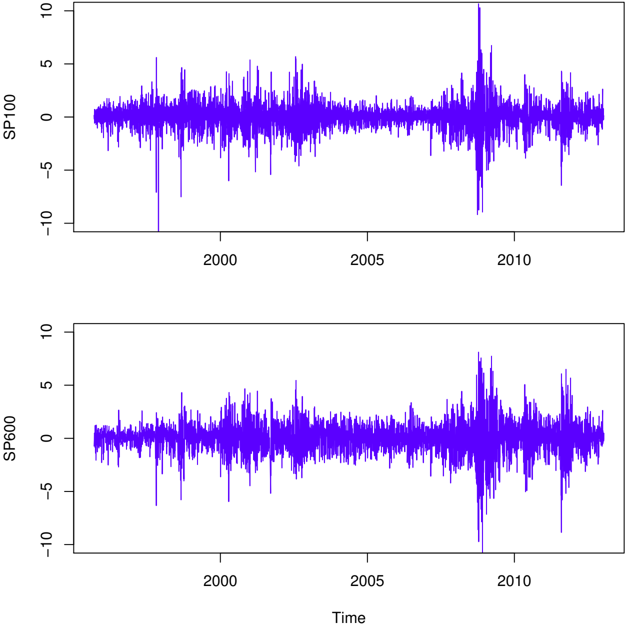
Dynamic Tail-Dependence and Correlation Modeling with Efficient Bayesian Approach

Feng Li

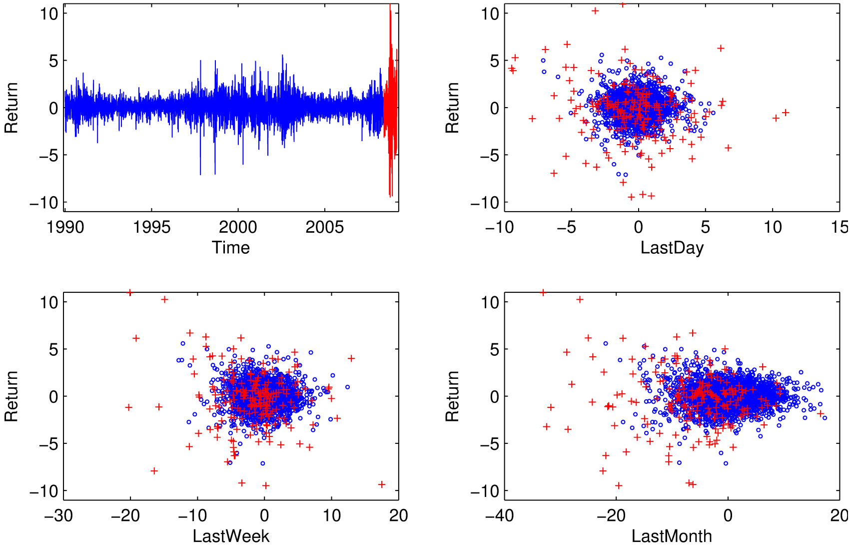
School of Statistics and Mathematics
Central University of Finance and Economics

Outline

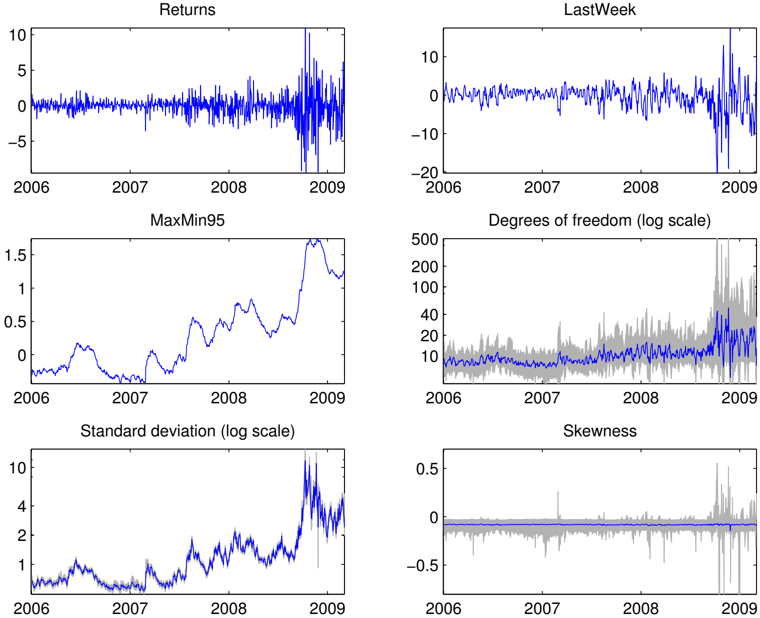
The stock market returns for two indices



Typical variables used in financial data



Individual Modeling the stock market returns



- 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**

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

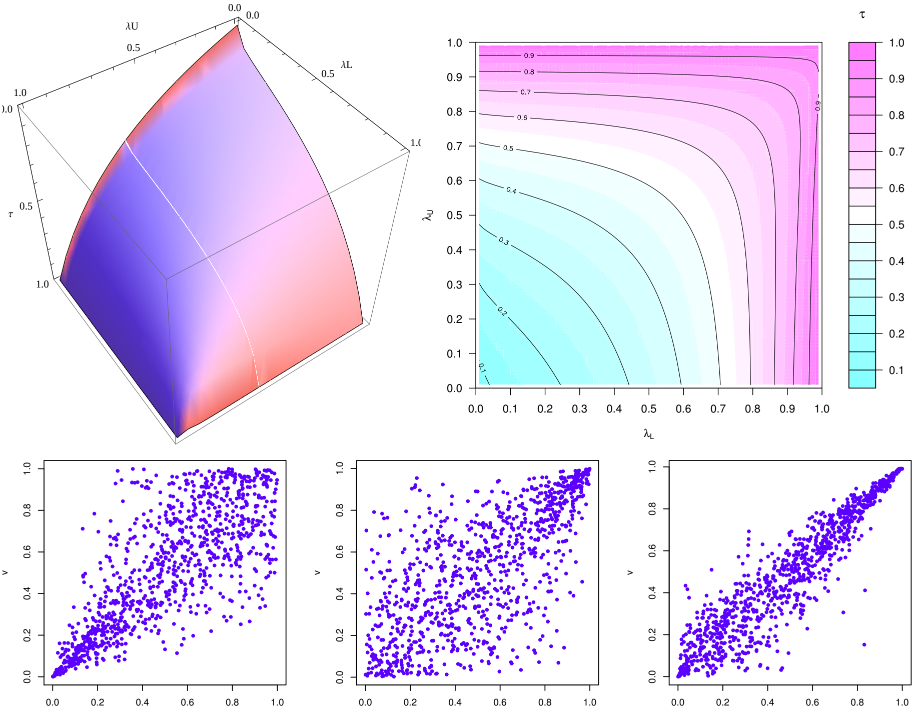
- Some facts:
  - The Kendall's τ is invariant w.r.t. **strictly** increasing transformations.
  - For all copulas in the elliptical class (Gaussian, *t*,...), $\tau = \frac{2}{\pi}\arcsin(\rho)$.
  - The Gaussian copula has zero tail dependence.
  - The student t copula has asymptotic upper tail dependence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

- The Joe-Clayton copula function

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

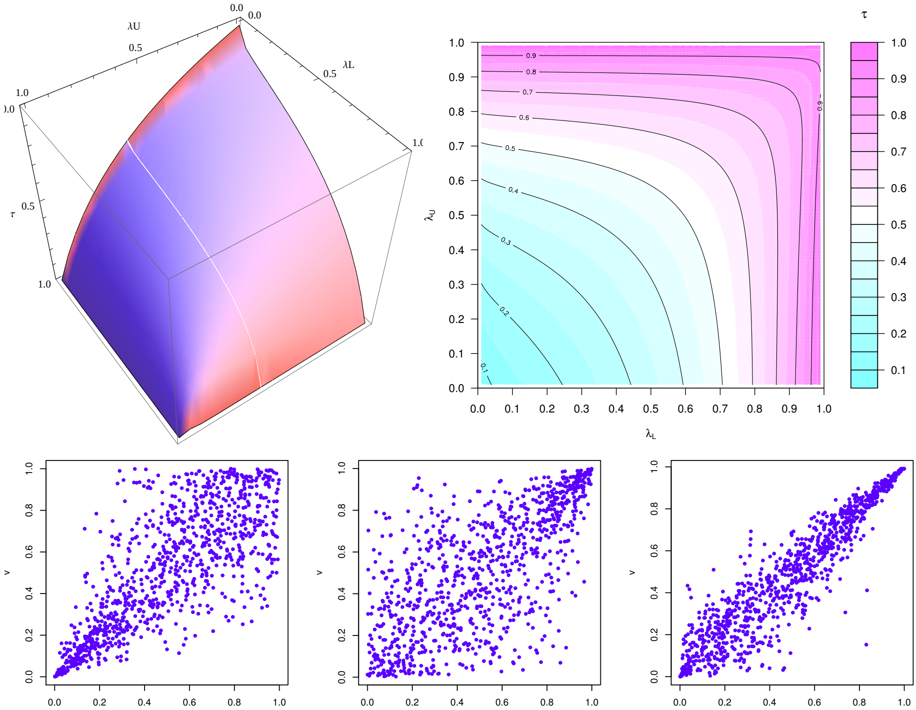
where $\theta \geqslant 1$, $\delta > 0$, $\bar{u} = 1 - u$, $\bar{v} = 1 - v$.

- 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.



λU
λL
τ
0.0
0.5
1.0
λ_U
λ_L
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
v

- **The motivation** i) The interpretation of correlation and tail-dependence. ii) Dynamical modeling tail-dependence and correlation.
- **Reparametrization**: We reparameterize copula as a function of tail-dependence and Kendall's tau $C(\mathbf{u}, \lambda_L, \tau)$.
- **Applicable Copulas**: Any copula can be equally well used with such reparameterization.
    - **Joe-Clayton Copula**: lower tail-dependence and upper tail-dependence are independent.
    - **Clayton Copula**: allow for modeling lower tail-dependence
    - **Gumbel Copula**: commonly used in extreme value theory.
    - **Multivariate *t* copula**: elliptical copula allows for tail-dependence with small df.



λU
λL
τ
λ_U
λ_L
v

The covariate-contingent copula model

Connecting density features with covariates

- All parameters are connected with covariates via known link function $\varphi(\cdot)$, (identity, log, logit, probit,...)

Components	Features	Linkage
Margins	mean	$\mu = \varphi_{\beta_u}^{-1}(X_u\beta_u)$,
	variance	$\sigma^2 = \varphi_{\beta_\sigma}^{-1}(X_\sigma\beta_\sigma)$,
	df	$\nu = \varphi_{\beta_\nu}^{-1}(X_\nu\beta_\nu)$,
	skewness	$s = \varphi_{\beta_s}^{-1}(X_s\beta_s)$,
Copula	lower tail-dependence	$\lambda_L = \varphi_{\lambda}^{-1}((X_u, X_\nu)\beta_{\lambda_L})$,
	upper tail-dependence	$\lambda_U = \varphi_{\lambda}^{-1}((X_u, X_\nu)\beta_{\lambda_u})$,
	Kendall's τ	$\tau = \varphi_{\tau}^{-1}((X_u, X_\nu)\beta_\tau)$.
	Covariance Matrix*	$\Sigma = \Sigma_0 + \kappa I$ where $\text{vech}(\mathbf{\Sigma_0}) = \varphi^{-1}([\mathbf{I} \otimes \mathbf{X}]\text{vec}\mathbf{B})$

* Cholesky decomposition (Huang et al., 2007) is possible but not interpretation friendly.

The covariate-contingent copula model I

↪ The Bayesian approach

- **The marginal models**
  - In principle, any combination of univariate marginal models can be used.
  - When there are discrete margins, data augmentation method can be used (Smith and Khaled, 2012).
  - We develop **R** package to allow for
    - mixtures of elliptical distributions (Li et al., 2010)
    - regression spline where the knots locations are treated as unknown parameters (Li and Villani, 2013).
  - In the continuous case, we use univariate model that each margin is from the student *t* distribution (Li et al., 2010).
- **The log Posterior**

$$\log p(\{\beta, \mathcal{I}\}|y, x) = c + \sum_{j=1}^{M} \left\{\log p(y_{.j}|\{\beta, \mathcal{I}\}_j, x_j) + \log p(\{\beta, \mathcal{I}_j\})\right\} + \log \mathcal{L}_C(u_{1:M}|\{\beta, \mathcal{I}\}_C, y, x) + \log p_C(\{\beta, \mathcal{I}\})$$

where

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

- **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 $(\tau,\ \mu,\ \sigma^2)$ 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.

The dynamic copula model I

↪ Sampling the posterior with an efficient MCMC scheme

- We update all the parameters **jointly** by using tailored Metropolis-Hastings within Gibbs.
- The proposal density for each parameter vector β is a multivariate t-density with $\mathrm{df} > 2$,

$$\beta_p|\beta_c \sim \mathbf{MVT}\left[\hat{\beta},\ -\left(\frac{\partial^2 \ln p(\beta|\mathbf{Y})}{\partial\beta\partial\beta'}\right)^{-1}\Bigg|_{\beta=\hat{\beta}},\ \mathrm{df}\right],$$

  where $\hat{\beta}$ is obtained by R steps ($R \leqslant 3$) Newton's iterations during the proposal with analytical gradients.
- This approach has some flavor of Hamiltonian MC when $R = 1$ (Thanks Rong Chen for pointing this out).
- **Bayesian variable selection** is carried out simultaneously.

The dynamic copula model II

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

- The Gibbs sampler for covariate-dependent copula.
- The notation $\{\beta_\mu, \mathcal{I}_\mu\}_{-m}$ indicates all other parameters in the model except $\{\beta_\mu, \mathcal{I}_\mu\}_m$. The updating order is column-wise from left to right. If dependent link functions are used, the updating should be ordered accordingly.

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) $\{\beta_\kappa, \mathcal{I}_\kappa\}_1 \vert \{\beta_\kappa, \mathcal{I}_\kappa\}_{-1}$	...	$(M.4)$ $\{\beta_\kappa, \mathcal{I}_\kappa\}_M \vert \{\beta_\kappa, \mathcal{I}_\kappa\}_{-M}$	

The dynamic copula model I

↬ The computational details

- **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.

$$
\begin{aligned}
\frac{\partial \log c(u_{1:M}, \lambda_L, \tau)}{\partial \lambda_L} =& \frac{\partial \log c(u_{1:M}, \theta, \delta)}{\partial \delta} \times \left(\frac{\partial \lambda_L}{\partial \delta}\right)^{-1} \\
&+ \frac{\partial \log c(u_{1:M}, \theta, \delta)}{\partial \theta} \times \left(\frac{\partial \lambda_L}{\partial \theta}\right)^{-1} \\
\frac{\partial \log c(u_{1:M}, \lambda_L, \tau)}{\partial \tau} =& \frac{\partial \log c(u_{1:M}, \theta, \delta)}{\partial \theta} \times \left(\frac{\partial \tau(\theta, \delta)}{\partial \theta}\right)^{-1} \\
&+ \frac{\partial \log c(u_{1:M}, \theta, \delta)}{\partial \delta} \times \left(\frac{\partial \tau(\theta, \delta)}{\partial \delta}\right)^{-1} \\
\frac{\partial \log c(u_{1:M}, , \lambda_L, \tau)}{\partial \varphi_m} =& \frac{\partial \log c(u_{1:M}, , \theta, \delta)}{\partial u_m} \times \frac{\partial u_m}{\partial \varphi_m} \\
&+ \frac{\partial \log p_m(y_m, \varphi_m)}{\partial \varphi_m}
\end{aligned}
$$

The dynamic copula model II

↪ The computational details

- The direct derivatives of CDF function and PDF functions with respect to their parameters are straightforward for most densities.
- Existing derivatives for PDF functions in marginal models:
  - Li et al. (2010) (mixtures of asymmetric student-t densities where asymmetric normal and symmetric student-t densities are its special cases),
  - Li et al. (2011) (gamma and log-normal models)
  - Villani et al. (2012) (negative binomial, beta and generalized Poisson models)
  - Li and Villani (2013) (spline model with knots location as unknown parameters). densities)
- Li (2015, JBES forthcoming) (derivatives for Joe-Clayton copula, Gumbel copula and multivariate t copula).

- **The bad news**: Evaluating the gradients are very time consuming if we do it sequentially, e.g.
    - when t copula is used, the tail-dependence for ith and jth margins (λ_{Lij}) are (Embrechts et al., 1997)

$$\lambda_{Lij} = \frac{\int_{\pi/4-\arcsin(\rho_{ij})/2}^{\pi/2} \cos^{\nu}(t)dt}{\int_0^{\pi} \cos^{\nu}(t)dt}$$

and ρ_{ij} is the correlation coefficient for ith and jth margins.

    - Kendall's τ of the Joe-Clayton copula is of the form

$$\tau(\theta,\delta) = \begin{cases} 1 - 2/[\delta(2-\theta)] + 4B\left(\delta+2, 2/\theta-1\right)/(\theta^2\delta), & 1 \leqslant \theta < 2; \\ 1 - \left[\psi(2+\delta) - \psi(1) - 1\right]/\delta, & \theta = 2; \\ 1 - 2/[\delta(2-\theta)] & \theta > 2 \\ \quad - 4\pi/\left[\theta^2\delta(2+\delta)\sin(2\pi/\theta)B\left(1+\delta+2/\theta, 2-2/\theta\right)\right], & \end{cases}$$

The dynamic copula model IV

↪ The computational details

- **The good news**: the gradient can be evaluated parallelly because we assume the observations are independent.
- Our parallel version code running on a 16-core CPU can speed up the computation at least **10X**.
- The code is written in **R** and is running on a Linux cluster with 80 cores and total 1TB RAM.
- We recompile R with Intel MKL library that greatly speed up the numerical computations.
- A rich class of multivariate models is implemented.
- Our tailored Metropolis-Hastings keeps the overall acceptance probability above **80%**.

Why not two-stage approach?

- The asymptotic relative efficiency of the two-stage estimation procedure depends on how close the copula is to the Fréchet bounds (Joe, 2005).
- The two-stage approach in estimating the multivariate DCC GARCH model is consistent but not fully efficient due to the limited information provided by the estimators (Engle and Sheppard, 2001).

Model Comparison

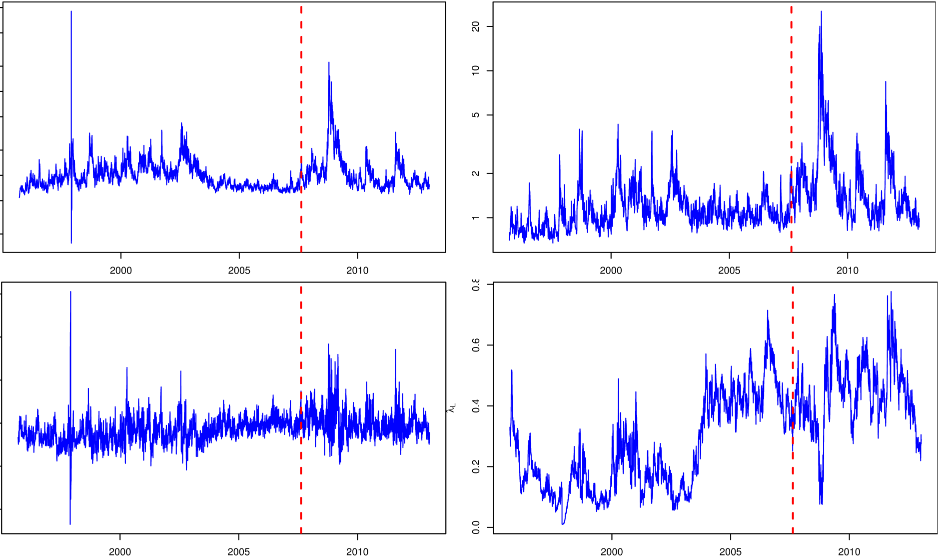
- We evaluating the model performance based on **out-of-sample prediction**.
- In our time series application, we estimate the model based on the 80% of historical data and then predict the last 20% data.
- We evaluate the quality of the one-step-ahead predictions using the **log predictive score** (LPS)

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

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

The stock returns, a revisit

The Kendall's τ (left column) and tail-dependence (right column) over time for SP100 (top row) and SP600 (bottom row)



The posterior copula plot

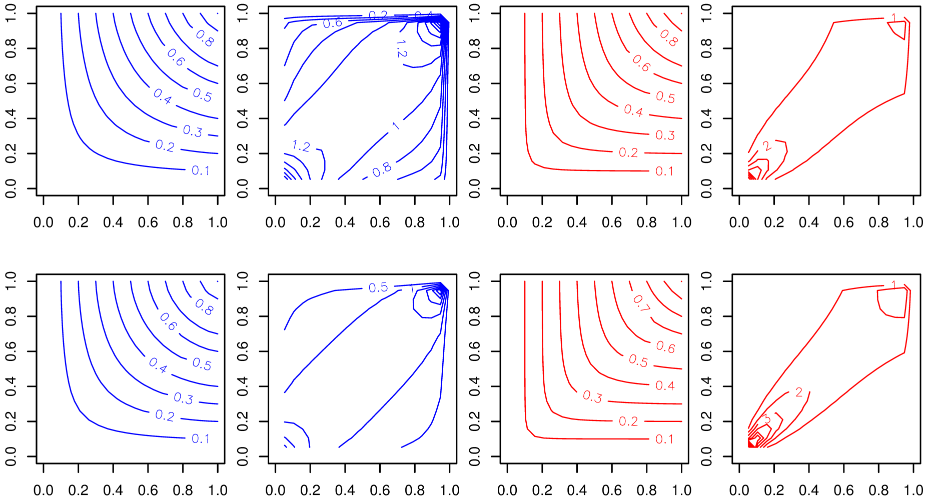


Table: Posterior summary of copula model with S&P100 and S&P600 data. In the copula component part, the first row and second row for β and $\mathcal{I}$ are the results for the combined covariates that are used in the first and second marginal model, respectively. The intercept are always included in the model.

	Intercept	RM1	RM5	RM20	CloseAbs95	CloseAbs80	MaxMin95	MaxMin80	CloseSqr95	CloseSqr80
					Copula component ($\mathcal{C}$)					
β_{λ_L}	−8.165	**−0.555**	1.793	**0.005**	−0.170	**0.110**	−0.667	−1.448	−0.636	0.050
		1.463	**0.405**	0.934	−2.138	−1.288	−1.954	−1.577	**−1.873**	−1.805
$\mathcal{I}_{\lambda_L}$	1.00	**0.98**	0.37	**0.63**	0.02	**0.61**	0.36	0.35	0.39	0.29
		1.00	**1.00**	0.00	0.30	0.35	0.40	0.00	**0.61**	0.34
β_{τ}	−1.726	0.181	−0.217	−0.304	−0.107	**0.115**	**0.005**	**−0.257**	**1.068**	0.037
		−0.191	**0.170**	0.274	0.144	**−0.051**	**−0.671**	0.059	−0.209	−0.181
$\mathcal{I}_{\tau}$	1.00	0.00	0.00	0.00	0.00	**0.90**	**0.99**	**1.00**	**0.85**	0.00
		1.00	**1.00**	0.00	0.00	**1.00**	**1.00**	0.00	0.00	0.00

The inefficiency factors for the parameters are all bellow 25.

Extensions and future work

1. We are working to extend the model in text mining and financial modeling that allow some margins are continuous but some are discrete.
   - One margin: A company's stock credited as A, A^{+} over time by Standard & Poor's.
   - The other margin: the stock returns over time
2. Efficient approximation of the posterior via sub-sampling to handle much bigger data. Several Big Data MCMC approaches have been already considered in Welling and Teh (2011), Korattikara et al. (2013), Teh et al. (2014), Bardenet et al. (2014), Maclaurin and Adams (2014), Minsker et al. (2014), Quiroz et al. (2014) and Strathmann et al. (2015) but not in such general model.

References I

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997), *Modelling extremal events*, vol. 33, Springer Science & Business Media.

Engle, R. F. and Sheppard, K. (2001), "Theoretical and Empirical properties of Dynamic Conditional Correlation Multivariate GARCH," *Working Paper*, 1–46.

Huang, J. Z., Liu, L., and Liu, N. (2007), "Estimation of large covariance matrices of longitudinal data with basis function approximations," *Journal of Computational and Graphical Statistics*, 16, 189–209.

Joe, H. (2005), "Asymptotic efficiency of the two-stage estimation method for copula-based models," *Journal of Multivariate Analysis*, 94, 401–419.

Li, F. and Villani, M. (2013), "Efficient Bayesian Multivariate Surface Regression," *Scandinavian Journal of Statistics*, 40, 706–723.

Li, F., Villani, M., and Kohn, R. (2010), "Flexible modeling of conditional distributions using smooth mixtures of asymmetric student t densities," *Journal of Statistical Planning and Inference*, 140, 3638–3654.

References II

— (2011), "Modeling Conditional Densities Using Finite Smooth Mixtures," in *Mixtures: estimation and applications*, eds. Mengersen, K., Robert, C., and Titterington, M., John Wiley & Sons Inc, Chichester, pp. 123–144.

Smith, M. and Khaled, M. (2012), "Estimation of Copula Models with Discrete Margins via Bayesian Data Augmentation," 107, 290–303.

Villani, M., Kohn, R., and Nott, D. J. (2012), "Generalized Smooth Finite Mixtures," 171, 121–133.

Thank you!

feng.li@cufe.edu.cn

http://feng.li/