Variational Inference for high dimensional structured factor copulas

Hoang Nguyen

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

Thursday 8th November, 2018

Motivation

- Understanding the dependence structure of financial time series is the key to perform portfolio allocation and risk management.
- The popularity of vine copulas makes them easier to derive the entire joint distribution. However, the number of parameters is explosive in high dimensions.
- In high dimensional financial time series, 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

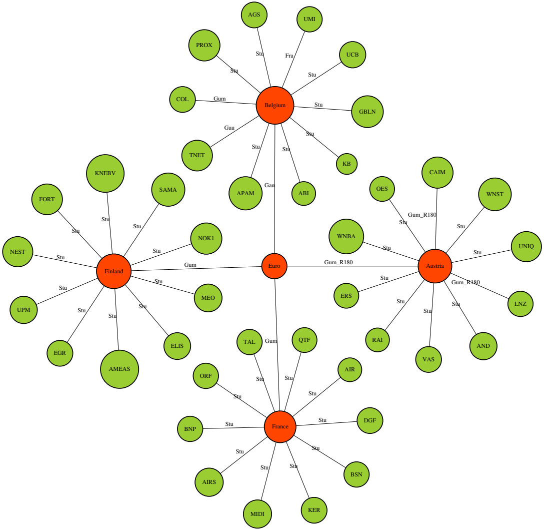


Figure: Nested factor copulas for the dependence structure of stock returns

Preliminary results

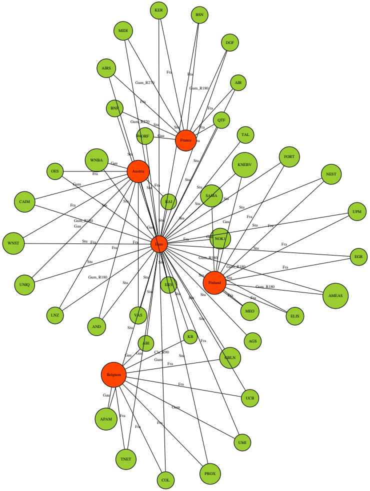


Figure: Bi-factor copulas for the dependence structure of stock returns

Go to Table

Introduction to Copulas

Consider a n-dimensional joint cdf F with marginals $F_1, ..., F_d$. There exists a copula C, such that

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

for all x_i in $[-\infty, \infty]$, $i = 1, ..., d$ and $u_i = F_i(x_i)$for all $i = 1, ..., d$,

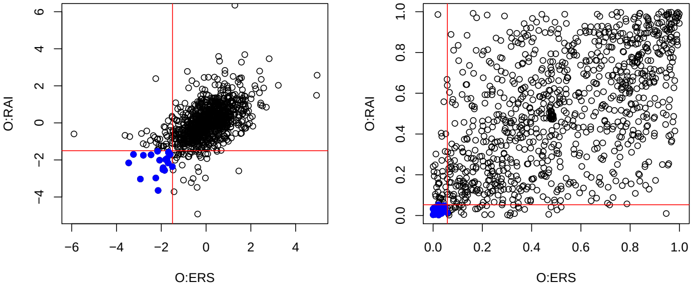


Figure: Copula function

Elliptical copulas

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

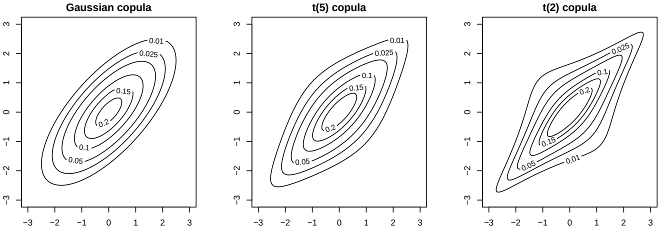


Figure: Contours of bivariate distributions with the same marginal standard normal

Archimedean copulas

Common Bivatiate 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	$\lambda_L = 0,\ \lambda_U = 2 - 2^{1/\theta}$	$\lambda_L = 0,\ \lambda_U = 2 - 2^{1/\theta}$

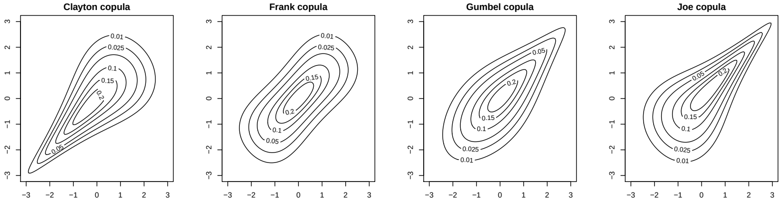


Figure: Contours of bivariate distributions with the same marginal standard normal

Factor copulas

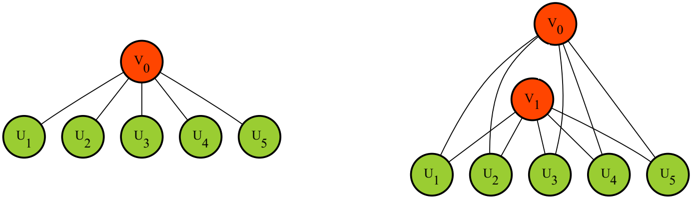


Figure: One factor and two factor copula models (Krupskii and Joe (2013))

Go to algorithm

Bifactor and nested factor copulas

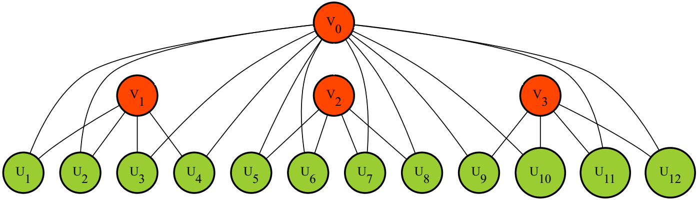


Figure: Bifactor copulas with $d = 12$ and $G = 3$ (Krupskii and Joe (2015a))

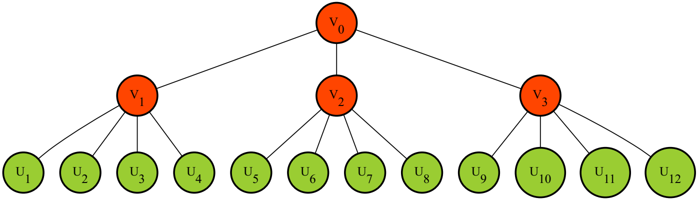


Figure: Nested factor copulas with $d = 12$ and $G = 3$ (Krupskii and Joe (2015a))

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 bifactor copula, we derive the posterior using the properties for vine copula,

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

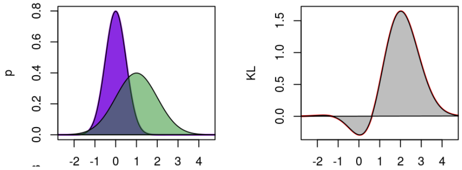
where $u_{i_g|v_0} = F(u_{i_g|v_0})$. Thus, it is computational expensive. We approximate the posterior by a proposal $q(\Theta|\lambda^*)$.

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

Kullback Leibler divergence

Variational Inference measures the different between two distributions using Kullback Leibler 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

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

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

Objective function

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

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

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 $\texttt{ELBO} = \texttt{log}p(u)$.

Optimization strategy

- 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,

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

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

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

Transformation functions in Stan package

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

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

One factor copula model

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 NUTS

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 NUTS							
One-factor	1097	491166	3762	5262	1083	2221	8535
Nested factor	1786	567862	5977	6098	1264	3478	20520
Bi-factor	17935	1758007	83680	177932	30345	236809	118729

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

One factor copula model

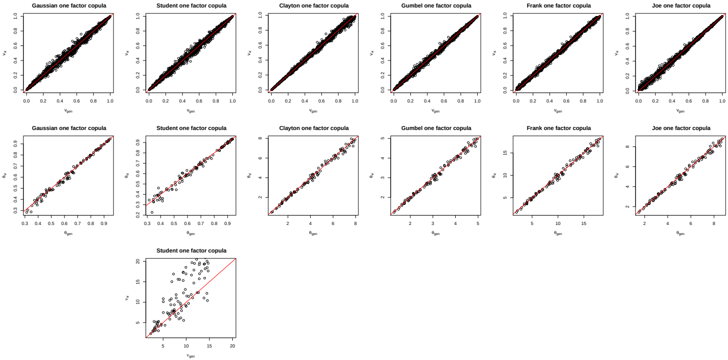


Figure: Variational inference for the one-factor copula models.

The figure compares the posterior means using variational approximation to the true generated values of the one factor copula models. In general, the posterior means are close to its true generated values.

One factor copula model

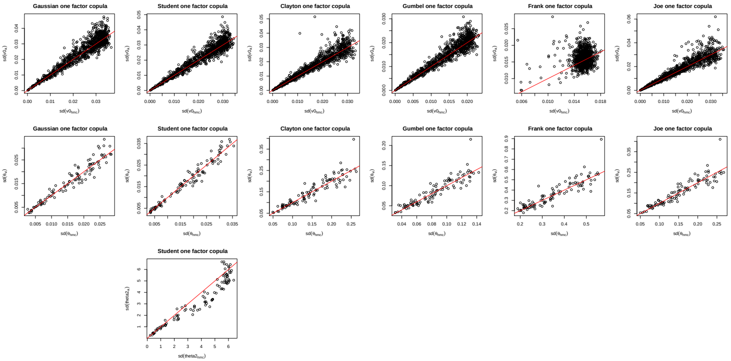


Figure: Comparison the standard deviations of VI and MCMC estimation for the one-factor copula models.

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

Bi-factor copula model

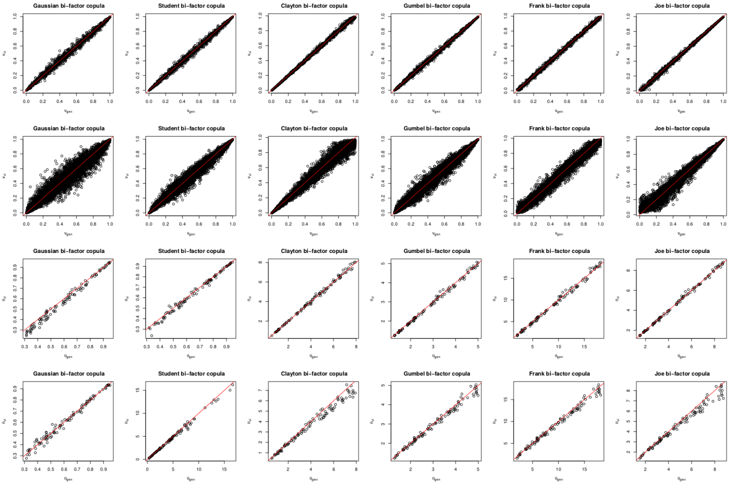


Figure: Variational inference for the bi-factor copula models.

The figure compares the posterior means using variational approximation to the true

Bi-factor copula model

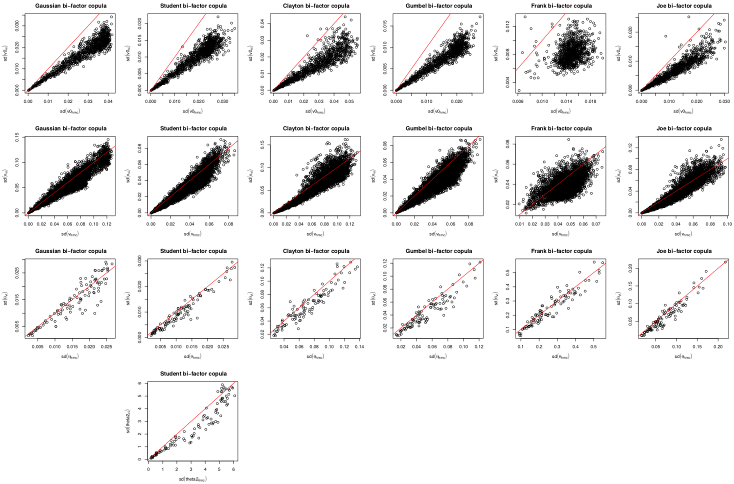
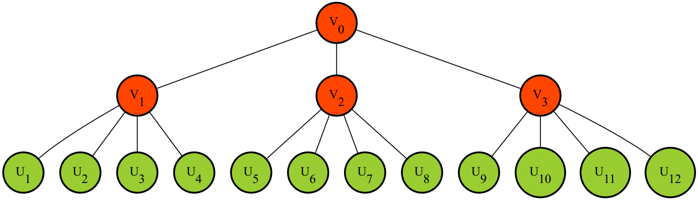


Figure: Comparison the standard deviations of VI and NUTS estimation for the bi-factor copula models.

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 $\bar{v}$.
3. Choose bivariate copula functions between u_i and $\bar{v}$ based on BIC.
4. If there are changes in the bivariate links, repeat (2)-(4) until convergence.



Model check

When the bivariate copula links are unknown, we can take advantage of the posterior modes of the latent variables to inspect the assumption of the initial links. The ideal bivariate function should minimize the BIC of each couple link.

Algorithm 1: Modified of the ADVI algorithm in (Kucukelbir et al., 2017). See https://github.com/hoanguc3m/vifcopula

Data: Copula data $u = \{u_{ti}\}$ 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 $\mu^{(i)} = 0, \omega^{(i)} = 0$;
  Variational inference to find μ^*, ω^* ;
  Obtain $\bar{v} = \mathbb{T}^{-1}(\bar{\tilde{v}})$;
  Select bivariate copula links between u and $\bar{v}$ based on minimum the BIC ;
  Reassign the copulas and estimate ;
end
Sample $\tilde{\Theta} \sim \Phi_N(\mu, exp(\omega))$, obtain $\Theta = \mathbb{T}^{-1}(\tilde{\Theta})$;
Return bivariate copula links and Θ samples ;

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

We illustrate an empirical example using d = 218 companies listed in G = 10 European countries from 01/01/2014 to 31/12/2017. 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,

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

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).

Empirical Illustration

Table: Model comparison of stock return dependence

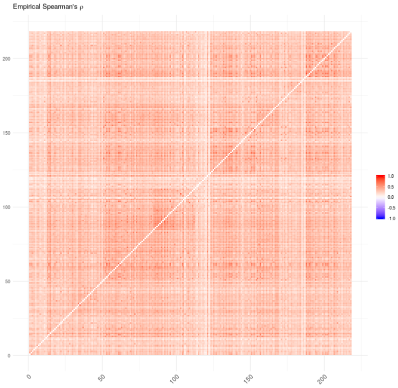
Structure	One-factor	Two-factor	Nested factor	Bi-factor
AIC	-102.4	-117.5	-111.9	**-117.7**
BIC	-100.5	-114.6	-109.9	**-114.9**
$\log p(u\|\theta)$	51.6	59.4	56.3	59.4
# bivariate links	218	434	228	412
# Gaussian	11	19	7	61
# Student-t	176	171	177	177
# Clayton (rotated)	0	0	0	4
# Gumbel (rotated)	20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

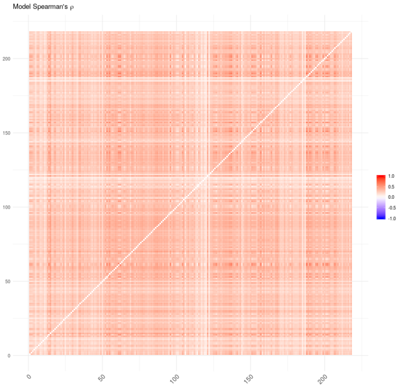
Empirical Illustration

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

$$\begin{aligned} \rho_L &= Cor\left((1-\frac{U_1}{p})^6, (1-\frac{U_2}{p})^6 \middle| U_1 < p, U_2 < p \right) \\ \rho_U &= Cor\left((1-\frac{1-U_1}{p})^6, (1-\frac{1-U_2}{p})^6 \middle| 1-U_1 < p, 1-U_2 < p \right) \end{aligned} \tag{4}$$

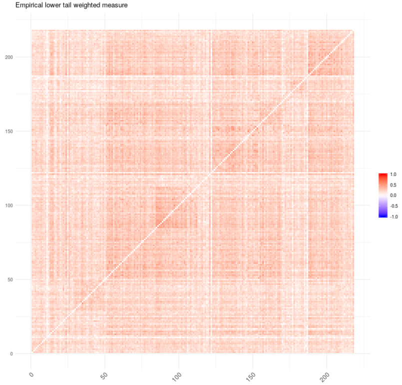
Empirical comparison

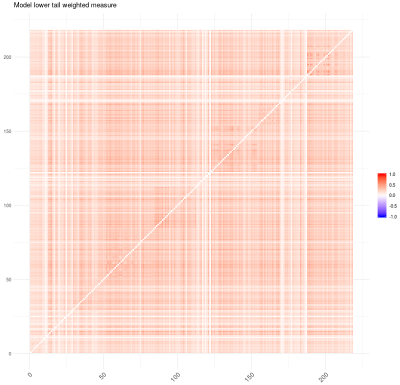




The average of the absolute difference of Spearman's ρ between the bi-factor copula model and that of the empirical copula data is 0.032

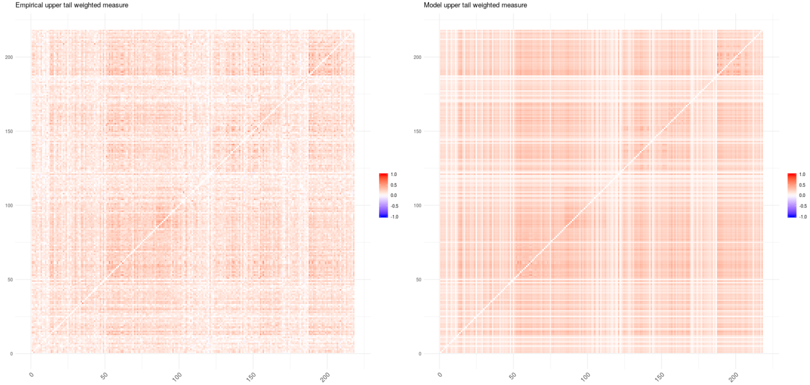
Empirical comparison





The average of the absolute difference of lower tail-weighted dependence measure between the bi-factor copula model and that of the empirical copula data is 0.059.

Empirical comparison



The average of the absolute difference of upper tail-weighted dependence measure between the bi-factor copula model and that of the empirical copula data is 0.073.

Conclusion and Discussion

- Fast variational inference for factor copula model in high dimensions.
- Recover the bivariate copula functions based on the posterior mode of the latent variables.
- 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.
- There are several strategies to extend for dynamic factor copula models.

Thank you

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.

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.

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. B. Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, pages 347–370, 1991.

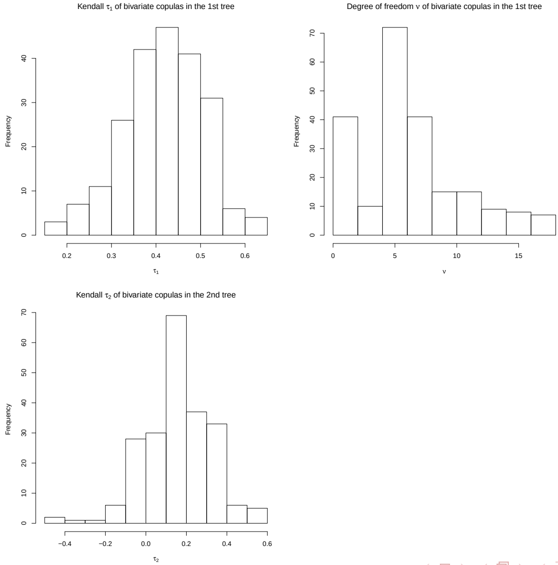
Bivariate copula families and their characteristics

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	$C_G(u,v;\theta)=\texttt{exp}\left[-\left\{(-\texttt{log}u)^\theta+(-\texttt{log}v)^\theta\right\}^{1/\theta}\right]$	$\pi_G(\theta)=\frac{1}{\theta^2}$	$\theta\in[1,\infty)$	$1-\frac{1}{\theta}$
	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}	$C_{J90}(u,v;\theta)=v-C_J(1-u,v;-\theta)$	$\pi_J(\theta)=\sum_{k=1}^{\infty}\frac{8(-\theta(k-1)+2-1/k)}{(\theta k-2)^2(\theta(k-1)-2)^2}$	$\theta\in(-\infty,-1]$	$1-4\sum_{k=1}^{\infty}\frac{1}{k(\theta k-2)(\theta(k-1)-2)}$
	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



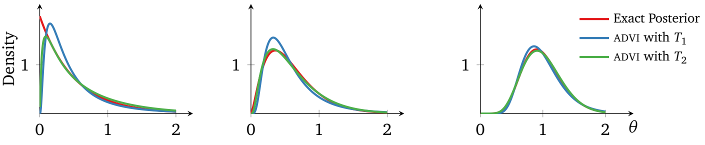
Sensitivity to Transformations

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