



**A Comparative Study on Large Multivariate Volatility Matrix Modeling for  
High-Frequency Financial Data**

Dongchen Jiang

A Thesis

Submitted to the Faculty

of

Worcester Polytechnic Institute

In partial fulfillment of the requirements for the

Degree of Master of Science

in

Applied Statistics

by

---

April 30, 2015

APPROVED :

---

Professor Jian Zou, Major Thesis Advisor

## Abstract

Modeling and forecasting the volatilities of high-frequency data observed on the prices of financial assets are vibrant research areas in econometrics and statistics. However, most of the available methods are not directly applicable when the number of assets involved is large, due to the lack of accuracy in estimating high-dimensional matrices. This paper compared two methodologies of vast volatility matrix estimation for high-frequency data. One is to estimate the Average Realized Volatility Matrix and to regularize it by banding and thresholding. In this method, first we select grids as pre-sampling frequencies, construct a realized volatility matrix using previous tick method according to each pre-sampling frequency and then take the average of the constructed realized volatility matrices as the stage one estimator, which we call the ARVM estimator. Then we regularize the ARVM estimator to yield good consistent estimators of the large integrated volatility matrix. We consider two regularizations: thresholding and banding. The other is Dynamic Conditional Correlation(DCC) which can be estimated for two stage, where in the first stage univariate GARCH models are estimated for each residual series, and in the second stage, the residuals are used to estimate the parameters of the dynamic correlation. Asymptotic theory for the two proposed methodologies shows that the estimator are consistent. In numerical studies, the proposed two methodologies are applied to simulated data set and real high-frequency prices from top 100 S&P 500 stocks according to the trading volume over a period of 3 months, 64 trading days in 2013. From the performances of estimators, the conclusion is that TARVM estimator performs better than DCC volatility matrix. And its largest eigenvalues are more stable than those of DCC model so that it is more appropriate in eigen-based analysis.

## **Acknowledgements**

I would like to express my very great appreciation to my thesis advisor, Dr. Jian Zou, who not only provided this meaningful topic but also put so much effort helping me understand essence of my thesis making it go the right way. I am also inspired and encouraged to explore and learn new knowledge during this period under his instruction. Thanks to Professor Zou for taking so much time helping me examine the thesis carefully and patiently. I wish to acknowledge the great programming advices provided by Hong Yan. Also, my special thanks are extended to all the professors who imparted to me with the valuable statistical knowledge and their assistance for the past three semesters.

# Contents

- 1 Introduction** **3**
  
- 2 Methodology** **6**
  - 2.1 Realized Volatility Process . . . . . 6
    - 2.1.1 Price Model . . . . . 6
    - 2.1.2 ARVM Estimator . . . . . 7
    - 2.1.3 Regularization of ARVM Estimator . . . . . 8
  - 2.2 DCC Process . . . . . 10
    - 2.2.1 DCC Model . . . . . 10
    - 2.2.2 Estimation . . . . . 11
  
- 3 Aysmptotic Theory** **14**
  - 3.1 Realized Volatility Process . . . . . 14
  - 3.2 DCC Process . . . . . 17
  
- 4 Numerical Studies** **19**
  - 4.1 The Simulation Model . . . . . 19
    - 4.1.1 Simulation of  $\gamma(t)$  . . . . . 20
    - 4.1.2 Simulation of the True Log Price  $\mathbf{X}(t)$  . . . . . 22
    - 4.1.3 Simulation of Observed Data  $\mathbf{Y}(t)$  . . . . . 23
    - 4.1.4 Simulation of Nonsynchronized Data . . . . . 23
  - 4.2 Two Methods in Simulation Data . . . . . 24

4.2.1	Realized Volatility Process . . . . .	24
4.2.2	DCC Method . . . . .	25
4.3	Simulation Result . . . . .	26
4.4	Example of Real Data . . . . .	32
<b>5</b>	<b>Conclusion and Discussion</b>	<b>35</b>

# Chapter 1

## Introduction

It is increasingly important in financial economics to estimate volatilities of asset returns. For low-frequency data, there exists extensive methods to estimate the volatility matrix by using GARCH, discrete stochastic volatility and diffusive stochastic volatility models. With the availability of high-frequency data, there is a surging interest on estimating volatilities using high-frequency returns directly. High-frequency data possess some unique features such as price discreteness, unequally spaced time intervals, non-synchronized trading, and leverage effect. Here, the non-synchronized issue refers to the fact that transactions for different assets often occur at distinct times and the high-frequency prices of different assets are recorded at mismatched time points. The field of high-frequency finance has been developed rapidly in past several years. And estimation of integrated volatility is one of the important issues. Realized volatility estimation process and improved GARCH(Generalized Autoregressive Conditional Heteroskedasticity) are often applied into model high-frequency financial data.

For realized volatility estimation process, estimation methods for the univariate case include realized volatility (RV) [Andersen et al. (2003), Barndorff-Nielsen and Shephard (2002)], bi-power realized variation (BPRV)[Barndorff-Nielsen and Shephard (2006)], two-time scale realized volatility (TSRV)[Zhang, Mykland and Ait-Sahalia (2005)], multiple-time scale realized volatility (MSRV) [Zhang(2006)], wavelet realized volatility (WRV)[Fan and Wang (2007)], kernel realized volatility (KRV)[Barndorff-Nielsen et al. (2008a)], pre-averaging realized volatility[Jacod et al. (2007)]

and Fourier realized volatility (FRV)[Mancino and Sanfelici (2008)]. In multiple assets part, researchers meet a non-synchronization issue and develop new methods to solving it. Hayashi and Yoshida(2005) and Zhang(2011) proposed to estimate integrated covolatility of the two assets based on overlap intervals and previous ticks, respectively. Barndorff-Nielsen et al. (2010) employed a refresh time scheme to synchronize the data and then applied a realized kernel to the synchronized data for estimating integrated covolatility. Christensen, Kinnebrock and Podolskij(2010) studied integrated covolatility estimation by the preaveraging approach.

For improved GARCH processes, the standard practice of using variance estimates based on daily returns has changed to accommodate the estimates of realized volatilities. The univariate GARCH models have met with widespread empirical success, for example, IGARCH( Integrated GARCH), EGARCH(Exponential GARCH)[Nelson(1991)], QGARCH(Quadratic GARCH)[ Sentana (1995)], TGARCH (Threshold GARCH)[ Zakoian (1994)],etc. The problems associated with the estimation of multivariate GARCH models also develop very fast. In multivariate part, three nonmutually exclusive approaches are as follows: (i) direct generalizations of the univariate GARCH model of Bollerslev(1986), for example, VEC(Vector Error-Correlation), BEKK[Baba, Engle, Kraft and Kroner] and factor models; (ii) linear combinations of univariate GARCH models, orthogonal models and latent factor models are included in this category; (iii) nonlinear combinations of univariate GARCH models, this category contains constant and dynamic conditional correlation models(DCC), the general dynamic covariance model and copula-GARCH models.

Most of these works focus on volatility estimation using high-frequency data for a single or small number of assets where volatility is either scalar or a small matrix. However, in reality, a large number of assets are usually involved in asset pricing, portfolio allocation and risk management. Due to the large number of elements in the volatility matrix, most existing methods estimator often behaves poorly. As dimension goes to infinity, the estimators such as sample covariance matrix and usual realized covolatility estimators are inconsistent because the eigenvalues and eigenvectors of the matrix estimators are far from the true targets[Johnstone(2001), Johnstone and Lu(2009), El Karoui(2007,2008),Bickel and Levina(2008a,2008b)]. Therefore, it is important to estimate integrated volatility of large size for the assets.

This paper compared two methodologies to estimate large volatility matrices considered on high-frequency prices. The first method is based on the realized volatility estimation process. The estimation procedure in the first method consists of two steps. First I select grids as pre-sampling frequencies, construct a realized volatility matrix and then take the average as the step one estimator, which is the ARVM(average realized volatility matrix) estimator. In next step, I apply two regularizations: banding and thresholding[Bickel and Levina(2008a,2008b)] to the covariance matrix estimation. As the number of assets goes larger, the volatility matrix becomes more decaying or sparser. The regularized ARVM estimators can provide better volatility estimation that will greatly enhance portfolio allocation and risk management. The second method focus on DCC(Dynamic Conditional Correlation) Multivariate GARCH model, which is first introduced in Engle(2001). The DCC model has two stage estimation, where in the first stage univariate GARCH models are estimated for each residual series, and in the second stage, residuals, transformed by their standard deviation estimated during the first stage, are used to estimate the parameters of the dynamic correlation. Then I combine estimations in two stages to get the volatility matrix. Also, the asymptotic theory for both methods are stated to show the consistency under different conditions. The proposed methods are fitted into both the simulation data and high-frequency data on 100 stocks from the top 100 according to trading volume S&P 500 trading assets .

The rest of the paper is organized as follows. The basic model for high-frequency data and the two proposed methods are presented in Chapter 2. Chapter 3 shows their asymptotic theories. Numerical studies are reported in Chapter 4. Chapter 5 makes conclusions.

# Chapter 2

## Methodology

Suppose there are  $p$  assets and their observed log price at time  $t_{il}$  in one day of the  $i$ th asset is  $Y_i(t_{il}), i = 1, \dots, p, l = 1, \dots, n_i$ . Due to the nonsynchronized problem, typically  $t_{il} \neq t_{jl}$  for any  $i \neq j$ . Let  $n_i$  be the sample size for asset  $i$ , then  $\sum_{i=1}^p n_i/p$ , is the average sample size of the  $p$  assets. Next I will illustrate two models and their estimation methodologies by reference to papers of Wang and Zou(2010) and Engle and Sheppard(2001).

### 2.1 Realized Volatility Process

#### 2.1.1 Price Model

In Diffusion processes, the true log prices at time  $t$  of  $p$  assets  $\mathbf{X}(t) = (X_1(t), \dots, X_p(t))^T$  obeys an Itô process,

$$d\mathbf{X}(t) = \boldsymbol{\mu}_t dt + \boldsymbol{\sigma}_t^T d\mathbf{B}_t, t \in [0, 1] \quad (2.1.1)$$

where  $\boldsymbol{\mu}_t = (\mu_1(t), \dots, \mu_p(t))^T$  is a drift taking values in  $\mathbb{R}^p$ ,  $\boldsymbol{\sigma}_t$  is a  $p$  by  $p$  matrix, and  $\mathbf{B}_t = (B_{1t}, \dots, B_{pt})^T$  is a  $p$ -dimensional standard Brownian motion. Both  $\boldsymbol{\mu}_t$  and  $\boldsymbol{\sigma}_t$  are continuous in  $t$ .  $\mathbf{X}_t$  has volatility matrix  $\boldsymbol{\gamma}(t) = (\gamma_{ij}(t))_{1 \leq i, j \leq p} = \boldsymbol{\sigma}_t^T \boldsymbol{\sigma}_t$ , the integrated volatility matrix for one day based on noisy and nonsynchronized observations  $Y_i(t_{il}), i = 1, \dots, p, l = 1, \dots, n_i$  is defined as

$$\boldsymbol{\Gamma} = (\Gamma_{ij})_{1 \leq i, j \leq p} = \int_0^1 \boldsymbol{\gamma}(\mathbf{t}) dt = \left( \int_0^1 \gamma_{ij}(t) dt \right)_{1 \leq i, j \leq p}$$

The high-frequency prices are usually mixed with micro-structure noise so that the observed log price  $Y_i(t_{il})$  is a noisy version of the corresponding true log price  $X_i(t_{il})$ . It is common to assume

$$Y_i(t_{il}) = X_i(t_{il}) + \varepsilon_i(t_{il}), \quad i = 1, \dots, p, l = 1, \dots, n_i \quad (2.1.2)$$

where  $\varepsilon_i(t_{il}), i = 1, \dots, p, l = 1, \dots, n_i$  are iid noise with mean zero and variance  $\eta_i$ , and  $\varepsilon_i(\cdot)$  and  $\mathbf{X}_i(\cdot)$  are independent with each other. The estimation of large volatility matrix for high-frequency prices by diffusion process has two stage. The first stage is to estimate the ARVM (Averaging Realized Volatility Matrix). The second stage is to regularize ARVM estimator by banding and thresholding.

### 2.1.2 ARVM Estimator

To understand the basic idea of realized volatility matrix estimation better, we first consider estimating one day integrated volatility matrix by averaging realized volatility matrix (ARVM) estimator [Wang and Zou(2010)]. Given a fixed integer  $m$ , suppose that  $\boldsymbol{\tau} = \{\tau_s, s = 1, \dots, m\}$  is a pre-determined sampling frequency. For asset  $i$ , define previous-tick times

$$\tau_{i,s} = \max \{t_{il} \leq \tau_s, l = 1, \dots, n_i\}$$

Based on  $\boldsymbol{\tau}$  we define realized co-volatility between assets  $i$  and  $j$  by

$$\hat{\Gamma}_{ij}(\boldsymbol{\tau}) = \sum_{s=1}^m [Y_i(\tau_{i,s}) - Y_i(\tau_{i,s-1})][Y_j(\tau_{j,s}) - Y_j(\tau_{j,s-1})] \quad (2.1.3)$$

and realized volatility matrix by

$$\hat{\mathbf{\Gamma}}(\boldsymbol{\tau}) = (\hat{\Gamma}_{ij}(\boldsymbol{\tau})) \quad (2.1.4)$$

Next, in this paper, the pre-determined sampling frequency  $\boldsymbol{\tau}$  is usually selected as regular grids.

For a fixed  $m$ , there are  $K = \lfloor n/m \rfloor$  classes of nonoverlap regular grids given by

$$\boldsymbol{\tau}^k = \{(s-1)/m, s = 1, \dots, m\} + (k-1)/n = \{(s-1)/m + (k-1)/n, s = 1, \dots, m\} \quad (2.1.5)$$

where  $k = 1, \dots, K$  and  $n$  is the average sample size for  $p$  assets. Then for each specific  $\tau^k$ , we define realized co-volatility matrix  $\hat{\Gamma}_{ij}(\tau^k)$  between assets  $i$  and  $j$  and realized volatility matrix  $\hat{\Gamma}(\tau^k)$  according to (2.1.3) and (2.1.4):

$$\hat{\Gamma}_{ij} = \frac{1}{K} \sum_{k=1}^K \hat{\Gamma}_{ij}(\tau^k), \quad \hat{\Gamma} = (\hat{\Gamma}_{ij}) = \frac{1}{K} \sum_{k=1}^K \hat{\Gamma}(\tau^k) \quad (2.1.6)$$

Then the averaging realized volatility matrix (ARVM) estimator denoted by  $\tilde{\Gamma}$  is given by

$$\tilde{\Gamma} = (\tilde{\Gamma}_{ij}) = \hat{\Gamma} - 2m\hat{\eta}, \quad (2.1.7)$$

where

$$\hat{\eta}_i = \frac{1}{2n_i} \sum_{l=1}^{n_i} [Y_i(t_{i,l}) - Y_i(t_{i,l-1})]^2. \quad (2.1.8)$$

are estimators of noise variance  $\eta_{il}$ , and  $\hat{\eta} = \text{diag}(\hat{\eta}_1, \dots, \hat{\eta}_p)$  is the estimator of  $\eta = \text{diag}(\eta_1, \dots, \eta_p)$ . Here  $\tau^k$  is used to subsample data and then we can take the average. The subsampling and averaging is to reduce the impact of microstructure noise and get a better ARVM estimator.

### 2.1.3 Regularization of ARVM Estimator

When  $p$  is small,  $\tilde{\Gamma}$  provides a good estimator for  $\Gamma$ . But  $\tilde{\Gamma}$  is not consistent for large  $p$ . The previous papers [Bickel and Levina(2008a,2008b), Johnstone(2001) and Johnstone and Lu(2009)] have already shown that the eigenvalues and the eigenvectors of  $\tilde{\Gamma}$  are far from those corresponding to  $\Gamma$ . Also the paper of Bickel and Levina(2008a,2008b) mentioned two assumptions of  $\Gamma$ : decay conditions and sparsity conditions. According to these two assumptions, we can regularize  $\tilde{\Gamma}$  with banding or thresholding. Next I will illustrate two assumptions and their corresponding regularization methods.

*Decay condition:* Assume that the elements of  $\Gamma$  decay when moving away from its diagonal,

$$|\Gamma_{ij}| \leq \frac{M}{1 + |i - j|^{\alpha+1}}, \quad 1 \leq i, j \leq p, \quad E[M] \leq C \quad (2.1.9)$$

where  $M$  is a positive random variable, and  $C$  and  $\alpha$  are positive generic constants.

For  $\mathbf{\Gamma}$  satisfying decay condition(2.1.9), its important terms are elements within a band along the diagonal, and the elements outside the band are negligible. Therefore, we regularize  $\tilde{\mathbf{\Gamma}}$  by banding a matrix, to keep only the elements in a band along its diagonal and replace others by zero. Define the Banding Averaging Realized Volatility Matrix(BARVM) estimator  $\mathcal{B}_b[\tilde{\mathbf{\Gamma}}]$  as following:

$$\mathcal{B}_b[\tilde{\mathbf{\Gamma}}] = (\tilde{\Gamma}_{ij}1(|i-j| \leq b)) \quad (2.1.10)$$

where  $b$  is a banding parameter,  $1(|i-j| \leq b)$  is the indicator of  $\{(i, j), |i-j| \leq b\}$ : So (2.1.10) can also be written as:

$$\mathcal{B}_b[\tilde{\Gamma}_{ij}] = \begin{cases} \tilde{\Gamma}_{ij} & |i-j| \leq b \\ 0 & o.w \end{cases} \quad (2.1.11)$$

*Sparsity condition:* Assume that  $\mathbf{\Gamma}$  satisfies

$$\sum_{j=1}^p |\Gamma_{ij}|^\delta \leq M\pi(p), \quad i = 1, \dots, p, \quad E[M] \leq C, \quad (2.1.12)$$

where  $M$  is a positive random variable,  $0 \leq \delta < 1$ , and  $\pi(p)$  is a deterministic function of  $p$  that grows very slowly in  $p$ .

For  $\mathbf{\Gamma}$  satisfying sparse condition(2.1.12), the small number of elements with large values are important. We need to locate those elements and estimate their values.. Therefore, we regularize  $\tilde{\mathbf{\Gamma}}$  by thresholding a matrix, to retain only the elements whose absolute values exceed a given thresholding value and to replace the others by zero. Define the Thresholding Averaging Realized Volatility Matrix(TARVM) estimator  $\mathcal{T}_\varpi[\tilde{\mathbf{\Gamma}}]$  as following:

$$\mathcal{T}_\varpi[\tilde{\mathbf{\Gamma}}] = (\tilde{\Gamma}_{ij}1(|\tilde{\Gamma}_{ij}| \geq \varpi)), \quad (2.1.13)$$

where  $\varpi$  is thresholding parameter.  $\varpi$  is the quantile of  $\tilde{\Gamma}$ . So (2.1.13) can also be written as:

$$\mathcal{T}_{\varpi}[\tilde{\Gamma}_{ij}] = \begin{cases} \tilde{\Gamma}_{ij} & |\tilde{\Gamma}_{ij}| \geq \varpi \\ 0 & o.w \end{cases} \quad (2.1.14)$$

REMARK Decay condition(2.1.9) is a special case of sparsity condition(2.1.12) with  $\delta = 1/(\alpha+1)$  and  $\pi(p) = \log p$  or  $1/(\alpha+1) < \delta < 1$  and  $\pi(p) = 1$ . In (2.1.12),  $\pi(p)$  can be 1,  $\log p$  and a small power of  $p$ .  $\delta = 0$  is the case that each row of  $\Gamma$  has at most  $M\pi(p)$  number of non-zero elements. In financial applications, sparsity is much more realistic than the decay assumption because stocks prices have no natural ordering character. Thus, in Chapter 5 real data part, TARVM estimator is designed to estimate  $\Gamma$ .

## 2.2 DCC Process

### 2.2.1 DCC Model

Consider the log returns  $\mathbf{r}(t_l)$ ,  $l = 1, \dots, n_i$  from  $p$  assets in multivariate GARCH model[Engle and Sheppard(2001)]. We condition on the sigma field, denoted by  $\mathcal{F}_{t_{l-1}}$ , generated by the past information until time  $(t_{l-1})$ . Let  $\boldsymbol{\theta}$  be a finite vector of parameters and write:

$$\mathbf{r}_{t_l} | \mathcal{F}_{t_{l-1}} = (\log(\mathbf{Y}_{t_l}) - \log(\mathbf{Y}_{t_{l-1}})) | \mathcal{F}_{t_{l-1}} = (\boldsymbol{\mu}_{t_l}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon}_{t_l}) | \mathcal{F}_{t_{l-1}}, \quad (2.2.1)$$

where  $\boldsymbol{\mu}_{t_l}(\boldsymbol{\theta})$  is the conditional mean vector and  $\boldsymbol{\varepsilon}_{t_l} = H_{t_l}^{1/2}(\boldsymbol{\theta})\mathbf{z}_{t_l}$ , where  $H_{t_l}^{1/2}(\boldsymbol{\theta})$  is a  $p \times p$  positive definite matrix. Assume that  $p \times 1$  random vector  $\mathbf{z}_{t_l}$  has  $E(\mathbf{z}_{t_l}) = 0$  and  $Var(\mathbf{z}_{t_l}) = I_p$ ,  $I_p$  is the identity matrix of order  $p$ . The conditional variance matrix of  $r_{t_l}$ :

$$Var(\mathbf{r}_{t_l}) | \mathcal{F}_{t_{l-1}} = Var_{t_{l-1}}(\boldsymbol{\varepsilon}_{t_l}) = H_{t_l}^{1/2} Var_{t_{l-1}}(\mathbf{z}_{t_l})(H_{t_l}^{1/2})' = H_{t_l}, \quad (2.2.2)$$

$$H_{t_l} \equiv D_{t_l} R_{t_l} D_{t_l} \quad (2.2.3)$$

where  $D_{t_l}$  is the  $p$  by  $p$  diagonal matrix of time varying standard deviations from univariate GARCH models with  $\sqrt{h_{iit_l}}$  on the  $i^{th}$  diagonal, i.e,  $D_{t_l} = \text{diag}(\sqrt{h_{11t_l}}, \dots, \sqrt{h_{ppt_l}})$ , here  $\sqrt{h_{iit_l}}$  is defined as any univariate GARCH( $p', q'$ ) model, so that

$$h_{iit_l} = \omega_i + \sum_{p'=1}^{P'_i} \alpha_{ip'} \varepsilon_{ii,t_l-p'}^2 + \sum_{q'=1}^{Q'_i} \beta_{iq'} h_{ii,t_l-q'}, \quad i = 1, \dots, p. \quad (2.2.4)$$

(2.2.3) is imposed by the usual GARCH restrictions for non-negativity and stationarity, such as non-negativity of variances and  $\sum_{p'=1}^{P'_i} \alpha_{ip'} + \sum_{q'=1}^{Q'_i} \beta_{iq'} < 1$ .

$R_{t_l} = (\rho_{ij,t_l})$  is the time varying correlation matrix. The proposed dynamic correlation structure is:

$$R_{t_l} = \text{diag}(q_{11,t_l}^{-1/2}, \dots, q_{pp,t_l}^{-1/2}) Q_{t_l} \text{diag}(q_{11,t_l}^{-1/2}, \dots, q_{pp,t_l}^{-1/2}) \quad (2.2.5)$$

where the  $N \times N$  symmetric positive definite matrix  $Q_{t_l} = (q_{ij,t_l})$  is given by:

$$Q_{t_l} = (1 - \sum_{m=1}^M \alpha_m - \sum_{n=1}^N \beta_n) \bar{Q} + \sum_{m=1}^M \alpha_m \mathbf{z}_{t_l-m} \mathbf{z}'_{t_l-m} + \sum_{n=1}^N \beta_n Q_{t_l-n}, \quad (2.2.6)$$

where  $\mathbf{z}_{t_l} = D_{t_l}^{-1} \varepsilon_{t_l}$ ,  $\bar{Q}$  is the  $p$  by  $p$  unconditional covariance of the standardized residuals resulting from the first stage estimation. And  $\alpha_m$  and  $\beta_n$  are non-negative scalar parameters satisfying  $\sum_{m=1}^M \alpha_m + \sum_{n=1}^N \beta_n < 1$ . Here, the assumption of normality in (2.2.1) gives rise to a likelihood function. Without this assumption, the estimator will still have the Quasi-Maximum Likelihood(QML) interpretation. Therefore, the corresponding volatility matrix  $\mathbf{\Gamma} = \sum_{l=1}^n H_{t_l}/n$ .

## 2.2.2 Estimation

The proposed DCC model can be estimated for two stage, where in the first stage univariate GARCH models are estimated for each residual series, and in the second stage, the residuals which are transformed by their standard deviation estimated during the first stage are used to estimate the parameters of the dynamic correlation. Although the normality of the innovations is rejected in most applications, the Gaussian Quasi-maximum likelihood estimator will still be a consistent estimator of  $\boldsymbol{\theta}$  by maximizing the log likelihood of multivariate normal distribution. QMLE is

consistent provided the conditional mean and the conditional variance, which has been proved by Jeantheau(1998). The log likelihood for the estimator can be expressed as:

$$\begin{aligned}
L &= -\frac{1}{2} \sum_{l=1}^n (p \log(2\pi) + \log |H_{t_l}| + \varepsilon'_{t_l} H_{t_l}^{-1} \varepsilon_{t_l}) \\
&= -\frac{1}{2} \sum_{l=1}^n (p \log(2\pi) + \log |D_{t_l} R_{t_l} D_{t_l}| + \varepsilon'_{t_l} D_{t_l}^{-1} R_{t_l}^{-1} D_{t_l}^{-1} \varepsilon_{t_l}) \\
&= -\frac{1}{2} \sum_{l=1}^n (p \log(2\pi) + 2 \log |D_{t_l}| + \log |R_{t_l}| + z'_{t_l} R_{t_l}^{-1} z_{t_l}) \\
&= -\frac{1}{2} \sum_{l=1}^n (p \log(2\pi) + 2 \log |D_{t_l}| + \varepsilon'_{t_l} D_{t_l}^{-1} D_{t_l}^{-1} \varepsilon_{t_l} - z'_{t_l} z_{t_l} + \log |R_{t_l}| + z'_{t_l} R_{t_l}^{-1} z_{t_l})
\end{aligned} \tag{2.2.7}$$

Let the parameters of the model,  $\boldsymbol{\theta}$ , be written in two groups  $(\boldsymbol{\phi}, \boldsymbol{\psi}) = (\phi_1, \phi_2, \dots, \phi_p, \psi)$ , where the elements of  $\phi_i$  correspond to the parameters of the univariate GARCH model for the  $i^{th}$  assets series,  $\phi_i = (\omega, \alpha_{1i}, \beta_{1i})$ . The log-likelihood can be written as the sum of a volatility part and a correlation part:

$$L(\boldsymbol{\theta}) = L(\boldsymbol{\phi}) + L(\boldsymbol{\psi}|\boldsymbol{\phi}) \tag{2.2.8}$$

The volatility term is

$$\begin{aligned}
L_V(\boldsymbol{\phi}|r_{t_l}) &= -\frac{1}{2} \sum_{l=1}^n (p \log(2\pi) + 2 \log |D_{t_l}| + \varepsilon'_{t_l} D_{t_l}^{-1} D_{t_l}^{-1} \varepsilon_{t_l}) \\
&= -\frac{1}{2} \sum_{l=1}^n (p \log(2\pi) + 2 \log |D_{t_l}| + \varepsilon'_{t_l} D_{t_l}^{-2} \varepsilon_{t_l}) \\
&= -\frac{1}{2} \sum_{l=1}^n (p \log(2\pi) + \sum_{i=1}^p (\log(h_{ii,t_l}) + \frac{(r_{i,t_l} - \mu_{i,t_l})^2}{h_{ii,t_l}})) \\
&= -\frac{1}{2} \sum_{l=1}^n (n \log(2\pi) + \sum_{l=1}^n (\log(h_{ii,t_l}) + \frac{(r_{i,t_l} - \mu_{i,t_l})^2}{h_{ii,t_l}}))
\end{aligned} \tag{2.2.9}$$

which is simply the sum of the log-likelihoods of the individual GARCH models for each of the assets. Once the first stage has been estimated, the second stage is estimated using the correctly

specified likelihood, conditioning on the parameters estimated in the first stage likelihood:

$$L_C(\boldsymbol{\psi}|\boldsymbol{\phi}, r_{t_l}) = -\frac{1}{2} \sum_{l=1}^n (-z'_{t_l} z_{t_l} + \log |R_{t_l}| + z'_{t_l} R_{t_l}^{-1} z_{t_l}). \quad (2.2.10)$$

The two-step approach to maximizing the likelihood is to find

$$\hat{\boldsymbol{\phi}} = \mathit{arg} \max L_V(\boldsymbol{\phi}), \quad (2.2.11)$$

and then take this value as given in the second stage:

$$\max_{\boldsymbol{\phi}} \{L_C(\boldsymbol{\psi}|\hat{\boldsymbol{\phi}})\}. \quad (2.2.12)$$

Under reasonable regularity conditions, consistency of the first step will ensure consistency of the second step. The maximum of the second step will be a function of the first step parameter estimates, so if the first step is consistent, then the second step will be consistent as long as the function is continuous in a neighborhood of the true parameters.

# Chapter 3

## Asymptotic Theory

### 3.1 Realized Volatility Process

The asymptotic theory for realized volatility process is referring to the paper of Wang and Zou(2010).First we fix some notations for the theoretical analysis. Given a  $p$ -dimensional vector  $\mathbf{x} = (x_1, \dots, x_p)^T$  and  $p$  by  $p$  matrix  $\mathbf{V} = (V_{ij})$ , define matrix norms as follows,

$$\|\mathbf{V}\|_2 = \sup\{\|\mathbf{V}\mathbf{x}\|_2, \|\mathbf{x}\|_2 = 1\}, \quad \|\mathbf{x}\|_2 = \left(\sum_{i=1}^p |x_i|^2\right)^{\frac{1}{2}}.$$

Then  $\|\mathbf{V}\|_2$  is equal to the square root of the largest eigenvalue of  $\mathbf{V}^T\mathbf{V}$ , where  $\mathbf{V}^T$  is the transpose of  $\mathbf{V}$ , and for symmetric  $\mathbf{V}$ ,  $\|\mathbf{V}\|_2$  is equal to its largest absolute eigenvalue.

Next I will state the assumptions for the asymptotic analysis[Wang and Zou(2010)].

A1: Assuming the following moment conditions on diffusion drift  $\boldsymbol{\mu}_t = (\mu_1(t), \dots, \mu_p(t))^T$  and diffusion variance  $\boldsymbol{\sigma}_t = (\sigma_{ij}(t))_{1 \leq i, j \leq p}$  in price model (2.1.1) and microstructure noise  $\varepsilon_i(t_{ij})$  in observed data model (2.1.2): for some  $\beta \geq 4$ ,

$$\max_{1 \leq i \leq p} \max_{0 \leq t \leq 1} E[|\gamma_{ii}(t)|^\beta] < \infty,$$

$$\max_{1 \leq i \leq p} \max_{0 \leq t \leq 1} E[|\mu_i(t)|^{2\beta}] < \infty,$$

$$\max_{1 \leq i \leq p} \max_{0 \leq t \leq 1} E[|\varepsilon_i(t)|^{2\beta}] < \infty.$$

A2: Each of  $p$  assets has at least one observaion between  $\tau_r^k$  and  $\tau_{r+1}^k$ . In the construction of ARVM estimator, I assume  $m = o(n)$ , where  $n = (n_1 + \dots + n_p)/p$  and

$$C_1 \leq \min_{1 \leq i \leq p} \frac{n_i}{n} \leq \max_{1 \leq i \leq p} \frac{n_i}{n} \leq C_2,$$

$$\max_{1 \leq i \leq p} \max_{1 \leq l \leq n_i} |t_{il} - t_{i,l-1}| = O(n^{-1}).$$

Assumption A1 is the minimal moment requirements for the price and microstructure noise. Assumption A2 is a technical condition that ensures adequate number of observations between grids and establishes the asymptotic theory for the proposed methodology.

**THEOREM 1.(Consistency)** Suppose models (2.1.1) and (2.1.2) satisfy assumptions A1 and A2, then for all  $1 \leq i, j \leq p$ ,

$$E(|\tilde{\Gamma}_{ij} - \Gamma_{ij}|^\beta) \leq C e_n^\beta, \quad (3.1.1)$$

where  $C$  is a generic constant that is not related to  $n$  and  $p$ , and the convergence rate  $e_n^\beta$  given below is equal to the sum of terms with powers of  $n$  and  $K = \lfloor n/m \rfloor$  which depend on whether the observed data in the model specification have micro-structure noise or not.

(1) If there is micro-structure noise in model (2.1.2),

$$e_n^\beta = (Kn^{-1/2})^{-\beta} + K^{-\beta/2} + (n/K)^{-\beta/2} + K^{-\beta} + n^{-\beta/2}.$$

Thus with  $K \sim n^{2/3}$  we have  $e_n \sim n^{-1/6}$ .

(2) If there is no micro-structure noise, i.e.  $\varepsilon_i(t_{il}) = 0$  and  $Y_i(t_{il}) = X_i(t_{il})$  in model (2.1.2),

$$e_n^\beta = (n/K)^{-\beta/2} + K^{-\beta} + n^{-\beta/2}.$$

Thus with  $K \sim n^{1/3}$  we have  $e_n \sim n^{-1/3}$ .

**REMARK 1.** Noise, nonsynchronization and discrete observations for continuous process  $\mathbf{X}(t)$  are three main factor that affect the convergence rate  $e_n$ . The true log-price  $\mathbf{X}(t)$  is not directly

observable due to micro-structure noise in high-frequency financial data. And as a continuous process,  $\mathbf{X}(t)$  is only observed with noise at discrete time points. Therefore the convergence rate  $e_n$  is slower than  $n^{-1/2}$ . In fact, the optimal convergence rate for the univariate noise case is  $n^{-1/4}$  the nonsynchronization for multiple assets also make the problem complicated. In Theorem 1, the convergence rates  $e_n^\beta$  is separated into three parts. The first part  $(Kn^{-1/2})^{-\beta} + K^{-\beta/2}$  in  $e_n^\beta$  is due to noise. Because  $\mathbf{X}(t)$  is observed at discrete time points, we need to discretize  $\mathbf{X}(t)$  and use its discretization to approximate integrated volatility matrix. The second part  $(n/K)^{-\beta/2}$  in  $e_n^\beta$  is contributed to the approximation error because of the discretizaion of  $\mathbf{X}(t)$ . The third part is  $K^{-\beta} + n^{-\beta/2}$  which is associated with nonsynchronization. All the proof of Theorem 1 is in paper of Wang and Zou(2010).

**THEOREM 2.**(Consistency of BARVM) Assume that  $\mathbf{\Gamma}$  satisfies decay condition(2.1.9). Then with the assumption A1 and A2 and models (2.1.1) and (2.1.2),

$$\|\mathcal{B}[\tilde{\mathbf{\Gamma}}] - \mathbf{\Gamma}\|_2 \leq \|\mathcal{B}[\tilde{\mathbf{\Gamma}}] - \mathbf{\Gamma}\|_\infty = O_P([e_n p^{1/\beta}]^{\alpha/(\alpha+1+1/\beta)}),$$

where banding parameter  $b$  of order  $(e_n p^{1/\beta})^{-1/(\alpha+1+1/\beta)}$  is selected.

**THEOREM 3.**(Consistency of TARVM) Assume that  $\mathbf{\Gamma}$  satisfies sparsity condition(2.1.13). Then with the assumption A1 and A2 and models (2.1.1) and (2.1.2),

$$\|\mathcal{T}[\tilde{\mathbf{\Gamma}}] - \mathbf{\Gamma}\|_2 \leq \|\mathcal{T}[\tilde{\mathbf{\Gamma}}] - \mathbf{\Gamma}\|_\infty = O_P(\pi(p)[e_n p^{2/\beta} h_{n,p}]^{1-\delta}),$$

where  $e_n$  is given in Theorem 1,  $\varpi = e_n p^{2/\beta} h_{n,p}$ , and  $h_{n,p}$  is any sequence converging to infinity arbitrarily slow with one exaple  $h_{n,p} = \log \log(n \wedge p)$ .

**REMARK 2.** For  $\mathbf{\Gamma}$  satisfying decay condition(2.1.9), the sparsity condition is held with  $\delta = 1/(\alpha + 1)$  and  $\pi(p) = \log p$ . The convergence rate in Theorem 2 corresponding to Theorem 3 under the sparsity condition has a leading factor of order  $[e_n p^{2/\beta}]^{\alpha/(\alpha+1)}$ . Also, the convergence rate in Theorem 3 is nearly equal to  $\pi(p)[e_n p^{2/\beta}]^{1-\delta}$ . Because under the noise condition  $e_n \sim n^{-1/6}$  and under the noiseless condition  $e_n \sim n^{-1/3}$ , our goal is to make  $e_n p^{2/\beta}$  converge to zero, then under the noise condition,  $p$  should grow more slowly than  $n^{\beta/12}$ , and under the noiseless case condition,

$p$  should grow more slowly than  $n^{\beta/6}$ . Comparing two convergence rate in Theorem 2 and Theorem 3, we could find that as  $\beta$  goes reasonably large, the two convergence rates are quite close to each other.

## 3.2 DCC Process

The proofs for consistency and asymptotic normality of the DCC estimators have already been provided by White(1994). In order to establish the consistency of the parameters estimated using the two stage procedure and guarantee the completeness of the probability space and measurability of the quasi-likelihood functions, I will state the following assumptions[Engle and Sheppard(2001)] close to those of White's paper.

A3:(i) For each  $\phi$  in  $\Phi$ ,  $E(\log f_1(r_{tl}, \phi))$  exists and is finite,  $l = 1, 2, \dots, n$ ,  $\log f_1(r_{tl}, \phi)$  obeys the strong uniform law of large numbers.

(ii) For each  $\theta = (\phi, \psi)$  in  $\Theta = \Phi \times \Psi$ ,  $E(\log f_2(r_{tl}, \theta))$  exists and is finite,  $l = 1, 2, \dots, n$ ,  $\log f_2(r_{tl}, \theta)$  obeys the strong ULLN.

A4:(i)  $\theta_0 = (\phi_0, \psi_0)$  is identifiably unique, interior in  $\Theta = \Phi \times \Psi$  uniformly in  $n$ ,  $\Theta$  is compact, and  $\theta_0$  satisfies the conditions of positive denitess of DCC.

(ii)  $\bar{L}_{1n}(\phi) = E(n^{-1} \sum_{l=1}^n \log f_1(r_{tl}, \phi))$  is  $O(1)$  uniformly on  $\Phi$ ,  $\bar{L}_{2n}(\phi) = E(n^{-1} \sum_{l=1}^n \log f_2(r_{tl}, \phi))$  is  $O(1)$  uniformly on  $\Theta$ .

A5: For all  $\phi$  in  $\Phi$ ,  $\nabla \bar{L}_{1n}(\phi) = E(\nabla L_{1n}(r^n, \phi)) < \infty$  where  $r^n = (r_{t_1}, r_{t_2}, \dots, r_{t_n})$ , the  $n$ - dimensional vector of observations. For all  $\theta$  in  $\Theta$ ,  $\nabla \bar{L}_{2n}(\theta) = E(\nabla L_{2n}(r^n, \theta)) < \infty$

A6:(i) For all  $\phi$  in  $\Phi$ ,  $\nabla^2 \bar{L}_{1n}(\phi) = E(\nabla^2 L_{1n}(r^n, \phi)) < \infty$ ,  $E(\nabla^2 L_{1n}(r^n, \phi))$  is continuous on  $\Phi$  uniformly in  $l = 1, 2, \dots, n$  and  $\nabla^2 \log f_1(r^n, \phi)$  obeys the strong ULLN.

(ii) For all  $\theta$  in  $\Theta$ ,  $\nabla^2 \bar{L}_{2n}(\theta) = E(\nabla^2 L_{2n}(r^n, \theta)) < \infty$ ,  $E(\nabla^2 L_{2n}(r^n, \cdot))$  is continuous on  $\Theta$  uniformly in  $l = 1, 2, \dots, n$  and  $\nabla^2 \log f_2(r^{tl}, \phi)$  obeys the strong ULLN.

A7:  $A_{11.n} = \nabla_{\phi\phi} \bar{L}_{1n}(\phi_0)$  is  $O(1)$  and uniformly negative definite.  $A_{22.n} = \nabla_{\psi\psi} \bar{L}_{2n}(\phi_0)$  is  $O(1)$  and uniformly negative definite.

A8:  $\{n^{-1/2} \nabla'_{\phi} \ln f_1(r_{tl}, \phi_0), n^{-1/2} \nabla'_{\phi} \ln f_1(r_{tl}, \phi_0)\}$  obeys the central limit condition with covariance matrix  $B_{0n}$ , and  $B_{0n}$  is  $O(1)$  and uniformly positive definite

THEOREM 4 (Consistency) Under assumptions A3-A7,  $\hat{\phi}_n \xrightarrow{p} \phi_0$  and  $\hat{\theta}_n = (\hat{\phi}_n, \hat{\psi}_n) \xrightarrow{p} \theta_0$ .

THEOREM 5 (Asymptotic Normality) Under assumptions A3-A8, for  $f_1$  and  $f_2$ ,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{A}{\rightsquigarrow} N(0, A_0^{-1} B_0 A_0'^{-1})$$

where  $A_0 = \begin{bmatrix} \nabla_{\phi\phi} \ln f_1(\phi_0) & 0 \\ \nabla_{\phi\psi} \ln f_2(\phi_0) & \nabla_{\psi\psi} \ln f_2(\phi_0) \end{bmatrix} = \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix}$  and

$$B_0 = \text{var} \left[ \sum_{l=1}^n \{n^{-1/2} \nabla'_{\phi} \ln f_1(r_{t_l}, \phi_0), n^{-1/2} \nabla'_{\psi} \ln f_2(r_{t_l}, \phi_0, \psi_0)\} \right] = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

REMARK 3 The conditions for consistency given in Theorem 4 are very weak and will be satisfied by numerous data generating processes. In Theorem 5, the asymptotic variance of  $\hat{\theta}_n$  is given by  $A_0^{-1} B_0 A_0^{-1}$ . For the first stage, by applying the partitioned inverse theorems for square matrices, the asymptotic variances of the GARCH parameters for each assets,  $\hat{\phi}_n$  are the standard robust covariance matrix estimators given by  $A_{11}^{-1} B_{11} A_{11}^{-1}$ , where  $A_{11}^{-1} B_{11} A_{11}^{-1}$  is a block diagonal matrix with the covariance matrix for the  $i^{th}$  univariate GARCH model on the  $i^{th}$  diagonal block. However, the asymptotic variance of the second stage DCC parameters is much more complicated.

## Chapter 4

# Numerical Studies

### 4.1 The Simulation Model

First, I will construct a simulation data set of  $n$  values for  $p$  assets' synchronized prices and volatility process for one day at  $t_l = l/n, l = 1, \dots, n$  according to section 2.1.1 and 2.1.2. Define  $\mathbf{Y}_t$  be observed data, from (2.1.2) we know that  $\mathbf{Y}(t)$  follows  $\mathbf{Y}_t = \mathbf{X}_t + \varepsilon_t$ , where  $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{pt}), \varepsilon_{it}$  are independent normal noise with mean zero and variance  $\eta_i$ , this part will be mentioned after the true log prices  $\mathbf{X}(t)$  are simulated. The true log prices  $\mathbf{X}(t)$  in (2.1.1) of  $p$  assets follows the following continuous time diffusion model with zero drift,

$$d\mathbf{X}_t = \boldsymbol{\sigma}_t^T d\mathbf{B}_t, \quad \mathbf{X}(t) = \mathbf{X}(0) + \int_0^t \boldsymbol{\sigma}_s^T d\mathbf{B}_s, \quad t \in [0, 1], \quad (4.1.1)$$

where  $\mathbf{B}_t = (B_{1t}, \dots, B_{pt})^T$  is a standard  $p$ -dimensional Brownian motion. Hence, the volatility of  $\mathbf{X}(t)$  is  $\boldsymbol{\gamma}(t) = \boldsymbol{\sigma}_t^T \boldsymbol{\sigma}_t = (\gamma_{ij}(t))_{1 \leq i, j \leq p}$ . Take  $\boldsymbol{\sigma}_t$  as a Cholesky decomposition of  $\boldsymbol{\gamma}(t)$ . So the first step is to simulate  $\boldsymbol{\gamma}(t)$ . Let  $\mathbf{Z}_l^c$  be a  $N(0, 1)$  random variable,  $\mathbf{Z}_l^0, \mathbf{U}_l^1$  and  $\mathbf{U}_l^2$  be three standard  $p$ -dimension Brownian motions whose coordinates are all i.i.d.  $N(0, 1)$  and independent of  $\mathbf{Z}_l^c$ . Partition  $\mathbf{Z}_l^0, \mathbf{U}_l^1$  and  $\mathbf{U}_l^2$  equally into 4 blocks respectively

$$\mathbf{Z}_l^0 = ([\mathbf{Z}_{1,l}^0]^T, [\mathbf{Z}_{2,l}^0]^T, [\mathbf{Z}_{3,l}^0]^T, [\mathbf{Z}_{4,l}^0]^T)^T,$$

$$\mathbf{U}_l^1 = ([\mathbf{U}_{1,l}^1]^T, [\mathbf{U}_{2,l}^1]^T, [\mathbf{U}_{3,l}^1]^T, [\mathbf{U}_{4,l}^1]^T)^T,$$

$$\mathbf{U}_l^2 = ([\mathbf{U}_{1,l}^2]^T, [\mathbf{U}_{2,l}^2]^T, [\mathbf{U}_{3,l}^2]^T, [\mathbf{U}_{4,l}^2]^T)^T,$$

Define  $\mathbf{Z}_{j,l}^i, i = 1, 2, j = 1, 2, 3, 4$  as eight  $p/4$ -dimensional vectors by  $\mathbf{Z}_l^i = ([\mathbf{Z}_{1,l}^i]^T, [\mathbf{Z}_{2,l}^i]^T, [\mathbf{Z}_{3,l}^i]^T, [\mathbf{Z}_{4,l}^i]^T)^T$ , where

$$\mathbf{Z}_{j,l}^1 = \rho_j \mathbf{Z}_{j,l}^0 + \sqrt{1 - \rho_j^2} \mathbf{U}_{j,l}^1, \quad \mathbf{Z}_{j,l}^2 = \rho_j \mathbf{Z}_{j,l}^0 + \sqrt{1 - \rho_j^2} \mathbf{U}_{j,l}^2. \quad (4.1.2)$$

We choose the following negative values for  $\rho_j, j = 1, 2, 3, 4$  to reflect the leverage effect,

$$\rho_j = \begin{cases} -0.62, 1 \leq j \leq p/4 \\ -0.50, p/4 \leq j \leq p/2 \\ -0.25, p/2 \leq j \leq 3p/4 \\ -0.30, 3p/4 \leq j \leq p \end{cases} \quad (4.1.3)$$

#### 4.1.1 Simulation of $\gamma(t)$

*Diagonal elements* The diagonal elements of  $\gamma(t)$  are generated from four common stochastic volatility models with leverage effect. The four volatility processes are geometric Ornstein-Uhlenbeck process, the volatility process in Nelson GARCH diffusion limit model[Wang(2002)], the sum of two CIR processes[Cox, Ingersoll and Ross(1985) and Barndorff-Nielsen and Shephard(2002)] and two-factor log-linear stochastic volatility process[Huang and Tauhen(2005)]. With  $\mathbf{B}_{t_l} = (B_{1t_l}, \dots, B_{pt_l})^T$  and  $\mathbf{W}_{t_l} = (W_{1t_l}, \dots, W_{pt_l})^T$  being simulated, by Euler method we use  $W_{it_l}$  to simulate each of four parts of  $\gamma_{ii}$ .

(1) For first  $p/4$  of  $\gamma_{ii}(t_l), i = 1, \dots, p/4$ .  $\mathbf{v}_{t_l} = (\log \gamma_{11}(t_l), \dots, \log \gamma_{p/4,p/4}(t_l))^T$ ,  $\mathbf{v}_{ii}(t_l)$  are drawn from the geometric Ornstein-Uhlenbeck model[Barndorff-Nielsen and Shephard(2002)],

$$\begin{aligned} dv_{ii}(t_l) &= -0.6(0.157 + v_{ii}(t_l))dt + 0.25dZ_{1,l}^1, \\ \mathbf{v}_{t_l} &= \mathbf{v}_{t_{l-1}} - 0.6(0.157 + \mathbf{v}_{t_{l-1}}) \frac{1}{n} + 0.25 \frac{1}{\sqrt{n}} \mathbf{Z}_{1,l}^1, \\ (\gamma_{11}(t_l), \dots, \gamma_{p/4,p/4}(t_l))^T &= \exp(\mathbf{v}_{t_l}). \end{aligned} \quad (4.1.4)$$

(2) For second  $p/4$  of  $\gamma_{ii}(t_l), i = p/4 + 1, \dots, p/2$ .  $\mathbf{v}_{t_l} = (\gamma_{p/4+1,p/4+1}(t_l), \dots, \gamma_{p/2,p/2}(t_l))^T, v_{ii}(t_l)$  are drawn from the volatility process in Nelson's GARCH diffusion limit model[Barndorff-Nielsen and Shephard(2002)],

$$\begin{aligned} dv_{ii}(t_l) &= [0.1 - v_{ii}(t_l)]dt + 0.2v_{ii}(t_l)dZ_{2,l}^1, \\ \mathbf{v}_{t_l} &= \mathbf{v}_{t_{l-1}} + (0.1 - \mathbf{v}_{t_{l-1}})\frac{1}{n} + \frac{0.2}{\sqrt{n}}\mathbf{v}_{t_{l-1}} \otimes \mathbf{Z}_{2,l}^1, \\ (\gamma_{p/4+1,p/4+1}(t_l), \dots, \gamma_{p/2,p/2}(t_l))^T &= \mathbf{v}_{t_l}. \end{aligned} \quad (4.1.5)$$

(3) For third  $p/4$  of  $\gamma_{ii}(t_l), i = p/2+1, \dots, 3p/4$ ,  $0.98(\mathbf{v}_{1,t_l} + \mathbf{v}_{2,t_l}) = (\gamma_{p/2+1,p/2+1}(t_l), \dots, \gamma_{3p/4,3p/4}(t_l))^T, v_{ii}(t_l)$  are drawn from the sum of two CIR processes[Barndorff-nielsen and Shephard(2002)],

$$\begin{aligned} dv_{1,t_l} &= 0.0429(0.108 - v_{1,t_l})dt + 0.1539\sqrt{v_{1,t_l}}dZ_{3,l}^1, \\ \mathbf{v}_{1,t_l} &= \mathbf{v}_{1,t_{l-1}} + 0.0429(0.108 - \mathbf{v}_{1,t_{l-1}})\frac{1}{n} + \frac{0.1539}{\sqrt{n}}\mathbf{v}_{1,t_{l-1}} \otimes \mathbf{Z}_{3,l}^1, \end{aligned} \quad (4.1.6)$$

$$\begin{aligned} dv_{2,t_l} &= 3.74(0.401 - v_{2,t_l})dt + 1.4369\sqrt{v_{2,t_l}}dZ_{3,l}^2, \\ \mathbf{v}_{2,t_l} &= \mathbf{v}_{2,t_{l-1}} + 3.74(0.401 - \mathbf{v}_{2,t_{l-1}})\frac{1}{n} + \frac{1.4369}{\sqrt{n}}\mathbf{v}_{2,t_{l-1}} \otimes \mathbf{Z}_{3,l}^2, \\ (\gamma_{p/2+1,p/2+1}(t_l), \dots, \gamma_{3p/4,3p/4}(t_l))^T &= 0.98(\mathbf{v}_{1,t_l} + \mathbf{v}_{2,t_l}) \end{aligned} \quad (4.1.7)$$

(4) For fourth  $p/4$  of  $\gamma_{ii}(t_l), i = 3p/4 + 1, \dots, p, v_{ii}(t_l)$  are drawn from the two-factor log-linear stochastic volatility model[Huang and Tauhen(2005)],

$$\begin{aligned} dv_{1,t_l} &= -0.00137v_{1,t_l}dt + dZ_{4,l}^1, \\ \mathbf{v}_{1,t_l} &= \mathbf{v}_{1,t_{l-1}} - 0.00137\mathbf{v}_{1,t_{l-1}}\frac{1}{n} + \frac{1}{\sqrt{n}}\mathbf{Z}_{4,l}^1, \end{aligned} \quad (4.1.8)$$

$$\begin{aligned} dv_{2,t_l} &= -1.386v_{2,t_l}dt + (1 + 0.25v_{2,t_l})dZ_{4,l}^2, \\ \mathbf{v}_{2,t_l} &= \mathbf{v}_{2,t_{l-1}} - 1.386\mathbf{v}_{2,t_{l-1}}\frac{1}{n} + \frac{1}{\sqrt{n}}(1 + 0.25\mathbf{v}_{2,t_{l-1}}) \otimes dZ_{4,l}^2, \end{aligned} \quad (4.1.9)$$

and

$$s - \exp(u) = \begin{cases} \exp^u & \text{if } u \leq \log(8.5) \\ 8.5\{1 - \log(8.5) + u^2/\log(8.5)\}^{1/2} & \text{if } u > \log(8.5) \end{cases} \quad (4.1.10)$$

With  $\gamma_{ii}(t_l)$  generated from above stochastic differential equations, we still need to adjust it to match them with those average realized volatility for top 100 S&P 500 stocks. Order the diagonal elements of the average of 64 trading daily ARVM estimators for the high-frequency data from top 100 S&P 500 stocks from the largest to the smallest. Then select the  $p$  largest diagonal elements to form a diagonal matrix of size  $p$  and multiply them by 1000. Denote the resulting diagonal matrix by  $\hat{\boldsymbol{\theta}}$ . Let  $\hat{\boldsymbol{\theta}}_1$  be the diagonal matrix with first, second, third and fourth  $p/4$  diagonal elements 1/7.44 1/0.98 1/2.82 1/1.83, respectively. To adjust the computed diagonal elements by multiplying  $\text{diag}(\gamma_{ii}(t_l))$  with  $\hat{\boldsymbol{\theta}}$  and  $\hat{\boldsymbol{\theta}}_1$ , i.e.,  $(\gamma_{11}(t_l), \dots, \gamma_{pp}(t_l))^T = \hat{\boldsymbol{\theta}} \text{diag}(\gamma_{11}(t_l), \dots, \gamma_{pp}(t_l)) \hat{\boldsymbol{\theta}}_1$ .

*Off-diagonal elements* With the diagonal elements of  $\boldsymbol{\gamma}(t)$ , we define its off-diagonal elements by

$$\gamma_{ij}(t_l) = \{k(t_l)\}^{|i-j|} \sqrt{\gamma_{ii}(t_l)\gamma_{jj}(t_l)}, \quad 1 \leq i \neq j \leq p, \quad (4.1.11)$$

where process  $k(t_l)$  is given by [Barndorff-Nielsen and Shephard(2002,2004)]

$$\begin{aligned} k(t) &= \frac{e^{2v(t)} - 1}{e^{2v(t)} + 1}, \\ dv(t) &= 0.03[0.64 - v(t)]dt + 0.118v(t)dZ_t^k \\ \text{i.e., } \mathbf{v}(t_l) &= \mathbf{v}(t_{l-1}) + 0.03[0.64 - \mathbf{v}(t_{l-1})]/n + 0.118\mathbf{v}(t_{l-1}) \otimes \mathbf{Z}_l^k / \sqrt{n}, \\ \mathbf{Z}_l^k &= \sqrt{0.96}\mathbf{Z}^c - 0.2(1, \dots, 1)\mathbf{Z}_l^0 / \sqrt{p}, \end{aligned} \quad (4.1.12)$$

$\mathbf{Z}^c$  is a standard 1-dimensional Brownian motion.

#### 4.1.2 Simulation of the True Log Price $\mathbf{X}(t)$

After obtaining the matrix  $\boldsymbol{\gamma}(t_l)$ , we compute its Cholesky decomposition to be  $\boldsymbol{\sigma}(t_l)$ . Then  $n$  values for the true log prices with leverage effect are as following:

$$\mathbf{X}_{t_l} = \mathbf{X}_{t_{l-1}} + \frac{1}{\sqrt{n}}[\boldsymbol{\sigma}_{t_{l-1}}]^T \mathbf{Z}_l^0, \quad t_l = l/n, \quad l = 1, \dots, n.$$

Note that  $\boldsymbol{\gamma}(t_l)$  should be positive definite so that we can get the Cholesky decomposition of  $\boldsymbol{\gamma}(t_l)$ , but the simulated data from Euler method cannot guarantee the positive definiteness of

$\gamma(t_l)$ , then here I use "nearPD" in R to find the nearest positive definite matrix of  $\gamma(t_l)$ . The basic idea is to find a matrix  $X$  that will minimize the Frobenius norm of  $\gamma - X$ , which is defined as  $\|\gamma - X\|_F = \sqrt{\sum_{i,j=1}^p |\gamma_{ij} - X_{ij}|^2}$  [Higham Nick(2002)]. Therefore,  $X$  can be used to take place of  $\gamma$  if  $\gamma$  is not positive definite.

### 4.1.3 Simulation of Observed Data $Y(t)$

Finally, observed data  $Y_{t_l}$  are obtained by adding to  $X_{t_l}$  normal noise  $\varepsilon_{t_l}$ ,  $l = 1, \dots, n$ , i.e.,

$$Y_{t_l} = X_{t_l} + \varepsilon_{t_l}, \quad l = 1, \dots, n,$$

where  $\varepsilon_{t_l} = (\varepsilon_{1,t_l}, \dots, \varepsilon_{p,t_l})$ ,  $\varepsilon_{i,t_l}$ , are independent normal noise with mean zero and  $Var(\varepsilon_{i,t_l}) = \eta_i$ . The standard deviation  $\sqrt{\eta_i}$  is chosen to reflect the empirical fact that relative noise level found in high frequency data. In our simulation study, we select three noise standard deviations which are translated into 0.002%, 0.004% and 0.065% of the average volatility or relative noise level, respectively. Thus, the three standard deviation are chosen as:  $0.002 \sqrt{\theta_i}, 0.005 \sqrt{\theta_i}, 0.012 \sqrt{\theta_i}$  which corresponding to low, medium and high noise level.  $\theta_i$  is the diagonal element of the average of 64 trading daily ARVM estimators for high-frequency data from top 100 S&P 500 stocks from the largest to the smallest.

### 4.1.4 Simulation of Nonsynchronized Data

For nonsynchronized data, the simulation procedure is quite similar to the above procedure. Instead of generating observations for the processes at  $n$  time points,  $\gamma(t_l)$ ,  $X_i(t_l)$  and  $Y_i(t_l)$  are simulated at  $3n$  time points  $t_l = l/(3n), l = 1, \dots, 3n$ . Next divide the  $3n$  time point  $t_l$  into  $n$  groups  $\{t_{3g-2}, t_{3g-1}, t_{3g}\}$ ,  $g = 1, \dots, n$  by grouping together three consecutive time points. From the simulated  $3n$  values of  $Y_i(t_l)$  we randomly select one time point from each group, then use the  $n$  selected values to form noisy observations for asset  $i$ . The selection procedure is applied to  $p$  assets. The obtained data are nonsynchronized due to random selection.

## 4.2 Two Methods in Simulation Data

### 4.2.1 Realized Volatility Process

In the synchronized case, after the simulation of  $\gamma(t_l)$ ,  $\mathbf{X}(t_l)$ ,  $\mathbf{Y}(t_l)$ , we numerically define integrated volatility matrix  $\mathbf{\Gamma}$  as  $\sum_{l=1}^n \gamma(t_l)/n$ , compute ARVM estimator  $\tilde{\mathbf{\Gamma}}$  according to (2.1.7) and calculate its banding  $\mathcal{B}_b[\tilde{\mathbf{\Gamma}}]$  and thresholding  $\mathcal{T}_\varpi[\tilde{\mathbf{\Gamma}}]$  as described in Section 2.1.3. Repeat the whole simulation procedure 500 times. The mean square error(MSE) of a matrix estimator is computed by averaging  $l_2$ -norms of the differences between the estimator and  $\mathbf{\Gamma}$  over 500 iterations,i.e,

$$MSE(\tilde{\mathbf{\Gamma}}) = \frac{1}{500} \sum_{i=1}^{500} \|\tilde{\mathbf{\Gamma}} - \mathbf{\Gamma}\|_2^2,$$

$$MSE(\mathcal{B}_b[\tilde{\mathbf{\Gamma}}]) = \frac{1}{500} \sum_{i=1}^{500} \|\mathcal{B}_b[\tilde{\mathbf{\Gamma}}] - \mathbf{\Gamma}\|_2^2,$$

$$MSE\mathcal{T}_\varpi[\tilde{\mathbf{\Gamma}}]) = \frac{1}{500} \sum_{i=1}^{500} \|\mathcal{T}_\varpi[\tilde{\mathbf{\Gamma}}] - \mathbf{\Gamma}\|_2^2,$$

MSEs of  $\tilde{\mathbf{\Gamma}}$ ,  $\mathcal{B}_b[\tilde{\mathbf{\Gamma}}]$  and  $\mathcal{T}_\varpi[\tilde{\mathbf{\Gamma}}]$  are used to evaluate the performance of these estimators.

In the nonsynchronized case, the true  $\mathbf{\Gamma}$  is computed by  $\sum_{g=1}^n \gamma(t_{3g})/n$ . But we use the nonsynchronized data to evaluate  $\tilde{\mathbf{\Gamma}}, \mathcal{B}_b[\tilde{\mathbf{\Gamma}}]$  and  $\mathcal{T}_\varpi[\tilde{\mathbf{\Gamma}}]$ , where the values of  $b$  and  $\varpi$  are selected by minimizing their respective MSEs. Again, we repeat the whole simulation procedure 500 times and evaluate MSEs of  $\tilde{\mathbf{\Gamma}}, \mathcal{B}_b[\tilde{\mathbf{\Gamma}}]$  and  $\mathcal{T}_\varpi[\tilde{\mathbf{\Gamma}}]$  based on the 500 repetitions.

Note that R is used to simulate the whole procedure, it is very time consuming when we compute the 500 times MSE. There is a helpful package called "parallel" in R which builds on the work of package "multicore"[Urbanek] and "snow"[Tierney] and provides dropin replacements for most of the functionality of those packages, with integrated handling of random number generation. Parallelism means running several computations at the same time and taking advantage of multiple cores or CPUs on a single system, or CPU on other system. This package handles running much larger chunks of computations in parallel. In this package, "parLapply" is parallel versions of "lapply", which can run several computations of given function at the same time. Therefore, it helps me to

save a lot of time by applying this package into programming.

#### 4.2.2 DCC Method

DCC Model is applied in nonsynchronized data set. So there are  $3n = 1200$  time points in one day. In stage one, the conditional variances for each of the univariate GARCH models need to be estimated at time point  $t_l$ ,  $l = 1, \dots, 3n$  and  $p = 100$  assets. If we select every univariate GARCH model by criteria for example AIC or BIC, the computation work is so large. For simplicity, I select  $P = 1$  and  $Q = 1$  to be the univariate GARCH model's parameters. So first, by fitting the real data set into GARCH(1,1) model, we can find the initial values for stage one. Here, we set all the univariate GARCH(1,1) models with the same initial value for simplicity. In stage two, because the statistical software R cannot allocate large vector space in the middle procedure, I use the pairwise volatility matrix between two different assets, the  $i^{th}$  asset and the  $j^{th}$  asset  $i \neq j$ ,  $i, j = 1, \dots, p$ , to approximate the  $100 \times 100$  volatility matrix between  $p = 100$  assets. Then there are  $p \times (p + 1)/2 = 5050$  pairs of two assets combination. In every pairs, DCC(1,1) with initial values of  $[0.09, 0.90]$ , DCC(2,1) with initial values of  $[0.045, 0.045, 0.90]$ , DCC(2,2) with initial values of  $[0.045, 0.045, 0.45, 0.45]$  and DCC(3,2) with initial values of  $[0.03, 0.03, 0.03, 0.45, 0.45]$  are estimated with simulation data. Although DCC(3,2) has the smallest MSE, at the same time only in stage two, the number of parameters that need to be estimated increase from 2 to 5. And this procedure will run 5050 times. And there are not very large difference among these MSEs of DCC models. Therefore, in order to make the computation faster, I select DCC(1,1) as the DCC model. In R, "ccgarch" package can implement the DCC model directly and the output gives us the matrix of  $D_{t_l}$  and  $R_{t_l}$ ,  $l = 1, \dots, n$  in one day. Then we can apply DCC(1,1) model into simulation data. After simple matrix computation, we will obtain the volatility matrix  $H_{t_l}$  for every  $t_l$ . Average  $H_{t_l}$ ,  $l = 1, \dots, n$  and define this resulting matrix  $\mathcal{D}[\tilde{\Gamma}]$  as the estimation of volatility matrix in given day. Here, define

$$MSE(\mathcal{D}[\tilde{\Gamma}]) = \| \mathcal{D}[\tilde{\Gamma}] - \Gamma \|_2^2.$$

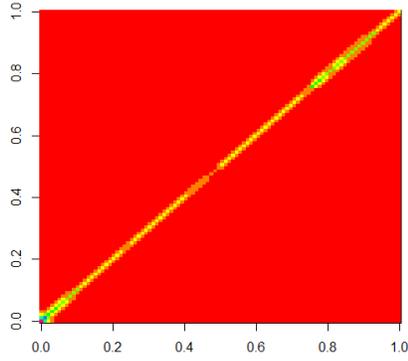
It takes a lot of time to run DCC model for large size of assets, so there is no repetition in this method.

### 4.3 Simulation Result

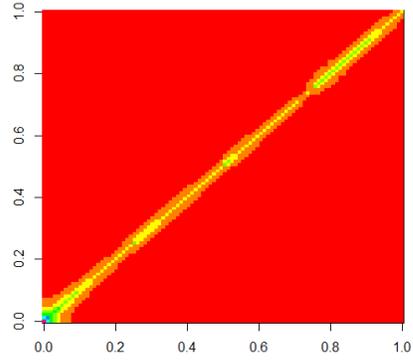
First, I focus on the discussion about the selection of two regularization methods under different conditions. Next compare the MSEs of the regularization methods with that of the DCC model for nonsynchronized high-frequency data. The simulation results and reports findings are based on  $n = 400$  and  $p = 100$  with  $m = 400$  or  $80$ , which is quite similar to the real data with 1-min and 5-min interval. In (4.1.13), five initial values of  $k(0) = 0.537, 0.762, 0.905, 0.964, 0.995$  are corresponding to five initial values of  $v(0) = 0.6, 1, 1.5, 2, 3$ .

From the previous paper[Wang and Zou(2010)], we know that  $k(t)$  is heavily influenced by its initial value and its whole path stays within a narrow band around the initial value. For every initial value of  $k(0)$ , we can compute its corresponding integrated volatility matrix  $\mathbf{\Gamma}$ . Figure 4.1 shows the images of  $\mathbf{\Gamma}$  with different initial values and the volatility matrix in DCC model. The image plots (a)-(e) indicate that decaying conditions effect the patterns of  $\mathbf{\Gamma}$  a lot. The significant elements of  $\mathbf{\Gamma}$  fall into a band along its diagonal and its off-diagonal elements outside the band are negligible. When  $k(t)$  is small, the decay is very fast and the band is very narrow. Because banding a matrix is to keep only the elements in a band along its diagonal and replace others by zero, then BARVM estimator will be more accurate than TARVM estimator under this condition. However, the larger  $k(t)$  goes, the slower the decay gets and the wider the band becomes. It indicates that BARVM and TARVM will both not perform well as  $k(t)$  increase. For example ,in Figure 1(e), when  $k(0) = 0.995$  is very large,  $\mathbf{\Gamma}$  becomes less sparse and more diffuse along its diagonal, then it will be more difficult to estiamte  $\mathbf{\Gamma}$ .

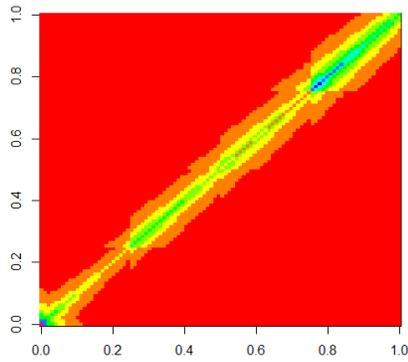
Table 4.1 summarize the MSEs of  $\tilde{\mathbf{\Gamma}}, \mathcal{B}_b[\tilde{\mathbf{\Gamma}}], \mathcal{T}_\infty[\tilde{\mathbf{\Gamma}}]$  according to five initial values of  $k(0)$ , three noise levels and two values of  $K$  for the case of synchronized data. From Table 4.1, we can find that MSEs of BARVM estimator are smaller than that of TARVM, and both of them are smaller than MSEs of ARVM, especially when  $k(0)$  is small, which confirm the conclusion from Figure 4.1 that decay has more effect on pattern than sparisy for small  $k(0)$ . In Figure 4.1, as  $k(0)$  increases, both decay and sparse pattern are so weak that it is hard to estimate  $\mathbf{\Gamma}$  at this time. This is because  $\mathbf{\Gamma}$  is not even nearly sparse for very large  $k(0)$ ,then we select almost all elements in  $\mathbf{\Gamma}$  by applying banding and thresholding methods. This may lead to the similarity among ARVM estimator, BARVM estimator



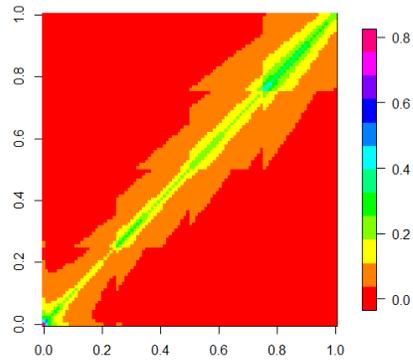
(a)  $\Gamma$  with  $k(0) = 0.537$



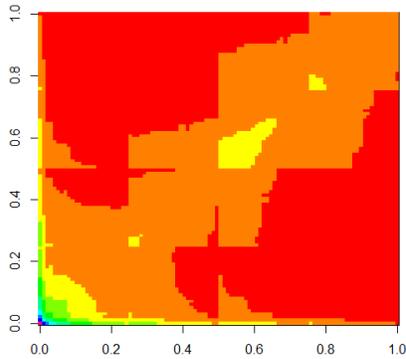
(b)  $\Gamma$  with  $k(0) = 0.762$



(c)  $\Gamma$  with  $k(0) = 0.905$



(d)  $\Gamma$  with  $k(0) = 0.964$



(e)  $\Gamma$  with  $k(0) = 0.995$

Figure 4.1: Image plots of matrix  $\Gamma$  generated with different initial values for  $k(t)$ . (a) – (e) correspond to the images of  $\Gamma$  with  $k(0) = 0.537, 0.762, 0.905, 0.964, 0.995$ , respectively. The colors is from rainbow, red, orange, yellow, green, blue, Indigo, Violet corresponding to the elements values from small to large.

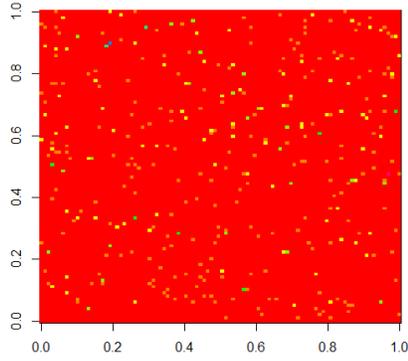
Table 4.1: MSEs of  $\tilde{\Gamma}$ ,  $\mathcal{B}_b[\tilde{\Gamma}]$  and  $\mathcal{T}_\varpi[\tilde{\Gamma}]$  for noisy synchronized data

Noise Level	Estimator	$K$	$k(0)$				
			0.537	0.762	0.905	0.964	0.995
Low	$\tilde{\Gamma}$	1	3.908	4.082	6.714	8.527	13.002
Low	$\tilde{\Gamma}$	5	8.047	11.394	14.87	16.468	29.075
Low	$\mathcal{B}_b[\tilde{\Gamma}]$	1	0.727	1.411	3.66	5.186	
Low	$\mathcal{T}_\varpi[\tilde{\Gamma}]$	1	0.826	2.351	5.001	7.81	
Low	$\mathcal{B}_b[\tilde{\Gamma}]$	5	1.509	2.631	5.491	9.487	
Low	$\mathcal{T}_\varpi[\tilde{\Gamma}]$	5	1.758	3.452	6.998	11.57	
Medium	$\tilde{\Gamma}$	1	4.12	4.377	6.921	9.269	13.551
Medium	$\tilde{\Gamma}$	5	8.312	11.624	15.04	16.883	30.746
Medium	$\mathcal{B}_b[\tilde{\Gamma}]$	1	0.784	1.568	3.955	6.233	
Medium	$\mathcal{T}_\varpi[\tilde{\Gamma}]$	1	0.898	2.611	5.374	8.22	
Medium	$\mathcal{B}_b[\tilde{\Gamma}]$	5	1.664	2.735	5.62	9.518	
Medium	$\mathcal{T}_\varpi[\tilde{\Gamma}]$	5	1.86	3.728	7.089	11.96	
High	$\tilde{\Gamma}$	1	4.26	4.577	6.989	9.564	14.26
High	$\tilde{\Gamma}$	5	8.619	12.07	15.433	17.265	32.822
High	$\mathcal{B}_b[\tilde{\Gamma}]$	1	0.81	1.754	4.142	6.771	
High	$\mathcal{T}_\varpi[\tilde{\Gamma}]$	1	0.918	2.81	5.557	8.613	
High	$\mathcal{B}_b[\tilde{\Gamma}]$	5	1.7	2.824	5.96	10.251	
High	$\mathcal{T}_\varpi[\tilde{\Gamma}]$	5	1.972	3.38	7.266	12.548	

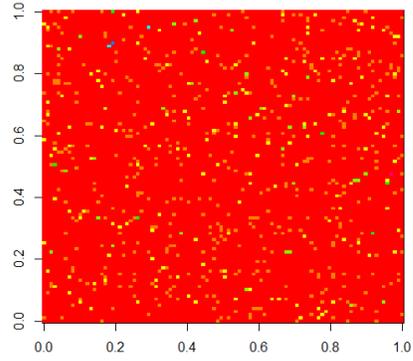
and TARVM estimator. Also, the MSEs of three estimators are more sensitive in initial values for  $k(0)$  than noise level. And for all three noise level, the estimators with  $K = 1$  perform better than that with  $K = 5$ . Therefore, in nonsynchronized data, I only choose the case for  $K = 1$ .

The significant decay pattern of  $\mathbf{\Gamma}$  will be changed if we randomly permute the rows and columns of  $\mathbf{\Gamma}$ , the resulting matrix no longer decays along its diagonal but retains the same sparsity. This procedure imitates the property of nonsynchronization in real high-frequency prices. Figure 4.2 plots the image of matrices obtained by randomly permuting rows and columns of  $\mathbf{\Gamma}$ . It shows that the significantly large elements are scattered all over the place but the decay patterns completely disappear. Because of the remaining sparsity pattern, TARVM estimator will perform better than BARVM estimator.

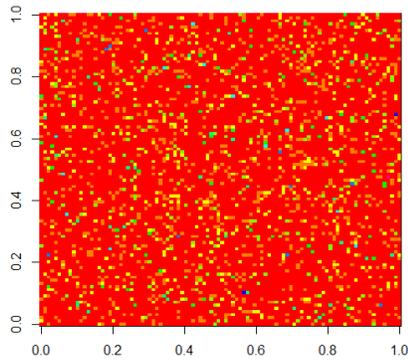
Also in the nonsynchronized data part, I apply DCC model into the data to obtain the MSEs. All the results for MSEs of  $\tilde{\mathbf{\Gamma}}$ ,  $\mathcal{D}[\tilde{\mathbf{\Gamma}}]$ ,  $\mathcal{B}_b[\tilde{\mathbf{\Gamma}}]$  and  $\mathcal{T}_\omega[\tilde{\mathbf{\Gamma}}]$  are presented in Table 4.2. The comparison of Table 4.1 and Table 4.2 shows that the MSEs in Table 4.2 are much larger than the corresponding ones in Table 4.1 for all three noise levels and five initial values of  $k(0)$  considered. From the comparison we can conclude that nonsynchronization contributes more on MSEs than noise. The MSEs of DCC model are larger than MSEs of TARVM estimator, but they are very similar with MSEs of  $\tilde{\mathbf{\Gamma}}$  and BARVM estimators. In DCC estimation stage two procedure, in order to make the computation feasible, I select the pairwise estimation between two different assets and combine all the element in  $100 \times 100$  matrix. To find the difference of volatility matrix between directly applying DCC model into  $p = 100$  assets and that of the method I proposed, I have tried this procedure in a smaller size of assets. I tried  $p = 10$  or  $20$  and compared the volatility matrix with that of the proposed DCC methods. The results show that the largest eigenvalue of estimated volatility matrix in proposed methods are a little larger than that of  $p = 10$  or  $20$ . And when  $p = 20$ , the largest eigenvalue is larger than that when  $p = 10$ . But if we compared the largest eigenvalue of estimated volatility matrix from directly apply DCC model into  $p = 10$  and  $20$  assets with those of BARVM estimator and TARVM estimator, TARVM estimator still perform better than DCC methods. Thus, we may deduce that when  $p = 100$ , the MSE, largest eigenvalue of estimated volatility matrix will be larger than the MSE if we apply DCC directly into 100 assets. But due to the software and the space



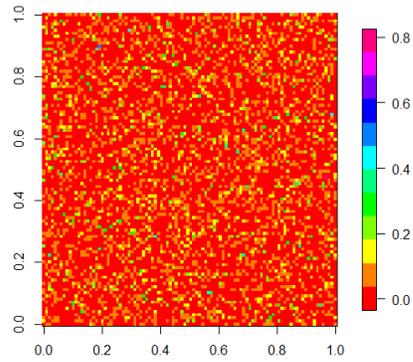
(a)  $\Gamma$  with  $k(0) = 0.537$



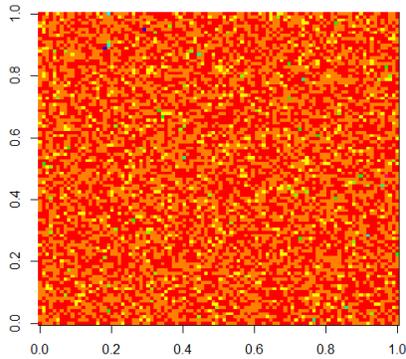
(b)  $\Gamma$  with  $k(0) = 0.762$



(c)  $\Gamma$  with  $k(0) = 0.905$



(d)  $\Gamma$  with  $k(0) = 0.964$



(e)  $\Gamma$  with  $k(0) = 0.995$

Figure 4.2: Image plots of the matrix obtained by randomly permuting rows and columns of  $\Gamma$  in Figure 4.1. (a) – (e) correspond to the images of randomly permuted  $\Gamma$  with  $k(0) = 0.537, 0.762, 0.905, 0.964, 0.995$ , respectively. The colors are from rainbow, red, orange, yellow, green, blue, Indigo, Violet corresponding to the elements values from small to large.

Table 4.2: MSEs of  $\tilde{\Gamma}$ ,  $\mathcal{D}[\tilde{\Gamma}]$ ,  $\mathcal{B}_b[\tilde{\Gamma}]$  and  $\mathcal{T}_\varpi[\tilde{\Gamma}]$  for noisy nonsynchronized data

Noise Level	Estimator	$K$	$k(0)$				
			0.537	0.762	0.905	0.964	0.995
Low	$\tilde{\Gamma}$	1	15.642	18.031	30.17	51.771	164.25
Low	$\mathcal{D}[\tilde{\Gamma}]$	1	14.335	17.681	29.87	51.054	166.115
Low	$\mathcal{B}_b[\tilde{\Gamma}]$	1	15.38	17.544	28.713	50.668	164.03
Low	$\mathcal{T}_\varpi[\tilde{\Gamma}]$	1	4.842	6.988	9.718	18.662	80.013
Medium	$\tilde{\Gamma}$	1	16.311	18.98	32.079	54.886	169.71
Medium	$\mathcal{D}[\tilde{\Gamma}]$	1	15.852	7.965	31.045	52.7	168.595
Medium	$\mathcal{B}_b[\tilde{\Gamma}]$	1	15.9	18.65	30.214	53.951	168.644
Medium	$\mathcal{T}_\varpi[\tilde{\Gamma}]$	1	4.998	7.24	10.03	19.108	85.375
High	$\tilde{\Gamma}$	1	17.56	20.885	35.049	57.441	174.26
High	$\mathcal{D}[\tilde{\Gamma}]$	1	17.611	19.405	33.082	56.3	170.005
High	$\mathcal{B}_b[\tilde{\Gamma}]$	1	16.341	19.773	34.528	57.05	173.42
High	$\mathcal{T}_\varpi[\tilde{\Gamma}]$	1	5.04	7.811	10.575	20.6	89.136

in computer, it is not feasible to implement DCC directly into 100 assets. Therefore, the MSEs of DCC may not perform as well as TARVM estimator.

## 4.4 Example of Real Data

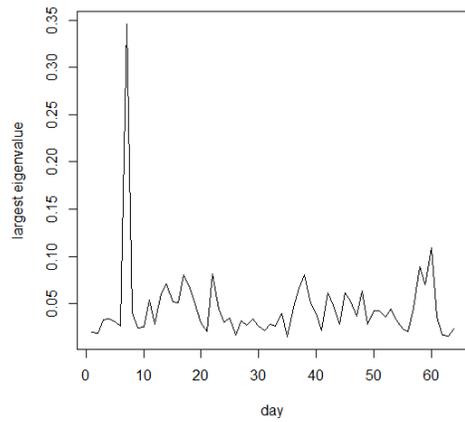
The real data set for our numerical studies is high-frequency tick by tick price data on 100 stocks in top 100 S&P 500 according to trading volume over 3 months, 64 trading days in 2013 from April 1<sup>st</sup> to June 28<sup>th</sup>. This data set is collected from the Wharton Data Service(WRDS) Database. Apply two proposed methods to the real data set for 64 trading days. In each day, we computed two part of estimator: one is the ARVM estimator  $\tilde{\Gamma}$  included its regularization estimator and the other one is DCC volatility matrix estimator  $\mathcal{D}[\tilde{\Gamma}]$ . Similar to the measurement of estimators in simulation, in real data set we can use Mean Squared Prediction Error(MSPE) to evaluate the performances of two methods. But since we don't have the true integrated volatility for real data and the basic model is under a stationarity assumptions on volatility in financial time series, then the realized volatility which is one day ahead of the current one is used to replace  $\Gamma$ .

For ARVM estimator, the pre-determined sampling frequencies were selected to correspond with 5 minute returns. This yielded 64 matrices of size 100 by 100 as ARVM estimators of integrated volatility matrices over the 64 trading days. Denote by  $\tilde{\Gamma}_d, d = 1, \dots, 64$ . The average of these 64 matrices is evaluated. Because stocks have no natural ordering, the decay assumption is not realistic for volatility matrices, then banding may not be appropriate for  $\tilde{\Gamma}$ . Therefore, thresholding is applied to regularize ARVM estimator. Here the MSPE is defined as

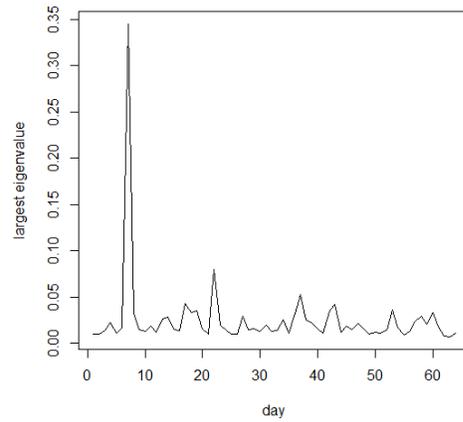
$$MSPE_a(\mathcal{T}_{\varpi_a}[\tilde{\Gamma}]) = \frac{1}{63} \sum_{d=1}^{63} \|\tilde{\Gamma}_{d+1} - \mathcal{T}_{\varpi_{d,a}}[\tilde{\Gamma}_d]\|_2^2, \quad (4.4.1)$$

where  $a \in (0, 1)$ ,  $\varpi_{d,a}$  is the  $a$ -quantile of the absolute entries of  $\tilde{\Gamma}_d$ . We selected the value of  $a$  by minimizing  $MSPE_a(\mathcal{T}_{\varpi_a}[\tilde{\Gamma}_d])$  over  $a \in (0, 1)$ . In calculation,  $a$  equals to 0.95, then for each of  $\tilde{\Gamma}_d$ , we retain its top 5% entries and replace the others by zero. After calculation, the MSPE of  $\mathcal{T}_{\varpi_{d,a}}[\tilde{\Gamma}]$  is 5.89.

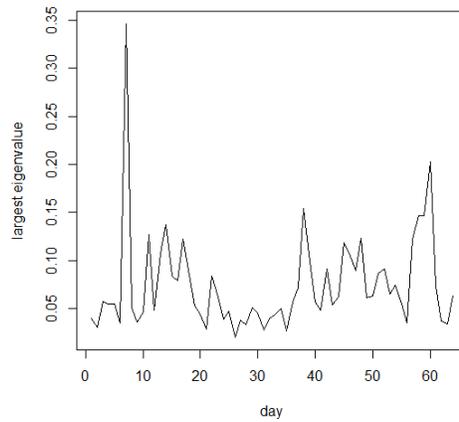
For DCC volatility matrix estimator, first by fitting the real data set into GARCH(1,1) model, we



(a) Largest eigenvalues of daily volatility matrices in DCC



(b) Largest eigenvalues of the thresholding daily realized volatility matrices



(c) Largest eigenvalues of daily realized volatility matrices

Figure 4.3: Plots of the largest eigenvalues of daily volatility matrices in DCC, the thresholded daily realized volatility matrices and realized volatility matrices for the high-frequency data from top 100 S&P 500 stocks.

can find the initial values for stage 1. From paper of Engle and Sheppard(2001), the initial value for stage 2 can be defined as  $\alpha = 0.005$  and  $\beta = 0.959$ , then apply the proposed method by daily data. We can still obtain 64 trading daily volatility matrices of size 100 by 100 in DCC model. The MSPE for  $\mathcal{D}[\tilde{\Gamma}]$  is defined as

$$MSPE(\mathcal{D}[\tilde{\Gamma}]) = \frac{1}{63} \sum_{d=1}^{63} \|\tilde{\Gamma}_{d+1} - \mathcal{D}[\tilde{\Gamma}_d]\|_2^2, \quad (4.4.2)$$

After calculation, the MSPE of  $\mathcal{D}[\tilde{\Gamma}]$  is 20.05. Compare two MSPEs of proposed methods, it is obvious that TARVM estimator performs better than DCC estimator.

For each of  $\tilde{\Gamma}_d$ ,  $\mathcal{T}_{\varpi_d, 0.95}[\tilde{\Gamma}_d]$  and  $\mathcal{D}[\tilde{\Gamma}_d]$ ,  $d = 1, \dots, 64$ , we computed its largest eigenvalues and collected them asset. Figure 4.3 displays the largest eigenvalues of daily volatility matrices in DCC, thresholding daily realized volatility matrices and daily realized volatility matrices. The plots shows that the reductions of the largest eigenvalues due to DCC method are less significant than that due to thresholding for 64 trading days. Daily realized volatility is the most unstable estimation among three estimators. Estimated volatility of DCC method is more volatile than that of TARVM estimator. In eigenbased analysis, the more unstable the largest eigenvalue, the more risk we will end up with a very misleading conclusion. So TARVM estimators have a better performance not only in estimation of large volatility matrix but also in the future eigenbased analysis.

## Chapter 5

# Conclusion and Discussion

This paper presents two methodologies of estimating vast volatility matrix. One is under diffusion process, first, to estimate the ARVM estimator, then to regularize this estimator by banding and thresholding. Banding is more suitable under the decaying condition while thresholding is more appropriate under the sparse condition. The other method is DCC model. The estimation of DCC model also includes two stage, where in the first stage univariate GARCH models are estimated for each residual series, and in the second stage, the residuals are used to estimate the parameters of the dynamic correlation. Both of these methods have asymptotic theory to guarantee the consistency of estimators. In numerical studies, we apply the proposed two methods into simulation high-frequency data and real high-frequency data set, respectively. The real data set is collected from high-frequency tick by tick price data on 100 stocks in top 100 S&P 500 over 64 trading days in 2013 from April 1<sup>st</sup> to June 28<sup>th</sup>. All MSEs and MSPEs of different estimators indicate that TARVM estimator has better performance than DCC estimator.

After we get the volatility matrix for high-frequency data, eigenbased analyses like clustering analysis, principal component analysis, factor analysis and portfolio allocations will be the next job to study the high-frequency data. Here, I will illustrate the effect of volatility matrix on principal component analysis. In computational terms the principal components are found by calculating the eigenvectors and eigenvalues of the data covariance matrix. This process is equivalent to finding the axis system in which the co-variance matrix is diagonal. The eigenvector with the largest eigenvalue

is the direction of greatest variation which may be the first component and will have significant influence on the future analysis, the one with the second largest eigenvalue is the (orthogonal) direction with the next highest variation and so on. Therefore, if the largest eigenvalue is volatile, then we may not get the real first component in PCA, and it will lead to misleading conclusions. For example, we still use our real data set of 100 assets high-frequency trading prices of top 100 S&P 500, and our goal is to find important components that affect the prices during April 1<sup>st</sup> to June 28<sup>th</sup> 64 trading days in 2013. If our estimated volatility matrix has unstable largest eigenvalue, then when we apply the principal component analysis into the estimated volatility matrix, we may get different and uncertain factors and that will lead to unreasonable decisions in practical life.

# Bibliography

- [1] Barndorff-Nielsen O.E., Shephard N. (2002), *Econometric Analysis of Realised Volatility and its Use in Estimating Stochastic Volatility Models*. J.Roy. Statist. Soc. Ser.B, Vol. 64, 253-280.
- [2] Barndorff-Nielsen O.E., Shephard N. (2004), *Econometric Analysis of Realized Covariance: High Frequency base Covariance, Regression and Correlation in Financial Economics*. Econometrica, Vol. 72, 885-925.
- [3] Barndorff-Nielsen O.E., Shephard N. (2006), *Econometrics of Testing for Jumps in Financial Econometrics using Bipower Variation*. Journal of Financial Econometrics, Vol. 4, 1-30.
- [4] Barndorff-Nielsen O.E., Hansen P.R., Lunde A. and Shephard N. (2008a), *Designing Realized Kernels to Measure the Ex-post Variation of Equity Prices in the Presence of Noise*. Econometrica, Vol. 76, 1481-1536.
- [5] Barndorff-Nielsen O.E., Hansen P.R., Lunde A. and Shephard N. (2008b), *Multivariate Realized Kernels: Consistent Positive Semi-definite Estimators of the Covariation of Equity Prices with Noise and Non-synchronous Trading* . preprint.
- [6] Bauwens L., Laurent S. (2006), *Multivariate GARCH Models: a Survey*. Journal of Applied Econometrics, Vol. 21, No. 1, 79-109.
- [7] Bickel P.J. and Levina E. (2008a), *Regularized Estimation of Large Covariance Matrices*. Ann. Statist, Vol. 36, No. 4, 199-227.
- [8] Bickel P.J. and Levina E. (2008b), *Covariance Regularization by Thresholding*. Ann. Statist, Vol. 36, No. 4, 2577-2604.

- [9] Comte F. and Lieberman O. (2000), *Asymptotic Theory for Multivariate GARCH Process*. Journal of Multivariate Analysis, Vol. 84, No. 2003, 61-84.
- [10] El Karoui N. (2008), *Operator Norm Consistent Estimation of Large Dimensional Sparse Covariance Matrices*. Ann. Staist, Vol. 36, No. 4, 2717-2756.
- [11] Engle R.F. and Sheppard K. (2001), *Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH*. NBER Working Paper, No. 8554.
- [12] Engle R. (2002), *Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autogressive Conditional Heteroskedasticity Models*. Journal of Business & Economic Statistics, Vol. 20, No. 3, 339-350.
- [13] Engle R.F and Sokalska M. (2012), *Forecasting Intraday Volatility in the US Equity Market. Multiplicative Component GARCH*. Journal of Financial Econometrics, Vol. 10, No. 1, 54-83.
- [14] Hayashi T. and Yoshida N. (2005), *On Covariance Estimation of Non-synchronously Observed Diffusion Process*. Bernoulli, Vol. 11, 359-379.
- [15] McAleer M., Hoti S. and Chan F. (2009), *Structure and Asymptotic Theory for Multivariate Asymmetric Conditional Volatility*. Econometric Reviews, Vol. 28, No. 5, 422-440.
- [16] Tao M., Wang Y., Yao Q. and Zou J. (2011), *Large Volatility Matrix Inference via Combining Low-Frequency and High-Frequency Approaches*. Journal of the American Statistical Association, Vol. 106, No. 495, 1025-1040.
- [17] Tsui A.K. and Tse Y.K. (1998), *A Multivariate GARCH Model with Time-Varing Correlations*. National University of Singapore.
- [18] Wang Y. and Zou J. (2010), *Vast Volatility Matrix Estimation For High-Frequency Financial Data*. The Annals of Statitics, Vol. 38, 943-978.
- [19] Zou J. and Wang Y. (2013), *Statistical methods for large portfolio risk management*. Statistica and its Interface, Vol. 6, No. 477-485.