Universidad Carlos III de Madrid

Variational Inference for high dimensional structured factor copulas

Hoang Nguyen

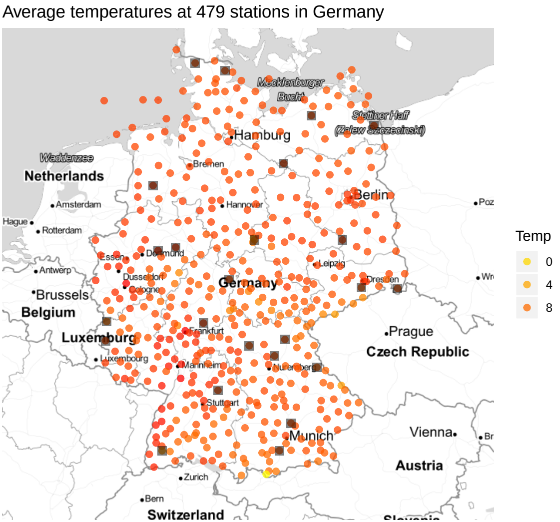
joint work with **M. Concepcion Ausin, Pedro Galeano**
Universidad Carlos III de Madrid

Monday 17th June, 2019

Motivation

- In many research fields, understanding the dependence structure is the key to predict the evolution of variables.
- Alternative to classical multivariate models, copulas can be easy to apply with a lot of flexibility.
- In high dimensions, there are few factors that contributes to the dependence structure.
  - The factor copula model also provides parsimonious and interpretable economic meanings
  - Complex structure needs an efficient inference method.

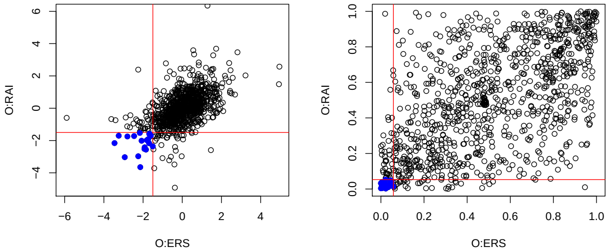
Preliminary results - Temperatures



1. Introduction to copulas
2. Variational Inference
3. Simulation
    - VI vs MCMC
    - Model check
4. Empirical Illustration
5. Conclusion

Introduction to Copulas

Copula is a n-dimensional joint cumulative distribution function (cdf) in the unit domains.



Let $F(x_1, ..., x_d|\theta)$ be a n-dimensional joint cdf with marginals $F_1, ..., F_d$ for all x_i in $[-\infty, \infty]$, and $u_i = F_i(x_i|\theta_i)$ for all $i = 1, ..., d$, (see Sklar (1959))

$$F(x_1, ..., x_d|\theta) = C(u_1, ..., u_d|\theta)$$

$$f(x_1, ..., x_d|\theta) = c(u_1, ..., u_d|\theta_C) \prod_{i=1}^{n} f(x_i|\theta_i)$$

Bivariate copulas - Elliptical copulas

$$\text{Gaussian Copula } C_R^{Ga}(u) = \Phi_R^n(\Phi^{-1}(u_1), .., \Phi^{-1}(u_d))$$
$$\text{Student Copula } C_R^{St}(u;\nu) = F_R^{MSt}(F^{-1}(u_1;\nu), .., F^{-1}(u_d;\nu))$$

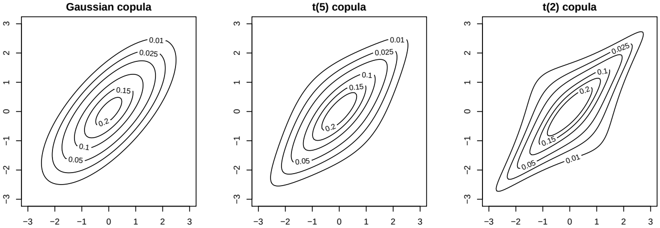


Figure: Contours of bivariate distributions with the same marginal standard normal

Bivariate copulas - Archimedean copulas

Common Bivariate Archimedean Copulas:

$$C(u_1, u_2) = \varphi^{-1}(\varphi(u_1) + \varphi(u_2))$$

Clayton (1978)	Frank (1979)	Gumbel (1960)	Joe (1993)
$\theta \geqslant 0$	$\theta \geqslant 0$	$\theta \geqslant 1$	$\theta \geqslant 1$
$\varphi(t) = t^{-\theta} - 1$	$\varphi(t) = -ln\frac{e^{-\theta t}-1}{e^{-\theta}-1}$	$\varphi(t) = (-lnt)^{\theta}$	$\varphi(t) = -log(1-(1-t)^{\theta})$
$\varphi^{-1}(s) = (1+s)^{-1/\theta}$	$\varphi^{-1}(s) = -\frac{ln(1+e^{-s}(e^{-\theta}-1))}{\theta}$	$\varphi^{-1}(s) = exp(-s^{1/\theta})$	$\varphi^{-1}(s) = (1-(1-e^{-s}))^{1/\theta}$
$\lambda_L = 2^{-1/\theta}, \lambda_U = 0$	No tail dependence	$\lambda_L = 0,\ \lambda_U = 2 - 2^{1/\theta}$	$\lambda_L = 0,\ \lambda_U = 2 - 2^{1/\theta}$

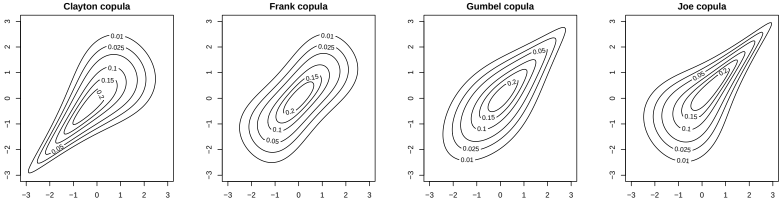


Figure: Contours of bivariate distributions with the same marginal standard normal

Factor copulas

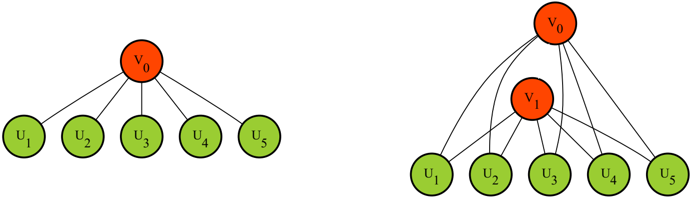


Figure: One factor and two factor copula models (Krupskii and Joe (2013))

Go to algorithm

Bi-factor and nested factor copulas

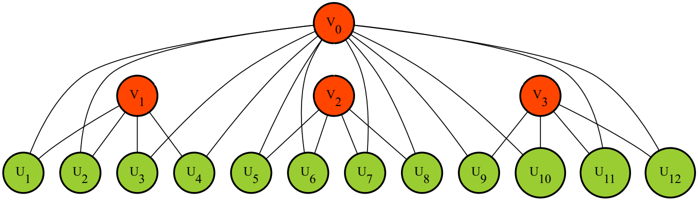


Figure: Bifactor copulas with $d = 12$ and $G = 3$ (Krupskii and Joe (2015))

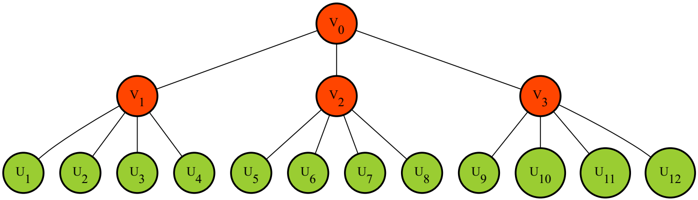


Figure: Nested factor copulas with $d = 12$ and $G = 3$ (Krupskii and Joe (2015))

Discussion

- Estimation of factor copula models in high dimensions is a challenging problem.
  - **Contribution:** Using the Variational approach to make fast inferences.
- In general, the hierarchical structure and group members can be specified based on the prior knowledge as the assumption of hierarchical models. However, the bivariate copula links are unknown.
  - **Contribution:** We derive an automatic procedure to recover the hidden dependence structure using the posterior modes of the latent variables.

Bayesian inference and Variational inference

Assuming that we have specify a factor copula structure together with bivariate linking copula in each tree layers.

- We are interested in the inference on the collection of latent variables and copula parameters $\Theta = \{v, \theta\}$ based on the observables $u = \{u_{1t}, \dots u_{dt}\}$
- The conditional copula density of one factor copulas is the following,

$$p(u_1, \dots, u_d | v_0; \theta) = \prod_{i=1}^{d} \left[\prod_{t=1}^{T} c_{U_i, V_0}(u_{ti}, v_{t0} | \theta_{0i}) \right],$$

- The the joint posterior density up to a normalized constant is

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

Bayesian inference and Variational inference

For bi-factor copula, we derive the posterior using the properties for vine copula,

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

where $u_{i_g|v_0} = F(u_{i_g|v_0})$. Thus, it is computational expensive. We approximate the posterior by a proposal $q(\Theta|\lambda^*)$, see Kucukelbir et al. (2017).

$$q(\Theta|\lambda^*) \approx p(\Theta|u)$$

VI vs MCMC - Inference time

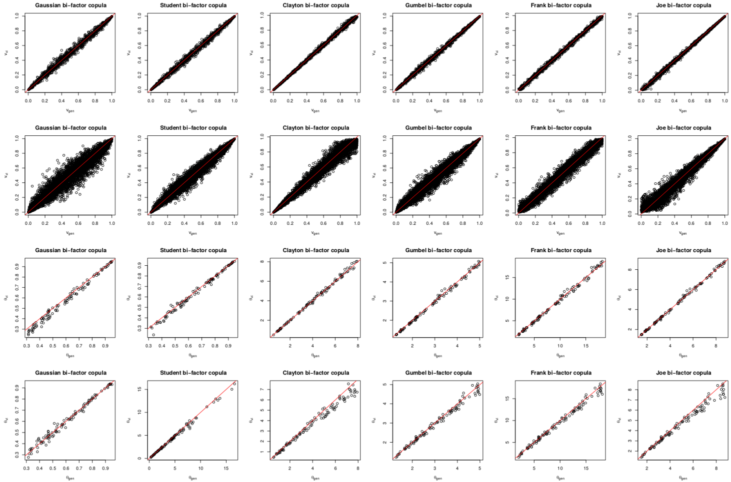
We generate a sample of d = 100 variables in G = 5 groups with T = 1000 time observations. Bivariate copula types are Gaussian, Student, Clayton, Gumbel, Frank, Joe (and their rotation 90, 180, 270 degree) and Mix copulas. Time is report in seconds using one core Intel i7-4770 processor.

Table: Time estimation using VI and MCMC

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

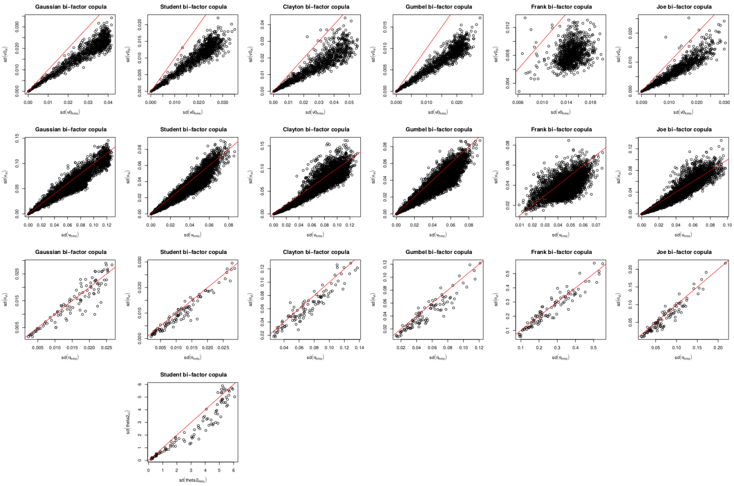
We report the time of estimation for the factor copula models using VI and MCMC for 1000 samples. The VI convergence time depends on its optimization parameters such as the number of MC samples, number of MC for calculating the gradients, tolerance, among others. The MCMC approach depends mainly on the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

Bi-factor copula model



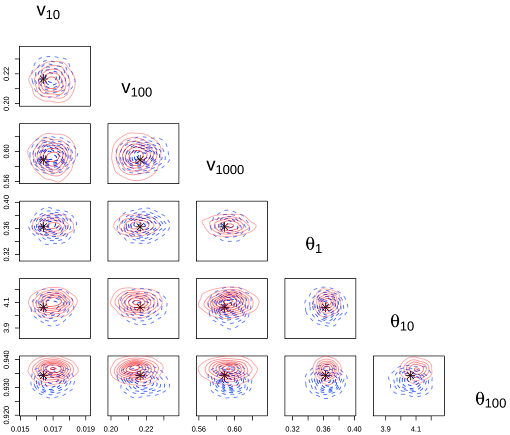
The figure compares the posterior means using variational approximation to the true generated values of the bi-factor copula models. For the bi-factor Student-t copula, the bivariate copulas in the second tree are mixed of other copula families due to the identification issue.

Bi-factor copula model



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

Why VI performs well in comparison to MCMC



The figure compares the posterior samples of VI (in red) and MCMC (in blue) for Gaussian one factor copula model.

Recover the dependence structure

1. Initialize random bivariate links.
2. Given a copula structure, we use VI to estimate the latent variables and obtain the posterior mode $\bar{v}$.

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

The BIC of the one factor copula could be derived as,

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

Recover the dependence structure

1. Initialize random bivariate links.
2. Given a copula structure, we use VI to estimate the latent variables and obtain the posterior mode $\bar{v}$.

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

The BIC of the one factor copula could be derived as,

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

3. Choose bivariate copula functions between u_i and $\bar{v}$ based on BIC.

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

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

4. If there are changes in the bivariate links, repeat (2)-(4) until convergence.

One factor copula model

Table: Model comparison for the one-factor copula models

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

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

Nested factor copula model

Table: Model comparison for the nested factor copula models

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

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

Bifactor copula model

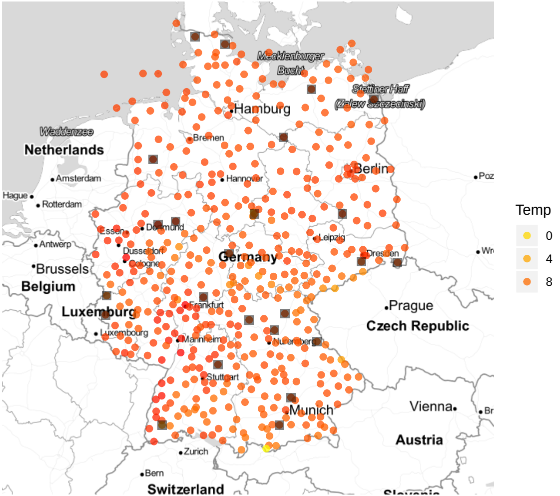
Table: Model comparison for the bi-factor copula models

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

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

Empirical Illustration - Temperatures

Average temperatures at 479 stations in Germany



Empirical Illustration - Temperatures

- Data: daily temperatures measured at 479 stations in Germany
- $T = 1094$ observations
- $G = 24$ groups that the distance among stations in each group at most 200 kilometers.
- Marginal distribution of temperatures using the ARMA(1,1) process, see Erhardt et al. (2015),

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

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

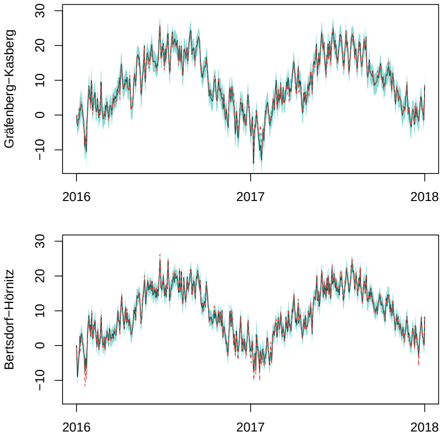
where $(\alpha_{0i}, \phi_i, \delta_i)$ are parameters of ARMA(1,1) process, (ν_i, γ_i) are the parameters of skew Student-t distribution, and $(\alpha_{ki}, \beta_{ki})$ are the slopes of Fourier exogenous regressors with different frequencies ($K = 2$ to minimize the AIC of marginal models).

Empirical Illustration - Temperatures

Table: Model comparison of daily temperature dependence

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

Empirical Illustration - Temperatures



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

Conclusion and Discussion

Contributions:

- Fast variational inference for factor copula model in high dimensions.
- Recover the bivariate copula functions based on the posterior mode of the latent variables.

Findings:

- Compared to MCMC, VI tends to be faster and easier to scale to large data.
- The posterior means of VI samples are similar to that of MCMC samples while the posterior standard deviations are only underestimated in the case of bi-factor copulas.

Extensions:

- There are several strategies to extend for dynamic factor copula models.

Thank you

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.

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, 2015.

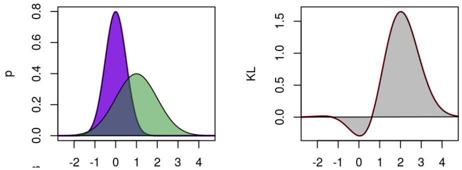
A. Kucukelbir, D. Tran, R. Ranganath, A. Gelman, and D. M. Blei. Automatic differentiation variational inference. *The Journal of Machine Learning Research*, 18(1):430–474, 2017.

M. Sklar. *Fonctions de Répartition à n Dimensions et Leurs Marges*. Université Paris 8, 1959.

Kullback Leibler divergence

In order to measure how well the approximation, Variational inference chooses the proposal distribution come close to the posterior in term of KL divergence:

$$KL(Q||P) = \int Q(x) \log \frac{Q(x)}{P(x)} dx \geq 0$$



Note that: $KL(Q||P) \neq KL(P||Q) \geq 0$

Objective function

We specify a family $\mathcal{Q}$ of densities as the proposal distribution

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

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

Objective function

We specify a family $\mathcal{Q}$ of densities as the proposal distribution

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

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

Because we cannot compute the KL, we optimize an alternative objective (Evidence lower bound) that is equivalent to the KL up to an added constant:

$$\begin{aligned}\arg\max_{\lambda} \mathtt{ELBO}(q) &= \mathbb{E}_q[\mathtt{log} p(u,\Theta)] - \mathbb{E}_q[\mathtt{log} q(\Theta;\lambda)] \\ &= \mathtt{log} p(u) - KL(q(\Theta;\lambda)||p(\Theta|u)) \leq \mathtt{log} p(u)\end{aligned} \quad (1)$$

such that when $q(\Theta;\lambda) = p(\Theta|u)$, we have $\mathtt{ELBO} = \mathtt{log} p(u)$.

Objective function

We specify a family $\mathcal{Q}$ of densities as the proposal distribution

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

Because we cannot compute the KL, we optimize an alternative objective (Evidence lower bound) that is equivalent to the KL up to an added constant:

$$\begin{aligned}\arg\max_{\lambda} \texttt{ELBO}(q) &= \mathbb{E}_q[\texttt{log}p(u,\Theta)] - \mathbb{E}_q[\texttt{log}q(\Theta;\lambda)] \\ &= \texttt{log}p(u) - KL(q(\Theta;\lambda)||p(\Theta|u)) \leq \texttt{log}p(u)\end{aligned} \tag{1}$$

such that when $q(\Theta;\lambda) = p(\Theta|u)$, we have $\texttt{ELBO} = \texttt{log}p(u)$.
Monte Carlo approximation for ELBO with $\Theta_{(s)} \sim q(\Theta;\lambda)$

$$\texttt{ELBO}(q) \approx \frac{1}{S}\sum_{s=1}^{S}\left[\texttt{log}p(u,\Theta_{(s)})\right] - \mathbb{E}_{q(\Theta)}[\texttt{log}q(\Theta;\lambda)].$$

Optimization strategy in Kucukelbir et al. (2017)

- Due to several restrictions on the parameters, we transform the constraint space of the copula parameters to the real coordinate space $\tilde{\Theta} = \{\tilde{\Theta}_j\} = \{\mathbb{T}_j(\Theta_j)\} = \mathbb{T}(\Theta)$, for $j = 1, \ldots, N$.
- We assume a product of univariate Gaussian densities as the proposal density,

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

- Then, we apply the stochastic optimization to maximize the ELBO.

$$\begin{aligned} \mathtt{ELBO}(q) &\approx \frac{1}{S}\sum_{s=1}^{S}\left[\mathtt{log}p(u, \mathbb{T}^{-1}(\tilde{\Theta}_{(s)})) + \mathtt{log}|\det J_{\mathbb{T}^{-1}}(\tilde{\Theta}_{(s)})|\right] \\ &- \mathbb{E}_{q(\tilde{\Theta})}[\mathtt{log}q(\tilde{\Theta}; \lambda)]. \\ \nabla_\lambda \mathtt{ELBO} &\approx \frac{1}{M}\sum_{m=1}^{M}\nabla_\lambda\left[\mathtt{log}p(u, \mathbb{T}^{-1}(\tilde{\Theta}_{(m)})) + \mathtt{log}|\det J_{T^{-1}}(\tilde{\Theta}_{(m)})|\right] \\ &- \nabla_\lambda \mathbb{E}_{q(\tilde{\Theta})}[\mathtt{log}q(\tilde{\Theta}; \lambda)] \end{aligned} \quad (3)$$

Transformation functions

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

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

Bivariate copula families and their characteristics

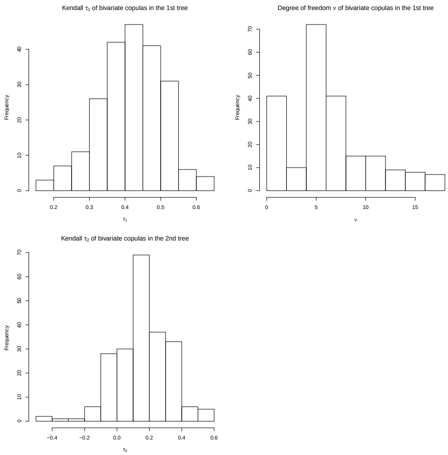
Table: Bivariate copula families and their characteristics

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	$C_G(u,v;\theta)=\exp\left[-\left\{(-\log u)^\theta+(-\log v)^\theta\right\}^{1/\theta}\right]$	$\pi_G(\theta)=\frac{1}{\theta^2}$	$\theta\in[1,\infty)$	$1-\frac{1}{\theta}$
	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	$C_{Fp}(u,v;\theta)=-\frac{1}{\theta}\log\left(1-\frac{(1-e^{-\theta u})(1-e^{-\theta v})}{1-e^{-\theta}}\right)$	$\pi_F(\theta)=\frac{4}{\theta^2}(1-B(\theta)+2D_1(\theta))$	$\theta\in(0,\infty)$	$1-\frac{4}{\theta}(1-D_1(\theta))$
	F_n	$C_{Fn}(u,v;\theta)=-\frac{1}{\theta}\log\left(1-\frac{(1-e^{-\theta u})(1-e^{-\theta v})}{1-e^{-\theta}}\right)$	$\approx Cauchy(\theta;0,6)$	$\theta\in(-\infty,0)$	
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}	$C_{J90}(u,v;\theta)=v-C_J(1-u,v;-\theta)$	$\pi_J(\theta)=\sum_{k=1}^{\infty}\frac{8(-\theta(k-1)+2-1/k)}{(\theta k-2)^2(\theta(k-1)-2)^2}$	$\theta\in(-\infty,-1]$	$1-4\sum_{k=1}^{\infty}\frac{1}{k(\theta k-2)(\theta(k-1)-2)}$
	J_{270}	$C_{J270}(u,v;\theta)=u-C_J(u,1-v;-\theta)$	$\approx\frac{2}{(\theta-2)^2}$		
Independence	I	$C_I(u,v)=uv$	-	-	0

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

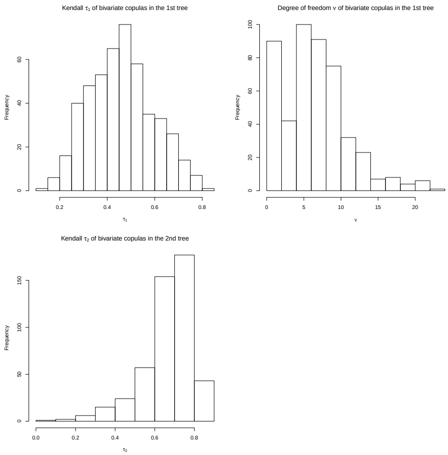
The table shows some common bivariate copula functions as well as their characteristics such as parameter ranges, and Kendall's τ correlation. We divide the symmetric copula functions into positive and negative Kendall's τ correlation copulas to prevent the identification issue of the factor copula models.

Empirical comparison - Stock returns



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

Empirical comparison - Temperatures



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

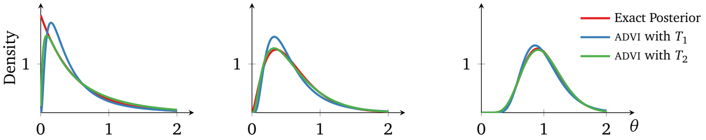
Sensitivity to Transformations

Consider a posterior density in the Gamma family, with support over $\mathbb{R}_{>0}$. Figure 9 shows three configurations of the Gamma, ranging from Gamma(1,2), which places most of its mass close to $\theta = 0$, to Gamma(10,10), which is centered at $\theta = 1$. Consider two transformations T_1 and T_2

$$T_1 : \theta \mapsto \log(\theta) \quad \text{and} \quad T_2 : \theta \mapsto \log(\exp(\theta) - 1),$$

both of which map $\mathbb{R}_{>0}$ to $\mathbb{R}$. ADVI can use either transformation to approximate the Gamma posterior. Which one is better?

Figure 9 show the ADVI approximation under both transformations. Table 2 reports the corresponding KL divergences. Both graphical and numerical results prefer T_2 over T_1. A quick analysis corroborates this. T_1 is the logarithm, which flattens out for large values. However, T_2 is almost linear for large values of θ. Since both the Gamma (the posterior) and the Gaussian (the ADVI approximation) densities are light-tailed, T_2 is the preferable transformation.



(a) Gamma(1,2) **(b)** Gamma(2.5,4.2) **(c)** Gamma(10,10)

Figure 9: ADVI approximations to Gamma densities under two different transformations.