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

Feng Li, Central University of Finance and Economics, Beijing 100081, China

Email: feng.li@cufe.edu.cn

Web: http://feng.li/

Contributions

Tail-dependence modeling based on copula with flexible marginal distributions is widely used in financial time series. Most of the available copula approaches for estimating tail-dependence are restricted within certain types of bivariate copulas due to computational complexity. We propose a general bayesian approach for jointly modeling high-dimensional tail-dependence for financial returns and related news information.

Our method allows for variable selection among the key words in news in the copula tail-dependence parameters. We apply an efficient sampling technique into the posterior inference where the likelihood function is estimated from a random subset of the data, resulting in substantially fewer density MCMC evaluations.

Modeling news sentiment

- A corpus about *Alibaba* is built sorted by date and labeled with Positive(P)/Negative(N)/Unknown(U).
- The vocabulary consists of 703 key words.
- We model the number of positive articles with **Poisson regression**

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

  where $\lambda = \exp(x'\beta)$ and x represents the words used in the news.
- This is an $n < p$ problem. An efficient Bayesian variable selection algorithm is used.
- Bayesian data augmentation (Smith and Khaled, 2012)is used for discrete marginal modeling.
- Other types of models, e.g., negative binomial regression (Villani et al., 2012), dynamic topic models (Blei and Lafferty, 2006) are applicable.

Modeling stock returns

- Use smooth mixture of asymmetric student's t densities to model stock returns (Li et al., 2010)
- Each of the four parameters μ, ϕ, λ and ν is connected to covariates as

$$\mu = \beta_{\mu 0} + x_t'\beta_\mu;\ \log\phi = \beta_{\phi 0} + x_t'\beta_\phi;$$
$$\log\lambda = \beta_{\lambda 0} + x_t'\beta_\lambda;\ \log v = \beta_{v0} + x_t'\beta_v.$$

- This makes it possible, e.g., to have the degrees of freedom smoothly varying over covariate space; to capture skewness and excess kurtosis with the mixing components.

Stocks & news sentiment dependence modeling

- **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 $C(\boldsymbol{u}, \lambda_L, \tau)$.
- **Copulas used**: i) *Joe-Clayton Copula*: lower tail-dependence and upper tail-dependence are independent. ii) *Clayton Copula*: allow for modeling lower tail-dependence iii) *Gumbel Copula*: commonly used in extreme value theory. iv) *Multivariate t copula*: elliptical copula allows for tail-dependence with small df.

Remarks and Extensions

- The code is written in **R** and is run on a Linux cluster with 96 cores and 5TB RAM in total.
- Computer code of this paper is available at `http://bitbucket.org/fli/`.
- We recompile R with Intel MKL library that greatly speeds up the numerical computation. Parallel computing of the analytical gradient is also implemented.
- A rich class of multivariate models are implemented.
- Our tailored Metropolis-Hastings keeps the overall acceptance probability above **80%**.
- It is possible to extend the model to high dimensional situation where multiple stocks and their news sentiment are considered jointly with vine copula (Panagiotelis et al., 2012).
- High dimensional tail-dependence can be constructed via conditional structure (Joe et al., 2010).

References

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.

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

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

Joe, H., Li, H., and Nikoloulopoulos, A. K. (2010), "Tail dependence functions and vine copulas," *Journal of Multivariate Analysis*, 101, 252–270.

Li, F. (2015), "Modeling covariate-contingent correlation and tail-dependence with copulas," *Jornal of Business and Economic Statistics*, forthcoming.

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.

Panagiotelis, A., Czado, C., and Joe, H. (2012), "Pair copula constructions for multivariate discrete data," *Journal of the American Statistical Association*, 107, 1063–1072.

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.

The posterior inference

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

Features	Linkage
lower tail-dependence	$\lambda_L = \varphi_\lambda^{-1}((X_u, X_v)\beta_{\lambda_L})$,
upper tail-dependence	$\lambda_U = \varphi_\lambda^{-1}((X_u, X_v)\beta_{\lambda_u})$,
Kendall's τ	$\tau = \varphi_\tau^{-1}((X_u, X_v)\beta_\tau)$.
Covariance Matrix*	$\Sigma = \Sigma_0 + \kappa I$ where $\mathrm{vech}(\boldsymbol{\Sigma_0}) = \varphi^{-1}([\boldsymbol{I} \otimes \boldsymbol{X}]\mathrm{vec}\boldsymbol{B})$

- **The priors** for the copula model are easy to specify due to our reparameterization.
  - It it **not easy** to specify priors directly on $\{\boldsymbol{\beta}, \boldsymbol{\mathcal{I}}\}$.
  - But it is **easy** to put 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 log Posterior** $\log p(\{\boldsymbol{\beta}, \boldsymbol{\mathcal{I}}\}|\boldsymbol{y}, \boldsymbol{x})$

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

- We follow Li (2015) and update all the parameters **jointly** by using tailored Metropolis-Hastings within Gibbs .

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 proposal density for each parameter vector β is a multivariate t-density with $df > 2$,

$$\boldsymbol{\beta}_p|\boldsymbol{\beta}_c \sim \boldsymbol{MVT}\left[\hat{\boldsymbol{\beta}},\ -\left(\frac{\partial^2 \ln p(\boldsymbol{\beta}|\boldsymbol{Y})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right)^{-1}\bigg|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}},\ df\right]$$

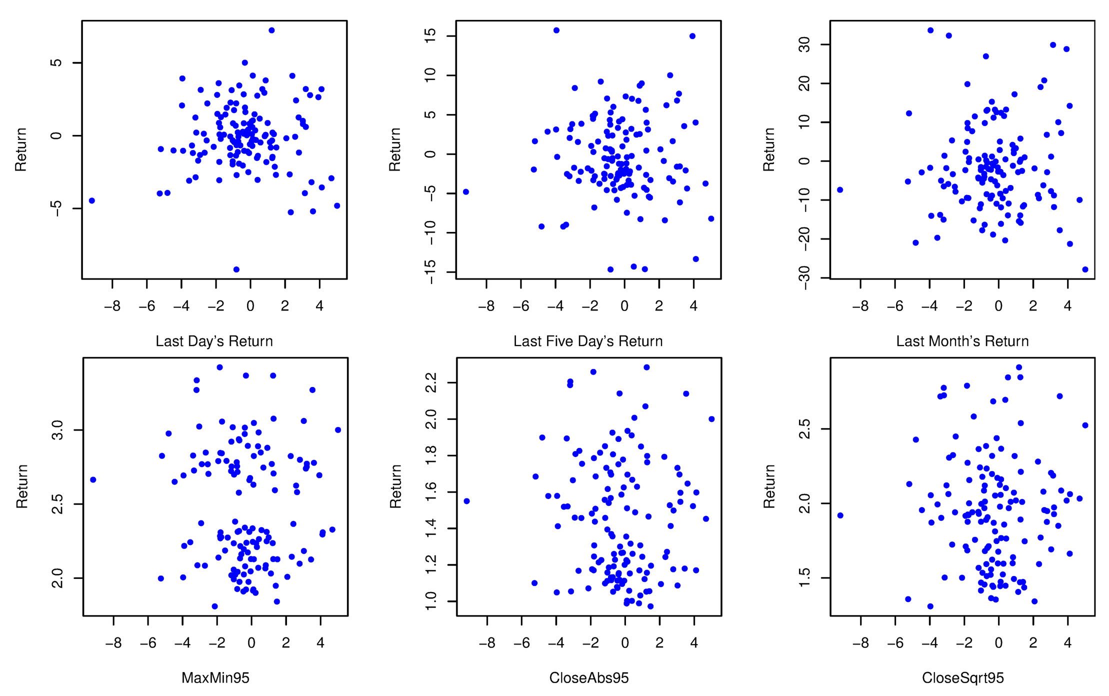
  where $\hat{\boldsymbol{\beta}}$ is obtained by R-step ($R \le 3$) Newton's iterations during the proposal with analytical gradients.
- This approach has some flavor of Hamiltonian MC when $R = 1$.
- Why not two-stage approach?
  - The asymptotic relative efficiency of the two-stage estimation procedure 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 always efficient due to the limited information provided by the estimators (Engle and Sheppard, 2001).

The Alibaba Stock Returns and Its News Sentiment

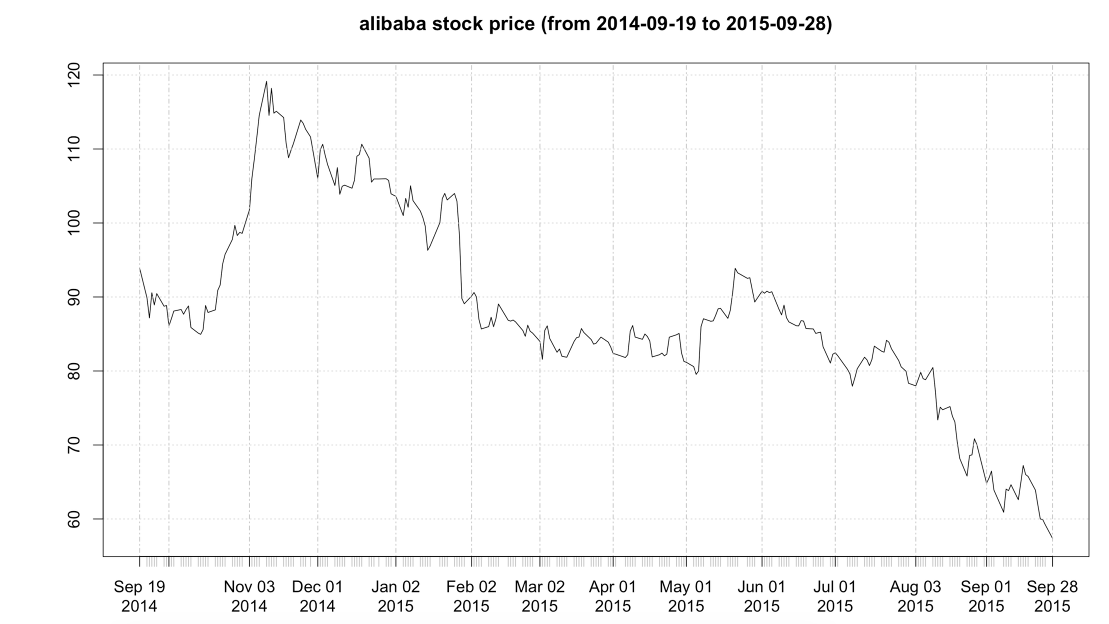
- **Scopes**
  - Is there any correlation between news information and stock returns? And is the correlation static?
  - Does there exist a way to join two models, say, one is discrete and the other one is continuous?
  - Can we find the co-movement between news and stocks?
  - What is the driven factor that causes the co-movement?
- 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.

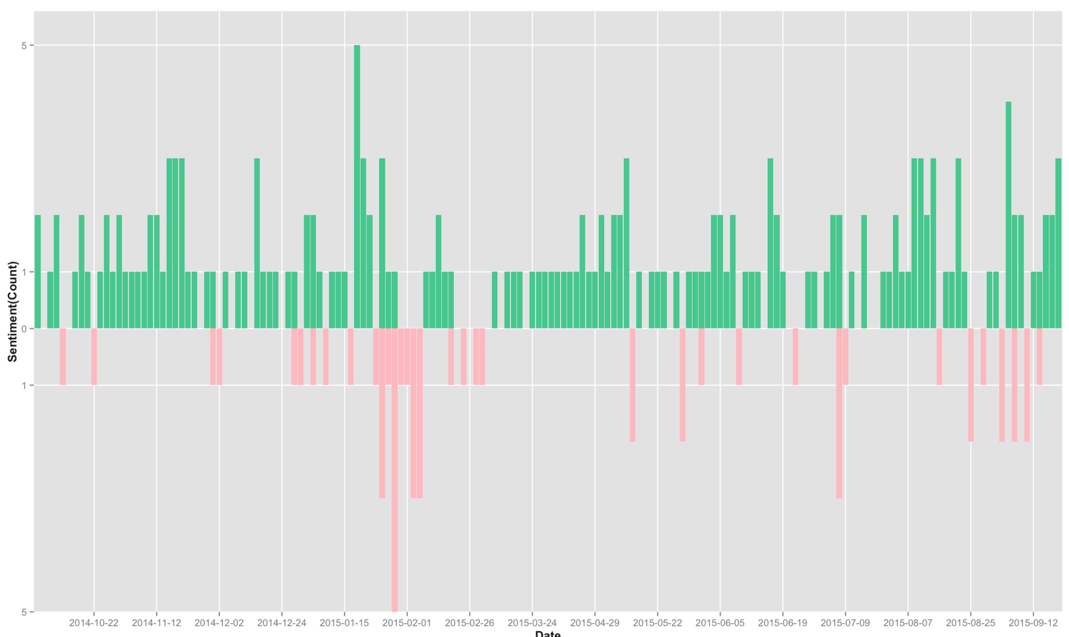
Date	P	N	U	上涨	下跌	打击	合作	增加	影响	违法
2014-10-15	2	0	1	0	0	1	3	0	1	0
2014-11-19	3	0	1	0	0	1	3	0	1	0
2015-01-28	3	3	3	0	0	2	5	1	2	4
2015-01-29	1	1	2	1	2	1	1	1	2	1
2015-01-30	1	7	2	0	1	3	1	1	5	5
2015-07-08	2	3	2	0	2	1	2	2	2	0

- Covariates used in mixture model for stock returns.

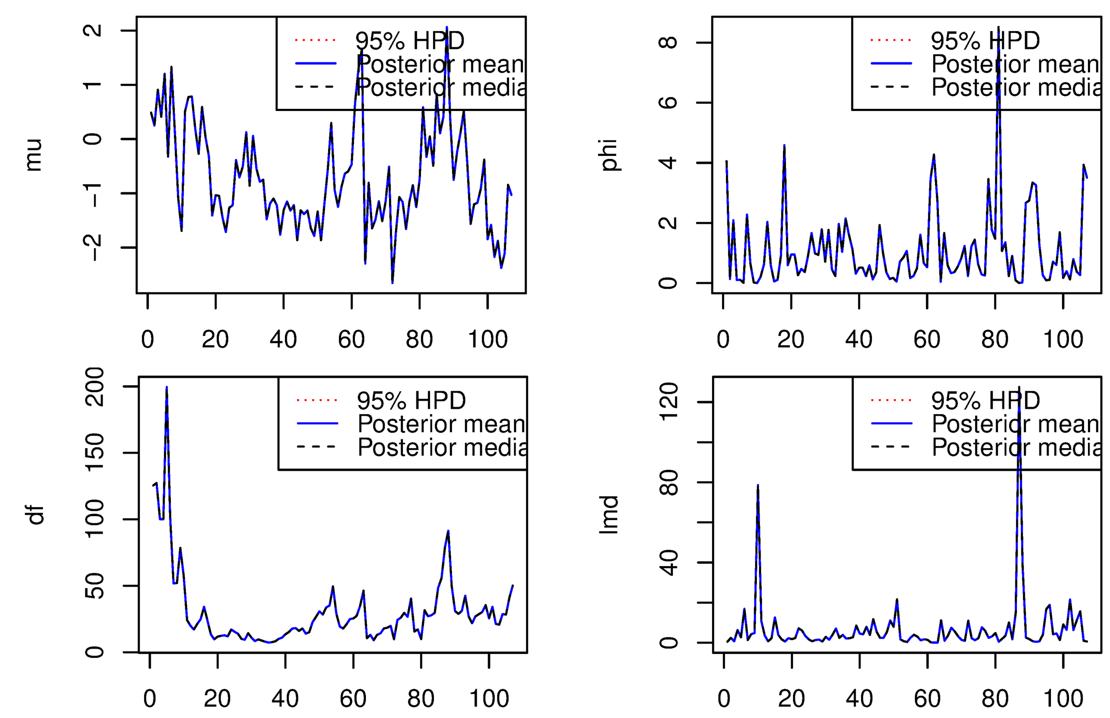


- The co-movement of the stock and news sentiment

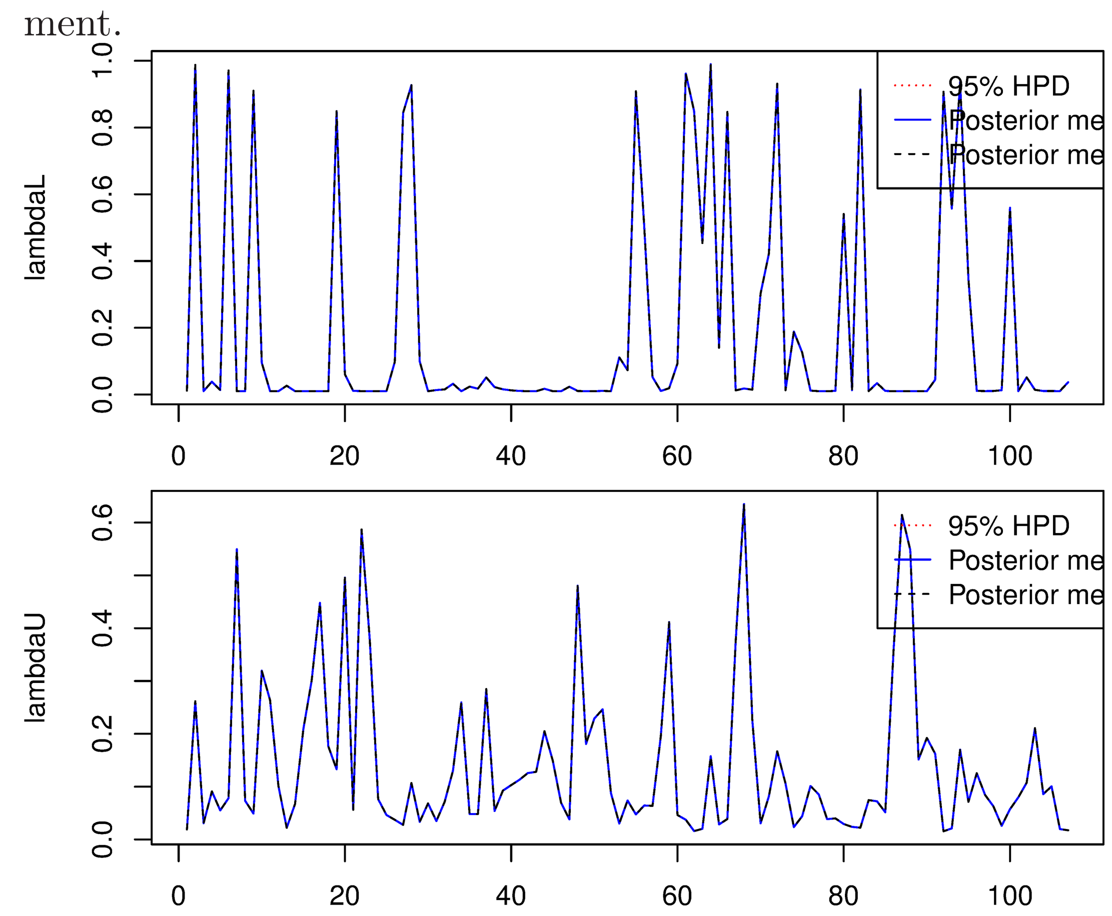




- The posterior mean for stock returns.



- The time-variant dependence of stock returns and new sentiment.



- We evaluate the quality of the one-step-ahead predictions using the log predictive density score (LPDS)

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

It can be shown that LPDS in a copula model equals the sum of their LPDSs in each marginal model and copula component, ie., $\mathrm{LPDS} = \sum_{i=1}^{M} \mathrm{LPDS}_i + \mathrm{LPDS}_C$.

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				