Variational Inference for high dimensional structured factor copulas

Hoang Nguyen[(a,b)], M. Concepción Ausín[(b,c)] and Pedro Galeano[(b,c)]

[(a)] Örebro University

[(b)] Department of Statistics - Universidad Carlos III de Madrid, Spain

[(c)] UC3M-Santander Big Data Institute - Universidad Carlos III de Madrid

May 6, 2020

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. A Bayesian procedure is employed in order to make fast inferences for multi-factor and structured factor copulas. To deal with the high dimensional structure, a Variational Inference (VI) algorithm is applied 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 limited time. Another issue of factor copula models is that the bivariate copula functions connecting the variables are unknown in high dimensions. An automatic procedure is derived to recover the hidden dependence structure. By taking advantage of the posterior modes of the latent variables, the bivariate copula functions are selected by minimizing the Bayesian Information Criterion (BIC). 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. The proposed procedure is illustrated with two high dimensional real data sets.

Keywords: Factor copulas; 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. (2018), 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), and financial time series dependence in Krupskii and Joe (2013), Creal and Tsay (2015), Oh and Patton (2017a), Nguyen et al. (2019), and Krupskii and Joe (2020) among others. In comparison to the truncated vine copulas proposed by Brechmann et al. (2012), the factor copula model is an alternative copula model which provides parsimonious and interpretable economic meanings.

In this paper, we propose a Bayesian procedure to make inferences for the multi-factor and structured factor copulas proposed in Krupskii and Joe (2013, 2015a). 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 often much faster. 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 two main 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. (2019) use reversible jump MCMC to select the factor copula links during the sampling process. Among those, only the Tan et al. (2019) 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. (2019). VI aims to approximate the model posterior distribution by a simpler distribution such that their distance in terms 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), Ong et al. (2018), Smith et al. (2020), 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 slightly underestimated in the bi-factor copula model. However, we greatly benefit from the speed of calculation that VI could be often much faster. VI has also been used to estimate vine copulas with discrete margins in Loaiza-Maya and Smith (2019) and semi-parametric regression using copulas in Smith and Klein (2019). VI have been applied effectively for models with latent variables in Tan and Nott (2018).

The Bayesian approach can also help to recover the hidden structure in factor copula models which shows an advantage over the proposed maximum likelihood estimation. 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 the latent variables are fixed at their posterior modes. We assign the new dependence structure to the data, reestimate the factor model and perform the copula selection until the bivariate copula linkages remain the same. We generate simulation studies in different contexts to show that the procedures of bivariate copula selection could be very 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 204 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 $X = (X_1, \ldots, X_d)'$ be the d-dimensional continuous random variable that we want to model their joint dependence structure and let $F_1(x_1), \ldots, F_d(x_d)$ 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, $C(u_1, \ldots, u_d) = F(x_1, \ldots, x_d)$, where $u_i = F_i(x_i)$, for $i = 1, \ldots, d$. It is well-known that $U_i = F_i(X_i) \sim \mathbb{U}(0,1)$, for $i = 1, \ldots, d$. Hence, copula helps to separate the marginal distributions from the dependence structure. Several examples of bivariate copula functions and their characteristics 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. Brechmann et al. (2012) proposed a truncated vine copula model while Smith et al. (2010) specify a prior over the level of vine copula models.

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. Next, 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.

Table 1: Bivariate copula families and their characteristics

Copula	Notation	Copula distribution function	Prior	Range	Kendall's τ
Gaussian	Gp	$C_{Gp}(u,v;\theta) = \Phi_2(\Phi^{-1}(u), \Phi^{-1}(v); \theta)$	$\pi_{Gp}(\theta) = \pi_{Gn}(\theta) = \frac{2}{\pi}\frac{1}{\sqrt{1-\theta^2}}$	$\theta \in (0,1)$	$\frac{2}{\pi}\texttt{arcsin}(\theta)$
	Gn	$C_{Gn}(u,v;\theta) = \Phi_2(\Phi^{-1}(u), \Phi^{-1}(v); \theta)$		$\theta \in (-1,0)$	
Student-t	Tp	$C_{Tp}(u,v;\theta,\nu) = T_2(T_\nu^{-1}(u), T_\nu^{-1}(v); \theta, \nu)$	$\pi_{Tp}(\theta) = \pi_{Tn}(\theta) = \frac{2}{\pi}\frac{1}{\sqrt{1-\theta^2}}$	$\theta \in (0,1), \nu \in (2,30)$	$\frac{2}{\pi}\texttt{arcsin}(\theta)$
	Tn	$C_{Tn}(u,v;\theta,\nu) = T_2(T_\nu^{-1}(u), T_\nu^{-1}(v); \theta, \nu)$	$\pi_T(\nu) = Gamma(\nu; 1, 0.1)$	$\theta \in (-1,0), \nu \in (2,30)$	
Clayton	C	$C_C(u,v;\theta) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1}{\theta}}$	$\pi_C(\theta) = \pi_{C180}(\theta) = \frac{2}{(\theta+2)^2}$	$\theta \in (0,\infty)$	$\frac{\theta}{\theta+2}$
	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	$C_G(u,v;\theta) = \texttt{exp}\left[-\left\{(-\texttt{log}u)^\theta + (-\texttt{log}v)^\theta\right\}^{1/\theta}\right]$	$\pi_G(\theta) = \frac{1}{\theta^2}$	$\theta \in [1,\infty)$	$1 - \frac{1}{\theta}$
	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	$C_{Fp}(u,v;\theta) = -\frac{1}{\theta}\texttt{log}\left(1 - \frac{(1-e^{-\theta u})(1-e^{-\theta v})}{1-e^{-\theta}}\right)$	$\pi_F(\theta) = \frac{4}{\theta^2}(1 - B(\theta) + 2D_1(\theta))$	$\theta \in (0,\infty)$	$1 - \frac{4}{\theta}(1 - D_1(\theta))$
	F_n	$C_{Fn}(u,v;\theta) = -\frac{1}{\theta}\texttt{log}\left(1 - \frac{(1-e^{-\theta u})(1-e^{-\theta v})}{1-e^{-\theta}}\right)$	$\approx Cauchy(\theta; 0, 6)$	$\theta \in (-\infty,0)$	
Joe	J	$C_J(u,v;\theta) = 1 - \left\{(1-u)^\theta + (1-v)^\theta - (1-u)^\theta(1-v)^\theta\right\}^{1/\theta}$	$\pi_J(\theta) = \sum_{k=1}^{\infty} \frac{8(\theta(k-1)+2-1/k)}{(\theta k+2)^2(\theta(k-1)+2)^2}$	$\theta \in [1,\infty)$	$1 - 4\sum_{k=1}^{\infty} \frac{1}{k(\theta k+2)(\theta(k-1)+2)}$
	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}	$C_{J90}(u,v;\theta) = v - C_J(1-u, v; -\theta)$	$\pi_{J90}(\theta) = \sum_{k=1}^{\infty} \frac{8(-\theta(k-1)+2-1/k)}{(\theta k-2)^2(\theta(k-1)-2)^2}$	$\theta \in (-\infty,-1]$	$1 - 4\sum_{k=1}^{\infty} \frac{1}{k(\theta k-2)(\theta(k-1)-2)}$
	J_{270}	$C_{J270}(u,v;\theta) = u - C_J(u, 1-v; -\theta)$	$\approx \frac{2}{(\theta-2)^2}$		
BB1	$BB1$	$C_{BB1}(u,v;\theta,\delta) = \{1 + [(u^{-\theta}-1)^\delta + (v^{-\theta}-1)^\delta]^{1/\delta}\}^{-1/\theta}$	$\pi_{BB1}(\theta) = Gamma(\theta; 0.25, 0.25)$	$\theta \in (0,\infty)$	$1 - \frac{2}{\delta(\theta+2)}$
	$BB1_{180}$	$C_{BB1_{180}}(u,v;\theta,\delta) = 1 - u - v + C_{BB1}(1-u, 1-v; \theta, \delta)$	$\pi_{BB1}(\delta) = Gamma(\delta - 1; 0.25, 0.25)$	$\delta \in [1,\infty)$	
	$BB1_{90}$	$C_{BB1_{90}}(u,v;\theta,\delta) = v - C_{BB1}(1-u, v; -\theta, -\delta)$	$\pi_{BB1_{90}}(\theta) = Gamma(-\theta; 0.25, 0.25)$	$\theta \in (-\infty,0)$	$1 - \frac{2}{\delta(\theta-2)}$
	$BB1_{270}$	$C_{BB1_{270}}(u,v;\theta,\delta) = u - C_{BB1}(u, 1-v; -\theta, -\delta)$	$\pi_{BB1_{90}}(\delta) = Gamma(-\delta - 1; 0.25, 0.25)$	$\delta \in (-\infty,-1]$	

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

The table shows some common bivariate copula functions as well as their characteristics such as parameter ranges, and Kendall's τ correlation. We divide the symmetric copula functions into positive and negative Kendall's τ correlation copulas to prevent the identifiability issue of the factor copula models, see discussion in Section 2.4. The BB1 copula will become the Gumbel copula when $\theta \to 0$ and it becomes the Clayton copula when $\delta = 1$.

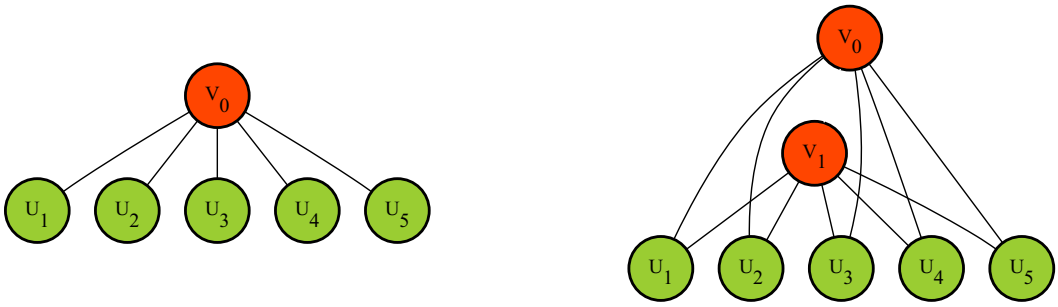

Figure 1: One-factor and two-factor copula models (Krupskii and Joe (2013))

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 $i = 1, \ldots, d$. Let us denote each bivariate copula density by $c_{U_i,V_0}(u_i, v_0; \boldsymbol{\theta_{0i}})$, each bivariate copula function by $C_{U_i,V_0}(u_i, v_0; \boldsymbol{\theta_{0i}})$, and each conditional distribution function of U_i given V_0 by $F_{U_i|V_0}(u_i|v_0; \boldsymbol{\theta_{0i}})$, where $\boldsymbol{\theta_{0i}}$ denotes the vector of of parameters of the bivariate copula. Then assuming that the variables $U_1, \ldots, U_d$ are conditionally independent given the latent variable V_0, as shown in Krupskii and Joe (2013), we have

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

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

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

where $\boldsymbol{\theta} = [\boldsymbol{\theta'_{01}}, \ldots, \boldsymbol{\theta'_{0d}}]'$ is the vector of copula parameters. Note that we have the conditional density $p(u_1, \ldots, u_d|v_0; \boldsymbol{\theta}) = c(u_1, \ldots, u_d|v_0; \boldsymbol{\theta})$ due to the uniform marginal of U_i, for $i = 1, \ldots, d$. 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 $\boldsymbol{\theta}$ through

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

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 $g = 1, \ldots, G$, 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 $i_g = 1, \ldots, d_g$. 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 $V = \{V_0, V_1, \ldots, V_G\}^{'}$ 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 $c_{U_{i_g},V_g}(u_{i_g}, v_g|\boldsymbol{\theta_{gi_g}})$ be the bivariate copula density of U_{i_g} and V_g and let $c_{V_g,V_0}(v_g, v_0|\boldsymbol{\theta_{0g}})$ be the bivariate copula density of V_g and V_0, for $g = 1, \ldots, G$, and $i_g = 1, \ldots, d_g$. Then, the conditional density function for the nested factor copulas is the following,

$$p(u_1, \ldots, u_d, v_1, \ldots, v_G|v_0; \boldsymbol{\theta}) = p(u_1, \ldots, u_d|v_1, \ldots, v_G; \boldsymbol{\theta}) p(v_1, \ldots, v_G|v_0; \boldsymbol{\theta}), \tag{2}$$

where $\boldsymbol{\theta} = [\boldsymbol{\theta}'_{\mathbf{01}}, \ldots, \boldsymbol{\theta}'_{\mathbf{0G}}, \boldsymbol{\theta}'_{\mathbf{11}}, \ldots, \boldsymbol{\theta}'_{\mathbf{Gd_G}}]'$ is the vector of copula parameters, and from Eq. (1), the conditional density for each layer is,

$$p(u_1, \ldots, u_d | v_1, \ldots, v_G; \boldsymbol{\theta}) = \prod_{g=1}^{G} \prod_{i_g=1}^{d_g} c_{U_{i_g}, V_g}(u_{i_g}, v_g | \boldsymbol{\theta_{g i_g}}),$$

and

$$p(v_1, \ldots, v_G | v_0; \boldsymbol{\theta}) = \prod_{g=1}^{G} c_{V_g, V_0}(v_g, v_0 | \boldsymbol{\theta_{0g}}),$$

respectively.

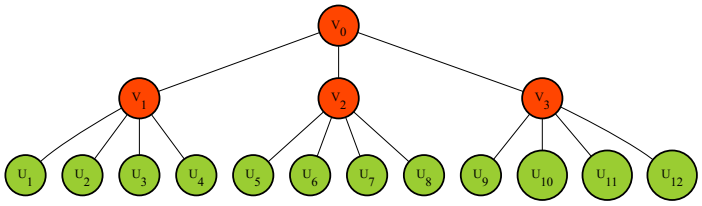

Figure 2: Nested factor copulas with $d = 12$ and $G = 3$ (Krupskii and Joe (2015a))

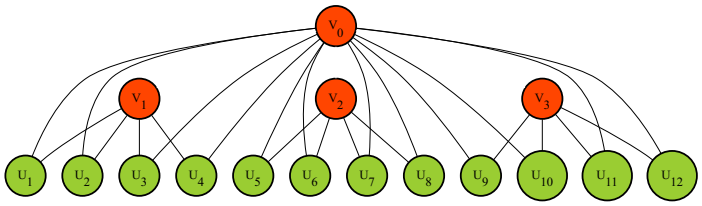

Figure 3: Bi-factor copulas with $d = 12$ and $G = 3$ (Krupskii and Joe (2015a))

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 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 $i_g = 1, \ldots, d_g$. Krupskii and Joe (2015a) assume that $V_0, \ldots, V_G$ 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 $c_{U_{i_g},V_0}(u_{i_g}, v_0|\boldsymbol{\theta_{0i_g}})$ be the bivariate copula density of U_{i_g} and V_0, and let $F_{U_{i_g}|V_0}(u_{i_g}|v_0, \boldsymbol{\theta_{0i_g}})$ be the conditional distribution of U_{i_g} given V_0, for $g = 1, \ldots, G$, and $i_g = 1, \ldots, d_g$. It is straightforward to show that the conditional variable $U_{i_g|v_0} = F(U_{i_g}|V_0, \boldsymbol{\theta_{0i_g}}) \sim \mathbb{U}(0, 1)$. Additionally, let $c_{U_{i_g},V_g|V_0}(u_{i_g|v_0}, v_g|\boldsymbol{\theta_{gi_g}})$ 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,

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

where $\boldsymbol{\theta} = [\boldsymbol{\theta'_{01}}, \ldots, \boldsymbol{\theta'_{0d}}, \boldsymbol{\theta'_{11}}, \ldots, \boldsymbol{\theta'_{Gd_G}}]'$ is the vector of copula parameters.

The bi-factor copula model requires $2d$ bivariate copula links and the computational expensive transformation $u_{i_g|v_0} = F_{U_{i_g}|V_0}(u_{i_g}|v_0, \boldsymbol{\theta_{0i_g}})$ 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 $c_{U_{i_g},V_g|V_0}(u_{i_g|v_0}, v_g)$ 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 regulated and they are all affected not only by a common economic latent factor but also by the country latent factors. Or one can assume that stocks are more dependent among industries, then divide them into different group sectors. The structured factor model could also handle hundreds of variables and capture well different behaviours in the upper and lower tails. The number of latent factors

have been discussed in Oh and Patton (2017a) who use the "scree plot" for selecting the number of factors. Our analysis is based on Krupskii and Joe (2013, 2015a) where the number of factors and the structured factor are set based on the prior assumption of hierarchical models and the model AIC and/or BIC are used to select which factor copula model is preferred. Even though, 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 a 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 correlated Gaussian copula and a positive correlated Frank copula are in the first group, while a negative correlated Gaussian copula and a negative correlated Frank copula 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. Vine structures with linking copulas with two or more parameters are harder to estimate than those with a single parameter because the likelihood has multiple modes and so will have, in general, the posterior distribution. 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 and the BB1 copula in the higher tree levels 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 τ 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 degrees of freedom parameter is high. We also note that the BB1 copula with parameters (θ, δ) will become the Gumbel copula

when $\theta \to 0$ and it becomes the Clayton copula when $\delta = 1$. 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 bivariate copulas.

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 VI 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 terms 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 $\boldsymbol{v_t} = [v_{t0}, \ldots, v_{tG}]^{'}$ for $t = 1, \ldots, T$ where T is the number of the observations. To account 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 τ follows the uniform $\mathbb{U}(0,1)$ and for negative bivariate copula τ follows the uniform $\mathbb{U}(-1,0)$. Then, we calculate the induced prior distributions for the one-parameter copulas using the transformation of the probability density functions. Table 1 shows the induced prior distributions for the bivariate copula parameters. For the Student-t copula, we let the degree of freedom follows the $Gamma(1, 0.1)$. Also for the BB1 copula, we assume the two parameters independently follow the Gamma and the shifted Gamma distributions so that the Kendall-τ of the BB1 copula is flat between $(0.2, 0.9)$.

In order to simplify the notations, we consider both latent variables and model copula parameters as the parameters of factor copula models as $\Theta = [\boldsymbol{v_1'}, \ldots, \boldsymbol{v_T'}, \boldsymbol{\theta}]^{'}$. Then, the prior density is $\pi(\Theta) = \prod_{t=1}^{T} \pi(\boldsymbol{v_t})\pi(\boldsymbol{\theta})$ where $\pi(\boldsymbol{\theta})$ is the product of independent prior densities of copula parameters in $\boldsymbol{\theta}$. Note that, we use the notation θ for a general parameter of a bivariate copula linkage and $\boldsymbol{\theta}$ for the vector of copula parameters in the factor copula models.

3.2 Posterior distributions

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

$$p(\mathcal{U}|\Theta) = \prod_{t=1}^{T} p(u_{t1}, \ldots, u_{td}|\boldsymbol{v_t}; \boldsymbol{\theta}).$$

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

$$\pi(\Theta|\mathcal{U}) \propto p(\mathcal{U}, \Theta) = \prod_{t=1}^{T} p(u_{t1}, \ldots, u_{td}|\boldsymbol{v_t}; \boldsymbol{\theta}) \prod_{t=1}^{T} \pi(\boldsymbol{v_t})\pi(\boldsymbol{\theta}). \tag{4}$$

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

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

Similarly, the posterior distribution of Θ for the nested factor copula is given by,

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

while for the bi-factor copula model is,

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

The bi-factor copula model requires the conditional transformation $u_{ti_g|v_0} = F(u_{ti_g}|v_{t0})$ whenever we generate new samples of $\boldsymbol{\theta_{0i}}$ 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, see Hoffman and Gelman (2014). 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 Θ, and let N be the total number of parameters in Θ. We assume that the posterior density is approximated by a proposal density $q(\Theta;\lambda)$, parameterized by a vector λ such that $q(\Theta;\lambda)$ is close to the joint posterior $p(\Theta|\mathcal{U})$. In order to find the values of the free parameters in λ, the VI approach tries to minimize the Kullback-Leibler divergence from the posterior $p(\Theta|\mathcal{U})$ to the proposal $q(\Theta;\lambda)$, i.e.:

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

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

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. (2019). Note that, the posterior $p(\Theta|\mathcal{U})$ 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,

$$\begin{aligned} \texttt{ELBO}(q) &= \mathbb{E}_q[\texttt{log}p(\mathcal{U},\Theta)] - \mathbb{E}_q[\texttt{log}q(\Theta;\lambda)] \\ &= \texttt{log}p(\mathcal{U}) - KL(q(\Theta;\lambda)||p(\Theta|\mathcal{U})) \leq \texttt{log}p(\mathcal{U}). \end{aligned} \tag{5}$$

such that when $q(\Theta;\lambda) = p(\Theta|\mathcal{U})$, we have $\texttt{ELBO} = \texttt{log}p(\mathcal{U})$. We obtain $p(\mathcal{U},\Theta)$ 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 $j = 1,\ldots,N$. Then we assume a product of univariate Gaussian densities as the proposal density,

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

where $\mu = \{\mu_j\}$ and $\sigma = \{\sigma_j\}$ for $j = 1,\ldots,N$. This factorized distribution is also called a Gaussian mean field distribution. Therefore, the computational cost is less than that corresponding to assuming that $\tilde{\Theta}$ is approximated using a multivariate Gaussian distribution with a full covariance matrix. Wainwright and Jordan (2008) show that the optimization problem will become more nonconvex if we allow for more dependences in the variational distribution. Kucukelbir et al. (2017) propose several transformation functions $\mathbb{T}$ to suit with different restrictions in the domain of Θ, see Table 2. Let $\omega = \{\omega_j\} = \texttt{log}(\sigma)$ to relax the positive constraint of σ, the variational parameters become $\lambda = \{\mu_j,\omega_j\}$, for $j = 1,\ldots,N$. The joint density $p(\mathcal{U},\tilde{\Theta})$ is derived based on the Jacobian of the inverse transformation, $J_{\mathbb{T}^{-1}}(\tilde{\Theta})$,

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

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,

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

where $\mathbb{E}_{q(\tilde{\Theta})}[-\texttt{log}q(\tilde{\Theta};\lambda)] = \sum_{j=1}^{N} 0.5\texttt{log}(2\pi e(\exp\omega_j)^2)$ is the closed-form formula of the entropy of multivariate Gaussian distribution and e is the base of the natural logarithm. For each factor copula model, the objective function ELBO corresponds to the joint posterior density up to a normalized constant $p(\mathcal{U}, \Theta)$. 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,

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

We use the stochastic gradient ascent to maximize the ELBO over $\lambda = \{\mu, \omega\}$. The gradients of ELBO with respect to parameters λ of the proposal distribution is denoted as ∇_λELBO. 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 λ follow the chain rule when we apply the transformations,

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

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 $\nabla_\Theta\texttt{log}p(\mathcal{U}, \Theta_{(m)})$ and ∇_λELBO. Then, the free variational parameter λ is updated along the gradient, $\lambda \leftarrow \lambda + \varrho\nabla_\lambda$ELBO with a vector step size ϱ. The vector of step size sequence or the learning rate ϱ 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.

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

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

Transformation function for different restrictions in constrain parameter space of Θ. The constrain domains of the copula families are showed in Table 1.

Algorithm 1 outlines the procedure of `ELBO` optimization. Starting with an initial copula structure 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` with respect to the variational parameter λ 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 Θ 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 Θ 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 $\bar{\boldsymbol{v}}$, to inspect the assumption of the initial

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

Result: Bivariate copula links, and samples of Θ 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 ∇_μ`ELBO` and ∇_ω`ELBO` ;

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

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

    Incremental iteration (i) ;

  end

  Select bivariate copula links between u and $\bar{\boldsymbol{v}}$ 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 Θ samples ;

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

links. The idea is to find the bivariate copula functions that minimize the BIC of bivariate copulas between $\mathcal{U}$ and $\bar{\boldsymbol{v}}$. In particular, we start with a random initial structure and obtain the posterior modes $\bar{\boldsymbol{v}}$. Then, we reassess the agreement of bivariate copulas between u_i and $\bar{\boldsymbol{v}}$. 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 $c_{U_i,V_0}(u_i, v_0; \boldsymbol{\theta_{0i}})$ be the current copula density of the link between variable U_i and V_0. The BIC for this bivariate link is,

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

where $\bar{\boldsymbol{v}}_0 = \{\bar{v}_{10}, \ldots, \bar{v}_{T0}\}$ is the vector of the posterior modes of the latent variables, $\hat{\boldsymbol{\theta}}_{0i}$ 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,

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

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

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

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

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

If there are any changes in the bivariate copula links ($c^{(*)}_{U_i,V_0} \neq c_{U_i,V_0}$ for any $i = 1,\ldots,d$), we update the copula structure and estimate the new model. The $\texttt{BIC}^{(*)}$ of the reestimated model will be at least as good as the $\underline{\texttt{BIC}}^{(*)}$. This guarantees that the model goodness of fit increases. We repeat the selection step until there is no change in the bivariate copula selection. Accordingly, the posterior modes of the latent variables help to find out the most appropriate candidate for each bivariate link. 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 $u_{i_g|v_0}$ 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 Monte Carlo study, 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 $i = 1, \ldots, d$. 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 all BB1 copulas. We call a mix of Gaussian, Student-t, Clayton, Gumbel, Frank, Joe, BB1 copula linkages as the Mix copulas. The experiments for the rotated Archimedean copulas have the similar results. The Kendall's τ correlation of each bivariate copula link is randomized uniformly in the range $[0.2, 0.8]$ and the tail dependence of the BB1 copula is restricted such that the lower and upper tail dependence coefficients are the same.

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 calculate the average of model selection criteria for the one-factor, the nested factor and the bi-factor copulas using 100 Monte Carlo simulations. 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 C a comparison between the posterior means using VI and the true generated values, see Figure 5. 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-t copula is quite flat when ν is high which makes the posterior samples of ν right skewed. Hence, 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 average of the model selection criteria for the one-factor Gaussian, Student-t, Clayton, Gumbel, Frank, Joe, BB1 and Mix copulas using 100 Monte Carlo simulations. The result of the first experiment is shown in the top and that of the second experiment is shown in the bottom of the table. 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 $\texttt{log}p(\mathcal{U}|\boldsymbol{\theta})$, see Appendix B. The value of ELBO, AIC, BIC, $\texttt{log}p(\mathcal{U}|\boldsymbol{\theta})$ are normalized for 1000

data observations. It usually took less than a minute to estimate the parameters of the one-factor models assuming the true generated structure, 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. The selection procedure is robust to the initial choice. It takes a few iterations of copula selection to reach a steady state where the bivariate links remains unchanged. One can make the selection procedure faster by assuming all bivariate Gaussian copulas as an initial choice. Alternatively, Krupskii and Joe (2015a) suggests the choice of linking copulas based on the measure of dependence in the tail. The accuracy rate is quite high, except for the Clayton, Joe and BB1 copulas. Bivariate copula selection often goes wrong between the Clayton copula and the survival Joe copula when the Kendal's τ is high and vice versa. Due to that reason, we restrict the maximum number of selection iterations under 10. Also for the BB1 copula, if the parameter θ is small, the procedure usually selects the Gumbel copula and the Clayton copula is chosen if the parameter δ is small. After all, we still archive similar values of the 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 22 copula types and select the best functions. It is about one second for each bivariate copula selection, and one can parallel the selection procedure to speed up the calculation.

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 6 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 average of model selection criteria for the nested factor copulas using 100 Monte Carlo simulations. In the second experiment with a random initial structure, it

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

Copula type	Gaussian	Student-t	Clayton	Gumbel	Frank	Joe	BB1	Mix
(a) Estimated models assuming the true generated bivariate copulas								
ELBO	31.2	33.0	75.4	67.8	56.6	77.0	41.8	54.1
AIC	-63.0	-66.3	-146.8	-134.8	-114.2	-150.0	-83.3	-107.5
BIC	-62.5	-65.3	-146.3	-134.3	-113.7	-149.5	-82.3	-106.9
$\log p(\mathcal{U}\|\boldsymbol{\theta})$	31.6	33.4	73.5	67.5	57.2	75.1	41.8	53.9
(b) Estimated models assuming the selected bivariate copulas using Algorithm 1								
# Selection iteration	3	7	10	3	3	10	9	8
% accuracy	98	69	68	98	99	61	63	80
ELBO	31.2	33.0	75.4	67.8	56.6	77.0	41.7	54.1
AIC	-63.0	-66.3	-146.8	-134.8	-114.2	-150.1	-83.3	-107.5
BIC	-62.5	-65.4	-146.3	-134.3	-113.7	-149.6	-82.4	-106.9
$\log p(\mathcal{U}\|\boldsymbol{\theta})$	31.6	33.3	73.5	67.5	57.2	75.1	41.8	53.9

We report the average model selection criteria of 100 Monte Carlo simulations for the one-factor copula models. Each simulated dataset contains 100 bivariate links with about 100 to 200 copula parameters. We use Gauss-Legendre quadrature integration over the latent space to obtain $\log p(\mathcal{U}|\boldsymbol{\theta})$, see Appendix B. The value of ELBO, AIC, BIC, $\log p(\mathcal{U}|\boldsymbol{\theta})$ are normalized for 1000 data observations.

also took a few iterations to make the choice of bivariate copulas unchanged. Again, the similarity of Clayton and survival Joe copula reduces the number of correctly specified links, and the BB1 copula is misspecified to the Clayton and the Gumbel copulas as shown in the bottom table. 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 75%. The time of inference is still quite comparable with the one-factor model, even in this case where we have much more latent parameters. More details are shown in Section 4.4.

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 identifiability issue of the bi-factor Student-t copulas and the bi-factor BB1 copulas, the bivariate copulas in the second tree are a combination of one parameter copula families, see the discussion in Section 2.4. Figure 7 in Appendix C shows the posterior means using the 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 the observable variables is used to make inferences on the common latent variable v_0, hence we obtain more certain estimation of v_0 than the group latent variables v_g. The computational cost significantly increases

Table 4: Model comparison for the nested factor copula models

Copula type	Gaussian	Student-t	Clayton	Gumbel	Frank	Joe	BB1	Mix
(a) Estimated models assuming the true generated bivariate copulas								
ELBO	26.2	27.8	69.4	61.8	50.8	71.0	36.5	47.3
AIC	-53.3	-56.8	-137.5	-124.0	-103.7	-140.6	-73.5	-95.2
BIC	-52.8	-55.7	-137.0	-123.5	-103.2	-140.1	-72.5	-94.5
$\texttt{log}p(\mathcal{U}\|\boldsymbol{\theta})$	26.8	28.6	68.9	62.1	52.0	70.4	37.0	47.7
(b) Estimated models assuming the selected bivariate copulas using Algorithm 1								
# Selection iteration	5	6	9	4	4	10	7	8
% accuracy	96	65	65	96	97	53	58	78
ELBO	26.1	27.8	69.5	61.8	50.7	71.2	36.4	47.3
AIC	-53.3	-56.7	-137.6	-124.0	-103.5	-140.6	-73.5	-95.1
BIC	-52.7	-55.8	-137.0	-123.5	-103.0	-140.0	-72.5	-94.5
$\texttt{log}p(\mathcal{U}\|\boldsymbol{\theta})$	26.7	28.6	68.9	62.1	51.9	70.4	36.9	47.7

We report the average model selection criteria of 100 Monte Carlo simulations 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 $\texttt{log}p(\mathcal{U}|\boldsymbol{\theta})$, see Appendix B. The value of ELBO, AIC, BIC, $\texttt{log}p(\mathcal{U}|\boldsymbol{\theta})$ are normalized for 1000 data observations.

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 average of the model selection criteria for the bi-factor copula models using 100 Monte Carlo simulations. 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 model selection criteria are quite close when the true structure and the recovered structure are compared. Even in this case, the copula selections are implemented 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 perform the copula selection for the copula links between $u_{i_g|v_0}$ and v_g in the second layer.

4.4 Comparison between VI and MCMC estimation

We incorporate the NUTS algorithm using the RStan package (2018) and consider the posterior samples as the benchmark of the MCMC approach. For each experiment of the factor copula models proposed in the previous section, we generate posterior 1000 samples via the NUTS algorithm with 500 burn-in. The chain quickly converges after a few hundred iterations. Figure 8, 9 and 10 in Appendix C compare the standard deviations of the posterior samples using the VI and the MCMC

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

Copula type	Gaussian	Student-t	Clayton	Gumbel	Frank	Joe	BB1	Mix
(a) Estimated models assuming the true generated bivariate copulas								
ELBO	56.4	84.7	141.2	126.9	105.4	144.0	93.3	106.7
AIC	-115.6	-172.0	-276.1	-256.0	-215.5	-281.7	-188.4	-211.4
BIC	-114.6	-170.5	-275.2	-255.0	-214.6	-280.8	-186.9	-210.3
$\log p(\mathcal{U}\|\boldsymbol{\theta})$	58.0	86.3	138.3	128.2	108.0	141.1	94.5	105.9
(b) Estimated models assuming the selected bivariate copulas using Algorithm 1								
# Selection iteration	4	10	10	7	4	10	10	9
% accuracy Tree 1	98	67	84	98	99	46	60	80
% accuracy Tree 2	96	84	20	64	95	3	84	69
ELBO	56.4	84.5	132.9	124.9	105.4	135.2	93.2	104.7
AIC	-115.6	-171.6	-264.4	-252.4	-215.3	-269.0	-188.2	-208.6
BIC	-114.6	-170.2	-263.4	-251.4	-214.3	-268.0	-186.8	-207.5
$\log p(\mathcal{U}\|\boldsymbol{\theta})$	58.0	86.1	132.4	126.4	107.9	134.7	94.4	104.5

We report the average model selection criteria of 100 Monte Carlo simulations for the bi-factor copula models. Each factor copula model contains 6 latent factors, about 200 bivariate links with about 200 to 300 copula parameters. We use Gauss-Legendre quadrature integration over the latent space to obtain $\log p(\mathcal{U}|\boldsymbol{\theta})$, see Appendix B. The value of ELBO, AIC, BIC, $\log p(\mathcal{U}|\boldsymbol{\theta})$ are normalized for 1000 data observations.

for different factor copula models given known bivariate copula linkages. In the one-factor model and the nested factor model, the standard deviations of the model parameters are similar using both methods. In the bi-factor copula models, we clearly observe that VI only underestimates the standard deviations of the latent variables which is in agreement with the literature of VI estimation, see Blei et al. (2017). However, this underestimation is acceptable as we are more interested in the copula parameter $\boldsymbol{\theta}$.

Figure 4 shows the contour plots of the posterior samples of some selected parameters using VI and MCMC for the bi-factor copula models. In general, VI approximates quite well the posterior distribution of the latent variable and the bivariate copula parameters. The plots also show that the MCMC posterior samples are only weakly correlated among the parameters of different bivariate copulas but there might be some correlation between the parameters of two-parameter bivariate copulas. That is because the conditional posterior of the parameter θ_i given the latent variable depends on only its bivariate copula density and prior density. Also the conditional posterior of the latent variables are independent due to the model specification.

There are several proposals to overcome the limitation of the Gaussian mean-field assumption. For examples, Ong et al. (2018) propose a factor covariance structure which can be useful when the covariance matrix has no conditional independent structure. But, it can be difficult to choose

number of factors and the computational cost will increase with the number of factors. Alternatively, as the covariance matrix of the MCMC posterior samples are sparse, one can assume a Gaussian variational proposal distribution with a sparse precision matrix, see Tan and Nott (2018). This approach also requires an additional computation cost which depends on the number of the parameters used in the Cholesky factor of the precision matrix. There also needs a restriction of the sparsity based on the specific models. Last but not least, after optimizing with a mean field proposal, Giordano et al. (2015) apply the linear response methods to correct for a more accurate covariance matrix. As our models cope with high dimensional variables, a product of univariate Gaussian distributions can be appropriate as the proposal in the VI.

Table 6 compares the time of computation between VI and the MCMC for 1000 posterior 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, and 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: Estimation time (in seconds) for the factor copula models using VI and NUTS

Copula type	Gaussian	Student-t	Clayton	Gumbel	Frank	Joe	BB1	Mix
(a) Time estimated (in seconds) using VI								
One-factor	10	500	25	40	10	20	100	110
Nested factor	20	800	30	45	15	25	120	150
Bi-factor	90	3400	160	270	90	150	250	600
(b) Time estimated (in seconds) using NUTS								
One-factor	1100	63000	3700	5200	1100	2200	31000	16000
Nested factor	1700	86000	4500	6100	1300	3500	33000	38000
Bi-factor	17000	710000	84000	122000	61000	85000	95000	114000

We report the average estimation time for the factor copula models using the VI and the NUTS given a known dependence structure ($d = 100, T = 1000$ for the one-factor copula and $G = 5$ for the structured factor copulas). The VI convergence time depends on its optimization parameters such as the number of MC samples, the number of MC for calculating the gradients, tolerance, among others. The NUTS approach depends mainly on the number of iterations. The computational cost of 1000 posterior samples is round up based on 5 simulations for the NUTS approach and 100 simulations for the VI approach.

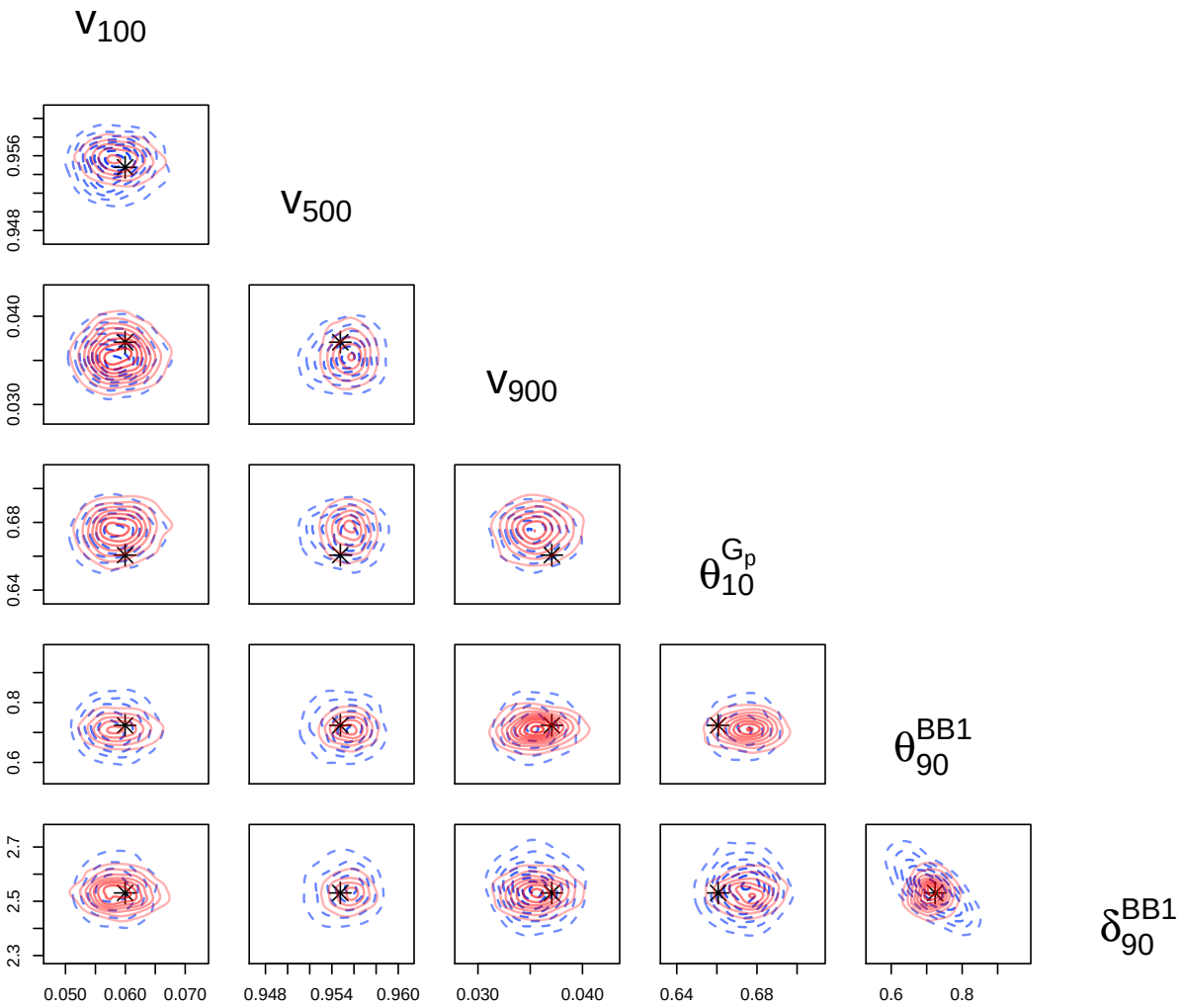

Figure 4: The contour plots of some selected parameters using the VI (red solid lines) and the MCMC (blue dashed lines) for the bi-factor copula models. The true values are marked as black stars. The selected parameters are the common latent variables at t = 100, 500, 900 and the copula parameters of the positive Gaussian and the BB1 copula that link the 10th and the 90th observable variables to the common latent variable. The MCMC posterior samples are only weakly correlated among the parameters of different bivariate copulas but there might be some correlation between the parameters of two-parameter bivariate copulas. In general, VI approximates quite well the posterior distribution for the bi-factor copula model.

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 204 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, $\mathcal{U} = (u_{t1}, \ldots, u_{td}) = (F_1(x_{t1}), \ldots, F_d(x_{td}))$. Then, we apply the selection procedure for the one-factor, the nested factor and the bi-factor copula models. The bi-factor model is preferred in both examples when we compare the model selection 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 at `https://www.dwd.de/EN/climate_environment/cdc/cdc_node.html`. 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 is at most 200 kilometers.

Erhardt et al. (2015) model the marginal distribution of temperatures using an 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 the ARMA(1,1) process, (ν_i, γ_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 the one-factor, the two-factor, the nested factor and the bi-factor copulas using the bivariate copula selection procedure, together with the two bi-factor copula models using two parameters bivariate copulas at the first level and Frank copulas at the second level. 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 the 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 two-factor copula models. We find that the bi-factor copula model is the most suitable model for the dependence of daily temperatures based on the model selection criteria.

Table 7: Model comparison for the daily temperature dataset

Structure	One-factor	Two-factor	Nested factor	Bi-factor		
				BB1-Frank	Student-Frank	Selection
AIC	-459.5	-612.9	-935.9	-944.7	-941.5	-962.5
BIC	-455.8	-607.4	-931.5	-938.1	-936.0	-956.2
$\log p(\mathcal{U}\|\boldsymbol{\theta})$	230.5	307.6	468.8	473.6	471.8	482.5
# bivariate links	479	903	503	958	958	957
# Gaussian (rotated)	0	146	26	0	0	358
# Student-t (rotated)	223	233	361	0	479	354
# Clayton (rotated)	0	5	0	0	0	1
# Gumbel (rotated)	150	203	18	0	0	140
# Frank (rotated)	0	239	0	479	479	46
# Joe (rotated)	0	0	0	0	0	1
# BB1 (rotated)	106	77	98	479	0	57
# Independence	0	55	0	0	0	1

We report the model comparison for daily temperature dataset of the one-factor, the two-factor, the nested factor and the bi-factor copulas using the bivariate copula selection procedure, together with the two bi-factor copula models using two parameters bivariate copulas at the first level and Frank copulas at the second level.

Next, we can take advantage of the factor copula models 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. For example, we sample the missing copula variable $u_{ti} \sim F(u_{ti}|v_{t0}, v_{tg})$ in the bi-factor copulas, and the predicted temperature x_{ti} using the marginal models as,

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

where $v_{t0}^{(n)}, v_{tg}^{(n)}$ are obtained from the posterior samples of the predicted latent variables using the VI approach. 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. Table 8 compares the mean absolute forecasting error (MAFE), the mean square forecasting error (MSFE), and the mean log predictive densities (LP) of the daily temperature at the 24 stations. In general, the structured factor copulas give better prediction in comparison to the one-factor and the two-factor copulas.

Table 8: Forecasting comparison for the daily temperature dataset

Structure	One-factor	Two-factor	Nested factor	Bi-factor		
				BB1-Frank	Student-Frank	Selection
MAFE	1.198	0.990	0.693	0.695	0.697	0.669
MSFE	2.643	1.888	1.102	1.084	1.092	1.047
LP	-1.976	-1.823	-1.454	-1.463	-1.475	-1.415

We report the average forecasting errors for daily temperature at the 24 stations using the one-factor, the two-factor, the nested factor and the bi-factor copulas. The bi-factor copula with the selected copula links using Algorithm 1 gives the best prediction.

5.2 Stock return dependence

We estimate the stock return dependence of 204 firms listed in 10 European countries as Austria (12 firms), Belgium (15 firms), Finland (18 firms), France (31 firms), Germany (28 firms), Ireland (10 firms), Italy (31 firms), Netherlands (18 firms), Portugal (11 firms), and Spain (30 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 (ν_i, γ_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 ϵ_{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 9 shows the summary statistics of the one-factor, the nested factor and the bi-factor copula models for stock return data. We divide firms into country groups and sector groups for model comparison. 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. By grouping the stock returns by 14 industries, the structured factor copula models can be improved in terms of minimizing AIC and BIC. The bi-factor model is still preferable according to the model selection 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.

Next, we compare the Spearman's ρ and the tail-weighted dependence measures of the factor copula models. Krupskii and Joe (2015b) propose the tail-weighted dependence as the correlation

Table 9: Model comparison of stock return dependence

Structure	One-factor	Two-factor	Nested factor		Bi-factor	
			by Country	by Industry	by Country	by Industry
AIC	-98.1	-113.0	-106.7	-113.6	-112.7	-119.4
BIC	-96.3	-110.2	-104.8	-111.8	-109.9	-116.7
$\texttt{log}p(\mathcal{U}\|\boldsymbol{\theta})$	49.4	57.1	53.7	57.2	56.9	60.2
# bivariate links	204	405	214	218	387	384
# Gaussian (rotated)	10	16	12	18	57	95
# Student-t (rotated)	155	144	151	153	156	158
# Clayton (rotated)	0	0	0	0	7	8
# Gumbel (rotated)	13	26	25	18	57	31
# Frank (rotated)	7	193	12	17	87	76
# Joe (rotated)	0	0	0	0	0	1
# BB1 (rotated)	19	26	14	12	20	12
# Independence	0	3	0	0	21	24
$\Delta\rho$	0.041	0.032	0.046	0.050	0.037	0.037
$\Delta\lambda_L$	0.059	0.085	0.082	0.086	0.057	0.063
$\Delta\lambda_U$	0.078	0.059	0.091	0.075	0.076	0.081

We report the model comparison for stock return dataset using the one-factor, the two-factor, the nested factor and the bi-factor copulas. Firms can be grouped by country or by sector due to the prior assumption. The bottom table shows the average of the absolute difference of Spearman's ρ and the tail-weighted dependence measures between the factor copula models and those of the empirical copula data.

of transformed variables where the joint movements in the tails have heavier weights,

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

The bottom of Table 9 shows the average of the absolute difference of Spearman's ρ and the tail-weighted dependence measures between the factor copula models and those of the empirical copula data. The results are quite similar among the factor copula models. The bi-factor copulas fit slightly better at the center of the distribution which results in a smaller difference of Spearman's ρ. However, it is suggested 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 MCMC posterior samples are only weakly correlated, hence a Gaussian mean field distribution can be an appropriate proposal for the VI. The 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. (2019) and Acar et al. (2019).

Acknowledgment

We thank Michael Wiper, Roberto Casarin, the Associate Editor and three anonymous referees for their helpful comments. The authors acknowledge financial support from the Spanish Ministry of Economy and Competitiveness, project numbers ECO2015-66593-P.

Appendix

A The step size

At iteration i, the gradient vector $g^{(i)} = \nabla_{\lambda^{(i)}} \texttt{ELBO}$ 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 $j = 1, \ldots, N$,

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

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

B Marginal likelihood calculation

We calculate the marginal likelihood of the factor copula models by integrating the augmented likelihood over the latent factor variable space and approximate the integral via the Gauss-Legendre quadrature. For example, in the one-factor copula model

$$\begin{aligned}
p(\mathcal{U}|\boldsymbol{\theta}) &= \prod_{t=1}^{T} c(u_{t1}, \ldots, u_{td}|\boldsymbol{\theta}) \\
&= \prod_{t=1}^{T} \int_0^1 \prod_{i=1}^{d} c_{U_i,V_0}(u_{ti}, v_{t0}|\boldsymbol{\theta_{0i}}) dv_{t0} \\
&= \prod_{t=1}^{T} \sum_{k=1}^{n_q} w_k \prod_{i=1}^{d} c_{U_i,V_0}(u_{ti}, \mathbf{x}_k|\boldsymbol{\theta_{0i}})
\end{aligned}$$

where $n_q = 25$ is the number of quadrature points, the set $\{\mathbf{x}_k\}$ is the nodes and the set $\{w_k\}$ is the quadrature weights. The choice of the approximation methods are discussed in Krupskii and Joe (2013, 2015a).

Then the model selection criteria of the factor copula models are valued at the posterior mean

of the copula parameters $\bar{\boldsymbol{\theta}}$

$$\begin{aligned}
\texttt{log}p(\mathcal{U}|\boldsymbol{\theta}) &= \log p(\mathcal{U}|\bar{\boldsymbol{\theta}}) \\
AIC &= -2\log p(\mathcal{U}|\bar{\boldsymbol{\theta}}) + 2K \\
BIC &= -2\log p(\mathcal{U}|\bar{\boldsymbol{\theta}}) + K\log(T)
\end{aligned}$$

where K is the number of the copula parameters $\bar{\boldsymbol{\theta}}$. The value of `ELBO` is calculated at the last value from the stochastic gradient algorithm. The posterior joint tail probability between variable U_i and U_j can also be derived by approximating the integral via the Gauss-Legendre quadrature,

$$\begin{aligned}
Pr(U_i < p, U_j < p|\mathcal{U}) &= \int_0^p \int_0^p p(u_i, u_j|\mathcal{U}) du_i du_j \\
&= \int_0^p \int_0^p \int_{\boldsymbol{\theta}\in R^n} p(u_i, u_j|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathcal{U}) d\boldsymbol{\theta} du_i du_j \\
&= \sum_{k_i=1}^{n_q} \sum_{k_j=1}^{n_q} \sum_{l=1}^{n_l} w_{k_i} w_{k_j} p(u_{k_i}, u_{k_j}|\boldsymbol{\theta}^{(\boldsymbol{l})})
\end{aligned}$$

where $n_q = 25$ is the number of quadrature points, the set $\{u_{k_i}, u_{k_j}\}$ is the nodes and the set $\{w_{k_i}, w_{k_j}\}$ is the quadrature weights, the $\boldsymbol{\theta}^{(\boldsymbol{l})}$ for $l = 1, \ldots, n_l$ are the posterior samples of the factor copula parameters.

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

Here, we show several comparisons between the posterior means using the VI and the true generated values for the one-factor, the nested factor and the bi-factor copula models, see Figure 5, 6, and 7. In general, the VI posterior means are close to their true generated values. The standard deviations of latent factors depend on the number of the series in the sector group.

Figure 8, 9 and 10 compare the standard deviations of the posterior samples using VI and MCMC for the one-factor, the nested factor and the 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 models.

In the bi-factor copula model, when the number of groups G increases and the number of the

observable variables are fixed, the inferences of the latent variables will become less accurate which also leads to inaccurate bivariate copula selection. This effect could be seen in both the MCMC approach and the VI. About the time of VI, it will not increase much when we try to compare the two factor copulas (G = 1) and the bi-factor copulas (G = 5).

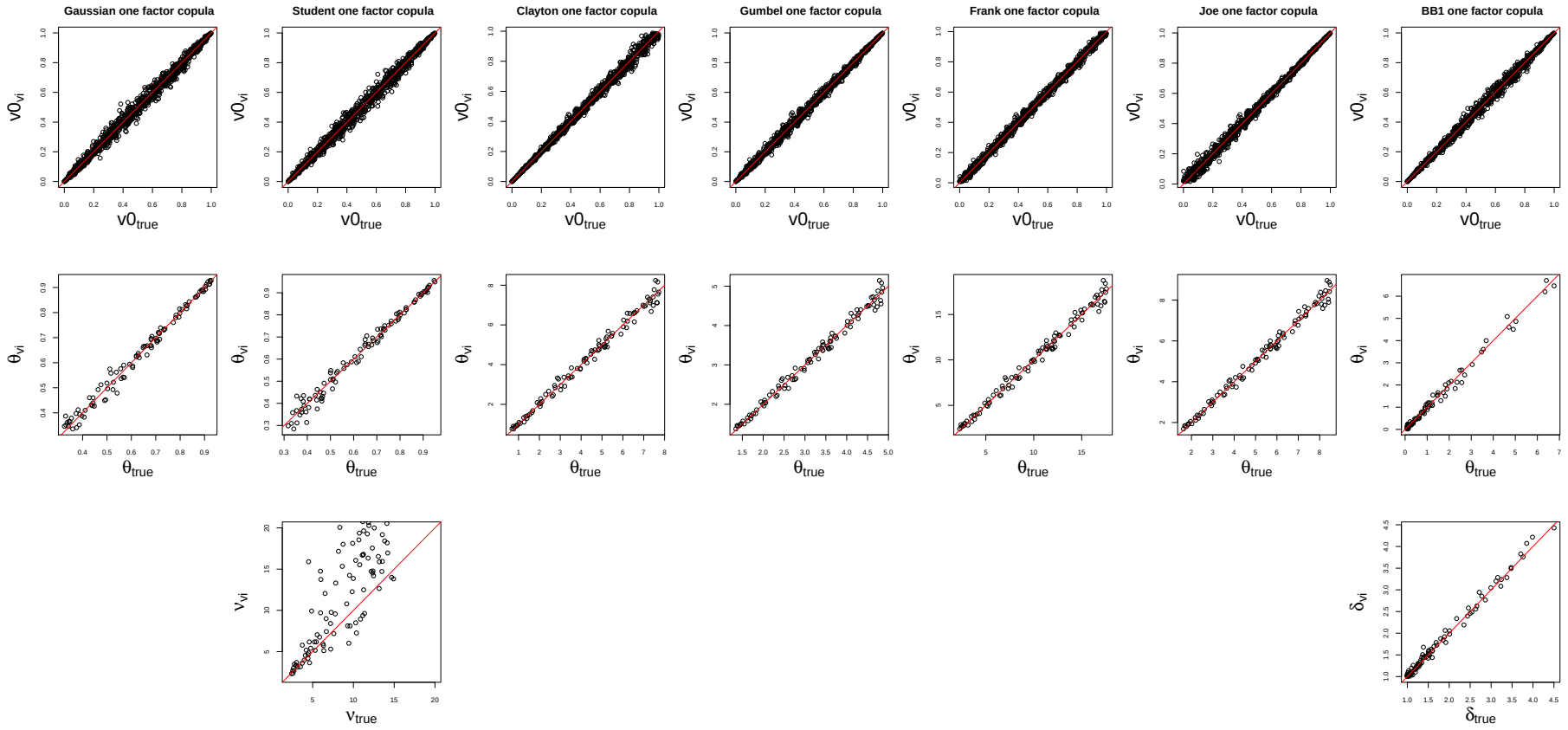

Figure 5: Variational inference for the one-factor copula models.

The figure compares the posterior means using VI 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.

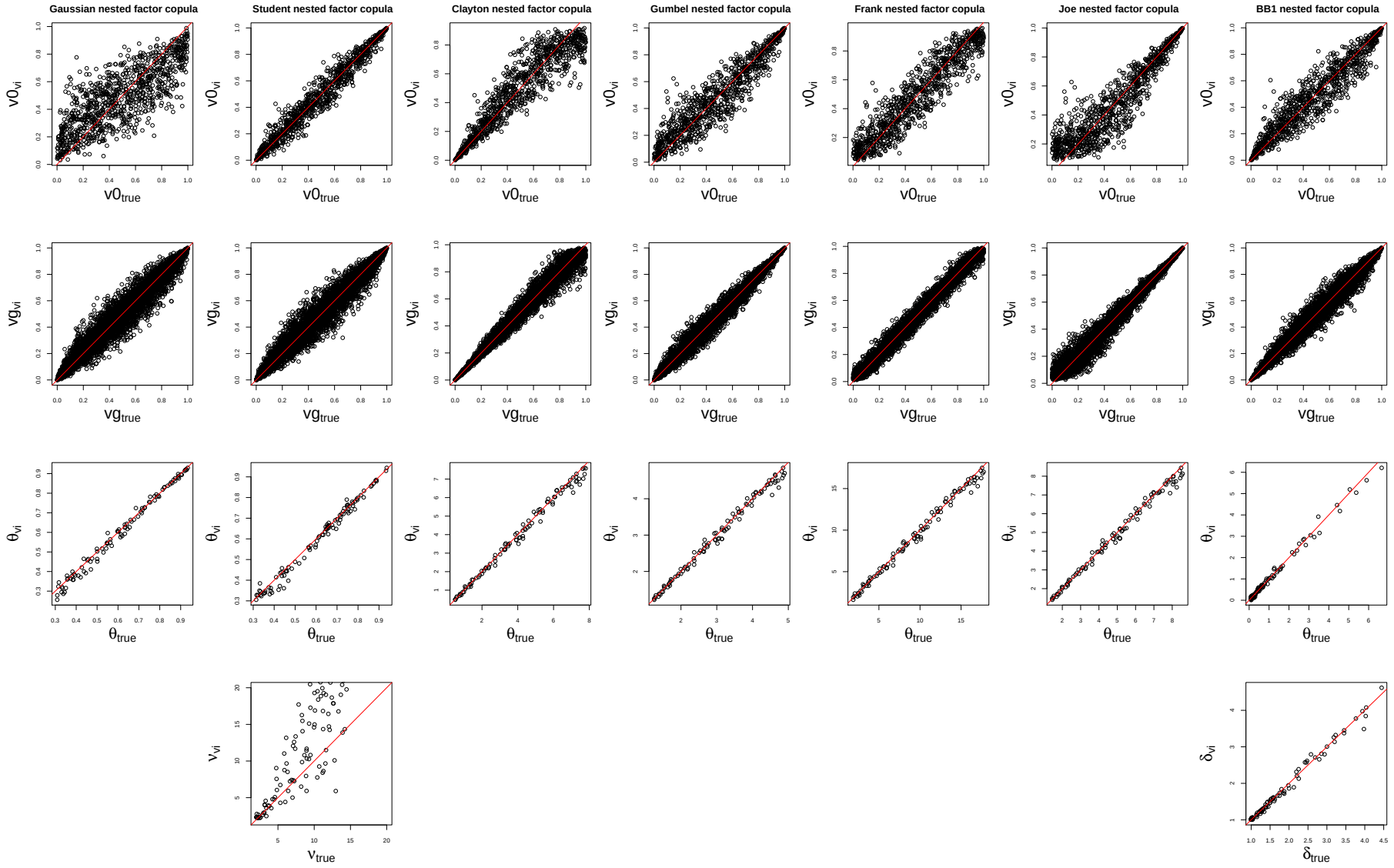

Figure 6: Variational inference for the nested factor copula models.

The figure compares the posterior means using the VI to the true generated values of 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.

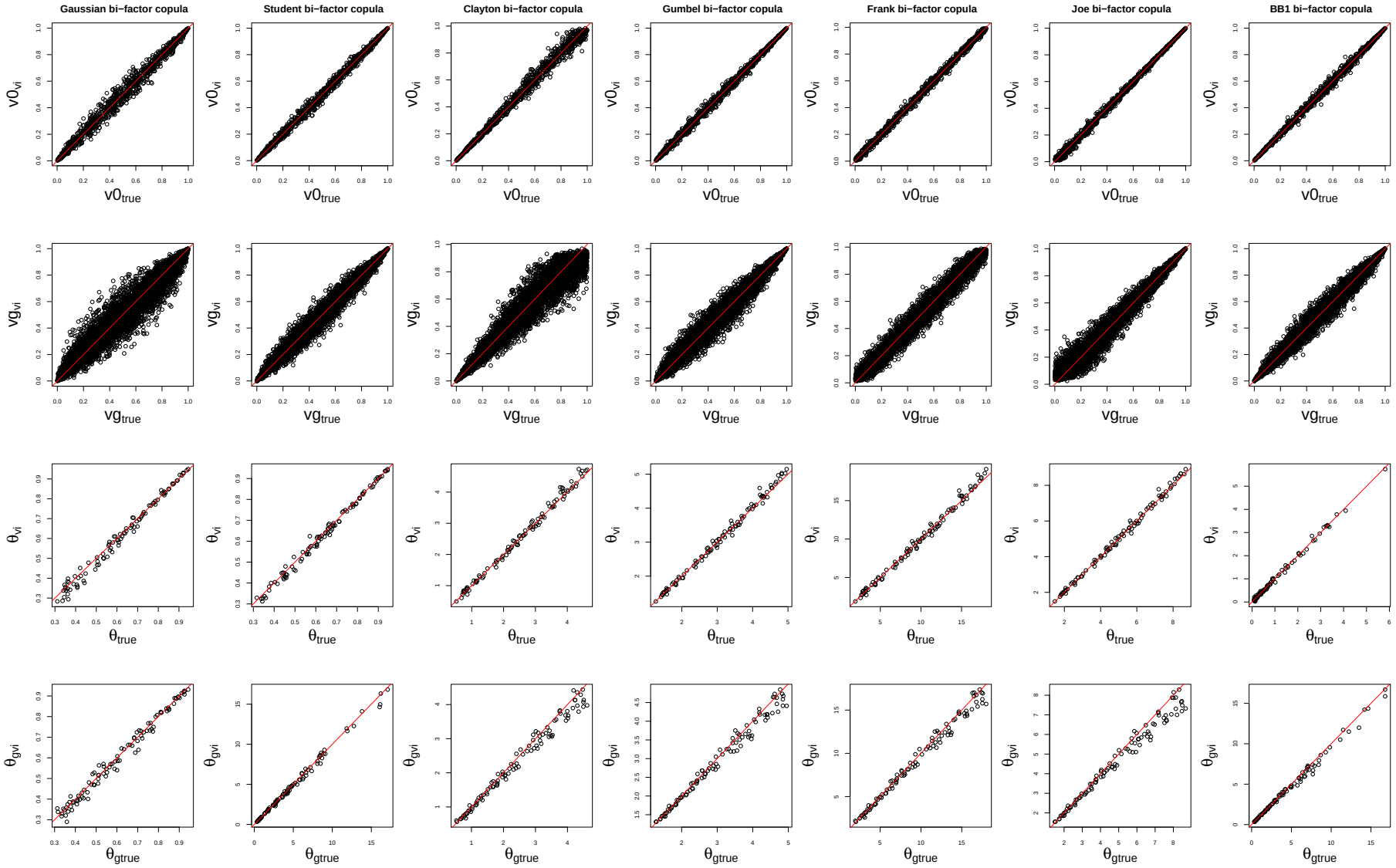

Figure 7: Variational inference for the bi-factor copula models.

The figure compares the posterior means using the VI to the true generated values of the bi-factor copula models. For the Student-t and BB1 bi-factor copula, the bivariate copulas in the second tree are a combination of one parameter copula families due to the identifiability issue.

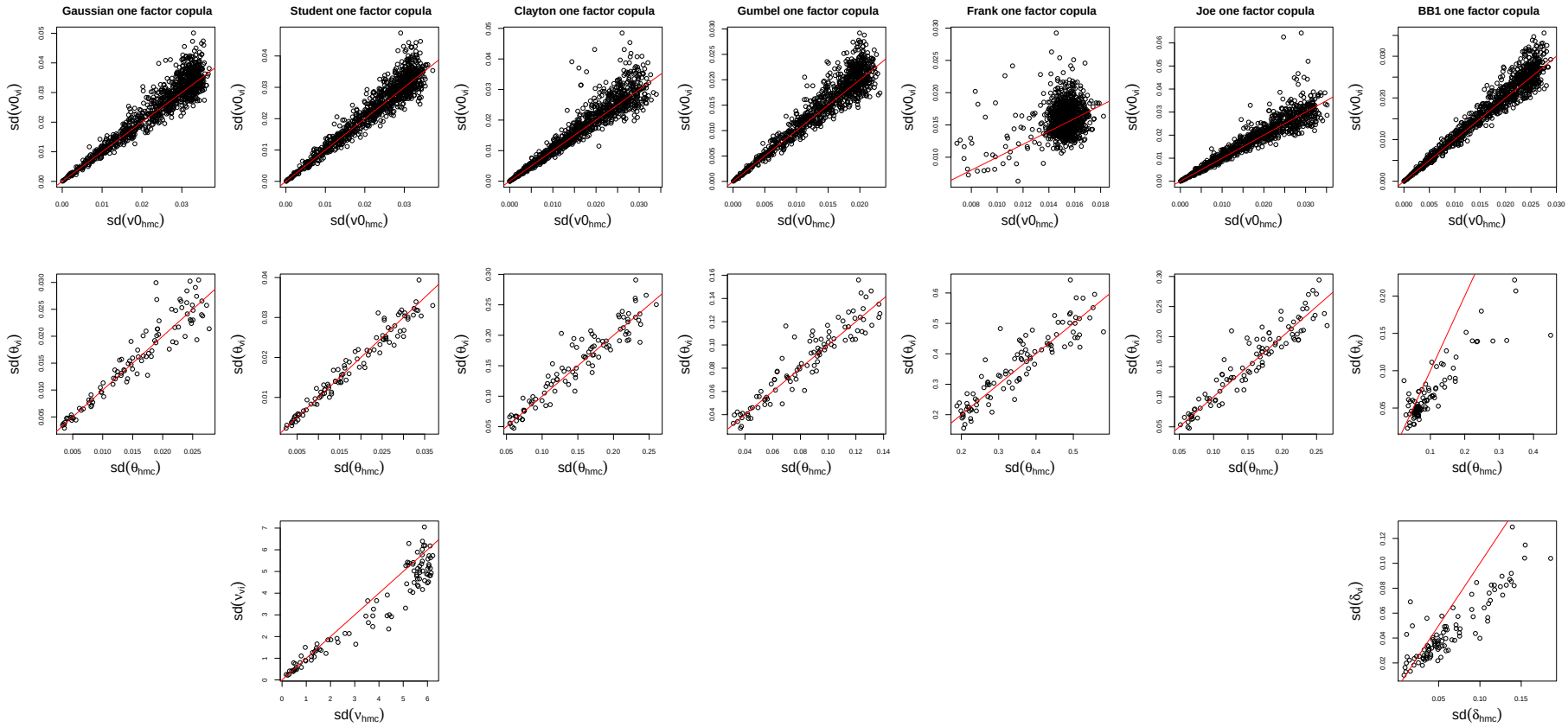

Figure 8: Comparison the standard deviations of the VI and the NUTS estimation for the one-factor copula models.

The figure compares the standard deviations of the VI and the NUTS estimation for the one-factor copula models. In the one-factor models, the standard deviations of the parameters θ are similar using both methods. The VI estimate of the parameters in the factor BB1 copula has a slightly lower variance than that of the MCMC approach.

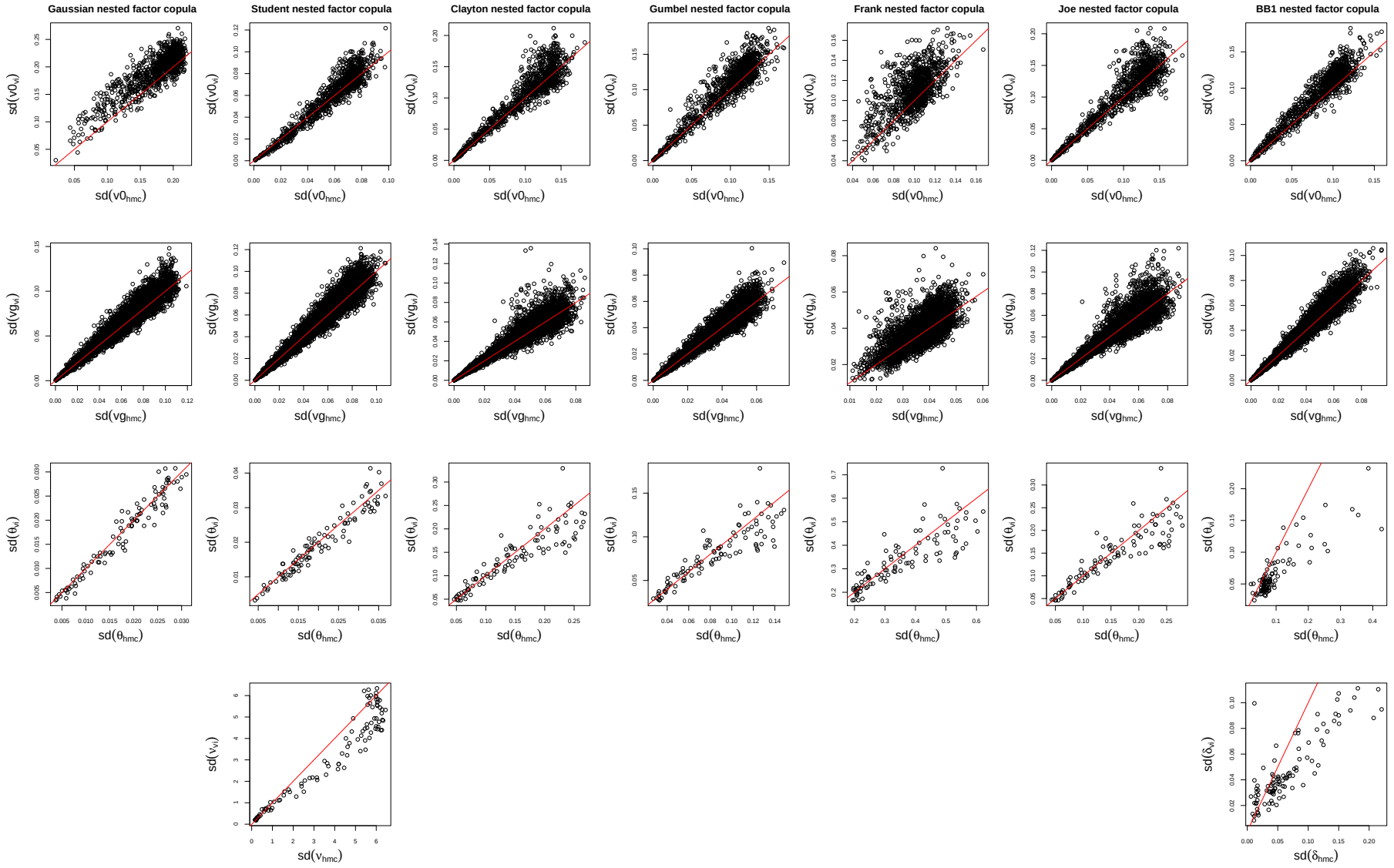

Figure 9: Comparison the standard deviations of VI and NUTS estimation for the nested factor copula models.

The figure compares the standard deviation of VI and NUTS estimation for the nested-factor copula models. In the nested-factor model, the standard deviations of the parameters θ are similar using both methods.

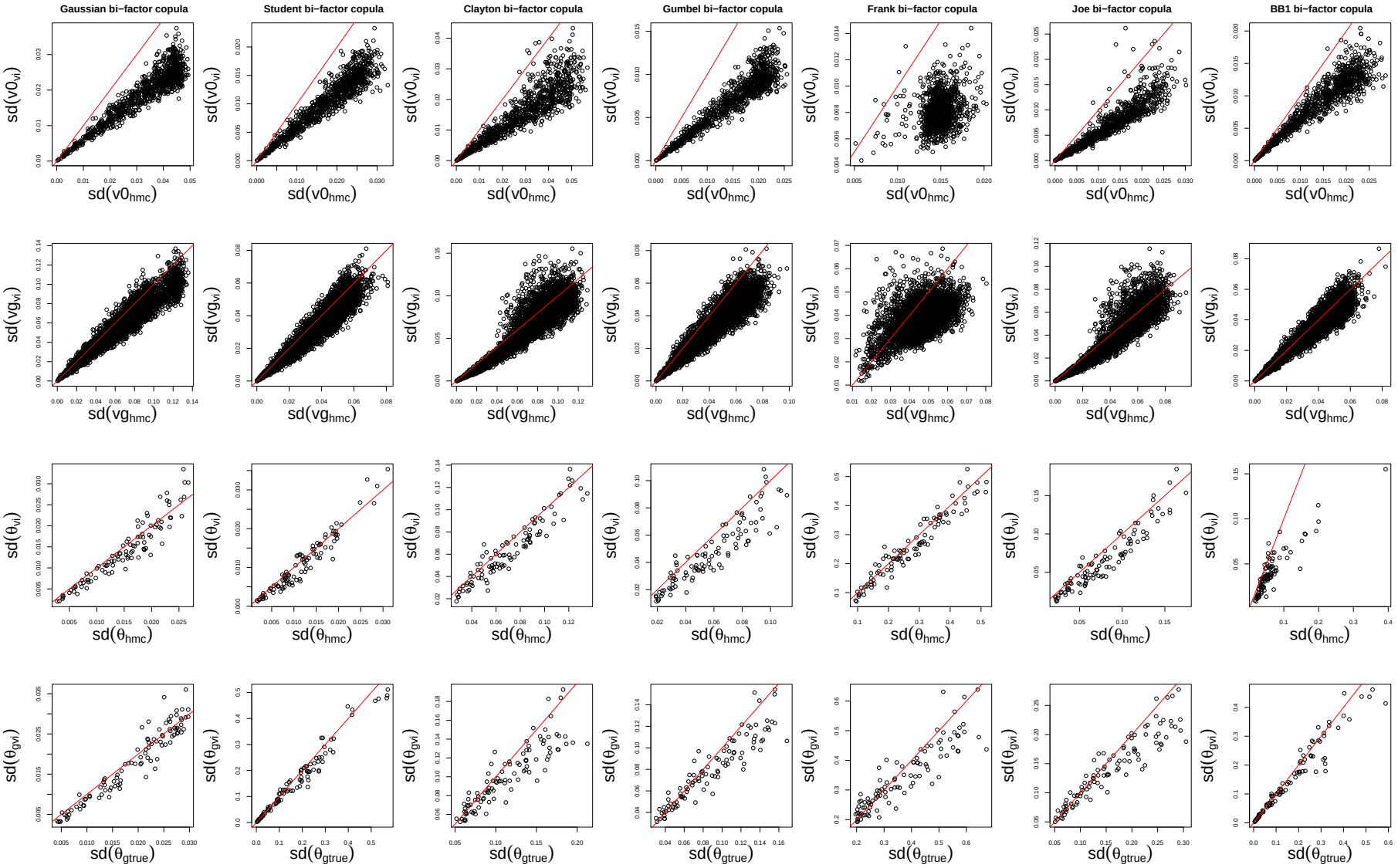

Figure 10: Comparison the standard deviations of the VI and the NUTS estimation for the bi-factor copula models.

The figure compares the standard deviations of VI and NUTS estimation for the bi-factor copula models. In the bi-factor model, the standard deviations of the v_0 are lower than that of the MCMC approach. It is acceptable because we are more interested in the copula parameters θ.

D Empirical illustration

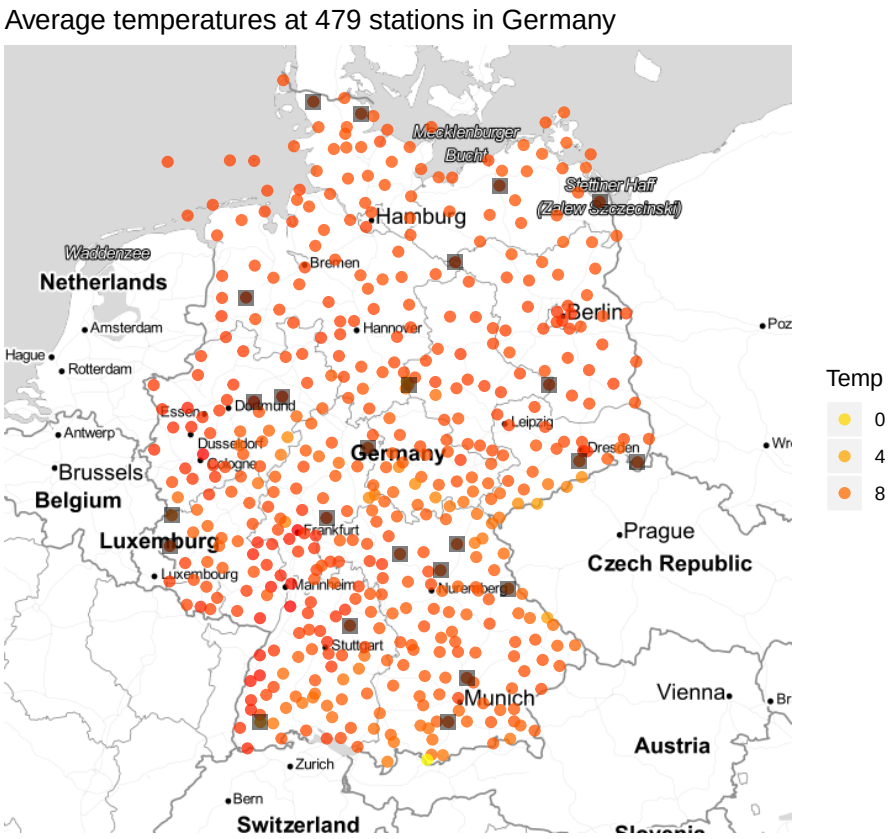

Figure 11: The locations of 479 stations in Germany.

The figure shows the locations of 479 stations in Germany. The color represents for the average daily temperature during 2013-2015. 24 stations are marked as black squares for prediction purpose.

E Monte Carlo study

In this section, we present a Monte Carlo study between the estimation of the copula parameters in the factor copulas model using the VI method and the MLE method. We generate 100 samples of $d = 10$ time series with $T = 500$ observation points from the Gumbel one-factor copula and 100 samples of $d = 30$ time series into $G = 5$ groups with $T = 500$ from Gumbel nested and Gumbel bi-factor copulas. Table 10 shows the averages of the mean square error (MSE) and the mean absolute error (MAE) of the estimation using VI and MLE. The MSE and MAE are quite

similar in case of the one-factor and nested factor models. The MLE is slightly better in case of the bi-factor copulas.

Table 10: Comparison of estimation using the VI and the MLE for the factor copula models

		Gumbel One-factor (d = 10, T = 500)	Gumbel Nested factor (d = 30, k = 5, T = 500)	Gumbel Bi-factor (d = 30, k = 5, T = 500)
VI	MSE	0.0237	0.0260	0.0435
	MAE	0.1129	0.1201	0.1608
MLE	MSE	0.0239	0.0241	0.0408
	MAE	0.1123	0.1086	0.1032

The table shows the averages of the mean square error (MSE) and the mean absolute error (MAE) of the estimation using VI and MLE for 100 simulations. The MSE and MAE are quite similar in case of the one-factor and nested factor models. The MLE is slightly better in case of the bi-factor 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.

E. F. Acar, C. Czado, and M. Lysy. Flexible dynamic vine copula models for multivariate time series data. *Econometrics and Statistics*, 12:181–197, 2019.

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*, 30(4):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.

R. J. Giordano, T. Broderick, and M. I. Jordan. Linear response methods for accurate covariance estimates from mean field variational bayes. In *Advances in Neural Information Processing Systems*, pages 1441–1449, 2015.

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 and H. Joe. Flexible copula models with dynamic dependence and application to financial data. *Econometrics and Statistics*, 2020.

P. Krupskii, R. Huser, and M. G. Genton. Factor copula models for replicated spatial data. *Journal of the American Statistical Association*, 113(521):467–479, 2018.

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 discrete-margined copula models with application to time series. *Journal of Computational and Graphical Statistics*, pages 1–17, 2019.

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*, 59 (2):347–370, 1991.

H. Nguyen, M. C. Ausín, and P. Galeano. Parallel Bayesian inference for high-dimensional dynamic factor copulas. *Journal of Financial Econometrics*, 17(1):118–151, 2019.

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.

V. M.-H. Ong, D. J. Nott, and M. S. Smith. Gaussian variational approximation with a factor covariance structure. *Journal of Computational and Graphical Statistics*, 27(3):465–478, 2018.

R. Ranganath, S. Gerrish, and D. M. Blei. Black box variational inference. *AISTATS*, 33:814–822, 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. *Uncertainty in Artificial Intelligence*, 16:647–656, 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.

M. Smith, A. Min, C. Almeida, and C. Czado. Modeling longitudinal data using a pair-copula decomposition of serial dependence. *Journal of the American Statistical Association*, 105(492): 1467–1479, 2010.

M. S. Smith and N. Klein. Bayesian inference for regression copulas. *arXiv preprint arXiv:1907.04529*, 2019.

M. S. Smith, R. Loaiza-Maya, and D. J. Nott. High-dimensional copula variational approximation through transformation. *Journal of Computational and Graphical Statistics*, (1):1–35, 2020.

B. K. Tan, A. Panagiotelis, and G. Athanasopoulos. Bayesian inference for the one-factor copula model. *Journal of Computational and Graphical Statistics*, 28(1):155–173, 2019.

L. S. Tan and D. J. Nott. Gaussian variational approximation with sparse precision matrices. *Statistics and Computing*, 28(2):259–275, 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.

M. J. Wainwright and M. I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, 1(1–2):1–305, 2008.

C. Zhang, J. Butepage, H. Kjellstrom, and S. Mandt. Advances in variational inference. *IEEE transactions on pattern analysis and machine intelligence*, 41:2008–2026, 2019.