Time-varying high-dimensional covariance matrices

Anastasios Th. Plataniotis and Petros Dellaportas

Athens University of Economics and Business

Department of Statistics

Greek Stochastics, Lefkada 2009

Outline

The problem

- We deal with N-dimensional nonstationary time series (asset returns) $\boldsymbol{Y}_t$, $t = 1, \cdots, T$, when the focus is shifted to modelling and predicting the covariances $\{\boldsymbol{\Sigma}_t; t = 1, \cdots, T\}$ of the individual vectors
- The number of parameters in $\boldsymbol{\Sigma}_t$ is $N(N+1)/2$ which grows quadratically in N.
- It is common to decompose $\boldsymbol{Y}_t = (y_{1t}, y_{2t}, \cdots, y_{Nt})$ the returns of N assets at time t as

$$\boldsymbol{Y}_t = \mu_t + \varepsilon_t, \tag{1}$$

where $\mu_t = E(\boldsymbol{Y}_t | \mathcal{F}_{t-1})$ is the conditional mean (predictor) of $\boldsymbol{Y}_t$ given the past information $\mathcal{F}_{t-1}$, and ε_t is the shock (innovation) at time t with the conditional covariance matrix $\boldsymbol{\Sigma}_t = \text{cov}(\varepsilon_t | \mathcal{F}_{t-1})$.
- Assume $\boldsymbol{Y}_t \equiv \varepsilon_t$, and focus on modeling the conditional covariance matrices $\boldsymbol{\Sigma}_t$, $t = 1, \cdots, T$.

Dynamic eigenvalue and eigenvector modelling

- We decompose $\mathbf{\Sigma}_t = \mathbf{P}_t \mathbf{\Lambda}_t \mathbf{P}_t^T$ and model $\mathbf{\Lambda}_t$ and $\mathbf{P}_t$ with an AR(1) process. Direct modelling of $\mathbf{P}_t$ is hard.

- Since $\mathbf{P}_t$ is a rotation matrix, it can be parameterised w.r.t. $N(N-1)/2$ Givens angles, each one belonging to matrix $\mathbf{G_{jt}}$:

$$\mathbf{P_t} = \prod_{j=1}^{\frac{N(N-1)}{2}} \mathbf{G_{jt}}$$

2-Dim

$$\boldsymbol{\Sigma}_t = \begin{pmatrix} \cos(\omega_t) & \sin(\omega_t) \\ -\sin(\omega_t) & \cos(\omega_t) \end{pmatrix} \begin{pmatrix} \lambda_{1t} & 0 \\ 0 & \lambda_2 t \end{pmatrix} \begin{pmatrix} \cos(\omega_t) & \sin(\omega_t) \\ -\sin(\omega_t) & \cos(\omega_t) \end{pmatrix}^T$$

Uniqueness

$$\lambda_{1t} > \lambda_{2t},\, -\frac{\pi}{2} < \omega_t < \frac{\pi}{2}$$

3-Dim

Ignoring t

$$\boldsymbol{\Sigma} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^T = \boldsymbol{G}\boldsymbol{\Lambda}\boldsymbol{G}^T = \boldsymbol{G}_{12}\boldsymbol{G}_{13}\boldsymbol{G}_{23}\boldsymbol{\Lambda}\boldsymbol{G}_{23}^T\boldsymbol{G}_{13}^T\boldsymbol{G}_{12}^T,$$

where $\boldsymbol{G} = \begin{pmatrix} \cos(\omega_{12}) & \sin(\omega_{12}) & 0 \\ -\sin(\omega_{12}) & \cos(\omega_{12}) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos(\omega_{13}) & 0 & \sin(\omega_{13}) \\ 0 & 1 & 0 \\ -\sin(\omega_{13}) & 0 & \cos(\omega_{13}) \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\omega_{23}) & \sin(\omega_{23}) \\ 0 & -\sin(\omega_{23}) & \cos(\omega_{23}) \end{pmatrix}$

N-Dim

$$\mathbf{P_t} = \prod_{k=1,l>k}^{N} \mathbf{G_{(kl)t}} = \prod_{k=1,l>k}^{N} \begin{pmatrix} \mathbf{I}_{k-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \cos(\omega_{kl,t}) & \mathbf{0} & \sin(\omega_{kl,t}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{l-k-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\sin(\omega_{kl,t}) & \mathbf{0} & \cos(\omega_{kl,t}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{N-l} \end{pmatrix}$$

Note

Every matrix contains only 4 elements with angles, ones in the diagonal, and everywhere else zeroes

The Model

- $\mathbf{Y} = \{\mathbf{Y_1}, \ldots, \mathbf{Y_T}\}$, $\mathbf{Y_t} = (y_{1t}, \ldots, y_{Nt})^T$, $\mathbf{Y_t} \sim MVN\left\{\mathbf{0}, \mathbf{G_t \Lambda_t G_t^T}\right\}$.
- Transformations: $h_{it} = \log \lambda_{it}$, $\delta_{it} = \log(\frac{\pi/2+\omega_{it}}{\pi/2-\omega_{it}})$, $i = 1, \ldots, N$, $t = 1, \ldots, T$

$$h_{i,t+1} = \mu_i^h + \phi_i^h \cdot (h_{it} - \mu_i^h) + \sigma_i^h \cdot \eta_{it}^h, \quad i = 1, \ldots, N$$

$$\delta_{j,t+1} = \mu_j^\delta + \phi_j^\delta \cdot (\delta_{jt} - \mu_j^\delta) + \sigma_j^\delta \cdot \eta_{jt}^\delta, \quad j = 1, \ldots, \frac{N(N-1)}{2}$$

where $\eta_{it}^h, \eta_{jt}^\delta \sim N\left\{0, 1\right\}$ independently, and we denote

$$\boldsymbol{\theta}_h = (\phi_1^h, \ldots, \phi_N^h, \sigma_1^h, \ldots, \sigma_N^h)$$

$$\boldsymbol{\theta}_\delta = (\phi_1^\delta, \ldots, \phi^\delta_{\frac{N(N-1)}{2}}, \sigma^\delta_{1\eta}, \ldots, \sigma^\delta_{\frac{N(N-1)}{2}\eta})$$

Estimation and model choice

- Based on Laplace approximations for large N, based on MCMC for small N.
- Inference is Bayesian, but can be also viewed as classical
- Exploit some interesting conditional independence structure of our parameterisation
- Achieve parsimony through Bayes Factors

Conditional independence [1]

Suppress t

$$
\begin{aligned}
\pi(\boldsymbol{h}|\boldsymbol{\delta}, \boldsymbol{Y}) \quad &\propto \quad \pi(h_1|\delta_{12}, \delta_{13}, \ldots, \delta_{1N}, \boldsymbol{Y}) \times \\
&\quad\; \pi(h_2|\delta_{12}, \delta_{13}, \ldots, \delta_{1N}, \delta_{23}, \delta_{24}, \ldots, \delta_{2N}, \boldsymbol{Y}) \times \\
&\quad\; \times \cdots \times \pi(h_{N-1}|\boldsymbol{\delta}, \boldsymbol{Y}) \times \pi(h_N|\boldsymbol{\delta}, \boldsymbol{Y})
\end{aligned}
$$

Conditional independence [1] cont.

For example, the log-likelihood for $N = 3$ will be

$$\begin{aligned} L &= c - \frac{1}{2}\sum_{t=1}^{T}\left(\log\left|\mathbf{P}_t\boldsymbol{\Lambda}_t\mathbf{P}_t^T\right| + \boldsymbol{Y}_t^T\left(\boldsymbol{G}_{12,t}\boldsymbol{G}_{13,t}\boldsymbol{G}_{23,t}\boldsymbol{\Lambda}_t\boldsymbol{G}_{23,t}^T\boldsymbol{G}_{13,t}^T\boldsymbol{G}_{12,t}^T\right)^{-1}\boldsymbol{Y}_t\right) \\ &= c - \frac{1}{2}\sum_{t=1}^{T}\left[\log|\boldsymbol{\Lambda}_t| + (\boldsymbol{G}_{23,t}\boldsymbol{Y}_t^*)^T\boldsymbol{\Lambda}_t^{-1}(\boldsymbol{G}_{23,t}\boldsymbol{Y}_t^*)\right] \end{aligned}$$

where $\boldsymbol{Y}_t^* = \boldsymbol{G}_{13,t}^T\boldsymbol{G}_{12,t}^T\boldsymbol{Y}$, and since

$$\boldsymbol{G}_{23,t} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\omega_{23,t}) & \sin(\omega_{23,t}) \\ 0 & -\sin(\omega_{23,t}) & \cos(\omega_{23,t}) \end{pmatrix}$$

$\boldsymbol{\lambda}_{1t}$ appears only as a $\sum_t \boldsymbol{Y}_{1t}^{*2}\boldsymbol{\lambda}_{1t}$ term in L so it is independent of $\boldsymbol{G}_{23}$.

Conditional independence [2]

Supress t

$$
\begin{aligned}
\pi(\delta_{12}|\boldsymbol{Y}) &\propto \pi(\delta_{12}|\boldsymbol{Y_1}, \boldsymbol{Y_2}) \\
\pi(\delta_{13}|\boldsymbol{Y}) &\propto \pi(\delta_{13}|\boldsymbol{Y_1}, \boldsymbol{Y_2}, \boldsymbol{Y_3}, \delta_{12}) \\
&= \pi(\delta_{13}|\boldsymbol{Y_1^*}, \boldsymbol{Y_2^*}, \boldsymbol{Y_3}, \delta_{12}), \;\; \boldsymbol{Y}^* = \boldsymbol{G}_{12}^T\boldsymbol{Y} \\
\pi(\delta_{14}|\boldsymbol{Y}) &\propto \pi(\delta_{14}|\boldsymbol{Y_1^*}, \boldsymbol{Y_2^*}, \boldsymbol{Y_3^*}, \delta_{12}\delta_{13}) \\
&\vdots
\end{aligned}
$$

Very convenient for MCMC implementation

Suggests to estimate Givens angles sequentially.

Information Matrix

Orthogonality

- $I(\boldsymbol{h}, \boldsymbol{\delta})$ is block diagonal so $\boldsymbol{h}$ and $\boldsymbol{\delta}$ are orthogonal, see Yang and Berger 1994, Annals of Statistics.
- The block $I(\boldsymbol{h})$ is diagonal as well, see Yang and Berger 1994, Annals of Statistics.
- The block $I(\boldsymbol{\delta})$ is diagonal in the case when $\boldsymbol{\delta}$ =0, see Daniels and Kass 2001, Biometrics.

The expected information matrix of $I(\boldsymbol{\delta})$ becomes diagonal, if we have transformed the data at first according to the sample eigenvector matrix, thus making $E(\boldsymbol{\delta}) = 0$. Therefore, we can perform separate maximizations for all the parameters in the spirit of Cox and Reid 1987, JRSSB and suffer small loss in the accuracy of our results.

Conditional independence [2] cont.

Supress t

$$
\begin{aligned}
\pi(\delta_{12}|\boldsymbol{Y}) &\propto \pi(\delta_{12}|\boldsymbol{Y_1},\boldsymbol{Y_2}) \\
\pi(\delta_{13}|\boldsymbol{Y}) &\propto \pi(\delta_{13}|\boldsymbol{Y_1},\boldsymbol{Y_3}) \\
&\vdots \\
\pi(\delta_{\frac{N(N-1)}{2}}|\boldsymbol{Y}) &\propto \pi(\delta_{\frac{N(N-1)}{2}}|\boldsymbol{Y_{N-1}},\boldsymbol{Y_N})
\end{aligned}
$$

Independence

This suggests to estimate Givens angles independently.

Estimation algorithm

- Perform a spectral decomposition $\boldsymbol{\Sigma} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^T$ based on sample estimates and work with the standardised vector $\boldsymbol{Y}^* = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{P}^T\boldsymbol{Y}$. Note that although $\boldsymbol{P}$ and $\boldsymbol{\Lambda}$ are substantially suboptimal estimators, the order of eigenvalues is retained in Y^*.
- Estimate separately (by running a 2-dim SV model) the marginal density of angles $\delta_{12}, \dots, \delta_{1N}$ and the corresponding $2 \times (N-1)$ $\boldsymbol{\theta}_\delta$ parameters (we exploit the first conditional independence property)
- Estimate the marginal densities of h_1 and $\boldsymbol{\theta}_{h_1}$ by running an 1-d SV model (here we exploit the orthogonality of h_1 with $\delta_{12}, \dots, \delta_{1N}$ and the the second conditional independence property).
- Proceed to the rest of rows as above

Laplace approximation for 1-D SV model

(based on Rue, Martino, Chopin, JRSSB, discussion on 15/10/08 in London)

$$\begin{aligned} y_t|h_t &\sim N\{0, \exp(\eta_t)\} \\ \eta_t &= \mu + h_t \\ h_t|h_1,\ldots,h_{t-1},\sigma,\phi &\sim N(\phi h_{t-1}, \sigma^2), \quad t = 1,\ldots,T \end{aligned}$$

Call $\boldsymbol{\theta} = (\log\frac{1+\phi}{1-\phi}, log(\sigma^2))$, $\boldsymbol{h} = (\mu, h_1, \ldots, h_T)$

$$\begin{aligned} \pi(\boldsymbol{h},\boldsymbol{\theta}|\boldsymbol{y}) &\propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{h}|\boldsymbol{\theta})\prod_i \pi(y_i|h_i,\boldsymbol{\theta}) \\ &\propto \pi(\boldsymbol{\theta})|Q(\boldsymbol{\theta})|^{T/2}\exp\left[-\frac{1}{2}\boldsymbol{h}^T Q(\boldsymbol{\theta})\boldsymbol{h} + \sum_i \log\{\pi(y_i|h_i,\boldsymbol{\theta})\}\right] \end{aligned}$$

We exploit the fact that $Q(\boldsymbol{\theta})$, the precision matrix of $\pi(\boldsymbol{h} \mid \boldsymbol{\theta})$, is very sparse.

Laplace approximation for 1-D SV model

Approximate with Laplace approximations first

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) \propto \frac{\pi(\boldsymbol{h}, \boldsymbol{\theta}, \boldsymbol{y})}{\tilde{\pi}(\boldsymbol{h} \mid \boldsymbol{\theta}, \boldsymbol{y})}\Big|_{\boldsymbol{h}=\boldsymbol{h}^*(\boldsymbol{\theta})}$$

where $\tilde{\pi}(\boldsymbol{h} \mid \boldsymbol{\theta}, \boldsymbol{y})$ is the Gaussian approximation to $\pi(\boldsymbol{h} \mid \boldsymbol{\theta}, \boldsymbol{y})$, and then

$$\tilde{\pi}(\boldsymbol{h}|\boldsymbol{y}) = \int \tilde{\pi}(\boldsymbol{h} \mid \boldsymbol{\theta}, \boldsymbol{y})\tilde{\pi}(\boldsymbol{\theta} \mid \boldsymbol{y})d\boldsymbol{\theta}$$

by (simple) numerical integration. Normality of the marginal densities is not necessarily assumed in either approximations

Rue Martino and Chopin (2008) run this algorithm in 11 seconds assuming t-errors -beats MCMC!

Laplace approximation for 2-D SV model

- The latent variables are $\boldsymbol{h} = (\boldsymbol{h}_1, \boldsymbol{h}_2, \boldsymbol{\delta})$ and $\boldsymbol{\theta} = (\phi_{h_1}, \sigma^2_{h_1}, \phi_{h_2}, \sigma^2_{h_2}, \phi_\delta, \sigma^2_\delta)$.
- The maximisations required w.r.t. $\boldsymbol{h}$ are achieved by exploiting the orthogonality of $\boldsymbol{h}_1, \boldsymbol{h}_2, \delta$.
- When we estimate the marginal density of $\boldsymbol{\theta}_{\delta_{12}}$ we use only rows $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$. The $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$ that are integrated out are not the same as those that we obtain from the full $\boldsymbol{Y}$ dataset, these will be obtained later when all angles in row 1 are obtained.

Model determination [1]

- Important to achieve parsimony, especially in $O(N^2)$ angles
- Bayes factor, DIC, predictive measures are available
- An approximation of marginal likelihood is immediately available as the normalising constant of $\pi(\boldsymbol{\theta}|\boldsymbol{Y})$ under the assumption of Normality
- A better approximation is achieved by numerical integration of $\pi(\boldsymbol{\theta}|\boldsymbol{Y})$.

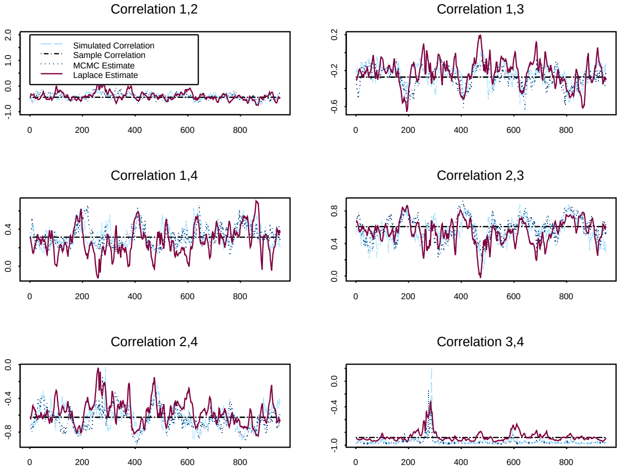
Model determination [2]

- For every latent angle vector δ_{ij}, we also estimate another, simpler model, in which the δ_{ij} is not an AR(1) process but has a Normal prior with zero mean and a unit information variance.
- In this case $\boldsymbol{\theta} = (\phi_{h_1}, \sigma^2_{h_1}, \phi_{h_2}, \sigma^2_{h_2}, \boldsymbol{\delta})$.
- Bayes factors are used according to Kass and Raftery (1985) critical values.

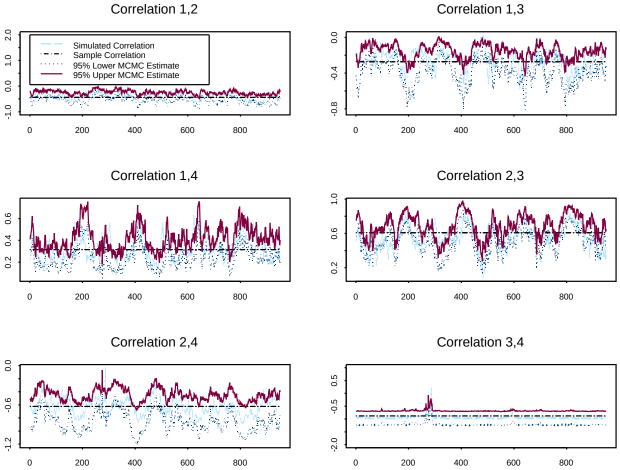
Applications

- We fitted our model to two datasets.
- The first consists of 945 simulated daily returns of 4 assets.
- The second consists of 348 daily returns of 14 stocks of FTSE 100 until June 12th 2009.
- First model has $4 \times 3/2 = 6$ angles while the second $14 \times 13/2 = 91$. Bayes factors gave correct evidence to support models with all 6 AR(1) angles for the simulated data and 9 (10%) AR(1) angles for the real data.

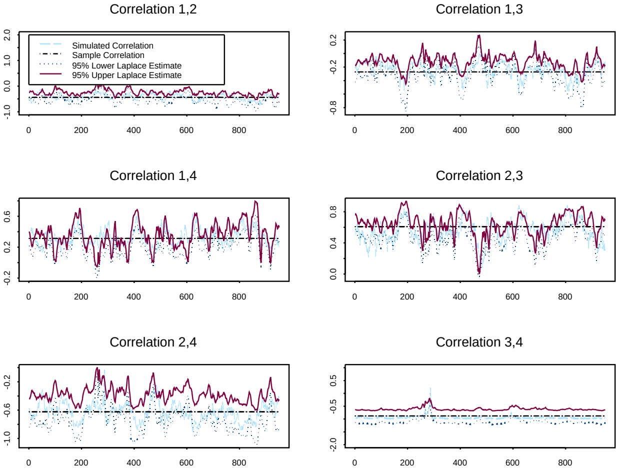
Correlations of the 4-simulated returns



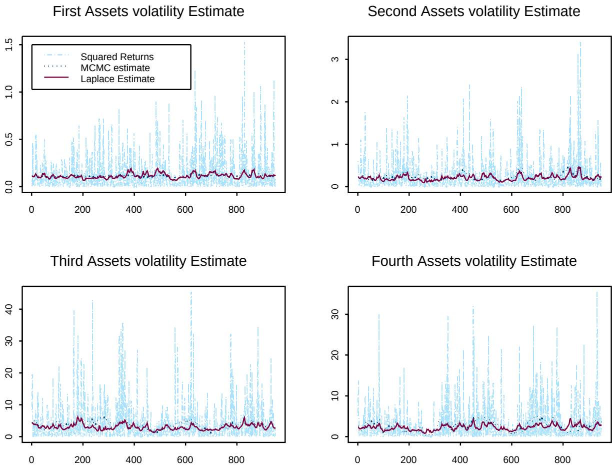
Correlations of the 4-simulated returns continued



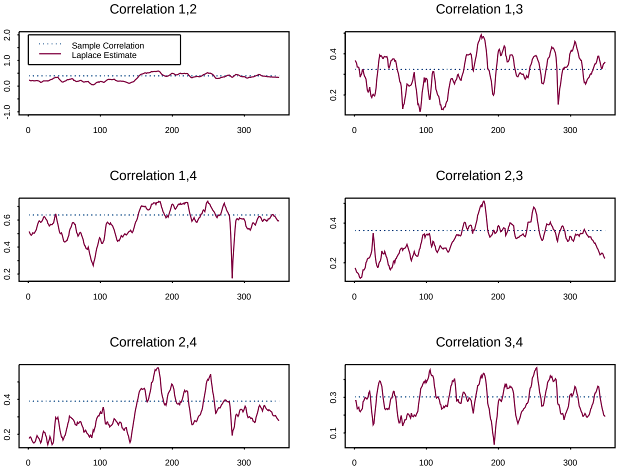
Correlations of the 4-simulated returns continued



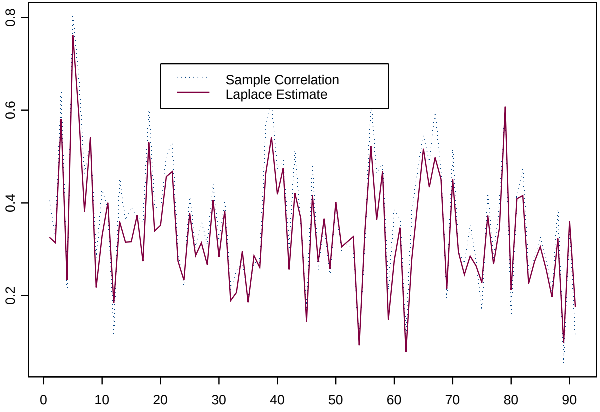
Volatilities of the 4-simulated returns



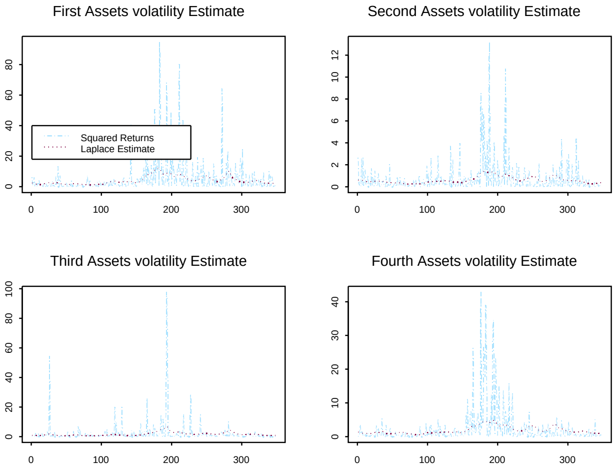
Correlations of the 14-stock data



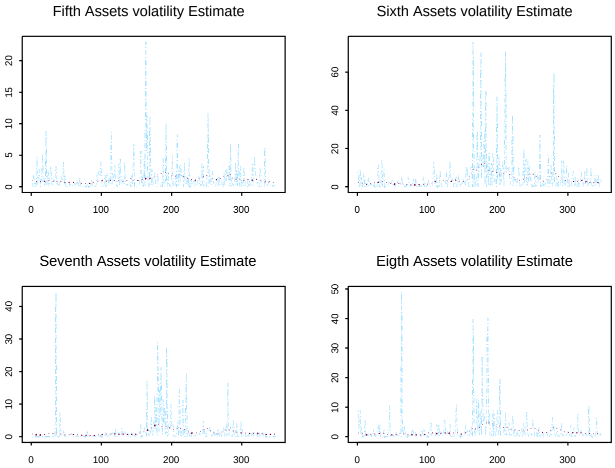
Correlations of the 14-stock data continued



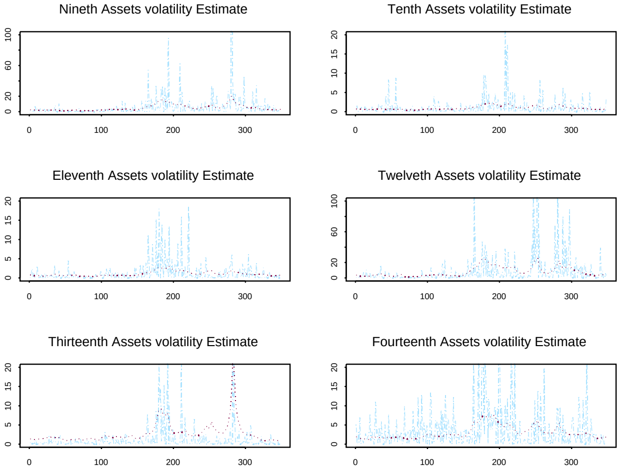
Volatilities of the 14-stock data



Volatilities of the 14-stock data



Volatilities of the 14-stock data



Current work

- Assessing the approximation error by comparing with MCMC output
- Investigate the most appropriate form of Laplace-type approximations
- Prediction theory