prod_{t=1}^T c_{V_g, V_0}(v_{tg}, v_{t0} | \theta_{0g}) \pi(\theta_{0g}) \right] \right],$$

while for the bi-factor copula model is,

$$p(\Theta|u) \propto \prod_{g=1}^G \prod_{i=1}^{d_g} \left[ \prod_{t=1}^T c_{U_{i_g}, V_g | V_0}(u_{t, i_g | v_{t0}}, v_{tg} | \theta_{gi_g}) \prod_{t=1}^T c_{U_{i_g}, V_0}(u_{ti_g}, v_{t0} | \theta_{0i_g}) \pi(\theta_{gi_g}) \pi(\theta_{0g}) \right].$$

The bi-factor copula model requires the conditional transformation  $u_{ti_g | v_{t0}} = F(u_{ti_g} | v_{t0})$  whenever we generate new samples of  $\theta_{0i}$  or  $v_{t0}$ , which makes the MCMC approach computationally expensive. However, as we can still calculate the joint posterior distribution of the factor copula parameters, it is feasible to employ the No-U-Turn Sampler (NUTS) proposed by Hoffman and Gelman (2014) to make inferences. In high dimensions, NUTS converges to the target distribution quicker than Metropolis or Gibbs sampling. Even though, we overcome such time demanding sampler by employing an approximated approach to the posterior distribution.

### 3.3 Variational Inference

Considering a factor copula model in Section 3.2, the parameter set is  $\Theta$ , and let  $N$  be the total number of parameters in  $\Theta$ . We assume that the posterior density is approximated by a proposal density  $q(\Theta; \lambda)$ , parameterized by a vector  $\lambda$  such that  $q(\Theta; \lambda)$  is close to the joint posterior  $p(\Theta|u)$ . In order to find the values of the free parameters in  $\lambda$ , the VI approach tries to minimize the Kullback-Leibler divergence from the posterior  $p(\Theta|u)$  to the proposal  $q(\Theta; \lambda)$ , i.e.:

$$\arg \min_{\lambda} \text{KL} (q(\Theta; \lambda) || p(\Theta|u)) = -\mathbb{E}_q[\log p(u|\Theta)] + \mathbb{E}_q[\log q(\Theta; \lambda)]$$

$$\text{such that } \text{supp}(q(\Theta; \lambda)) \subseteq \text{supp}(p(\Theta|u))$$

where the support of the proposal  $q(\Theta; \lambda)$  is a subset of the support of the posterior. The proposal  $q(\Theta; \lambda)$  needs to be simple for a tractable approximation, and also it should be expressive in order to match closely with the posterior, see Zhang et al. (2018). Note that, the posterior  $p(\Theta|u)$  is only defined up to a normalizing constant, then we can not perform the optimization directly. Instead, we maximize the evidence lower bound (ELBO) as an equivalent objective function,

$$\begin{aligned} \arg \max_{\lambda} \text{ELBO}(q) &= \mathbb{E}_q[\log p(u, \Theta)] - \mathbb{E}_q[\log q(\Theta; \lambda)] \\ &= \log p(u) - KL(q(\Theta; \lambda) || p(\Theta|u)) \leq \log p(u) \end{aligned} \quad (5)$$

such that when  $q(\Theta; \lambda) = p(\Theta|u)$ , we have  $\text{ELBO} = \log p(u)$ . We obtain  $p(u, \Theta) = p(u|\Theta)\pi(\Theta)$  from Eq. (4). Following Kucukelbir et al. (2017), we apply an Automatic Differentiation Variational Inference (ADVI) algorithm to maximize the objective function ELBO using noisy estimates of its gradients. Firstly, we transform the constrained parameter space to the real coordinate space using a transformation function  $\mathbb{T}_j$  for each parameter,  $\tilde{\Theta} = \{\tilde{\Theta}_j\} = \{\mathbb{T}_j(\Theta_j)\} = \mathbb{T}(\Theta)$ , for  $j = 1, \dots, N$ . Then we assume a product of univariate Gaussian densities as the proposal density,

$$q(\tilde{\Theta}; \mu, \sigma^2) = \phi_N(\tilde{\Theta}; \mu, \sigma^2) = \prod_{j=1}^N \phi(\tilde{\Theta}_j; \mu_j, \sigma_j^2), \quad (6)$$

where  $\mu = \{\mu_j\}$  and  $\sigma = \{\sigma_j\}$  for  $j = 1, \dots, N$ . This factorized distribution is also called a mean field distribution. Kucukelbir et al. (2017) propose several transformation functions  $\mathbb{T}$  to suit with different restrictions in the domain of  $\Theta$ , see Table 2. Let  $\omega = \{\omega_j\} = \log(\sigma)$  to relax the positive constraint of  $\sigma$ , the variational parameters become  $\lambda = \{\mu_j, \omega_j\}$ , for  $j = 1, \dots, N$ . The joint density  $p(u, \tilde{\Theta})$  is derived based on the Jacobian of the inverse transformation,  $J_{\mathbb{T}^{-1}}(\tilde{\Theta})$ ,

$$p(u, \tilde{\Theta}) = p(u, \mathbb{T}^{-1}(\tilde{\Theta})) |\det J_{\mathbb{T}^{-1}}(\tilde{\Theta})|. \quad (7)$$

So, instead of optimizing in the constrained parameter space, we optimize the variational parameters in the real space  $\mathcal{R}^N$ . Substituting Eq. (7) into the ELBO Eq. (5), the objective function becomes,

$$\arg \max_{\lambda} \text{ELBO}(q) = \arg \max_{\lambda=\{\mu, \omega\}} \mathbb{E}_{q(\tilde{\Theta})} \left[ \log p(u, \mathbb{T}^{-1}(\tilde{\Theta})) + \log |\det J_{\mathbb{T}^{-1}}(\tilde{\Theta})| \right] - \mathbb{E}_{q(\tilde{\Theta})} [\log q(\tilde{\Theta}; \lambda)] \quad (8)$$

where  $\mathbb{E}_{q(\tilde{\Theta})}[-\log q(\tilde{\Theta}; \lambda)] = \sum_{j=1}^N 0.5 \log(2\pi e(\exp \omega_j)^2)$  is the closed-form formula of the entropy of multivariate Gaussian distribution. For each factor copula model, the objective function ELBO corresponds to the joint posterior density up to a normalized constant  $p(u, \Theta)$ . The Monte Carlo method is used to evaluate the ELBO by sampling  $S$  samples  $\tilde{\Theta}_{(s)} \sim \Phi_N(\mu, (\exp \omega)^2)$  and plugging Eq. (4) into Eq. (8) as,

$$\text{ELBO}(q) \approx \frac{1}{S} \sum_{s=1}^S \left[ \log p(u, \mathbb{T}^{-1}(\tilde{\Theta}_{(s)})) + \log |\det J_{\mathbb{T}^{-1}}(\tilde{\Theta}_{(s)})| \right] - \mathbb{E}_{q(\tilde{\Theta})}[\log q(\tilde{\Theta}; \lambda)]. \quad (9)$$

We use the stochastic gradient ascent to maximize the ELBO over  $\lambda = \{\mu, \omega\}$ . The gradients of ELBO with respect to parameters  $\lambda$  of the proposal distribution is denoted as  $\nabla_{\lambda} \text{ELBO}$ . Using a similar trick, the noisy gradient is approximated by drawing  $M$  samples  $\tilde{\Theta}_{(m)} = \mu + \exp(\omega)\eta_{(m)}$ , where  $\eta_{(m)} \sim \Phi(0, I_N)$ , and taking the average over the gradients of each sample point. Kucukelbir et al. (2017) recommend that  $M = 1$  is sufficient for this purpose. The derivatives of the ELBO with respect to  $\lambda$  follow the chain rule when we apply the transformations,

$$\begin{aligned} \nabla_{\lambda} \text{ELBO} &= \nabla_{\lambda} \mathbb{E}_{q(\tilde{\Theta})} \left[ \log p(u, \mathbb{T}^{-1}(\tilde{\Theta})) + \log |\det J_{\mathbb{T}^{-1}}(\tilde{\Theta})| \right] - \nabla_{\lambda} \mathbb{E}_{q(\tilde{\Theta})}[\log q(\tilde{\Theta}; \lambda)] \\ &\approx \frac{1}{M} \sum_{m=1}^M \nabla_{\lambda} \left[ \log p(u, \mathbb{T}^{-1}(\tilde{\Theta}_{(m)})) + \log |\det J_{\mathbb{T}^{-1}}(\tilde{\Theta}_{(m)})| \right] - \nabla_{\lambda} \mathbb{E}_{q(\tilde{\Theta})}[\log q(\tilde{\Theta}; \lambda)] \\ &\approx \frac{1}{M} \sum_{m=1}^M \left[ \nabla_{\Theta} \log p(u, \Theta_{(m)}) \nabla_{\tilde{\Theta}} \mathbb{T}^{-1}(\tilde{\Theta}_{(m)}) + \nabla_{\tilde{\Theta}} \log |\det J_{\mathbb{T}^{-1}}(\tilde{\Theta}_{(m)})| \right] \nabla_{\lambda} \tilde{\Theta}_{(m)} - \nabla_{\lambda} \mathbb{E}_{q(\tilde{\Theta})}[\log q(\tilde{\Theta}; \lambda)] \end{aligned} \quad (10)$$

Schepsmeier and Stöber (2014) derive the partial derivatives of a copula density  $c(u, v|\theta)$  w.r.t. its arguments. Hence, it is straightforward to obtain  $\nabla_{\Theta} \log p(u, \Theta_{(m)})$  and  $\nabla_{\lambda} \text{ELBO}$ . Then, the free variational parameter  $\lambda$  is updated along the gradient,  $\lambda \leftarrow \lambda + \varrho \nabla_{\lambda} \text{ELBO}$  with a vector step size  $\varrho$ . The vector of step size sequence or the learning rate  $\varrho$  needs to be adaptively adjusted corresponding to the variance of the gradient. When the variance of the gradient is large, we expect the step size to be small and vice-versa. Kucukelbir et al. (2017) modify the RMSPROP sequence in Tieleman and Hinton (2012) to guarantee that the step size consequence decays sufficiently, see Appendix A.

Algorithm 1 outlines the procedure of ELBO optimization. Starting with an initial copula struc-

Table 2: Transformation functions from a constraint domain to the real domain

Parameter range	$\tilde{\Theta} = \mathbb{T}(\Theta) \in \mathbb{R}$	$\Theta = \mathbb{T}^{-1}(\tilde{\Theta})$	$J_{\mathbb{T}^{-1}}(\tilde{\Theta}) = \frac{\partial \mathbb{T}^{-1}(\tilde{\Theta})}{\partial \tilde{\Theta}}$
$\theta \in [0, 1]$	$\tilde{\theta} = \log\left(\frac{\theta}{1-\theta}\right)$	$\theta = \frac{\exp \tilde{\theta}}{1+\exp \tilde{\theta}}$	$J = \frac{\exp \tilde{\theta}}{(1+\exp \tilde{\theta})^2}$
$\theta \in [0, \infty]$	$\tilde{\theta} = \log(\theta)$	$\theta = \exp \tilde{\theta}$	$J = \exp \tilde{\theta}$
$\theta \in [L, \infty]$	$\tilde{\theta} = \log(\theta - L)$	$\theta = \exp \tilde{\theta} + L$	$J = \exp \tilde{\theta}$
$\theta \in [-1, 0]$	$\tilde{\theta} = \log\left(\frac{-\theta}{1+\theta}\right)$	$\theta = -\frac{\exp \tilde{\theta}}{1+\exp \tilde{\theta}}$	$J = -\frac{\exp \tilde{\theta}}{(1+\exp \tilde{\theta})^2}$
$\theta \in [-\infty, 0]$	$\tilde{\theta} = \log(-\theta)$	$\theta = -\exp \tilde{\theta}$	$J = -\exp \tilde{\theta}$
$\theta \in [-\infty, U]$	$\tilde{\theta} = \log(U - \theta)$	$\theta = U - \exp \tilde{\theta}$	$J = -\exp \tilde{\theta}$
$\theta \in [L, U]$	$\tilde{\theta} = \log\left(\frac{\theta-L}{U-\theta}\right)$	$\theta = \frac{L+U\exp \tilde{\theta}}{1+\exp \tilde{\theta}}$	$J = \frac{(U-L)\exp \tilde{\theta}}{(1+\exp \tilde{\theta})^2}$
$\theta \in [-\infty, \infty]$	$\tilde{\theta} = \theta$	$\theta = \tilde{\theta}$	$J = 1$

Transformation function for different restrictions in constrain parameter space of  $\Theta$

ture and the Gaussian proposal distribution. We sample from the proposal and obtain the values  $\tilde{\Theta}_{(m)}$ . After that, we calculate the noisy gradient of ELBO w.r.t. to the variational parameter  $\lambda$  in Eq. (10), and update them along the gradient line using adaptive step sizes. We calculate the value of ELBO using Eq. (9) and stop if the relative change is less than a threshold. We recover the distribution of  $\Theta$  using the inverse transformation  $\Theta = \mathbb{T}^{-1}(\tilde{\Theta})$  when sampling  $\tilde{\Theta} \sim \Phi_N(\mu, \exp(\omega)^2)$ . In general, the variational distribution of  $\Theta$  is non-Gaussian due to the Jacobian transformation.

### 3.4 Model check

Given a structured factor copula, it is straightforward to derive the ELBO and optimize using the ADVI algorithm. However, when the bivariate copula links are unknown, we can take advantage of the posterior modes of the latent variables, called  $\bar{v}$ , to inspect the assumption of the initial links. The idea is to find the bivariate copula functions that minimize the BIC of bivariate copulas between  $u$  and  $\bar{v}$ . In particular, we start with a random initial structure and obtain the posterior modes  $\bar{v}$ . Then, we reassess the agreement of bivariate copulas between  $u_i$  and  $\bar{v}$ . The ideal bivariate function should minimize the BIC of each couple link. By selecting these functions, the model goodness of fit increases. For example, in one-factor copula model, let  $c_{U_i, V_0}(u_i, v_0; \theta_{0i})$  be the current copula



**Data:** Copula data  $u = \{u_{ti}\}$  and a structured copula model

**Result:** Bivariate copula links, and samples of  $\Theta$  from the proposal distribution

Initial bivariate copula links;

**while** *Any change in copula types* **do**

Initialize  $i = 0$ , vector  $\mu^{(i)} = 0, \omega^{(i)} = 0$  ;

**while** *Change in ELBO is above some threshold* **do**

Draw M samples  $\eta_m \sim \Phi(0, I_N)$ ;

Obtain  $\tilde{\Theta}_{(m)} = \mu^{(i)} + \exp(\omega^{(i)})\eta_{(m)}$ ;

Obtain the noisy gradient  $\nabla_{\mu}\text{ELBO}$  and  $\nabla_{\omega}\text{ELBO}$  ;

Update  $\mu^{(i+1)} \leftarrow \mu^{(i)} + \varrho_{\mu}^{(i)} \nabla_{\mu}\text{ELBO}$  ;

Update  $\omega^{(i+1)} \leftarrow \omega^{(i)} + \varrho_{\omega}^{(i)} \nabla_{\omega}\text{ELBO}$  ;

Incremental iteration (i) ;

**end**

**Select bivariate copula links between  $u$  and  $\bar{v}$  based on minimum the BIC ;**

Reassign the copulas and estimate ;

**end**

Sample  $\tilde{\Theta} \sim \Phi_N(\mu, \exp(\omega))$ , obtain  $\Theta = \mathbb{T}^{-1}(\tilde{\Theta})$  ;

Return bivariate copula links and  $\Theta$  samples ;

**Algorithm 1:** Modified of the ADVI algorithm in (Kucukelbir et al., 2017)

density of the link between variable  $U_i$  and  $V_0$ . The BIC for this bivariate link is,

$$\text{BIC}_i = -2\log c_{U_i, V_0}(u_i, \bar{v}_0; \hat{\theta}_{0i}) + n_i \log(T)$$

where  $\bar{v}_0 = \{\bar{v}_{10}, \dots, \bar{v}_{T0}\}$  is the posterior modes of the latent variables,  $\hat{\theta}_i$  is the maximum likelihood estimator of the bivariate copula, and  $n_i$  is the number of parameters in  $c_{U_i, V_0}$ . The BIC of the one factor copula could be derived as,

$$\begin{aligned} \text{BIC} &= -2 \sum_{i=1}^d \log c_{U_i, V_0}(u_i, \bar{v}_0; \hat{\theta}_{0i}) + (T + \sum_{i=1}^d n_i) \log(T) \\ &= \sum_{i=1}^d \text{BIC}_i + T \log(T). \end{aligned}$$

In the selection step, we propose a new bivariate link  $c_{U_i, V_0}^{(*)}(u_i, v_0; \theta_{0i})$  among possible copula functions (see Table 1) which minimizes the bivariate copula BIC,

$$\underline{\text{BIC}}_i^{(*)} = -2\log c_{U_i, V_0}^{(*)}(u_i, \bar{v}_0; \theta_{0i}^{(*)}) + n_i^{(*)}\log(T) \leq \text{BIC}_i$$

where  $n_i^{(*)}$  is the number of parameters and  $\theta_{0i}^{(*)}$  is the maximum likelihood estimation of  $c_{U_i, V_0}^{(*)}$ . Therefore,

$$\underline{\text{BIC}}^{(*)} = \sum_{i=1}^d \underline{\text{BIC}}_i^{(*)} + T\log(T) \leq \sum_{i=1}^d \text{BIC}_i + T\log(T) = \text{BIC}$$

If there are any changes in the bivariate copula links, we update the copula structure and estimate the new model. The  $\text{BIC}^{(*)}$  of the new model will be at least as good as the  $\underline{\text{BIC}}^{(*)}$ . This guarantees that the model goodness of fit increases. Until there is no difference in the bivariate copula selection, we reach convergence. Accordingly, the posterior modes of the latent variables help to find out the most appropriate candidate for each bivariate links. For the bi-factor models, one can perform bivariate copula selection jointly for all tree layers, however it requires exponential computation. Here, we prioritize the bivariate copula selection of the first tree at root. Then we obtain the transformation  $u_{ig|v_0}$  and implement the selection procedure for higher layers. In Section 4, we illustrate this trade-off for accuracy in different simulated contexts. Algorithm 1 summarizes the procedure for selecting bivariate copula links in the factor copula models.

## 4 Data simulation

In this section, we illustrate the VI algorithm with simulated data sets. We generate samples of  $d = 100$  time series with  $T = 1000$  observation points. First, we randomize the latent variable  $v$  uniformly in the unit range  $[0, 1]$ , then we simulate copula data  $u$  based on the conditional bivariate copulas of  $u_i$  and  $v$ , for  $i = 1, \dots, d$ . The chosen bivariate links are all Gaussian copulas, all Student- $t$  copulas, all Clayton copulas, all Gumbel copulas, all Frank copulas, all Joe copulas and mixing among bivariate copulas. The experiments for the rotated Archimedean copulas have the similar results. The Kendall's  $\tau$  correlation of each bivariate copula link is randomized uniformly in the range  $[0.2, 0.8]$ .

In the first experiment, we estimate the factor models given their known structure, while in the

second experiment, we start with a random bivariate copula structure and let Algorithm 1 search for the most appropriate bivariate links in the factor copula models. We consider the copula families shown in Table 1 that contains 7 common bivariate copula functions and their rotations. We repeat the experiment 100 times and calculate the average of statistical criteria. We report the time in seconds using one Intel Core i7-4770 @ 3.40GHz processor. The computation is implemented in the `vifcopula` package.

#### 4.1 One-factor copula model

We report in Appendix B a comparison between the posterior means using variational approximation and the true generated values, see Figure 6. In general, the VI posterior means are close to their true values. The posterior means of the degree of freedom parameters in the Student- $t$  factor copulas are less reliable when  $\nu > 10$ . This is due to the fact that the likelihood of the Student copula is quite flat when  $\nu$  is high which makes the posterior samples of  $\nu$  right skewed. Also in general, the posterior medians or posterior modes are closer to the true values. We also compare with the posterior samples from the MCMC approach and the results are similar. Table 3 shows the summary of statistical criteria for the one-factor Gaussian, Student- $t$ , Clayton, Gumbel, Frank, Joe and Mix copulas. The result of the first experiment is shown in Panel (a) and that of the second experiment is shown in Panel (b). Each factor copula model contains 100 bivariate links with about 100 to 200 copula parameters. We use Gauss-Legendre quadrature integration over the latent space to obtain  $\log p(u|\theta)$ . The value of ELBO, AIC, BIC,  $\log p(u|\theta)$  are normalized for 1000 data observations. It usually took less than a minute to estimate the copula parameters at the correct structure model, more details are reported later in Section 4.4. In the second experiment, we start with random bivariate links as an initial dependence structure and let Algorithm 1 recover the original structure, it takes a few iterations of bivariate copula selection to converge. The accuracy rate is quite high, except for the Clayton and Joe. Bivariate copula selection often goes wrong between the Clayton copula and the survival Joe copula when the Kendall's  $\tau$  is high and vice versa. Due to that reason, we restrict the maximum number of selection iterations under 10. After all, we still archive a decent level of the ELBO, AIC and BIC even with misspecified models. The time of estimation gets longer in the case of unknown structure mainly because we need to perform the bivariate copula selection for  $d = 100$  bivariate links across all possible copula

functions. In general, given copula data  $u$  and the estimated latent factor  $v_0$ , we need to calculate the BIC of  $d$  bivariate links for 18 copula types and select the best functions. It is about one second for each bivariate selection, and one can parallel the procedure to speed up the calculation.

Table 3: Model comparison for the one-factor copula models

Copula type	Gaussian	Student- $t$	Clayton	Gumbel	Frank	Joe	Mix
<i>(a) Initial at the correct structure</i>							
ELBO	31.3	32.6	75.2	67.9	56.6	77.1	56.2
AIC	-63.2	-65.5	-146.4	-134.8	-114.3	-149.9	-111.5
BIC	-62.7	-64.5	-146.0	-134.3	-113.8	-149.4	-110.9
$\log p(u \theta)$	31.7	32.9	73.3	67.5	57.2	75.1	55.9
<i>(b) Initial at a random structure</i>							
# Selection iteration	3	5	10	2	3	10	7
% accuracy	99	80	70	99	99	61	85
ELBO	31.3	32.6	75.2	67.9	56.6	77.2	56.2
AIC	-63.2	-65.5	-146.5	-134.8	-114.3	-149.9	-111.5
BIC	-62.7	-64.6	-146.0	-134.3	-113.8	-149.5	-111.0
$\log p(u \theta)$	31.7	32.9	73.3	67.5	57.2	75.1	55.9

We report the statistical criteria for the one-factor copula models. Each factor copula model contains 100 bivariate links with about 100 to 200 copula parameters. We use Gauss-Legendre quadrature integration over the latent space to obtain  $\log p(u|\theta)$ . The value of ELBO, AIC, BIC,  $\log p(u|\theta)$  are normalized for 1000 data observations.

## 4.2 Nested factor copula model

We randomly divide  $d = 100$  time series into  $G = 5$  groups and simulate data from the nested factor copulas. Figure 7 in the Appendix compares the posterior means of the nested factor model to the true generated values. The estimates of common factor  $v_0$  fluctuate significantly around the true means which brings a lot of uncertainty. Note that in this case, we have only 5 group latent variables to infer the distribution of their common factor. Hence, the estimates will be improved when increasing the number of latent groups. We also expect the selection of bivariate links among the latent variables to be less accurate than that between the observable variables and the latent group variables. Table 4 shows the summary of statistical criteria for nested factor copulas. In the second experiment with a random initial structure, it also took a few iterations of bivariate copula selection to converge. Again, the similarity of Clayton and survival Joe copula reduces the number of correct specified links, as shown in Panel (b). The accuracy of bivariate copula functions between observable variables and group latent factors is still high. If the bivariate copulas are mixed, we still

have the accuracy rate at 83%. The time of inference is still quite comparable with the one-factor model, even in this case we have much more latent parameters. More details are shown in Section 4.4.

Table 4: Model comparison for the nested factor copula models

Copula type	Gaussian	Student- <i>t</i>	Clayton	Gumbel	Frank	Joe	Mix
<i>(a) Initial at the correct structure</i>							
ELBO	25.9	27.9	69.3	61.3	50.7	70.3	49.7
AIC	-52.9	-56.8	-137.2	-122.9	-103.5	-139.1	-99.9
BIC	-52.3	-55.8	-136.7	-122.4	-103.0	-138.5	-99.3
$\log p(u \theta)$	26.5	28.6	68.7	61.6	51.8	69.6	50.1
<i>(b) Initial at a random structure</i>							
# Selection iteration	4	5	10	3	4	10	8
% accuracy	96	77	75	99	97	52	83
ELBO	25.8	27.8	69.3	61.2	50.6	70.3	49.7
AIC	-52.7	-56.7	-137.0	-122.8	-103.2	-138.9	-99.8
BIC	-52.2	-55.8	-136.5	-122.3	-102.7	-138.3	-99.2
$\log p(u \theta)$	26.5	28.5	68.6	61.5	51.7	69.5	50.0

We report the statistical criteria for the nested factor copula models. Each factor copula model contains 6 latent factors, 105 bivariate links with about 105 to 210 copula parameters. We use Gauss-Legendre quadrature integration over the latent space to obtain  $\log p(u|\theta)$ . The value of ELBO, AIC, BIC,  $\log p(u|\theta)$  are normalized for 1000 data observations.

### 4.3 Bi-factor copula model

Similar to the nested factor copula experiments, we randomly divide  $d = 100$  variables into  $G = 5$  groups. For the identification issue of bi-factor Student-*t* copulas, the bivariate copulas in the second tree are mixed of other copula families, see Section 2.4. Figure 8 in Appendix B shows the posterior mean using VI approximation in comparison to the true generated values. The posterior means of  $v_0$  are close to their true values while that of  $v_g$  vary in a greater range. Note that, all information of variables is used to make inference on the common latent variable  $v_0$ , hence we obtain more certainty of  $v_0$  than the group latent variables  $v_g$ . Computational time significantly increases because we need to obtain the intensive transformation of  $u_{i_g|v_0} = F(u_{i_g}|v_0)$  when calculating ELBO and its derivatives. As a consequence, if the bivariate links in the first tree are not correctly specified, the links in the second tree would likely be misspecified. Table 5 shows the statistical criteria of bi-factor copula models. We obtain quite a good accuracy in both tree layers. Except for the bi-factor Joe copulas, we could at least recover 70% of bivariate links in bi-factor copula models. The statistical criteria are quite close when the correct structure and the recovered structure are

compared. Even in this case, we perform the copula selection separately for each tree layer. We prioritize the copula selection in the first layer and obtain the conditional variable  $u_{i_g|v_0}$ . Then, we obtain the copula selection for the copula links between  $u_{i_g|v_0}$  and  $v_g$  in the second layer.

Table 5: Model comparison for the bi-factor copula models

Copula type	Gaussian	Student- $t$	Clayton	Gumbel	Frank	Joe	Mix
<i>(a) Initial at the correct structure</i>							
ELBO	56.2	83.8	140.9	126.2	105.0	143.2	107.6
AIC	-115.1	-170.4	-275.0	-254.4	-214.7	-279.4	-212.7
BIC	-114.1	-168.9	-274.0	-253.4	-213.7	-278.5	-211.6
$\log p(u \theta)$	57.8	85.5	137.7	127.4	107.5	139.9	106.6
<i>(b) Initial at a random structure</i>							
# Selection iteration	4	9	9	5	9	10	9
% accuracy Tree 1	99	76	84	99	99	46	85
% accuracy Tree 2	97	83	19	66	92	2	67
ELBO	56.2	83.8	132.2	124.0	104.8	134.2	105.5
AIC	-115.1	-170.1	-263.1	-250.6	-214.0	-266.8	-209.7
BIC	-114.1	-168.8	-262.1	-249.6	-213.0	-265.8	-208.7
$\log p(u \theta)$	57.7	85.3	131.7	125.5	107.2	133.6	105.1

We report the statistical criteria for the bi-factor copula models. Each factor copula model contains 6 latent factors, 200 bivariate links with about 200 to 300 copula parameters. We use Gauss-Legendre quadrature integration over the latent space to obtain  $\log p(u|\theta)$ . The value of ELBO, AIC, BIC,  $\log p(u|\theta)$  are normalized for 1000 data observations.

#### 4.4 Comparison between VI and MCMC estimation

We incorporate the NUTS algorithm using RStan package (2018) and consider the posterior samples as the benchmark of the MCMC approach. For each experiment of the factor copula models, we generate 1000 samples via the NUTS algorithm with 500 burn-in. The chain quickly converges after a hundred iterations. Figure 9, 10 and 11 in Appendix B compare the standard deviations of the posterior samples using VI and MCMC for different factor copula models. In one-factor model and nested factor model, the standard deviations of the parameters  $\theta$  are similar using both methods. In bi-factor copula models, we clearly observe that VI underestimates the standard deviations which is corresponding with the literature of VI estimation, see Blei et al. (2017). However, the underestimation is acceptable for the copula parameters  $\theta$ .

Table 6 compares the time of computation between VI and the MCMC for 1000 samples. As expected, the computational time of variational approximation is much less than that of the

MCMC approach. While the VI convergence time depends on the optimization parameters such as the number of MC samples, the number of MC for calculating the gradients, tolerance, among others. The NUTS depends mainly on the number of iterations. Despite that it is difficult to compare the efficiency of VI and MCMC samples, we obtain a quite reasonable estimate in the limited time with the VI approach. Especially, the complexity of the Student- $t$  copula function makes it really computational expensive for the MCMC approach.

Table 6: Time estimation using VI and NUTS

Copula type	Gaussian	Student- $t$	Clayton	Gumbel	Frank	Joe	Mix
<i>(a) Time estimated (s) using VI</i>							
One-factor	10	509	24	41	11	17	98
Nested factor	20	783	25	43	13	19	121
Bi-factor	90	3366	165	267	87	154	596
<i>(b) Time estimated (s) using NUTS</i>							
One-factor	1097	491166	3762	5262	1083	2221	8535
Nested factor	1786	567862	5977	6098	1264	3478	20520
Bi-factor	17935	1758007	83680	177932	30345	236809	118729

We report the time of estimation for the factor copula models using VI and NUTS for 1000 samples. The VI convergence time depends on its optimization parameters such as the number of MC samples, number of MC for calculating the gradients, tolerance, among others. The NUTS approach depends mainly on the number of iterations.

## 5 Empirical illustration

In this section, we show several examples of high dimensional real data sets. In the first example, we model the dependence of temperatures measured at 479 stations in Germany while in the second example, we model the dependence of 218 stock returns from 10 European countries. We follow the two-stage estimation procedure for estimating the copula parameters, see Joe (2005), and Chen and Fan (2006). In the first stage, we find appropriate models for marginal time series and the second stage, we obtain the copula data using the cdf function,  $u = (u_{t1}, \dots, u_{td}) = (F_1(x_{t1}), \dots, F_d(x_{td}))$ . Then, we apply the selection procedure for one-factor, two-factor, nested factor and bi-factor copula models. The bi-factor model is preferred in both examples when we compare the statistical criteria.

## 5.1 Temperature dependence

We estimate the joint dependence of daily temperatures measured at 479 stations in Germany. The data is collected from the German Meteorological Service. The time period is selected from 01/01/2013 to 31/12/2015 which results in  $T = 1094$  observations. In order to divide stations into groups, we apply a hierarchical clustering approach based on the geographic distance, see Murtagh and Contreras (2011). We obtain 24 station groups where the distance among stations in each group at most 200 kilometers.

Erhardt et al. (2015) model the marginal distribution of temperatures using the ARMA(1,1) process with the skew Student- $t$  distribution of Fernández and Steel (1998) for the noise,

$$x_{ti} = \alpha_{0i} + \sum_{k=1}^K \left( \alpha_{ki} \sin \left( \frac{2\pi kt}{365.25} \right) + \beta_{ki} \cos \left( \frac{2\pi kt}{365.25} \right) \right) + \phi_i x_{t-1,i} + \epsilon_{ti} + \delta_i \epsilon_{t-1,i},$$

$$\epsilon_{ti} \sim F_{Skew-t}(\nu_i, \gamma_i, \sigma_i),$$

where  $(\alpha_{0i}, \phi_i, \delta_i)$  are parameters of ARMA(1,1) process,  $(\nu_i, \gamma_i)$  are the parameters of skew Student- $t$  distribution, and  $(\alpha_{ki}, \beta_{ki})$  are the slopes of Fourier exogenous regressors with different frequencies to account for the seasonal effect. We choose the value of  $K = 2$  to minimize the AIC of marginal models. We estimate the ARMA(1,1) model using MLE and obtain  $\hat{\vartheta}_i = \{\hat{\alpha}_{0i}, \hat{\phi}_i, \hat{\delta}_i, \hat{\alpha}_{1i}, \hat{\beta}_{1i}, \hat{\alpha}_{2i}, \hat{\beta}_{2i}, \hat{\nu}_i, \hat{\gamma}_i, \hat{\sigma}_i\}$ . We check if the choice of a skew Student- $t$  distribution is suitable for the data by performing the Kolmogorov-Smirnov, Anderson-Darling, and Neyman's smooth tests for the goodness of fit. All series passed the tests with  $p$ -values greater than 0.05. The estimates of ARMA(1,1) process for each marginal series are summarized in the Online Appendix. Then, we obtain the copula data using the cdf transformation of the residuals,  $u_{ti} = F_{Skew-t}(\epsilon_{ti} | \hat{\nu}_i, \hat{\gamma}_i, \hat{\sigma}_i)$ . Table 7 shows the summary statistics for one-factor, two-factor, nested factor and bi-factor copulas using the bivariate copula selection procedure. There are such a large proportion of Student- $t$  copulas that all models reveal strong upper and lower tail dependence between the observable and the common factor. This results in a strong tail dependence of temperatures among stations. The second layer of the two-factor copula and bi-factor copula aim to capture the asymmetric tail dependence and the remain correlation. The relationship among the groups in nested factor copula also confirms this asymmetric dependence. These findings are similar to Erhardt et al. (2015) using a truncated vine



copula for a similar problem. In general, the structured factor copula models are more preferred than the multi-factor models. We find that the bi-factor copula model is the most suitable model for the dependence of daily temperatures based on statistical criteria.

Table 7: Model comparison of daily temperature dependence

Structure	One-factor	Two-factor	Nested factor	Bi-factor
AIC	-458.9	-611.4	-935.7	-961.5
BIC	-455.4	-606.2	-931.4	-955.3
$\log p(u \theta)$	230.1	306.7	468.7	482.0
# bivariate links	479	929	503	958
# Gaussian	1	204	26	365
# Student- $t$	286	210	442	389
# Clayton (rotated)	0	6	0	1
# Gumbel (rotated)	192	332	35	154
# Frank (rotated)	0	172	0	47
# Joe (rotated)	0	5	0	2
# Independence	0	29	0	0

Next, we can take advantage of the bi-factor copula model to predict the temperatures at different stations from 01/01/2016 to 31/12/2017. We assume that the temperatures are missing at 24 stations during that period, one station for each group sector. Using the information from the 455 remained stations, we want to infer about the temperatures at the missing locations. The idea is first to make inferences on the latent variables during the prediction period, then we predict the missing temperatures given the latent variable samples. We employ VI to infer the common latent variable and group latent variables. Next, we sample the missing copula variable  $u_{ti} \sim F(u_{ti}|v_{t0}, v_{tg})$  and the predicted temperature  $x_{ti}$  using the marginal models as,

$$\begin{aligned}
u_{ti}^{(n)} &\sim F(u_{ti}|v_{t0}^{(n)}, v_{tg}^{(n)}), \\
\epsilon_{ti}^{(n)} &= F_{S_{kew-t}}^{-1}(u_{ti}^{(n)}; \nu_i, \gamma_i, \sigma_i), \\
x_{ti}^{(n)} &= \hat{\alpha}_{0i} + \sum_{k=1}^K \left( \hat{\alpha}_{ki} \sin \left( \frac{2\pi kt}{365.25} \right) + \hat{\beta}_{ki} \cos \left( \frac{2\pi kt}{365.25} \right) \right) + \hat{\phi}_i x_{t-1,i}^{(n)} + \epsilon_{ti}^{(n)} + \hat{\delta}_i \epsilon_{t-1,i}^{(n)},
\end{aligned}$$

where  $v_{t0}^{(n)}, v_{tg}^{(n)}$  are obtained from the variational samples of the predicted latent variables. Note that, each predictive sample  $x_{ti}^{(n)}$  requires the previous sample  $x_{t-1,i}^{(n)}$ , hence the uncertainty is accumulated along the prediction period. Figure 4 shows the prediction of temperatures at Grafenberg-Kasberg and Bertsdorf-Hörnitz stations. They are chosen such that in Grafenberg-Kasberg group,

we have the information of the 34 other stations while in Bertsdorf-Hörnitz group, we only have the information of the 7 other stations. In general, the predicted mean temperatures together with the 95% prediction interval are quite accurate. We have a better estimation in Grafenberg-Kasberg group because we have more certain estimations of the latent variable in the larger group. The empirical temperatures are included in the prediction interval most of the times.

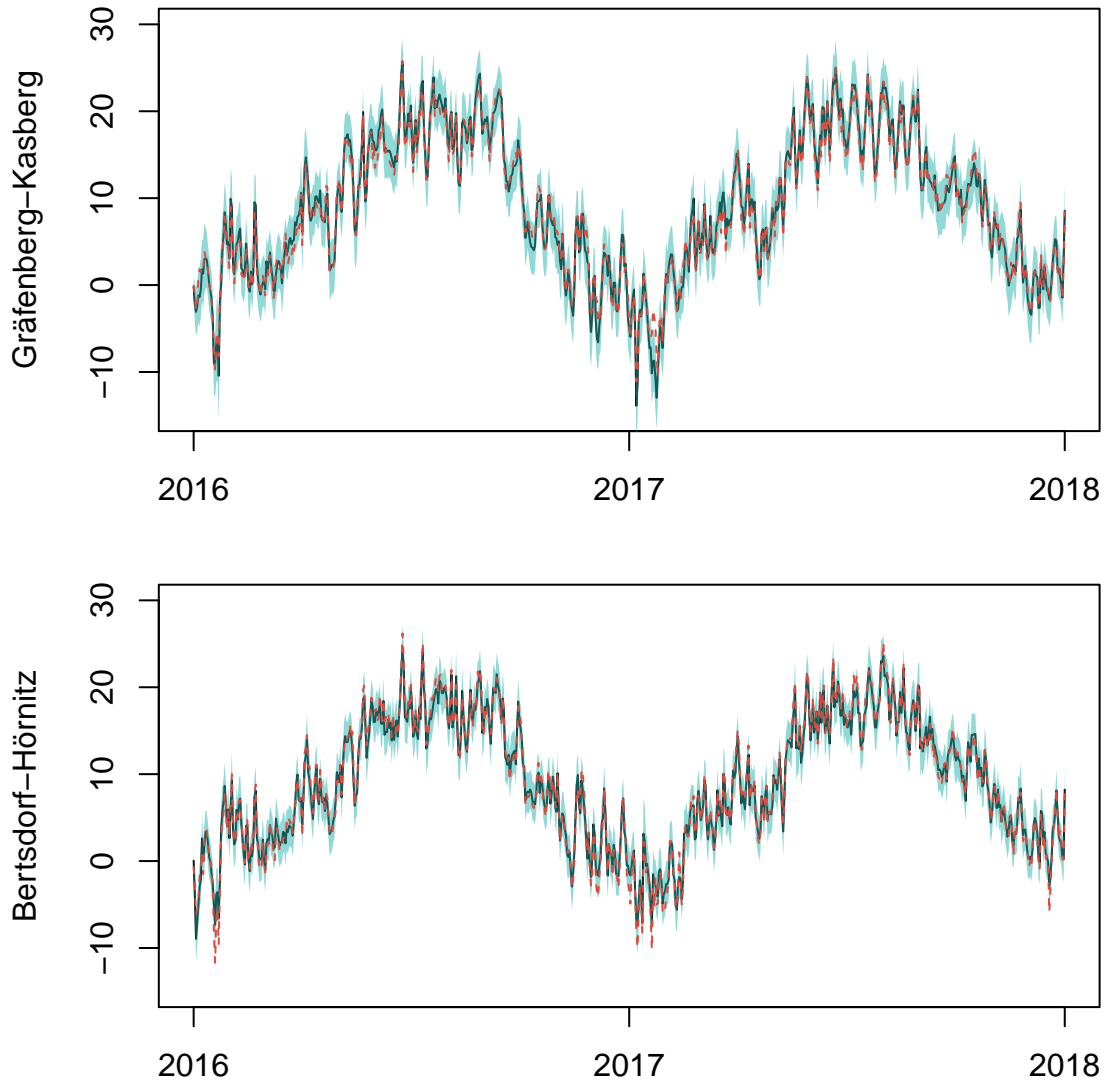


Figure 4: The prediction of temperatures using the estimated bi-factor copula model

The figure shows prediction of temperatures using the estimated bi-factor copula model at Grafenberg-Kasberg and Bertsdorf-Hörnitz stations. They are chosen such that in Grafenberg-Kasberg group, we have the information of 34 other stations while in Bertsdorf-Hörnitz group, we only have the information of 7 other stations. The predicted mean temperatures is the black line, and the measured temperatures is the red dash line. The online version of this figure is in color.

## 5.2 Stock return dependence

We estimate the stock return dependence of 218 firms listed in 10 European countries as Austria (12 firms), Belgium (16 firms), Finland (22 firms), France (34 firms), Germany (28 firms), Ireland (10 firms), Italy (31 firms), Netherlands (19 firms), Portugal (15 firms), Spain (31 firms). The selected companies are large blue-chip stocks which are the constituents of stock market indices in 10 European countries. We take data samples from 01/01/2014 to 31/12/2017 which results in  $T = 1020$  daily observations. We first filter out the conditional mean and variance of all the marginal stock returns using the AR(1) - EGARCH(1,1) process with the skew Student- $t$  distribution of Fernández and Steel (1998) for the innovations,

$$\begin{aligned} x_{ti} &= \mu_i + \phi_i x_{i,t-1} + \sigma_{it} \epsilon_{it} \\ \log(\sigma_{it}^2) &= \alpha_{0i} + \alpha_{1i} \epsilon_{i,t-1} + \delta_i (|\epsilon_{i,t-1}| - E(|\epsilon_{i,t-1}|)) + \beta_i \log(\sigma_{i,t-1}^2) \\ \epsilon_{it} &\sim F_{Skew-t}(\nu_i, \gamma_i) \end{aligned}$$

where  $(\nu_i, \gamma_i)$  are shape parameter and skewness parameter of the skew Student- $t$  distribution, and  $(\alpha_{0i}, \alpha_{1i}, \beta_i, \delta_i)$  are the parameters of exponential GARCH model, see Nelson (1991). We estimate the AR(1) - EGARCH(1,1) model using MLE and obtain the set of marginal parameters  $\hat{\vartheta}_i = \{\hat{\mu}_i, \hat{\phi}_i, \hat{\alpha}_{0i}, \hat{\alpha}_{1i}, \hat{\beta}_i, \hat{\delta}_i, \hat{\nu}_i, \hat{\gamma}_i\}$ . We also check the goodness of fit of the distribution of residuals  $\epsilon_{it}$  as in the previous example using Kolmogorov-Smirnov, Anderson-Darling, and Neyman's smooth tests. Every series passes with  $p$ -values greater than 0.05. We obtain the copula data using the cdf transformation of the residuals,  $u_{it} = F_{Skew-t}(\epsilon_{it} | \nu_i, \gamma_i)$ , and model their joint dependence.

Table 8 shows the summary statistics of one-factor, two-factor, nested factor and bi-factor copula models for stock return data. As expected, we observe a strong tail dependence in the one-factor copula model. The links between copula variables and the common latent variables are mostly Student- $t$  copulas with low degrees of freedom. The two-factor model outperforms the nested factor model which suggests that the joint dependence of European stock returns is only weakly affected by geographic clusters. The bi-factor model is still preferable according to the statistical criteria however the improvement is not substantial. The bivariate links in the second tree layer of the bi-factor model are mostly Gaussian and Frank copula, thus most of the tail dependence has been captured in the first tree layer.

Table 8: Model comparison of stock return dependence

Structure	One-factor	Two-factor	Nested factor	Bi-factor
AIC	-102.4	-117.5	-111.9	-117.7
BIC	-100.5	-114.6	-109.9	-114.9
$\log p(u \theta)$	51.6	59.4	56.3	59.4
# bivariate links	218	434	228	412
# Gaussian	11	19	7	61
# Student- $t$	176	171	177	177
# Clayton (rotated)	0	0	0	4
# Gumbel (rotated)	20	37	34	51
# Frank (rotated)	11	207	10	115
# Joe (rotated)	0	0	0	1
# Independence	0	2	0	24

Next, we compare the Spearman's  $\rho$  and the tail-weighted dependence measures of the bi-factor copula model to that of the empirical copula data. Krupskii and Joe (2015b) propose the tail-weighted dependence as the correlation of transformed variables where the joint movements in the tails have heavier weights,

$$\begin{aligned}\rho_L &= Cor \left( \left(1 - \frac{U_1}{p}\right)^6, \left(1 - \frac{U_2}{p}\right)^6 \middle| U_1 < p, U_2 < p \right) \\ \rho_U &= Cor \left( \left(1 - \frac{1 - U_1}{p}\right)^6, \left(1 - \frac{1 - U_2}{p}\right)^6 \middle| 1 - U_1 < p, 1 - U_2 < p \right)\end{aligned}\tag{11}$$

Figure 5 shows the heat maps of the Spearman's  $\rho$  matrix in the first row and the lower and upper tail-weighted dependence measures in the last two rows. The figures have a similar pattern and the bi-factor copula model can capture well the correlation and the tail dependence. The average of the absolute difference of Spearman's  $\rho$  between the bi-factor copula model and that of the empirical copula data is 0.032 while the average difference of lower tail-weighted dependence measure is 0.059 and upper tail-weighted dependence measure is 0.073. The results suggest that maybe more complex copula models are necessary to account for the asymmetric tail dependence.

## 6 Conclusion

In this paper, we employ the ADVI algorithm in Kucukelbir et al. (2017) to estimate the multi-factor and structured factor copula models. Estimation have been be carried out for high dimensional variables. The posterior means of VI samples are similar to that of MCMC samples while the

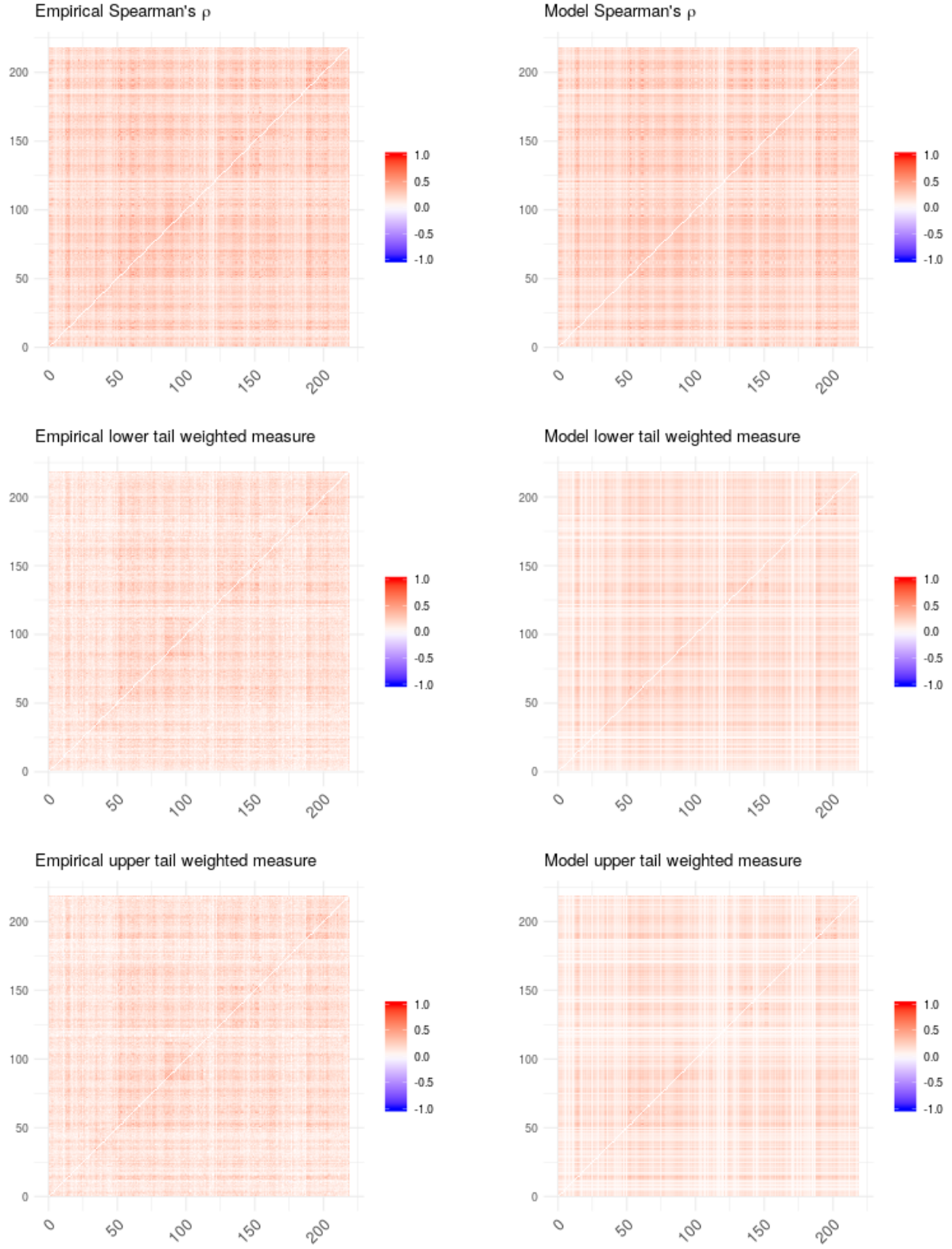


Figure 5: The Spearman's  $\rho$  and the tail-weighted dependence measures of the empirical copula and bi-factor copula model

posterior standard deviations are only underestimated in the case of bi-factor copulas. Instead, VI is much less computational expensive than the MCMC approach. Due to the fast estimation, we take advantage of the posterior modes of the latent variables to reassess the assumption of bivariate copula links and select better bivariate copula functions based on minimizing the BIC. The selection procedure performs quite well with different simulated data. In the future, we can allow for switching variables among the group sectors in the structured copula models or extend the static structured factor model to a dynamic factor copula model as proposed in Nguyen et al. (2017).

## Appendix

### A The step size

At iteration  $i$ , the gradient vector  $g^{(i)} = \nabla_{\lambda^{(i)}} \text{ELBO}$  and step size vector  $\varrho^{(i)}$  are used to update the parameters of copula model. The idea is to downscale the gradient by the exponential moving average of the second moment of the gradient. The parameter with a large derivative will have a quickly decline learning rate. We have for each element  $\varrho_j^{(i)}$ , where  $j = 1, \dots, N$ ,

$$\begin{aligned}\varrho_j^{(i)} &= i^{-0.5+\epsilon} \times \left(1 + \sqrt{s_j^{(i)}}\right)^{-1} \\ s_j^{(i)} &= 0.1g_j^{2(i)} + 0.9s_j^{(i-1)}\end{aligned}$$

where  $\epsilon = 10^{-16}$  to guarantee the step size decay, and an initialization  $s_j^{(1)} = g_j^{2(1)}$  at  $i = 1$ .

### B VI estimation in comparison to the true generated values and the MCMC approach

In Appendix B, we show several comparisons between the posterior means using variational approximation and the true generated values for one-factor, nested factor and bi-factor copula models, see Figure 6, 7, 8. In general, the VI posterior means are close to their true values. The standard deviations of latent factors depend on the number of the series in the sector group.

Figure 9, 10 and 11 compare the standard deviations of the posterior samples using VI and MCMC for one-factor, nested factor and bi-factor copula models. The posterior modes of VI estimation are similar to that of MCMC samples while the standard deviations are only underestimated in the bi-factor copula model.

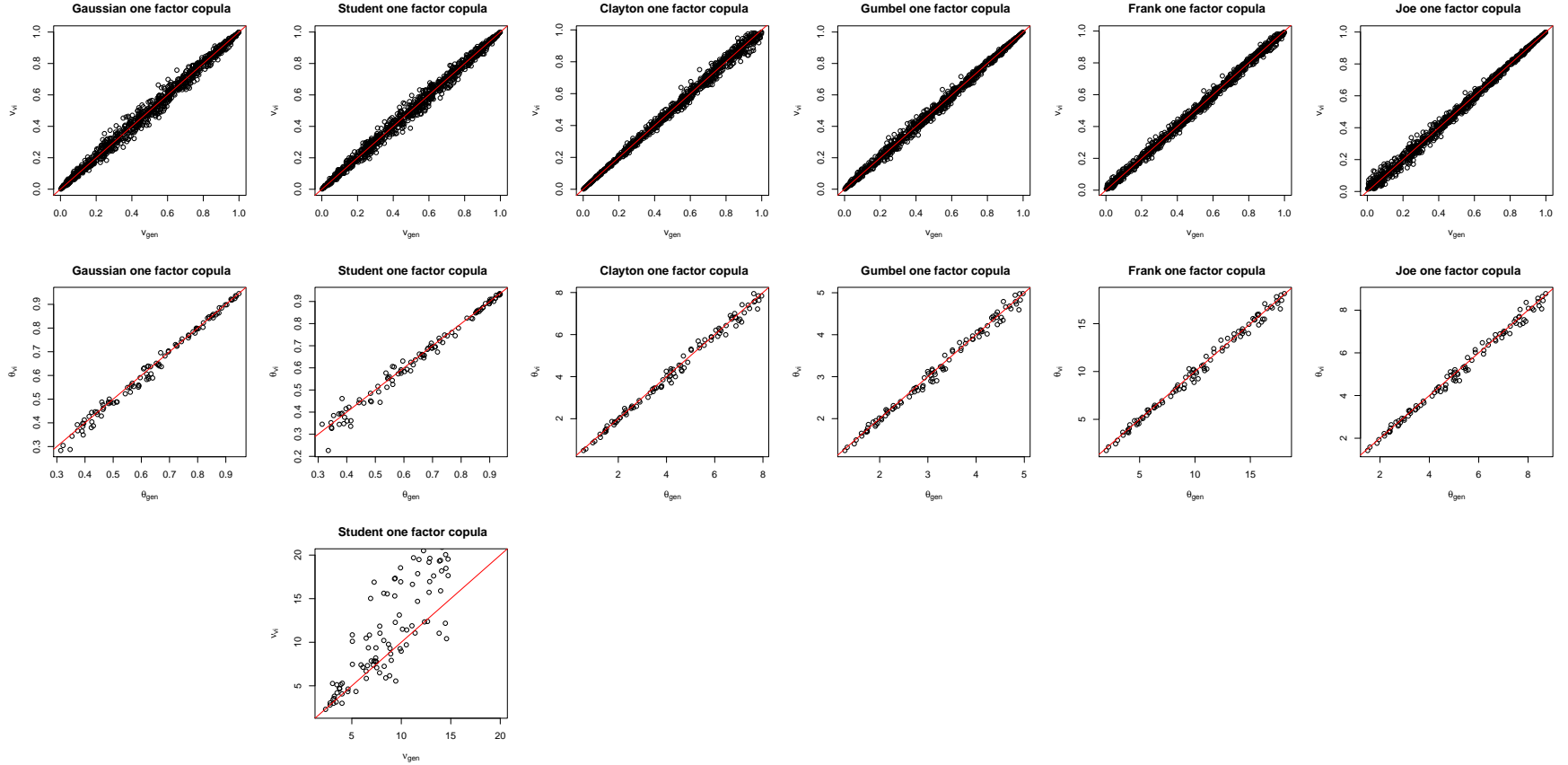


Figure 6: Variational inference for the one-factor copula models.

The figure compares the posterior means using variational approximation to the true generated values of the one factor copula models. In general, the posterior means are close to its true generated values. The posterior means of the degree of freedom in the Student- $t$  factor copula are less reliable when  $\nu > 10$  due to the flat likelihood.



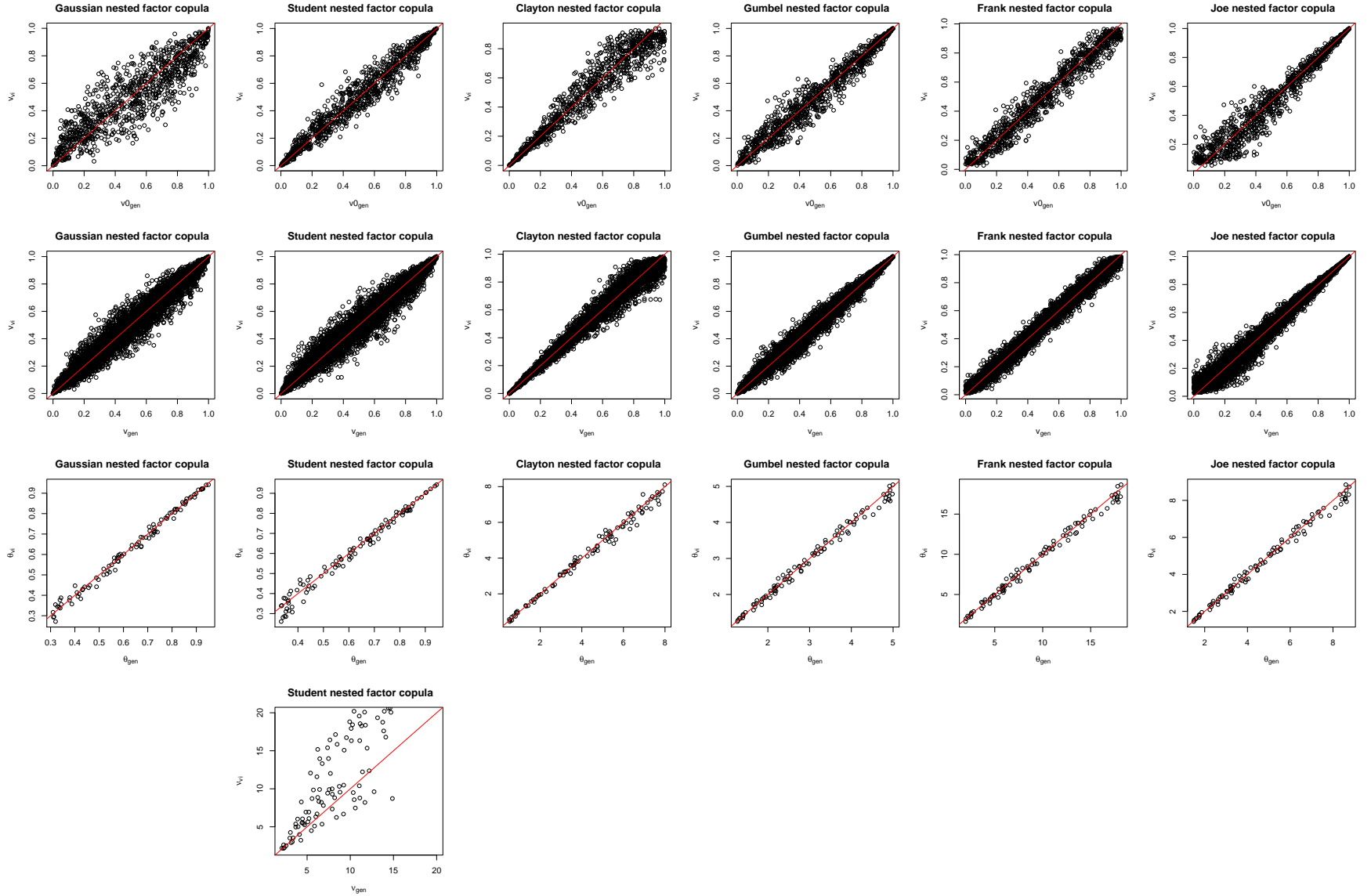


Figure 7: Variational inference for the nested factor copula models.

The figure compares the posterior means using variational approximation to the true generated values for the nested factor copula models. The common factor  $v_0$  fluctuates significantly around the true mean due to a small number of group latent variables.

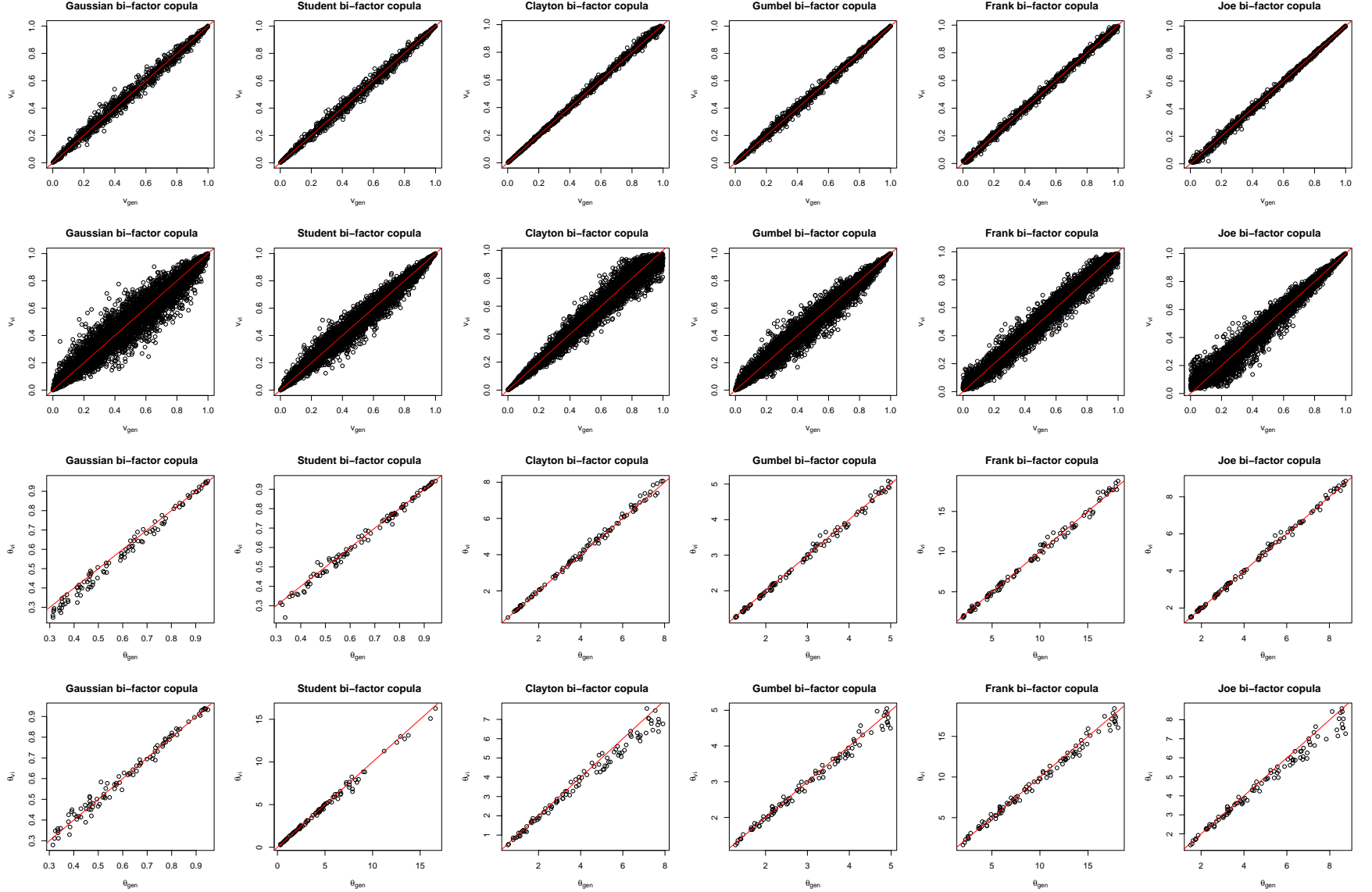


Figure 8: Variational inference for the bi-factor copula models.

The figure compares the posterior means using variational approximation to the true generated values of the bi-factor copula models. For the bi-factor Student- $t$  copula, the bivariate copulas in the second tree are mixed of other copula families due to the identification issue.

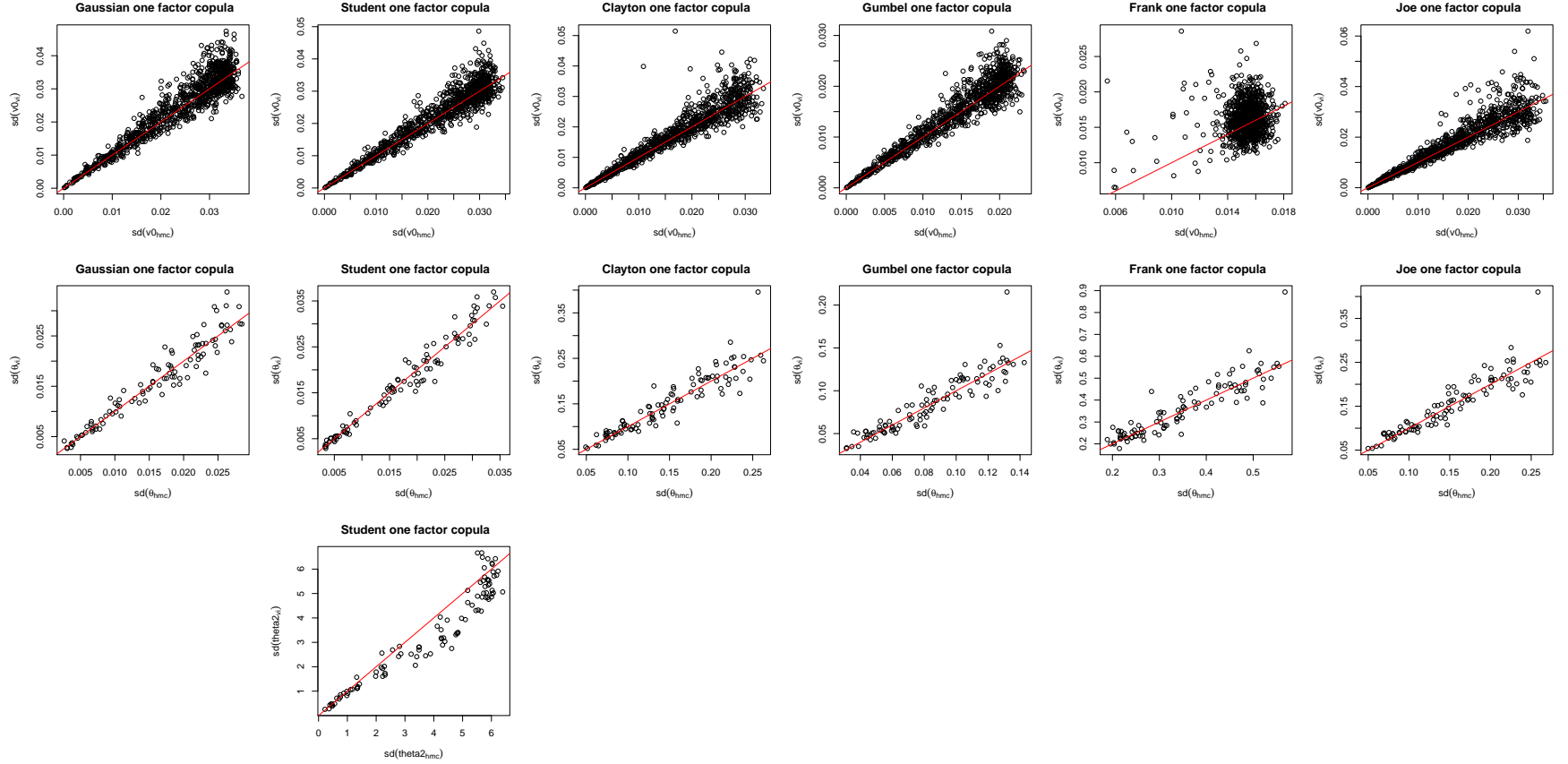


Figure 9: Comparison the standard deviations of VI and NUTS estimation for the one-factor copula models.

The figure compares the standard deviations of VI and NUTS estimation for one-factor copula models. In the one-factor models, the standard deviations of the parameters  $\theta$  are similar using both methods. The VI estimate of the degree of freedom in the factor Student- $t$  copula has a slightly lower variance than that of the MCMC approach.

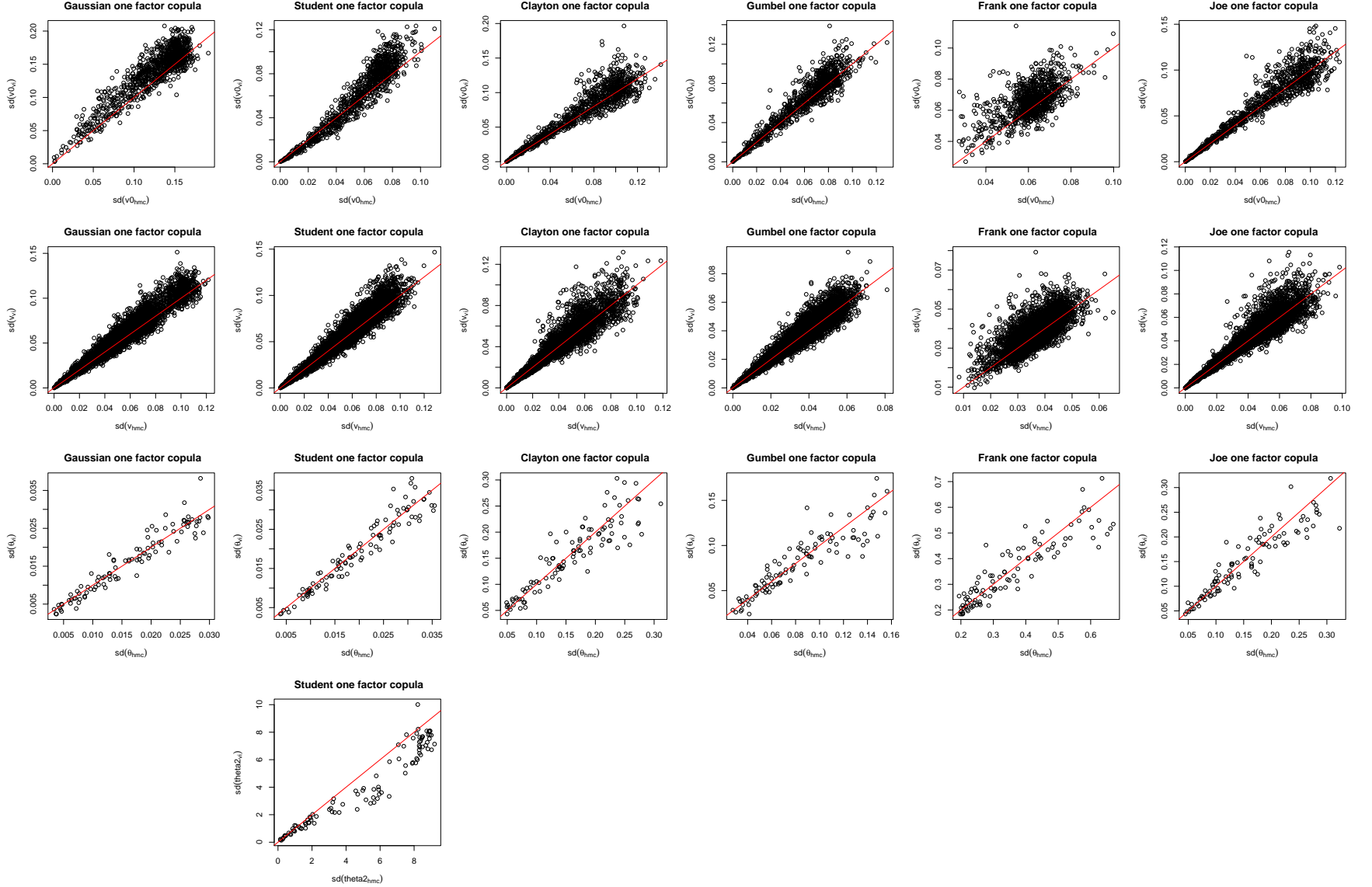


Figure 10: Comparison the standard deviations of VI and NUTS estimation for the nested factor copula models.

The figure compares the standard deviation of VI and NUTS estimation for nested-factor copula models. In the nested-factor model, the standard deviations of the parameters  $\theta$  are similar using both methods. The VI estimate of the degree of freedom in the factor Student- $t$  copula has a slightly lower variance than that of the MCMC approach.

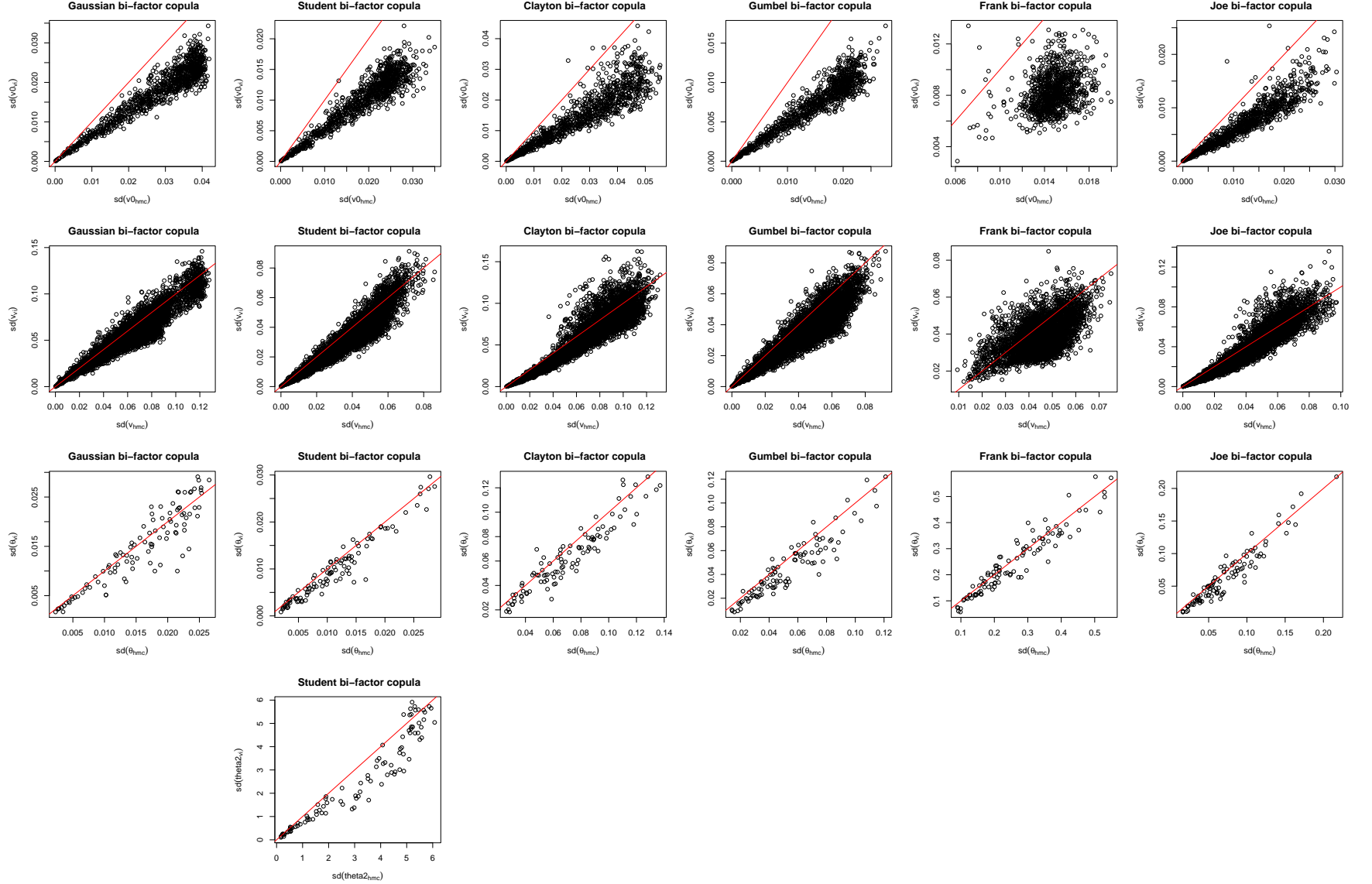


Figure 11: Comparison the standard deviations of VI and NUTS estimation for the bi-factor copula models.

The figure compares the standard deviations of VI and NUTS estimation for bi-factor copula models. In the bi-factor model, the standard deviations of the parameters  $\theta$  and  $v$  are lower than that of the MCMC approach. It is acceptable because we are more interested in the copula parameters  $\theta$ .

## C Empirical illustration

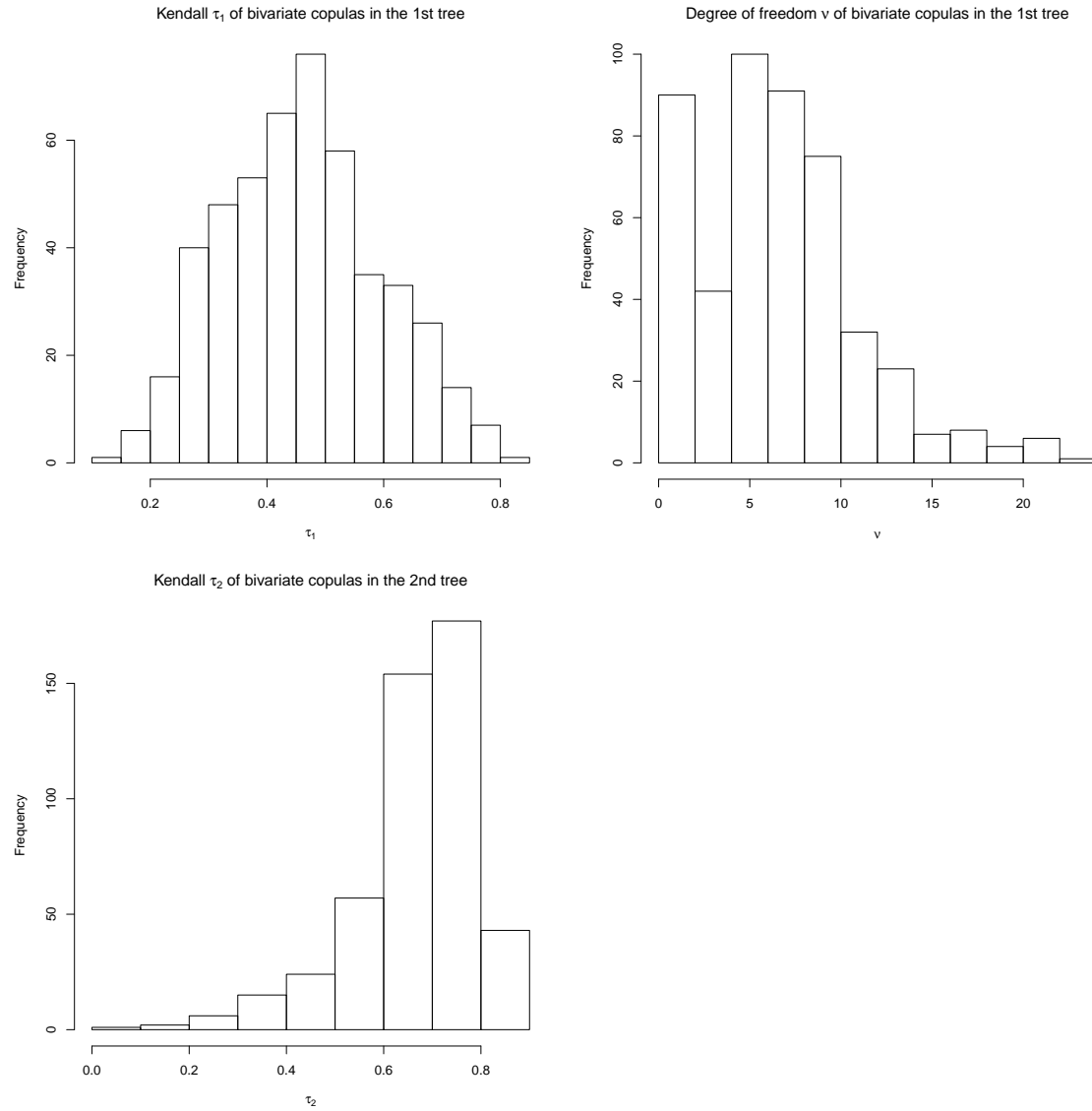


Figure 12: Histogram of the Kendall- $\tau$  correlation and degree of freedom  $\nu$  of bivariate copulas in temperature data.

The figure shows the histogram of the Kendall- $\tau$  correlation in the bi-factor model estimated at the posterior means. The first tree layer is on left top corner and the second tree is on left bottom corner. The histogram of degree of freedom  $\nu$  is on top right corner. Note that,  $\nu = 0$  represents for non Student- $t$  bivariate copulas.

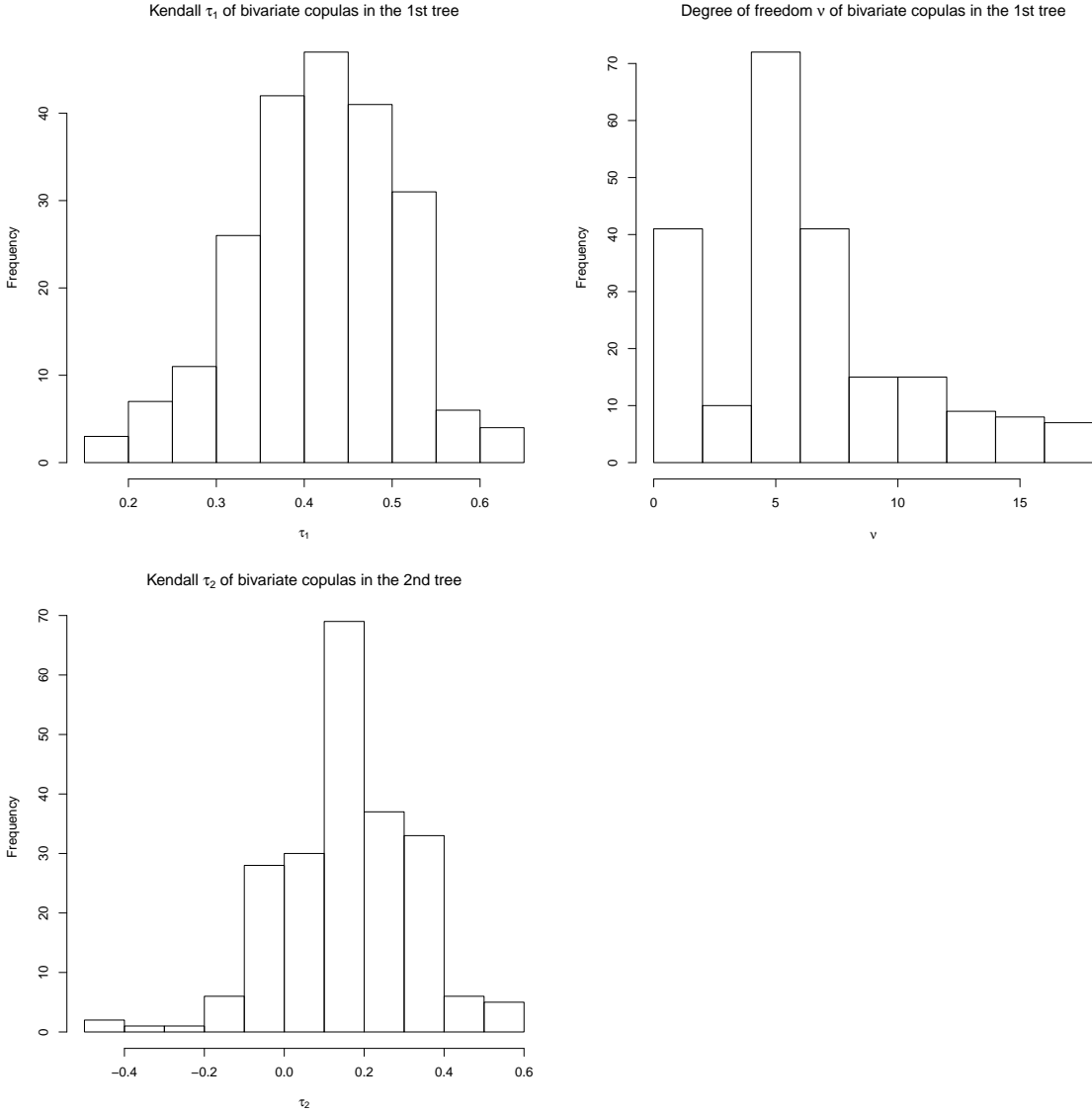


Figure 13: Histogram of the Kendall- $\tau$  correlation and degree of freedom  $\nu$  of bivariate copulas in stock return data.

The figure shows the histogram of the Kendall- $\tau$  correlation in the bi-factor model estimated at the posterior means. The first tree layer is on left top corner and the second tree is on left bottom corner. The histogram of degree of freedom  $\nu$  is on top right corner. Note that,  $\nu = 0$  represents for non Student- $t$  bivariate copulas.

## References

- K. Aas, C. Czado, A. Frigessi, and H. Bakken. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182–198, 2009.
- T. Bedford and R. M. Cooke. Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, 32(1-4):245–268, 2001.
- T. Bedford and R. M. Cooke. Vines: A new graphical model for dependent random variables. *Annals of Statistics*, pages 1031–1068, 2002.
- D. M. Blei, A. Kucukelbir, and J. D. McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877, 2017.
- E. C. Brechmann, C. Czado, and K. Aas. Truncated regular vines in high dimensions with application to financial data. *Canadian Journal of Statistics*, 40(1):68–85, 2012.
- H. Chen, R. MacMinn, and T. Sun. Multi-population mortality models: A factor copula approach. *Insurance: Mathematics and Economics*, 63:135–146, 2015.
- X. Chen and Y. Fan. Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *Journal of Econometrics*, 135(1-2):125–154, 2006.
- D. D. Creal and R. S. Tsay. High dimensional dynamic stochastic copula models. *Journal of Econometrics*, 189(2):335 – 345, 2015. *Frontiers in Time Series and Financial Econometrics*.
- T. M. Erhardt, C. Czado, and U. Schepsmeier. R-vine models for spatial time series with an application to daily mean temperature. *Biometrics*, 71(2):323–332, 2015.
- C. Fernández and M. F. Steel. On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93(441):359–371, 1998.
- I. H. Haff, K. Aas, and A. Frigessi. On the simplified pair-copula construction—simply useful or too simplistic? *Journal of Multivariate Analysis*, 101(5):1296–1310, 2010.



- M. D. Hoffman and A. Gelman. The No-U-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15(1):1593–1623, 2014.
- J. Hull and A. White. Valuation of a CDO and an n-th to default CDS without Monte Carlo simulation. *The Journal of Derivatives*, 12(2):8–23, 2004.
- H. Joe. Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, 94(2):401–419, 2005.
- H. Joe. *Dependence modeling with copulas*. CRC Press, 2014.
- P. Krupskii and M. G. Genton. Factor copula models for data with spatio-temporal dependence. *Spatial Statistics*, 22:180–195, 2017.
- P. Krupskii and H. Joe. Factor copula models for multivariate data. *Journal of Multivariate Analysis*, 120:85 – 101, 2013.
- P. Krupskii and H. Joe. Structured factor copula models: Theory, inference and computation. *Journal of Multivariate Analysis*, 138:53–73, 2015a.
- P. Krupskii and H. Joe. Tail-weighted measures of dependence. *Journal of Applied Statistics*, 42(3):614–629, 2015b.
- P. Krupskii, R. Huser, and M. G. Genton. Factor copula models for replicated spatial data. *Journal of the American Statistical Association*, 0(ja):0–0, 2016.
- A. Kucukelbir, D. Tran, R. Ranganath, A. Gelman, and D. M. Blei. Automatic differentiation variational inference. *The Journal of Machine Learning Research*, 18(1):430–474, 2017.
- D. Lee and H. Joe. Multivariate extreme value copulas with factor and tree dependence structures. *Extremes*, pages 1–30, 2017.
- R. Loaiza-Maya and M. S. Smith. Variational bayes estimation of time series copulas for multivariate ordinal and mixed data. *arXiv preprint arXiv:1712.09150*, 2018.
- J. S. Murray, D. B. Dunson, L. Carin, and J. E. Lucas. Bayesian Gaussian copula factor models for mixed data. *Journal of the American Statistical Association*, 108(502):656–665, 2013.

- F. Murtagh and P. Contreras. Methods of hierarchical clustering. *arXiv preprint arXiv:1105.0121*, 2011.
- D. B. Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, pages 347–370, 1991.
- H. Nguyen, M. C. Ausín, and P. Galeano. Parallel bayesian inference for high dimensional dynamic factor copulas. *UC3M Working Papers Statistics and Econometrics*, WP17-07, 2017.
- A. K. Nikoloulopoulos and H. Joe. Factor copula models for item response data. *Psychometrika*, 80(1):126–150, 2015.
- D. H. Oh and A. J. Patton. Modelling dependence in high dimensions with factor copulas. *Journal of Business & Economic Statistics*, 35(1), 2017a.
- D. H. Oh and A. J. Patton. Time-varying systemic risk: Evidence from a dynamic copula model of CDS spreads. *Journal of Business & Economic Statistics*, 2017b. doi: 10.1080/07350015.2016.1177535.
- R. Ranganath, S. Gerrish, and D. M. Blei. Black box variational inference. In *AISTATS*, pages 814–822, 2014.
- RStan package. RStan: the R interface to Stan, 2018. URL <http://mc-stan.org/>. R package version 2.17.3.
- F. J. Ruiz, M. K. Titsias, and D. M. Blei. Overdispersed black-box variational inference. *arXiv preprint arXiv:1603.01140*, 2016.
- B. Schamberger, L. F. Gruber, and C. Czado. Bayesian inference for latent factor copulas and application to financial risk forecasting. *Econometrics*, 5(2):21, 2017.
- U. Schepsmeier and J. Stöber. Derivatives and fisher information of bivariate copulas. *Statistical Papers*, 55(2):525–542, 2014.
- M. Sklar. *Fonctions de répartition à n dimensions et leurs marges*. Université Paris 8, 1959.
- B. K. Tan, A. Panagiotelis, and G. Athanasopoulos. Bayesian inference for the one-factor copula model. *Journal of Computational and Graphical Statistics*, (just-accepted), 2018.

- T. Tieleman and G. Hinton. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning*, 4(2):26–31, 2012.
- C. Zhang, J. Butepage, H. Kjellstrom, and S. Mandt. Advances in variational inference. *arXiv preprint arXiv:1711.05597*, 2018.