Bayesian Modeling Tail-Dependence *of* Stock Returns and News Sentiment with Copulas

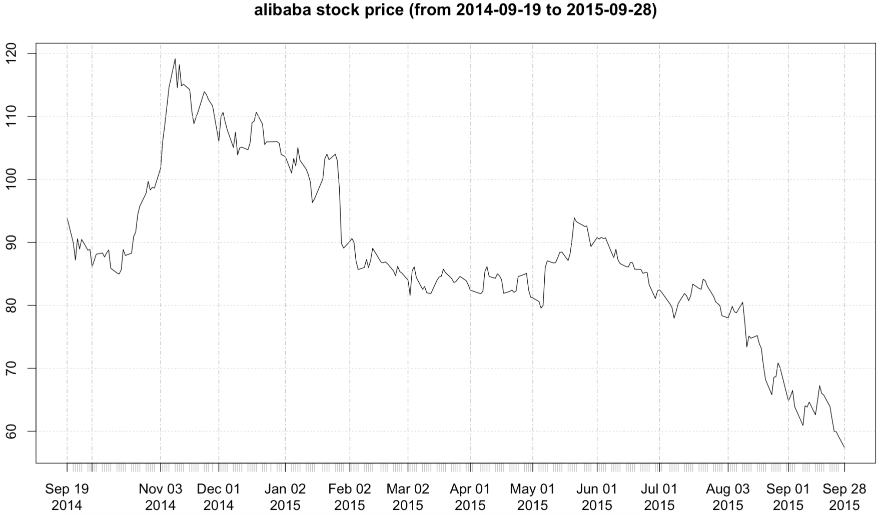
Feng Li

School of Statistics and Mathematics
Central University of Finance and Economics

Revised on July 27, 2016

Outline

The daily stock returns for Alibaba



Text information related to Alibaba

财新网 Caixin.com

Search

阿里巴巴 搜索 帮助？

您搜索 **阿里巴巴** 获得大约 **647** 条查询结果，以下是第 **1 - 20** 条。 （搜索用时 0.18117423 秒）

所有结果
财新网
杂志
图片
博客
视听
数字说
专题

时间不限
一天内
一周内
一月内
一年内
自定日期范围

全文
正文
作者
标题

联姻东风日产不到两年 恒大三次提出回购要求（要闻）[2015-11-27]

2月1日至2016年1月31日。 2014年6月，恒大俱乐部增资扩股引入阿里巴巴。阿里巴巴（中国）网络技术有限公司注资12亿，持有恒大足球俱乐部40%股权。广州恒大队更名为广州恒大淘宝队。 2014年

阿里影业积攒超级IP 年产电影三部左右（TMT）[2015-11-26]

的整条电影产业链。此前阿里影业还收购了粤科软件，阿里巴巴还对优酷土豆发起私有化邀约，也有意与阿里影业产生协同，如何贯通业务是阿里影业首先要解决问题。 C2B影视娱乐内容投融资平台娱乐宝就是一个众筹平台

互联网跨界融合中的"共生金融"新模式（王永利）[2015-11-26]

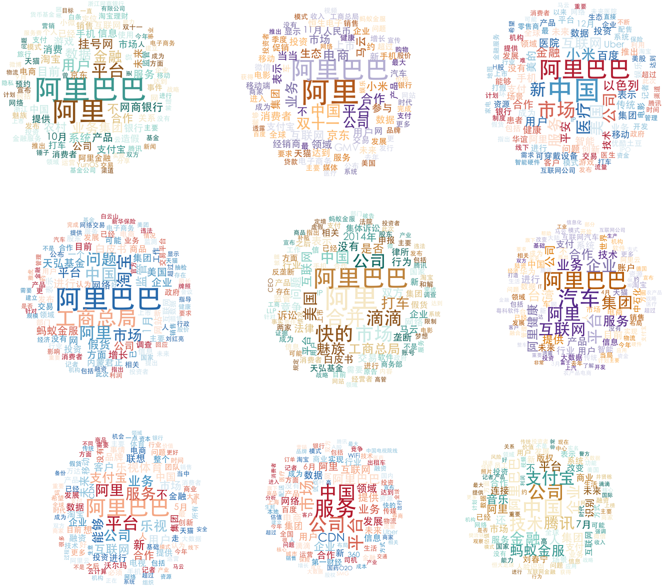
，不仅新兴的互联网企业（如阿里巴巴、腾讯、百度等）依托其核心竞争力不断拓展经营范围，形成跨业经营的共生经济模式，而且一些商业物流企业、房地产企业等也运用互联网技术进行改造，加快共生经济模式的发展（如京

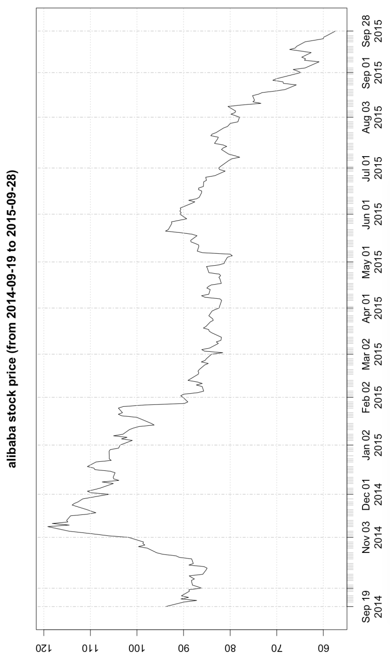
中国优步瞄准企业服务和二三线市场（TMT）[2015-11-26]

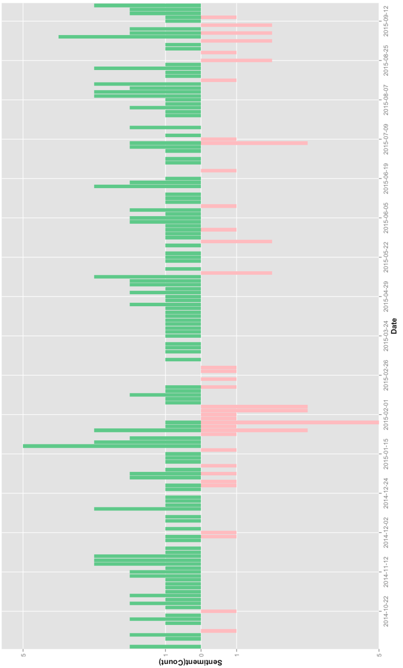
。"2015年加满油、蓄势待发，2016年才是真正全速前进的一年。"柳甄说。就在发布会前一天，她刚与阿里巴巴旗下钉钉签署合作协议，钉钉的员工将通过优步实现上下班通勤等，并正式宣布针对企业客户的U4B（即

- 647 new articles about Alibaba from Sep 23, 2014 to Sep 22, 2015.

The evolution of topics in a sequentially organized corpus of news




alibaba stock price (from 2014-09-19 to 2015-09-28)
120
110
100
90
80
70
60
Sep 19 2014
Nov 03 2014
Dec 01 2014
Jan 02 2015
Feb 02 2015
Mar 02 2015
Apr 01 2015
May 01 2015
Jun 01 2015
Jul 01 2015
Aug 03 2015
Sep 01 2015
Sep 28 2015


Sentiment(Count)
Date
2014-10-22
2014-11-12
2014-12-02
2014-12-24
2015-01-15
2015-02-01
2015-02-26
2015-03-24
2015-04-29
2015-05-22
2015-06-05
2015-06-19
2015-07-09
2015-08-07
2015-08-25
2015-09-12

Individual modeling the stock market returns

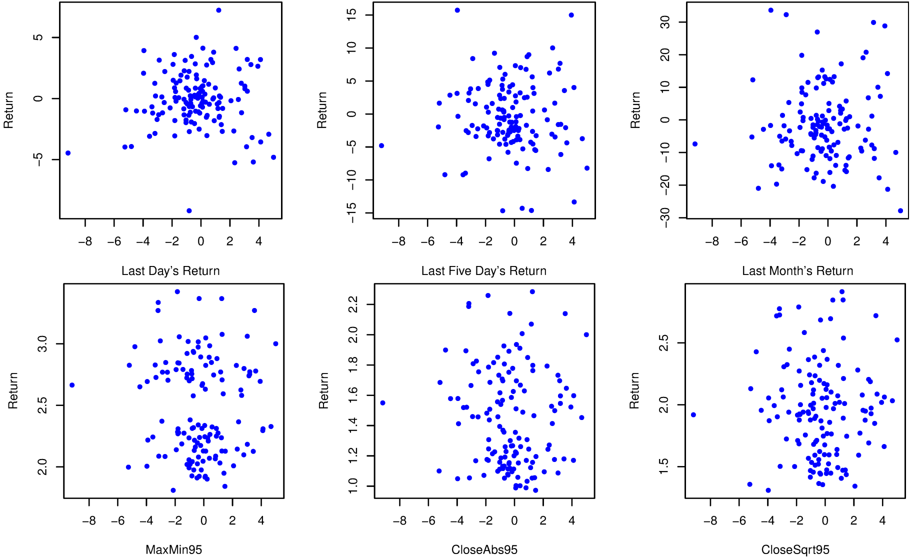


Figure: Covariates used to capture stock's changes

Use smooth mixture of asymmetric student's t densities to model stock returns

- The split-t density is

$$c \cdot \kappa(\mu, \phi, \nu) I(y \leqslant \mu) + c \cdot \kappa(\mu, \lambda\phi, \nu) I(y > \mu),$$

where $\kappa(\mu, \phi, \nu) = \left(\frac{\nu}{\nu + \frac{(y-\mu)^2}{\phi^2}}\right)^{(\nu+1)/2}$ is the kernel of student t density and c is the normalization constant.

- Each of the four parameters μ, ϕ, λ and ν are connected to covariates as

$$\begin{aligned} \mu &= \beta_{\mu 0} + x_t' \beta_\mu \\ \ln \phi &= \beta_{\phi 0} + x_t' \beta_\phi \\ \ln \lambda &= \beta_{\lambda 0} + x_t' \beta_\lambda \\ \ln \nu &= \beta_{\nu 0} + x_t' \beta_\nu \end{aligned}$$

but any smooth link function can equally well be used in the MCMC methodology.

- This make it possible e.g. to have the degrees of freedom smoothly

Individual modeling text information

- We obtain full articles for Alibaba Inc. from financial news site caixin.com with web scraping techniques. Covariates used in Poison model are financial key words appeared in those articles.

Date	P	N	U	上涨	下跌	打击	合作	增加	影响	违法
2014-10-15	2	0	1	0	0	1	3	0	1	0
2014-11-19	3	0	1	0	0	1	3	0	1	0
2015-01-28	3	3	3	0	0	2	5	1	2	4
2015-01-29	1	1	2	1	2	1	1	1	2	1
2015-01-30	1	7	2	0	1	3	1	1	5	5
2015-07-08	2	3	2	0	2	1	2	2	2	0

Figure: Covariates used to capture stock's changes

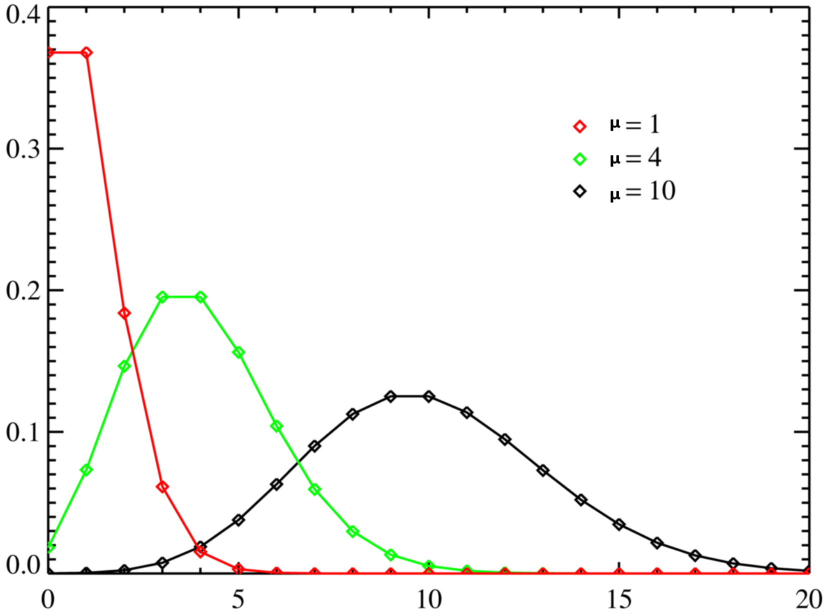
Individual modeling text information

- A corpus about Alibaba are built sorted by date and makrked with P/N/U.
- The vocabulary consists of 703 key words.
- We model the texts with **Poisson regression** for illustrative purpose.

$$\Pr(X{=}k) = \frac{\lambda^k e^{-\lambda}}{k!},$$

  where $\lambda = exp(x'\beta)$

- This is an $n < p$ problem. An efficient Bayesian variable selection algorithm is used.
- Other types of models e.g. negative binomial regression (Villani et al., 2012), dynamic topic models (Blei and Lafferty, 2006) are possible.



0.4
0.3
0.2
0.1
0.0
0
5
10
15
20
μ = 1
μ = 4
μ = 10

The Poisson model I

- We model the mean value (positive) of Y_i with covariates X_1, X_2,...,X_k

$$\mu_i = E(Y_i) = exp(\beta_1 + \beta_2 X_2 + ... + \beta_k X_k)$$

or alternatively we write the model as

$$p(y_i) = \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!}, \text{ where } \mu_i = \beta_1 + \beta_2 X_2 + ... + \beta_k X_k$$

- The interpretation of the model
  - How frequently the event happens to ith observations on average?

$$\mu_i = E(Y_i) = exp(\beta_1 + \beta_2 X_2 + ... + \beta_k X_k)$$

  - What is the probability the event happens exactly y_i times to ith observation?

$$p(y_i) = \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!}$$

The Poisson model II

- What is the probability the event happens at most y_i times to ith observation?

$$\sum_{l=0}^{y_i} \frac{\mu_i^l e^{-\mu_i}}{l!}$$

- What is the probability the event happens at lest y_i times to ith observation?

$$1 - \sum_{l=0}^{y_i} \frac{\mu_i^l e^{-\mu_i}}{l!}$$

- Estimate the Poisson model with maximum likelihood method
- The likelihood

$$p(y_1, y_2, ...y_n) = \prod_{i=1}^{n} p(y_i)$$

The Poisson model III

- The log likelihood

$$\log p(y_1, y_2, ...y_n) = \sum_{i=1}^{n} \log p(y_i) = \sum_{i=1}^{n} [y_i \log(u_i) - u_i - \log(y_i!)]$$
$$= \sum_{i=1}^{n} [y_i(\beta_1 + \beta_2 X_2 + ... + \beta_k X_k)$$
$$- \exp(\beta_1 + \beta_2 X_2 + ... + \beta_k X_k) - \log(y_i!)]$$

- Then maximize $\log p(y_1, y_2, ...y_n)$ with respect to $\beta_1, \beta_2, ..., \beta_k$

Is that enough?

- Are there any correlations between news information and stock returns?
- Does there exist a way to joint two models, say one is discrete and the other one is continuous?
- Can we find the co-movement between news and stocks?

- 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.

- **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.

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	$\mu = \varphi^{-1}_{\beta_u}(X_u \beta_u)$,
	variance	$\sigma^2 = \varphi^{-1}_{\beta_\sigma}(X_\sigma \beta_\sigma)$,
	df	$\nu = \varphi^{-1}_{\beta_\nu}(X_\nu \beta_\nu)$,
	skewness	$s = \varphi^{-1}_{\beta_s}(X_s \beta_s)$,
Copula	lower tail-dependence	$\lambda_L = \varphi^{-1}_{\lambda}((X_u, X_\nu)\beta_{\lambda_L})$,
	upper tail-dependence	$\lambda_U = \varphi^{-1}_{\lambda}((X_u, X_\nu)\beta_{\lambda_u})$,
	Kendall's τ	$\tau = \varphi^{-1}_{\tau}((X_u, X_\nu)\beta_\tau)$.
	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 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

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

The 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 dynamic copula model III

The computational details

- **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

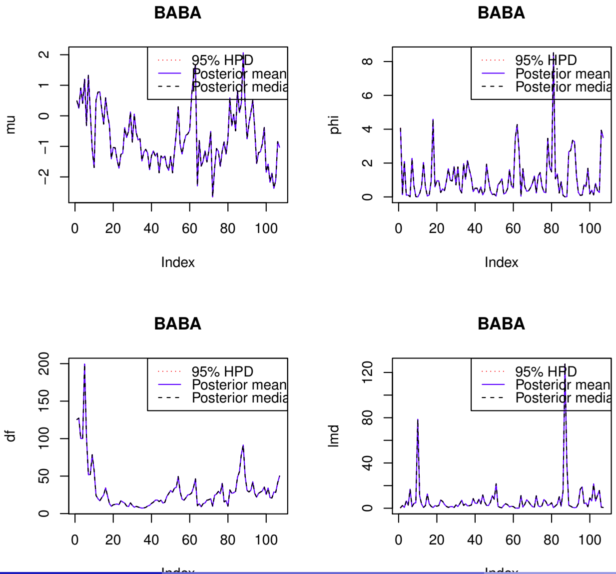
- **The good news**: the gradient can be evaluated parallelly because we assume the observations are independent.
- Our parallel version code running on a 16-core CPU can speed up the computation at least **10X**.
- The code is written in **R** and is running on a Linux cluster with 80 cores and total 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%**.

The stock returns, a revisit

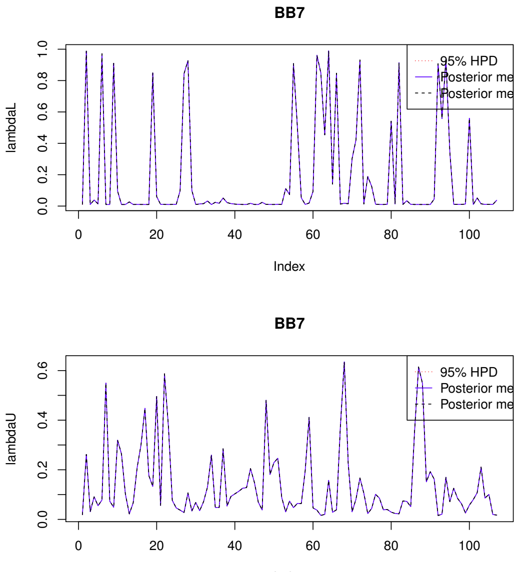
Model comparison

Components Combination $(M_1 + M_2 + C)$		Copulas (*reparameterizations*) Joe-Clayton $(\lambda_L,\ \lambda_U)$	Clayton (τ)	Gumbel (τ)	*t*-Copula $(\tau,\ \nu)$
		(*joint modeling approaches*)			
SPLIT-*t*	M_1	−1743.12	−1741.04	−1754.36	−1741.47
Poison	M_2	−1435.98	−1468.25	−1485.68	−1430.07
	C_x	837.50	690.22	797.78	792.14
	Joint	−**2344.12**	−2523.75	−2448.14	−2380.12
SPLIT-*t*	M_1	−1747.99	−1747.15	−1754.61	−1782.37
Poison	M_2	−1434.22	−1449.95	−1446.84	−1658.09
	C_0	779.14	654.46	780.33	703.96
	Joint	−**2411.06**	−2547.14	−2421.15	−2736.49
		(*two-stage modeling approaches*)			
SPLIT-*t*	M_1	−1740.10	−1741.05	−1737.73	−1741.47
Poison	M_2	−1428.39	−1436.63	−1427.83	−1433.41
	C_x	819.63	694.84	781.39	788.22
	Joint	−**2346.61**	−2483.93	−2392.13	−2389.41
GARCH	M_1	−1948.07	−1948.07	−1948.07	−1948.07
Poison	M_2	−1673.85	−1673.85	−1673.85	−1673.85
	C_x	702.35	530.48	810.39	791.55
	Joint	−2919.57	-3091.44	−2811.53	−2830.37
SV	M_1	−2166.90	−2154.18	−2168.17	−2179.36
Poison	M_2	−1811.36	−1844.57	−1808.61	−1808.24
	C_x	964.37	698.30	1012.10	1053.19
	Joint	−3013.90	−3300.46	−2964.68	−2934.40
		(*bivariate volatility models*)			
DCC-GARCH	−2730.78				
SV	−2999.63				

The posterior mean, variance, skewness and kurtosis for Alibaba stock returns



The lower tail dependence λ_L (up) and upper tail-dependence (down) over time for Alibaba stock returns and its news (bottom row)



Extensions and future work

1. We are working to extend the model for high-dimensional response variables.
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

Blei, D. M. and Lafferty, J. D. (2006), "Dynamic topic models," in *Proceedings of the 23rd international conference on Machine learning*, ACM, pp. 113–120.

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.

References II

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.

— (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," *Journal of the American Statistical Association*, 107, 290–303.

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

Thank you!

feng.li@cufe.edu.cn

http://feng.li/