

Forecasting Volatility in Stock Market Using GARCH Models

By

Xiaorong Yang

Submitted to the graduate degree program in Mathematics
and the Graduate Faculty of the University of Kansas
in partial fulfillment of the requirements for the degree of
Master of Arts.

Chairperson

Committee members*

*

*

*

Date defended: _____

The Thesis Committee for Xiaorong Yang certifies
That this is the approved Version of the following thesis:

**Forecasting Volatility in Stock Market
Using GARCH Models**

Committee:

Chairperson*

Date approved: _____

Acknowledgements

This master's thesis is based on my summer project (2008). So, first I would like to thank my advisor---Prof. Bozenna Pasik-Duncan for funding this project and providing the most valuable instructions.

Then I am grateful to Prof. Tyrone Duncan and Prof. Heping He for their willingness to serve on my master's thesis committee.

Also, I would like to thank my parents, my mother-in-law and my wife for supporting me spiritually all the time.

Finally, my special thanks go to my friends---Xianping Li and Weilin Xiao for their guidance in MATLAB. Without their help, my master's thesis could not have been finished.

Lawrence, November 2008

Xiaorong Yang

Abstract

Forecasting volatility has held the attention of academics and practitioners all over the world. The objective for this master's thesis is to predict the volatility in stock market by using *generalized autoregressive conditional heteroscedasticity*(GARCH) methodology. A detailed explanation of GARCH models is presented and empirical results from Dow Jones Index are discussed. Different from other literatures in this field, this paper studies forecasting volatility from a new perspective by comparing GARCH(P,Q) model with GJR-GARCH(P,Q) model and EGARCH(P,Q) model. GJR-GARCH(P,Q) model turns out to be more powerful than GARCH(P,Q) model due to catching some leverage effects successfully. This makes our prediction more reliable and accurate. This paper also shows that both GARCH(P,Q) model and GJR-GARCH(P,Q) model are good choices for dealing with heteroscedastic time series.

Keywords: Volatility Heteroscedastic time series

GARCH (P,Q) model GJR-GARCH(P,Q) model EGARCH(P,Q) model

contents**1 Introduction**

1.1	Volatility	1
1.2	Why Forecasting Volatility is so important.....	3
1.3	Heteroscedastic time series.....	5
1.4	Some popular Volatility Models.....	6
1.4.1	Exponentially Weighted Moving Average (EWMA).....	6
1.4.2	Stochastic Volatility Models(SV).....	7
1.4.3	ARCH Volatility Models.....	7
1.4.4	GARCH Volatility Models.....	8

2 GARCH Models

2.1	Statistics.....	10
2.2	Correlation.....	11
2.2.1	Autocorrelation.....	11
2.2.2	Partial autocorrelation.....	11
2.3	Stochastic Processes.....	11
2.3.1	Autoregressive process.....	11
2.3.2	Moving average process.....	12
2.3.3	Stationarity process.....	12
2.4	The form of GARCH Models.....	13
2.4.1	The form of GARCH(P,Q) Models.....	13
2.4.2	The form of EGARCH(P,Q) Models.....	14
2.4.3	The form of GJR-GARCH(P,Q) Models.....	15
2.5	Why GARCH Models.....	16

3 GARCH models Identification, estimation and diagnostic checking

3.1	GARCH models Identification.....	17
3.1.1	Use of the ACF and PACF.....	17
3.1.2	Use of model selection criteria AIC and BIC.....	17
3.2	GARCH models Estimation.....	18
3.3	GARCH models Diagnostic Checking.....	20

4 Empirical Application

4.1	Data Analysis.....	21
4.2	Model Identification.....	25
4.3	Model estimation.....	25
4.4	Model Diagnostic Checking.....	27
4.5	Model Simulation.....	31
4.6	Model Forecast.....	32

5 Conclusion

6 Future research
References

1 Introduction

1.1 Volatility

Volatility most frequently refers to the standard deviation of the continuously compounded returns of a financial instrument with a specific time horizon. It is often used to quantify the risk of the instrument over that time period. Historical volatility is the volatility of a financial instrument based on historical returns. This phrase is used particularly when it is wished to distinguish between the actual volatility of an instrument in the past, and the current volatility implied by the market.

For a financial instrument whose price follows a Gaussian random walk, or Wiener process, the volatility increases as time increases. Conceptually, this is because there is an increasing probability that the instrument's price will be farther away from the initial price as time increases. However, rather than increase linearly, the volatility increases with the square-root of time as time increases, because some fluctuations are expected to cancel each other out, so the most likely deviation after twice the time will not be twice the distance from zero. Volatility is typically expressed in annualized terms. The annualized volatility σ is the standard deviation σ of the instrument's logarithmic returns in a year. The generalized volatility σ_T for time horizon T in years is expressed as: $\sigma_T = \sigma\sqrt{T}$ (see

[13]). Note that the formula used to annualize returns is not deterministic, but is an extrapolation valid for a random walk process whose steps have finite variance. Generally, the relation between volatility in different time scales is more complicated, involving the Lévy stability exponent α : $\sigma_T = T^{\frac{1}{\alpha}} \sigma$. If $\alpha = 2$ you get the Wiener process scaling relation.

More broadly, volatility refers to the degree of (typically short-term) unpredictable change over time of a certain variable. It may be measured via the standard deviation of a sample, as mentioned above. However, price changes actually do not follow Gaussian distributions. Better distributions used to describe them actually have “fat tails” although their variance remains finite. Therefore, other metrics may be used to describe the degree of spread of the variable. As such, volatility reflects the degree of risk faced by someone with exposure to that variable. Volatility does not imply direction. This is due to the fact that all changes are squared. An instrument that is more volatile is likely to increase or decrease in value more than one that is less volatile.

Volatility is often viewed as a negative in that it represents uncertainty and risk. However, volatility can be good in that if one shorts on the peaks, and buys on the lows one can make money, with greater money coming with greater volatility. The possibility for money to be made via volatile markets is how short term market

players like day traders hope to make money, and is in contrast to the long term investment view of buy and hold. In today's markets, it is also possible to trade volatility directly, through the use of derivative securities such as options and variance swaps.

1.2 Why Forecasting Volatility is so important?

Volatility plays an important role in financial markets and has held the attention of academics and practitioners over the last two decades.

First of all, volatility has become a key input to many investment decisions and portfolio creations. Investors and portfolio managers have certain levels of risk which they can bear. A good forecast of the volatility of asset prices over the investment holding period is a good starting point for assessing investment risk. Volatility is the most important variable in the pricing of derivative securities, whose trading volume has quadrupled in recent years. To price an option, we need to know the volatility of the underlying asset from now until the option expires. In fact, the market convention is to list option prices in terms of volatility units. Nowadays, one can buy derivatives that are written on volatility itself, in which case the definition and measurement of volatility will be clearly specified in the derivative contracts. In these new contracts, volatility now becomes the underlying "asset." So a volatility forecast is needed to price such derivative

contracts.

Next, financial risk management has taken a central role since the first Basle Agreement was established in 1996. This effectively makes volatility forecasting a compulsory risk-management exercise for many financial institutions around the world. Banks and trading houses have to set aside reserve capital of at least three times that of value-at-risk (VaR), which is defined as the minimum expected loss with a 1% confidence level for a given time horizon (usually one or ten days). Sometimes, a 5% critical value is used. Such VaR estimates are readily available given volatility forecast, mean estimate, and a normal distribution assumption for the changes in total asset value. When the normal distribution assumption is disputed, which is very often the case, volatility is still needed in the simulation process used to produce the VaR figures. Financial market volatility can have a wide repercussion on the economy as a whole. The incidents caused by the terrorists' attack on September 11, 2001, and the recent financial crisis in the United States have caused great turmoil in financial markets and a negative impact on the world economy. This is clear evidence of the important link between financial market uncertainty and public confidence.

Finally, policy makers often rely on market estimates of volatility as a barometer for the vulnerability of financial markets and the economy. In the United States,

the Federal Reserve explicitly takes into account the volatility of stocks, bonds, currencies, and commodities in establishing its monetary policy ([13]). The Bank of England is also known to make frequent references to market sentiment and option implied densities of key financial variables in its monetary policy meetings.

1.3 Heteroscedastic time series

It's common knowledge that types of assets experience periods of high and low volatility. That is, during some periods prices go up and down quickly, while during other times they might not seem to move at all. Periods when prices fall quickly (a crash) are often followed by prices going down even more, or going up by an unusual amount. Also, a time when prices rise quickly (a bubble) may often be followed by prices going up even more, or going down by an unusual amount. The converse behavior can last for a long time as well. Most typically, extreme movements are presaged by larger movements than usual. This is termed autoregressive conditional heteroscedasticity. Of course, whether such large movements have the same direction, or the opposite, is more difficult to say.

Many financial time series turn out to be heteroscedastic, meaning that its variance vary with time. *Heteroscedastic time series* have some special characteristics([7]pp.32-34). The first is *fat-tail behavior*. The probability

distributions of time series often exhibit fatter tails than the standard normal (i.e. $kurtosis > 3$). The second is *volatility clustering*. Large changes tend to follow large changes, and small changes tend to follow small changes. In either case, the changes from one period to the next are typically of unpredictable sign. Thirdly, they have *squared series autocorrelation*. Although the ACF (Autocorrelation Function) of time series themselves are largely uncorrelated, the ACF of their squared series exhibit some correlation. Finally, they may have leverage effect. This effect often results in observed asset returns being negatively correlated with changes in volatility. For certain asset classes, volatility tends to rise in response to lower than expected returns and to fall in response to higher than expected returns.

1.4 Some popular Volatility Models

In this section, we introduce various popular time series volatility models that use the historical information set to formulate volatility forecasts.

1.4.1 Exponentially Weighted Moving Average Models(EWMA)

An exponentially weighted moving average model applies weighting factors which decrease exponentially. The weighting for each older data point decreases exponentially, giving much more importance to recent observations while still not discarding older observations entirely. The EWMA model is given by([12])

$$\sigma_n^2 = \lambda \sigma_{n-1}^2 + (1 - \lambda) u_{n-1}^2$$

Where $0 < \lambda < 1$, σ_n is the volatility on day n , and u_i is the daily return for a specific day.

1.4.2 Stochastic Volatility Models(SV)

Jacquier, Polson, and Ross(1994)([16]) came up with a univariate stochastic volatility model. The mean and volatility equations of an asset return r_t are

$$r_t = \beta_0 + \beta_1 x_{1t} + \dots + \beta_p x_{pt} + \sigma_t \varepsilon_t$$

$$\ln \sigma_t^2 = \alpha_0 + \alpha_1 \ln \sigma_{t-1}^2 + v_t$$

Where $\{x_{it} | i = 1, \dots, p\}$ are explanatory variables, $\beta_j (j = 1, \dots, p)$ are parameters, $\{\varepsilon_t\}$ is a Gaussian white noise sequence with mean 0 and variance 1, $\{v_t\}$ is also a Gaussian white noise sequence with mean 0 and variance σ_v^2 , and $\{\varepsilon_t\}$ and $\{v_t\}$ are independent. Also, we assume $|\alpha_1| < 1$ so that the log volatility process $\ln \sigma_t^2$ is stationary.

1.4.3 ARCH Volatility Models

The first model that provides a systematic framework for volatility modeling is the ARCH model of Engle(1982)([2]). The basic idea of ARCH models is that(a) the mean-corrected asset return a_t is serially uncorrelated, but dependent, and(b) the dependence of a_t can be described by a simple quadratic function of its lagged values. Specifically, an ARCH(m) model assumes that

$$a_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \dots + \alpha_m a_{t-m}^2,$$

Where $\{\varepsilon_t\}$ is a sequence of independent and identically distributed random variables with mean zero and variance 1, $\alpha_0 > 0$ and $\alpha_i \geq 0$ for $i > 0$. In practice, ε_t is often assumed to follow the standard normal or standardized Student- t distribution.

From the structure of the model, it is seen that large past squared shocks $\{a_{t-i}^2\}_{i=1}^m$ imply a large conditional variance σ_t^2 for the mean-corrected return a_t . Consequently, a_t tends to assume a large value. This means that, under the ARCH framework, large shocks tend to be followed by another large shock.

1.4.4 GARCH Volatility Models

GARCH stands for *generalized autoregressive conditional heteroscedasticity*. You can think of heteroscedasticity as time-varying variance(volatility). Conditional implies a dependence on the observations of the immediate past, and autoregressive describes a feedback mechanism that incorporates past observations into the present. So, GARCH is a mechanism that includes past variance in the explanation of future variances. More specifically, GARCH is a time-series technique that you use to model the serial dependence of volatility. Compared with other time-series models, GARCH models can provide a better description for heteroscedastic time series (we will discuss this in section 2).

Bollerslev(1986)([1])developed GARCH as a generalization of Engle's(1982) ([2])original ARCH volatility modeling technique. He designed it to offer a more parsimonious model (using fewer parameters) that lessens the computational burden. Nelson (1991)([3]) proposed an exponential GARCH (EGARCH) model, based on a logarithmic expression of the conditional variability in the variable under analysis. Later, a number of modifications were derived from this method. One of them is GJR-GARCH model of Glosten, Jagannathan and Runkle(1993)([4]) which can accommodate the asymmetry in the response of the variance to a shock. The recent application of the GARCH model in the capital market was studied by Christoffersen, P., and K. Jacobs (2004)([5]) Giovanni B-A, Robert F. E. and Loriano M.(2008)([6]), GARCH models are widely applied to diverse fields such as risk management, portfolio management, option pricing and asset allocation etc.

This paper applies GARCH methods to stock market and builds a GARCH model for the volatility of DowJones Index daily closing value log-return series. The rest of the paper is organized as follows. Section 2 gives the form of the GARCH models. Section 3 introduces the methodology of identification, estimation and diagnostic checking for GARCH models. Section 4 and section 5 provide an empirical application for GARCH models.

2 GARCH Models

2.1 Statistics

A brief description of the statistics that will be used is presented. In all cases below, X is a discrete valued stochastic variable, k is the summation index and $p_x(k)$ is the probability that X is taking value k . A more detailed description is presented in Hamilton (1994)([14]).

The first moment is the population mean and is defined as

$$E(X) = \sum_k k p_x(k) = \mu$$

The noncentral second moment is then defined as

$$E(X^2) = \sum_k k^2 p_x(k) = \text{Var}(X) + E^2(X)$$

Noncentral moments are then in the general case defined as

$$E(X^r) = \sum_k k^r p_x(k) \quad r = 1, 2, 3, \dots$$

Skewness is defined as

$$\frac{E((X - \mu)^3)}{(\text{Var}(X))^{3/2}}$$

A variable with positive skewness is more likely to have is values far above the mean value than far below. For a normal distribution the skewness is zero.

Kurtosis is defined as

$$\frac{E((X - \mu)^4)}{(\text{Var}(X))^2}$$

For a normal distribution the kurtosis is 3. A distribution with a kurtosis greater than 3 has more probability mass in the tails, so called “fat tails”, or leptokurtic.

2.2 Correlation

2.2.1 Autocorrelation

The j th autocorrelation is defined as the j th autocovariance divided by the variance:

$$\text{corr}(X_t, X_{t-j}) = \frac{E[(X_t - \mu_t)(X_{t-j} - \mu_{t-j})]}{\sqrt{\text{Var}(X_t)\text{Var}(X_{t-j})}} \quad (\text{Hamilton 1994})([14])$$

2.2.2 Partial autocorrelation

The partial autocorrelation is also a useful tool in the identification of a time series. It is defined as the last coefficient $\alpha_m^{(m)}$ in the following linear projection of Y on the m most recent values:

$$Y_t = \alpha_0^{(m)} + \alpha_1^{(m)}Y_{t-1} + \alpha_2^{(m)}Y_{t-2} + \dots + \alpha_m^{(m)}Y_{t-m} \quad t = 1, 2, \dots, m \quad (\text{Hamilton 1994})([14])$$

If the process were a true AR(p) process the coefficients with lags greater than m would be zero.

2.3 Stochastic Processes

2.3.1 Autoregressive process

First we consider an AR(1) process

$$Y_t = \alpha_0 + \beta_1 Y_{t-1} + \varepsilon_t$$

where α_0 and β_1 can be any real constants. ε_t is a white noise process, that

is, $E(\varepsilon_t) = 0$ and $E(\varepsilon_t \varepsilon_s) = \begin{cases} \sigma^2 & \text{for } t = s \\ 0 & \text{for } t \neq s \end{cases}$. An AR(1) process can also be

generalized to an AR(p) process.([14])

$$Y_t = \alpha_0 + \sum_{j=1}^p \beta_j Y_{t-j} + \varepsilon_t$$

2.3.2 Moving average process

A moving average process of order one MA(1) is described as([14])

$$Y_t = \alpha_0 + \alpha_1 \varepsilon_{t-1} + \varepsilon_t$$

where α_0 and α_1 could be any real constants, and ε_t is a white noise process .

This can of course also be considered in the general MA(q) case. Then the

equation becomes

$$Y_t = \alpha_0 + \sum_{j=1}^q \alpha_j \varepsilon_{t-j} + \varepsilon_t$$

2.3.3 Stationary process

A stationary process ([12]) is a stochastic process whose joint probability

distribution does not change when shifted in time or space. As a result, parameters

such as mean and variance, if they exist, also do not change over time or position.

Formally, let X_t be a stochastic process and let $F_{X_{t_1}, \dots, X_{t_k}}(x_{t_1}, \dots, x_{t_k})$ represent the

cumulative distribution function of the joint distribution of X_t at times t_1, \dots, t_k .

Then X_t is said to be stationary if, for all k , for all τ , and for all t_1, \dots, t_k ,

$$F_{X_{t_1}, \dots, X_{t_k}}(x_{t_1}, \dots, x_{t_k}) = F_{X_{t_1+\tau}, \dots, X_{t_k+\tau}}(x_{t_1}, \dots, x_{t_k}) .$$

2.4 The form of GARCH Models

2.4.1 The form of GARCH(P,Q) Models

The general GARCH(P,Q) model([7]) includes two parts:

(i) Conditional Mean Equation

$$y_t = C + \varepsilon_t + \sum_{i=1}^R \phi_i y_{t-i} + \sum_{j=1}^M \theta_j \varepsilon_{t-j} + \sum_{k=1}^{N_x} \beta_k x(t, k) \quad (2-1)$$

This is a general ARMAX(R, M, N_x) model. It applies to all variance models with autoregressive coefficients $\{\phi_i\}$, moving average coefficients $\{\theta_j\}$, innovations $\{\varepsilon_t\}$, and stationary return series $\{y_t\}$.

X is an explanatory regression matrix in which each column is a time series. $X(t, k)$ denotes the t th row and k th column of this matrix.

The eigenvalues $\{\lambda_i\}$ associated with the characteristic AR polynomial

$$\lambda^R - \phi_1 \lambda^{R-1} - \phi_2 \lambda^{R-2} - \dots - \phi_R$$

must lie inside the unit circle to ensure stationarity.

Similarly, the eigenvalues associated with the characteristic MA

polynomial $\lambda^M + \phi_1 \lambda^{M-1} + \phi_2 \lambda^{M-2} + \dots + \phi_M$ must lie inside the unit circle to ensure invertibility.

(ii) Conditional Variance Equation

The general GARCH(P,Q) model for the conditional variance of innovations is

$$\sigma_t^2 = K + \sum_{i=1}^P G_i \sigma_{t-i}^2 + \sum_{j=1}^Q A_j \varepsilon_{t-j}^2 \quad (2-2)$$

with constraints

$$\sum_{i=1}^P G_i + \sum_{j=1}^Q A_j < 1$$

$$K > 0$$

$$G_i \geq 0 \quad i = 1, 2, \dots, P$$

$$A_j \geq 0 \quad j = 1, 2, \dots, Q$$

The basic GARCH(P,Q) model is a symmetric variance process, in that it ignores the sign of the disturbance.

In particular, when P=Q=1, we have GARCH(1,1) model as follows:

$$(i) \quad y_t = C + \varepsilon_t \quad (2-3)$$

$$(ii) \quad \sigma_t^2 = K + G_1 \sigma_{t-1}^2 + A_1 \varepsilon_{t-1}^2, \quad 0 \leq G_1, A_1 \leq 1, \quad (G_1 + A_1) < 1 \quad (2-4)$$

2.4.2 The form of EGARCH(P,Q) Models

The general EGARCH(P,Q) model ([7]) for the conditional variance of the innovations, with leverage terms and an explicit probability distribution assumption, is

$$\log \sigma_t^2 = k + \sum_{i=1}^P G_i \log \sigma_{t-1}^2 + \sum_{j=1}^Q A_j \left[\frac{|\varepsilon_{t-j}|}{\sigma_{t-j}} - E \left\{ \frac{|\varepsilon_{t-j}|}{\sigma_{t-j}} \right\} \right] + \sum_{j=1}^Q L_j \left(\frac{\varepsilon_{t-j}}{\sigma_{t-j}} \right)$$

Where $E \left\{ \frac{|\varepsilon_{t-j}|}{\sigma_{t-j}} \right\} = E \left(\frac{|\varepsilon_{t-j}|}{\sigma_{t-j}} \right) = \sqrt{\frac{2}{\pi}}$ for the Gaussian distribution, and

$$E \left\{ \frac{|\varepsilon_{t-j}|}{\sigma_{t-j}} \right\} = E \left(\frac{|\varepsilon_{t-j}|}{\sigma_{t-j}} \right) = \sqrt{\frac{v-2}{\pi}} \left(\frac{\Gamma(\frac{v-1}{2})}{\Gamma(\frac{v}{2})} \right)$$

for the Student's t distribution, with degrees of freedom $\nu > 2$.

$$\sum_{i=1}^P G_i + \sum_{j=1}^Q A_j + \frac{1}{2} \sum_{j=1}^Q L_j < 1$$

$$k > 0$$

$$G_i \geq 0 \quad i=1,2,\dots,P$$

$$A_j \geq 0 \quad j=1,2,\dots,Q$$

$$A_j + L_j \geq 0 \quad j=1,2,\dots,Q$$

The GARCH Toolbox software treats EGARCH(P,Q) models as ARMA(P,Q) models for $\log \sigma_t^2$. Thus, it includes the stationarity constraint for EGARCH(P,Q) models by ensuring that the eigenvalues of the characteristic polynomial $\lambda^P - G_1 \lambda^{P-1} - G_2 \lambda^{P-2} - \dots - G_P$ are inside the unit circle.

2.4.3 The form of GJR-GARCH(P,Q) Models

The general GJR-GARCH(P,Q) model([7]) for the conditional variance of the innovations with leverage terms is

$$\sigma_t^2 = k + \sum_{i=1}^P G_i \sigma_{t-i}^2 + \sum_{j=1}^Q A_j \varepsilon_{t-j}^2 + \sum_{j=1}^Q L_j S_{t-j} \varepsilon_{t-j}^2 \quad (2-5)$$

$$\text{Where } S_{t-j} = \begin{cases} 1 & \text{if } \varepsilon_{t-j} < 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{and}$$

$$\sum_{i=1}^P G_i + \sum_{j=1}^Q A_j + \frac{1}{2} \sum_{j=1}^Q L_j < 1$$

$$k > 0$$

$$G_i \geq 0 \quad i=1,2,\dots,P$$

$$A_j \geq 0 \quad j=1,2,\dots,Q$$

$$A_j + L_j \geq 0 \quad j=1,2,\dots,Q$$

2.5 Why GARCH Models

Theoretically, GARCH models can describe heteroscedastic time series very well. Let's consider GARCH(1,1) model and recall three characteristics of heteroscedastic time series mentioned in section 1.3. First, GARCH(1,1) model can describe fat-tail behavior. We can prove that the kurtosis of GARCH(1,1) model is greater than 3 ([8]). Second, GARCH(1,1) model can reflect volatility clustering. By equation (2-4) we can see that a large σ_{t-1}^2 or ε_{t-1}^2 gives rise to a large σ_t^2 . This means that a large σ_{t-1}^2 tends to be followed by another large σ_t^2 , generating the behavior of volatility clustering. Third, by equation (2-4) it is easy to see that the squared series autocorrelation exists. For this reason, the three characteristics are also called GARCH effects.

GARCH models are consistent with various forms of efficient market theory. These theories state that asset returns observed in the past cannot improve the forecasts of asset returns in the future. Since GARCH innovations $\{\varepsilon_t\}$ are serially uncorrelated, GARCH modeling does not violate efficient market theory.

3 GARCH models Identification, estimation and diagnostic checking

3.1 GARCH models Identification

3.1.1 Use of the Autocorrelation and Partial Autocorrelation Functions

(ACF and PACF)

Time series identification with the ACF and PACF([15])

Shape of ACF	Indicated Model
Exponential, decaying to zero	Autoregressive model. Use the PACF plot to identify the order of autoregressive model.
Alternating positive and negative, decaying to zero	Autoregressive model. Use the PACF plot to identify the order of autoregressive model.
One or more spikes, rest are essentially zero	Moving average model. Order identified by where plot becomes zero.
Decay, starting after a few lags	Mixed and autoregressive and moving average model.
All zero or close to zero	Data is essentially random.
High values at fixed intervals	Include seasonal autoregressive term.
No decay to zero	Series is not stationary.

3.1.2 Use of model selection criteria AIC and BIC([9] pp.200–201)

AIC(Akaike Information Criterion) and BIC(Bayesian Information Criterion) are proposed by Akaike(1974)[10] and Schwarz(1978)[11] respectively. In the implementation of this approach, a range of potential GARCH(P,Q) models is estimated by maximum likelihood methods, and for each, a criterion such as AIC(normalized by sample size n), given by

$$AIC_{p,q} = \frac{-2 \ln(\text{maximized likelihood}) + 2r}{n} \approx \ln(\hat{\sigma}_a^2) + r \frac{2}{n} + \text{constant}$$

Or the related BIC given by

$$BIC_{p,q} = \ln(\hat{\sigma}_a^2) + r \frac{\ln(n)}{n}$$

is evaluated, where $\hat{\sigma}_a^2$ denotes the maximum likelihood estimate of σ_a^2 , and $r = p + q + 1$ denotes the number of parameters estimated in the model, including a constant term. In the criteria above, the first term essentially corresponds to minus $2/n$ times the log of the maximized likelihood, while the second term is a “penalty factor” for inclusion of additional parameters in the model. In the information criteria approach, models that yield a minimum value for the criteria are to be preferred, and the AIC or BIC values are compared among various models as the basis for selection of the model. Hence since the BIC criterion imposes a greater penalty for the number of estimated model parameters than does AIC, use of minimum BIC for model selection would always result in a chosen model whose number of parameters is no greater than that chosen under AIC. It is also important to note that we should difference given time series as many times as is needed to produce stationarity before we use AIC and BIC criterion.

3.2 GARCH models Estimation

Use of Maximum likelihood Estimation([7] pp.172–173)

To estimate the parameters of GARCH models, we use MATLAB7.0 garchfit

function. The garchfit function calls the appropriate log-likelihood objective function to estimate the model parameters using *maximum likelihood estimation(MLE)*. The chosen log-likelihood objective function proceeds as follows:

(i) Given the vector of current parameter values and the observed data Series, the log-likelihood function infers the process innovations (residuals) by *inverse filtering*. This inference operation rearranges the conditional mean equation (2-1) to solve for the current innovation ε_t :

$$\varepsilon_t = -C + y_t - \sum_{i=1}^R \phi_i y_{t-i} - \sum_{j=1}^M \theta_j \varepsilon_{t-j} - \sum_{k=1}^{N_x} \beta_k x(t, k)$$

This equation is a whitening filter, transforming a (possibly) correlated process y_t into an uncorrelated white noise process ε_t .

(ii) The log-likelihood function then uses the inferred innovations ε_t to infer the corresponding conditional variances σ_t^2 via recursive substitution into the previous conditional variance equation (2-2).

(iii) Finally, the function uses the inferred innovations and conditional variances to evaluate the appropriate log-likelihood objective function. If ε_t is Gaussian, the log-likelihood function(*LLF*) is

$$LLF = -\frac{T}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^T \log \sigma_t^2 - \frac{1}{2} \sum_{t=1}^T \varepsilon_t^2 \sigma_t^{-2}$$

If ε_t is Student's t, the log-likelihood function is

$$LLF = T \log \left\{ \frac{\Gamma\left[\frac{(v+1)}{2}\right]}{\pi^{1/2} \Gamma\left(\frac{v}{2}\right)} (v-2)^{-1/2} \right\} - \frac{1}{2} \sum_{t=1}^T \log \sigma_t^2 - \frac{v+1}{2} \sum_{t=1}^T \log \left[1 + \frac{\varepsilon_t^2}{\sigma_t^2 (v-2)} \right]$$

where T is the sample size, that is, the number of rows in the series $\{y_t\}$. The

degrees of freedom v must be greater than 2.

The conditional mean equation (2-1) and the conditional variance equations (2-2) are recursive, and generally require presample observations to initiate inverse filtering. For this reason, the objective functions shown here are referred to as *conditional log-likelihood functions*.

The iterative numerical optimization repeats the previous three steps until it satisfies suitable termination criteria.

3.3 GARCH models Diagnostic Checking

Use of Ljung-Box Q-Statistic lack-of-fit Test ([7])

The Ljung-Box lack-of-fit hypothesis test is based on the Q-statistic

$$Q = N(N+2) \sum_{k=1}^L \frac{r_k^2}{(N-k)}$$

Where N =sample size, L =the number of autocorrelation lags included in the statistic, and

r_k^2 is the squared sample autocorrelation at lag k . Once you fit a univariate model to an observed time series, you can use the Q-statistic as a lack-of-fit test for a

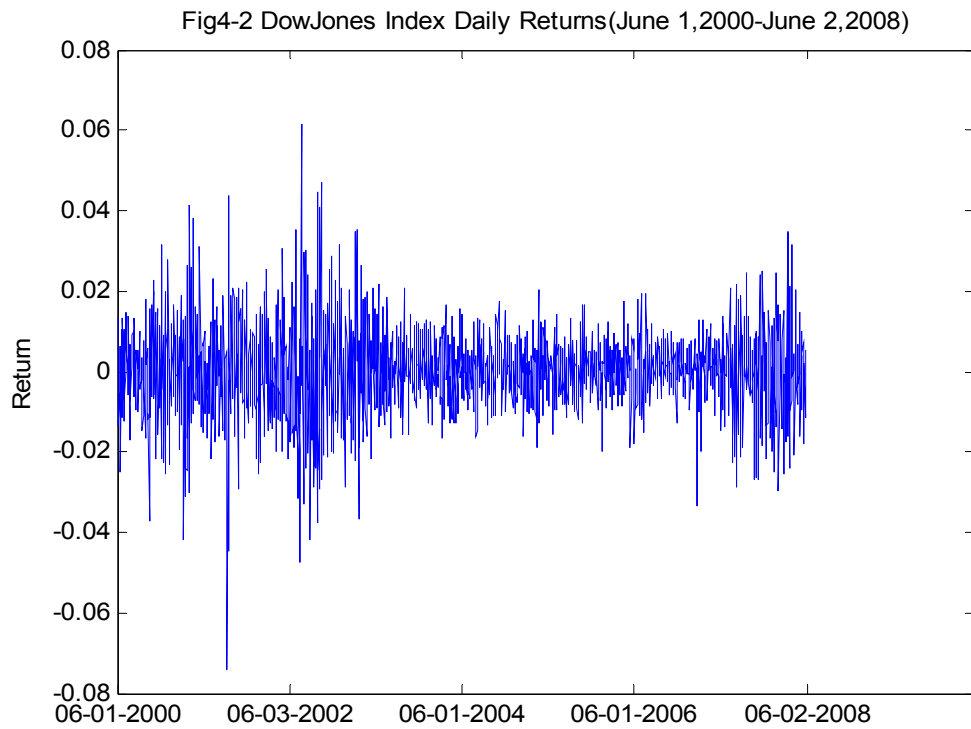
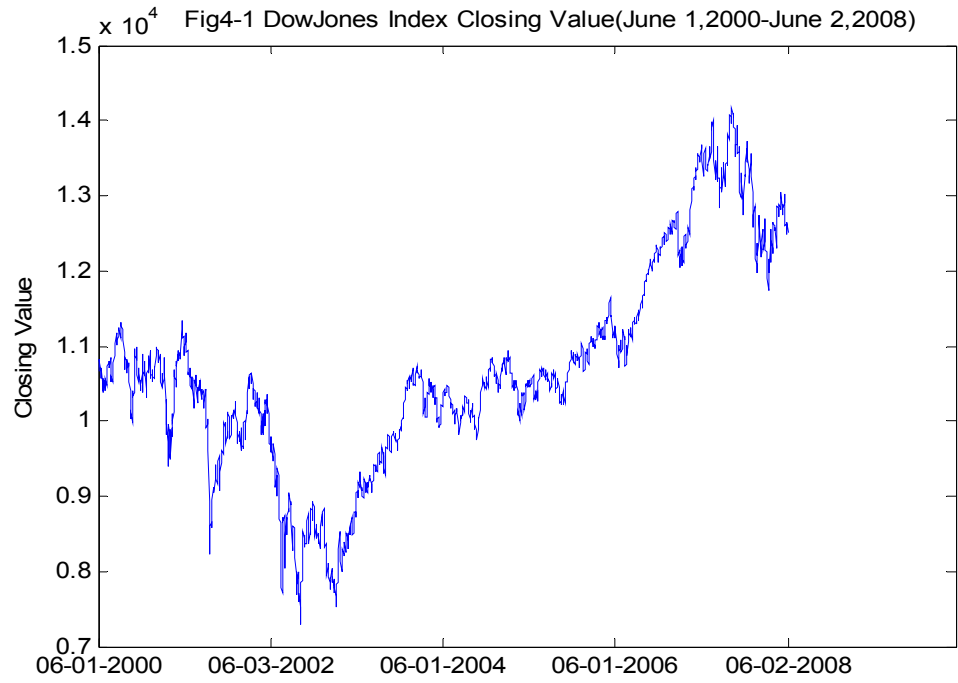
departure from randomness. Under the null hypothesis that the model fit is adequate, the test statistic is asymptotically chi-square distributed. If Boolean decision vector $H=0$ with highly significant p-values, also Q-statistic value less than corresponding chi-square critical value, then we accept the null hypothesis. That is, the model fit is good. Conversely, If $H=1$ with p-values=0 and Q-statistic value greater than corresponding chi-square critical value, then we reject the null hypothesis.

4 Empirical Application

4.1 Data Analysis

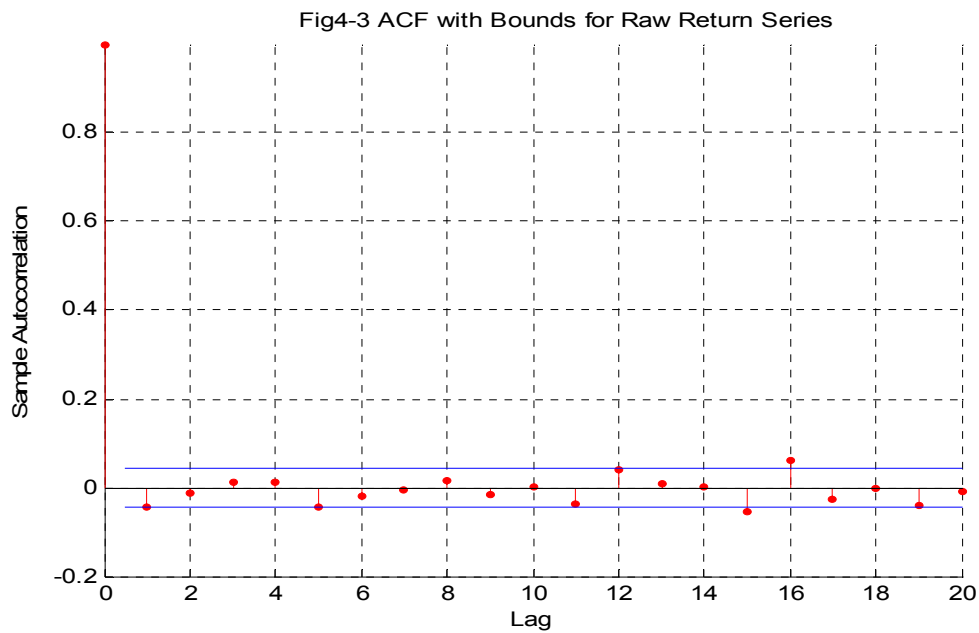
The data series comes from <http://finance.yahoo.com/> and consists of daily continuously closing price on the Dow Jones Index from June 1, 2000 through June 2, 2008 for a total of 2011 observations (Fig4-1). We denote successive price observations made at times t and $t + 1$ as P_t and P_{t+1} , respectively, then continuous compounding transforms a price series $\{ P_t \}$ into a return series $\{ y_t \}$ (Fig4-2)

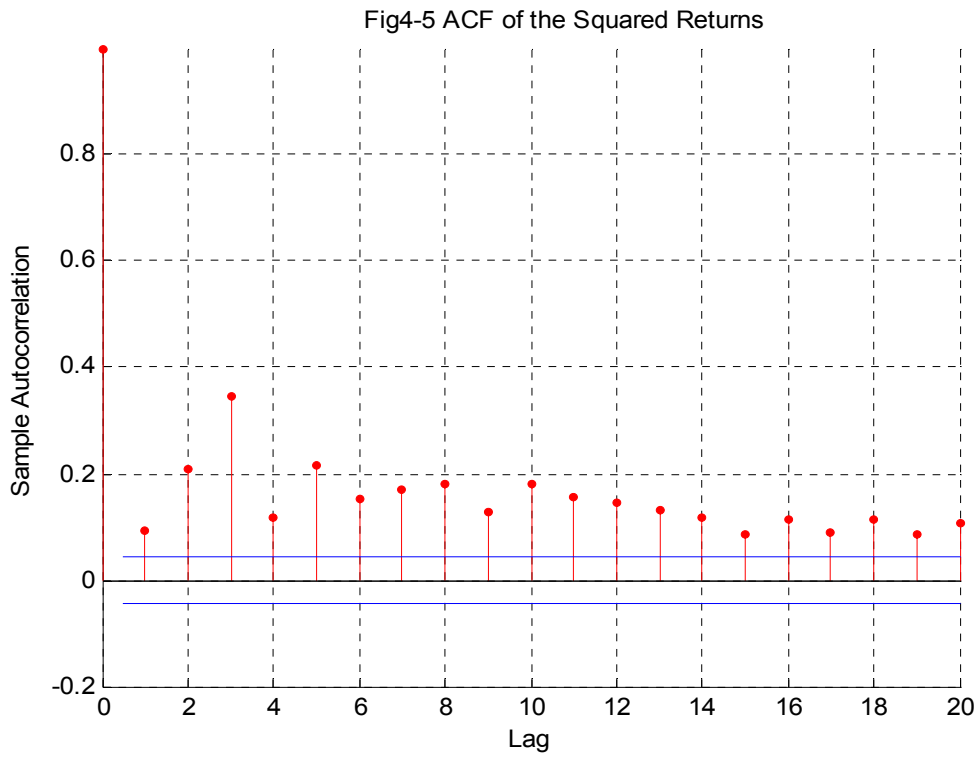
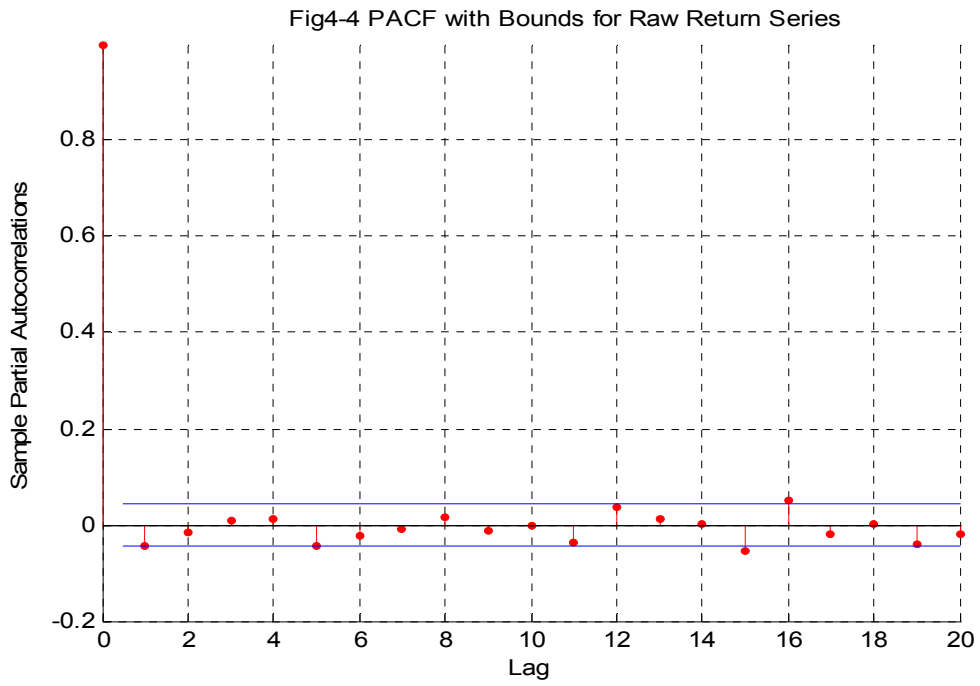
$$\text{as } y_t = \log \frac{P_{t+1}}{P_t} = \log P_{t+1} - \log P_t.$$



We use MATLAB7.0 to analyze the data series. The result is as follows:

- (i) The kurtosis of sequence $\{y_t\}$ $k=6.675589>3$, so the fat-tail behavior exists.
- (ii) From Fig 4-2 we can see that sequence $\{y_t\}$ shows volatility clustering.
- (iii) Fig4-5 shows that the ACF of the squared return sequence $\{y_t^2\}$ exhibits some correlation although the ACF and PACF of sequence $\{y_t\}$ are largely uncorrelated(Fig4-3 and Fig4-4).





By (i)~(iii), sequence $\{y_t\}$ is a heteroscedastic time series. Also it is stationary because its ACF and PACF die out quickly. So, we can build a GARCH model for it.

4.2 Model Identification

We use AIC and BIC criterion and list the result as follows:

Table 1 GARCH models AIC and BIC Comparison

	GARCH(1,1)	GARCH(1,2)	GARCH(2,1)
AIC	$1.0 e^{+004} * (-1.313796)$	$1.0 e^{+004} * (-1.312641)$	$1.0 e^{+004} * (-1.312441)$
BIC	-1.3109925	-1.310398	-1.309638

By Table 1, it is easy to see that GARCH(1,1) model gets the minimum AIC and BIC value. So, we select GARCH(1,1) model. Also, we can get the same result by using ACF and PACF.

4.3 Model estimation

Table 2 Parameters estimation for GARCH(1,1) model

Mean: ARMAX(0,0,0); Variance: GARCH(1,1)

Conditional Probability Distribution: Gaussian

Number of Model Parameters Estimated: 4

Parameter	Value	Standard Error	T Statistic
-----	-----	-----	-----
C	0.00038114	0.00018811	2.0262
K	1.1176e-006	2.0607e-007	5.4233
GARCH(1)	0.91523	0.0089953	101.7461
ARCH(1)	0.075114	0.0077822	9.6519

It is easy to check that all parameters are significant at significance level 0.05.

Also, all parameters satisfy the assumption of GARCH(P,Q) models.

Substituting these estimates in the equation (2-3) and (2-4), we get the

GARCH(1,1) model that best fits the observed data:

$$y_t = 0.00038114 + \varepsilon_t$$

$$\sigma_t^2 = 1.1176e^{-006} + 0.91523\sigma_{t-1}^2 + 0.075114\varepsilon_{t-1}^2$$

where $G_1 = GARCH(1) = 0.91523$ and $A_1 = ARCH(1) = 0.075114$. In addition,

$C = 0.00038114$ and $k = K = 1.1176e^{-006}$.

In order to catch leverage effect, we also consider EGARCH(1,1) model and GJR(1,1) model.

Table 3 Parameters estimation for EGARCH(1,1) model

Mean: ARMAX(0,1,0); Variance: EGARCH(1,1)

Conditional Probability Distribution: Gaussian

Number of Model Parameters Estimated: 6

Parameter	Value	Standard Error	T Statistic
C	0.00011544	0.00016726	0.6902
MA(1)	-0.058125	0.023843	-2.4378
K	-0.15587	0.019707	-7.9094
GARCH(1)	0.98333	0.0021583	455.5999
ARCH(1)	0.08484	0.015674	5.4127
Leverage(1)	-0.11129	0.008602	-12.9379

Where

$G_1 = GARCH(1) = 0.98333$, $A_1 = ARCH(1) = 0.08484$ and $L_1 = Leverage(1) = -0.11129$.

Since $K < 0$ and $G_1 + A_1 + 1/2L_1 = 1.0125 > 1$ and $A_1 + L_1 = -0.0265 < 0$, the parameters

G_1, A_1 and L_1 do not meet the requirements for EGARCH(P,Q) model. So, we

ignore EGARCH(1,1) model.

Table 4 Parameters estimation for GJR-GARCH(1,1) model

Mean: ARMAX(0,1,0); Variance: GJR(1,1)

Conditional Probability Distribution: Gaussian

Number of Model Parameters Estimated: 6

Parameter	Value	Standard Error	T Statistic
C	0.00010187	0.00017227	2.5914
MA(1)	-0.059174	0.024619	-2.4036
K	1.0796e-006	1.6728e-007	6.4537
GARCH(1)	0.92668	0.0089196	103.8928
ARCH(1)	0	0.0086692	0.0000
Leverage(1)	0.1245	0.012513	9.9491

It is easy to see that all parameters are significant at significance level 0.05.

Also, all parameters satisfy the assumption of GJR-GARCH(P,Q) models.

Plug these estimates in the equation (2-1) and (2-5), we get the GARCH(1,1)

model that best fits the observed data:

$$y_t = 0.00010187 - 0.059174\varepsilon_{t-1} + \varepsilon_t$$

$$\sigma_t^2 = 1.0796e - 006 + 0.92668\sigma_{t-1}^2 + 0.1245S_{t-1}\varepsilon_{t-1}^2$$

Where $S_{t-1} = \begin{cases} 1 & \text{if } \varepsilon_{t-1} < 0 \\ 0 & \text{otherwise} \end{cases}$.

4.4 Model Diagnostic Checking

Table 5 GARCH(1,1) model Diagnostic Checking(significance level $\alpha = 0.05$)

lag	H	p-value	Q-Statistic	Critical value
10	0	0.1482	14.5789	18.3070
15	0	0.1842	19.6896	24.9958
20	0	0.4073	20.8290	31.4104

Table 6 GJR-GARCH(1,1) model Diagnostic Checking($\alpha = 0.05$)

lag	H	p-value	Q-Statistic	Critical value
10	0	0.2621	12.3532	18.3070
15	0	0.2917	17.4671	24.9958
20	0	0.5438	18.6629	31.4104

From Table 5 and table 6 we can see, at significance level 0.05, for all lags 10,15 and 20, we have Boolean decision vector $H=0$ with highly significant p-values, also Q-statistic value less than corresponding chi-square critical value. So, we accept the null hypothesis. That is, the model fit is good.

Alternatively, we can test the goodness-of-fit for this GARCH(1,1) model by the following way:

First, we inspect the relationship between the innovations (residuals) derived from the fitted model, the corresponding conditional standard deviations, and the observed returns (Fig4-6).

Fig4-6 comparing innovations, conditional standard deviations and observed returns

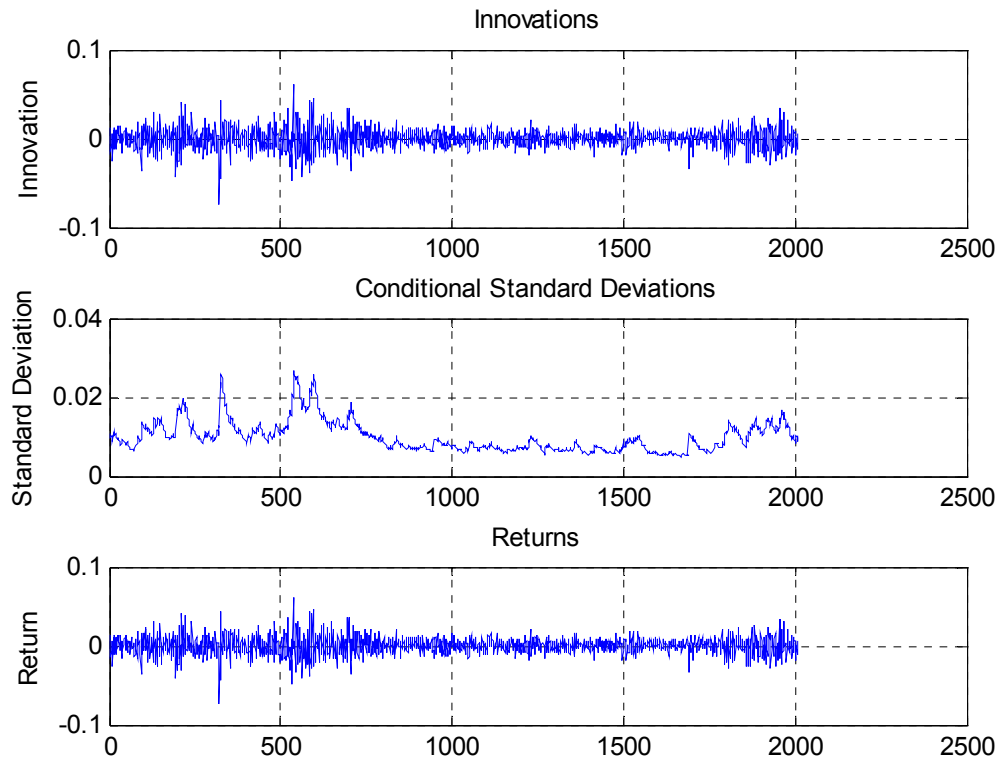
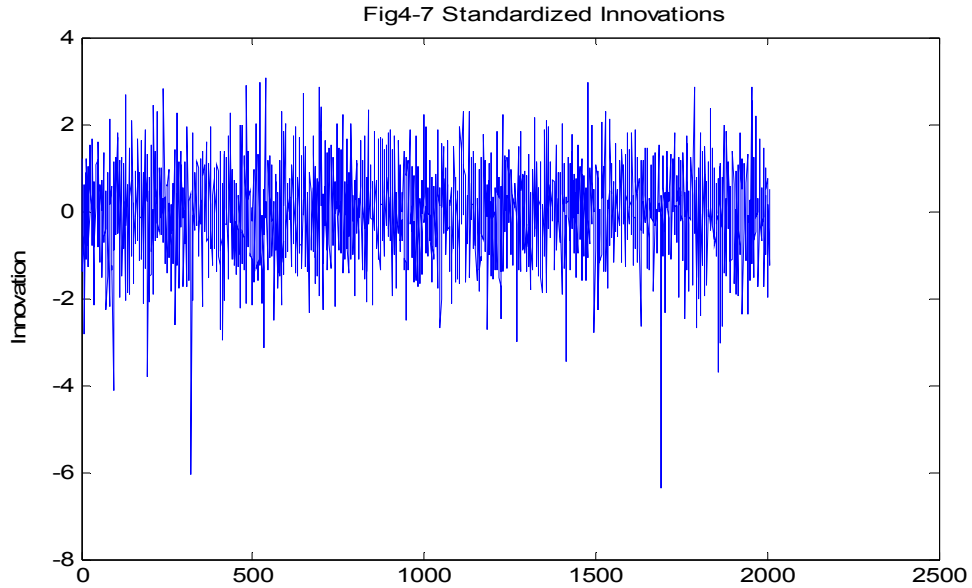
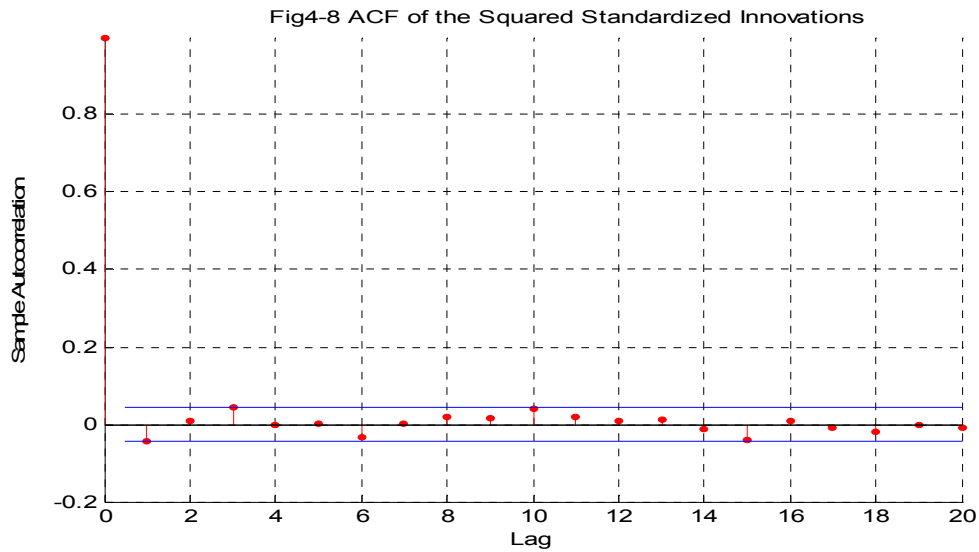


Fig4-6 shows that both the innovations (shown in the top plot) and the returns (shown in the bottom plot) exhibit volatility clustering. Also, the sum, $G1 + A1 = 0.91523 + 0.075114 = 0.990344 < 1$, is close to the integrated, non-stationary boundary given by the constraints associated with Equation (2-4).

Second, we plot the standardized innovations (the innovations divided by their conditional standard deviation) as Fig4-7:



The standardized innovations appear generally stable with little clustering. Third, we plot the ACF of the squared standardized innovations as Fig4-8:



The standardized innovations also show no correlation. Now compare the ACF of the squared standardized innovations in the Fig4-8 to the ACF of the squared returns before fitting the default model(Fig4-5). The comparison shows that our

GARCH(1,1) model sufficiently explains the heteroscedasticity in the raw returns. Similarly, we can test the goodness-of-fit for the GJR-GARCH(1,1) model by this way.

4.5 Model Simulation

We use this GARCH(1,1) model and GJR-GARCH(1,1) model to generate a single path of 1000 observations starting from the initial MATLAB random number generator state(Fig4-9 and Fig4-10). Each result is a single realization of 1000 observations each for the innovations $\{\varepsilon_t\}$, conditional standard deviations $\{\sigma_t\}$, and returns $\{y_t\}$ processes.

Fig4-9 GARCH(1,1) model simulation(a single path of 1000 observations)

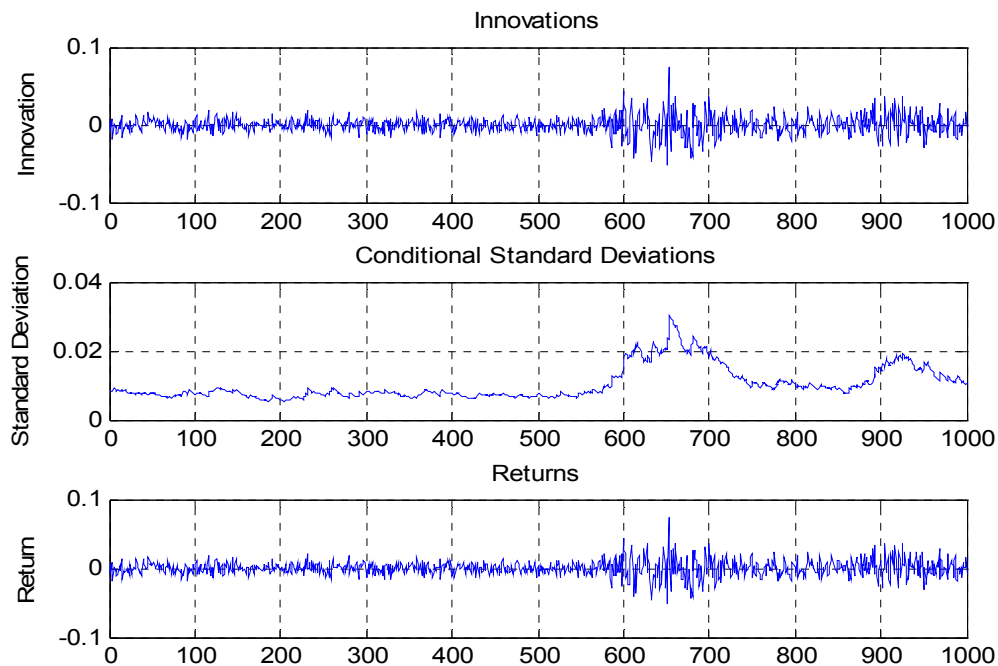
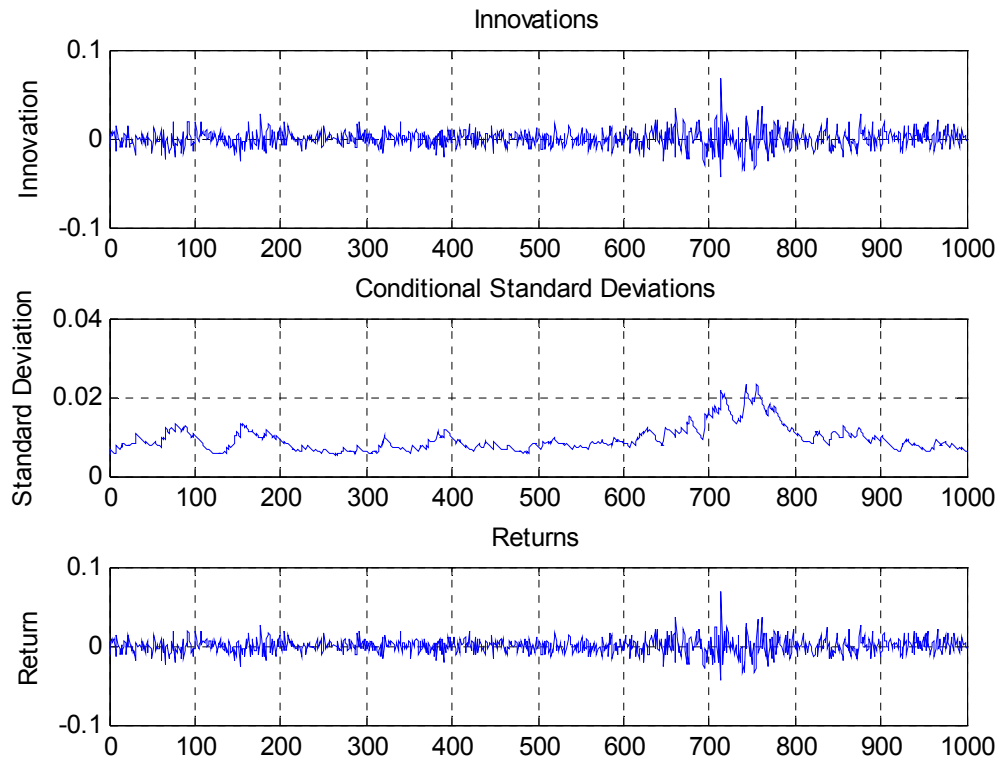


Fig4-10 GJR-GARCH(1,1) model simulation(a single path of 1000 observations)



4.6 Model forecast

We use MATLAB forecasting engine *garchpred* to compute minimum mean square error(MMSE) forecasts of the conditional mean of returns $\{y_t\}$ and the conditional standard deviation of the innovations $\{\varepsilon_t\}$ in each period of a user-specified forecast horizon. Also, we can compute the volatility forecasts of returns for assets held for multiple periods. For example, we can forecast the standard deviation of the return we would obtain if we purchased shares in a mutual fund that mirrors the performance of the Dow Jones Index today, and sold

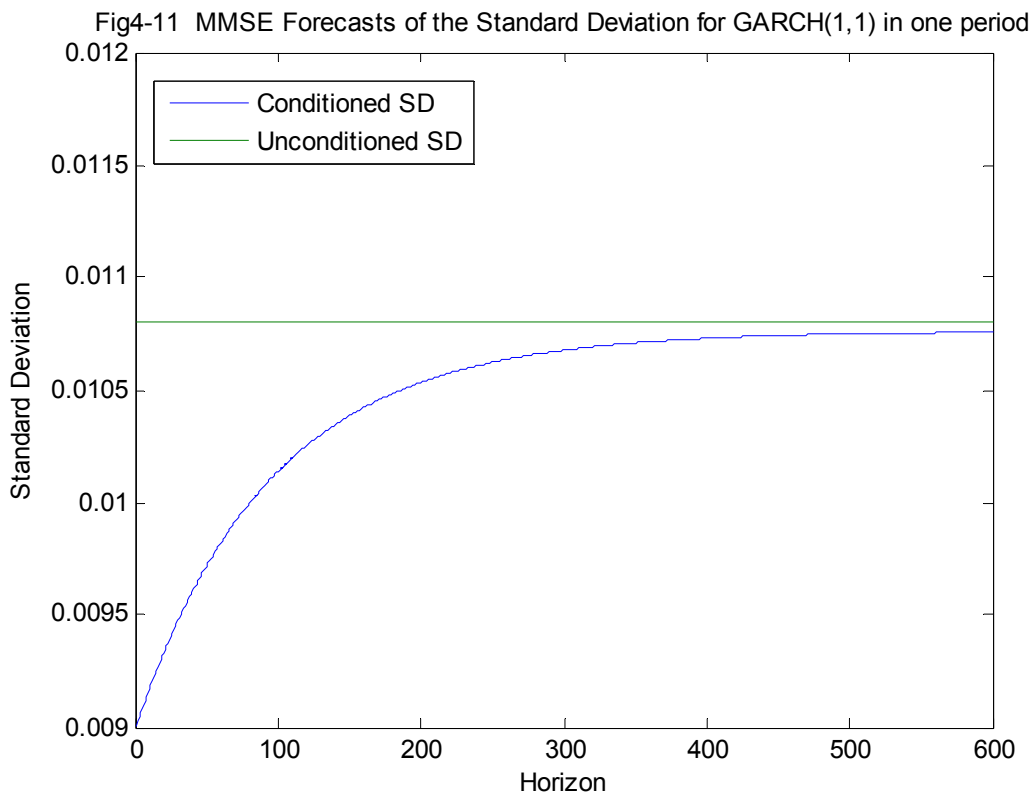
it 30 days from now . The MMSE can be obtained by the formula

$$Var_t[\sum_{i=1}^s y_{t+i}] = \sum_{i=1}^s [(1 + \sum_{j=1}^{s-i} \psi_j)^2 E_t(\sigma_{t+i}^2)]$$

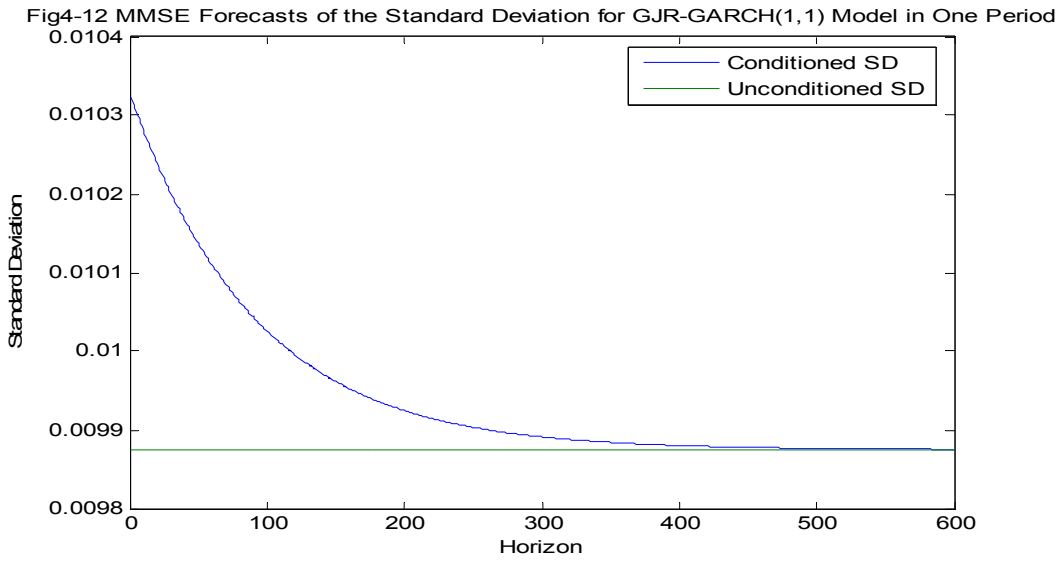
Where

- S is the forecast horizon of interest (Numperiods)
- ψ_j is the coefficient of the j th lag of the innovations process in an infinite-order MA representation of the conditional mean model.

4.6.1 MMSE forecasts for one-period

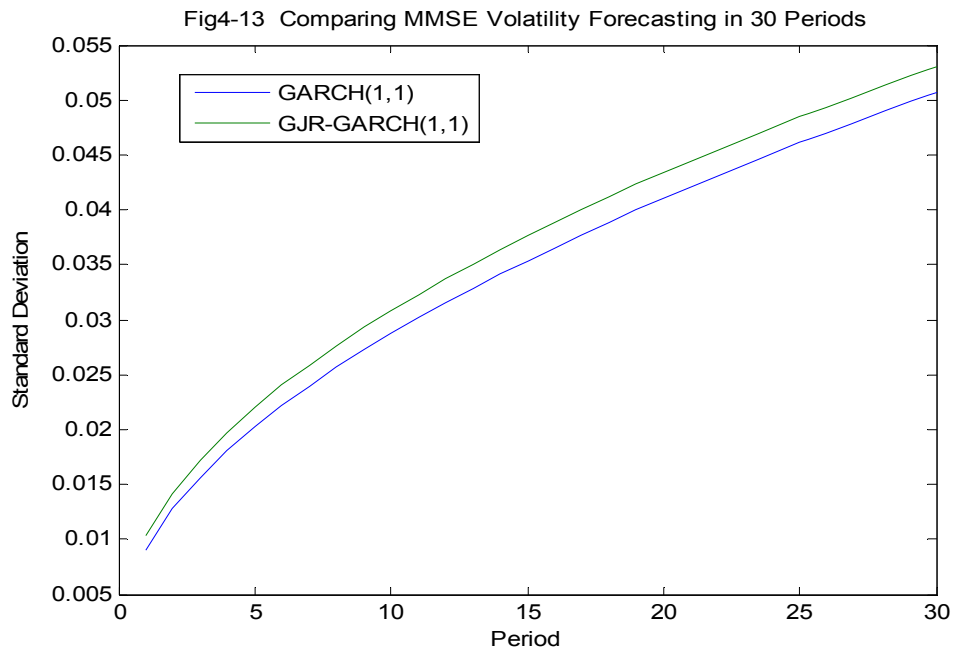


For GARCH(1,1) model, Fig4-11 shows that MMSE forecasts of the conditioned standard deviation increasingly converges to its unconditioned standard deviation.



For GJR-GARCH(1,1) model, Fig4-12 tells us that MMSE forecasts of the conditioned standard deviation decreasingly converges to its unconditioned standard deviation.

4.6.1 MMSE forecasts for multiple periods



5 Conclusion

Both GARCH(P,Q) model and GJR-GARCH(P,Q) model are good choices for forecasting volatility in financial market, especially for describing heteroscedastic time series. It should not be neglected that GARCH(P,Q) model responds equally to positive and negative shocks. To overcome its weakness, we use GJR-GARCH(1,1) model and catch some leverage effects successfully which makes our prediction more reliable and accurate. However, recent empirical studies of high-frequency financial time series indicate that the tail behavior of GARCH models remains too short even with standardized Student-t innovations([9] pp.94–95). So, GARCH models are only part of a solution. To make financial decisions, it is always necessary to connect GARCH models with other methods such as fundamental analysis. For instance, fundamental analysis can examine all relevant factors affecting the stock price in order to determine an intrinsic value for that stock.

6 Future research

The model for the mean equation used in this thesis uses a constant for the return series. This is consistent with the efficient market theory, despite the fact that investors also require extra return for taking extra risk. One improvement to this model could be to do further analysis of the model for the mean equation.

Besides, to make our prediction more accurate, we can compare MMSE method with other methods such as Median Squared Error (MedSE) and Mean Absolute Error(MAE).

References

- [1] Bollerslev, T., "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of GARCH*, Vol. 31, 1986, pp. 307–327.
- [2] Engle, Robert F., "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, Vol.50, 1982, pp. 987–1007.
- [3] Nelson, D.B., "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, Vol. 59, 1991, pp. 347–370.
- [4]Glosten, L., Jagannathan, R. and Runkle, D., "On the relation between expected value and the volatility of the nominal excess return on stocks," *Journal of Finance* 48, 1993, pp.1779–1801.
- [5]Christoffersen, P., and K. Jacobs., "Which GARCH Model for Option Valuation?" *Management Science*50, 2004, pp.204–21.
- [6]Giovanni Barone-Adesi, Robert F. Engle and Loriano Mancini. , "A GARCH Option Pricing Model with Filtered Historical Simulation," *Review of Financial Studies*, 21(3), 2008, pp.1223–1258.
- [7]MATLAB7.0., "GARCH toolbox user's guide," .
- [8]Ruey S. Tsay., "Analysis of Financial Time Series," *Wiley* (2002).
- [9] George E. P. Box, Gwilym M. Jenkins and Gregory C. Reinsel., "Time Series Analysis Forecasting and Control," *Prentice Hall*, 3rd edition (1994)
- [10] Akaike, H., " A new look at the statistical model identification," *IEEE Trans. Automatic Control*, AC-19, 1974, pp.716–723.
- [11] Schwarz, G., "Estimating the dimension of a model," *Ann. Statist.*, 6, 1978, pp.461–464.
- [12] <http://en.wikipedia.org/wiki>
- [13] Nasar, Sylvia. "For Fed, a New Set of Tea Leaves," *New York Times*(1992)
- [14] Hamilton, James D. and Rauli Susmel. "Autoregressive Conditional Heteroskedasticity and Changes in Regime," *J. Econometrics* 64:1–2, pp. 307–33.(1994)
- [15] www.itl.nist.gov
- [16] Jacquier, Polson & Rossi, "Bayesian Analysis of Stochastic Volatility", *Journal of Business and Economics Statistics*, 1994