Stony Brook University

Modeling Intra-day Markets with an application of Risk Management and Optimal Order Execution

A Dissertation Presented

by

Yikang Chai

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Applied Mathematics and Statistics

Stony Brook University

May 2014

Stony Brook University

The Graduate School

Yikang Chai

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Svetlozar Rachev– Dissertation Advisor
Professor of AMS department

James Glimm- Chairperson of Defense
Professor of AMS department

Aaron Kim –Member of Defense Committee
Assistant Professor of AMS department

Keli Xiao–Member of Defense Committee
Assistant Professor of College of Business

This dissertation is accepted by the Graduate School

Charles Taber
Dean of the Graduate School

Modeling Intra-day Markets with an application of Risk Management and Optimal Order Execution

by

Yikang Chai

Doctor of Philosophy

in

Applied Mathematics and Statistics

Stony Brook University

2014

Financial time series data exhibits heavy tailed, volatility clustering and long range dependence style facts. Traditional Gaussian distribution assumption based model failed to explain these phenomena. A unified framework model proposed in this thesis, fractionally integrated autoregressive moving average (FARIMA) and fractionally integrated generalized autoregressive conditional heteroskedasticity (FIGARCH) with multivariate generalized hyperbolic distribution (MGHD), trying to capture all these phenomena together. We also examined this model by using intra-day market dataset to backtest of various risk measure. With rise of high frequency trading and algorithm trading in recent years, trading volume hugely increased and markets became more volatile. Order execution is the main concern for traders, especially in the case of liquidation of big orders. We illustrate how the optimal order execution strategy behaves under the assumption that market price dynamics follows high volatile (non-Gaussian) markets with volatility clustering and log-range dependence characteristics.

Dedication Page

To my parents and wife

Table of Contents

Chapter 1 Introduction 1

Chapter 2 Long Memory 3

2.1 Basics concepts in time series 3

2.2 Definition of long memory 4

2.3 Long memory and self-similarity 4

2.4 Fractional integration in discrete time 5

Chapter 3 FARIMA-FIGARCH model 10

3.1 FARIMA-FIGARCH structure 10

3.2 Parameter estimation of FARIMA-FIGARCH model 12

3.2.1 Estimation of parameter $\boldsymbol{d}$ 12

3.2.2 Estimation of FARIMA-FIGARCH 13

3.3 Forecasting 15

3.4 Argument of Long Memory 16

Chapter 4 Generalized Hyperbolic Model 18

4.1 Definition and Properties 19

4.2 Parametrization 24

4.3 Parameter Estimation 26

4.3.1 Univariate Standardized GH Estimation 26

4.3.2 Multivariate GH Estimation 27

4.4 Simulation 29

4.5 Practical Issue for High Dimension 29

Chapter 5 Backtesting Procedure and Empirical Results 32

5.1 Model Description 32

5.2 Backtesting Procedure 33

5.3 Empirical Result 36

5.3.1 Dataset 36

5.3.2 In-sample distribution of parameter d 37

5.3.3 Out-sample test 41

5.4 Computing Technology 54

Chapter 6 Optimal Order Execution 56

6.1 Optimal execution problem 56
6.2 The Almgren Chriss execution scheme 57
6.3 Extension of Almgren Chriss model 58
6.4 Calibration of market impact parameter 60
6.5 Numerical example 61
References 68
Appendix 71

List of Figures

Figure 1: MA coefficients for two different processes 8
Figure 2: Autocorrelation coefficients for different processes 9
Figure 3: Comparison two dimensional Gaussian and GH 21
Figure 4: Tail behavior of Gaussian and GH with lambda=-0.2 24
Figure 5: Modified Bessel function with different order 31
Figure 6: Histogram of parameter d for 1 min 38
Figure 7: Histogram of parameter d for 5 min 39
Figure 8: Histogram of parameter d for 10 min 40
Figure 9: Histogram of parameter d for 30 min 40
Figure 10: Histogram of parameter d for 60 min 41
Figure 11: FARIMA-FIGARCH model backtesting 500 periods for hourly data 42
Figure 12: ARMA-GARCH model backtesting 500 periods for hourly data 43
Figure 13: Portfolio of 30 equal weighted assets backtesting 47
Figure 14: Portfolio of 60 equal weighted assets backtesting 48
Figure 15: Portfolio of 120 equal weighted assets backtesting 49
Figure 16: Portfolio of 240 equal weighted assets backtesting 50
Figure 17: Portfolio of 475 equal weighted assets backtesting 51
Figure 18: Bactseting of FARIMA-FIGARCH model for 5 min 52
Figure 19: Backtesting of ARMA-GARCH for 5min 53
Figure 20: Computer parallel architecture 55
Figure 21: 5 min price chart for AAPL and AMZN from 20090729 to 20090831 61
Figure 22: scatter plot of log return between AAPL and AMZN 62
Figure 23: 5 min normalized volume and return 62
Figure 24: Conditional volatility 63
Figure 25: Filtered innovations 64
Figure 26: Efficient frontier of execution cost 65
Figure 27: Trading schedule for AAPL with different λ 67

List of Tables

Table 1: Quantiles of parameter d for different frequency data and window size 37
Table 2: statistical test for FARIMA-FIGARCH at confident level alpha =0.01 42
Table 3: statistical test for ARMA-GARCH at confident level alpha=0.01 43
Table 4: violation number of VaR for different asset number 45
Table 5: statistical test for FIARMA-FIGARCH at confident level alpha=0.01 for different asset number 45
Table 6: statistical test of FARIMA-FIGARCH for 5min 52
Table 7: statistical test of ARMA-GARCH for 5min 53
Table 8: Timing table 54
Table 9: Estimated parameters of FARIMA-FIGARCH model for AAPL and AMZN 63

List of Abbreviations

ARMA	Autoregressive Moving Average
GARCH	Generalized AutoRegressive Conditional Heteroskedasticity
FARIMA	Fractionally Autoregressive Integrated Moving Average
FIGARCH	Fractionally Integrated GARCH
GHD	Generalized Hyperbolic Distribution
LRD	Long Range Dependence

Chapter 1 Introduction

The study of statistical properties of financial or economic time series data shows many stylized facts which seem to be common among several of markets, instruments and periods: heavy tails, volatility clustering, long range dependence, etc.. Although Gaussian distribution assumption is very popular in industry application, people already know that empirical distribution of asset returns exhibits excess kurtosis, namely, non-Gaussian or heavy tailed distribution. Mandelbrot (1960) is the first to attack Gaussian assumption when he used stable distribution to model cotton commodity returns. One would ask why we concerned about heavy tailed distribution. The one major concern is the financial crash or rare event from risk management prospective. From financial market history, we learned the financial crash is not that "rare", at least not as rare as Gaussian based model suggest. One can refer to Rachev and Kim (2010) for a good illustration. There are several popular alternatives for heavy tailed distributions, i.e. student-t distribution, Pareto distribution, classical tempered stable distribution, etc.. Meanwhile, there are some issues when we apply these distributions to real applications, for example, infinite high order moments or non-existence of probability distribution function.

Another important observation of financial return time series is the dependence properties. This means the return series is correlated, which forms the foundation of forecast of time series model. Classical time series model, autoregressive moving average (ARMA), was used to study the financial time series. Based on the specification of ARMA-type model, one expects that autocorrelation or dependence among large lags decays very fast which refers to short-memory. But this is not the case for some cases in empirical observed series which has slower decay rate, namely, hyperbolic decay. Long memory model, fractionally integrated ARMA (FARIMA), was proposed to study this kind of time series (see Granger and Joyeux, 1980, Bollerslev& Jubinski 1999). Although there is debatable existence of long memory in return series data, we argued that FARIMA is a broader class model which contains ARMA-type model as a sub-class model.

Volatility clustering is another well-known observation in financial market. As noted by Mandelbrot [3], "large changes tend to be followed by large changes, of either sign, and small changes tend to be followed small changes". Quantitatively, this phenomenon means the autocorrelation of absolute returns or squared returns display a slower decaying rate. This was modeled by variance of error terms in time series context, Bollerslev (1986) proposed GARCH model to explain volatility clustering phenomenon and Baillie (1996) proposed long memory generalization of GARCH model, namely, fractionally integrated GARCH (FIGARCH). With combination of FARIMA and FIGARCH model, we expect this model has the ability to capture long memory feature in returns series and also volatility series.

In portfolio level analysis, we also concern the dependence structure between each marginal asset. We use multivariate generalized hyperbolic distribution (MGHD) to model dependence structure among innovations for each marginal. The innovations for each marginal asset are filtered out based on FARIMA-FIGARCH structure. Barndorff-Nielsen (1977) first proposed GH distribution which is generated by mean-variance mixture structure. It is a very popular model since the flexibility of its parameters represents a broad class of distribution, and also includes semi-heavy tailed behavior. The GHD is closed under linear transformation. This means that portfolio return, a linear combination of each marginal return with different weight, is still a GHD. This enables us to calculate various risk measure based on one dimensional GH probability density function. We perform backtest based on VaR and CVaR of an equally weighted portfolio by using intra-day dataset with different frequency data for intra-day market.

With rise of high frequency trading and automated algorithm trading in recent years, optimal order execution is the main concern for a trader, especially in the case of liquidation of big orders of financial institutions, such as mutual funds and pension funds. Transaction cost of liquidation of portfolio position is the major concern for a portfolio manager, especially when a big order needs to be executed. The seminal work of Almgren R. and Chriss N. (2000) worked out a framework to define an optimal trading strategy in terms of trading cost, a variable defined to capture the characteristic of a trading strategy. A very interesting thing is to see that how an optimal trading strategy behaves during high volatile (non-Gaussian) markets with volatility clustering and log-range dependence. In this thesis, we provide a simple demonstration to illustrate this scenario in terms of efficient frontier of transaction cost.

This thesis organizes as following. Chapter 2 introduces concept and some properties of long memory. We present FARIMA-FIGARCH model in chapter 3. Generalized hyperbolic distribution is detailed in chapter 4. Backtesting procedure and empirical results are presented in chapter 5. We review optimal execution framework and illustration of a simple numerical study is demonstrated in chapter 6.

Chapter 2 Long Memory

2.1 Basics concepts in time series

A stochastic process X_t is a collection of random variables with time ordering index. X_t is discrete when $t \in N$ and continuous when $t \in R$. In practice, we usually observe a set of realization of X_t, say $(x_1, x_2, \dots x_N)$.

If the joint distribution associate with m observations $x_{t_1}, x_{t_2}, \dots, x_{t_m}$, is the same as observations $x_{t_1+k}, x_{t_2+k}, \dots, x_{t_m+k}$, then the stochastic process X_t is strictly stationary.

Under strictly stationary assumption, the covariance between x_t and x_{t-k} must be same for any t. It's only determined by k, lags between them. The autocovariance at lag k of X_t is defined as

$$\gamma_k = cov(x_t, x_{t-k}) = E[(x_t - \mu)(x_{t-k} - \mu)]$$

$$\mu = E[x_k]$$

And autocorrelation at lag k is

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

The spectral density is defined as

$$f(\omega) = \sum_{k=-\infty}^{\infty} \gamma_k \, e^{-2\pi i \omega k} \quad for \ -\infty < \omega < \infty$$

One can see that $f(\omega)$ is the discrete version of Fourier transform of γ_k. This is why spectral density is considered as frequency domain analysis. We also can show that

$$\gamma_k = \int_{-\infty}^{\infty} e^{2\pi i \omega k} \, \mathrm{f}(\omega) \mathrm{d}\omega$$

This means spectral density is related to autocovariance by Fourier transform.

2.2 Definition of long memory

Long memory is also called long range dependence. There are several related definitions of long memory in the literature. In this thesis, we take the definition below (see McLeod & Hipel (1978)) :

Definition 2.1. *Assume X_t is a discrete time process with autocorrelation coefficient ρ_j at lag j , then we say that the process contains long memory if*

$$\lim_{n\to\infty} \sum_{i=-n}^{n} |\rho_i| = \infty \tag{2.1}$$

By this definition, the absolute autocorrelations are non-summable. While short memory process could be defined as

$$\lim_{n\to\infty} \sum_{i=-n}^{n} |\rho_i| = k \tag{2.2}$$

And the autocorrelation of long memory process has representation (see Brockwell & Davis (1995)), we also show the detail later.

$$\rho_i \sim C|i|^{2d-1} \qquad i \to \infty, \quad 0 < d < \frac{1}{2} \tag{2.3}$$

By definition, the spectral density of X_t could be calculated as

$$f_X(\omega) = \frac{1}{2\,\pi} + \frac{1}{\pi} \sum_{k=1}^{\infty} \rho_k \cos(k\omega) \tag{2.4}$$

It is clear that spectral density of X_t is infinite at $\omega = 0$. This also means that the spectral density has a pole at $\omega = 0$.

2.3 Long memory and self-similarity

There is close relationship between long memory and self-similarity. In this section, we will explore this relationship. We also clarify that the concept of self-similarity defined in continuous time. Meanwhile, The FARIMA-FIGARCH model is about discrete time process.

Definition 2.2. *A stochastic process $(X_t)_{t\geq 0}$ is said to be self-similar if there exists $H > 0$ such that for any scaling factor $c > 0$, the processes $(X_{ct})_{t\geq 0}$ and $(C^H X_t)_{t\geq 0}$ have the same law*:

$$(X_{ct})_{t\geq 0} \overset{d}{=} (C^H X_t)_{t\geq 0} \tag{2.5}$$

Where H is called the self-similarity exponent of the process X.

To demonstrate the relationship between long memory and self-similarity, we define a continuous stochastic process Y_t, with self-similarity parameter H. Assume $Y_0 = 0$ with probability 1, $E[Y_t] = 0$ and the increments of Y_t are stationary, i.e. $X_t = Y_t - Y_{t-1}$ is stationary with $Var(X_t) = \sigma^2$. Then we have:

$$Cov(X_t,\ X_{t+k}) = Cov(X_1,\ X_{1+k})$$

$$= \frac{1}{2}[E((Y_{k+1} - Y_0)^2) + E((Y_{k-1} - Y_0)^2) - 2\,E((Y_k - Y_0)^2)]$$

Then by property of self-similarity, the correlation are given by

$$Corr(X_t, X_{t+k}) = \rho(k) = \frac{1}{2}[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\ ]$$

$$= \frac{1}{2}k^{2H}[\left(1 + \frac{1}{k}\right)^{2H} - 2 + \left(1 - \frac{1}{k}\right)^{2H}]$$

By Taylor expansion, we have

$$\rho(k) \sim H(2H-1)k^{2H-2} \qquad k \to \infty$$

Therefore, $\lim_{n\to\infty} \sum_{k=-n}^{k-n} \rho(k)$ exist if and only if $H < \frac{1}{2}$, and the case of $H > \frac{1}{2}$ implies that the series does not converge, we have X_t is a long memory process. And also from formula(2.3), we see that there is a relationship between parameters H and d :

$$H = d + \frac{1}{2} \tag{2.6}$$

Usually, value of 0.5 of H indicates the absence of long rang dependence. The degree of long range dependence getting stronger when H approach 1; and H less than 0.5 indicates anti-persistency, which means the process fluctuates violently. Therefore, the corresponding value of d suggests the same feature of a random process.

2.4 Fractional integration in discrete time

Following Hosking (1981), the $(\frac{1}{2} - H)$th derivative of a Brownian motion is called a fractional Brownian motion in continuous time. The discrete time analog of a Brownian motion is the random walk with Gaussian innovations, and a fractional Brownian motion is fractionally differenced Gaussian noise. If we define the discrete time version of Brownian motion as

$$\Delta x_t = (1-L)^1 x_t = \epsilon_t$$

Where L is the lag-operator, and ϵ_t is Gaussian $i.i.d(0, \sigma_\epsilon^2)$ random variable. Then we have the corresponding definition of fractional difference based on binomial expansion, see Granger and Joyeux (1980) :

$$\Delta^d = (1-L)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-L)^k \tag{2.7}$$

The convergence of this expansion follows theorem from Lindstrøm (1995). We also want to point out that binominal expansion (2.7) is nothing else but the Taylor series expansion of $f(z) = (1-z)^d$ at $z = 0$, one could possible choose the same Taylor series expansion at different point of z, which could get a different stochastic variable follows different probability law with the one in binominal expansion. This is still an open question to see how the different expansion would affect the structure of fraction integration.

Now, we can write the fractionally integrated process as following:

$$(1-L)^d x_t = x_t - dLx_t + \frac{1}{2} d(d-1)L^2 x_t + \cdots = \epsilon_t$$

$$x_t = d\, x_{t-1} - \frac{1}{2} d(d-1)x_{t-2} + \cdots + \epsilon_t = \sum_{j=1}^{\infty} \pi_j x_{t-j} \tag{2.8}$$

where

$$\pi_k = \frac{d\,(1-d)(2-d)\ldots(k-1-d)}{k!} = \frac{d\Gamma(k-d)}{\Gamma(1-d)\Gamma(k+1)}$$

And $\Gamma(\,)$ denotes gamma function. By using Stirling's formula, one can show that

$$\pi_k \sim -\frac{k^{-d-1}}{\Gamma(-d)} \qquad as\ \ k \to \infty$$

Therefore the $AR(\infty)$ coefficients of the integrated process x_t is π_k. If we consider the condition of invertibility with the existence of $AR(\infty)$ representation, we can show that $d > -1/2$ is a sufficient condition, see Hosking (1981). In order to check MA representation of x_t, we can rewrite equation (2.8) as :

$$x_t = (1-L)^{-d}\epsilon_t = \left[1 + d\,L + \frac{1}{2}\, d(d+1)L^2 + \cdots\right]\epsilon_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} \tag{2.9}$$

where

$$\psi_0 = 1$$

$$\psi_k = \frac{d(d+1)(d+2)\ldots(d+k-1)}{k!} = \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)} \quad , k = 1, 2, 3, \ldots.$$

Therefore ψ_k is the $MA(\infty)$ coefficients of x_t. Follow Stirling's formula, one can show that

$$\psi_k \sim \frac{k^{d-1}}{\Gamma(d)} \quad as \quad k \to \infty$$

And the variance of x_t is

$$Var(x_t) = \sigma^2 \sum_{k=0}^{\infty} \psi_k^2 = \sigma^2 \sum_{k=0}^{M} \psi_k^2 + \sigma^2 \sum_{k=M+1}^{\infty} \psi_k^2$$

$$\cong \sigma^2 \sum_{k=0}^{M} \psi_k^2 + \sigma^2 \sum_{k=M+1}^{\infty} \frac{k^{2d-2}}{\Gamma^2(d)} \quad \text{for large } M$$

From this expression, we can see that the variance x_t is finite if and only if

$$\sum_{\boldsymbol{k=M+1}}^{\infty} \frac{\boldsymbol{k^{2d-2}}}{\boldsymbol{\Gamma(d)}} < \infty$$

This means $2d - 2 < -1$, that is $d < 1/2$.

In order to get a good understanding of the difference between long memory fractionally integrated process and short memory process, we present an example of a $AR(1)$ process with coefficient 0.5 and a fractionally integrated process with $d = 0.4$. In figure 1, we plot their $MA(\infty)$ coefficient for illustration. As a process can be interpreted as the convolution of MA coefficients and independent random innovation (shocks). Actually, in signal process context, these MA coefficients is called impulse response weights which could be think of how much effects for different shocks. From the figure, we can see that the MA coefficients of fractionally integrated process decays much slower than the AR(1) process. This illustrates exactly how long memory process behaves when it hit by a shock. This is also why we call it "long memory", since the shocks has longer time effect.

We can also get the autocorrelation coefficient by definition above:

$$Cov(x_t, x_{t-k}) = \gamma_k = \frac{(-1)^k \Gamma(1-2d)}{\Gamma(1-k-d)\Gamma(1+k-d)}$$

And

$$\gamma_0 = \frac{\Gamma(1-2d)}{\Gamma(1-d)^2} \quad for \ -\frac{1}{2} < d < \frac{1}{2}$$

The autocorrelation coefficient:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{(-1)^k \Gamma(1-d)^2}{\Gamma(1-k-d)\Gamma(1+k-d)} \tag{2.10}$$

And it can be show by Stirling's formula:

$$\rho_k \sim \frac{\Gamma(1-d)}{\Gamma(d)} k^{2d-1} \qquad as \ k \to \infty \tag{2.11}$$

We can see that this formula corresponds to the formula (2.3) in section 2.1. Let's compare this to $AR(1)$ process, say,

$$x_t = ax_{t-1} + \epsilon_t$$

$$cov(x_t, x_{t-k}) = \gamma_k = \frac{a^k}{1-a^2}$$

$$Var(x_t) = \gamma_0 = \frac{1}{1-a^2}$$

$$\rho_k = \frac{\gamma_k}{\gamma_0} = a^k$$

We can see that autocorrelation of $AR(1)$ process decay geometrically. In figure 2, we plot the autocorrelation coefficient for the same process in figure 1. We can clear see that fractionally integrated process has slowly decay rate than $AR(1)$ process.

From all the discussion above, we see that the autoregressive coefficients, the moving average coefficients and the autocorrelation coefficients of the fractionally integrated process all decay as $k^{-d}, d > 0\ , k \to \infty$. This is hyperbolic decay rate which is slower than the exponentially decay rate of short memory process.

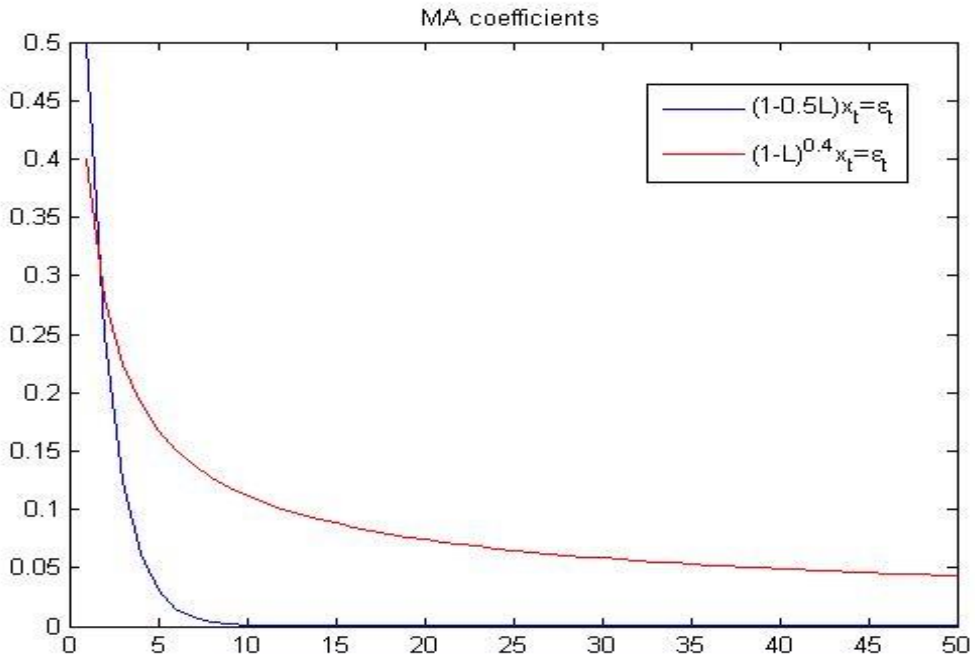


Figure 1: MA coefficients for two different processes

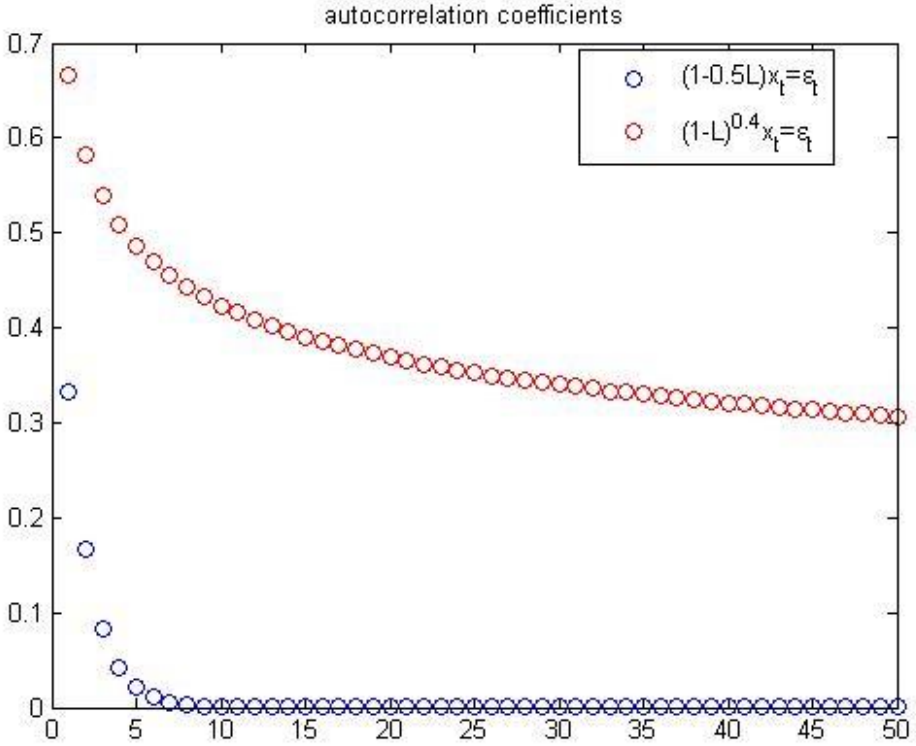


Figure 2: Autocorrelation coefficients for different processes

Chapter 3 FARIMA-FIGARCH model

3.1 FARIMA-FIGARCH structure

From the discussion in chapter 2, we see that there is only one parameter d in the definition of fractionally integrated process x_t , see formula (2.8). And the slow decay or long memory property all depend on this particular parameter. In order to generalize the fractionally integrated process, people add the feature of modeling the short run dynamics, say, ARMA-type model. With this motivation, Granger & Joyeux (1980) and Hosking (1981) propose the FARIMA structure.

A stochastic process y_t is said to be fractionally autoregressive integrated moving average (FARIMA) or denoted as FARIMA(a, d_0, m) if it is defined as

$$\phi(L)(1-L)^{d_0}(y_t-\mu)=\theta(L)\epsilon_t \tag{3.1}$$

$$\phi(L)=1-\sum_{j=1}^{a}\phi_j L^j$$

$$\theta(L)=1+\sum_{j=1}^{m}\theta_j L^j$$

$$\epsilon_t \sim i.i.d(0,\sigma^2)$$

Where μ is the unconditional mean of process y_t , $\phi(L)$ is the autoregressive operator and $\theta(L)$ is the moving average operator. The usual stationary and invertibility conditions of ϕ_j and θ_j for ARMA model are assumed and $d_0 \in (-0.5, 0.5)$ to ensure stationary and invertibility of fractionally integrated part.

To model the volatility clustering phenomena which is an important empirical observation in the financial market, Bollerslev (1986) proposed the ARMA-type model on the conditional volatility series of ϵ_t , GARCH model. If we assume $h_t = \mathrm{Var}(\epsilon_t | \Omega_{t-1})$, then GARCH(p, q) is defined as

$$h_t=\omega+\alpha(L)\epsilon_t^2+\beta(L)h_t \tag{3.2}$$

$$\alpha(L)=\sum_{j=1}^{q}\alpha_j L^j$$

$$\beta(L)=\sum_{j=1}^{p}\beta_j L^j$$

Where $\omega=\sigma^2(1-\alpha(1)-\beta(1))$ and $\sigma^2=Var(\epsilon_t)$.

Another characteristic of volatility series is the persistence, see Ding, Granger, and Engle (1993), and Harvey (1993). With same idea of FARIMA structure, one can use fractionally integrated process with GARCH to model persistence, which refers to fractionally integrated GARCH model (FIGARCH) see Baillie, Bollerslev and Mikkelsen (1996). There is different specification of FIGARCH model, see Chuang (1999). If we rewrite GARCH(p,q) in (3.2) as:

$$\big(1-\alpha(L)-\beta(L)\big)\epsilon_t^2=\omega+\big(1-\beta(L)\big)V_t \tag{3.3}$$

Where $V_t=\epsilon_t^2-h_t$ are innovations since they have zero mean and uncorrelated. To ensure process ϵ_t^2 is stationary, we assume the roots of polynomials of $1-\alpha(L)-\beta(L)$ and $1-\beta(L)$ lie outside the unit circle. Baillie, Bollerslev and Mikkelsen proposed FIGARCH(p,d,q) model based on (3.3):

$$\psi(L)(1-L)^d\epsilon_t^2=\omega+\big(1-\beta(L)\big)V_t \tag{3.4}$$

where

$$\psi(L)=1-\textstyle\sum_{j=1}^{q}\psi_j L^j$$

And assume $d\in(0,1)$ to ensure the stationary. To get explicit expression of conditional variance h_t ,we can rewrite (3.4) as :

$$\big(1-\beta(L)\big)h_t=\omega+\big(1-\beta(L)\big)\epsilon_t^2-\psi(L)(1-L)^d\epsilon_t^2 \tag{3.5}$$

But there is one problem of specification in (3.4), since it does not have similar structure to FARIMA model, in terms of $(1-L)^d$ operator. Because operator $(1-L)^{d_0}$ applies to μ in (3.1) , but $(1-L)^d$ not applies to ω in (3.4). Therefore, we can rewrite GARCH(p,q) in another way to avoid this problem:

$$\big(1-\alpha(L)-\beta(L)\big)(\epsilon_t^2-\sigma^2)\ =\big(1-\beta(L)\big)V_t \tag{3.6}$$

And the corresponding FIGARCH(p,d,q) defined as:

$$\psi(L)(1-L)^d(\epsilon_t^2-\sigma^2)=\big(1-\beta(L)\big)V_t \tag{3.7}$$

And the corresponding conditional variance h_t can be expressed as:

$$\big(1-\beta(L)\big)h_t=\big(1-\beta(L)\big)\epsilon_t^2-\psi(L)(1-L)^d(\epsilon_t^2-\sigma^2) \tag{3.8}$$

Compare specification of (3.4) and (3.7), we have

$$\omega=\psi(L)(1-L)^d\sigma^2 \tag{3.9}$$

Chung (1999) argued that the specification (3.7) is better than (3.4) in terms of parameter estimation. Following their track, we use specification (3.7) in this thesis.

To sum up, we combine structure (3.1) and (3.7) as our FARIMA(a, d_0, m)-FIGARCH(p, d, q) model:

$$\phi(L)(1-L)^{d_0}(y_t-\mu)=\theta(L)\epsilon_t \tag{3.10}$$

$$\psi(L)(1-L)^{d}(\epsilon_t^2-\sigma^2)=\big(1-\beta(L)\big)V_t$$

$$\epsilon_t \sim i.i.d(0,\sigma^2)$$

where

$$\phi(L)=1-\textstyle\sum_{j=1}^{a}\phi_j L^j$$

$$\theta(L)=1+\textstyle\sum_{j=1}^{m}\theta_j L^j$$

$$\psi(L)=1-\textstyle\sum_{j=1}^{q}\psi_j L^j$$

$$\beta(L)=\textstyle\sum_{j=1}^{p}\beta_j L^j$$

$$V_t=\epsilon_t^2-h_t \text{ and } \mathrm{h_t}=\mathrm{Var}(\epsilon_\mathrm{t}\mid\Omega_{\mathrm{t}-1}) \quad .$$

3.2 Parameter estimation of FARIMA-FIGARCH model

3.2.1 Estimation of parameter d

There are some simple tools to detect long memory, such as, R/S plot, variance plot and variogram, etc.. But they all just for diagnostic purpose. In this thesis, we employ semi-parametric method for estimation of d, GPH log-periodogram estimator, see Geweke and Porter-Hudak (1983).

We can rewrite FARIMA process in (3.1) as :

$$(1-L)^d y_t = w_t \tag{3.11}$$

$$\phi(L)w_t=\theta(L)\epsilon_t$$

w_t is an ARMA process, and from equation (3.11), we can write

$$h_y(\omega) = \left|1 - e^{-i\omega}\right|^{-2d} h_w(\omega) \tag{3.12}$$

where $h_y(\omega)$ is the spectrum of y . Take logarithm of (3.12) , we have

$$\ln\left(h_y(\omega)\right) = -2d \ln\left(\left|2\, sin\left(\frac{\omega}{2}\right)\right|\right) + \ln\left(\frac{h_w(\omega)}{h_w(0)}\right) + \ln(h_w(0))$$

Adding the periodogram of y to both sides

$$\ln\left(I_y(\omega)\right) = \ln\left(h_w(0)\right) - 2d \ln\left(\left|2 \sin\left(\frac{\omega}{2}\right)\right|\right) + \ln\left(\frac{h_w(\omega)}{h_w(0)}\right) + \ln\left(\frac{I_y(\omega)}{h_y(0)}\right) \tag{3.13}$$

Then we can estimate d by linear regression base on equation (3.13).

If we have data sequence $y_0, y_1, \dots, y_{n-1}$, then periodogram is calculated by:

$$I_j = \frac{1}{2\pi n}\left|\sum_{t=0}^{n-1} y_t e^{i\,\omega_j\, t}\right|^2 \ , \ j = 1, \dots, m \ ; \ where \ \ \omega_j = \frac{2\pi j}{n}$$

And define $X_j = \log\left|2 \sin\left(\frac{\omega_j}{2}\right)\right|$, then estimator of d could be written as

$$\hat{d} = \frac{-0.5 \sum_{j=1}^{m}\left(X_j - \bar{X}\right)\log I_j}{\sum_{j=1}^{m}\left(X_j - \bar{X}\right)^2} \qquad \text{where } \ \bar{X} = \frac{1}{m}\sum_{k=1}^{m} X_k \tag{3.14}$$

As Clifford M. Hurvich (1998) suggested, we choose $m = n^{1/2}$ which makes this estimator has minimized mean squared error.

3.2.2 Estimation of FARIMA-FIGARCH

We follow [Chung 1999] to use approximate maximum likelihood estimation (AMLE) for our parameters estimation procedure. Given the sample $y_1, y_2, \dots, y_T$ follow FARIMA-FIGARCH model, if we further assume that the innovation ϵ_t are normal distributed, $\epsilon_t \sim N(0, h_t)$, then the approximate log likelihood function is :

$$\ln L(\eta) = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T} \ln h_t - \frac{1}{2}\sum_{t=1}^{T} \frac{\epsilon_t^2}{h_t} \tag{3.15}$$

where η is parameter set needs to be estimeated.

The computation of this log-likelihood function require the values of ϵ_t and h_t . Follow FARMA-FIGARCH structure defined in (3.10), we first filter ϵ_t through two steps:

Step 1: $\xi_t = \sum_{i=0}^{t-1} \pi_i(d_0)\,(y_{t-i} - \mu)$ where $\pi_k = \frac{d\Gamma(k-d)}{\Gamma(1-d)\Gamma(k+1)}$

Step 2: $\epsilon_t = \xi_t - \sum_{i=1}^{a} \phi_i \xi_{t-i} - \sum_{i=1}^{m} \theta_i \epsilon_{t-i}$ (3.16)

Once obtain ϵ_t ,we can compute conditional variance h_t .

Step 3: $\zeta_t = \sum_{i=0}^{t-1} \pi_i(d)(\epsilon_t^2 - \sigma^2)$

Step 4: $h_t = \sum_{i=1}^{p} \beta_i h_{t-i} - \sum_{i=1}^{p} \beta_i \epsilon_{t-i}^2 + \epsilon_t^2 - \zeta_t + \sum_{i=1}^{q} \psi_i \zeta_{t-i}$ (3.17)

For FARIMA$(1, d_0, 1)$-FIGARCH$(1, d, 1)$, we have parameter set $= (\phi, d_0, \theta, \psi, d, \beta)$, six parameters. In order to get optimal local maximum for AMLE optimization procedure, we need to get optimal initial estimates of parameter set. We can rewrite FARIMA-FIGARCH in (3.10) as :

$$(1-L)^{d_0}(y_t - \mu) = W_t$$

$$\phi(L)W_t = \theta(L)\epsilon_t \quad (3.18)$$

And

$$(1-L)^{d}(\epsilon_t^2 - \sigma^2) = S_t$$

$$\psi(L)S_t = \big(1 - \beta(L)\big)V_t \quad (3.19)$$

Where W_t and S_t are ARMA process.

As discussion of semi-parametric estimator of d in last section, we can take advantage of this and estimate an ARMA process on the filtered data. We propose procedure as following:

Step 1: Estimate d_0 based on sequence of y_t .

Step 2: Filter W_t from $y_t - \mu$; i.e. $W_t = (1-L)^{d_0}(y_t - \mu)$.

Step 3: Fit W_t to the ARMA model, and get initial estimates of (ϕ, θ) .

Step 4: Filter ε_t from the FARIMA model based on the parameters from steps 1, 2, 3.

Step 5: Estimate d based on $\varepsilon_t^2 - \sigma^2$.

Step 6: Filter h_t as a function of (ψ, β) based on the formula in (3.17).

Step 7: Fit ε_t and h_t to the AMLE log-likelihood function given by equation (3.15).

3.3 Forecasting

After calibrate model, we need to have one step forward-looking prediction or generate scenarios. Let's review ARMA(a,m)-GARCH(p,q) model first:

$$y_t = \sum_{j=1}^{a} \phi_j \, y_{t-j} + \sum_{j=1}^{m} \theta_j \epsilon_{t-j} + \epsilon_t \tag{3.20}$$

$$h_t = \omega + \sum_{j=1}^{q} \alpha_j \epsilon_{t-j}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}$$

$$\epsilon_t = \sqrt{h_t}\, u_t$$

Suppose we are at time t now, i.e. information up to time t. The next step model would be:

$$y_{t+1} = \sum_{j=1}^{a} \phi_j \, y_{t+1-j} + \sum_{j=1}^{m} \theta_j \epsilon_{t+1-j} + \epsilon_{t+1} = E[y_{t+1}|\Omega_t] + \epsilon_{t+1} \tag{3.21}$$

$$h_{t+1} = \omega + \sum_{j=1}^{q} \alpha_j \epsilon_{t+1-j}^2 + \sum_{j=1}^{p} \beta_j h_{t+1-j}$$

$$\epsilon_{t+1} = \sqrt{h_{t+1}}\, u_{t+1}$$

Where $E[y_{t+1}|\Omega_t]$ and h_{t+1} are deterministic up to time t. Therefore the randomness all come from u_{t+1} which we will model separately with heavy-tailed distribution. Also, we can obtain scenarios based on the scenarios of u_{t+1}.

Now let's go to FARIMA(a,d_0,m)-FIGARCH(p,d,q) model. The difference between them is the fractional integrated part $(1-L)^d$ which could expand as an infinite sum:

$$(1-L)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} L^j = 1 + \sum_{j=1}^{\infty} \pi_j(d) L^j \tag{3.22}$$

Therefore,

$$\phi(L)(1-L)^d = \left(1 - \sum_{j=1}^{a} \phi_j L^j\right)\left(1 + \sum_{j=1}^{\infty} \pi_j(d) L^j\right) = 1 - \sum_{j=1}^{\infty} C_j L^j \tag{3.23}$$

Where C_j could be calculate by convolution of polynomials of ϕ_j and π_j . In real application, we need to truncate the infinite sum in (3.23) to some large lag term, say, $K \gg \max(a,m)$. Then we can rewrite FARIMA as:

$$\phi(L)(1-L)^{d_0} y_t = \theta(L)\epsilon_t$$

$$\Rightarrow y_t = \sum_{j=1}^{K} C_j y_{t-j} + \sum_{j=1}^{m} \theta_j \epsilon_{t-j} + \epsilon_t \tag{3.24}$$

This is exactly the ARMA model with different order, compare to equation (3.20). In the same manner, we can rewrite FIGARCH equation:

$$\psi(L)(1-L)^d(\epsilon_t^2-\sigma^2)=\big(1-\beta(L)\big)V_t$$

$$\Rightarrow \big(1-\beta(L)\big)h_t=-\beta(L)\epsilon_t^2+\epsilon_t^2-\psi(L)(1-L)^d(\epsilon_t^2-\sigma^2)$$

$$\Rightarrow h_t=\omega+\sum_{j=1}^{p}\beta_j(h_{t-j}-\epsilon_{t-j}^2)+\epsilon_t^2-\psi(L)(1-L)^d\epsilon_t^2$$

$$\Rightarrow h_t=\omega+\sum_{j=1}^{p}\beta_j(h_{t-j}-\epsilon_{t-j}^2)+\epsilon_t^2-(\epsilon_t^2-\sum_{j=1}^{K}C_j'\,\epsilon_{t-j}^2)$$

$$\Rightarrow h_t=\omega+\sum_{j=1}^{p}\beta_j(h_{t-j}-\epsilon_{t-j}^2)+\sum_{j=1}^{K}C_j'\,\epsilon_{t-j}^2 \qquad (3.25)$$

Where C_j' is calculated from $\psi(L)(1-L)^d$. Combine (3.24) and (3.25), we obtain the explicit equation for FARIMA-FIGARCH and next step equation:

$$y_{t+1}=\sum_{j=1}^{K}C_j y_{t+1-j}+\sum_{j=1}^{m}\theta_j\epsilon_{t+1-j}+\epsilon_{t+1}=E[y_{t+1}|\Omega_t]+\epsilon_{t+1} \qquad (3.26)$$

$$h_{t+1}=\omega+\sum_{j=1}^{p}\beta_j(h_{t+1-j}-\epsilon_{t+1-j}^2)+\sum_{j=1}^{K}C_j'\,\epsilon_{t+1-j}^2$$

$$\epsilon_{t+1}=\sqrt{h_{t+1}}\,u_{t+1}$$

Similarly, $E[y_{t+1}|\Omega_t]$ and h_{t+1} are deterministic up to time t, and all the randomness come from u_{t+1}. Scenario generation for y_{t+1} will be dependent on scenarios of u_{t+1}.

3.4 Argument of Long Memory

There are a lot of arguments about the existence of long memory or long range dependence of stock returns in the literature. Mandelbrot (1971) firstly suggested that stock returns exhibits long range dependence and subsequently many empirical studies in R/S analysis confirm that. But such test have been criticized by Lo (1991), he pointed out one may yield a different result after accounting for short range dependence. Although the absence of solid evidence of long range dependence in returns, it seems that the presence of long range dependence of volatility is more accepted. Therefore, many authors suggested that models such as Fractionally Integrated GARCH model could capture the long range dependence of volatility. This is also why we are using FARIMA-FIGARCH model to capture long range dependence of returns and volatility together in this thesis, as far as we know, this is the only statistical model could be used systematically.

We show that long memory process would have slow decay of autocorrelation in chapter 2. But testing for long range dependence by examining sample autocorrelation may fail to work when the asset returns have features of heavy tails and long range dependence, see J.Bearan (1994).

And slow decay of sample autocorrelation functions may possibly arise from structural change, this could lead to spurious long memory. There is no reliable technique to differentiate structural change and long memory in the literature now. But we do know there is similar effect between fractional integration and structural change. Diebold and Inoue (1999) show that stochastic regime switching and related model may be confused with long memory, as long as only a small amount of regime switch occurs. Transformations of random walks may also produce processes very similar to long memory processes. Granger and Ding (1996) provide an example of a process of nonlinear transformation of random walk,which has slowly decaying autocorrelations.

There are some argument that long memory model has parsimonious representation of data, since there is only parameter d to describe long memory. If one look at the transfer function of FARIMA process, one could hope high order ARMA model would approximate the FARIMA arbitrarily close in terms of rationality of transfer function. But this require a lot of parameters to be estimated in ARMA model, this could be done by using state space model for high order parameters, and also need some technique to reduce dimension of parameters in order to ensure prediction power of the model, for example, lattice or optimal balance technique in signal processing field.

Chapter 4 Generalized Hyperbolic Model

It is well known that the returns of most financial assets have heavy tails or semi-heavy tails. In risk management context, one is more concerned about the loss or negative tail part of the portfolio. And the most widely used risk measure is Value at Risk (VaR) or Conditional VaR (CVaR) which all builded up the distribution of underling financial instrument. The traditional Gaussian assumption would greatly underestimate these risk measure which would lead to great risk to large financial institution or investors. Generalized hyperbolic distributions (GHD) possess semi-heavy tails. Compared to other heavy tails distribution, such as stable distributions, GHD have finite moments which could be applied to pricing theory. And the probability distribution function of GHD is known explicitly which would be easily applied in practice compare to other tempered stable distributions.

Old E. Barndorff-Nielsen (1977) introduced this distribution first and applied it to model grain size distribution of windblown sands. GHD turns out to be a very broad class distribution, and contains many special cases, such as, hyperbolic distribution, normal inverse Gaussian, Student-t, Variance-gamma, etc.. The normal mean-variance mixture structure makes it have more flexibility to model multivariate distributions.

In this thesis, we are trying to model the standardized innovation in (3.10) as generalized hyperbolic distribution. The standardized innovation is defined as

$$u_t = \frac{\epsilon_t}{\sqrt{h_t}}$$

The FARIMA-FIGACH structure will capture the correlation and long range dependence effect between time lags. And modeling u_t will enables us to capture the external effects, such as heavy tails. In portfolio level, we also need to model the dependence structure between different assets of portfolio. Assume we have N instruments in our portfolio, and then we can get vector of standardized innovation $(u_1, u_2, \dots u_N)$ which can be filtered out for each asset based on FARIMA-FIGARCH structure. Under elliptical symmetry distribution assumption, such as multivariate Gaussian distribution, it implies that all one dimensional distributions are symmetric, which contradicts the empirical observation. Normal mean-variance mixture type distribution could enable us to add asymmetry to different dimension and model non-linear dependence, i.e. tail dependence in multivariate case. In this thesis, we suggest multivariate generalized hyperbolic distribution (MGHD) would be a good candidate for modeling dependence structure.

4.1 Definition and Properties

Definition 4.1 X is said to have a multivariate GH distribution if

$$\boldsymbol{X} \overset{d}{=} \mu + W\gamma + \sqrt{W}A\boldsymbol{Z} \tag{4.1}$$

Where

1. $\mathrm{Z} \sim N_{\boldsymbol{k}}(0, \mathrm{I_k})$
2. $A \in \mathbb{R}^{d\times k}$
3. $\mu, \gamma \in \mathbb{R}^{d}$
4. $W \geq 0$ is a scalar valued random variable and has GIG (Generalized Inverse Gaussian distribution)

There are different parametrizations for GH distribution, we will review this in next section. For the definition above, we say $X \sim GH(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$ which has six parameters. (λ, χ, ψ) are the parameters for GIG distribution, and determine the shape of the GH distribution. From definition above, we can see that it determines how much weight is assigned to the tails and to the center. In general, the larger those parameters the closer the GH distribution is to the normal distribution. μ is the location parameter. $\Sigma = AA'$ is the dispersion matrix. γ is the skewness parameter.

We can see that conditional distribution of $X|W = w$ is normal distribution:

$$X|W = w \sim N_d(\mu + w\,\gamma, w\,\Sigma) \tag{4.2}$$

Where $\Sigma = AA'$.

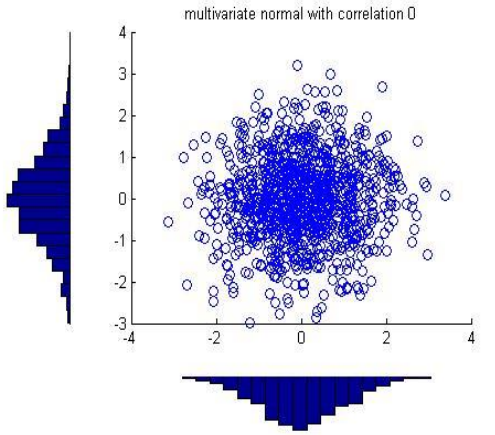
Also, we have expected value and variance as below:

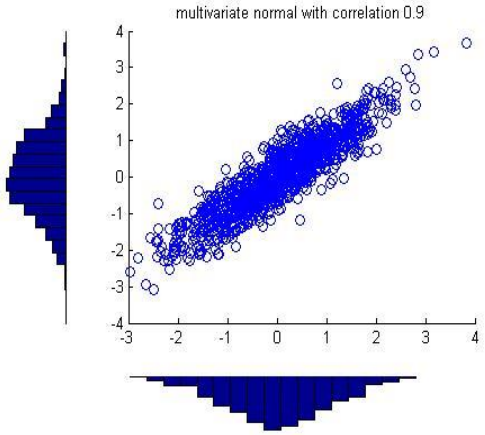
$$E[X] = E\big[E[X|W]\big] = \mu + E[W]\gamma \tag{4.3}$$

$$var[X] = E\big[cov[X|W]\big] + cov[E[X|W]]$$

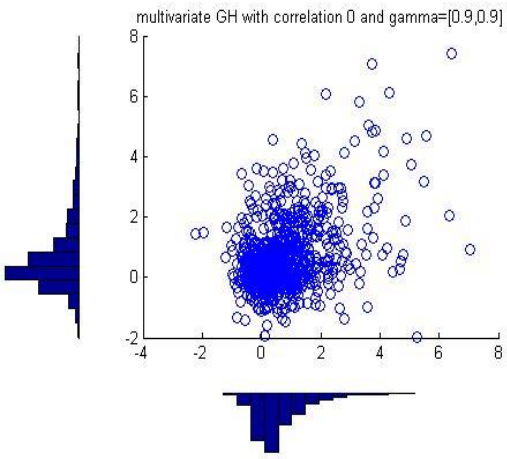
$$= var(W)\gamma\gamma' + E[W]\,\Sigma$$

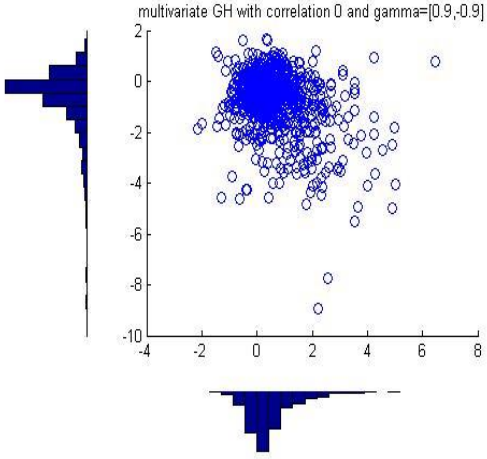
This normal mean-variance mixture structure can let GH have more flexibility to model dependence relationship between different dimension compare to Multivariate Gaussian distribution. To illustrate this, we show different parameter combination of γ and Σ in graphs below.





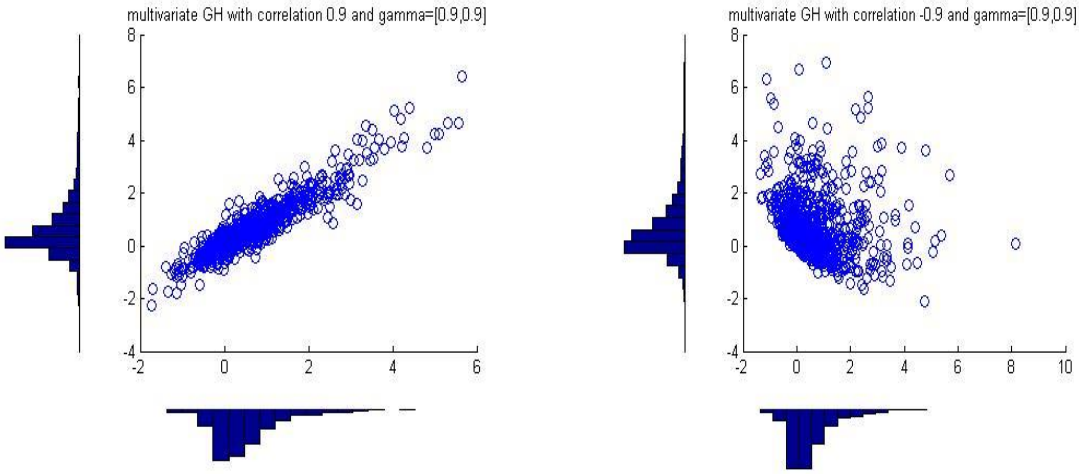
(a): Multivariate Gaussian distribution with correlation 0 and 0.9





Correlation 0 and $\gamma = [0.9,0.9]$

Correlation 0 and $\gamma = [0.9, -0.9]$



Correlation 0.9 and $\gamma = [0.9,0.9]$ Correlation -0.9 and $\gamma = [0.9,0.9]$

(b): Multivaraite GH distribution with different parameter

Figure 3: Comparison two dimensional Gaussian and GH

Definition 4.2 The density of a Generalized Inverse Gaussian (GIG) distribution is given by:

$$f_{GIG}(w;\lambda,\chi,\psi) = \left(\frac{\psi}{\chi}\right)^{\frac{\lambda}{2}} \frac{w^{\lambda-1}}{2\,K_\lambda(\sqrt{\chi\psi})} \exp\{-\frac{1}{2}(\frac{\chi}{w} + \psi w)\} \tag{4.4}$$

Where $K_\lambda()$ is modified Bessel function of the third kind and parameters satisfying

$\chi > 0, \psi \geq 0,\ \lambda < 0$

$\chi > 0, \psi > 0,\ \lambda = 0$

$\chi \geq 0, \psi > 0,\ \lambda > 0$.

The moment generating function of the GIG distribution is determined by

$$M_{GIG}(t) = \left(\frac{\psi}{\psi-2t}\right)^{\frac{\lambda}{2}} \frac{K_\lambda(\sqrt{\chi(\psi-2t)})}{K_\lambda(\sqrt{\chi\psi})} \tag{4.5}$$

The moment of a GIG random variable X can be found

$$E[X^n] = \left(\frac{\chi}{\psi}\right)^{\frac{n}{2}} \frac{K_{\lambda+n}(\sqrt{\chi\psi})}{K_\lambda(\sqrt{\chi\psi})} \tag{4.6}$$

And also,

$$E[\ln X] = \frac{dE[X^{\alpha}]}{d\alpha}|_{\alpha=0} \tag{4.7}$$

Noted that GIG contains some other class of distribution as special case, such as

Inverse Gaussian distribution for $\lambda = -0.5$

Gamma distribution for $\chi > 0 \; and \; \lambda > 0$

Inverse Gamma distribution for $\psi = 0 \; and \; \lambda < 0$

Properties

Density

Since the condition distribution of X given W is Gaussian distribution, see formula (4.2). Then we can find GH density by conditional density respect to W.

$$f_X(x) = \int_0^{\infty} f_{(X|W)}(x|w) f_W(w) \; dw \tag{4.8}$$

$$= \int_0^{\infty} \frac{e^{(x-\mu)'\Sigma^{-1}\gamma}}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}w^{\frac{d}{2}}} \exp\left\{-\frac{Q(x)}{2w} - \gamma'\Sigma^{-1}\gamma\frac{w}{2}\right\} f_W(w) \; dw$$

$$= \frac{\left(\sqrt{\psi/\chi}\right)^{\lambda}(\psi+\gamma'\Sigma^{-1}\gamma)^{\frac{d}{2}-\lambda}}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}K_{\lambda}(\sqrt{\chi\psi})} \; \frac{K_{\lambda-\frac{d}{2}}\left(\sqrt{(\chi+Q(x))(\psi+\gamma'\Sigma^{-1}\gamma)}\right) e^{(x-\mu)'\Sigma^{-1}\gamma}}{\left(\sqrt{(\chi+Q(x))(\psi+\gamma'\Sigma^{-1}\gamma)}\right)^{\frac{d}{2}-\lambda}}$$

Where $Q(x) = (x-\mu)'\Sigma^{-1}(x-\mu)$ and domain of (λ, χ, ψ) is the same as in GIG distribution, $\mu, \gamma \in \mathbb{R}^d$; $\Sigma \in \mathbb{R}^{d\times d}$.

Moment generating function

The moment generating function is calculated based on the moment generating function of GIG, see (4.5).

$$M_{GH}(t) = E\big[E[\exp(t'X)\,|W]\big] = e^{t'\mu}E\left[\exp\left\{W\left(t'\gamma + \frac{1}{2}\,t'\Sigma t\right)\right\}\right] \tag{4.9}$$

$$= e^{t'\mu}\left(\frac{\psi}{\psi - 2t'\gamma - t'\Sigma\, t}\right)^{\frac{\lambda}{2}} \frac{K_{\lambda}\left(\sqrt{\chi(\psi - 2t'\gamma - t'\Sigma\, t)}\right)}{K_{\lambda}(\sqrt{\chi\psi})}$$

with constraint $\chi \geq 2t'\gamma + t'\Sigma\, t$ for positivity.

Linear transformation

If $X \sim GH_d(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$ and $Y = BX + b$, where $B \in \mathbb{R}^{k \times d}$ and $b \in \mathbb{R}^k$, then $Y \sim GH_k(\lambda, \chi, \psi, B\mu + b, B\Sigma \mathrm{B}', B\gamma)$.

This means that GHD is closed under linear transformation and it can be proved by using characteristic function.

Proof

The characteristic function of X is

$$\phi_X(t) = E\left[E\left[e^{i\,t'X}\middle|W\right]\right] = E\left[e^{i\,t'(\mu + W\,\gamma) - 1/2\,W\,t'\Sigma\,t}\right] = e^{i\,t'\mu}\widehat{H}(1/2\;t'\Sigma\;t - i\;t'\gamma\;)$$

Where $\widehat{H}(\theta) = \int_0^\infty e^{-\theta v} d\,H(v)$ denotes the Laplace-Stieltjes transformation of the distribution function H of W. And the characteristic function of $Y = BX + b$ is

$$\phi_Y(t) = E\left[E\left[e^{i\,t'(BX+b)}\middle|W\right]\right] = E\left[e^{i\,t'(B\,(\mu + W\,\gamma)+b) - 1/2\,W\,t'\mathrm{B}\,\Sigma\,\mathrm{B}'\,t}\right]$$

$$= e^{i\,t'(B\,\mu + b)}\widehat{H}(1/2\;t'\mathrm{B}\,\Sigma\,\mathrm{B}'\;t - i\;t'B\gamma\;)$$

Therefore, $BX + b \sim GH_k(\lambda, \chi, \psi\,, B\mu + b, B\Sigma\,B', B\gamma)$.

Tail Behavior

The generalized hyperbolic distribution have semi-heavy tails, and in particular

$$gh(x; \lambda, \alpha, \beta, \delta) \sim |x|^{\lambda - 1} \exp\big((\mp\alpha + \beta)x\big) \quad as \quad x \to \pm\infty \tag{4.10}$$

up to a multiplicative constant (see Barndorff-Nielsen 1981). In figure below, we show the qqplot between Gaussian sample data and GH sample data and log-density of them. We can see that there is difference in the tail part.

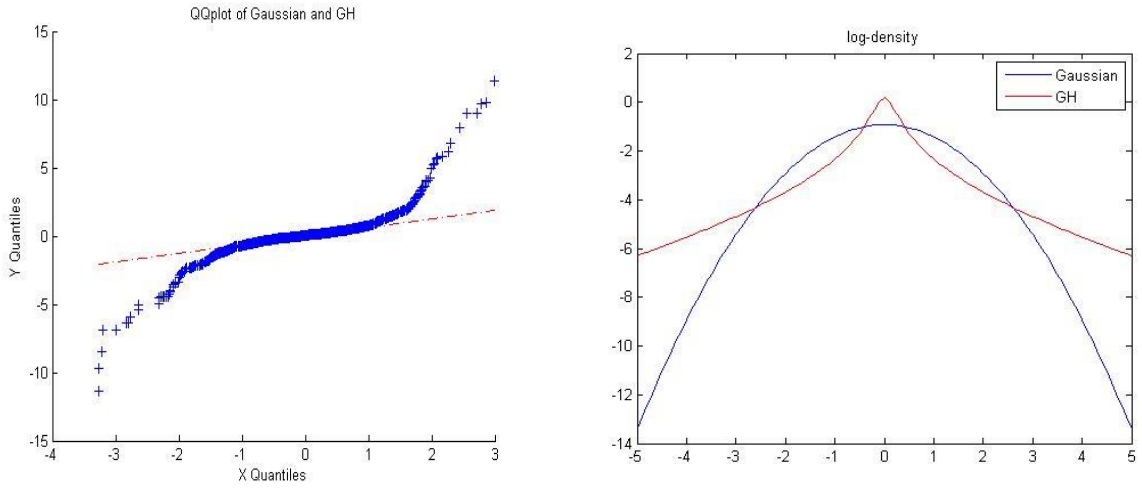


Figure 4: Tail behavior of Gaussian and GH with lambda=-0.2

4.2 Parametrization

There are alternative parametrization of GH distribution for different use. We will review three of them briefly since they are related to our parameter estimation application.

From definition in (4.1), we have parametrization $(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$. But this parametrization has a drawback of identification problem. Because the distribution $GH_d\ (\lambda, \chi, \psi, \mu, \Sigma, \gamma)$ and $GH_d\ (\lambda, \chi/k, k\psi, \mu, k\Sigma, k\gamma)$ are identical for any $k > 0$. Therefore, an identification problem occurs when we fit GH to a dataset. This problem could be solved by adding some constraint to the parameter set. In some application, one could require determinant of dispersion matrix to be 1, i.e. $|\Sigma| = 1$.

In this thesis, we use another more elegant way to eliminate the degree of freedom. We require the expected value of the GIG variable W to be one, i.e. $E[W] = 1$. This also makes the interpretation of the skewness parameter γ easier and the fitting procedure becomes faster. By the moment formula of GIG in (4.6), we define

$$E[W] = \sqrt{\frac{\chi}{\psi}}\,\frac{K_{\lambda+1}(\sqrt{\chi\psi})}{K_{\lambda}(\sqrt{\chi\psi})} = 1 \tag{4.11}$$

And set

$$\bar{\alpha} = \sqrt{\chi\psi} \tag{4.12}$$

Then we have parametrization $(\lambda, \bar{\alpha}, \mu, \Sigma, \gamma)$. It follows identity

$$\psi = \bar{\alpha}\frac{K_{\lambda+1}(\bar{\alpha})}{K_{\lambda}(\bar{\alpha})} \quad and \quad \chi = \bar{\alpha}\frac{K_{\lambda}(\bar{\alpha})}{K_{\lambda+1}(\bar{\alpha})} \tag{4.13}$$

The original parametrization is $(\lambda, \alpha, \mu, \Delta, \delta, \beta)$ when GH was firstly introduced by Barndorff-Nielsen (1977), the corresponding density as following:

$$f_X(x) = \frac{\left(\alpha^2 - \beta'\Delta\,\beta\right)^{\lambda/2}}{(2\pi)^{\frac{d}{2}}\sqrt{\Delta}\,\alpha^{\lambda-\frac{d}{2}}\delta^{\lambda}K_{\lambda}\left(\delta\sqrt{\alpha^2-\beta'\Delta\,\beta}\right)} \times \frac{K_{\lambda-\frac{d}{2}}\left(\alpha\sqrt{\delta^2+(x-\mu)'\Delta(x-\mu)}\right)e^{\beta'(x-\mu)}}{\left(\sqrt{\delta^2+(x-\mu)'\Delta(x-\mu)}\right)^{d/2-\lambda}} \tag{4.14}$$

And there is also an identification problem which can be solved by requiring $|\Delta| = 1$. The corresponding mixture representation is :

$$X|W = w \sim N_d(\mu + w\beta\Delta, w\Delta)$$

$$W \sim GIG\ (\lambda\,, \delta^2, \alpha^2 - \beta'\Delta\,\beta)$$

In the univariate case density (4.14) reduce to

$$f_X(x) = \frac{\left(\alpha^2 - \boldsymbol{\beta^2}\right)^{\lambda/2}}{\sqrt{2\pi}\ \alpha^{\lambda-\frac{1}{2}}\delta^{\lambda}K_{\lambda}\left(\delta\sqrt{\alpha^2-\beta^2}\right)} \times \frac{K_{\lambda-\frac{1}{2}}\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right)e^{\beta(x-\mu)}}{\left(\sqrt{\delta^2+(x-\mu)^2}\right)^{1/2-\lambda}} \tag{4.15}$$

If we define $\bar{\alpha} = \alpha\delta\,, \bar{\beta} = \beta\delta$, then the parametrization $(\lambda, \bar{\alpha}, \bar{\beta}\,, \mu)$ is scale and location invariant for GH with respect to parameters $\lambda, \bar{\alpha}, \bar{\beta}$. To prove this, suppose a linear transformation $Y = a\,X + b$ and $\sim GH_1(\lambda, \alpha, \beta, \delta, \mu)$, then by property of linear transformation of GH, Y is also GH with parameters $\lambda^+ = \lambda\,, \alpha^+ = \alpha/\,|a|\,, \beta^+ = \beta/|a|,\ \delta^+ = \delta\,|a|$ and $\mu^+ = a\mu + b$. We can see that $\alpha^+\delta^+ = \alpha\delta$ and $\beta^+\delta^+ = \beta\delta$.

We can use formulas to switch between different parametrization. We list them as below

$(\lambda, \chi, \psi, \mu, \Sigma, \gamma) \to (\lambda, \bar{\alpha}, \mu, \Sigma, \gamma)$: $\bar{\alpha} = \sqrt{\chi\psi}$ and set $\sqrt{\frac{\chi}{\psi}}\,\frac{K_{\lambda+1}\left(\sqrt{\chi\psi}\right)}{K_{\lambda}\left(\sqrt{\chi\psi}\right)} = 1$ (4.16)

$(\lambda, \chi, \psi, \mu, \Sigma, \gamma) \to (\lambda, \alpha, \mu, \Delta, \delta, \beta)$:

$$\Delta = |\Sigma|^{-\frac{1}{d}}\,\Sigma\ \,, \beta = \Sigma^{-1}\gamma$$

$$\delta = \sqrt{\chi|\Sigma|^{\frac{1}{d}}}\ \,, \alpha = \sqrt{|\Sigma|^{-\frac{1}{d}}(\psi + \gamma'\Sigma^{-1}\gamma)} \tag{4.17}$$

$(\lambda, \alpha, \mu, \Delta, \delta, \beta) \to (\lambda, \chi, \psi, \mu, \Sigma, \gamma)$:

$$\Sigma = \Delta\,,\ \gamma = \Delta\,\beta\,, \chi = \delta^2\,, \psi = \alpha^2 - \beta'\Delta\,\beta \tag{4.18}$$

$(\lambda, \alpha, \mu, \delta, \beta) \to \left(\lambda, \bar{\alpha}, \bar{\beta}, \mu\right)$ for univariate case :

$$\bar{\alpha} = \alpha\delta \, , \bar{\beta} = \beta\delta \tag{4.19}$$

4.3 Parameter Estimation

4.3.1 Univariate Standardized GH Estimation

Univaraite GH can be fitted by maximum likelihood estimation. One can use parametrization (4.8) or (4.15). Suppose we have observations $x_1, x_2, \dots x_n$. By using parametrization (4.15), then the log-likelihood function is

$$L = \sum_{i=1}^{n} \log f(x_i) = n\, log(a) + \left(\frac{\lambda}{2} - \frac{1}{4}\right) \sum_{i=1}^{n} \log(\delta^2 - (x_i - \mu)^2)$$

$$+ \sum_{i=1}^{n} [\log\left(K_{\lambda-0.5}\left(\alpha\sqrt{\delta^2 + (x_i - \mu)^2}\right)\right) + \beta(x_i - \mu)]$$

Therefore, the estimator could be obtained by using numerical optimizer directly.

In order to modeling standardized innovation sequence u_t in FARIMA-FIGARCH model, we have to maintain the constraint that $E[u_t] = 0, var[u_t] = 1$, i.e. we need to define standardized GH distribution. Of course, one can force any parametrization with constraint mean 0 and variance 1. For example, if we use parametrization $(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$, and by formula (4.3), we have

$$E[X] = \mu + \gamma = 0$$

$$var[X] = var(W)\gamma^2 + \Sigma = 1 \tag{4.20}$$

Then we have

$$\Sigma = 1 - var(W)\gamma^2 = 1 - (E[W^2] - 1)\gamma^2$$

$$= 1 + \gamma^2 - \gamma^2 \left(\frac{\chi}{\psi}\right) \frac{K_{\lambda+2}\left(\sqrt{\chi\psi}\right)}{K_{\lambda}\left(\sqrt{\chi\psi}\right)}$$

The domain of Σ is positive, apparently, there is some value γ would fail to satisfy this condition. One certainly could put extra constraint to fulfill this during the optimization routine. But this is not the elegant way while there is a better way to avoid this problem by using scale and location invariant parametrization $\left(\lambda, \bar{\alpha}, \bar{\beta}, \mu\right)$. We have expected value and variance formula as following:

$$E[X] = \mu + \frac{\bar{\beta}\delta}{\zeta}\frac{K_{\lambda+1}(\zeta)}{K_{\lambda}(\zeta)}$$

$$Var[X] = \delta^2\left(\frac{K_{\lambda+1}(\zeta)}{\zeta K_{\lambda}(\zeta)} + \frac{\bar{\beta}^2}{\bar{\alpha}^2-\bar{\beta}^2}\left[\frac{K_{\lambda+2}(\zeta)}{K_{\lambda}(\zeta)} - \left(\frac{K_{\lambda+1}(\zeta)}{K_{\lambda}(\zeta)}\right)^2\right]\right)$$

Where $\zeta = \sqrt{\bar{\alpha}^2 - \bar{\beta}^2}$. It is easy to see that if we set

$$\mu = -\frac{\bar{\beta}\delta}{\zeta}\frac{K_{\lambda+1}(\zeta)}{K_{\lambda}(\zeta)}$$

$$\delta = \left(\frac{K_{\lambda+1}(\zeta)}{\zeta K_{\lambda}(\zeta)} + \frac{\bar{\beta}^2}{\bar{\alpha}^2-\bar{\beta}^2}\left[\frac{K_{\lambda+2}(\zeta)}{K_{\lambda}(\zeta)} - \left(\frac{K_{\lambda+1}(\zeta)}{K_{\lambda}(\zeta)}\right)^2\right]\right)^{-\frac{1}{2}}$$

Then $GH(\lambda, \bar{\alpha}, \bar{\beta})$ have mean 0 and variance 1, and denoted as $stdGH(\lambda, \bar{\alpha}, \bar{\beta})$. This parametrization can also be switched with other form parametrization by formula (4.19).

4.3.2 Multivariate GH Estimation

A modified EM algorithm is used in this thesis, it is called multi-cycle, expectation, conditional estimation (MCECM) algorithm, see McNeil, Frey and Embrects, 2005. We introduce EM scheme and present different steps of MCECM algorithm as following.

Assume we have i.i.d data sample $x_1, x_2, \dots, x_n$; $x_i \in \mathbb{R}^d$, with parameter set $\Theta = (\lambda, \bar{\alpha}, \mu, \Sigma, \gamma)$. The MLE is trying to maximize

$$\ln L(\Theta; x_1, x_2, .., x_n) = \sum_{i=1}^{n} \ln f_X(x_i; \Theta) \tag{4.21}$$

Suppose we also have observation for $w_1, w_2, .. w_n$, then we have an augmented likelihood function:

$$\ln \hat{L}(\Theta; x_1, x_2, \dots, x_n, w_1, w_2, \dots, w_n) = \sum_{i=1}^{n} \ln f_{X|W}(x_i|w_i; \mu, \Sigma, \gamma) + \sum_{i=1}^{n} \ln f_W(w_i; \lambda, \bar{\alpha}) \tag{4.22}$$

If we have estimates for latent variable w_i, then we can maximize (4.22), and updating parameter set repeatedly. This is exactly how EM scheme works.

E-step: Calculate the conditional expectation of likelihood function (4.22) given the data sample $x_1, x_2, \dots, x_n$ and the current estimates of parameters $\Theta^{[k]}$:

$$Q\left(\Theta, \Theta^{[k]}\right) = E\left[\ln \hat{L}(\Theta; x_1, x_2, \dots, x_n, w_1, w_2, \dots, w_n)|x_1, .., x_n; \Theta^{[k]}\right] \tag{4.23}$$

M-step: Maximize the objective function (4.23) with respect to Θ to get the next step estimates $\Theta^{[k+1]}$.

We can see that there are two parts summation of (4.22). We can further decompose (4.23) into:

$$Q\left(\Theta, \Theta^{[k]}\right) = E\left[\sum_{i=1}^{n} L_1 | x_1, .., x_n; \Theta^{[k]}\right] + E\left[\sum_{i=1}^{n} L_2 | x_1, .., x_n; \Theta^{[k]}\right] \quad (4.24)$$

Where

$$L_1 = \ln f_{X|W}(x_i | w_i; \mu, \Sigma, \gamma)$$

$$L_2 = \ln f_W(w_i; \lambda, \bar{\alpha})$$

We know that $X|W$ is normal distribution, see (4.2) , then we have explicit expression of L_1 :

$$L_1 = -\frac{d}{2}\ln 2\pi - \frac{1}{2}\ln(w_i|\Sigma|) - \frac{1}{2}(x_i - \mu - w_i\gamma)' w_i^{-1}\Sigma^{-1}(x_i - \mu - w_i\gamma) \quad (4.25)$$

And W is GIG distribution:

$$L_2 = \frac{\lambda}{2}\ln\left(\frac{\psi}{\chi}\right) - \ln\left(2K_\lambda\sqrt{\psi\chi}\right) + (\lambda - 1)\ln w_i - \frac{\chi}{2}\frac{1}{w_i} - \frac{\psi}{2}w_i \quad (4.26)$$

Use (4.12), L_2 is a function of $(\lambda, \bar{\alpha})$.

Take the partial derivative respect to each parameter of Θ, we will deduce the update equation for each parameter. Let

$$\eta_i^{[k]} = E\left[w_i | x_i; \Theta^{[k]}\right] \text{ and } \bar{\eta}^{[k]} = \frac{1}{n}\sum_{i=1}^{n}\eta_i^{[k]} \quad (4.27)$$

$$\delta_i^{[k]} = E\left[w_i^{-1} | x_i; \Theta^{[k]}\right] \text{ and } \bar{\delta} = \frac{1}{n}\sum_{i=1}^{n}\delta_i^{[k]} \quad (4.28)$$

$$\xi_i^{[k]} = E\left[\ln w_i | x_i; \Theta^{[k]}\right] \quad (4.29)$$

The conditional variable $w_i|x_i$ is still a GIG random variable, see Appendix 2. Therefore these expectations could be calculated by moments of GIG, see (4.6). Then we have update estimate equation for (μ, γ, Σ) , see details in Appendix 3:

$$\gamma^{[k+1]} = \frac{1}{n}\frac{\sum_{i=1}^{n}\delta_i^{[k]}(\bar{x} - x_i)}{\bar{\eta}^{[k]}\bar{\delta}^{[k]} - 1} \quad (4.30)$$

$$\mu^{[k+1]} = \frac{1}{n}\frac{\sum_{i=1}^{n}(\delta_i^{[k]}x_i - \gamma^{[k+1]})}{\bar{\delta}^{[k]}} \quad (4.31)$$

$$\Sigma^{[k+1]} = \frac{1}{n}\sum_{i=1}^{n}\delta_i^{[k]}\left(x_i - \mu^{[k+1]}\right)\left(x_i - \mu^{[k+1]}\right)' - \bar{\eta}^{[k]}\gamma^{[k+1]}\gamma^{[k+1]\prime} \quad (4.32)$$

We present steps MCECM algorithm as following:

(1) Select initial values for parameter set $\Theta^{[k]}$ and tolerance ϵ.
(2) Calculate $\chi^{[k]}$ and $\psi^{[k]}$ based on value of $\bar{\alpha}^{[k]}$ by using formula(4.13).
(3) Use (4.27), (4.28), (4.29) to calculate $\eta_i^{[k]}$ and $\delta_i^{[k]}$ and their average.
(4) If an symmetric model, set $\gamma = 0$, otherwise set $\gamma^{[k+1]}$ by (4.30).
(5) Update μ and Σ by (4.31) and (4.32).
(6) Set $\Theta^{[k,2]} = (\lambda^{[k]}, \bar{\alpha}^{[k]}, \mu^{[k+1]}, \Sigma^{[k+1]}, \gamma^{[k+1]})$ and calculate weights $\eta_i^{[k,2]}, \delta_i^{[k,2]}, \xi_i^{[k,2]}$.
(7) Maximize the second summand of (4.24) with density (4.26) with respect to $\lambda, \bar{\alpha}$. Update $\Theta^{[k,2]}$ by using $\lambda^{[k+1]}$ and $\bar{\alpha}^{[k+1]}$.
(8) Check the updated distance $D = \left\|\Theta^{[k]} - \Theta^{[k+1]}\right\|_{Frobenius}$. Go back to step (2) if $D > \epsilon$; otherwise return $\Theta^{[k+1]}$ as final result.

4.4 Simulation

We can use mixing structure (4.1) to generate random variable $GH_d(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$. To draw N independent d-dimensional random samples from the Multivariate GH distribution, we use the following algorithm:

1. Draw N independent d-dimensional normal random samples from $N(0, \Sigma)$ by using Cholesky decomposition $\Sigma = AA'$.
2. Draw N independent GIG random samples from $w \sim GIG(\lambda, \chi, \psi)$.
3. Mixing them by formula (4.1) to obtain d-dimensional GH random variable.

We refer to Wolfgang H. and Josef L. (2013) for details to generate GIG random variables.

4.5 Practical Issue for High Dimension

Although we have algorithm to perform multivariate GH estimation, there is some numerical issue in practical application for large system. One of the major issues is coming from the high order of Modified Bessel function $K_\lambda(x)$. Of course, one can reduce dimension of the data by using dimension reduction technique, such as principal component analysis (PCA). But the issue still present in some cases. We noticed that there is $K_\lambda\left(\sqrt{\chi\psi}\right)$ term in GIG log likelihood or any moments of GIG, see (4.26) and (4.5). And the conditional density

$$W|X \sim GIG(\lambda - d/2, Q(x) + \chi\,, \psi + \gamma'\Sigma^{-1}\gamma) \tag{4.33}$$

This means we need to evaluate Modified Bessel function with base order $\lambda - d/2$. When dimension goes high, $d/2 \gg \lambda$ and $\sqrt{(Q(x)+\chi)(\psi+\gamma'\Sigma^{-1}\gamma)}$ will dependent on the data, but usually it is small compare to $\lambda - d/2$.

The modified Bessel function has integral representation

$$K_\lambda(x) = \frac{1}{2}\int_0^\infty w^{\lambda-1} \exp\left\{-\frac{1}{2}x(w+w^{-1})\right\} dw \ , \quad x > 0$$

And has symmetric property respect to λ

$$K_\lambda(x) = K_{-\lambda}(x)$$

Figure 3 shows Modified Bessel function with different order value λ. We can see that the function value grows very quick when x is small. And this is exactly the problem in high dimensional multivariate GH estimation. As far as we know, there is no good solution for compute high order Bessel function. Therefore, we are using asymptotic approximation formula for small x as below:

$$K_\lambda(x) \sim \frac{\Gamma(\lambda)}{2}\left(\frac{2}{x}\right)^\lambda \quad for\ \lambda > 0\ \ and\ \ 0 < |x| \ll \sqrt{\lambda+1}$$

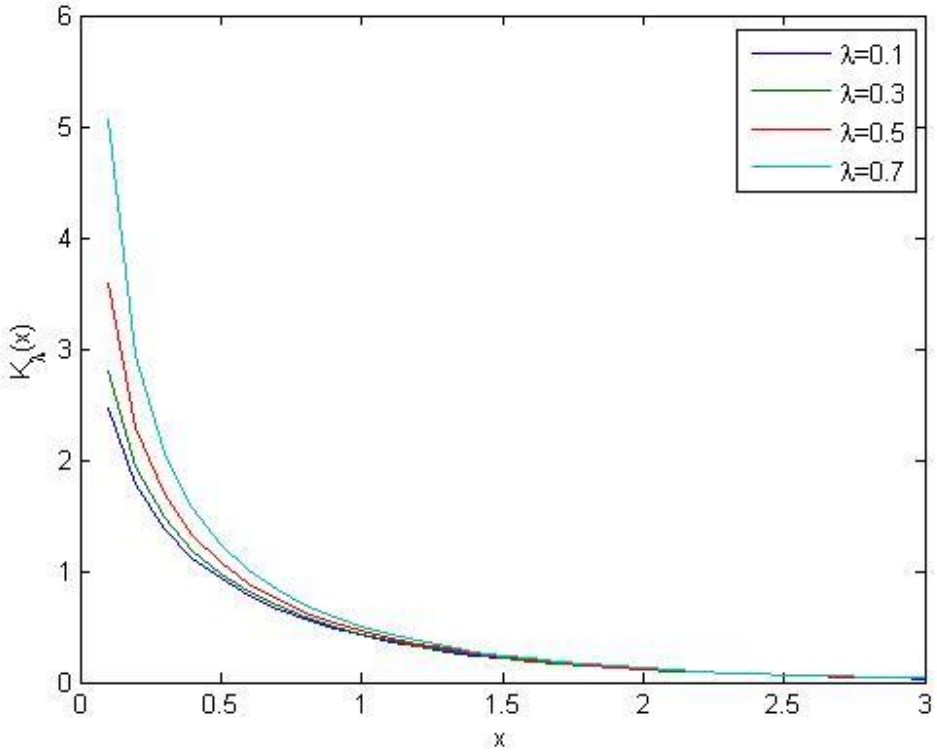


Figure 5: Modified Bessel function with different order

Chapter 5 Backtesting Procedure and Empirical Results

5.1 Model Description

Suppose we have N assets in our portfolio and have log-return time series data for each asset.

$$r_i = (r_{i1}, r_{i2}, \dots, r_{it})' , i = 1,2.., N$$

Then we use FARIMA $(1, d_0, 1)$-FIGARCH $(1, d, 1)$ as our marginal model to capture long memory feature of r_i and assume u_t follows stdGH distribution which defined in section 4.3.1 to capture heavy tailed feature. From forecasting formula we can see that all the randomness come from u_{t+1} , see equation (3.26). In portfolio level, we can write as

$$r_p = \mu_p + \sigma_p U \tag{5.1}$$

Where

$$r_p = (r_1, r_2, \dots, r_N)$$

$$\mu_p = diag\big(E\big[r_{1,t+1}\big|\Omega_t\big], \dots, E\big[r_{N,t+1}\big|\Omega_t\big]\big)$$

$$\sigma_p = diag(\sqrt{h_{1,t+1}}, .., \sqrt{h_{N,t+1}})$$

Upon time t, μ_p and σ_p is deterministic. Therefore, portfolio dependence structure is coming from U and we assume U follows multivariate GH distribution:

$$U \sim GH(\lambda, \chi, \psi, \mu, \gamma, \Sigma)$$

By the property of GH under linear transformation, see section 4.1. r_p is a N dimensional GH distribution:

$$r_p \sim GH_N(\lambda, \chi, \psi, \mu_{port}, \gamma_{port}, \Sigma_{port})$$

Where

$$\mu_{port} = \mu_p + \sigma_p \mu$$

$$\gamma_{port} = \sigma_p \gamma$$

$$\Sigma_{port} = \sigma_p \, \Sigma \, \sigma_p'$$

Suppose holding position in our portfolio is $w = (w_1, w_2, .., w_N)'$. Then the portfolio return is:

$$P = \sum_{i=1}^{N} w_i r_i = w' r_p$$

Again, by property of GH, R is a one dimensional GH distribution:

$$P \sim GH_1(\lambda, \chi, \psi, w'\mu_{port}, w'\gamma_{port}, w'\sigma_p \, \Sigma \, \sigma_p' w) \tag{5.2}$$

Then we have its density function explicitly, $f_P(x)$, by formula (4.8). Therefore, we can calculate risk measure, Value at Risk (VaR) and Conditional VaR (CVaR).

VaR at confident level α is defined as:

$$\alpha = P(\, P < -VaR_\alpha) = \int_{-\infty}^{-VaR_\alpha} f_P(x)\, dx$$

In real application, we need to approximate this integral numerically by choosing some lower bound value, say, x_l.

$$\int_{-\infty}^{-VaR_\alpha} f_P(x)\, dx \;\approx\; \int_{x_l}^{-VaR_\alpha} f_P(x)\, dx \approx \sum_{j=0}^{M} f_p(x_j)\Delta x \tag{5.3}$$

Where $x_j = x_l + j\, \Delta x$, and we choose $x_l = -5\, std(P)$, i.e. 5 times of standard deviation of GH distribution with parameters in (5.2). Given value of α, we can solve VaR_α satisfy (5.3).

CVaR is also called expected shortfall. $CVaR_\alpha$ is defined as:

$$CVaR_\alpha(P) = \frac{1}{\alpha} \int_0^\alpha VaR_x(P)\, dx$$

This also could be written as

$$CVaR_\alpha(P) = -\frac{1}{\alpha} \left(E\left[P\, 1_{\{P<P_\alpha\}}\right] + P_\alpha \left(\alpha - P(P \le P_\alpha)\right)\right)$$

Where P_α is the lower α quantile.

5.2 Backtesting Procedure

We perform backtest to our portfolio tail forecast by using VaR which calculated by formula (5.3). Suppose we backtesting T periods, say, $t_1, t_2, \dots, t_T$. At each time point t_i, we use previous H periods historical data to fit model, $t_{i-1}, t_{i-2}, \dots, t_{t-H}$ and make forecast for period t_i. Then we can compare forecast $VaR_\alpha(\, r_{t_i})$ and the realized observation r_{t_i} for $i = 1,2,..,T$. In risk management, we care about the quality of forecasted VaR sequence, i.e. violations number of testing sequence, and independence of violation sequence.

The first statistical test is Kupiec (1995) test, also called unconditional coverage test. Basically, it tests if the violation of sequence occurs with probability p or not. Given the indicator sequence $\{I_t\}$:

$$I_i = \begin{cases} 0 & \text{if } \mathrm{VaR}_\alpha(\mathrm{r_i}) \leq -\mathrm{r_i} \\ 1 & \text{if } \mathrm{VaR}_\alpha(\mathrm{r_i}) > -\mathrm{r_i} \end{cases}$$

Given independence of sequence of I_t, the null hypothesis is $E[I_t] = p$ and the alternative is $E[I_t] \neq p$. The likelihood under the null hypothesis is

$$L(p; I_1, .., I_T) = (1-p)^{n_0} p^{n_1}$$

and under the alternative

$$L(\pi; I_1, .., I_T) = (1-\pi)^{n_0} \pi^{n_1}$$

Then the likelihood ratio test could be formed as

$$LR_{uc} = -2 \log \frac{L(p; I_1, .., I_T)}{L(\hat{\pi}; I_1, .., I_T)} \sim \chi^2(1) \tag{5.4}$$

where $\hat{\pi}$ is the MLE estimate of π, i.e. $\hat{\pi} = n_1/(n_0 + n_1)$.

The assumption of Kupiec test is the independence of sequence I_t, Christoffersen (1998) propose an extended method to test the independence by using first order Markov chain of binary sequence I_t with transition probability matrix

$$M_1 = \begin{bmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{bmatrix}$$

where $\pi_{ij} = P(I_t = j | I_{t-1} = i)$. Therefore, approximate likelihood function is

$$L(M_1; I_1, .. I_T) = (1-\pi_{01})^{n_{00}} \pi_{01}^{n_{01}} (1-\pi_{11})^{n_{10}} \pi_{11}^{n_{11}}$$

Where n_{ij} is the number of observations with value i followed by j. The MLE estimator of π_{ij} is simply the ratio of corresponding counts

$$\widehat{M_1} = \begin{bmatrix} \frac{n_{00}}{n_{00}+n_{01}} & \frac{n_{01}}{n_{00}+n_{01}} \\ \frac{n_{10}}{n_{10}+n_{11}} & \frac{n_{11}}{n_{10}+n_{11}} \end{bmatrix}$$

Under the independent hypothesis

$$M_2 = \begin{bmatrix} 1-\pi_2 & \pi_2 \\ 1-\pi_2 & \pi_2 \end{bmatrix}$$

The likelihood would be

$$L(M_2; I_1, .. I_T) = (1-\pi_2)^{n_{00}+n_{10}} \pi_2^{n_{01}+n_{11}}$$

MLE of M_2 is

$$M_2 = \frac{n_{01}+n_{11}}{n_{00}+n_{10}+n_{01}+n_{11}}$$

Therefore, the LR test could be formulated as

$$LR_{ind} = -2\log\frac{L(\widehat{M_2}; I_1, .. I_T)}{L(\widehat{M_1}; I_1, .. I_T)} \sim \chi^2(1) \tag{5.5}$$

If combine unconditional coverage test and independence test, we have

$$LR_{cc} = -2\log\frac{L(p; I_1, .. I_T)}{L(\widehat{M_1}; I_1, .. I_T)} \sim \chi^2(2) \tag{5.6}$$

And we also have identity $LR_{cc} = LR_{uc} + LR_{ind}$.

The likelihood test above all based on sequence of violation sequence I_t, which only take two possible values. Berkowitz (2001) propose a density evaluation method to extract more information from the data. First, we transform data series r_t to uniform random variable by

$$x_t = \int_{-\infty}^{r_t} f(u)du$$

Which is the cumulative density function (CDF) of r_t. Under our model setting, this is the CDF of GH distribution. Then transform x_t to standard normal random variable $z_t = \Phi^{-1}(x_t) \sim N(0,1)$. To test independence, one can construct test against a first order autoregressive:

$$z_t - \mu = \rho(z_{t-1} - \mu) + \epsilon_t$$

Then the log-likelihood function of this equation could be rewritten as

$$L(\mu, \sigma^2, \rho) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log\left(\frac{\sigma^2}{1-\rho^2}\right) - \frac{\left(z_t - \mu/(1-\rho)\right)^2}{2\sigma^2/(1-\rho^2)}$$

$$-\frac{T-1}{2}\log 2\pi - \frac{T-1}{2}\log(\sigma^2) - \sum_{t=2}^{T}\left(\frac{(z_t - \mu - \rho z_{t-1})^2}{2\sigma^2}\right)$$

Where σ^2 is the variance of ϵ_t. Then a likelihood test could be constructed for testing independence of data series r_t

$$LR_{ind} = -2\big(L(\hat{\mu}, \hat{\sigma}^2, 0) - L(\hat{\mu}, \hat{\sigma}^2, \hat{\rho})\big) \sim \chi^2(1) \tag{5.7}$$

Where $\hat{\mu}$, $\hat{\sigma}^2$ could be approximated by mean and variance of z_t and $\hat{\rho}$ could be calculated as sample autocorrelation with lag 1. Within this framework, Berkowitz also propose a likelihood test based on censored normal distribution to test tail part of r_t which corresponding to large losses. Defining a tail by cutting off point by $= \Phi^{-1}(\alpha)$, for example, $VaR = \Phi^{-1}(0.05) = -1.64$. This means any observations not in the tail will be truncated and define a new variable as

$$z_t^* = \begin{cases} \text{VaR} & \text{if } \ \text{z}_\text{t} \geq \text{VaR} \\ \text{z}_\text{t} & \text{if } \ \text{z}_\text{t} < VaR \end{cases}$$

Then joint log-likelihood function of μ and σ is

$$L(\mu, \sigma | z^*) = \sum_{z_t^* < VaR} \log \frac{1}{\sigma} \phi\left(\frac{z_t^* - \mu}{\sigma}\right) + \sum_{z_t^* = VaR} \log\left(1 - \Phi\left(\frac{VaR - \mu}{\sigma}\right)\right)$$

$$= \sum_{z_t^* < VaR} \left(-\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma}(z_t^* - \mu)^2\right) + \sum_{z_t^* = VaR} \log\left(1 - \Phi\left(\frac{VaR - \mu}{\sigma}\right)\right)$$

Therefore, the likelihood ratio test for tail part could be formulated as

$$LR_{tail} = -2\big(\, L(0,1) - L(\hat{\mu}, \hat{\sigma}^2)\big) \sim \chi^2(2) \tag{5.8}$$

To sum up, we have five likelihood ratio tests for our backtesting procedure:

1. Kupiec test for unconditional coverage: LR_{uc} in (5.4).
2. Christoffersen test for independence of violation: LR_{ind} in (5.5).
3. Christoffersen test for independence and conditional coverage together: LR_{cc} in (5.6).
4. Berkowitz test for independence : LR_{Bind} in (5.7).
5. Berkowitz test for tail: LR_{tail} in (5.8).

They should be enough to test the quality of forecast VaR sequence under our model setting.

5.3 Empirical Result

In this section, we will present the in-sample and out-sample backtesting result, and analyze them based on statistical test.

5.3.1 Dataset

Our intra-day dataset is extracted from TAQ (Trade and Quote) database from NYSE. All the data has been adjusted by split and dividend which comes from yahoo finance website. We are using S&P 500 component stock data from Jan. 2010 to Apr. 2013 and all the data during the trading hours eastern time from 9:30 am to 4:00 pm for each trading day. We have aggregate the second price data to different frequency log-return data. We only choose the stocks which have

complete data series span entire periods. Under this selection rule, we totally have 472 stocks in our testing dataset.

5.3.2 In-sample distribution of parameter d

In this section, we will fit in-sample data over some periods to the marginal model, i.e. FIARIMA-FIGARCH model with stdGH distribution. We choose order 1 in FARIMA- memory FIGARCH structure, i.e. FARIMA$(1, d_0, 1)$-FIGARCH$(1, d, 1)$. Firstly, we exam the existence of long memory feature depends on the estimated parameter d for each frequency. We choose rolling window size to be 1000 and 2000. In table 2, we show the quantiles of parameter d in FARIMA-FIGARCH structure. We can see that 50% of d in FARIMA is less than 0, which means the market is very volatile. In meanwhile, d of FIGARCH exhibits different behavior for different frequency data. For 1 min, 10 min, 60 min frequency, d is average bigger than data of frequency 5 min and 10 min. And we also plot the histogram of parameter d and give us a sense that how the distribution of parameter d looks like.

Table 1: Quantiles of parameter d for different frequency data and window size

Frequency/ RollWindow	FARIMA $d \in (-0.5, 0.5)$				FIGARCH $d \in (0, 1)$			
	25%	50%	75%	100%	25%	50%	75%	100%
1min/1000	-0.1136	-0.0322	0.0343	0.4999	0.2233	0.3035	0.3797	0.7692
1min/2000	-0.0739	-0.0187	0.0335	0.4999	0.2258	0.3121	0.3878	0.8167
5min/1000	-0.0677	-0.0147	0.0370	0.4999	0.0740	0.1765	0.2976	0.6213
5min/2000	-0.0556	-0.0132	0.0341	0.4999	0.0733	0.1473	0.2352	0.5330
10min/1000	-0.0668	-0.0276	0.0084	0.2934	0.0154	0.0398	0.0808	0.4455
10min/2000	-0.0430	-0.0174	0.0140	0.4999	0.0330	0.0650	0.1097	0.4597

30min/1000	-0.0288	0.0049	0.0781	0.4999	0.1118	0.1523	0.1945	1.0000
30min/2000	-0.0355	-0.0056	0.0325	0.4999	0.0803	0.1051	0.1339	0.9907
60min/1000	-0.2237	-0.0859	-0.0059	0.1087	0.0654	0.1363	0.2022	0.4967
60min/2000	-0.0744	-0.0247	-0.0004	0.1204	0.1164	0.1560	0.1962	0.3599

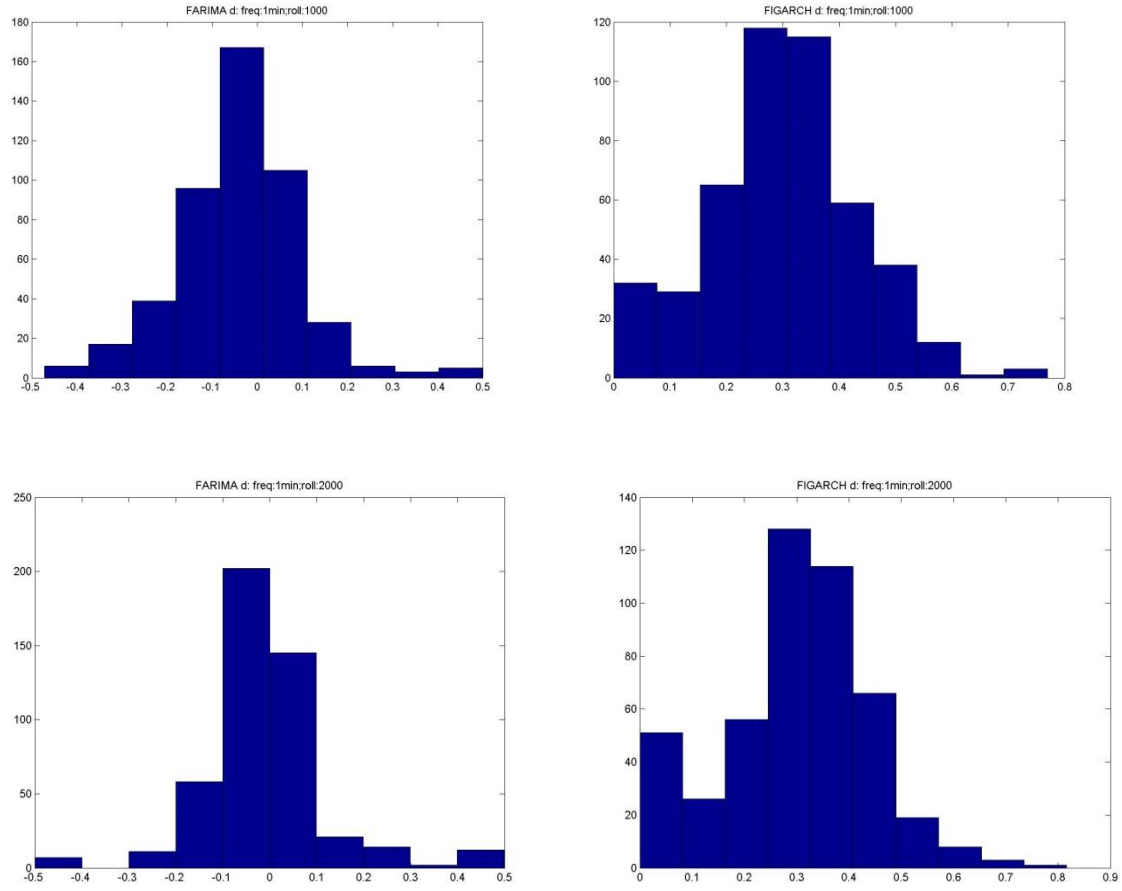


Figure 6: Histogram of parameter d for 1 min

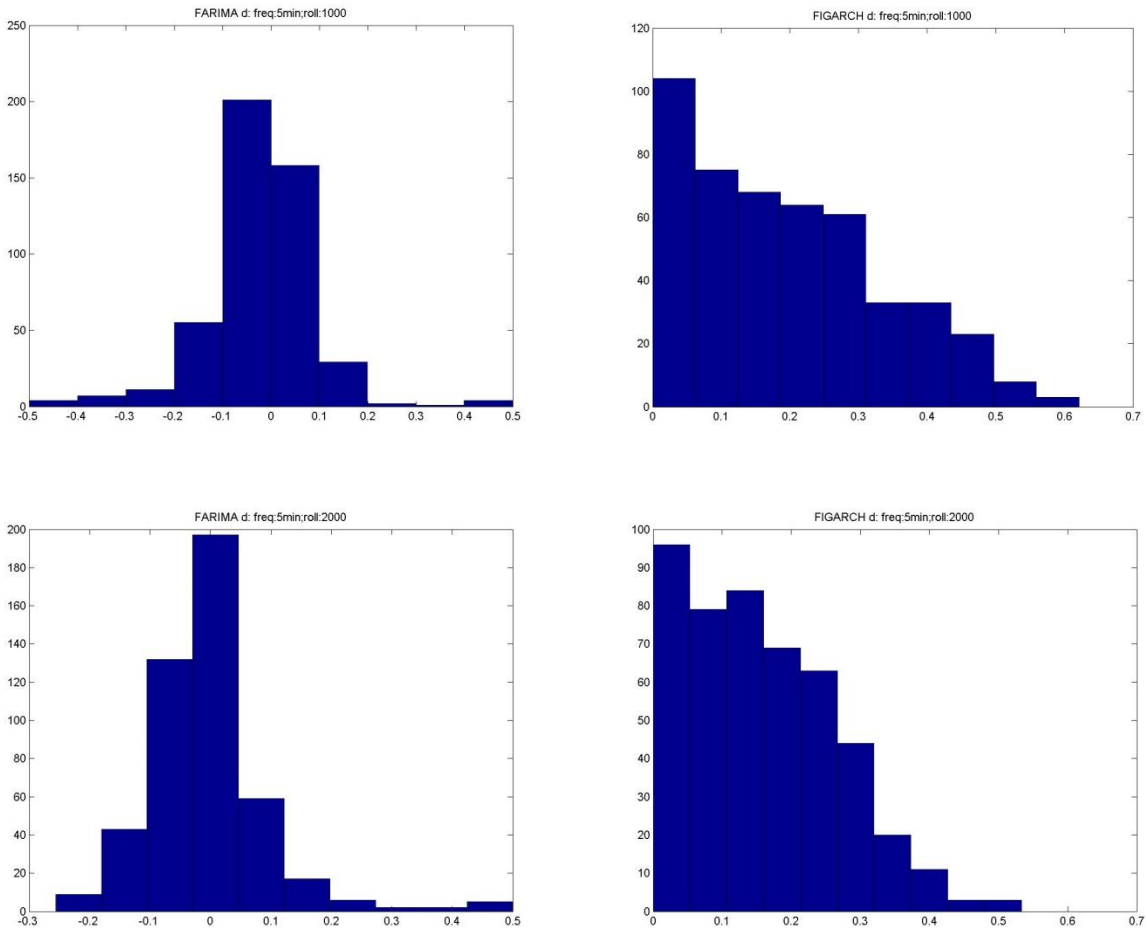
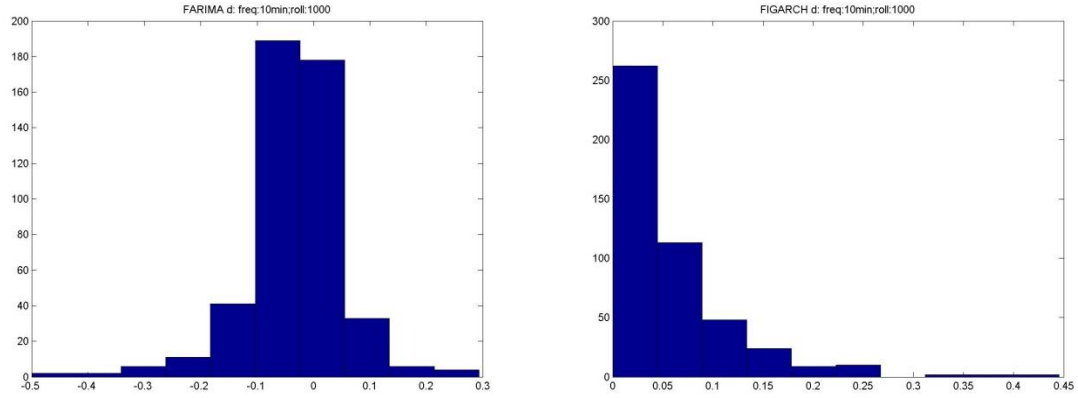


Figure 7: Histogram of parameter d for 5 min



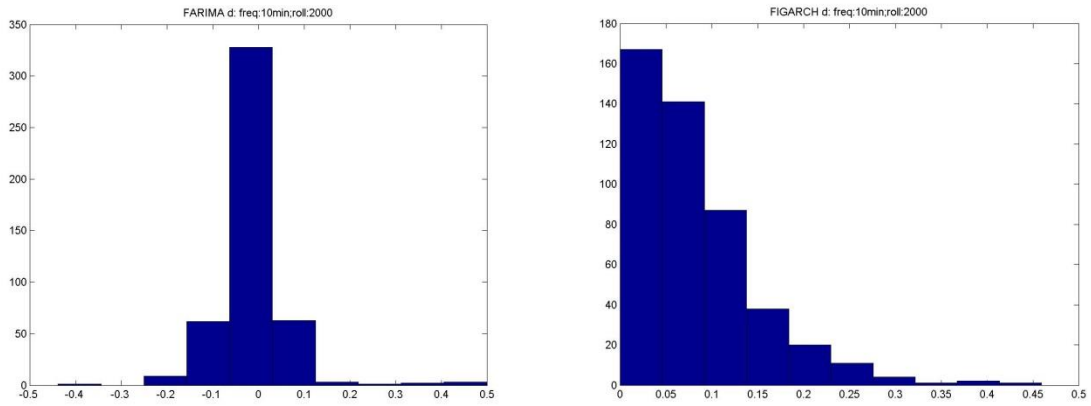


Figure 8: Histogram of parameter d for 10 min

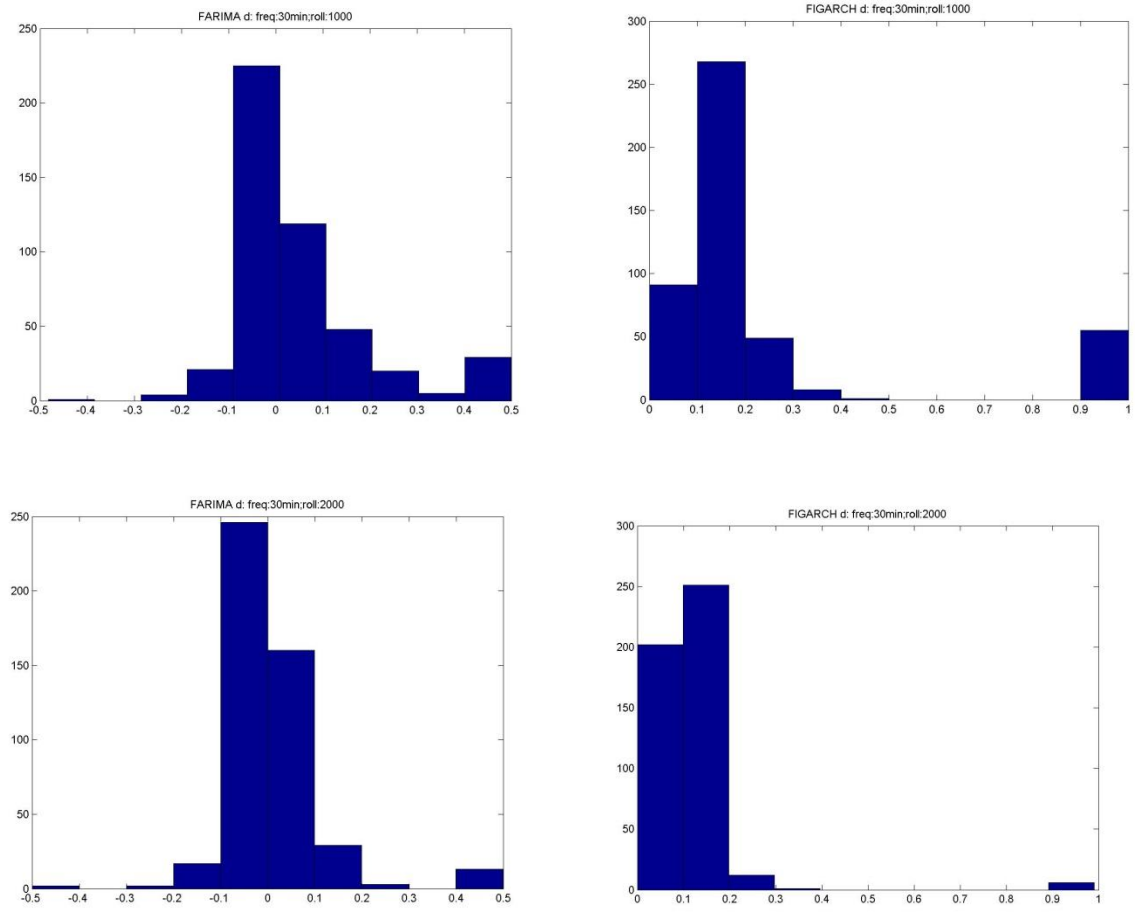


Figure 9: Histogram of parameter d for 30 min

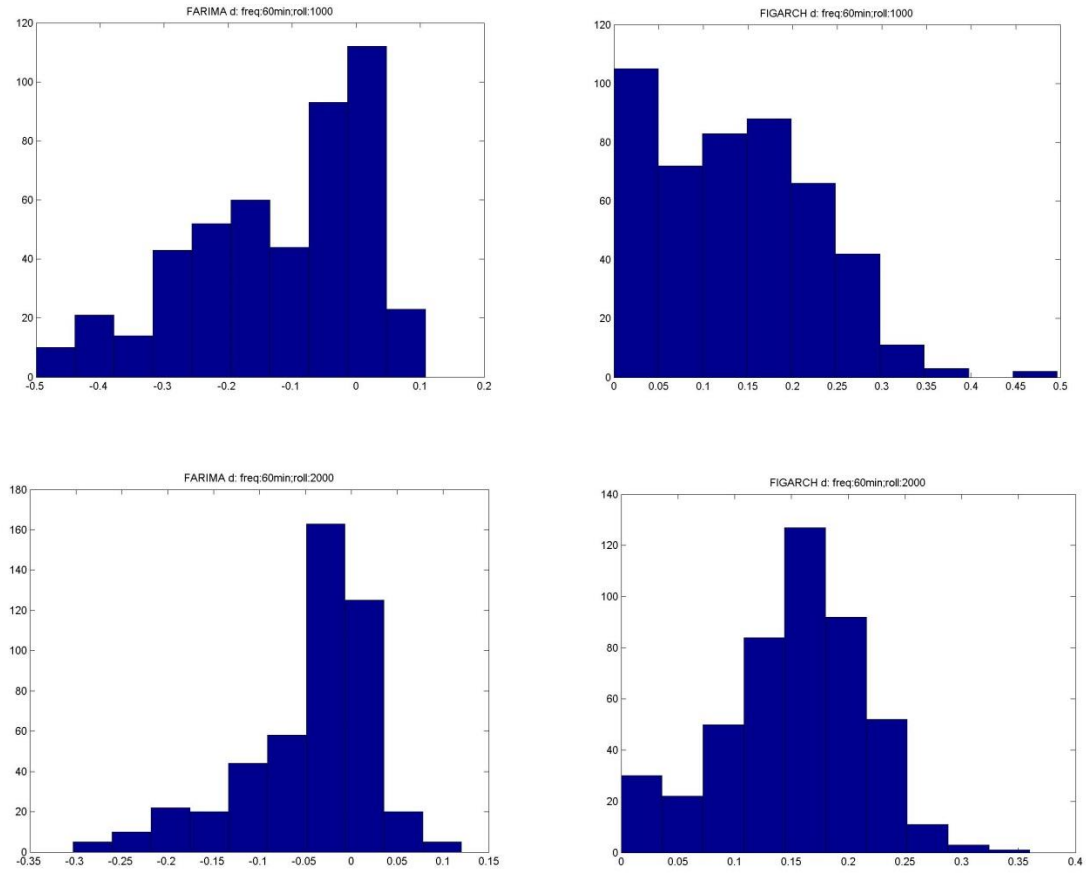


Figure 10: Histogram of parameter d for 60 min

5.3.3 Out-sample test

We perform backtesting procedure in section 5.2 over some periods. For each period, we use rolling window size H as input data length to fit our model. In order to exam how large H is good enough, we use the backtesting framework that Basel III required and combine with the statistical test as our checking tools.

Firstly, we compare ARMA-GARCH model and FARIMA-FIGARCH model combined with multivariate Gaussian distribution and multivariate generalized hyperbolic distribution for innovations. We backtest 500 total periods for portfolio of 30 assets from 20120613 to 20121001 for hourly data. The portfolio is equally weighted. We plot the portfolio return series and also

VaR and CVaR series in the following graph. In table 2 and 3, we present the statistical test for each case. We can see that the violation number of VaR of GH is less than Gaussian case. And the Gaussian fails for Berkowitz tail test in both models which confirms markets has heavy tails feature and tail dependence. And multivariate GH fails Berkowitz independent test for ARMA-GARCH model which can be viewed as market is more volatile and FIARIMA-FIGARCH model has the ability to modeling volatile market condition.

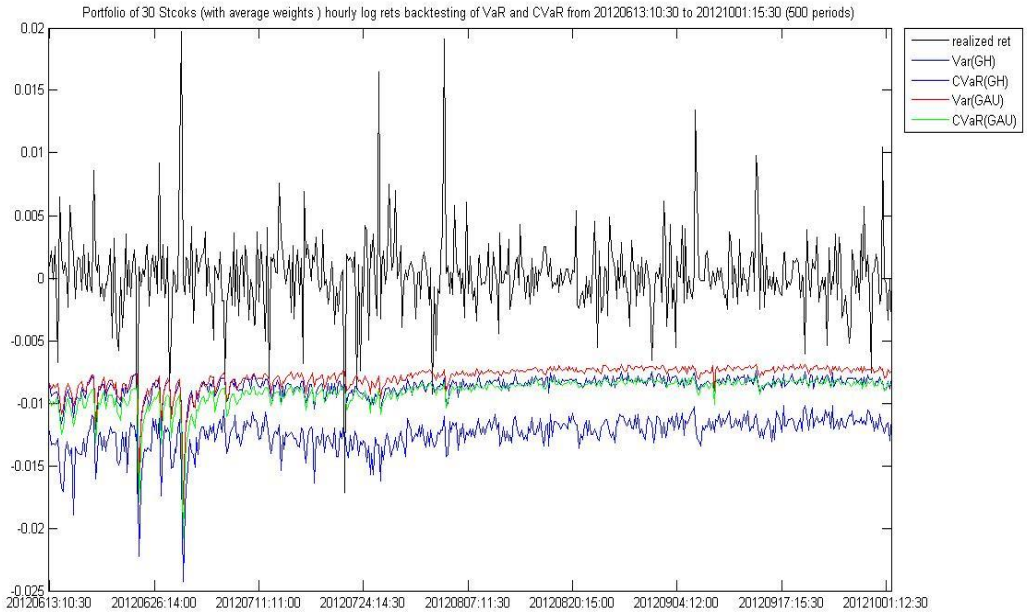


Figure 11: FARIMA-FIGARCH model backtesting 500 periods for hourly data

Table 2: statistical test for FARIMA-FIGARCH at confident level alpha =0.01

Statistics(p-value)	VaR(GH)	VaR(GAUSSIAN)
# Violation	6	8
Kupeic test	0.18988(0.663016)	1.53828(0.214874)
Independent test	0.145753(0.702628)	0.260174(0.610001)

Christoffersen test	0.335633(0.845509)	1.79845(0.406885)
Berkowitz tail test	1.65934(0.436193)	**21.2985(2.37183e-05)**
Berkowitz independent test	4.32045(0.037657)	3.89963(0.0482967)

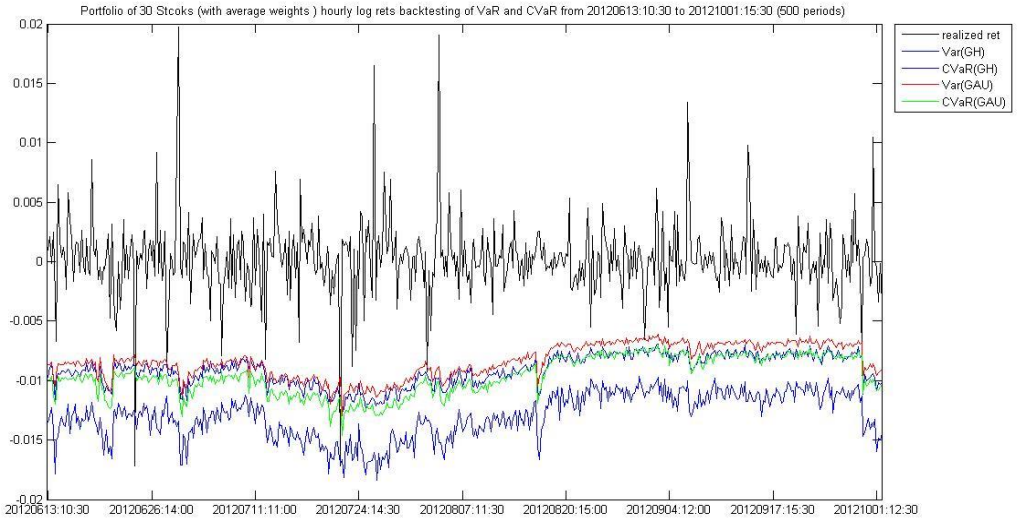


Figure 12: ARMA-GARCH model backtesting 500 periods for hourly data

Table 3: statistical test for ARMA-GARCH at confident level alpha=0.01

Statistics(p-value)	VaR(GH)	VaR(GAUSSIAN)
# Violation	3	6
Kupeic test	0.943116(0.331478)	0.18988(0.663016)
Independent test	0.0362175(0.849067)	0.145753(0.702628)
Christoffersen test	0.979334(0.612831)	0.335633(0.845509)
Berkowitz tail test	0.250309(0.88236)	**11.9359(0.00255946)**
Berkowitz independent test	**7.43331(0.00640276)**	5.79872(0.0160378)

The very practical question is that how many assets our model still working without reduce dimension, i.e. how high dimensions we can go for our model and how much data is sufficient as our input to our model. Our second experiment is trying to answer this question. We backtest our model for different asset numbers and different rolling window size over 30 minutes frequency data. We choose candidate asset number as 30, 60, 120, 240, 475, and rolling window size as 1000 and 2000. Violation numbers are presented in Table 4 and statistical test result in Table 5. GH and Gaussian are all in the green zone, and GH is always better than Gaussian in terms of violation number. We can see that the Gaussian case fails the Berkowitz tail test all the time. And when the dimension goes high, some GH case fails the Berkowitz tail test, but bigger rolling window size is better than small one, which means the more data we have, the better result we get. This makes sense since for high dimension problem, parameters that need to be estimated grows, and more data would help us to perform better estimation step. For this particular test, we see that GH passes all the test except 475 assets with 1000 rolling window size. This indicates that we need more data to fit model when dimension goes high, otherwise PCA should be applied to reduce dimension.

Table 4: violation number of VaR for different asset number

Asset/rollingWindow	VaR(GH)	VaR(GAUSSIAN)
30/1000	8	9
30/2000	6	10
60/1000	7	9
60/2000	5	8
120/1000	7	8
120/2000	6	10
240/1000	7	9
240/2000	6	10
475/1000	8	9
475/2000	7	9

Table 5: statistical test for FIARMA-FIGARCH at confident level alpha=0.01 for different asset number

Asset/	Kupeic		Independent		Christoffersen		Berkowitz indep.		Berkowitz tail	
rollWindow	GH	GAU	GH	GAU	GH	GAU	GH	GAU	GH	GAU
30/1000	0.4337 (0.5102)	0.1045 (0.7465)	0.1290 (0.7194)	0.1635 (0.6860)	0.5628 (0.7547)	0.2680 (0.8746)	2.6746 (0.1020)	4.0002 (0.0455)	0.9698 (0.6158)	42.6703 **(0.0000)**
30/2000	1.8862 (0.1696)	0 (1.0000)	0.0724 (0.7878)	0.2020 (0.6531)	1.9587 (0.3756)	0.2020 (0.9039)	0.2668 (0.6055)	0.6734 (0.4119)	1.2897 (0.5247)	27.3054 **(0.0000)**
60/1000	1.0156 (0.3136)	0.1045 (0.7465)	0.0987 (0.7534)	0.1635 (0.6860)	1.1143 (0.5728)	0.2680 (0.8746)	0.7817 (0.3766)	1.7521 (0.1856)	0.7494 (0.6875)	33.1539 **(0.0000)**

60/2000	3.0937 (0.0786)	0.4337 (0.5102)	0.0503 (0.8226)	0.1290 (0.7194)	3.1440 (0.2076)	0.5628 (0.7547)	0.0145 (0.9040)	0.2292 (0.6321)	0.6229 (0.7324)	34.5427 **(0.00)**
120/1000	1.0156 (0.3136)	0.4337 (0.5102)	0.0987 (0.7534)	0.1290 (0.7194)	1.1143 (0.5728)	0.5628 (0.7547)	0.7156 (0.3976)	1.1228 (0.2893)	2.2824 (0.3194)	32.9544 **(0.0000)**
120/2000	1.8862 (0.1696)	0 (1.0000)	0.0724 (0.7878)	0.2020 (0.6531)	1.9587 (0.3756)	0.2020 (0.9039)	0.0565 (0.8122)	0.2016 (0.6534)	1.1378 (0.5662)	30.0265 **(0.0000)**
240/1000	1.0156 (0.3136)	0.1045 (0.7465)	0.0987 (0.7534)	0.1635 (0.6860)	1.1143 (0.5728)	0.2680 (0.8746)	0.6258 (0.4289)	0.7962 (0.3722)	4.6990 (0.0954)	27.0501 **(0.0000)**
240/2000	1.8862 (0.1696)	0 (1.0000)	0.0724 (0.7878)	0.2020 (0.6531)	1.9587 (0.3756)	0.2020 (0.9039)	0.0452 (0.8317)	0.1178 (0.7315)	2.1418 (0.3427)	28.4817 **(0.0000)**
475/1000	0.4337 (0.5102)	0.1045 (0.7465)	0.1290 (0.7194)	0.1635 (0.6860)	0.5628 (0.7547)	0.2680 (0.8746)	0.6949 (0.4045)	0.7706 (0.3800)	13.1429 **(0.0014)**	26.3796 **(0.000)**
475/2000	1.0156 (0.3136)	0.1045 (0.7465)	0.0987 (0.7534)	0.1635 (0.6860)	1.1143 (0.5728)	0.2680 (0.8746)	0.0712 (0.7897)	0.1360 (0.7123)	4.1307 (0.1268)	27.9068 **(0.0000)**

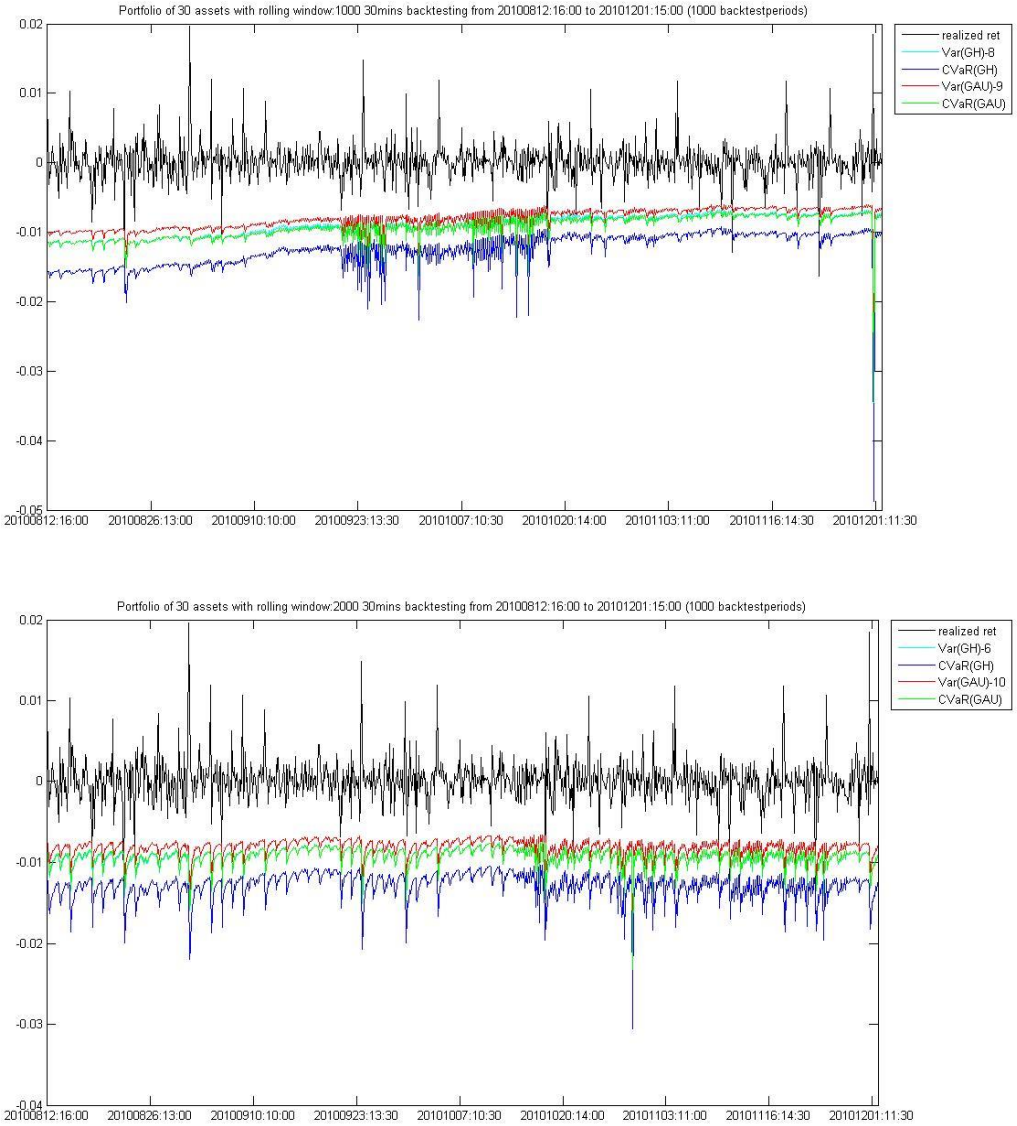


Figure 13: Portfolio of 30 equal weighted assets backtesting

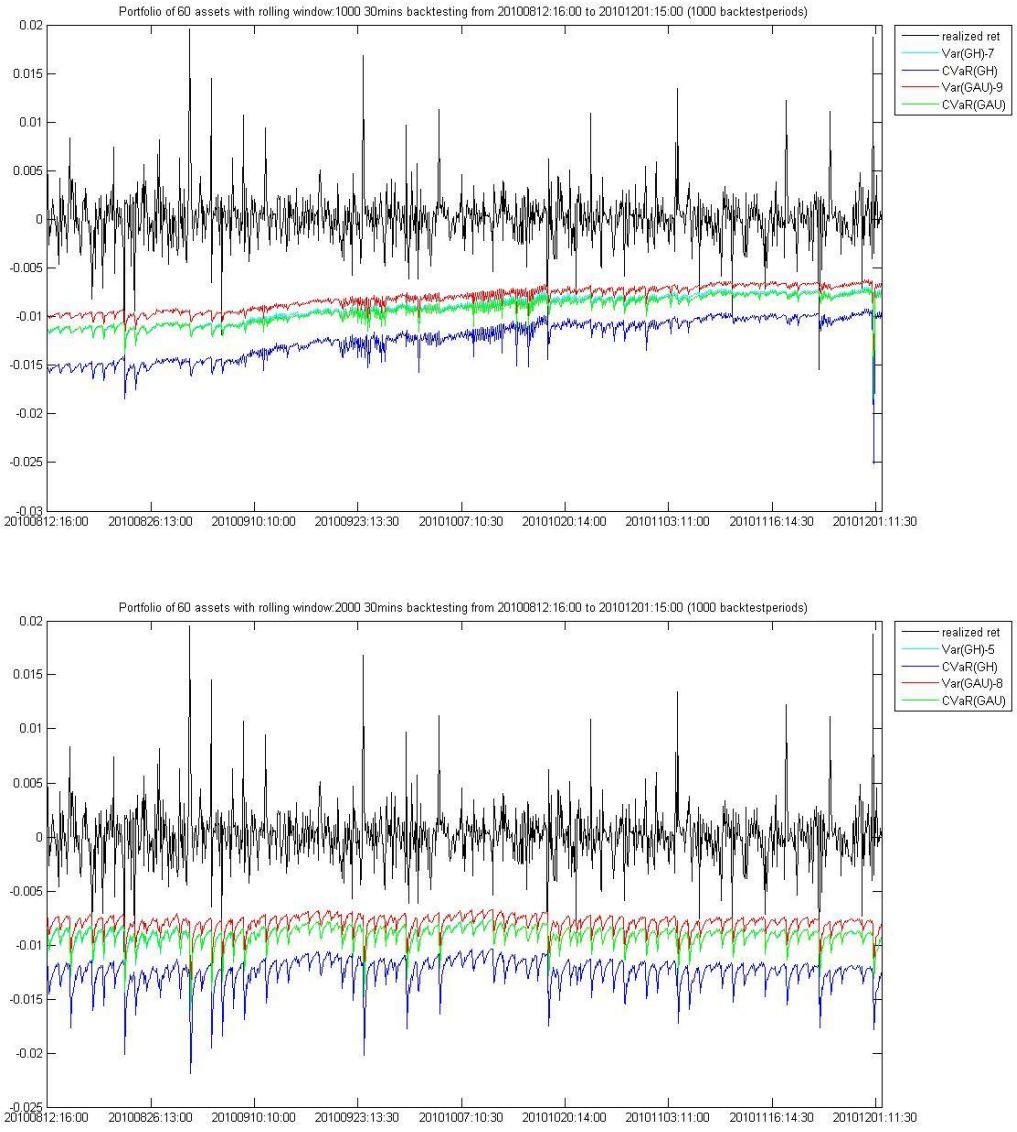


Figure 14: Portfolio of 60 equal weighted assets backtesting

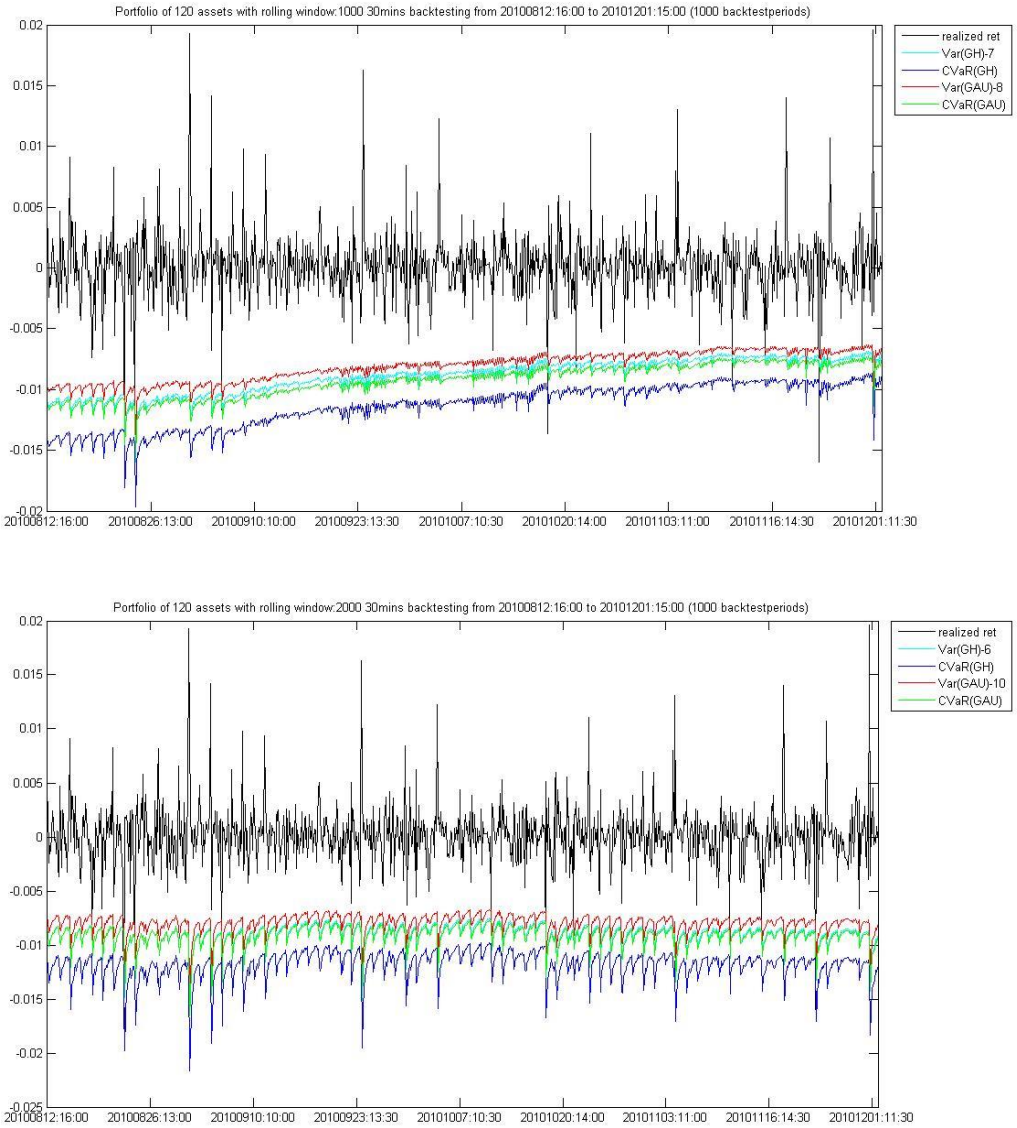


Figure 15: Portfolio of 120 equal weighted assets backtesting

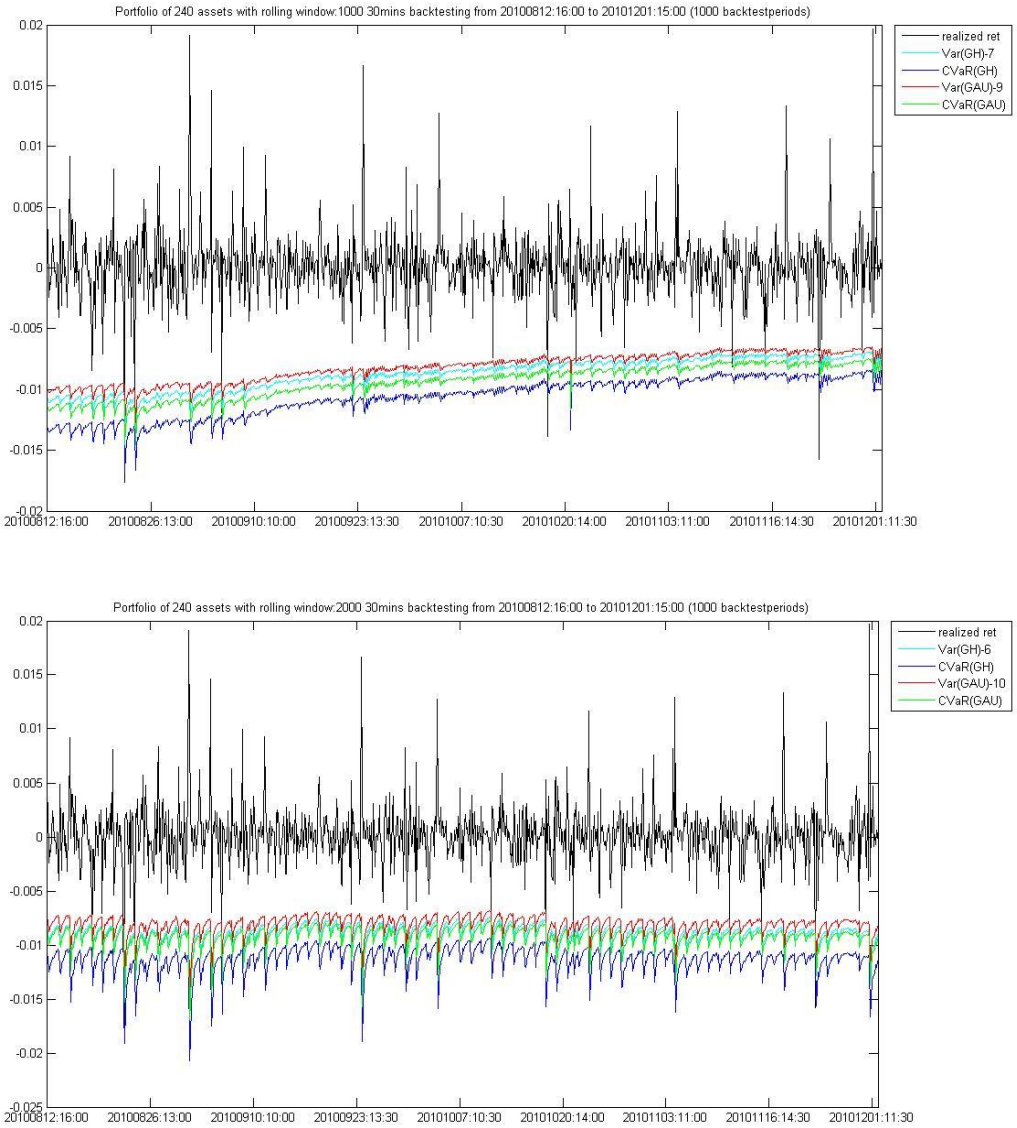


Figure 16: Portfolio of 240 equal weighted assets backtesting

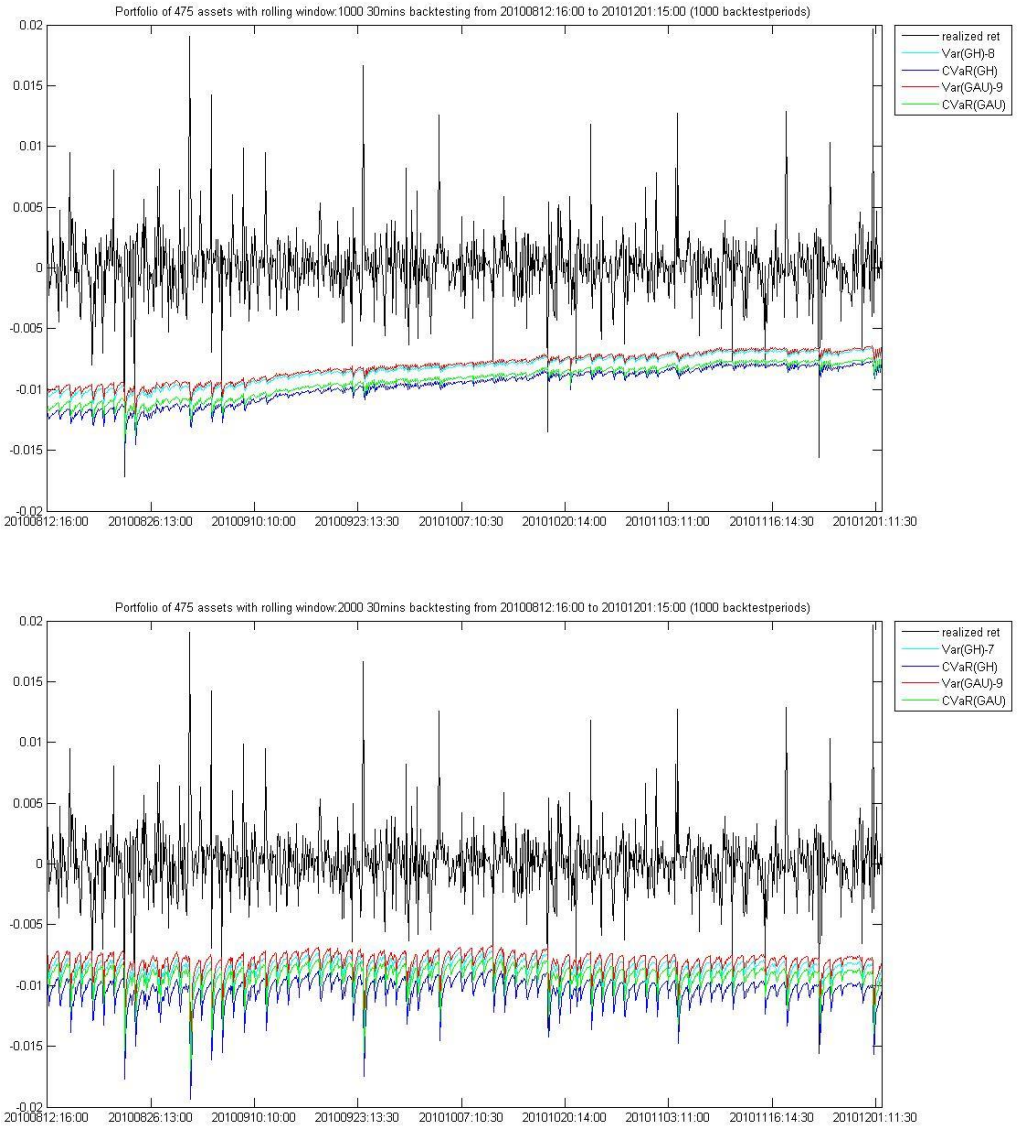


Figure 17: Portfolio of 475 equal weighted assets backtesting

We perform another backtesting of S&P 500 components (472 assets) for 5 minutes frequency data. The time span is from 20110208 to 20110228 for 1000 total backtesting periods. We also use PCA to reduce the dimension for multivariate GH and Gaussian model, and choose 200 as subspace dimension. All the model fitted with 1000 historical data. In table 6, we can see the violation number of Gaussian model in the yellow zone now while GH model still in the green zone. Gaussian still fails Berkowitz tail test.

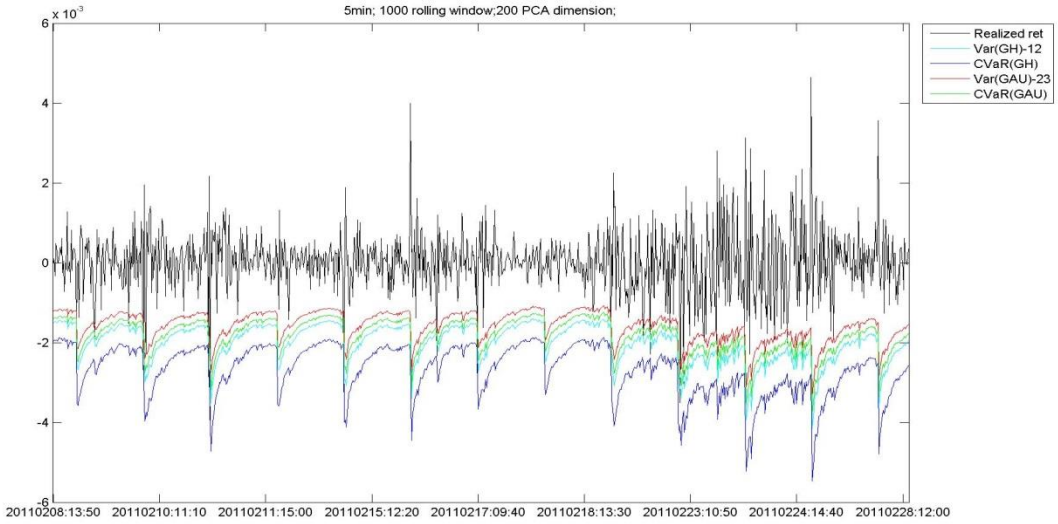


Figure 18: Bactseting of FARIMA-FIGARCH model for 5 min

Table 6: statistical test of FARIMA-FIGARCH for 5min

Statistics(p-value)	VaR(GH)	VaR(GAUSSIAN)
# Violation	12	23
Kupeic test	0.3798 (0.5377)	12.4853**(0.0004)**
Independent test	0.2915 (0.5893)	0.3517(0.5532)
Christoffersen test	0.6713 (0.7149)	12.8369 **(0.0016)**
Berkowitz tail test	0.7293(0.6944)	92.7608(**0.0000**)
Berkowitz independent test	0.5192(0.4712)	0.3366(0.5618)

For comparison purpose, we also perform the backtesting for ARMA-GARCH model. All other parameter settings is the same as backtesting above. From statistical result in table 7, we can see that violation number of Gaussian in the yellow zone and worse than the FARIMA-FIGARCH case. This indicates that FARIMA-FIGARCH model is better than ARMA-GARCH for more volatile market, i.e. higher frequency intra-day markets.

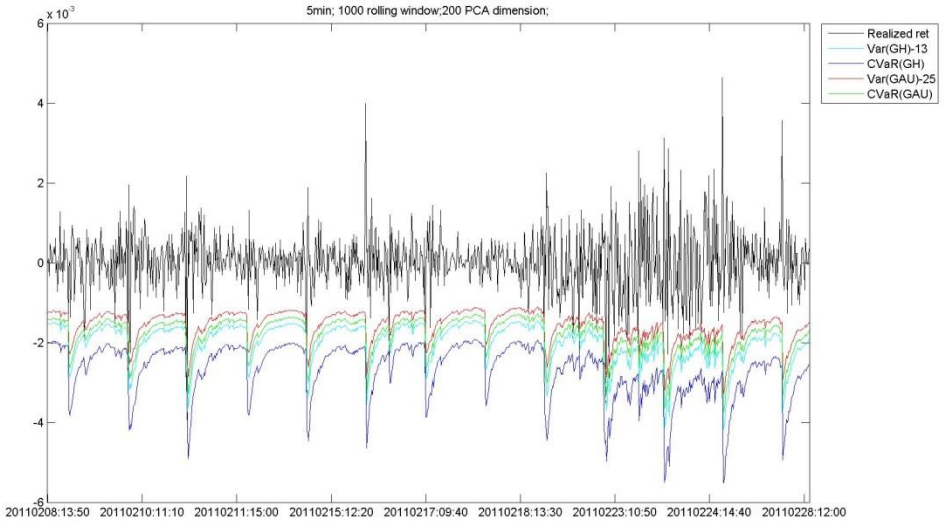


Figure 19: Backtesting of ARMA-GARCH for 5min

Table 7: statistical test of ARMA-GARCH for 5min

Statistics(p-value)	VaR(GH)	VaR(GAUSSIAN)
# Violation	13	25
Kupeic test	0.8306 (0.3621)	16.0430**(0.0001)**
Independent test	0.3425 (0.5584)	2.0627 (0.1509)
Christoffersen test	1.1730 (0.5563)	18.1057**(0.0001)**
Berkowitz tail test	0.8540 (0.6525)	90.9549(**0.0000**)
Berkowitz independent test	0.7766 (0.3782)	0.3145 (0.5749)

Based on all these out-sample backtesting result, we can clearly draw conclusion that FARIMA-FIGARCH with MGH model is consistently better than ARMA-GARCH with Gaussian model in the intra-day markets. Since our model has the capability to capture long range dependence and heavy tails feature together.

5.4 Computing Technology

In order to perform intra-day updates for risk measure calculation, we need high speed computation technology. Taking advantage of development of parallel computing in recent years, we implemented our model in C++ with parallel computing feature.

Based on trivial parallel scheme for the marginal model, it's easy to make it parallel. Suppose we have N assets, and M cores, $N \gg M$. Then each core performs certain amount assets calculation, and M of them happens in the same time. To assign the workload evenly among cores, we assign workload to a core either $[N/M]$ or $[N/M]+1$, where $[x]$ is the largest integer smaller than x.

The parallel algorithm is implemented by Message Passing Interface (MPI). We use a Dell Precision T7600 computer, which uses a dual node 8-core 3.1 Ghz (16 cores in total) with Intel Xeon E5-2687W chips. Only 8 cores are effectively available through use of MPI. The computer has 16GB memory. The computer also has two Nvidia GPU cards, one also can use GPU to get more speed up. Figure 6 shows similar computer architecture, and we can see that multiple cores communicated with each other through a shared memory block.

For the dependence estimation step, there are only matrix operations in the algorithm, therefore we use Openblas, high speed linear algebra open source library, to perform calculation. It can be found at http://www.openblas.net/. Openblas use shared memory threads to speed up calculation, in our case, there are two threads in each cores, and we have 16 threads totally. We report a timing study in table below, marginal model is FARIMA$(1, d_0, 1)$-FIGARCH$(1, d, 1)$ with stdGH distribution, dependence model is multivariate GH distribution, and window size is 1000 data points, 8 cores are used for parallelization.

Table 8: Timing table

Settings	500 assets	4000 assets
Marginal / cold start	110.08 sec	791.61 sec
Marginal / warm start	80.78 sec	590.10 sec
Dependence / cold start	2.75 sec	492.39 sec
Dependence / warm start	0.37 sec	49.88 sec

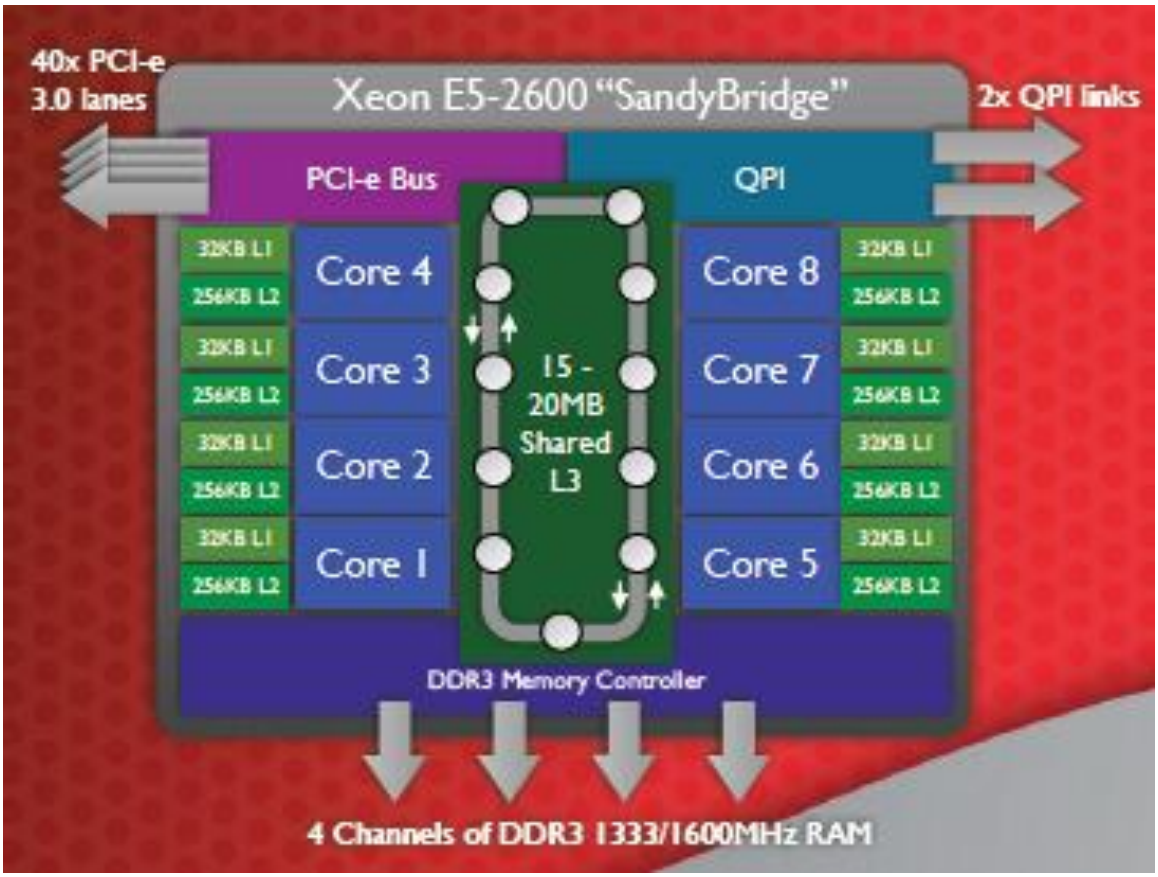


Figure 20: Computer parallel architecture

Chapter 6 Optimal Order Execution

Classical Portfolio theory worked out optimal asset allocation and optimal portfolio construction under market efficiency assumption. In reality, however, liquidation of a portfolio position does not come for free. Transaction cost became the key ingredient for implementing an investment decision. There are many kinds of transaction cost in the real world trading environment. Some are observable directly: brokerage commissions, fees (e.g. clearing and settlement costs) and taxes. When there is an order need to be executed, a trader also consider market impact, timing risk, and opportunity cost, all these components are stochastic variables and very much depend on the execution algorithm. Especially, when an institutional investor needs to liquidate a large position, his main concern is the market impact caused by his trades. Potential poor execution can erode portfolio performance substantially.

In high frequency trading context, market structure is more volatile and sensitive to "big" order. As a trader, one always wants to minimize market impact from his trades, therefore he can come back to exploit profit again and again. An obvious way to avoid market impact is simply to trade slowly, and then market can recover between trades. However, slower trading raise the fear of market volatility, and price may move towards hurting your trade. Therefore, there is always a tradeoff between the market impact costs of rapid execution and the volatility risk of slow execution. An optimal execution scheduling model tends to deal with this problem and finds the optimal way in terms of trading strategy.

6.1 Optimal execution problem

In this section, we will review classical optimal execution problem and introduce the some definitions and terminologies in this context. We also review Almgren R. and Chriss N. (2000) frame work.

Suppose we hold a block of X units of a security that needs to be completely liquidated before time T. We divide T into N intervals of length $\tau = T/N$, and define the discrete times $t_k = k\tau$,

for $k = 0, \dots, N$. A trading trajectory is a list $x_0, \dots, x_N$, where x_k is the number of units that we plan to hold at time t_k. Apparently, our initial holding is $x_0 = X$, and $x_N = 0$.

Equivalently, we can define a trading schedule as N dimensional vector $v = (v_1, v_2, \dots, v_N)$, where $v_k = x_{k-1} - x_k$ is the number of units that we will sell between t_{k-1} and t_k. From the definition above, we have the relationship between x_k and v_k as follow:

$$X = \sum_{j=1}^{N} v_j$$

$$x_k = X - \sum_{j=1}^{k} v_j = \sum_{j=k+1}^{N} v_j \,, \qquad k = 0, \dots, N \tag{6.1}$$

Note that we are not imposing that the sign of each v_k is the same as the sign of X. For example, if we consider a sell program, $X > 0$, then we could have some $v_k < 0$, which means buy transaction. For simplicity, we consider a sell program $X > 0$.

Each trading schedule has an execution cost. Suppose that the initial security price is P_0 and P_k is the price at the start of interval k. Note that the actual execution price for v_k not exactly is P_k. In fact, commission fees and, bid-ask spread raise the price for buy and lower it for sell. Therefore, we define $\tilde{P}_k$ as the effective price at which shares are actually traded at every step. The relationship between P_k and $\tilde{P}_k$ will be specified through market impact function, we will discuss this later.

Now we define the execution cost $C(v)$ as the difference between the total money we pay or receive during the execution and the initial market value of the position. Express it as formula:

$$C(v) = \sum_{k=1}^{N} v_k \tilde{P}_k - XP_0 = \sum_{k=1}^{N} v_k (\tilde{P}_k - P_0) \tag{6.2}$$

It is important to notice that $C(v)$ is a stochastic variable since $\tilde{P}_k$ is a stochastic process. Then it is reasonable to define its expected value $E[C(v)]$ and variance $Var[C(v)]$.

6.2 The Almgren Chriss execution scheme

Almgren and Chriss (2001) propose an optimal execution model which assume that the price of the stock at step k is equal to the previous price plus a linear market impact term and a random shock :

$$P_k = P_{k-1} + \theta v_k + \eta_k \qquad \text{where } \eta \sim i.i.d(0, \sigma) \tag{6.3}$$

They also consider the effective price $\tilde{P}_k$ and model it as a temporary impact function:

$$\tilde{P}_k = P_k + \rho v_k + sign(v_k) * \frac{S}{2} \tag{6.4}$$

where the term $sign(v_k) * S/2$ is due to the bid ask spread S and ρv_k represents a linear temporary impact.

We will get execution cost $C(v)$ by substitute (6.3) and (6.4) into (6.2) :

$$C(v) = \sum_{k=1}^{N} v_k \tilde{P}_k - XP_0 = \sum_{k=1}^{N} (\eta_k + \theta v_k) \sum_{j=k+1}^{N} v_j + \sum_{k=1}^{N} (\frac{sign(v_k)S}{2} + \rho v_k) v_k \quad (6.5)$$

and the expected execution cost is :

$$E[C(v)] = \frac{\theta}{2} X^2 + \left(\rho - \frac{\theta}{2}\right) \sum_{k=1}^{N} v_k^2 + \frac{S}{2} \sum_{k=1}^{N} |v_k| \quad (6.6)$$

The variance of the execution cost is:

$$Var[C(v)] = E[(C(v) - E[C(v)])^2] = E\left[\left(\sum_{k=1}^{N} \eta_k \sum_{j=k}^{N} v_j\right)^2\right] \quad (6.7)$$

Since we assume η_k are independent, so we have $E[\eta_i \eta_j] = 0$, then :

$$Var[C(v)] = \sigma^2 \sum_{k=1}^{N} \left(\sum_{j=k}^{N} v_j\right)^2 \quad (6.8)$$

Similar to portfolio theory, Almgren and Chriss consider the following optimal scheduling solution:

$$v^* = \arg\min_v (E[C(v)] + \lambda \, Var[C(v)]) \quad (6.9)$$

where λ is the coefficient of risk aversion. The set of solutions to this problem for different values of λ is called optimal frontier.

One can get the closed form solution for problem (6.9) :

$$v_k = A \cosh(\beta \, (T - t_k)) \quad (6.10)$$

where A is a normalization constant. The coefficient β is given by :

$$\beta = \sqrt{\frac{\lambda \sigma^2}{\rho}} \quad , \lambda > 0 \quad (6.11)$$

6.3 Extension of Almgren Chriss model

From Almgren Chriss model, we can see that opportunity cost $C(v)$ is determined by market price process (6.3) and temporal impact function (6.4). We generalize model (6.3) to include heavy-tails, volatility clustering and long range dependence. We assume the log-returns:

$$r_k = P_k - P_{k-1} = \theta v_k + \eta_k$$

where r_k follows FARIMA-FIGARCH process with Generalized Hyperbolic distribution.

Suppose we consider a portfolio with N assets, then we have asset log-return:

$$r(t) = \big(r_1(t), \dots, r_N(t)\big) \quad , \quad t = 1, \dots, T \tag{6.12}$$

where T is the time we need to liquidate position for each asset.

In our model, we assume $r(t)$ follows FARIMA-FIGARCH process with dependent innovations with multivariate generalized hyperbolic distribution with dimension N. For each time period t, we can fit model for $r(t)$, and forecast the log-prices

$$S(s) = \big(S_1(s), \dots, S_N(s)\big)' \ , for\ s > t \tag{6.13}$$

Now we have trading schedule matrix for portfolio with N assets:

$$V = (v(1), \dots, v(T))' \tag{6.14}$$

$$v(t) = (v_1(t), \dots, v_N(t))$$

$$v_i(t) = x_i(t-1) - x_i(t)$$

From definition above, we can see that dimension of V is $T \times N$. The vector of actual execution log-prices at time t is

$$Se(t) = S(t) - \theta \cdot v(t) \tag{6.15}$$

where $\theta = (\theta_1, \dots, \theta_N)$ is the vector of coefficients of market impact for each asset and $\cdot$ denotes the element by element product of two vector.

And the vector of actual execution prices at time t is :

$$\begin{aligned} Pe(t) &= (Pe_1(t), \dots, Pe_N(t)) \\ &= \exp(Se(t)) \\ &= (\exp\big(Se_1(t)\big), \dots, \exp(Se_N(t))) \end{aligned} \tag{6.16}$$

Now we define the total execution cost of portfolio with N assets is :

$$C(V) = C(v_1) + \cdots + C(v_N) \tag{6.17}$$

where execution cost of each asset $C(v_i)$ is defined as formula (6.5) :

$$C(v_i) = \big(v_i(1)Pe_i(1) + \cdots + v_i(T)Pe_i(T)\big) - X_i P_i(0) \tag{6.18}$$

We will use Monte Carlo to simulate M-trajectories (M=10000) for N-dimensional vectors of log-price (t) :

$$S(t,m) = (S_1(t,m), \dots, S_N(t,m),\ m = 1, \dots, M; t = 1, \dots, T. \tag{6.19}$$

Correspondingly, we have scenarios for execution log-price for each asset:

$$Se_i(t,m) = S_i(t,m) - \theta_i v_i(t),\ i = 1, \dots, N \tag{6.20}$$

And

$$Pe(t,m) = (Pe_1(t,m), \dots, Pe_N(t,m)) \tag{6.21}$$

where $Pe_i(t,m) = \exp(Se_i(t,m))$.

Finally, we have scenarios for portfolio execution cost $C(V)$ for a given trading schedule $C(V,m), m = 1, \dots, M$.

Since the random variable $C(V)$ is not Gaussian and symmetric, we also extend Markowitz Mean-variance approach (formula (6.9)) to a min-CVaR framework :

$$V^* = \arg\min(E[C(V)] + \lambda\, CVaR[C(V), \alpha]) \tag{6.22}$$

where λ is a constant penalize only big positive deviation of the execution cost, $C(V)$.

6.4 Calibration of market impact parameter

As we show our model in the last section, the only parameter we need to calibrate is the temporal market impact parameter θ. This parameter tells us how much impact to the price by transaction volume.

Suppose we look at the series of times t_n as separated by a constant time interval. And define the series of aggregated volumes v_n in terms of the volumes v^{tt} of the single transactions:

$$v_n = \sum_{[t_n, t_{n+1}]} v_i^{tt} \tag{6.23}$$

where v_i^{tt} is the singed volumes are traded between time t_n and t_{n+1}. The quantity v_n is the volume imbalance over the n-th time interval. This quantity may be very big in absolute value during time intervals of high market activity. Therefore, we use the normalized version, v_n^{nor}, defined as :

$$v_n^{nor} = \frac{\sum_{[t_n, t_{n+1}]} v_i^{tt}}{\sum_{[t_n, t_{n+1}]} |v_i^{tt}|} \tag{6.24}$$

When there is no trade in the n-th interval, we define $v_n^{nor} = 0$. All this definition follows F. Lillo and J.D. Farmer (2004).

We need to calibrate the impact function $f(v^{nor})$ of the normalized volume imbalance. It is defined as the expected return in a constant interval conditioned to a normalized volume imbalance v_n^{nor} in the same interval :

$$f(v^{nor}) = E[r_n | v_n^{nor}] \tag{6.25}$$

we use a linear function to fit this impact function, say,

$$f(v^{nor}) = \theta v^{nor} \tag{6.26}$$

Where θ is positive for a sell program.

6.5 Numerical example

In this section, we show a numerical example for the procedure we proposed in previous section. We extract all the trading and quoting data for AAPL and AMZN from TAQ database for one month, 20090729 to 20090831. And we aggregate price to 5 minutes interval log return data. Graph below shows the price and scatter plot of log return data.

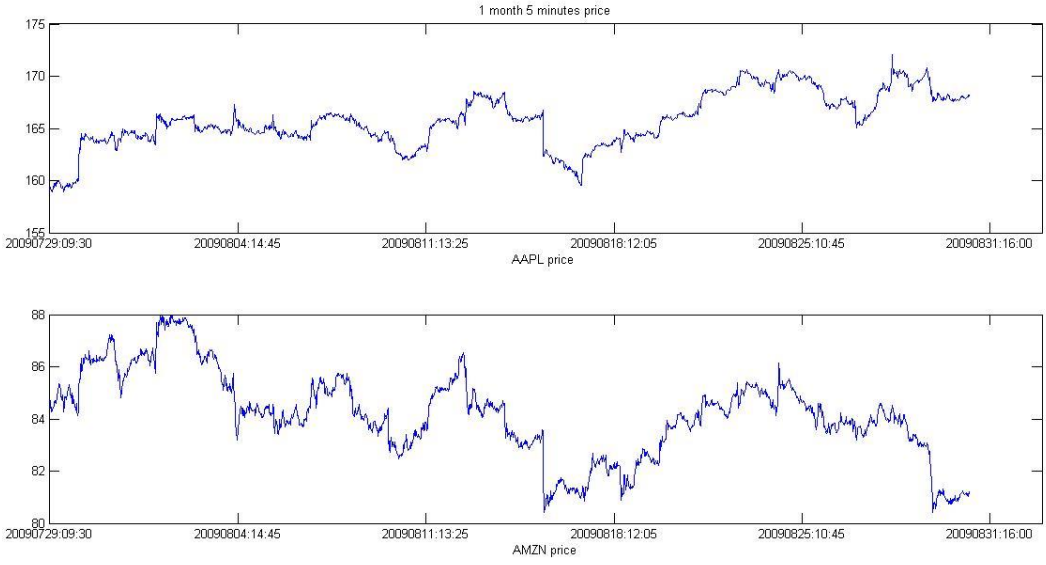


Figure 21: 5 min price chart for AAPL and AMZN from 20090729 to 20090831

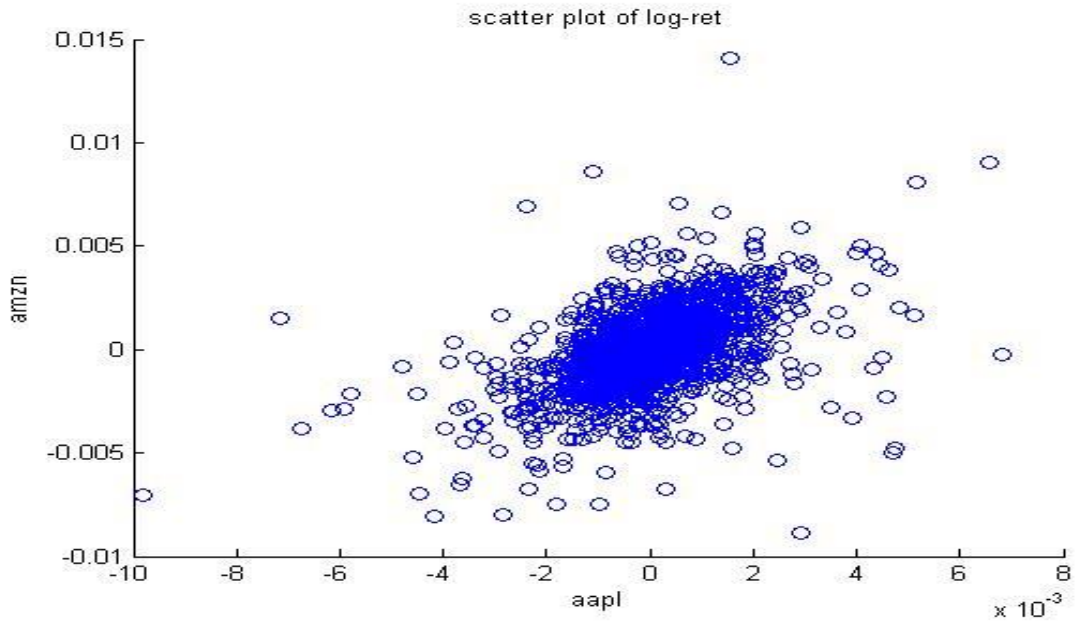


Figure 22: scatter plot of log return between AAPL and AMZN

We also need to calibrate the market impact parameter θ in formula (6.26). In order get the normalized trading volume, we are required to classify the trading volume for every transaction in 5 min interval. We are using Lee and Ready (1991) algorithm to infer the trading direction. This algorithm requires comparing the trade price with the prevailing quote so that identify the trade as buyer or seller initiated. Prevailing quote is defined as the quote immediately before the trade. There is no rules to measure what is the "immediately", some researchers using 2 seconds before trade or 5 seconds before trades. We are using 5 seconds before trade as our prevailing quotes. Graph below shows the relationship between return and normalized volume for 5 minutes interval.

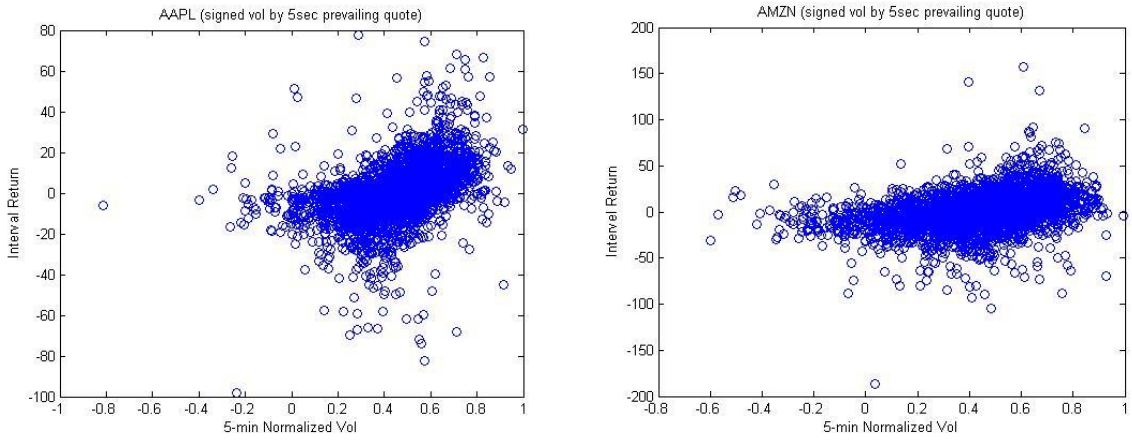


Figure 23: 5 min normalized volume and return

By running simple linear regression, we calibrate their market impact parameter: 26.98 bp for AAPL and 30.87 bp for AMZN. Then we fit data (1843 observations) to FARIMA-FIGARCH and multivariate GH model, we present the estimated parameters and the heteroskedasticity volatility and filtered innovations below:

Table 9: Estimated parameters of FARIMA-FIGARCH model for AAPL and AMZN

Parameter	μ	AR	MA	d	ω	GARCH	ARCH	d_0
AAPL	6.5121e-006	0.6575	-0.7130	0.0220	1.8956e-007	0.0636	5.5508e-006	0.3405
AMZN	-1.8646e-005	0.2628	-0.1734	-0.0933	6.8028e-007	0.1734	0.0867	0.2495

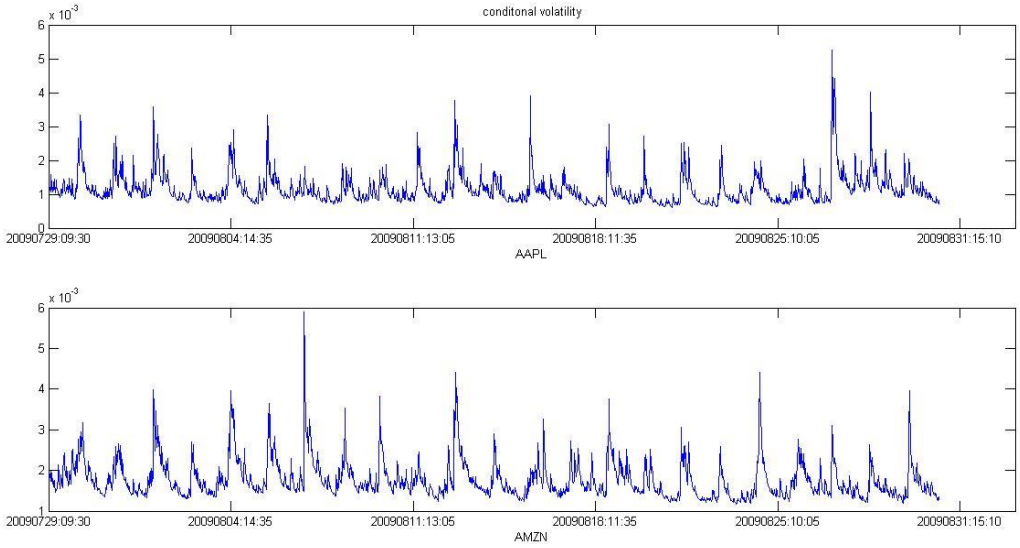


Figure 24: Conditional volatility

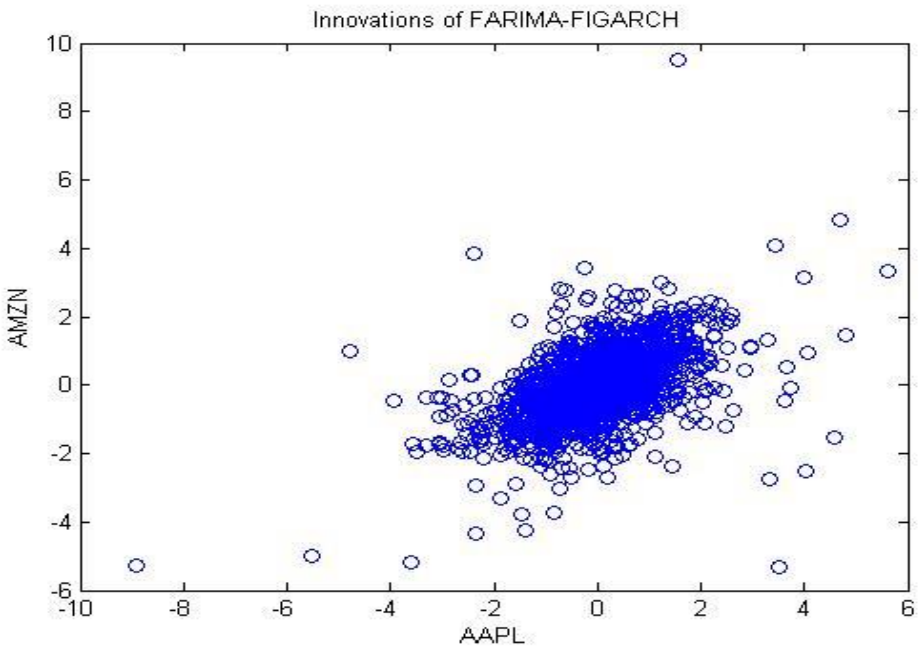


Figure 25: Filtered innovations

And GH parameters: $\lambda = -2.2952$, $\chi = 2.5906, \psi = 3.9004e-005, \mu = (-0.0365, 0.0396)'$ $\gamma = (0.0362, -0.0359)', \Sigma = (1.0779, 0.5395; 0.5395, 1.0104)$.

Suppose we need to sell some shares of AAPL and AMZN in the next 50 minutes. Then we can generate 10-step scenarios of log returns based on fitted FARIMA-FIGARCH and multivariate GH model. And we can have scenarios of transaction cost $C(V)$ based on formula (6.15)-(6.18) for some particular scheduling V. Then we can find optimal schedule by solving optimization problem in (6.22). When we vary the value of λ , we can construct the efficient frontier of transaction cost, as shown below. Each point on this frontier corresponding to a particular trading schedule, i.e. trading strategy. Trader would move along with this frontier based on their risk profile. The execution risk getting bigger when move to the upper right corner which can be consider as a risky trading strategy, while the bottom left corner represents a risk conservative trading strategy. In figure below, we plot the efficient frontier for our model with 10000 scenarios of log returns, and also random pick three point on the frontier with their corresponding trading scheduling. We can see that there is more than 50% volume happen in the first step for the blue point, which means risk conservative trading strategy, since traders with great uncertainty of the market dynamics in the future. When risk profile of trading getting higher, one move along to the upper right corner of the efficient frontier, more trading happen in afterward of the trading schedule, see green point.

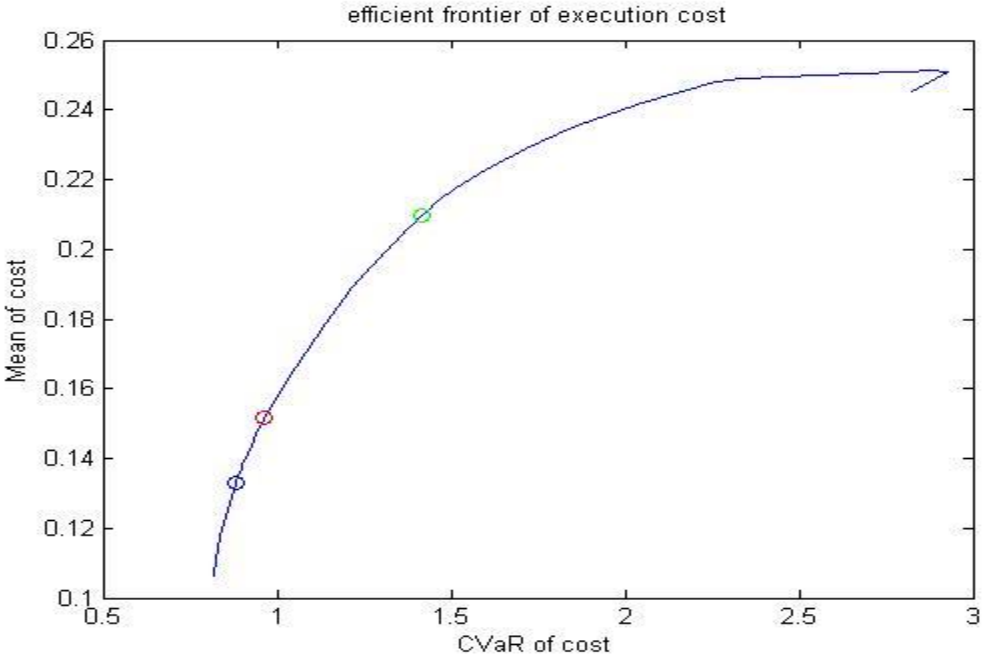
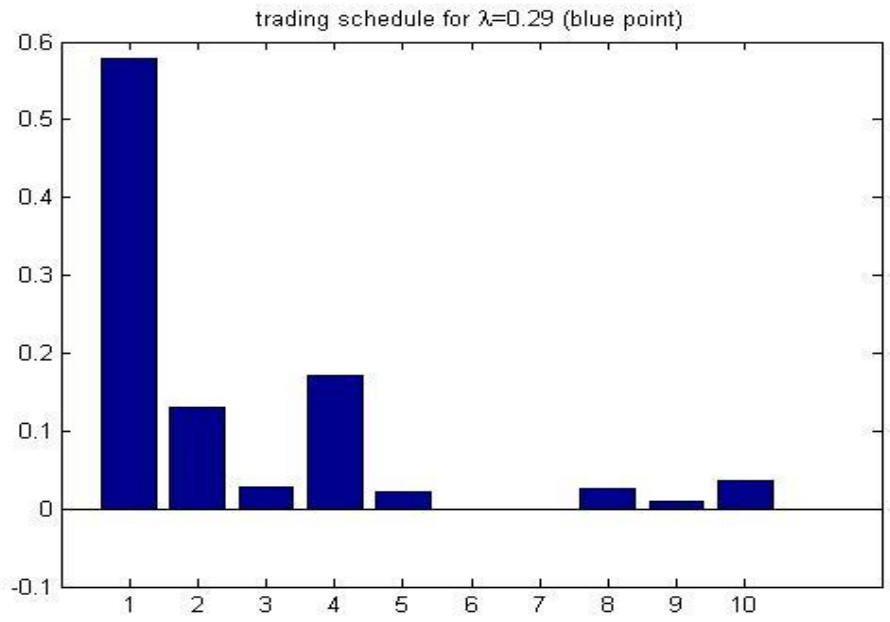
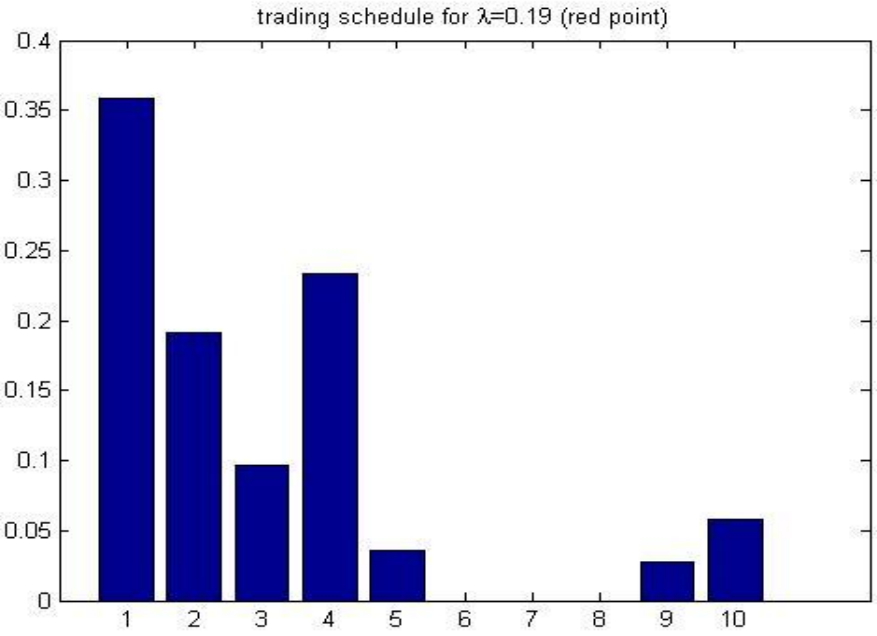


Figure 26: Efficient frontier of execution cost


trading schedule for λ=0.29 (blue point)
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
1
2
3
4
5
6
7
8
9
10


trading schedule for λ=0.19 (red point)
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
1
2
3
4
5
6
7
8
9
10

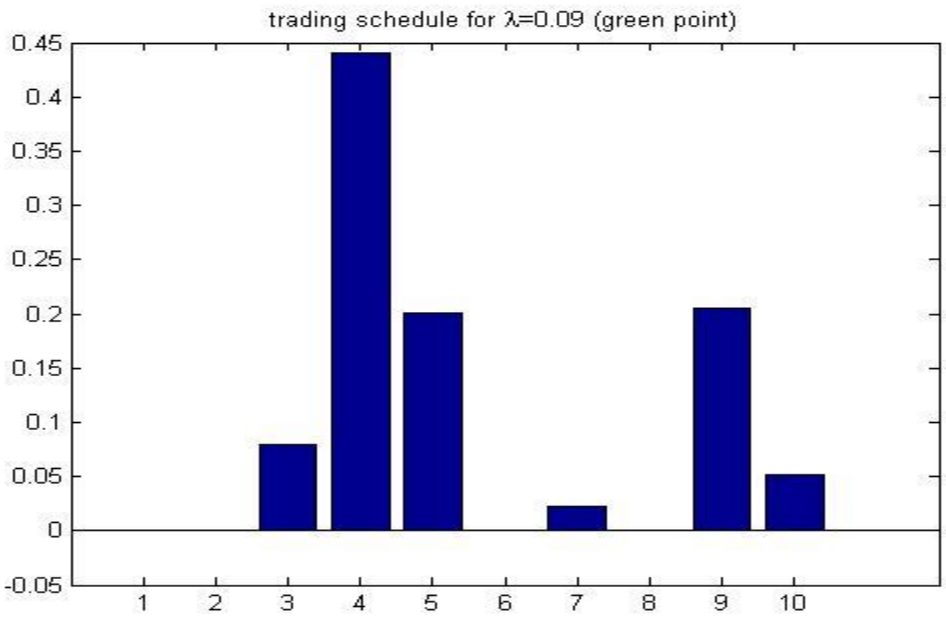


Figure 27: Trading schedule for AAPL with different λ

References

1. R.CONT, *Empirical properties of asset returns: Stylized facts and statistical issue*, Quant. Finance, 1(2001), pp. 1-14.

2. Rachev, S.T., Kim, Y.S., Bianchi, M.L., and Fabozzi, F.J. (2010), 'Time series analysis for financial market meltdowns', *Journal of Banking & Finance*, 35.

3. B.B. Mandelbrot, The variation of certain speculative prices, Journal of Business, XXXVI (1963), pp. 392-417.

4. McLeod, A. & Hipel, K.W. (1978), Preservation of the rescaled adjusted range reassessment of the hurst phenomenon, *Water Resources Research* 14(3), 491-508.

5. Geweke, J. & Hansen, H. (1983), The estimation and application of long memory time series models, *Journal of Time Series Analysis* 4(4), 221-238.

6. Robinson, P. (1995), *Log-periodogram regression of time series with long range dependence*, The Annals of Statistics 23(3), 1048-1072.

7. Hosking, J.R.M. (1981), *Fractional differencing* , Biometrika 68(1), 165-176

8. Lo, A.W. (1991), long term memory in stock market price, Econometricca 59(5), 1297-1313.

9. Granger, C.J. (1980), *long memory relationship and the aggregation of dynamic models*, Journal of Econometrics 14, 227-238.

10. Granger, C.J. & Joyeux, R. (1980), *An introduction to long memory time series models and fractional differencing*, Journal of Time Series Analysis 1(1), 15-29.

11. Jenkins, G. M., and D. G. Watts, *Spectral Analysis and Its Applications*, Holden-Day, San Francisco, 1968.

12. A. Lo, *Long term memory in stock market prices*, Econometrica, 59 (1991), pp. 1279-1313.

13. J. Beran, *Statistics for long memory processes*, vol. 61 of Monographs on Statistics and Applied Probability, Chapman and Hall, New York, 1994.

14. Baillie, R., Bollerslev, T. and Mikkelsen, H. (1996), *Fractionally integrated generalized autoregressive conditional heteroskedasticity*, Journal of Econometrics 74, 3-30.

15. Diebold, F. and Inoue, A. (1999), Long memory and structural change.

16. Geweke, J. and Porter-Hadak, D. (1983), The estimation and application of long memory time series models, Journals of Time Series Analysis 4(4), 221-238.

17. Granger, C.J and Joyeux, R (1980), *An introduction to long memory time series models and fractional differencing*, Journal of Time Series Ananlysis 1(1), 15-29.

18. Granger, C. (1980), *Long memory relationships and the aggregation of dynamic models*, Journal of Econometrics 14, 227-238.

19. Clifford M. Hurvich, Rohit Deo and Julia Brodsky, *The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long memory time series*, Journal of Time Series Analysis, Vol 19, No.1, 19-46.

20. Ching-Fat Chung (1999), *Estimating the Fractionally Integrated GARCH model*, Institute of Economics, Academia Sinica.

21. Peter Lildholdt (2000), *Long memory and ARFIMA modelling*, Department of Economics, University of Aarhus and Centre for Dynamic Modelling in Economics.

22. Wolfgang H and Josef L., *Generating generalized inverse Gaussian random variates*, Institute for statistics and Mathematics, Wirtschafts Uinversity.

23. Milton Abramowitz and Irene A. Stegun. *Handbook of mathematical functions*. Dover books on mathematics. Dover, New York, 9. Dover printing edition, 1972.

24. O.E. Barndorff-Nielsen. *Exponentially decreasing distributions for the logarithm of particle size*. Proceedings of the Royal Society of London. Series A, Mathematical and Physical Science, 353(1674):401-419,1977.

25. O.E. Barndorff-Nielsen. And O. Halgreen. Infinite divisibility of the hyperbolic and generalized inverse Gaussian distribution. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 38(4):309-311, 1977.

26. Rama Cont and Peter Tankov. *Financial Modeling with Jump Process*. Chapman & Hall/CRC Financial Mathematics Series, 2003.

27. Alexander J. McNeil, Rüdiger Frey and Paul Embrechts. *Quantitative risk management: Concepts, techniques and tools*. Princetion University Press, Princeton, 2005.

28. Karsten Prause. The Generalized Hyperbolic Model: Estimation Financial Derivatives and Risk Measures. PhD thesis, Universität Freiburg i. Br., 1999.

29. Peter F. Christoffersen. Evaluating Interval Forecasts, International Economic Review, Vol. 39, No. 4, Symposium on Forecasting and Empirical Methods in Macroeconomics and Finance (Nov.,1998),841-862.

30. Jeremy Berkowitz. *Testing Density Forecasts, with application to risk management*. Journal of Business & Economic Statistics, Oct.,2001, Vol. 19, No.4.

31. R. Almgren and N. Chriss. *Optimal execution of portfolio transactions*, *Journal Risk*, 3:5-40, 2001.

32. H.M. Markowitz. *Portfolio selection: Efficien diversification of investments*. Yale University Press, New Haven, CT,1968.

33. E. Busseti and F. Lillo. *Calibration of optimal execution of financial transactions in the presence of transient market impact*. arXiv:1206.0682v1[q-fin.TR] 4 Jun.,2012.

34. Admai, A. and P. Pfleiderer. *A theory of intraday patterns : Volume and price variability. Rev. Financial Studies* 1,3-40.

35. J.M. Lorenz. *Optimal trading algorithms: Portfolio transactions, multiperiod portfolio selection, and competitive online search*. Ph.D thesis, ETH Zürich, 2008.

36. J.P. Bouchoud, Y. Gefen, M.Potters, and M. Wyart. *Fluctuations and response in financial markets: the subtle nature of 'random' price changes*. Quantitative Finance Vol.4,176-190, 2004.

37. M.C. Charles and Mark J. Ready. Inferring trade direction form intraday data. Journal of finance, Vol. 46, No. 2, 733-746, 1991.

38. B. Chakrabarty, P.C. Moulton and A. Shkilko. Short Sales, Long Sales, and the Lee-Ready Trade Classification Algorithm Revisited. *Journal of financial markets*,2012.

39. F. Lillo and J.D. Farmer. The long memory of the efficient market. *Stud. Nonliear Dyn.* E., 8(3):1-31, 2004.

Appendix

1. Spectrum of fractionally integrated process

x_t is fractional integrated process, then by definition

$$x_t = (1-L)^{-d}\epsilon_t \quad -0.5 < d < 0.5 \text{ for stationary}$$

If two stationary process are linear related, say,

$$x_t = \phi(L)\epsilon_t$$

Then

$$\sigma_x^2 f_x(\omega) = \sigma_\epsilon^2 \left|\phi\left(e^{-i\omega}\right)\right|^2 f_\epsilon(\omega) \quad \text{for } -\pi \le \omega \le \pi$$

Therefore, we have

$$\begin{aligned}
\sigma_x^2 f_x(\omega) &= \sigma_\epsilon^2 \left|\left(1-e^{-i\omega}\right)^{-d}\right|^2 f_\epsilon(\omega) \\
&= \sigma_\epsilon^2 f_\epsilon(\omega)\left(1-e^{-i\omega}\right)^{-d}\left(1-e^{i\omega}\right)^{-d} \\
&= \sigma_\epsilon^2 f_\epsilon(\omega)(2-2\cos(\omega))^{-d} \qquad \text{since } e^{i\omega} = \cos(\omega) + i\sin(\omega) \\
&= \sigma_\epsilon^2 f_\epsilon(\omega)\left(4\sin^2\left(\frac{\omega}{2}\right)\right)^{-d} \\
&= \sigma_\epsilon^2 f_\epsilon(\omega)\left(2\sin\left(\frac{\omega}{2}\right)\right)^{-2d}
\end{aligned}$$

If Assume ϵ_t is normal , then $f_\epsilon(\omega) = 1$ and spectrum of x_t is

$$h_x(\omega) = \sigma_\epsilon^2\left(2\sin\left(\frac{\omega}{2}\right)\right)^{-2d}$$

2. Conditional density of $W|X$

Suppose $X \sim GH(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$ and W is the mixing GIG variable.
Then we have $X|W = w \sim N_d(\mu + w\,\gamma, w\,\Sigma)$

$$\begin{aligned}
f_{w|x}(w) &= \frac{f_{X|W}(x)\, f_W(w)}{f_X(x)} \\
&= \frac{\left(\frac{\gamma'\Sigma^{-1}\gamma+\psi}{Q(x)+\chi}\right)^{0.5(\lambda-d/2)}\left(w^{\lambda-d/2-1}\exp\left\{-\frac{1}{2}\left(\frac{Q(x)+\chi}{w}+w(\gamma'\Sigma^{-1}\gamma+\psi)\right)\right\}\right)}{2\,K_{\lambda-d/2}\left(\sqrt{(Q(x)+\chi)(\gamma'\Sigma^{-1}\gamma+\psi)}\right)}
\end{aligned}$$

Compare this to GIG density(4.4), we can see this is still a GIG variable:

$$W|X \sim GIG(\lambda - d/2, Q(x) + \chi\,, \psi + \gamma'\Sigma^{-1}\gamma)$$

3. Derivation of update equation for $(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\gamma})$

In order to get update equation, we need to maximize first part of (4.24).

$$\begin{aligned} L &= E\left[\sum_{i=1}^{n} L_1 | x_1, .., x_n; \Theta^{[k]}\right] \\ &= const - \frac{1}{2}E[\sum_{i=1}^{n} w_i] - \frac{n}{2}\ln|\Sigma| \\ &\quad -\frac{1}{2}E[\,\sum_{i=1}^{n}(x_i - \mu - w_i\gamma)' w_i^{-1}\Sigma^{-1}(x_i - \mu - w_i\gamma)] \end{aligned}$$

Take the partial derivative with respect to each parameter, and set them to zero:

For μ :

$$\frac{\partial L}{\partial \mu} = 0$$

$$\Rightarrow E[\,\sum_{i=1}^{n} w_i^{-1}\,\Sigma^{-1}(x_i - \mu - w_i\gamma)] = 0$$

$$\Rightarrow \hat{\mu} = \frac{E[\sum_{i=1}^{n}\,(w_i^{-1}x_i - \gamma)\,]}{E[\,\sum_{i=1}^{n}\,w_i^{-1}\,]}$$

$$\Rightarrow \hat{\mu} = \frac{\sum_{i=1}^{n}(\delta_i\, x_i - \gamma)}{\sum_{i=1}^{n}\,\delta_i} = \frac{\sum_{i=1}^{n}(\delta_i\, x_i - \gamma)}{n\,\bar{\delta}} \qquad (7.1)$$

For :

$$\frac{\partial L}{\partial \gamma} = 0$$

$$\Rightarrow E[\sum_{i=1}^{n} w_i^{-1}\Sigma^{-1}(x_i - \mu - w_i\gamma)w_i] = 0$$

$$\Rightarrow \sum_{i=1}^{n} x_i - n\,\mu - \gamma\, E[\sum_{i=1}^{n} w_i] = 0 \quad (7.2)$$

$$\Rightarrow \bar{x} - \mu - \gamma\bar{\eta} = 0$$

$$\Rightarrow \hat{\gamma} = \frac{1}{n}\,\frac{\sum_{i=1}^{n}\delta_i(\bar{x} - x_i)}{\bar{\eta}\,\bar{\delta} - 1} \quad \text{by plug in (7.1)}$$

For Σ :

$\frac{\partial L}{\partial \Sigma} = 0$

Since $\frac{\partial}{\partial \Sigma} \ln|\Sigma| = \Sigma^{-1}$; $a = tr(a)$ for scale ; $tr(AB) = tr(BA)$

$\Rightarrow n\,\Sigma^{-1} + E\left[\frac{\partial}{\partial \Sigma}\sum_{i=1}^{n} tr(\,(x_i - \mu - w_i\gamma)(x_i - \mu - w_i\gamma)' w_i^{-1}\Sigma^{-1})\right]$

$\Rightarrow n\,\Sigma^{-1} - E[\sum_{i=1}^{n}(x_i - \mu - w_i\gamma)(x_i - \mu - w_i\gamma)' w_i^{-1}\,\Sigma^{-1}\Sigma^{-1}] = 0$

$\Rightarrow n\,\Sigma = \sum_{i=1}^{n} \delta_i (x_i - \mu)(x_i - \mu)' - 2\sum_{i=1}^{n}(x_i - \mu)\gamma' + \sum_{i=1}^{n} \eta_i \gamma\gamma'$

By (7.2), we have $\sum_{i=1}^{n}(x_i - \mu)\gamma' = \sum_{i=1}^{n} \eta_i \gamma\gamma'$

$\Rightarrow \hat{\Sigma} = \sum_{i=1}^{n} \delta_i (x_i - \mu)(x_i - \mu)' - \sum_{i=1}^{n} \eta_i \gamma\gamma'$