Improving forecasting performance using covariate-dependent copula models

Yanfei Kang

yanfeikang@buaa.edu.cn
http://yanfei.site/

School of Economics and Management
Beihang University

Outline

What is a copula?

- The word “copula” means **linking**.
- **Sklar’s theorem**
  Let H be a multi-dimensional distribution function with marginal distribution functions $F_1(x_1), ..., F_m(x_m)$. Then there exists a function C (**copula function**) such that

$$\begin{aligned} H(x_1, ..., x_m) =& C(F_1(x_1), ..., F_m(x_m)) \\ =& C\left(\int_{-\infty}^{x_1} f(z_1)dz_1, ..., \int_{-\infty}^{x_m} f(z_m)dz_m\right) = C(u_1, ..., u_m). \end{aligned}$$

  Furthermore, if $F_i(x_i)$ are continuous, then C is unique, and the derivative $c(u_1, ..., u_m) = \partial^m C(u_1, ..., u_m)/(\partial u_1 ... \partial u_m)$ is the **copula density**.

Measuring correlation and tail dependence

↪ Kendall's τ and tail-dependences

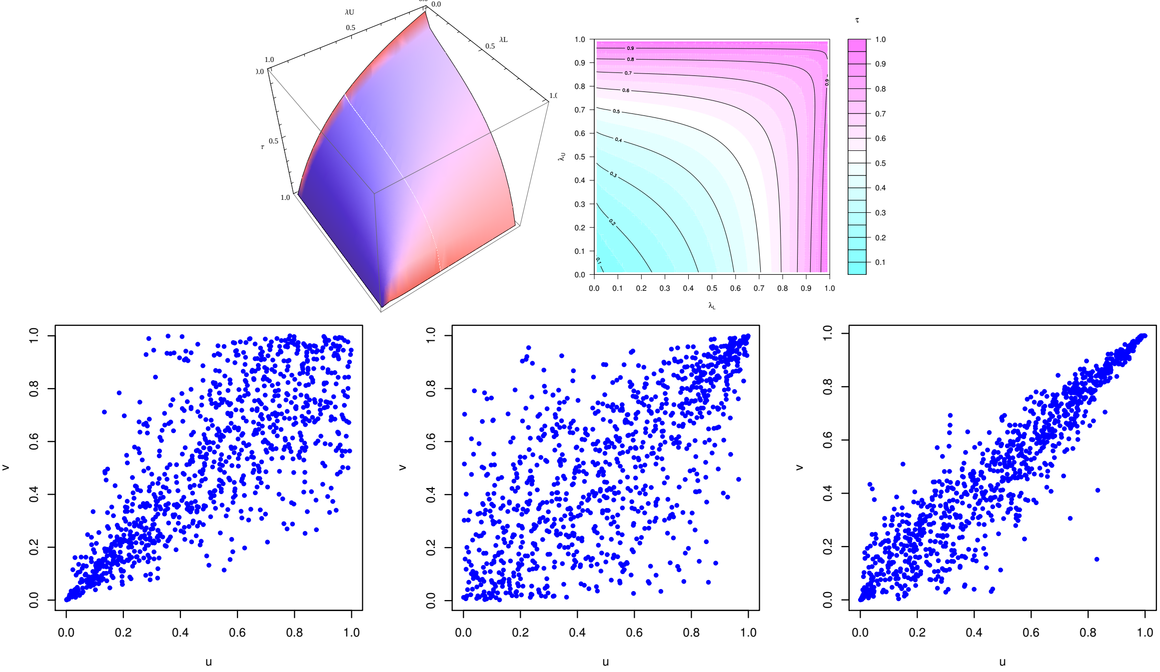
- The **Kendall's** τ can be written in terms of copula function:

$$\tau = 4\int\int F(x_1, x_2) dF(x_1, x_2) - 1 = 4\int\int C(u_1, u_2) dC(u_1, u_2) - 1.$$

- As well as the bivariate lower and upper **tail dependences**

$$\lambda_L = \lim_{u \to 0^+} Pr(X_1 < F_1^{-1}(u) | X_2 < F_2^{-1}(u)) = \lim_{u \to 0^+} \frac{C(u, u)}{u},$$
$$\lambda_U = \lim_{u \to 1^-} Pr(X_1 > F_1^{-1}(u) | X_2 > F_2^{-1}(u)) = \lim_{u \to 1^-} \frac{1 - C(u, u)}{1 - u}.$$

- Some facts:
  - The Kendall's τ is invariant w.r.t. **strictly** increasing transformations.
  - For all copulas in the elliptical class (Gaussian, t,...), $\tau = \frac{2}{\pi} arcsin(\rho)$.
  - The Gaussian copula has zero tail dependence.
  - The student t copula has asymptotic upper tail dependence even for negative and zero correlations. The tail dependence decreases when degrees of freedom increases.


λU
λL
τ
λ_U
λ_L
u
v

Covariate-dependent copula models
↛ The Joe-Clayton copula example

- The Joe-Clayton copula function

$$C(u, v, \theta, \delta) = 1 - \left[1 - \left\{\left(1 - \bar{u}^{\theta}\right)^{-\delta} + \left(1 - \bar{v}^{\theta}\right)^{-\delta} - 1\right\}^{-1/\delta}\right]^{1/\theta}$$

where $\theta \geqslant 1$, $\delta > 0$, $\bar{u} = 1 - u$, $\bar{v} = 1 - v$.

- Some properties:
    - $\lambda_L = 2^{-1/\delta}$ does not depend on $\lambda_U = 2 - 2^{-1/\theta}$.
    - $\tau = 1 - 4\int_0^{\infty} s \times (\varphi'(s))^2 ds$ is calculated via Laplace transform.

Covariate-dependent copula models

↪ The reparameterized copula model

- **Reparametrization**: We reparameterize the copula as a function of tail-dependence and/or Kendall's tau $C(\boldsymbol{u}, \lambda_L, \tau)$.
- **Link with covariates**: All copula features in the k:th and l:th margins can be connected with covariates

$$\tau_{kl} = l_{\tau}^{-1}(\boldsymbol{X}_{kl}\beta_{\tau}),$$
$$\lambda_{kl} = l_{\lambda}^{-1}(\boldsymbol{X}_{kl}\beta_{\lambda})$$

- **Applicable Copulas**: Any copula can be equally well used with such reparameterization when there is closed form of tail-dependence and Kendall's τ.
    - **Archimedean copulas**: Joe-Clayton, Clayton, Gumbel,...
    - **Elliptical copulas**: Gaussian and t copulas
- Marginal models we have used
    - Mixture of asymmetric student's-t distributions.
    - GARCH models
    - stochastic volatility (SV) models.
    - Poisson regression models.

The Bayesian approach

- **The log Posterior**

$$\log p(\{\beta, \mathfrak{I}\}|\boldsymbol{y}, \boldsymbol{x}) = \mathrm{c} + \sum_{j=1}^{M} \{\log p(\boldsymbol{y}_{.j}|\{\beta, \mathfrak{I}\}_j, \boldsymbol{x}_j) + \log p(\{\beta, \mathfrak{I}_j\})\}$$
$$+ \log \mathcal{L}_C(\boldsymbol{u}_{1:M}|\{\beta, \mathfrak{I}\}_C, \boldsymbol{y}, \boldsymbol{x}) + \log p_C(\{\beta, \mathfrak{I}\})$$

where

- $\{\beta\}$ are the coefficient in the linking function,
- $\{\mathfrak{I}\}$ are the corresponding variable selection indicators.
- $\{\beta, \mathfrak{I}\}$ can be estimated jointly via Bayesian approach.
- $\boldsymbol{u}_j = F_j(y_j)$ is the CDF of the j:th marginal model.

The Bayesian approach

- **The priors** for the copula model are easy to specify due to our reparameterization.
    - It it **not easy** to specify priors directly on $\{\beta, \mathcal{I}\}$
    - But it is **easy** to puts prior information on the model parameters features (τ, μ, σ^2) and then derive the implied prior on the intercepts and variable selection indicators.
    - When variable selection is used, we assume there are no covariates in the link functions *a priori*.
- **The posterior** inference is straightforward although the model is very complicated.

$\looparrowright$ Sampling the posterior with an efficient MCMC scheme

- We update all the parameters **jointly** by using tailored Metropolis-Hastings within Gibbs. This is more efficient compared to the two-stage inference according to our study.
- **Taming the Beast:** the analytical gradients require the derivative for the copula density and marginal densities which can be conveniently decomposed via the chain rule that greatly reduces the complexity of the the gradient calculation.
- **Bayesian variable selection** is carried out simultaneously.
- The Gibbs sampler for covariate-dependent copula.

Margin component (1)	...	Margin component (M)	Copula component (C)
(1.1) $\{\beta_\mu, \mathcal{I}_\mu\}_1 \vert \{\beta_\mu, \mathcal{I}_\mu\}_{-1}$	...	$(M.1)$ $\{\beta_\mu, \mathcal{I}_\mu\}_M \vert \{\beta_\mu, \mathcal{I}_\mu\}_{-M}$	$(C.1)$ $\{\beta_\lambda, \mathcal{I}_\lambda\}_C \vert \{\beta_\lambda, \mathcal{I}_\lambda\}_{-C}$
(1.2) $\{\beta_\phi, \mathcal{I}_\phi\}_1 \vert \{\beta_\phi, \mathcal{I}_\phi\}_{-1}$	...	$(M.2)$ $\{\beta_\phi, \mathcal{I}_\phi\}_M \vert \{\beta_\phi, \mathcal{I}_\phi\}_{-M}$	$(C.2)$ $\{\beta_\tau, \mathcal{I}_\tau\}_C \vert \{\beta_\tau, \mathcal{I}_\tau\}_{-C}$
(1.3) $\{\beta_\nu, \mathcal{I}_\nu\}_1 \vert \{\beta_\nu, \mathcal{I}_\nu\}_{-1}$	...	$(M.3)$ $\{\beta_\nu, \mathcal{I}_\nu\}_M \vert \{\beta_\nu, \mathcal{I}_\nu\}_{-M}$	
(1.4) $\{\beta_\kappa, \mathcal{I}_\kappa\}_1 \vert \{\beta_\kappa, \mathcal{I}_\kappa\}_{-1}$	...	$(M.4)$ $\{\beta_\kappa, \mathcal{I}_\kappa\}_M \vert \{\beta_\kappa, \mathcal{I}_\kappa\}_{-M}$	

Model Comparison

- We evaluating the model performance based on **out-of-sample prediction**.
- In our time series application, we estimate the model based on the 80% of historical data and then predict the last 20% data.
- We evaluate the quality of the one-step-ahead predictions using the **log predictive score** (LPS)

$$
\begin{aligned}
\mathrm{LPS} =& \log p(D_{(T+1):(T+p)}|D_{1:T}) \\
=& \sum_{i=1}^{p} \log \int p(D_{T+i}|\theta, D_{1:(T+i-1)}) p(\theta|D_{1:(T+i-1)}) \mathrm{d}\theta
\end{aligned}
$$

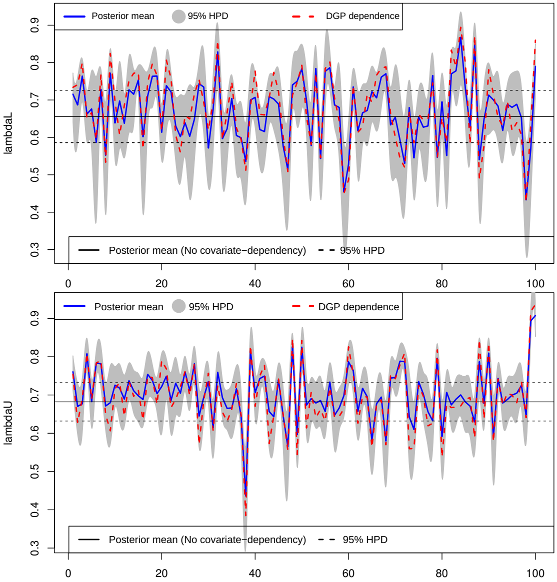
where $D_{a:b}$ is the dataset from time a to b and θ are the model parameters.

Simulation

Table 4: LPS of four-fold cross-validation for Joe-Clayton copula with 16 DGP settings and 64 simulations based on different combination of lower tail-dependence and upper tail-dependence, respectively. Each dataset consists of 1,000 observations with given mean ($\bar{\lambda}_L$ and $\bar{\lambda}_U$) and standard deviation (0.1) for lower and upper tail-dependences. Each dataset is estimated with four models (*J. + CD.*, *J. + Const.*, *T.+ CD.* and *T. + Const.*) and the LPS for the best model is marked in bold.

DGP settings		$\bar{\lambda}_U^{(DGP)} = 0.3$		$\bar{\lambda}_U = 0.5$		$\bar{\lambda}_U = 0.7$		$\bar{\lambda}_U = 0.9$	
	MCMC	*CD.*	*Const.*	*CD.*	*Const.*	*CD.*	*Const.*	*CD.*	*Const.*
$\bar{\lambda}_L^{(DGP)} = 0.3$	*J.*	**−519.56**	−520.91	**−506.90**	−508.95	**−427.72**	−432.35	**−273.93**	−306.99
	T.	−523.25	−522.00	−510.60	−511.75	−444.32	−439.68	−310.67	−321.38
$\bar{\lambda}_L = 0.5$	*J.*	**−501.33**	−502.57	**−468.30**	−471.97	**−424.30**	−436.54	**−244.02**	−268.56
	T.	−510.51	−507.29	−476.68	−474.30	−446.38	−451.83	−299.08	−314.36
$\bar{\lambda}_L = 0.7$	*J.*	**−440.81**	−454.16	**−424.20**	−439.24	**−380.30**	−390.38	**−243.16**	−244.78
	T.	−457.76	−460.83	−440.01	−440.70	−397.72	−402.37	−283.96	−295.11
$\bar{\lambda}_L = 0.9$	*J.*	**−228.83**	−256.11	**−218.61**	−294.52	**−241.21**	−255.13	**−210.11**	−269.86
	T.	−244.01	-294.00	−292.74	−317.60	−280.67	−289.88	−259.15	−297.25

Simulation



Posterior mean
95% HPD
DGP dependence
Posterior mean (No covariate−dependency)
95% HPD
lambdaL
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0
20
40
60
80
100
Posterior mean
95% HPD
DGP dependence
Posterior mean (No covariate−dependency)
95% HPD
lambdaU
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0
20
40
60
80
100

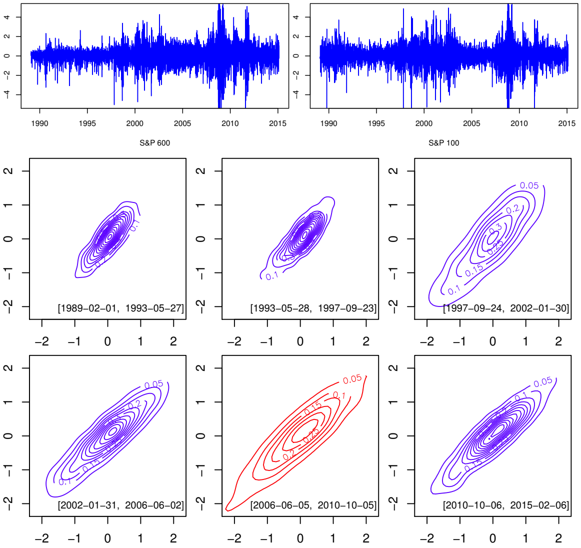
Improving forecasting performance with covariate-dependent tail-dependence

↪ Log predictive density score comparison

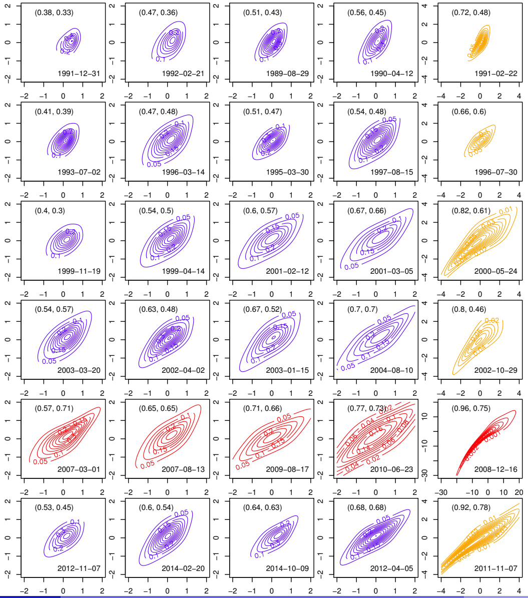
Margins	LPS decomposition	Reparameterized Copulas Joe-Clayton	Clayton	Gumbel	t-Copula
		(*Joint modeling approach*)			
SPLIT-t	M_1	−1743.12	−1741.04	−1754.36	−1741.47
	M_2	−1435.98	−1468.25	−1485.68	−1430.07
	$C(CD.)$	837.50	690.22	797.78	792.14
	Global	−2344.12	−2523.75	−2448.14	−2380.12
SPLIT-t	M_1	−1747.99	−1747.15	−1754.61	−1782.37
	M_2	−1434.22	−1449.95	−1446.84	−1658.09
	$C(Const.)$	779.14	654.46	780.33	703.96
	Global	−2411.06	−2547.14	−2421.15	−2736.49
		(*Two-stage modeling approach*)			
SPLIT-t	M_1	−1740.10	−1741.05	−1737.73	−1741.47
	M_2	−1428.39	−1436.63	−1427.83	−1433.41
	$C(CD.)$	819.63	694.84	781.39	788.22
	Global	−2346.61	−2483.93	−2392.13	−2389.41
GARCH(1,1)	M_1	−1948.07	−1948.07	−1948.07	−1948.07
	M_2	−1673.85	−1673.85	−1673.85	−1673.85
	$C(CD.)$	702.35	530.48	810.39	791.55
	global	−2919.57	-3091.44	−2811.53	−2830.37
SV	M_1	−2166.90	−2154.18	−2168.17	−2179.36
	M_2	−1811.36	−1844.57	−1808.61	−1808.24
	$C(CD.)$	964.37	698.30	1012.10	1053.19
	Global	−3013.90	−3300.46	−2964.68	−2934.40
		(*Bivariate volatility models*)			
Bivariate DCC-GARCH		−2730.78			
Bivariate SV		−2999.63			

Improving forecasting performance with covariate-dependent tail-dependence

↛ The S&P 100 and S&P 600 and their empirical copulas



Contour plots of the posterior densities



Working in progress

- Modeling multivariate covariate-dependent structure via the vine copula.
- Looking into more efficient inference techniques, VB?
- Including probabilistic topic model to model the texts margins.

Thank you!

yanfeikang@buaa.edu.cn

http://yanfei.site/