

Time Series Analysis with ARIMA/GARCH Model

Name: HU WEI STUDENT NO.:54206763

1. Introduction

Recently, there has been much work on modeling the series data, especially in financial markets. The purposes of time series analysis are to understand or model the stochastic mechanism, which gives rise to an observed series, and to predict or forecast the future values of a series based on the history of that series. Generally, there are two primary methods in series analysis, which are frequency-domain and time-domain. The former is based mostly on Fourier Transform while the latter closely investigates the autocorrelation of the time series. Time domain series make huge use of Box-Jenkins and ARCH/GARCH methods to perform forecast of the series.

In this project, this research generally investigates the financial time series such as the price and return of stock index NASDAQ Composite Index using ARIMA and GARCH methods.

2. Some common features of financial time series

2.1 High volatility of price

The basic information we can capture from financial market is price. And the basic feature of financial time series, such as price of NASDAQ Composite in picture 1, is its relatively high volatility, which usually changes through time. The picture 1 illustrates the daily price of NASDAQ Composite from 01-Jan-2001 to 01-Jan-2015, from which we can get that the volatility of the stock index is obvious.

2.2 Brownian Motion Theory

Due to the geometrically randomness of stock price, we can investigate the log-price instead the original price of stock index by using the formula below

$$r_t = \log(P_t) - \log(P_{t-1})$$

The lower part in Picture 1 illustrates the graph of log-return of NASDAQ Composite, from which we can observe the volatility clustering, giving us a hint that they may not be independently and identically distributed—otherwise the variance would be constant over time.

2.3 The ARCH model

2.3.1 Autoregressive process (AR process)

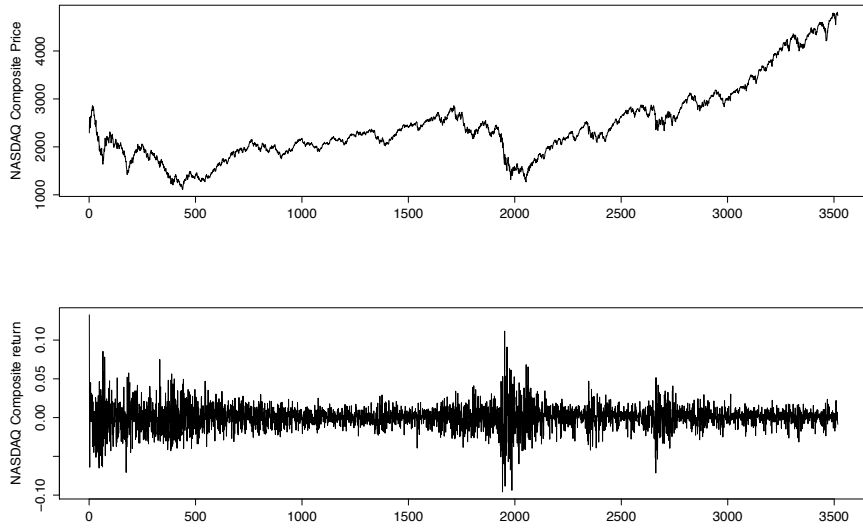
Time domain method is established and implemented by observing the

autocorrelation of the time series. For stationary time series, we can use AR process to investigate. Assume Y_t is modeled as weighted average of past observations plus a white noise “error”. The formula is as follow

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \cdots + \phi_p(Y_{t-p} - \mu) + \epsilon_t$$

ϵ_t is White Noise with zero mean and variance σ_ϵ^2 .

Picture 1 NASDAQ Composite Price and Log-Return



2.3.2 Moving Average process (MA process)

Assume time series Y_t is modeled as a weighted average (moving average) of the past value of white noise process ϵ_t rather than of past value. The formula is as follow

$$Y_t - \mu = \epsilon_t - \theta_1\epsilon_{t-1} - \cdots - \theta_q\epsilon_{t-q}$$

ϵ_t is also White Noise with zero mean and variance σ_ϵ^2 . If we combine the AR process and MA process together, we can get the ARMA process.

If the assumption of stationary of the time series can't sustain, we can make use of differencing of the random variables to convert the non-stationary process into stationary process. A time series $\{Y_t\}$ is said to follow an integrated autoregressive moving average (ARIMA) model if the d-th difference $\nabla^d Y_t$ is a stationary ARMA process.

2.3.3 Autoregression Conditional Heteroscedasticity process (ARCH)

The ARMA time series models are unsatisfactory for modeling volatility changes. And if the conditional variance of time series is inconstant. We need to take use of ARCH model. The formula of ARCH(q) model is

$$a_t = \epsilon_t \sigma_t; \sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i a_{t-i}^2$$

ϵ_t is White Noise(0, σ_ϵ^2), $\alpha_0 \geq 0$, $0 \leq \alpha_1 < 1$. ARCH(q) process is uncorrelated and it has a constant mean and a constant unconditional variance.

2.3 The General Autoregression Conditional Heteroscedasticity process (GARCH model)

In the GARCH model, assume $\eta_t = a_t^2 - \sigma_t^2$ and $E(\eta_t | a_{t-1}, \dots) = 0$, we get

$$a_t^2 = \alpha_0 + \sum_{i=1}^{\max(q,p)} (\alpha_i + \beta_i) a_{t-i}^2 + \eta_t - \sum_{i=1}^p \beta_i \eta_{t-i}^2$$

in which the past value of σ_t process are fed back. a_t is uncorrelated with a stationary mean and variance. Condition is $\sum_{i=1}^q \alpha_i + \sum_{i=1}^p \beta_i < 1$. Generally, we can combine the ARIMA model and GARCH model, such as AR(1)/GARCH(1,1) as follow $Y_t - \mu = \phi(Y_{t-1} - \mu) + a_t$; $a_t = \epsilon_t \sigma_t$; $\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$

3. Empirical Example

We still use the daily price of NASDAQ Composite from 01-Jan-2001 to 01-Jan-2015 as time series data to do the research and we mainly study the log-return.

3.1 Model Specification

Autocorrelation of time series is very essential to study. The graphs of the Autocorrelation Functions (ACF), Partial ACF of the original value, absolute value and squared value of log-return are investigated, which can be found in Picture 2.

The dashed horizontal lines are intended to give critical values for testing whether or not the autocorrelation coefficients are significantly different from zero. The pictures display some significant autocorrelations and hence provide some evidence that the daily SP500 returns are not constant and ARIMA model need to be

Table 1 The Sample EACF of Log-Return of NASDAQ Composite taken into consideration.

AR/MA													
0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	x	x	o	o	o	o	o	o	o	o	o	x	x
1	x	x	o	o	o	o	o	o	o	o	o	x	x
2	x	o	o	o	o	o	o	o	o	o	o	x	o
3	x	o	x	o	x	o	o	o	o	o	o	x	o
4	x	x	x	x	x	o	o	o	o	o	o	o	o
5	x	x	x	x	x	o	o	o	o	o	o	o	o
6	x	o	x	x	o	o	o	o	o	o	o	o	o
7	x	x	x	x	o	o	o	o	o	o	o	o	o

By extended autocorrelation (EACF) method, we get table 1. Of course, the sample EACF will never be this clear-cut, from the table we can get that

ARMA(2,1) might be taken consideration.

We can also use Dickey-Fuller Unit-Root Test to test the stationary of the time series.

```
Augmented Dickey-Fuller Test
data: spreturn
Dickey-Fuller = -14.141, Lag order = 15, p-value = 0.01
alternative hypothesis: stationary
```

From the result, the p-value is smaller than 0.05, which means that we can significantly reject the null hypothesis.

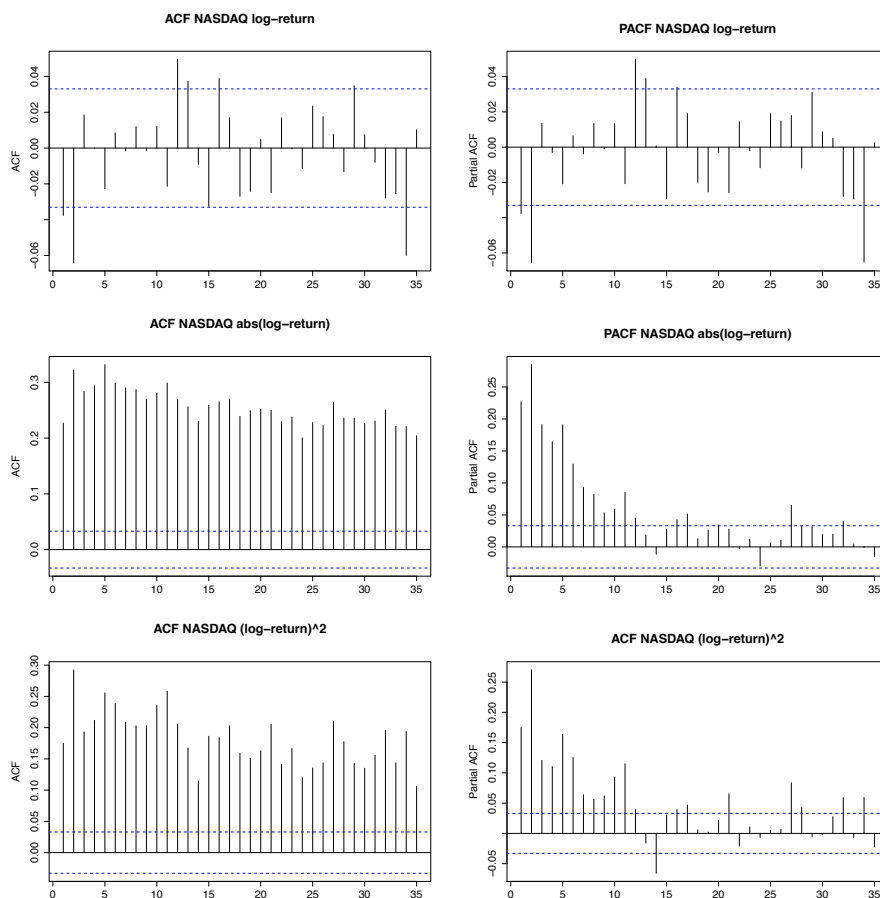
To select the parameters of ARIMA and GARCH model, we need use Akaike Information Criterion (AIC) to select the best model. In fitting ARIMA/GARCH model, the idea of parsimony is important in which the model should have as small parameters as possible yet still be capable of explaining the series.

Table 1 illustrates that the ARIMA(2,1,1) and GARCH(1,2) models can best fitted the NASDAQ Composite log-return time series.

Table 2 Model Selection

Model	AIC	Model	AIC
ARIMA(0,1,0)	-16760.57	GARCH(0,1)	-19579.5
ARIMA(1,1,0)	-17696.15	GARCH(0,2)	-20090.15
ARIMA(0,1,1)	-19282.28	GARCH(0,3)	-20283.28
ARIMA(1,1,1)	-19285.33	GARCH(0,4)	-20477.03
ARIMA(0,1,2)	-19286.08	GARCH(0,5)	-20617.52
ARIMA(2,1,0)	-18234.27	GARCH(0,6)	-20680.27
ARIMA(1,1,2)	-19290.64	GARCH(0,7)	-20717.4
ARIMA(2,1,1)	-19298.71	GARCH(0,8)	-20738.42
ARIMA(2,1,2)	-19297.33	GARCH(0,9)	-20746.57
ARIMA(0,2,0)	-12954.44	GARCH(0,10)	-20751.64
ARIMA(1,2,0)	-14850.49	GARCH(1,1)	-20777.02
ARIMA(0,2,1)	-16744.65	GARCH(1,2)	-20833.31
ARIMA(1,2,1)	-17679.17	GARCH(1,3)	-20751.98
ARIMA(0,2,2)	-19233.36		
ARIMA(2,2,0)	-15958.87		
ARIMA(1,2,2)	-18753.1		
ARIMA(2,2,1)	-18216.49		
ARIMA(2,2,2)	-19236.37		

picture 2 ACF/PACF of log-return



Here are the general information of ARIMA(2,1,1) and GARCH(1,2) models.

Table 3 Summary of model ARIMA(2,1,1)

```
Call:
arima(x = spreturn, order = c(2, 1, 1))
Coefficients:
      ar1      ar2      ma1
-0.0405 -0.0668 -0.9995
s.e.    0.0170  0.0170  0.0019
sigma^2 estimated as 0.0002409: log likelihood = 9653.35,
AIC = -19300.71, BIC=-19274.05
```

Table 4 Summary of model GARCH(1,2)

```
Call:
garch(x = res.arima211, order = c(1, 2))
Model:
GARCH(1,2)
Residuals:
      Min       1Q   Median       3Q      Max
-5.6575 -0.5666  0.1015  0.6534  4.4947
Coefficient(s):
      Estimate Std. Error t value Pr(>|t|)
a0 2.400e-06  4.548e-07  5.278 1.31e-07 ***
a1 2.911e-02  5.940e-03  4.901 9.55e-07 ***
a2 6.618e-02  1.042e-02  6.353 2.11e-10 ***
```

```

b1 8.921e-01  1.020e-02  87.450  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Diagnostic Tests:
  Jarque Bera Test
data:  Residuals
X-squared = 137, df = 2, p-value < 2.2e-16
  Box-Ljung test
data:  Squared.Residuals
X-squared = 3.8482, df = 1, p-value = 0.0498
    
```

And the formula below illustrate the ARIMA(2,1,1)/GARCG(1,2) model to fit the NASDAQ Composite log-return time series

$$Y_t - Y_{t-1} = -0.0405(Y_{t-1} - Y_{t-2}) - 0.0668(Y_{t-2} - Y_{t-3}) - 0.9995a_{t-1} + a_t$$

$$a_t^2 = \epsilon_t^2(2.400e - 06 + 2.911e - 02a_{t-1}^2 + 6.618e - 02 a_{t-1}^2 + 8.921e - 01\sigma_{t-1}^2)$$

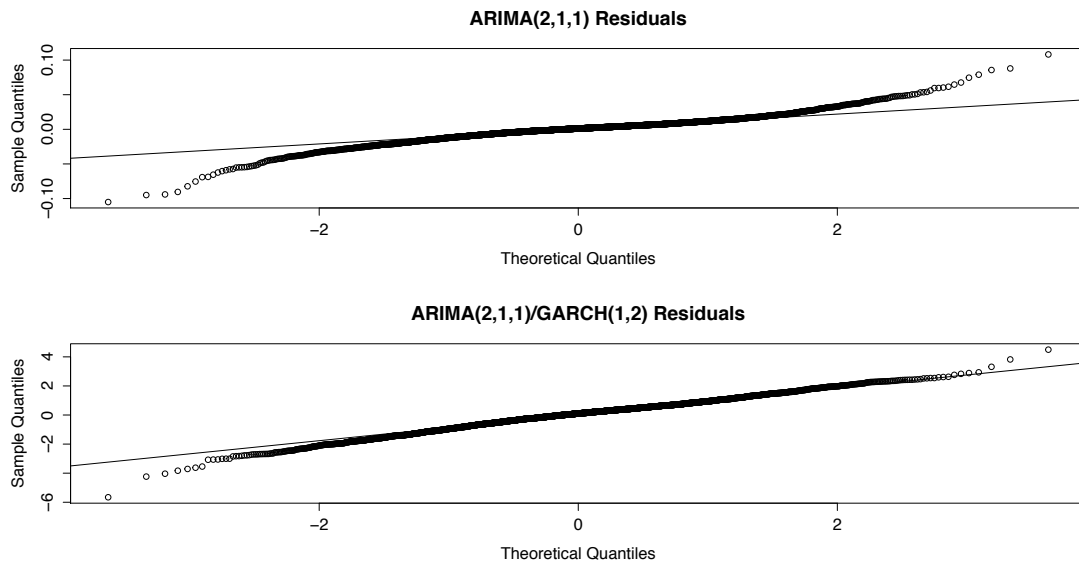
3.2 Model Fitting

Using a stationary GARCH(1,1) model as example, the Log-likelihood function is $L(\alpha_0, \alpha_1, \beta) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^n \left\{ \log(\sigma_{t-1|t-2}^2) + \frac{a_t^2}{\sigma_{t|t-1}^2} \right\}$. The Log-Likelihood of ARIMA(2,1,1) can be found in table 4. And from table 4, we can get the estimation of different parameters with the $\Pr(>|t|)$ values all close to zero.

3.3 Model Diagnostics

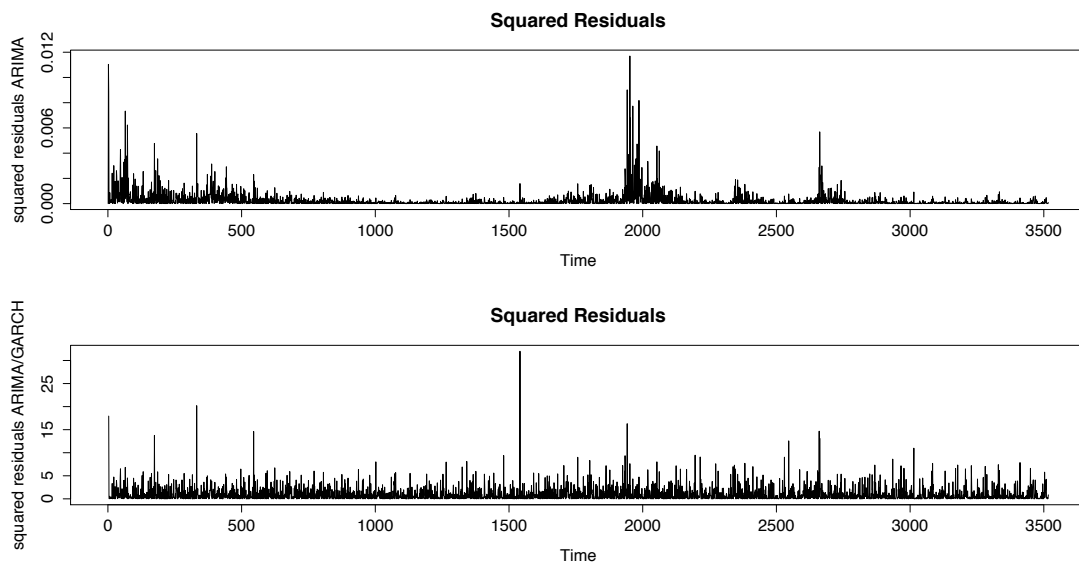
The procedure includes observing residual plot, checking Ljung-Box result and its ACF & PACF diagram. Because the normality assumption of the innovations can be explored by plotting the QQ plot, we can firstly observe the residuals graph and its QQ-Plot of ARIMA(2,1,1)/GARCH(1,2) models.

Picture 3 QQ-Plot of Residuals



The plots in picture 3 show that residuals seem to be roughly normally distributed although some points remain off the line. We can also get that the residuals of ARIMA/GARCH model are more normally distributed, compared to residuals of ARIMA model. In picture 4 is the residual graph of ARIMA(2,1,1) and GARCH(1,2) models. We can observe that the residual graphs do not have any significant lag, indicating ARIMA and mixed models are good models to represent the series. However, the distribution of mixed model is more even.

Picture 4 Residuals Graph

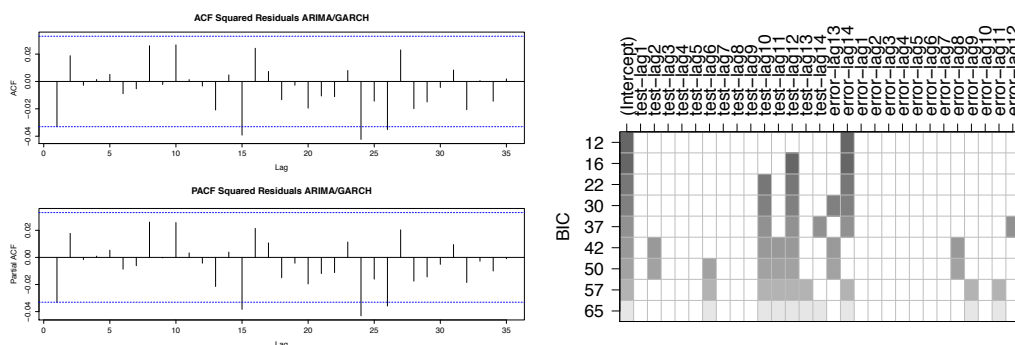


From the Box-Ljung Test, we can find that in the degree of freedom 6,12 and 18, the p-values are all larger than 0.05, which means that we cannot reject the hypothesis that the residuals are independent. So selected model seems to be appropriate.

Table 5 Box-Ljung test of ARIMA(2,1,1)/GARCH(1,2)

```
data: argar.res
X-squared = 7.5039, df = 6, p-value = 0.2767
X-squared = 14.08, df = 12, p-value = 0.2956
X-squared = 18.902, df = 18, p-value = 0.3979
```

Picture 5 ACF/PACF and ARMA-subset plot of ARIMA/GARCH model



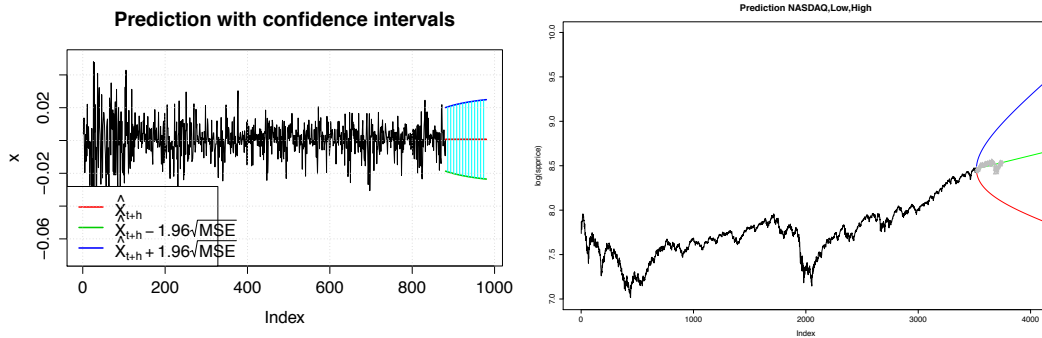
The ACF and PACF plot show that the residuals have no trend or pattern and except for lag 1, 15 and 24, residuals of most lags are below the critical values, showing that the autocorrelation coefficients are not significantly different from zero.

However, the ARMA-subsets plot show that the mixed model is still not that perfect, as the residuals of lag 14 rather than 1 or 2 has the lowest BIC value.

3.4 Prediction

We can use the ARIMA(2,1,1)/GARCH(1,2) model to predict the return and price of NASDAQ COMPOSITE index, which can be used to trade in the financial market. Moreover, using time series method to predict has already been widely used and accepted. In the right part of picture 6, the grey points are the real data which are not been used to develop the mixed model in this research and highly coincide with the prediction of ARIMA/GARCH model.

Picture 6 Prediction of Return and Price



4. Conclusion

On the one hand, ARMA model focuses on analyzing time series, which does not reflect recent changes of information. And it is more suitable to be applied for stationary series and in order to study non-stationary series, difference and transform (such as log transformation) need to be used. On the other hand, ARIMA is often used together with ARCH/GARCH model, which can better measure volatility of the series. By using the mixed models, people can forecast future values with up-to-date information.

In this research, we study the NASDAQ INDEX time series by using ARIMA/GARCH model with a few techniques to select parameters and tests to diagnose the model, which indicate that this mixed model can fit the series at some level. We also use this model to predict the return and price, which are coincide with real data.