

ISSN 2090-3359 (Print)
ISSN 2090-3367 (Online)



Advances in Decision Sciences

Volume 22(A)
22nd Anniversary Special Issue
December 2018

Michael McAleer
Editor-in-Chief
University Chair Professor
Asia University, Taiwan



Published by Asia University, Taiwan

ADS@ASIAUNIVERSITY

Bayesian Shrinkage Estimation of Time-varying Covariance Matrices in Financial Time Series*

Mike K. P. So **

Department of Information Systems, Business Statistics and Operations Management
Hong Kong University of Science and Technology

Wing Ki Liu

Department of Information Systems, Business Statistics and Operations Management
Hong Kong University of Science and Technology

Amanda M. Y. Chu

Department of Social Sciences
Education University of Hong Kong

December 2018

* The authors would like to thank Michael McAleer and a reviewer for invaluable comments and suggestions. This work was supported by the Hong Kong RGC General Research Fund (grant number 604313).

** Corresponding author: immkpso@ust.hk

Abstract

Modeling financial returns is challenging because the correlations and variance of returns are time-varying and the covariance matrices can be quite high-dimensional. In this paper, we develop a Bayesian shrinkage approach with modified Cholesky decomposition to model correlations between financial returns. We reparameterize the correlation parameters to meet their positive definite constraint for Bayesian analysis. To implement an efficient sampling scheme in posterior inference, hierarchical representation of Bayesian lasso is used to shrink unknown coefficients in linear regressions. Simulation results show good sampling properties that iterates from Markov chain Monte Carlo converge quickly. Using a real data example, we illustrate the application of the proposed Bayesian shrinkage method in modeling stock returns in Hong Kong.

Keywords: Bayesian shrinkage; dynamic correlations; GARCH; lasso; Markov chain Monte Carlo.

JEL: C11, C32, C58, G17, G32.

1. Introduction

Financial return modeling is an important topic in financial econometric research. The covariance matrices of financial returns are known to vary over time. When modeling financial time series, research including Jebabli et al. (2014), Billio et al. (2016), Dyhrberg (2016), Laurent et al. (2016) and Hendrych and Cipra (2016) has proposed models assuming changing covariance matrices. To model the changing covariance matrix of financial returns in markets, on top of accounting for the time-varying feature of variances and correlations, we propose a Bayesian shrinkage method to model the covariances of financial returns while imposing sparsity in dynamic covariances. An efficient shrinkage method is helpful because the number of returns in the data set can be quite large, resulting in a large number of parameters to be estimated.

To start with, the model for time-varying variances has to be determined. Engle (1982) proposed an autoregressive conditional heteroscedastic (ARCH) model and described the maximum likelihood estimators. Bollerslev (1986) generalized the ARCH and proposed the GARCH(p, q) model. Assume that ψ_t is an information up to time t . Franses and Ghijssels (1999) proposed jumps at $t = \tau$ and stated that certain return observations are additive outliers and cannot fit GARCH models. For extensions and applications of the GARCH model, we refer to Billio et al. (2016), Dyhrberg (2016), Laurent et al. (2016), Corrêa et al. (2016), Narayan et al. (2016), Smetanina (2017) and Valera et al. (2018).

The GARCH model is commonly used for modeling time series and has sparked discussions about possible extensions and applications owing to the interpretative advantage of allowing the current variance to be expressed as the weighted average of squared returns and variances in the past. In our model, the GARCH(1,1) structure is applied so that the current variance can be expressed as a weighted average of the squared returns and variances in the previous period. To enable changing correlations between the returns, we allow the standardized residuals for different stocks from the GARCH(1,1) model to be correlated at any time t .

Then, in markets, an important research question is how we model the correlations between returns by using the standardized residuals in the GARCH(1,1) model. To impose a

positive definite condition in correlation matrices, we first link the correlation matrix to the covariance matrix in an equation. By assuming that \mathbf{X} is a random vector and S is a diagonal matrix in which the diagonal elements are the square root of the diagonal elements of $cov(\mathbf{X})$, we have:

$$corr(\mathbf{X}) = S^{-1}cov(\mathbf{X})S^{-1}.$$

With this link between the correlation and covariance matrices, we may consider various kinds of covariance modeling. A popular covariance modeling approach is to apply reparametrization by modified Cholesky decomposition. Pourahmadi (1999) proposed that as covariance matrices Σ are positive definite and symmetric, there is a unique lower triangular matrix T with diagonal entries equal to 1 such that:

$$T\Sigma T^{\top} = D.$$

Suppose that $i > j$ and the (i, j) th off-diagonal entries of T is $-\phi_{i,j}$. Reparametrization allows us to write regression equation $Y_i - \mu_i = \sum_{j=1}^{i-1} \phi_{i,j} (Y_j - \mu_j) + \epsilon_i$. The reparametrization is equivalent to regressing Y_i on Y_{i-1}, \dots, Y_1 , provides statistical interpretation and is unconstrained. Pourahmadi (2000) studied the asymptotic distribution of the maximum likelihood estimators of the parameters. For the applications and extensions of modified Cholesky decomposition, we refer to Matsuo et al. (2011), Xu and Mackenzie (2012), Zhang and Leng (2012) and Hendrych and Cipra (2016). Modified Cholesky decomposition is widely discussed and often applied because it provides a linear regression interpretation and avoids the positive definite constraint of covariance matrices.

Owing to these advantages, we apply modified Cholesky decomposition. In the later sections, when implementing our model, no constraints except those in the updating equations are imposed. To achieve optimal convergence rates in the sampling of the

parameters, least absolute shrinkage and selection operator (lasso), which shrinks the parameters towards their means, is applied to our model because it improves estimation accuracy and allows for easier interpretation of results.

The remainder of this paper is organized as follows. Section 2 describes the shrinkage method in our model. Section 3 introduces the dependence between stocks, and the incorporation of modified Cholesky decomposition. Section 4 introduces the sampling scheme. Section 5 presents and interprets the simulation results. Section 6 discusses the real sampling results.

2. Shrinkage Method

One major breakthrough in parameter estimation is that Tibshirani (1996) proposed lasso, which shrinks some of the coefficients to zero, thus improving the prediction accuracy and helping determine a smaller subset of variables. Assume that $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \dots, \hat{\beta}_p)'$. The lasso estimate is as follows:

$$(\hat{\alpha}, \hat{\boldsymbol{\beta}}) = \underset{\alpha, \boldsymbol{\beta}}{\operatorname{argmin}} \left\{ \sum_{i=1}^N (y_i - \alpha - \sum_j \beta_j x_{ij})^2 \right\} \quad \text{subject to} \quad \sum_j |\beta_j| \leq t.$$

The constraint $\sum_j |\beta_j|$ can be represented by adding a penalty term $\sum_j \lambda_j |\beta_j|$ and therefore the lasso is equivalent to imposing independent double exponential priors:

$$f(\beta_j) = \frac{1}{2\tau} \exp\left(-\frac{|\beta_j|}{\tau}\right).$$

Park and Casella (2008) proposed the hierarchical representation of lasso:

$$\begin{aligned}
\mathbf{y}|\mu, \mathbf{X}, \boldsymbol{\beta}, \sigma^2 &\sim N_n(\mu \mathbf{1}_n + \mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n), \\
\boldsymbol{\beta}|\sigma^2, \tau_1^2, \dots, \tau_p^2 &\sim N_p(\mathbf{0}_p, \sigma^2 \mathbf{D}_\tau), \\
\mathbf{D}_\tau &= \text{diag}(\tau_1^2, \dots, \tau_p^2), \\
\sigma^2, \tau_1^2, \dots, \tau_p^2 &\sim \pi(\sigma^2) d\sigma^2 \prod_{j=1}^p \frac{\lambda^2}{2} \exp\left(-\frac{\lambda^2 \tau_j^2}{2}\right) d\tau_j^2.
\end{aligned}$$

They suggested that by integrating $\tau_1^2, \dots, \tau_p^2$, we can obtain the double exponential priors and the hierarchical representation allows a simple and practical implementation of Gibbs samplers. For the lasso parameter λ , they considered using empirical Bayes by marginal likelihood or imposing gamma priors. For extensions and applications of lasso, we refer to Chen et al. (2011), Guo et al. (2012a), Schmidt and Makalic (2013), Benoit et al. (2013) and Zhang and Biswas (2015).

Group lasso and adaptive lasso are extensions of lasso. Yuan and Lin (2006) compared group lasso, group least angle selection and group non-negative garotte. The results in the four examples indicated that group least angle selection and group lasso are better than the traditional stepwise method. For the extensions and applications of group lasso, please refer to Huang et al. (2010), Kyung et al. (2010) and Hefley et al. (2017). Zou (2006) proved that lasso estimates are not consistent and proposed an adaptive lasso. For extensions of adaptive lasso, we refer to Mutshinda and Sillanpää (2010) and Mutshinda and Sillanpää (2012). Zou and Hastie (2005) proposed the elastic net that generalizes lasso. The elastic net with standardized predictors can be expressed as:

$$\hat{\boldsymbol{\beta}} = (1 + \lambda_2) \underset{\boldsymbol{\beta}}{\text{argmin}} \left\{ \sum_i^N (y_i - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda_2 \sum_{j=1}^p \beta_j^2 + \lambda_1 \sum_{j=1}^p |\beta_j|^2 \right\}.$$

They explained that when λ_2 is equal to zero, the elastic net is equivalent to lasso. For the extensions of the elastic net, we refer to Zou and Zhang (2009), Kyung et al. (2010), Gefang (2014) and Zhang et al. (2017).

Other shrinkage methods include smoothly clipped absolute deviation (SCAD) and group SCAD. Fan and Li (2001) proposed SCAD, a quadratic spline function, and proved its oracle property. Kim et al. (2008) developed an optimization algorithm and proved the asymptotic properties of the estimator in high dimensions. Wang et al. (2016) described the asymptotic properties of SCAD

estimators in the quadratic approximation of objective function. For group SCAD and its oracle properties, we refer to Wang et al. (2007), Hu et al. (2014) and Guo et al. (2015). Lasso and SCAD-related shrinkage methods are considered for our sampling scheme. Firstly, lasso-related shrinkage methods, such as adaptive lasso, group lasso, and elastic net, have been widely discussed because lasso shrinks the parameters towards their means, which improves estimation accuracy and allows for easier interpretation of results. For these advantages, lasso is applied to our model in a Bayesian framework. As the number of time points can be large, group lasso is used to group correlation coefficients at each time point. Secondly, SCAD-related shrinkage methods are widely discussed because of their oracle properties. However, SCAD-related shrinkage methods have not been extended for application in a Bayesian framework and therefore the approach cannot be used in our model.

3. Model

The GARCH model has been widely used for modeling financial returns to incorporate time-varying conditional volatility. In our model, r_{tk} , the financial return of the k -th stock at time t , is modeled as:

$$r_{tk} = \sqrt{h_{tk}} \tilde{y}_{tk},$$

where \tilde{y}_{tk} follows standard normal distribution and h_{tk} is the conditional variance of r_{tk} given as follows:

$$\begin{aligned} h_{tk} &= \gamma_{0k} + \gamma_{1k} r_{t-1,k}^2 + \xi_{1k} h_{t-1,k}, \\ \gamma_{0k} &> 0, \gamma_{1k}, \xi_{1k} \geq 0, \gamma_{1k} + \xi_{1k} < 1. \end{aligned} \quad (1)$$

The return of one stock can be dependent on those of other stocks at any time t and thus one objective of this paper is to develop a model for capturing changing dependence structures between

returns. Let $\tilde{y}_t = (\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_p)$, where p is the total number of stocks. To model the changing correlation matrix of \tilde{y}_t , we specify:

$$\tilde{y}_t = S_t^{-1}y_t, \quad (2)$$

where S_t is a diagonal matrix with the k -th diagonal element given by the standard deviation of y_{tk} . From Equation (2), we have $cov(\tilde{y}_t) = S_t^{-1}\Sigma_t S_t^{-1}$ where Σ_t is the covariance matrix of y_t .

Our goal is therefore to estimate Σ_t while imposing possible sparsity on Σ_t . We write:

$$y_t | \mathfrak{F}_{t-1}, \Sigma_t \sim N(\mathbf{0}, \Sigma_t),$$

where \mathfrak{F}_t is the information up to time t and so $var(y_t | \mathfrak{F}_{t-1})$, the conditional covariance matrix of y_t is given by Σ_t . Following Pourahmadi (1999), we reparameterize Σ_t using a modified Cholesky decomposition:

$$I = T_t \Sigma_t T_t^{-1},$$

where I is identity matrix and T_t is a lower triangular matrix with diagonal entries equal to one at time t . The off-diagonal entry at the k -th row and j -th column of T_t is $-\phi_{t,kj}$ when $j < k$. The coefficient $\phi_{t,kj}$ is uniquely defined because there are only $p(p-1)/2$ off-diagonal entries in the Cholesky decomposition. Our model generalizes cases for different covariance matrices. Σ_t can take various forms according to the values of $\phi_{t,k}$. If $\phi_{t,kj}$ for all t, k are equal to zero, Σ_t will be an identity matrix for all t and the model will be reduced to an independent multivariate GARCH model.

By applying Cholesky decomposition to the covariance matrix, we provide unconstrained reparametrization and regression interpretation. Σ_t is constrained to be positive definite but ϕ_{tk} is unconstrained. For the regression interpretation, the central idea is to write:

$$y_{tk} = \sum_{j=1}^{k-1} \phi_{t,kj} y_{tj} + \epsilon_{tk} \quad k = 1, \dots, p, \quad (3)$$

where ϵ_{tk} is iid $N(0,1)$. The variance of ϵ_{tk} is set to 1 for normalization. As Equation (3) takes the form of a regression equation, ϕ_{tk} is similar to regression coefficients in terms of interpretation. Modified Cholesky decomposition is a standard approach for modeling covariance matrices. Nino-Ruiz et al. (2017) used modified Cholesky decomposition to estimate the background error covariance matrices in Kalman filter. Feng et al. (2016) and Lv and Guo (2017) applied Cholesky decomposition to responses in longitudinal analysis.

4. Bayesian Shrinkage Estimation

Define $\phi_{tk} = (\phi_{t,k1}, \dots, \phi_{t,kk-1})'$, $\mathbf{y}_k = (\gamma_{0k}, \gamma_{1k}, \xi_{1k})'$, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2)'$ and $\mathbf{r} = (\tau_{1,1}, \dots, \tau_{N,p})'$. Denote the unknown parameter vector $(\mathbf{y}_1', \dots, \mathbf{y}_p', \boldsymbol{\alpha}', \phi_{12}', \dots, \phi_{Np}', \tau_{21}^2, \dots, \tau_{p,p-1}^2)'$ by $\boldsymbol{\theta}$.

a. Prior for imposing sparsity

In this subsection, we develop a dynamic shrinkage prior for ϕ_{tk} to possibly impose sparsity in dynamic conditional correlations. A Bayesian lasso has been shown to be efficient in the literature when applied to different models because it can shrink the predictors. Chen et al. (2011) proposed a Reversible-jump Markov chain Monte Carlo algorithm (RJMCMC)-based Bayesian lasso. Their idea was that the prior in the Bayesian lasso provides shrinkage and RJMCMC allows movements between models with different numbers of predictors. In the simulation study, Chen et al. (2011) showed that the RJMCMC-based Bayesian lasso has the smallest mean square errors among the methods they considered. Li et al. (2011) presented genome-wide association studies and suggested a mixed effect model with imposed Bayesian lasso prior and observed that the Bayesian lasso shrank the small effects of genes to zero. Guo et al. (2012b) proposed a semi-parametric structural equation model

and imposed a Bayesian lasso prior. They found that their proposed model without spline performed better than the model without splines and the shrinkage effects of the Bayesian lasso lead to a correct model. Therefore, our goal is to shrink ϕ_{tk} towards its prior mean μ_{tk} and achieve sampling accuracy by applying a Bayesian lasso. Following Park and Casella (2008), we choose the following Bayesian lasso prior:

$$\begin{aligned}\phi_{tk}|\tau_k^2 &\sim N(\mu_{tk}, H_k), \\ \tau_{kj}^2|\lambda_t &\sim \text{Exp}(\lambda_t^2/2),\end{aligned}$$

where $\mu_{tk} = (\mu_{t,k1}, \dots, \mu_{t,kk-1})^T$, $H_k = \text{diag}(\tau_{k1}^2, \dots, \tau_{kk-1}^2)$ and $\text{Exp}(\theta)$ denotes the exponential distribution with mean $\frac{1}{\theta}$. The idea is that by integrating $\tau_{k1}^2, \dots, \tau_{kk-1}^2$, the prior becomes:

$$\pi(\phi_{tk}|\lambda_t) = \prod_{j=1}^{k-1} \frac{\lambda_t}{2} e^{-\lambda_t|\phi_{t,kj} - \mu_{t,kj}|},$$

which can perform shrinkage by adjusting λ_t . This prior is formed by independent

$\text{Laplace}(\mu_{t,kj}, \lambda_t^{-1})$. As the prior mean of $\phi_{t,kj}$ is $\mu_{t,kj}$, the iterates of $\phi_{t,kj}$ are expected to approach $\mu_{t,kj}$ when the prior variance $2\lambda_t^{-2}$ is small or λ_t is large. To enable dynamic shrinkage of ϕ_{tk} , μ_{tk} is written as:

$$\begin{aligned}\mu_{tk} &= \alpha_0 + \alpha_1\mu_{t-1,k} + \alpha_2\phi_{t-1,k}, \\ 0 &\leq \alpha_1 + \alpha_2 \leq 1,\end{aligned}\quad (4)$$

which is the weighted average of $\mu_{t-1,k}$ and $\phi_{t-1,k}$, implying that $\mu_{t,k}$ is an exponentially moving average of $\phi_{1,k}, \dots, \phi_{t-1,k}$. The above indicates that $\mu_{t,k}$, the prior mean of $\phi_{t,k}$, varies with t . We assume stationarity in $\mu_{t,k}$, implying that the mean of $\mu_{t,k}$ is $\alpha_0/(1 - \alpha_1 - \alpha_2)$.

b. Parameter estimation

Let $\theta_{-\gamma}$ be the parameter vector without the element γ . Equation (3) can be expressed as:

$$\frac{r_{tk}\sqrt{\text{var}(y_{tk})}}{\sqrt{h_{tk}}} = \sum_{j=1}^{k-1} \frac{r_{tk}\sqrt{\text{var}(y_{tk})}}{\sqrt{h_{tk}}} + \epsilon_{tk}.$$

The conditional likelihood function is:

$$\begin{aligned} p(\mathbf{r}|\theta) &= \prod_{t=1}^N \prod_{k=1}^p \frac{\sqrt{\text{var}(y_{tk})}}{\sqrt{2\pi}\sqrt{h_{tk}}} \exp\left(-\frac{1}{2} \left(\frac{r_{tj}}{\sqrt{h_{tj}}} \sqrt{\text{var}(y_{tj})} - \sum_{u=1}^{j-1} \frac{r_{tu}}{\sqrt{h_{tu}}} \sqrt{\text{var}(y_{tu})} \right)^2\right) \\ &= \prod_{t=1}^N \prod_{k=1}^p \frac{\sqrt{\text{var}(y_{tk})}}{\sqrt{2\pi}\sqrt{h_{tk}}} g_{tj}, \end{aligned}$$

where $g_{tj} = \exp\left(-\frac{1}{2} \left(\frac{r_{tj}}{\sqrt{h_{tj}}} \sqrt{\text{var}(y_{tj})} - \sum_{u=1}^{j-1} \frac{r_{tu}}{\sqrt{h_{tu}}} \sqrt{\text{var}(y_{tu})} \right)^2\right)$. Overall, the following prior for θ is adopted:

$$p(\theta) \propto \prod_{t=1}^N \prod_{k=2}^p P(\phi_{tk}|\tau_k^2) \prod_{k=2}^p \prod_{j=2}^{k-1} P(\tau_{kj}^2),$$

where $\tau_k^2 = (\tau_{k,1}^2, \dots, \tau_{k,k-1}^2)$.

The parameters to be estimated in the model are α , γ_k for $k = 1, \dots, p$, τ_{kj}^2 for $j = 1, \dots, k - 1$ and $k = 1, \dots, p$, and ϕ_{tj} for $t = 1, \dots, N$ and $j = 2, \dots, p - 1$. The other parameters become known after deriving the above parameters. Subsection 4.2.1 uses Gibbs sampling to sample τ_{kj}^2 because the posterior distribution is standard. Subsection 4.2.2 applies the Metropolis Hastings (MH) algorithm with random kernel and the blocking approach proposed by Chen and So (2006) to sample γ_k because MH algorithm allows the sampling of nonstandard posterior distributions and the approach can take into consideration the covariances between the parameters. Subsection 4.2.3 uses the multiple-try Metropolis algorithm with random kernel and the blocking approach proposed by Chen and So (2006) to sample α as the multiple-try Metropolis algorithm enables the sampling of nonstandard posterior distributions without settling the iterates in local regions and the approach can take into consideration the covariances between the parameters. Subsection 4.2.4 applies the MH algorithm with independence kernel to sample ϕ_{tj} because MH algorithm allows the sampling of nonstandard posterior distributions and the independence kernel can increase sampling efficiency.

i. Sampling τ_{kj}^2

We apply Gibbs sampling to τ_{kj}^2 because the posterior distribution follows generalized inverse Gaussian distribution. The posterior distribution of τ_{kj}^2 is:

$$\begin{aligned}
P(\tau_{kj}^2 | \mathbf{r}, \boldsymbol{\theta}_{-\tau_{kj}^2}) &\propto P(\mathbf{r} | \boldsymbol{\theta}) \prod_{t=1}^N P(\phi_{tk} | \tau_k^2) P(\tau_{kj}^2) \\
&\propto \prod_{t=1}^N P(\phi_{tk} | \tau_k^2) P(\tau_{kj}^2) \\
&\propto \prod_{t=1}^N \tau_{kj}^{-1} \exp(-(\phi_{tkj} - \mu_{tkj})/2\tau_{kj}^2) * \exp(-\lambda_t^2 \tau_{kj}^2/2) \\
&\propto (\tau_{kj}^2)^{-(N/2-1)-1} \exp(-\frac{1}{2} (\sum_{t=1}^N ((\phi_{tkj} - \mu_{tkj})^2/\tau_{kj}^2 + \lambda_t^2 \tau_{kj}^2))).
\end{aligned}$$

ii. Sampling γ_k

The posterior distribution of γ_k is:

$$\begin{aligned}
& P(\boldsymbol{\gamma}_k | \mathbf{r}, \boldsymbol{\theta}_{-(\gamma_{0k}, \gamma_{1k}, \xi_{1k})}) \\
& \propto P(\mathbf{r} | \boldsymbol{\theta}) \\
& \propto \prod_{t=1}^N \prod_{j=k}^p \frac{1}{\sqrt{h_{tj}}} g_{tj}.
\end{aligned} \tag{5}$$

The MH algorithm is applied to sample $\boldsymbol{\gamma}_k$ because the posterior distribution in Equation (5) is nonstandard. The procedure for sampling $\boldsymbol{\gamma}_k$ is

1. At iteration i , generate a point $\boldsymbol{\gamma}_k^*$ from $\boldsymbol{\gamma}_k^* = \boldsymbol{\gamma}_k^{(i-1)} + \boldsymbol{\epsilon}$ where $\boldsymbol{\gamma}_k^{(i-1)}$ is the $(i - 1)$ th estimate and $\boldsymbol{\epsilon} \sim N(0, \boldsymbol{\alpha}_{\boldsymbol{\gamma}_k} \boldsymbol{\Omega}_{\boldsymbol{\gamma}_k})$.
2. Accept $\boldsymbol{\gamma}_k^*$ with probability $\min\{1, P(\boldsymbol{\gamma}_k^*)/P(\boldsymbol{\gamma}_k^{(i-1)})\}$.

The parameters are expected to be strongly correlated because they are GARCH parameters in Equation (1). Therefore, we follow the approach used by Chen and So (2006). During the first M iterations, $\boldsymbol{\Omega}_{\boldsymbol{\gamma}_k}$ is selected to be a diagonal matrix. And $\boldsymbol{\alpha}_{\boldsymbol{\gamma}_k}$ is tuned to achieve an acceptance rate between 25% and 50%. After M iterations, $\boldsymbol{\Omega}_{\boldsymbol{\gamma}_k}$ is revised to be a sample covariance matrix of the first M iterates to improve the convergence rate.

An advantage of the approach used by Chen and So (2006) is that the covariance matrix between the parameters is taken into consideration in the sampling scheme, speeding up convergence. The scalar term is selected to obtain an acceptance rate between 25% and 50%, which is considered optimal.

iii. Sampling $\boldsymbol{\alpha}$

The posterior distribution of $\boldsymbol{\alpha}$ is:

$$\begin{aligned}
& P(\boldsymbol{\alpha}|\mathbf{r}, \boldsymbol{\theta}_{-\boldsymbol{\alpha}}) \\
& \propto P(\mathbf{r}|\boldsymbol{\theta}) \prod_{t=1}^N \prod_{k=1}^p P(\phi_{tk}|\tau_k^2) \\
& \propto \prod_{t=1}^N \prod_{k=1}^p P(\phi_{tk}|\tau_k^2) \\
& \propto \prod_{t=1}^N \prod_{k=2}^p \exp\left(-\frac{1}{2}(\phi_{tk} - \mu_{tk})' H_k^{-1} (\phi_{tk} - \mu_{tk})\right).
\end{aligned}$$

The posterior distribution of $\boldsymbol{\alpha}$ is not standard. We also observed that sampling of $\boldsymbol{\alpha}$ has a multi-modal problem in the preliminary test. Therefore, a multiple-try Metropolis algorithm with random walk kernel is applied to sample $\boldsymbol{\alpha}$ because we wish to tackle the multi-modal problem by applying an algorithm such that the iterates are settled in global regions instead of local regions. The sampling procedure of $\boldsymbol{\alpha}$ is

1. At iteration i , draw $\boldsymbol{\alpha}_{i,1}, \dots, \boldsymbol{\alpha}_{i,r}$ from $N(\boldsymbol{\alpha}_i, a\Omega)$, where r is the number of trials.
2. Select $\boldsymbol{\alpha}_i^*$ from $\{\boldsymbol{\alpha}_{i,1}, \dots, \boldsymbol{\alpha}_{i,r}\}$ according to the probability:
- 3.

$$\frac{f(\boldsymbol{\alpha}_i^*)}{f(\boldsymbol{\alpha}_{i,1}) + \dots + f(\boldsymbol{\alpha}_{i,r})}.$$

4. Draw $\boldsymbol{\alpha}_{i,1}^*, \dots, \boldsymbol{\alpha}_{i,r-1}^*$ from $N(\boldsymbol{\alpha}_i^*, a\Omega)$. Set $\boldsymbol{\alpha}_{i,r}^* = \boldsymbol{\alpha}_i$.
5. Accept $\boldsymbol{\alpha}_i^*$ with probability:
- 6.

$$\min\left\{1, \frac{f(\boldsymbol{\alpha}_{i,1}) + \dots + f(\boldsymbol{\alpha}_{i,r})}{f(\boldsymbol{\alpha}_{i,1}^*) + \dots + f(\boldsymbol{\alpha}_{i,r}^*)}\right\}.$$

Liu et al. (2000) proposed the above multiple-try Metropolis sampler and showed that while Metropolis chains were stuck in local maxima, most multiple-try Metropolis chains settled in modal regions. So (2006) presented two examples and showed that multiple-try Metropolis sampler enables the iterates to move from one mode to another mode, speeding up convergence. Pandolfi et al. (2014) proposed a generalized multiple-try reversible jump algorithm and found that it is more

efficient than reversible jump algorithm. Therefore, our hope is that the multiple-try algorithm will resolve the challenge of the multi-modal problem by settling the iterates in global regions.

To account for the correlations between α_0 , α_1 and α_2 , we follow the approach used by Chen and So (2006). During the first M iterations, Ω is a diagonal matrix and a is tuned so that we have an acceptance rate between 25% and 50%. After M iterations, a sample covariance matrix is selected to be Ω .

iv. Sampling ϕ_{tj}

A major challenge of sampling ϕ_{tj} is that the number of time points and dimension of the returns can be very large, and so the number of parameters that need to be estimated is potentially huge. With low sampling efficiency of ϕ_{tj} , we can get a low overall convergence rate because the number of parameters is potentially huge and ϕ_{tj} are key regression parameters for modeling correlations between stocks. Because of the nonstandard posterior distribution of ϕ_{tj} , we opt to use the MH algorithm to sample ϕ_{tj} , an algorithm that may increase the program run time substantially when a large number of parameters are sampled. The posterior distribution of ϕ_{tj} is:

$$\begin{aligned}
 & p(\phi_{tj} | \mathbf{r}, \boldsymbol{\theta}_{-\phi_{tk}}), \\
 & \propto P(\mathbf{r} | \boldsymbol{\theta}) \prod_{w=t}^N P(\phi_{wj} | \tau_k^2) \\
 & \propto \prod_{k=1}^p \sqrt{\text{var}(y_{tk})} g_{tk} \prod_{w=t}^N \exp\left(-\frac{1}{2}(\phi_{wj} - \mu_{wj})' H_j^{-1} (\phi_{wj} - \mu_{wj})\right).
 \end{aligned} \tag{6}$$

The posterior distribution of ϕ_{tj} is reduced to a Gaussian distribution if $\sqrt{\text{var}(y_{tk})}$ outside the exponential functions in Equation (6) are equal to 1. The reduced form is used as the independence kernel because it has a similar form to the posterior distribution and follows a normal distribution, which will increase the efficiency of the sampling of ϕ_{tj} . Assume that \mathbf{z}_j is a $j \times 1$ vector containing only 1s. The independence kernel can be written as:

$$\begin{aligned}
& q(\phi_{tj}) \\
& \propto P(\mathbf{r}|\boldsymbol{\theta}) \prod_{w=t}^N P(\phi_{wj}|\tau_k^2) \\
& \propto \exp\left(-\frac{1}{2}\left(\frac{r_{tj}}{\sqrt{h_{tj}}}\sqrt{\text{var}(y_{tj})} - \sum_{u=1}^{j-1} \frac{r_{tu}}{\sqrt{h_{tu}}}\sqrt{\text{var}(y_{tu})}\right)^2\right) \quad (7) \\
& * \prod_{w=t}^N \exp\left(-\frac{1}{2}(\phi_{wj} - \mu_{wj})'H_j^{-1}(\phi_{wj} - \mu_{wj})\right) \\
& \propto \exp\left(-\frac{1}{2}(\phi_{tj} - m_{tj})'V_{tj}^{-1}(\phi_{tj} - m_{tj})\right),
\end{aligned}$$

Where

$$\begin{aligned}
Y_{tj} &= (y_{t,1}, \dots, y_{t,j-1})', \\
V_{tj} &= \left(\left(\frac{(1 - \alpha_1^{2(N-t)})\alpha_2^2}{1 - \alpha_1^2} + 1\right)H_j^{-1} + Y_{t,j}Y_{t,j}'\right)^{-1}, \\
m_{tj} &= V_{tj}^{-1}(y_{t,j}Y_{t,j} - \left(\frac{(1 - \alpha_1^{2(N-t)})\alpha_1\alpha_2}{1 - \alpha_1^2} - 1\right)H_j^{-1}\mu_{t,j} + \sum_{u=t+1}^N \alpha_1^{u-t-1}\alpha_2 H_j^{-1}\phi_{u,j} \\
&\quad - \sum_{u=t+1}^{N-1} \sum_{h=t+1}^u \alpha_1^{2u-h-t}\alpha_2^2 H_j^{-1}\phi_{h,j} - \sum_{u=t+1}^N \frac{(1 - \alpha_1^{u-t})\alpha_1^{u-1-t}\alpha_0\alpha_2}{1 - \alpha_1} H_j^{-1}z_j), \\
z_j &= (1, \dots, 1)'.
\end{aligned}$$

For detailed derivation of the posterior distribution, please refer to Appendix 8. The sampling procedure is a standard MH algorithm which can be written as

1. At iteration i , generate ϕ_{tj}^* from $q(\phi_{tj})$.
2. Accept ϕ_{tj}^* with probability $\min\left\{1, \frac{f(\phi_{tj}^*)q(\phi_{tj}^{(i-1)})}{f(\phi_{tj}^{(i-1)})q(\phi_{tj}^*)}\right\}$, where $\phi_{tj}^{(i-1)}$ is $(i-1)$ th iterate of

ϕ_{tk} .

Good statistical properties and faster convergence can be acquired by using an independence kernel. Mengersen and Tweedie (1996) suggested that the independence kernel has geometric and uniform ergodicity properties. Robert and Casella (2013) stated that the chain is irreducible, aperiodic and ergodic if, and only if, the independent kernel is positive on the support of the posterior distribution. Bolstad (2009) presented a mixed normal target distribution and compared the

independence chain with the random walk chain. He found that the independence kernel enabled the chain to move more quickly.

We now describe the key recursive relationships for the summations, which enables a reduction of the amount of time required for the implementation of the MH algorithm. It can be problematic to compute the summations when more time series data are taken into consideration and the number of time points is very large. This problem can be solved by computing the summations using the recursive relationships between the summations. Let:

$$A_{t,j} = \sum_{u=t+1}^N \alpha_1^{u-t-1} \alpha_2 \phi_{u,j},$$

$$B_{t,j} = \sum_{u=t+1}^{N-1} \sum_{h=t+1}^u \alpha_1^{2u-h-t} \alpha_2^2 H_j^{-1} \phi_{h,j}.$$

The summations can be calculated as follows:

$$\begin{aligned} A_{t-1,j} &= \sum_{u=(t-1)+1}^N \alpha_1^{u-(t-1)-1} \alpha_2 \phi_{u,j} \\ &= \alpha_1 \sum_{u=t+1}^N \alpha_1^{u-t-1} \alpha_2 \phi_{u,j} + \alpha_2 \phi_{t,j} \\ &= \alpha_1 A_{t,j} + \alpha_2 \phi_{t,j}, \\ B_{t-1,j} &= \sum_{u=t}^{N-1} \sum_{h=t}^N \alpha_1^{2u-h-(t-1)} \alpha_2^2 H_j^{-1} \phi_{h,j} \\ &= \sum_{u=t+1}^{N-1} \alpha_1^{2u-h-t+1} \alpha_2^2 H_j^{-1} \phi_{h,j} + \sum_{u=t}^{N-1} \alpha_1^{2u-t-t+1} \alpha_2^2 H_j^{-1} \phi_{t,j} \\ &= \alpha_1 B_{t,j} + \sum_{u=t}^{N-1} \alpha_1^{2u-2t+1} \alpha_2^2 H_j^{-1} \phi_{t,j} \\ &= \alpha_1 B_{t,j} + \frac{\alpha_1 \alpha_2^2 (1 - \alpha_1^{2(N-t)})}{1 - \alpha_1^2} H_j^{-1} \phi_{t,j}. \end{aligned}$$

The initial value of $A_{t,j}$ is $A_{N-1,j}$ which is $\alpha_2 \phi_{N,j}$. The initial value of $B_{t,j}$ is $B_{N-2,j}$ which is $\alpha_1 \alpha_2^2 \phi_{N-1,j}$. At each iteration, we computed the summations by the recursive relationships right before sampling ϕ_{tj} for $t = 1, \dots, N$ and $j = 2, \dots, p - 1$. Without using the recursive relationships, it takes $O(N^2 p)$ time to finish one iteration. It requires $O(N)$ summations for each t and k . As there are N time points and p dimensions, the amount of time needed is $O(N^2 p)$. But by applying the recursive relationships in our calculations, the amount of time needed for one iteration is reduced to $O(Np)$. For each j , our implementation needs $O(N)$ summations because of the recursive relationship. As the dimension is p , the amount of time needed for our implementation is $O(Np)$. Hence, our implementation allows the sampling of larger amounts of time series data within a limited time and reduces the number of computations required.

5. Simulation Experiment

In order to investigate the efficiency of the Bayesian method in Section 4, we implement the Markov chain Monte Carlo (MCMC) algorithm in several settings. In the first setting, we set the true values of the parameters as follows:

$$\begin{aligned} \alpha_0 &= 0.1, & \alpha_1 &= 0.8, & \alpha_2 &= 0.1, \\ \gamma_{0k} &= 0.1, & \gamma_{1k} &= 0.1, & \xi_{1k} &= 0.8 \quad \text{for } k = 1, 2, 3, 4, 5. \\ \lambda_t &= 1. \end{aligned}$$

The dimension in the simulation is five and λ_t are set to be 1. The length of the time series is 2,000. The goal of this simulation is to examine whether the sampling scheme is effective and efficient. The constraints of the parameters in our model include $0 \leq \alpha_1 + \alpha_2 < 1$, $\gamma_{0,k} > 0$ and $0 \leq \gamma_{1k} + \xi_{1k} < 1$ for $k = 1, \dots, 5$. The parameters, γ_{0k} for $k = 1, 2, 3, 4, 5$ are set to be 0.1 and γ_{1k} and ξ_{1k} for $k = 1, 2, 3, 4, 5$ are set such that the sum of γ_{1k} and ξ_{1k} for $k = 1, 2, 3, 4, 5$ is close to 1. To study the efficiency of the MCMC sampling scheme, 30,000 iterations are run and the first 15,000 are discarded to eliminate the impact of initial values.

The second setting further explores the sampling accuracy. The parameter setting of the second experiment is as follows:

$$\begin{aligned}\alpha_0 &= 0.01, & \alpha_1 &= 0.9, & \alpha_2 &= 0.05, \\ \gamma_{0k} &= 0.05, & \gamma_{1k} &= 0.1, & \xi_{1k} &= 0.85 \quad \text{for } k = 1,2,3,4,5. \\ \lambda_t &= 1.\end{aligned}$$

The sum of α_1 and α_2 is 0.95, which is closer to 1 than in the first setting. The sum of γ_{1k} and ξ_{1k} for $k = 1,2,3,4,5$ is 0.95, which is slightly smaller than 1. The parameter γ_{0k} for $k = 1,2,3,4,5$ is 0.05, which is slightly greater than 0. The third setting has parameters given as follows:

$$\begin{aligned}\alpha_0 &= 0.01, & \alpha_1 &= 0.8, & \alpha_2 &= 0.1, \\ \gamma_{0k} &= 0.05, & \gamma_{1k} &= 0.1, & \xi_{1k} &= 0.85 \quad \text{for } k = 1,2,3,4,5. \\ \lambda_t &= 1.\end{aligned}$$

Similar to the second setting, the γ_{0k} , γ_{1k} and ξ_{1k} are set to be 0.05, 0.1 and 0.85 respectively. The parameters α_1 and α_2 are chosen to be different from the second setting. The third and fourth parameter settings are presented in Table 2.

We generated 100 replications for the four parameter settings to study the performance of the Bayesian estimation method. Table 1 and Table 2 show the mean, median and standard deviation of the posterior mean estimators based on the replications that were successfully run. In general, the means of the posterior mean estimators are within one standard deviation of the respective true values. Most of the standard deviations of the posterior mean estimators are reasonably small. The above experiment indicates that the Bayesian method is able to reliably estimate the unknown parameters of the model with dynamic sparsity.

6. Real Data

In order to study the performance of the model in real data, we collected five constituent stocks of the Hang Seng Index in Hong Kong. The real data set consists of 2,461 percentage returns of 0001 HK Equity, 0101 HK Equity, 0011 HK Equity, 1044 HK Equity and 1088 HK Equity between 31 July, 2007 and 21 July, 2017 from Yahoo Finance. The duration of the data set is ten years and there are no missing data. Table 3 shows the mean, maximum, median and minimum of percentage returns for the five stocks.

The MCMC algorithm was implemented on the real data set. The total number of iterations is 30,000 and the number of burn-in iterations is 15,000, implying that we kept the last 15,000 iterates for posterior analysis. We adopted $\lambda_t = 1$ for all t to observe whether this λ_t can lead to satisfactory results. When λ_t is larger, the deviation of ϕ_{tk} from its prior mean is penalized more heavily and the movement of the iterates is smaller. When λ_t is small, the effect of shrinkage is small and ϕ_{tk} may not be shrunk towards the prior mean μ_{tk} . Figure 1 and Figure 2 report the trace plots, histograms, and autocorrelation plots of the real data experiment. The trace plots show that the iterates converge very quickly to values that are not their initial values. As the iterates have only one mode in the histograms, there is no multi-modal problem.

The autocorrelations drop rapidly with the number of lags, proving that the current iterates are independent of previous iterates. Table 4 shows the posterior mean and standard deviation estimates from MCMC iterates. The means and standard deviations of the parameters are calculated by the iterates after discarding the first 15,000 iterations. The GARCH parameter estimates are close to our parameter settings in the simulation experiments where the estimates of γ_{1k} for $k = 1, 2, 3, 4, 5$ are between 0.05 and 0.2 and the estimates of ξ_{1k} are between 0.8 and 0.95. These GARCH parameter estimates are typical in econometric time series modeling that we observed high volatility persistence from $\gamma_{1k} + \xi_{1k}$ for $k = 1, 2, 3, 4, 5$.

As ϕ_{tk} captures the time-varying correlations between different stocks, from the estimates of α_0 , α_1 , and α_2 , we can learn the time series properties of the changing correlations. From the estimate of α_1 , which is 0.9314, the prior mean of ϕ_{tk} is very much dependent on the previous day prior mean, $\mu_{t-1,k}$. The sum of α_1 and α_2 is close to 1, indicating that μ_{tk} , the prior mean of ϕ_{tk} , is highly persistent. In other words, the impact of $\phi_{t-j,k}$ on ϕ_{tk} decays slowly with time lag j .

Equation (3) enables us to infer the dependence between the stocks. Figure 3(a) shows the estimates of ϕ_{tk} for $k = 2, 3, 4, 5$ over time. It can be seen that the estimates of ϕ_{tk} are generally different from zero and the range of ϕ_{tk} is from -1 to 2, meaning that the correlations between the stocks are generally different from zero and are significant. Figures 3(c) and 3(b) also report the variances of the stocks and the correlations between the stocks over time. The variances and correlations are calculated by the estimates in Table 4 with h_{1k} set at the long-run variance, $\gamma_{0k}/(1 - \gamma_{1k} - \xi_{1k})$ determined by the estimates of γ_{0k} , γ_{1k} and ξ_{1k} for $k = 1, 2, 3, 4, 5$. Figure 3 also presents time series plots of ρ_{tij} , the correlation between stock i and stock j at time t . We observe that the majority of the time-varying correlations are positive in the period 2007-2017. High persistence in correlations, where the time-varying correlations exhibit slowly-moving stochastic trends, is

observed. In Figure 3(c), we observe an obvious peak in 2008 around the period of global financial instability due to the “financial tsunami”. Interestingly, the estimated correlations are relatively lower in mid 2008 and started to climb after the financial tsunami.

7. Discussion

The proposed Bayesian shrinkage method allows the covariance matrix of multiple returns variables to change over time. The main idea is to express the prior mean of ϕ_{tk} as a weighted average of $\phi_{t-1,k}$ and $\mu_{t-1,k}$, the prior mean at time $t - 1$.

The real data sampling has achieved good convergence properties. In Table 4, the standard deviations are small after disregarding the initial burn-in period, meaning that after the initial burn-in period the iterates have converged. From the trace plots in Figure 1 and Figure 2, we can see that the MCMC iterates converge quite quickly. It is expected that an efficient Bayesian sampling strategy for modeling possible sparsity in ρ_{tij} via ϕ_{tk} is essential for effective modeling of time-varying correlations in financial markets.

Table 1

Summary statistics of the posterior mean estimators based on the replications that were successfully run for the first and second parameter settings

Parameter	1 st parameter setting				2 nd parameter setting			
	True value	Mean	Median	SD	True value	Mean	Median	SD
α_0	0.010	0.038	0.034	0.023	0.010	0.017	0.016	0.007
α_1	0.900	0.753	0.772	0.116	0.800	0.731	0.743	0.078
α_2	0.050	0.069	0.067	0.022	0.100	0.111	0.110	0.024
γ_{01}	0.050	0.065	0.064	0.016	0.050	0.066	0.064	0.016
γ_{11}	0.100	0.113	0.092	0.014	0.100	0.110	0.109	0.016
ξ_{11}	0.850	0.825	0.827	0.027	0.850	0.826	0.828	0.027
γ_{02}	0.050	0.061	0.060	0.015	0.050	0.063	0.061	0.015
γ_{12}	0.100	0.107	0.152	0.021	0.100	0.110	0.109	0.016
ξ_{12}	0.850	0.835	0.837	0.025	0.850	0.829	0.831	0.026
γ_{03}	0.050	0.062	0.061	0.015	0.050	0.060	0.059	0.015
γ_{13}	0.100	0.107	0.109	0.017	0.100	0.106	0.105	0.015
ξ_{13}	0.850	0.833	0.835	0.026	0.850	0.835	0.837	0.025
γ_{04}	0.050	0.063	0.062	0.014	0.050	0.067	0.066	0.016
γ_{14}	0.100	0.109	0.113	0.019	0.100	0.113	0.112	0.016
ξ_{14}	0.850	0.831	0.833	0.024	0.850	0.822	0.823	0.026
γ_{05}	0.050	0.060	0.059	0.014	0.050	0.063	0.061	0.015
γ_{15}	0.100	0.105	0.085	0.015	0.100	0.108	0.107	0.015
ξ_{15}	0.850	0.832	0.834	0.023	0.850	0.828	0.829	0.025

Table 2

Summary statistics of the posterior mean estimators based on the replications that were successfully run for the third and fourth parameter settings

Parameter	3 rd parameter setting				4 th parameter setting			
	True value	Mean	Median	SD	True value	Mean	Median	SD
α_0	0.010	0.017	0.016	0.007	0.100	0.051	0.049	0.018
α_1	0.800	0.735	0.749	0.081	0.800	0.866	0.868	0.030
α_2	0.100	0.109	0.107	0.024	0.100	0.082	0.081	0.014
γ_{01}	0.050	0.056	0.056	0.010	0.100	0.118	0.116	0.023
γ_{11}	0.200	0.211	0.210	0.024	0.100	0.107	0.107	0.014
ξ_{11}	0.700	0.682	0.683	0.034	0.800	0.783	0.785	0.030
γ_{02}	0.050	0.057	0.056	0.010	0.100	0.117	0.116	0.020
γ_{12}	0.200	0.212	0.211	0.024	0.100	0.108	0.107	0.012
ξ_{12}	0.700	0.679	0.681	0.034	0.800	0.784	0.786	0.026
γ_{03}	0.050	0.055	0.054	0.009	0.100	0.118	0.117	0.016
γ_{13}	0.200	0.207	0.206	0.023	0.100	0.107	0.106	0.010
ξ_{13}	0.700	0.686	0.687	0.033	0.800	0.784	0.785	0.021
γ_{04}	0.050	0.056	0.056	0.009	0.100	0.117	0.116	0.013
γ_{14}	0.200	0.214	0.213	0.023	0.100	0.106	0.106	0.008
ξ_{14}	0.700	0.678	0.679	0.032	0.800	0.785	0.785	0.018
γ_{05}	0.050	0.056	0.055	0.009	0.100	0.118	0.117	0.012
γ_{15}	0.200	0.205	0.204	0.021	0.100	0.106	0.106	0.007
ξ_{15}	0.700	0.680	0.681	0.031	0.800	0.784	0.784	0.016

Table 3

Summary statistics of percentage returns

Stock	Mean	SD	min	median	max
0001 HK Equity	0.0261	1.9847	-13.9763	0.0000	15.4151
0101 HK Equity	-0.0035	2.3540	-10.0587	0.0000	17.9938
0011 HK Equity	0.0349	1.5417	-12.2893	0.0000	13.1792
1044 HK Equity	0.0444	2.0474	-17.0855	0.0000	9.8812
1088 HK Equity	-0.0077	2.7596	-26.8474	0.0000	26.2277

Table 4**Summary statistics of real data sampling**

Parameter	Posterior mean	Posterior standard deviation
α_0	0.0017	0.0002
α_1	0.9314	0.0038
α_2	0.0635	0.0040
γ_{01}	0.0842	0.0211
γ_{11}	0.0709	0.0112
ξ_{11}	0.9160	0.0140
γ_{02}	0.0947	0.0218
γ_{12}	0.0667	0.0107
ξ_{12}	0.9261	0.0119
γ_{03}	0.0571	0.0085
γ_{13}	0.1467	0.0139
ξ_{13}	0.8485	0.0141
γ_{04}	0.3008	0.0868
γ_{14}	0.1540	0.0283
ξ_{14}	0.8076	0.0385
γ_{05}	0.2580	0.0495
γ_{15}	0.0680	0.0104
ξ_{15}	0.8959	0.0154
τ_{21}	0.1852	0.0299
τ_{31}	0.1296	0.0304
τ_{32}	0.0373	0.0128
τ_{41}	0.1313	0.0359
τ_{42}	0.1111	0.0212
τ_{43}	0.0996	0.0180
τ_{51}	0.2208	0.0372
τ_{52}	0.1933	0.0297
τ_{53}	0.2174	0.0277
τ_{54}	0.2332	0.0297

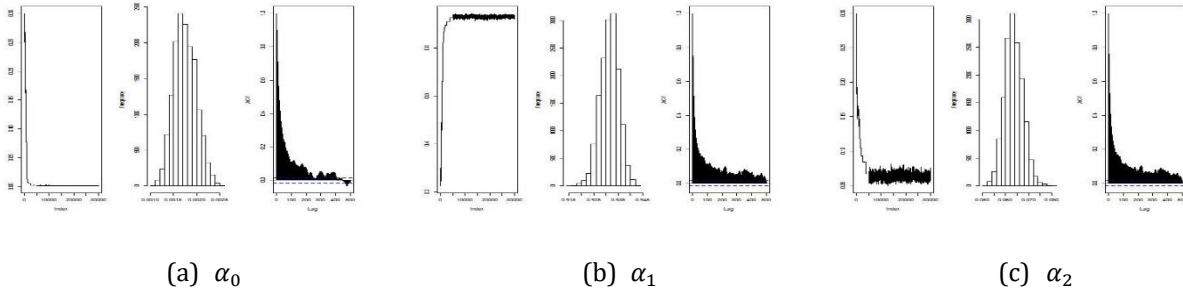


Figure 1

Trace plots, histograms and autocorrelation plots of α for real data sampling

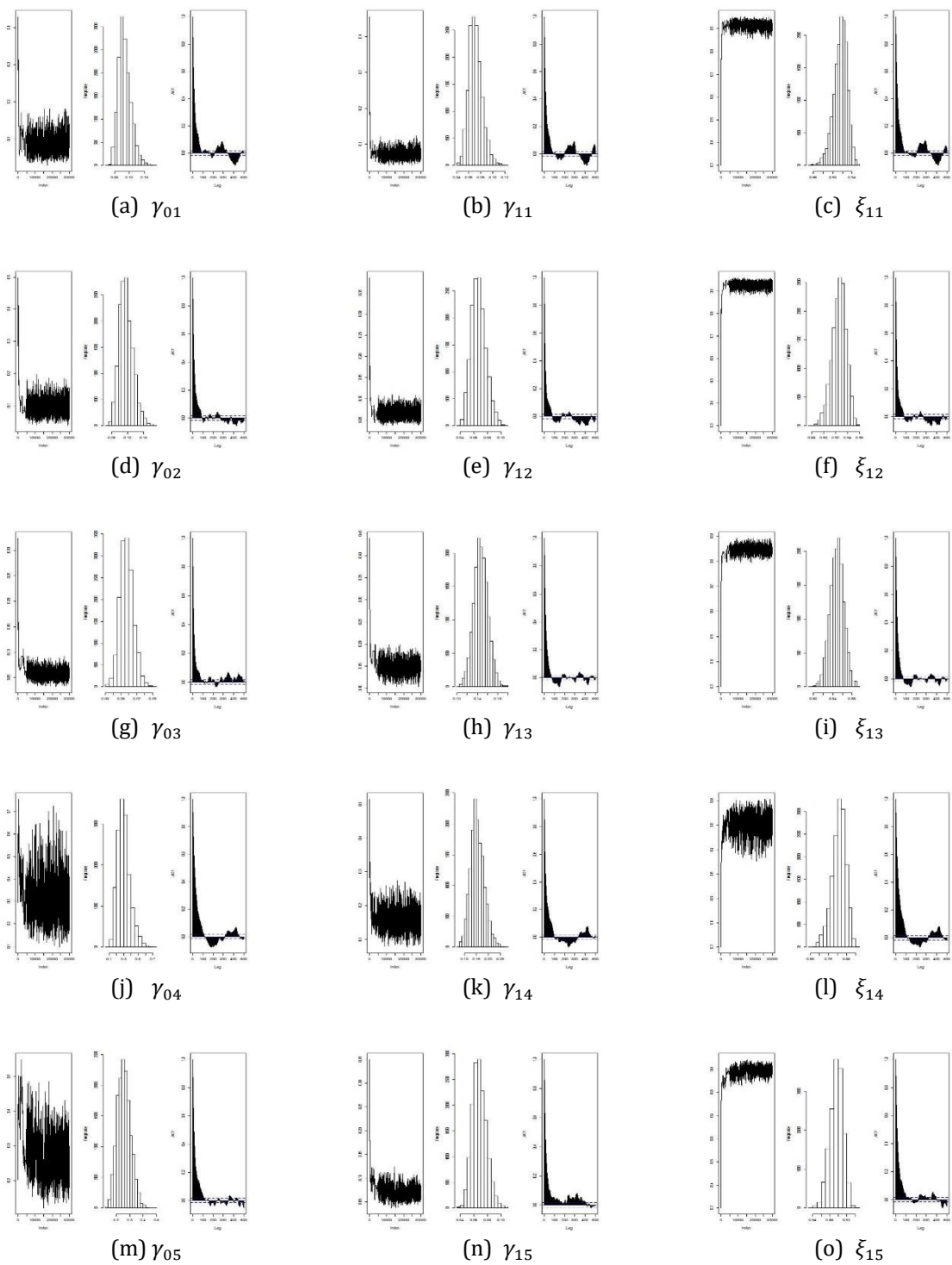
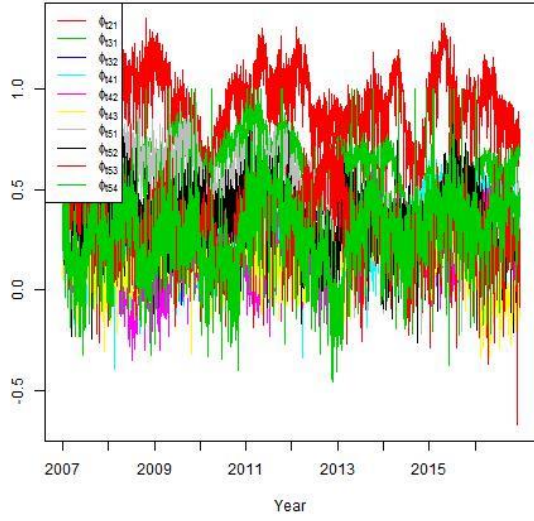
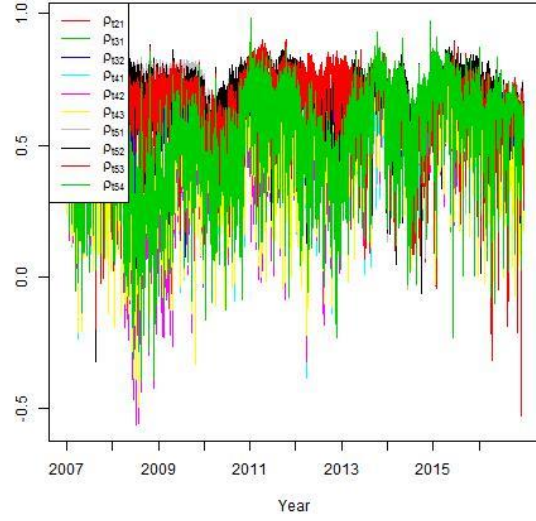


Figure 2

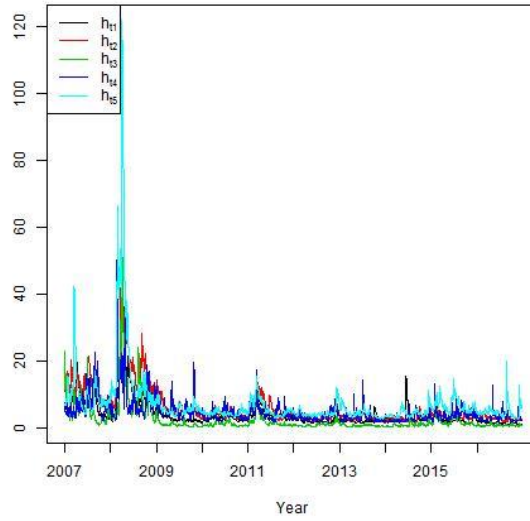
Trace plots, histograms and autocorrelation plots of GARCH parameters for real data sampling



(a) Estimate of ϕ_{tk}



(b) Implied correlations of the stocks



(c) Implied variance of the stocks

Figure 3

Implied ϕ_{tk} , correlations and variances generated by estimates

A. Proposal density derivation of ϕ_{tj}

Let $Y_{tj} = (y_{t,1}, \dots, y_{t,j-1})'$ and z_j be a vector containing only 1s. The proposal density function of ϕ_{tj} is:

$$\begin{aligned}
& q(\phi_{tj}) \\
& \propto P(\mathbf{r}|\boldsymbol{\theta}) \prod_{w=t}^N P(\phi_{wj}|\tau_k^2) \\
& \propto P(\mathbf{r}|\boldsymbol{\theta}) \prod_{w=t}^N P(\phi_{wj}|\tau_k^2) \\
& \propto \exp\left(-\frac{1}{2}\left(\frac{r_{tj}}{\sqrt{h_{tj}}}\sqrt{\text{var}(y_{tj})} - \sum_{u=1}^{j-1} \frac{r_{tu}}{\sqrt{h_{tu}}}\sqrt{\text{var}(y_{tu})}\right)^2\right) * \prod_{w=t}^N \exp\left(-\frac{1}{2}(\phi_{wj} - \mu_{wj})'H_j^{-1}(\phi_{wj} - \mu_{wj})\right) \\
& \propto \exp\left(-\frac{1}{2}(y_{tj} - \phi_{tj}'Y_{tj})^2\right) * \exp\left(-\frac{1}{2}(\phi_{tj} - \mu_{tj})'H_j^{-1}(\phi_{tj} - \mu_{tj})\right) \\
& \quad * \prod_{w=t+1}^N \exp\left(-\frac{1}{2}(\phi_{wj} - \mu_{wj})'H_j^{-1}(\phi_{wj} - \mu_{wj})\right) \\
& \propto \exp\left(-\frac{1}{2}(\phi_{tj}'Y_{tj}Y_{tj}'\phi_{tj} - 2\phi_{tj}'y_{tj}Y_{tj})\right) * \exp\left(-\frac{1}{2}(\phi_{tj}'H_j^{-1}\phi_{tj} - 2\phi_{tj}'H_j^{-1}\mu_{tj})\right) \\
& \quad * \prod_{w=t+1}^N \exp\left(-\frac{1}{2}(\phi_{wj} - \mu_{wj})'H_j^{-1}(\phi_{wj} - \mu_{wj})\right)
\end{aligned}$$

From Equation (4), we have:

$$\begin{aligned}
\mu_{wj} &= \alpha_0 + \alpha_1\mu_{w-1,k} + \alpha_2\phi_{w-1,k} \\
&= \alpha_0 + \alpha_2\phi_{w-1,j} + \alpha_1(\alpha_0 + \alpha_1\mu_{w-2,k} + \alpha_2\phi_{w-2,k}) \\
&= \alpha_0(1 + \alpha_1 + \dots + \alpha_1^{w-t-1})z_j + \alpha_2 \sum_{u=t}^{w-1} \alpha_1^{w-1-u} \phi_{u,j} + \alpha_2^{w-t}\mu_{t,j} \\
&= \frac{\alpha_0(1 - \alpha_1^{w-t})}{1 - \alpha_1} z_j + \alpha_2 \sum_{u=t}^{w-1} \alpha_1^{w-1-u} \phi_{u,j} + \alpha_1^{w-t}\mu_{t,j}
\end{aligned}$$

Therefore, the proposal density function of ϕ_{tj} is:

$$\begin{aligned}
& q(\phi_{tj}) \\
& \propto \exp\left(-\frac{1}{2}(\phi_{tj}'Y_{tj}Y_{tj}'\phi_{tj} - 2\phi_{tj}'y_{tj}Y_{tj})\right) * \exp\left(-\frac{1}{2}(\phi_{tj}'H_j^{-1}\phi_{tj} - 2\phi_{tj}'H_j^{-1}\mu_{tj})\right) \\
& \quad * \prod_{w=t+1}^N \exp\left(-\frac{1}{2}(\phi_{wj} - \mu_{wj})'H_j^{-1}(\phi_{wj} - \mu_{wj})\right) \\
& \propto \exp\left(-\frac{1}{2}(\phi_{tj}'y_{tj}Y_{tj}' - 2\phi_{tj}'y_{tj}Y_{tj})\right) * \exp\left(-\frac{1}{2}(\phi_{tj}'H_j^{-1}\phi_{tj} - 2\phi_{tj}'H_j^{-1}\mu_{tj})\right) \\
& \quad * \prod_{w=t+1}^N \exp\left(\frac{1}{2}(2\phi_{wj}'H_j^{-1}(\alpha_2 \sum_{u=t}^{w-1} \alpha_1^{w-1-u} \phi_{u,j}))\right) \\
& \quad * \prod_{w=t+1}^N \exp\left(-\frac{1}{2}\left(2\frac{\alpha_0(1-\alpha_1^{w-t})}{1-\alpha_1}z_j'H_j^{-1}(\alpha_2 \sum_{u=t}^{w-1} \alpha_1^{w-1-u} \phi_{u,j})\right)\right) \\
& \quad * \prod_{w=t+1}^N \left(-\frac{1}{2}\left(\alpha_2 \sum_{u=t}^{w-1} \alpha_1^{w-1-u} \phi_{u,j}\right)'(\alpha_2 \sum_{u=t}^{w-1} \alpha_1^{w-1-u} \phi_{u,j})\right) \\
& \quad * \prod_{w=t+1}^N \exp\left(-\frac{1}{2}\left(2\alpha_1^{w-t}\mu_{t,j}'H_j^{-1}(\alpha_2 \sum_{u=t}^{w-1} \alpha_1^{w-1-u} \phi_{u,j})\right)\right) \\
& \propto \exp\left(-\frac{1}{2}(\phi_{tj}'Y_{tj}Y_{tj}'\phi_{tj} - 2\phi_{tj}'y_{tj}Y_{tj})\right) * \exp\left(-\frac{1}{2}(\phi_{tj}'H_j^{-1}\phi_{tj} - 2\phi_{tj}'H_j^{-1}\mu_{tj})\right) \\
& \quad * \prod_{w=t+1}^N \exp\left(\frac{1}{2}(2\phi_{wj}'H_j^{-1}(\alpha_2\alpha_1^{w-1-t}\phi_{t,j}))\right) \\
& \quad * \exp\left(-\frac{1}{2}\left(2\sum_{w=t+1}^N \frac{\alpha_0\alpha_2(1-\alpha_1^{w-t})}{1-\alpha_1}z_j'H_j^{-1}(\alpha_1^{w-1-t}\phi_{t,j})\right)\right) \\
& \quad * \prod_{w=t+1}^N \exp\left(-\frac{1}{2}\left(\alpha_2^2\alpha_1^{2w-2-2t}\phi_{t,j}'H_j^{-1}\phi_{t,j}\right)\right) \\
& \quad * \prod_{w=t+2}^N \exp\left(-\frac{1}{2}\left(2\alpha_2^2\left(\sum_{u=t+1}^{w-1} \alpha_1^{w-1-u} \phi_{u,j}\right)'\alpha_1^{w-1-t}H_j^{-1}\phi_{t,j}\right)\right) \\
& \quad * \prod_{w=t+1}^N \exp\left(-\frac{1}{2}\left(2\alpha_1^{w-t}\mu_{t,j}'H_j^{-1}(\alpha_2\alpha_1^{w-1-t}\phi_{t,j})\right)\right) \\
& \propto \exp\left(-\frac{1}{2}(\phi_{t,j}'(Y_{tj}Y_{tj}' + (\alpha_2^2\frac{1-\alpha_1^{2(N-t)}}{1-\alpha_1^2} + 1)H_j^{-1})\phi_{t,j})\right) \\
& \quad * \exp\left(-\frac{1}{2}\left(-2\right)(Y_{tj}'y_{tj}\phi_{tj} - \left(\frac{\alpha_1\alpha_2(1-\alpha_1^{2(N-t)})}{1-\alpha_1^2} - 1\right)\mu_{t,j}'H_j^{-1}\phi_{tj} - \sum_{w=t+1}^{N-1} \sum_{u=t+1}^w \alpha_1^{2w-u-t} \alpha_2^2\phi_{u,j}'H_j^{-1}\phi_{t,j}\right. \\
& \quad \left. - \sum_{w=t+1}^N \frac{(1-\alpha_1^{w-t})\alpha_1^{w-1-t}\alpha_0\alpha_2}{1-\alpha_1}z_j'H_j^{-1}\phi_{tj} + \sum_{w=t+1}^N \alpha_1^{w-t-1}\alpha_2\phi_{w,j}'H_j^{-1}\phi_{t,j})\right)
\end{aligned}$$

By grouping the terms, we obtain Equation (7).

References

- Benoit, D. F., Alhamzawi, R., and Yu, K. (2013). Bayesian Lasso Binary Quantile Regression. *Computational Statistics*, 28 (6), 2861–2873.
- Billio, M., Casarin, R., and Osuntuyi, A. (2016). Efficient Gibbs Sampling for Markov Switching GARCH Models. *Computational Statistics and Data Analysis*, 100, 37–57.
- Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31 (3), 307–327.
- Bolstad, W. M. (2009). Markov Chain Monte Carlo Sampling from Posterior. In *Understanding Computational Bayesian Statistics*. Hoboken, New Jersey, USA: John Wiley & Sons, Inc.
- Chen, C. W. S., and So, M. K. P. (2006). On a Threshold Heteroscedastic Model. *International Journal of Forecasting*, 22 (1), 73–89.
- Chen, X., Wang, Z. J., and Mckeown, M. J. (2011). A Bayesian Lasso via Reversible-Jump MCMC. *Signal Processing*, 91 (8), 1920–1932.
- Corrêa, J. M., Neto, A. C., Júnior, L. A. T., Franco, E. M. C., and Faria, A. E. (2016). Time Series Forecasting with the WARIMAX-GARCH Method. *Neurocomputing*, 216, 805–815.
- Dyhrberg, A. H. (2016). Bitcoin, Gold and the Dollar – a GARCH Volatility Analysis. *Finance Research Letters*, 16, 85–92.
- Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50 (4), 987–1007.
- Fan, J., and Li, R. (2001). Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association*, 96 (456), 1348–1360.
- Feng, S., Lian, H., and Xue, L. (2016). A New Nested Cholesky Decomposition and Estimation for the Covariance Matrix of Bivariate Longitudinal Data. *Computational Statistics and Data Analysis*, 102, 98–109.

- Franses, P. H., and Ghijssels, H. (1999). Additive Outliers, GARCH and Forecasting Volatility. *International Journal of Forecasting*, 15 (1), 1–9.
- Gefang, D. (2014). Bayesian Doubly Adaptive Elastic-Net Lasso for VAR Shrinkage. *International Journal of Forecasting*, 30 (1), 1–11.
- Guo, R., Zhu, H., Chow, S.M., and Ibrahim, J. G. (2012a). Bayesian Lasso for Semiparametric Structural Equation Models. *Biometrics*, 68 (2), 567–577.
- Guo, R., Zhu, H., Chow, S.M., and Ibrahim, J. G. (2012b). Bayesian Lasso for Semiparametric Structural Equation Models. *Biometrics*, 68 (2), 567–577.
- Guo, X., Zhang, H., Wang, Y., and Wu, J. L. (2015). Model Selection and Estimation in High Dimensional Regression Models with Group SCAD. *Statistics and Probability Letters*, 103, 86–92.
- Hefley, T. J., Hooten, M. B., Hanks, E. M., Russell, R. E., and Walsh, D. P. (2017). The Bayesian Group Lasso for Confounded Spatial Data. *Journal of Agricultural, Biological and Environmental Statistics*, 22 (1), 42–59.
- Hendrych, R., and Cipra, T. (2016). On Conditional Covariance Modelling: An Approach Using State Space Models. *Computational Statistics and Data Analysis*, 100, 304–317.
- Hu, J., Xin, X., and You, J. (2014). Model Determination and Estimation for the Growth Curve Model via Group SCAD Penalty. *Journal of Multivariate Analysis*, 124, 199–213.
- Huang, J., Horowitz, J. L., and Wei, F. (2010). Variable Selection in Nonparametric Additive Models. *The Annals of Statistics*, 38 (4), 2282–2313.
- Jebabli, I., Arouri, M., and Teulon, F. (2014). On the Effects of World Stock Market and Oil Price Shocks on Food Prices: An Empirical Investigation Based on TVP-VAR Models with Stochastic Volatility. *Energy Economics*, 45, 66–98.
- Kim, Y., Choi, H., and Oh, H. S. (2008). Smoothly Clipped Absolute Deviation on High Dimensions. *Journal of the American Statistical Association*, 103 (484), 1665–1673.

- Kyung, M., Gill, J., Ghosh, M., and Casella, G. (2010). Penalized Regression, Standard Errors, and Bayesian Lassos. *Bayesian Analysis*, 5 (2), 369–411.
- Laurent, S., Lecourt, C., and Palm, F. C. (2016). Testing for Jumps in Conditionally Gaussian ARMA–GARCH Models, a Robust Approach. *Computational Statistics and Data Analysis*, 100, 383–400.
- Li, J., Das, K., Fu, G., Li, R., and Wu, R. (2011). The Bayesian Lasso for Genome-Wide Association Studies. *Bioinformatics*, 27 (4), 516–523.
- Liu, J. S., Liang, F., and Wong, W. H. (2000). The Multiple-Try Method and Local Optimization in Metropolis Sampling. *Journal of the American Statistical Association*, 95 (449), 121–134.
- Lv, J., and Guo, C. (2017). Efficient Parameter Estimation via Modified Cholesky Decomposition for Quantile Regression with Longitudinal Data. *Computational Statistics*, 32 (3), 947–975.
- Matsuo, T., Nychka, D. W., and Paul, D. (2011). Nonstationary Covariance Modeling for Incomplete Data: Monte Carlo EM Approach. *Computational Statistics and Data Analysis*, 55 (6), 2059–2073.
- Mengersen, K. L., and Tweedie, R. L. (1996). Rates of Convergence of the Hastings and Metropolis Algorithms. *The Annals of Statistics*, 24 (1), 101–121.
- Mutshinda, C. M., and Sillanpää, M. J. (2010). Extended Bayesian Lasso for Multiple Quantitative Trait Loci Mapping and Unobserved Phenotype Prediction. *Genetics*, 186 (3), 1067–1075.
- Mutshinda, C. M., and Sillanpää, M. J. (2012). A Decision Rule for Quantitative Trait Locus Detection Under the Extended Bayesian Lasso Model. *Genetics*, 192 (4), 1483–1491.
- Narayan, P. K., Liu, R., and Westerlund, J. (2016). A GARCH Model for Testing Market Efficiency. *Journal of International Financial Markets, Institutions and Money*, 41, 121–138.
- Nino-Ruiz, E., Sandu, A., and Deng, X. (2017). A Parallel Implementation of the Ensemble Kalman Filter Based on Modified Cholesky Decomposition. *Journal of Computational Science*.

- Pandolfi, S., Bartolucci, F., and Friel, N. (2014). A Generalized Multiple-Try Version of the Reversible Jump Algorithm. *Computational Statistics and Data Analysis*, 72, 298–314.
- Park, T., and Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association*, 103 (482), 681–686.
- Pourahmadi, M. (1999). Joint Mean-Covariance Models with Applications to Longitudinal Data: Unconstrained Parameterisation. *Biometrika*, 86 (3), 677–690.
- Pourahmadi, M. (2000). Maximum Likelihood Estimation of Generalised Linear Models for Multivariate Normal Covariance Matrix. *Biometrika*, 87 (2), 425–435.
- Robert, C., and Casella, G. (2013). *Monte Carlo Statistical Methods*. New York: Springer Science & Business Media
- Schmidt, D. F., and Makalic, E. (2013). Estimation of Stationary Autoregressive Models with the Bayesian Lasso. *Journal of Time Series Analysis*, 34 (5), 517–531.
- Smetanina, E. (2017). Real-Time GARCH. *Journal of Financial Econometrics*, 15 (4), 561–601.
- So, M. K. P. (2006). Bayesian Analysis of Nonlinear and Non-Gaussian State Space Models via Multiple-Try Sampling Methods. *Statistics and Computing*, 16 (2), 125–141.
- Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58 (1), 267–288.
- Valera, H. G. A., Holmes, M. J., and Hassan, G. M. (2018). Is Inflation Targeting Credible in Asia? A Panel GARCH Approach. *Empirical Economics*, 54 (2), 523–546.
- Wang, L., Chen, G., and Li, H. (2007). Group SCAD Regression Analysis for Microarray Time Course Gene Expression Data. *Bioinformatics*, 23 (12), 1486–1494.
- Wang, M., Song, L., and Wang, X. (2016). Quadratic Approximation via the SCAD Penalty with a Diverging Number of Parameters. *Communications in Statistics - Simulation and Computation*, 45 (1), 1-16.

- Xu, J., and Mackenzie, G. (2012). Modelling Covariance Structure in Bivariate Marginal Models for Longitudinal Data. *Biometrika*, 99 (3), 649–662.
- Yuan, M., and Lin, Y. (2006). Model Selection and Estimation in Regression with Grouped Variables. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 68 (1), 49–67.
- Zhang, W., and Leng, C. (2012). A Moving Average Cholesky Factor Model in Covariance Modelling for Longitudinal Data. *Biometrika*, 99 (1), 141–150.
- Zhang, Y., and Biswas, S. (2015). An Improved Version of Logistic Bayesian Lasso for Detecting Rare Haplotype-Environment Interactions with Application to Lung Cancer. *Cancer Informatics* 14s2.
- Zhang, Y., Ye, W., and Zhang, J. (2017). A Generalized Elastic Net Regularization with Smoothed ℓ_q Penalty for Sparse Vector Recovery. *Computational Optimization and Applications*, 68 (2), 437–454.
- Zou, H., and Hastie, T. (2005). Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 67 (2), 301–320.
- Zou, H. (2006). The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, 101 (476), 1418–1429.
- Zou, H., and Zhang, H. H. (2009). On the Adaptive Elastic-Net with a Diverging Number of Parameters. *The Annals of Statistics*, 37 (4), 1733–1751.