2012-01-01

ARMA-GARCH Model applied to Exchange-Traded Funds

Rebecca Davis
University of Texas at El Paso, rdavis4@miners.utep.edu

ARMA-GARCH MODELS APPLIED TO EXCHANGE-TRADED FUNDS

REBECCA DAVIS

Department of Mathematical Sciences

APPROVED:

Maria Christina Mariani, Ph.D., Chair

Ming-Ying Leung, Ph.D.

Laura Serpa, Ph.D.

Benjamin C. Flores, Ph.D.

Interim Dean of Graduate School

to my

GOD and KING

with love

ARMA-GARCH MODELS APPLIED TO EXCHANGE-TRADED FUNDS

by

REBECCA DAVIS

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

Department of Mathematical Sciences

THE UNIVERSITY OF TEXAS AT EL PASO

August 2012

Acknowledgements

My greatest appreciation is to the Almighty God. Without Him I would not have made it this far, despite all the odds. I would also like to express my heart-felt gratitude to my advisor, Dr Maria Christina Mariani who has been an enormous help and suport for this past year. She has truely been the epitome of an advisor in every sense. Her love and concern is simply remarkable. I also wish to thank my thesis committee members, Dr. Ming-Ying Leung and Dr. Laura Serpa both at The University of Texas at El Paso, for their guided council, constant constructive critisms and direction on this thesis journey. I have learnt valuable lessons (both academically and non-academically) which I will forever cherish in my heart.

To the Statistics faculty of the mathematical sciences department, University of Texas at El Paso, I am forever indebted for the valuable amount of knowledge acquired these past two years. I am over-whelmed by their patience in the transmit of knowledge and also believing constantly in me as a student. I'm grateful to Dr. Joan Staniswalis who has truely been "in loco parentis". Her love, patience, kindness and selflesness constantly reminded me of my mom and comforted me even in a foreign land. I am also thankful to Dr. Hongling Yang, Maria Barraza-Rios, Maria Salayandia and Rachael Konrardy for the support they extended to me in diverse ways. Finally, I must thank my dear family and friends for all the support and encouragement throughout the program.

Abstract

In this paper, time-varying volatility of some of the leading exchange-traded funds are studied. The ARMA mean equation with GARCH errors is used to model the series correlations and the conditional heteroscadesticity in the asset returns. The conditional distributions of the standardized residuals are assumed to be skew-generalized error distribution. The high kurtosis and fat tail of the returns, were captured in all the data by fitting an ARMA-GARCH model with the conditional distribution of, skew-generalized error distribution. Furthermore, the sample cross-correlations of these significant exchange-traded funds and the corresponding financial indices they mimic were computed. The empirical conclusion was that, the exchange-traded funds have statistical behavior similar to that of the corresponding financial indices that they mimic.

Keywords: Exchange-Traded Funds, GARCH, Generalized Error Distribution

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Volatility measures the likely magnitude of price changes over a given period and is expressed as a percentage of the underlying market price. It is the most important factor in option pricing. Volatility of the returns of an asset is a measurement of the risk for investment and it provides important information for making good decisions. There are two basic types of price volatility: Historical volatility and Implied volatility. Historical volatility, as its name suggests, is the range that prices have traded in over a given period in the past. Historical volatility allows us to see how prices have behaved under known market conditions. From this, we may be able to build a confidence level to help us in assessing the predictability of current situations. Implied volatility is the range that prices are expected to trade over a given period in the future. Implied values are calculated by inputting option premiums into an option pricing model (see [5], [18]).

The standard Black-Scholes model for option pricing, has an assumption that the volatility is constant. The dissagreement between the Black-Scholes option prices and the market-traded option prices, the smile curve, can be accounted for by a stochastic volatility model. Modeling volatility as a stochastic process is motivated a priori by empirical studies of stock price returns in which volatility is observed to exhibit random characteristics. Since the introduction of ARCH models by Engle (1982), there has been a veritable explosion of papers analyzing models of changing volatility. Some of the more popular variants of models of changing volatility have proved to be various forms of GARCH models. In these models, the volatility process is stochastic and is modeled to be dependent upon both the past volatility and past innovations. These models have been used in many applications of

stock return data, interest rate data, foreign exchange data etc.

A type of security that has grown increasingly popular in recent years is the exchange-traded fund (ETF). ETFs are securities that closely resemble index funds, but can be bought and sold throughout the day, purchased on margin, or even sold short, just like common stocks. These investment vehicles allow investors a convenient way to purchase a broad basket of securities in a single transaction. Essentially, ETFs offer the convenience of a stock along with the diversification of a mutual fund. ETFs have been very attractive to market players ever since the first one (SPDR) was born in 1993, however most of the theoretical and empirical research have not paid enough attention to it.

The empirical probability distributions for financial asset returns always exhibit some characteristics which are called stylized facts (Tavares, et al., 2008): One of them is volatility clustering or persistence: large changes tend to followed by large changes, and small changes tend to followed by small changes are observed in the asset returns. Another is that fat tail exits in the probability distribution of the assets returns that the kurtosis exceeds the value of the normal distribution which is 3. This fat tail phenomenon is referred to as excess kurtosis, and the returns time series which exhibit fat tail are often called leptokurtic. Excess kurtosis may be as a result of volatility clustering, which is a type of heteroscedasticity or the presence of non-Gaussian asset returns distribution (Bai, et al., 2003). Last but not the least, financial asset returns have asymmetric distributions, which is also called leverage effect (Black, 1976). This means that, volatility increases less when the change is positive than when the change is negative. That's why good news have lesser impact on the prices of stocks than bad news.

In the initial assumption in ARCH model and GARCH model, the conditional distribution of the innovation was Gaussian. However, it is not adequate to capture the skewness and fairly high kurtosis present in the financial time series. As a result, GARCH models with

non-Gaussian errors are mostly required to capture the observed fat-tailed behavior in asset returns. The GARCH model with student's t-distribution as conditional distribution was first introduced by Bollerslev (1987). The Generalized error distribution(GED) was suggested to be used in GARCH models by Nelson(1991). Further more, due to the skewness observed in the asset returns, which the GARCH model with normal conditional distribution cannot capture, the skewed generalized error distribution is used as the conditional distribution of the ARMA-GARCH model in this research.

Based on our objective, this thesis is structured as follows. Chapter 2 provides a brief description of the ARMA-GARCH model with Gaussian distribution and the Generalized error distribution. Chapter 3 is made up of the analysis of data and chapter 4 gives the observations and discussion of findings and suggestions.

Chapter 2

Models

In this chapter, we provide definition of some important characteristics in time series and a brief explanation of the ARMA-GARCH model. This Chapter also talks about the alternative non-gaussian conditional distributions that can be used for the GARCH model.

2.1 Stationarity

One of the most important assumptions to make about the structure of a stochastic process before making statistical inferences is that of stationarity. The basic idea of stationarity is that the probability laws that govern the behaviour of the process do not change over time (see [4]). If a process is strictly (or strongly) stationary, the joint probability distribution remains the same when shifted in time. A stochastic process is said to be weakly (or second-order) stationary if it has a fixed mean and a constant variance. If a series is stationary, the magnitude of the autocorrelation attenuates fairly rapidly, whereas if the series is nonstationary or integrated, the autocorrelation diminishes gradually over time (see [19]).

Differencing is the easiest way to make a nonstationary mean stationary. In real life time series, there might be possible shifts in both the mean and the dispersion over time. The mean may be edging upwards, and the variability may be increasing. If the mean is changing, the trend is removed by differencing once or twice. If the variability is changing, the process may be made stationary by logarithmic transformation. In financial literature, the difference of the natural logarithms (called the returns) is the transformation usually applied. There are other family of transformations, the power transformations, which were

introduced by Box and Cox (1964).

2.2 Stationary Time Series Models

It is assumed that the series being studied have already been made stationary by means of the methods introduced in previous section. Basically, there are three kinds of models: the moving average model (MA), the autoregressive model (AR) and the autoregressive moving average model (ARMA). They are used to describe stationary time series.

Let Y_t denote the observed time series and e_t represent an unobserved white noise series, that is, a sequence of indentically indepently distributed random variable with zero mean. A series Y_t such that

$$Y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + ... + \theta_q e_{t-q} \tag{2.1}$$

is called a moving average of order q (MA(q)). Slutsky (1927) and Wold (1938) were the first to consider the moving average models. Autoregressive Processes are regresions on themselves. A series Y_t that satisfies

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + e_t \tag{2.2}$$

is called a pth order autoregressive process. When a series Y_t is partly autoregressive and partly moving average, we obtain a quite general time series process. If

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + ... + \theta_q e_{t-q} \tag{2.3}$$

we say that Y_t is a mixed autoregressive moving average process of order p and q respectively (ARMA(p,q))

2.3 Sample Autocorrelation Functions

One important tool for assessing the degree of dependence in observed data is the sample autocorrelation function (sample ACF) of the data. The sample autocorrelation function

of an obseerved series $Y_1, Y_2, ..., Y_n$, is given by

$$r_k = \frac{\sum_{t=k+1}^{n}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2}, k = 1, 2, ... \tag{2.4}$$

2.4 Heteroscedasticity

The return series of an asset is often a serially uncorrelated sequence with zero mean, even as it exhibits volatility clustering. This suggests that the conditional variance or conditional volatility of the return, denoted by $\sigma^2_{t|t-1}$ where the subscript $t-1$ signifying that the conditioning is upon returns through time $t-1$, is not constant. If the return is known, the squared return is an unbiased estimator of $\sigma^2_{t|t-1}$. When a series has a large squared returns, it may foretell a relatively volatile period.

2.4.1 The ARCH(1) Model

The ARCH model is a regression model with the conditional volatility, the response variable as a function of the past lags of the squared return, the covariates. The ARCH(1) model assumes that the return series r_t is generated as,

$$r_t = \sigma^2_{t|t-1}\epsilon_t \tag{2.5}$$

$$\sigma^2_{t|t-1} = \omega + \alpha r^2_{t-1} \tag{2.6}$$

where α and ω are unknown parameters, ϵ_t is a sequence of independently and identically distributed random variables each with zero mean and a unit variance, making the conditional variance of r_t equals $\sigma^2_{t|t-1}$. ϵ_t is independent of $r_{t-j}, j = 1, 2, ...$ The conditional variance in equation 2.5 is a latent variable (cannot be observed directly) so it is relevant to replace it by something observable. Substituting $\sigma^2_{t|t-1} = r^2_t - \eta_t$ into equation 2.6 gives,

$$r^2_t = \omega + \alpha r^2_{t-1} + \eta_t \tag{2.7}$$

where η_t is a serially uncorrelated series with a mean of zero and uncorrelated with past returns. Thus, the squared return satisfies the AR(1) model under the assumption of an

ARCH(1) model for the return series. If the return series is stationary with variance σ^2, then taking expectation on both sides of equation 2.7 results in

$$\sigma^2 = \omega + \alpha\sigma^2 \tag{2.8}$$

That is, $\sigma^2 = \omega/(1-\alpha)$ and hence $0 \leq \alpha < 1$ is a necessary and sufficient condition for the weak stationarity of the ARCH(1) model. The parameters ω and α can be restricted to be nonnegative since the squared return must be nonnegative.

2.4.2 GARCH Models

The future conditional variances only involves the most resent squared returns when using the ARCH(1) model for forcasting. However, by including all past square returns with lesser weight for more distant volatilities may improve the accuracy of forcasting. The generaliaed form of equation 2.6 is the ARCH(q) model which is given by,

$$\sigma^2_{t|t-1} = \omega + \alpha_1 r^2_{t-1} + \alpha_2 r^2_{t-2} + ... + \alpha_q r^2_{t-q} \tag{2.9}$$

q is called, the ARCH order. p lags of conditional variance was introduced into this model and the combined model is called the generalized autoregressive conditional heteroscedasticity, GARCH(p,q) model. p is called, the GARCH order. The GARCH(p,q) model can be expressed as

$$\sigma^2_t = \omega + \sum_{i=1}^{q} \alpha_i r^2_{t-i} + \sum_{i=1}^{p} \beta_i \sigma^2_{t-i|t-i-1} \tag{2.10}$$

Assuming that the return process is stationary and taking expectation of both sides of equation 2.10, we get the unconditional variance σ^2

$$\sigma^2 = \omega + \sigma^2 \sum_{i=1}^{max(p,q)} (\alpha_i + \beta_i)$$

$$\sigma^2 = \frac{\omega}{1 - \sum_{i=1}^{max(p,q)} (\alpha_i + \beta_i)} \tag{2.11}$$

The above equation is finite if $\sum_{i=1}^{max(p,q)} (\alpha_i + \beta_i) < 1$. This is the condition necessary and sufficient for the weak stationarity of a GARCH(p,q) model.

2.4.3 Extended GARCH Models

The GARCH model is based on the assumption that, the conditional mean of the time series is zero but this is not always the case in financial time series. In general, the conditional mean structure may be modeled by an ARMA model and the white noise term of the ARMA model modeled by a GARCH model. If Y_t is the time series, then the described model (ARMA(u,v)-GARCH(p,q) model) is given by equation 2.12 below.

$$\begin{aligned} Y_t = & \quad \phi_1 Y_{t-1} + ... + \phi_u Y_{t-u} + e_t + \theta_1 e_{t-1} + ... + \theta_v e_{t-v} \\ e_t = & \quad \sigma^2_{t|t-1} \epsilon_t \\ \sigma^2_{t|t-1} = & \quad \omega + \alpha_1 e^2_{t-1} + ... + \alpha_q e^2_{t-q} + \beta_1 \sigma^2_{t-1|t-2} + ... + \beta_p \sigma^2_{t-p|t-p-1} \end{aligned} \tag{2.12}$$

2.4.4 Alternative conditional distributions

To completely specify a GARCH-type model an assumption about the error distribution should be made. As it was mentioned before, it is more appropriate to assume that the errors have a heavy tailed distribution rather than Gaussian distribution. Beside the Gaussian conditional distribution of the error term ϵ_t, two alternative non-Gaussian distributions can be considered: Student-t distribution and generalized error distribution (GED).

Standardized Student's t-distribution for the standardized errors, $z_t = \epsilon_t/\sigma_t$ with v degree of freedom can be expressed as

$$f(z_t|v) = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{\Pi(v-2)}\Gamma(\frac{v}{2})} \left(1 + \frac{z_t^2}{v-2}\right)^{-\frac{v+1}{2}} \tag{2.13}$$

where $\Gamma(.)$ is the gamma function, and $v > 2$ is the shape parameter.

The Generalized error distribution can be expressed as

$$f(z_t|v) = \frac{v}{\lambda_v 2^{(v+1)/v}\Gamma(1/v)} exp\left\{-\frac{1}{2}\left|\frac{x}{\lambda_v}\right|^v\right\}, \lambda_v = \left[\frac{2^{-2/v}\Gamma(1/v)}{\Gamma(3/v)}\right]^{1/2} \tag{2.14}$$

where v is a positive shape parameter governing the thickness of the tail behaviour of the distribution. For $v = 1$, GED reduces to the double exponential distribution (Laplace distribution). For $v = 2$, GED reduces to the standard normal distribution and for $v \to \infty$ to the continuous uniform distribution.

Chapter 3

Model Illustration and Data Analysis

In this chapter, the behavior of the following Exchange-Traded Funds is studied: PowerShares QQQ^{TM} (ticker: QQQQ), S&P 500 SPDR (ticker: SPY), Dow Diamond (ticker: DIA) and compare them to the behavior of the indices they are designed to track, namely, the Nasdaq-100 (NDX), the S&P 500, and the Dow Jones Industrial Average (DJI) respectively. The ARMA-GARCH model is fitted and the unconditional volatility is estimated using this model. The relationship between each ETF and its index is examined for QQQQ from March 10, 1999 to November 8, 2011, SPY from January 29, 1993 to November 8, 2011 and DIA from January 20, 1998 to November 8, 2011. Daily ETF prices and index data are provided by the New York Stock Exchange and the NASDAQ Stock Exchange and are freely available for download from the Yahoo Finance web site.

3.1 Comparing NDX and QQQQ

The time series plots of both data are represented graphically in Figure 3.1 below.

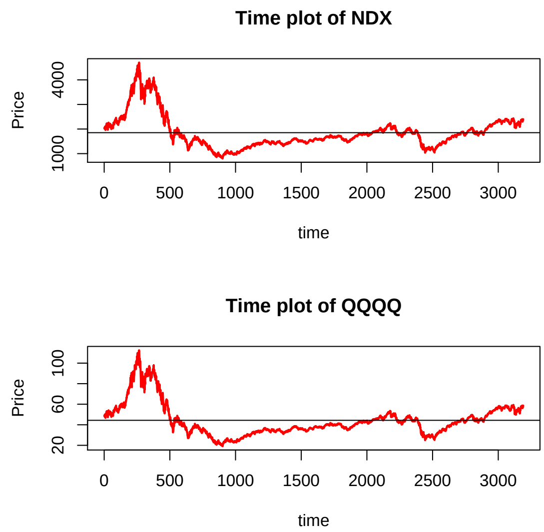

Figure 3.1: Time series plots of daily NDX and QQQQ prices

The time series plot looks similar except that, the price of the NDX index is higher than that of the Exchange Traded Fund, QQQQ. From the time plots of the series in Figure 3.1, there is no obvious trend in both series but they are not stationary. We will therefore transform the data and analyze the continuously compounded percentage rates of return (adjusted for dividends) that are calculated by taking the first differences of the logarithm of series (P_t is the adjusted price for each stock index at time t): $r_t = [ln(P_t) - ln(P_{t-1})] * 100$. The return is being multiplied by 100 in this work to reduce approximation errors in the data analysis. The transformed data and its distribution for the NDX index are represented graphically in Figure 3.2 and Figure 3.3 and that of QQQQ in Figure 3.4 and Figure 3.5 respectively.

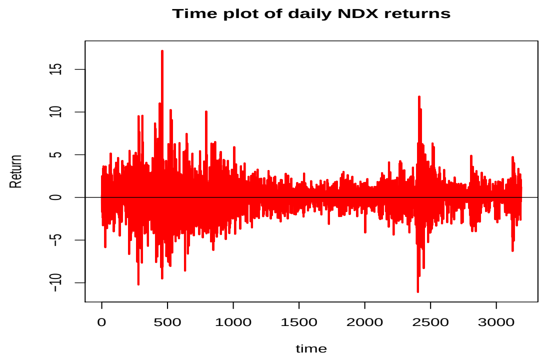

Figure 3.2: Time series plot of daily NDX returns

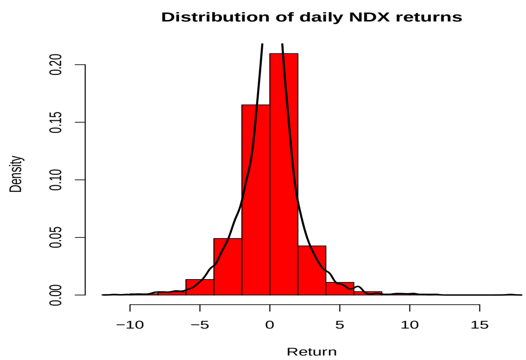

Figure 3.3: Distribution of daily NDX returns

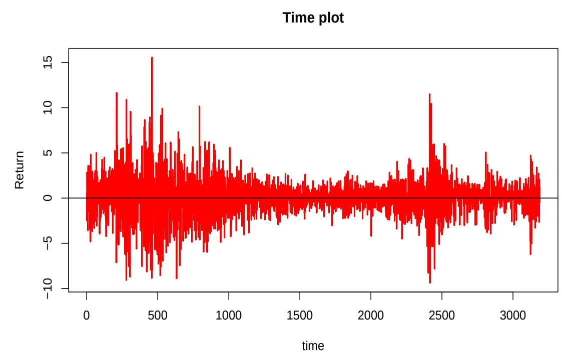

Figure 3.4: Time series plot of daily QQQQ returns

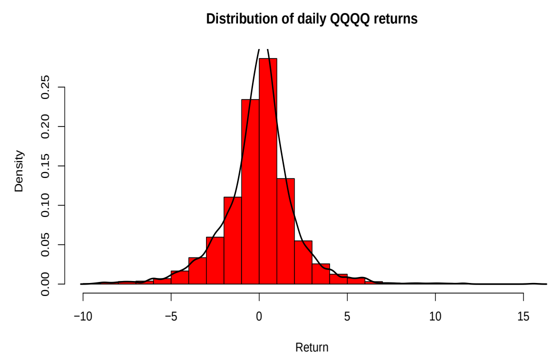

Figure 3.5: Distribution of daily QQQQ returns

By performing the Augmented Dickey-Fuller unit-root test on the NDX daily returns, the ADF test statistic is -14.1393 for lag order 14 (the software chose the lag order of 14) and a p-value of 0.01 is recorded. The same test on the QQQQ daily returns gives a ADF test statistic of -13.9173 and a p-value of 0.01. With stationarity as the alternative hypothesis, we reject the null hypothesis that there is a unit-root in the series. Hence we conclude that both series are stationary.

3.1.1 Model Specification

The estimated autocorrelation and the partial autocorrelation functions (ACF and PACF) for the NDX return series are illustrated in Figure 3.6 below and that for the QQQQ returns in Figure 3.7.

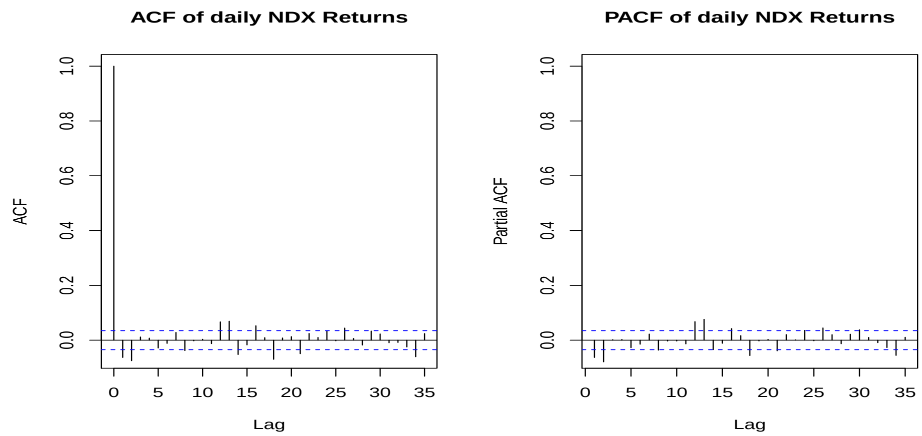

Figure 3.6: ACF and PACF of daily NDX returns

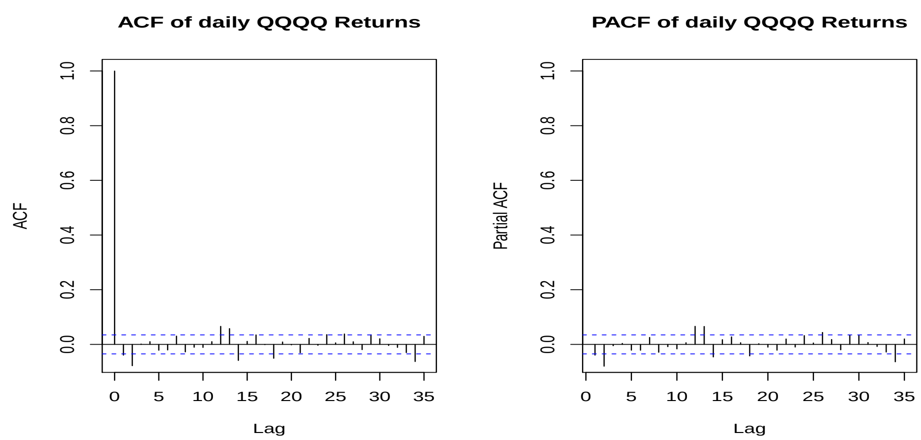

Figure 3.7: ACF and PACF of daily QQQQ returns

The critical values for testing whether the autocorrelation or partial autocorrelation coeffi-

cient is significantly different from zero or not are the dashed horiontal lines in Figure 3.6 and Figure 3.7. The ACF of both daily returns suggests an MA(2) model since the autoccorelation is significantly different from zero at lags 1 and 2. The PACF also gives a strong evidence to support an AR(2) model. However, non of these plots is very useful in detecting the order of ARMA models. We therefore take a look at the extended autocorrelation functions shown in Figure 3.8 and Figure 3.9 to see the mixed ARMA models that it suggests.

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

Figure 3.8: EACF of daily NDX returns

```
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  x
1 x x o o o o o o o o o  o  x  x
2 x x o o o o o o o o o  o  x  o
3 x o x o o o o x o o o  o  x  x
4 x x x x o o o o o o o  o  x  o
5 x x x x x o o o o o o  o  x  x
6 x o x x x o o o o o o  o  x  x
7 x o x x x x x o o o o  o  x  x
```

Figure 3.9: EACF of daily QQQQ returns

Figure 3.8 supports models like ARMA(2,0), ARMA(2,1) and ARMA(2,2) for the NDX returns. For the QQQQ returns, Figure 3.9 supports models like ARMA(0,2), ARMA(1,2) and ARMA(2,2). The software was used to fit all models suggested by the plots and the results are shown in Table 3.1 for the NDX returns and Table 3.2 for the QQQQ returns.

Table 3.1: ARMA(p,q) Model Selection for NDX returns

ARMA(p,q)	AIC
ARMA(0,2)	13766.06
ARMA(2,0)	13765.48
ARMA(2,1)	13767.40
ARMA(2,2)	13769.32

Table 3.2: ARMA(p,q) Model Selection for QQQQ returns

ARMA(p,q)	AIC
ARMA(0,2)	13642.41
ARMA(1,2)	13644.34
ARMA(2,0)	13642.19
ARMA(2,2)	13646.17

From Table 3.1 and Table 3.2 we assume an AR(2) model (or ARMA(2,0) model) for both returns, since it has the minimum AIC value and all of its estimated parameters are significant. Our model for the NDX and the QQQQ returns series are given in equation 3.1 and equation 3.2 respectively.

$$y_t = 0.006323 - 0.068092 y_{t-1} - 0.079194 y_{t-2} + e_t \tag{3.1}$$

$$y_t = 0.007132 - 0.042470 y_{t-1} - 0.079828 y_{t-2} + e_t \tag{3.2}$$

3.1.2 Model Diagnostics

To check whether the model assumptions are supported by the data, we take a look at Figure 3.10 and Figure 3.11 below.

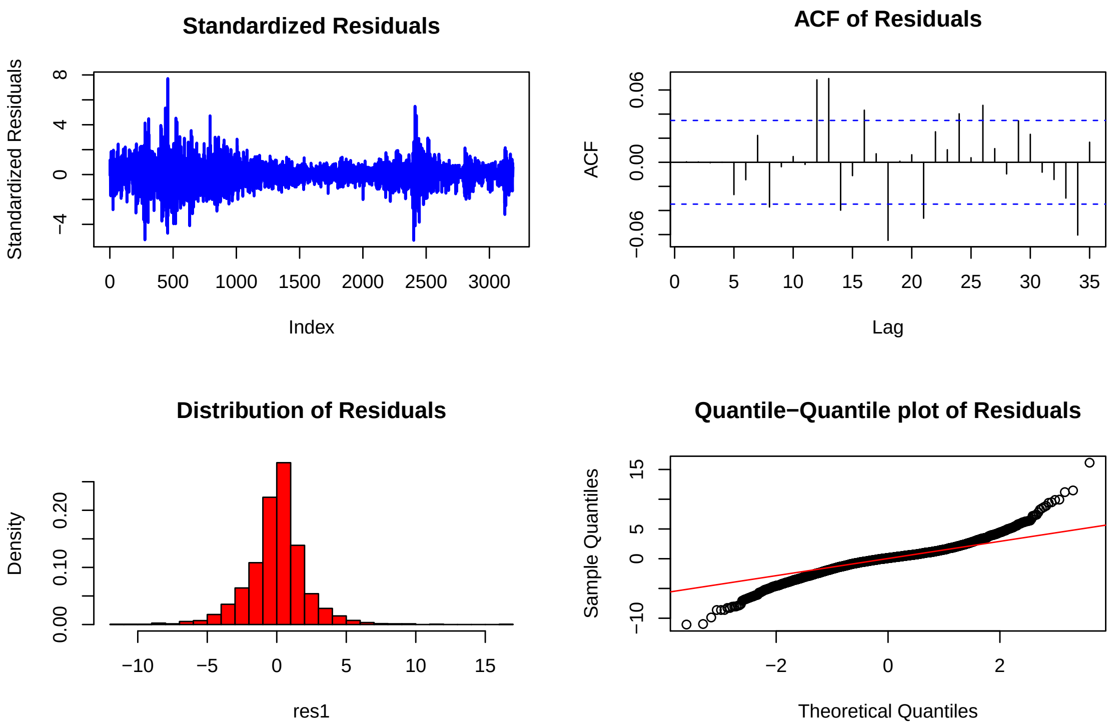

Figure 3.10: Diagnostics of ARMA(2,0) for daily NDX returns

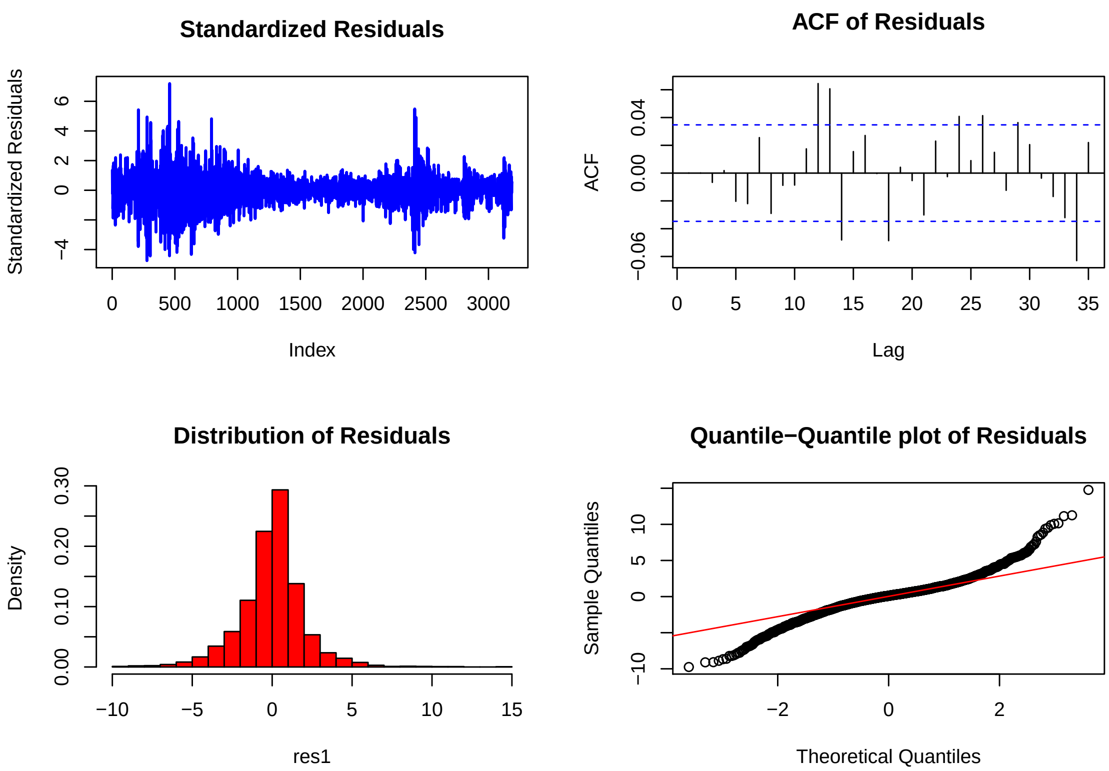

Figure 3.11: Diagnostics of ARMA(2,0) for daily QQQQ returns

Inspecting the time plot of the standardized residuals, there is evidence of volatility clustering, as large (small) fluctuations are usually succeeded by large (small) fluctuations. Looking at the ACF plot of the residuals of this model, there is no evidence of autocorrelation in the residuals for small lags. The normal QQ plot of the residuals has fat tails and hence shows departure from normality. This idea is supported by the histogram of the residuals which is slightly skewed to the left and has high kurtosis than a normal distribution.

3.1.3 Model of Heteroscedasticity

The analysis continue with a test for the presence of an ARCH in the residuals of the specified ARMA (2, 0) model. First we looked at ACF and PACF of the squared residuals. Figure 3.12 and Figure 3.13 show the ACF and PACF of the squared residuals for the NDX and QQQQ respectively. It appears that there is some dependency in the residuals.

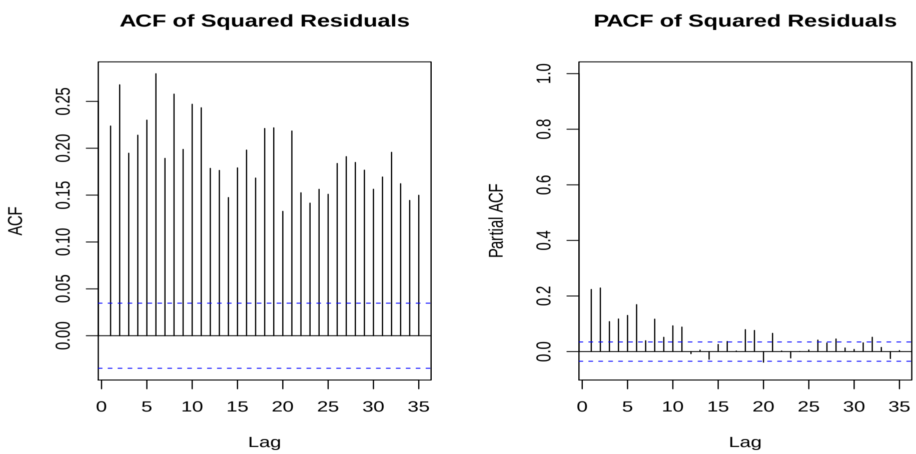

Figure 3.12: ACF and PACF of Squared Residuals of fitted ARMA(2,0) Model for NDX

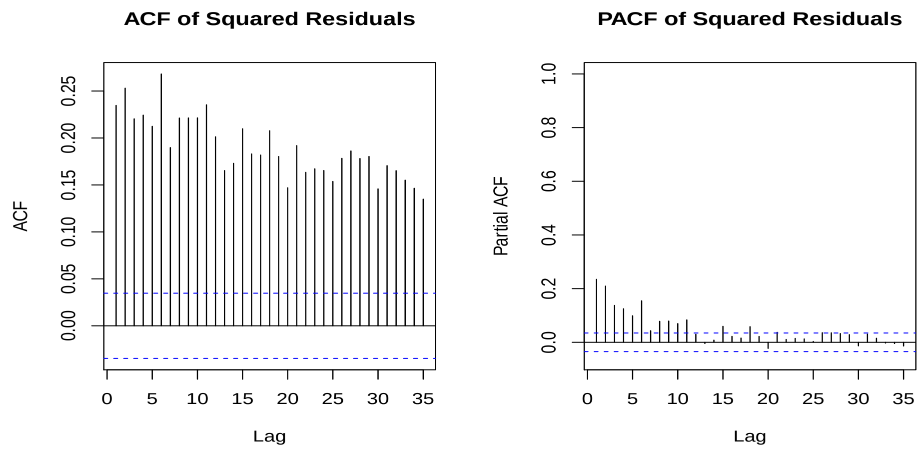

Figure 3.13: ACF and PACF of Squared Residuals of fitted ARMA(2,0) Model for QQQQ

Using the Box-Ljung test, we formally test for ARCH effect in the squared residuals. A test statistics of 1610.675 is recorded for NDX and 1655.524 for QQQQ (both are chi-squared with degrees of freedom of 10). These values indicate the presence of ARCH effect in both returns. The Jarque-Bera test for normality is used to confirm that the residuals of the fitted models are not normally distributed. As a result of these, we fit a GARCH model with a skewed generalized error distribution to the residuals of the fitted ARMA(2,0) model of the NDX returns. Some results obtained are shown in Table 3.3 and Table 3.4 below.

Table 3.3: ARMA-GARCH Model selection for NDX returns

Model	AIC	BIC
ARMA(2,0)-GARCH(1,1)	3.816181	3.831402
ARMA(2,0)-GARCH(1,2)	3.816767	3.833891
ARMA(2,0)-GARCH(2,1)	3.810776	3.827900
ARMA(2,0)-GARCH(2,2)	3.811094	3.830120

Table 3.4: ARMA-GARCH Model selection for QQQQ returns

Model	AIC	BIC
ARMA(2,0)-GARCH(1,1)	3.767220	3.782441
ARMA(2,0)-GARCH(1,2)	3.767787	3.784911
ARMA(2,0)-GARCH(2,1)	3.764233	3.781357
ARMA(2,0)-GARCH(2,2)	3.764770	3.783796

The ARMA(2,0) mean with GARCH(2,1) variance is chosen for both returns since it has minimum AIC and minimum BIC values. From Figure 3.14 and Figure 3.15 below, it is clear that the GARCH(2,1) model is adequate for describing the heteroscedasticity of the series.

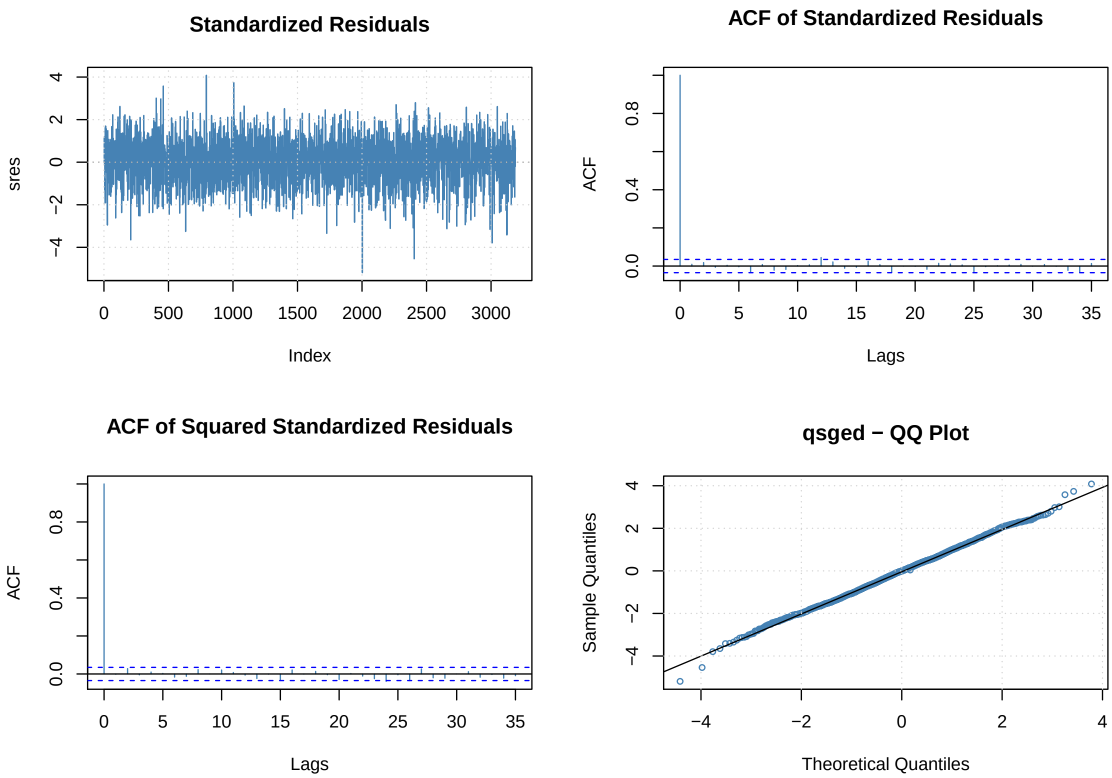

Figure 3.14: Diagnostics of ARMA(2,0)-GARCH(2,1) model for daily NDX returns

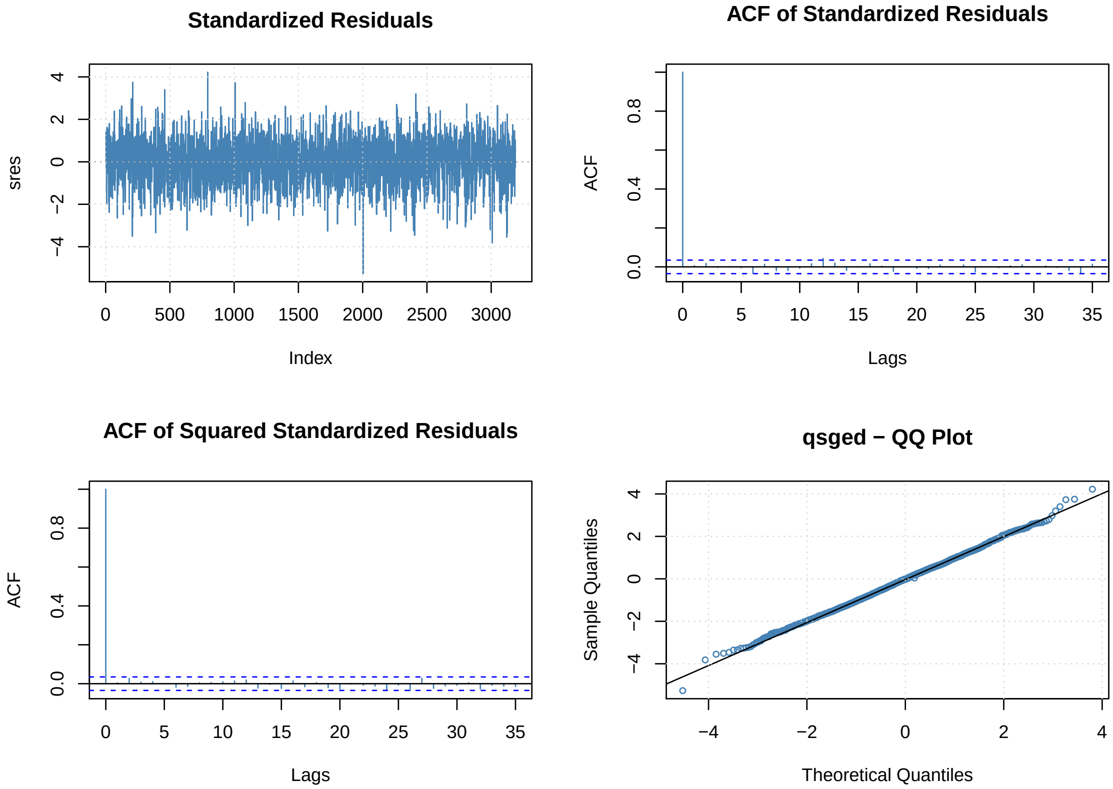

Figure 3.15: Diagnostics of ARMA(2,0)-GARCH(2,1) model for daily QQQQ returns

The maximum likelihood estimates of the ARMA(2,0)-GARCH(2,1) with skewed generalised distribution were obtained using the R software and the conditional mean and conditional variance equations for the NDX returns are shown in Equation 3.3 and for QQQQ returns in Equation 3.4.

$$\begin{aligned} y_t =& 0.05119 - 0.05764 y_{t-1} - 0.05629 y_{t-2} + e_t \\ \sigma^2_{t|t-1} =& 0.01722 + 0.09098 e^2_{t-2} + 0.9066 \sigma^2_{t-1|t-2} \end{aligned} \tag{3.3}$$

$$
\begin{aligned}
y_t =& 0.057707 - 0.031658 y_{t-1} - 0.055894 y_{t-2} + e_t \\
\sigma^2_{t|t-1} =& 0.016009 + 0.007247 e^2_{t-1} + 0.081766 e^2_{t-2} + 0.908827 \sigma^2_{t-1|t-2}
\end{aligned}
\tag{3.4}
$$

The coditional standard deviations are displayed in Figure 3.16 and Figure 3.17 below.

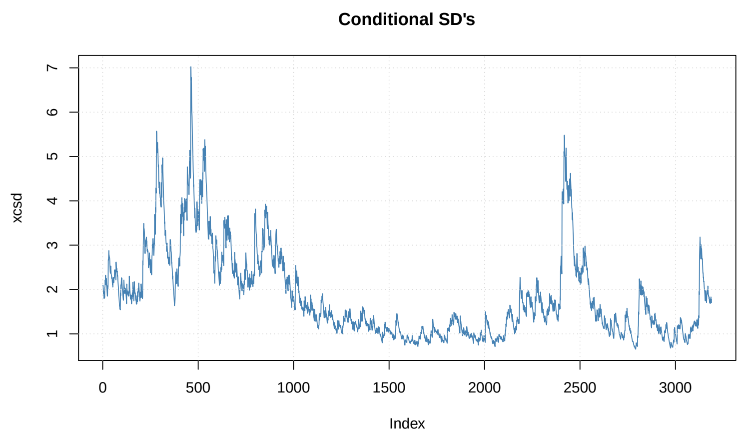

Figure 3.16: Coditional Standard Deviation for daily NDX returns

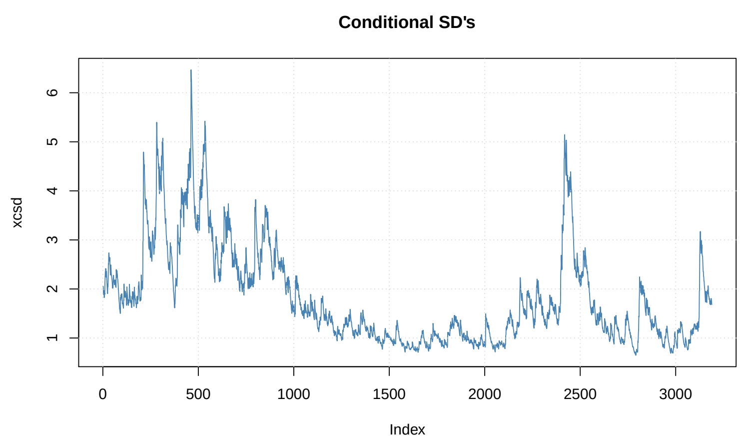

Figure 3.17: Coditional Standard Deviation for daily QQQQ returns

The fitted ARMA(2,0)-GARCH(2,1) model is weakly stationary and the unconditional (long-run) variance is given by equation 2.11. The standard deviation which is the volatility, is equal to the square root of the variance. This value is 2.6675% for NDX and 2.7224% for QQQQ. Remember this measure is calculated on the basis of the daily data.

3.2 Comparing S&P 500 and SPY

The time series plots of both data are represented graphically in Figure 3.18 below.

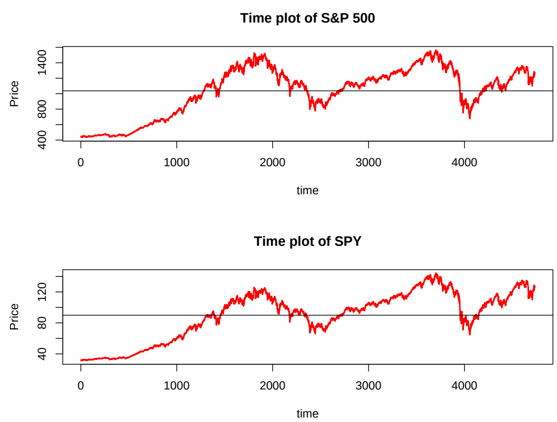

Figure 3.18: Time series plots of daily S&P 500 and SPY prices

The time series plots look similar but the price of the S&P 500 index is higher than that of the Exchange Traded Fund, SPY. From the time plots of the series in Figure 3.18, both series are not stationary. We will therefore transform the data and analyze the continuously compounded percentage rates of return (adjusted for dividends). The transformed data and its distribution for the S&P 500 index are represented graphically in Figure 3.19 and Figure 3.20 and that of SPY in Figure 3.21 and Figure 3.22 respectively.

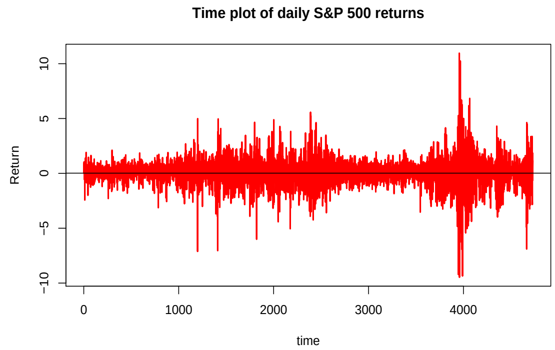

Figure 3.19: Time series plot of daily S&P 500 returns

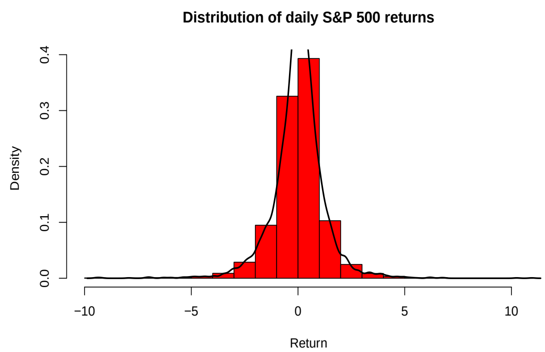

Figure 3.20: Distribution of daily S&P 500 returns

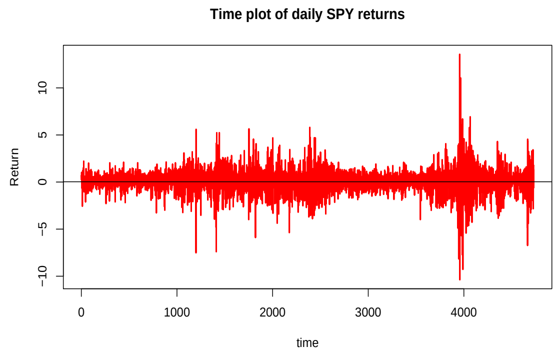

Figure 3.21: Time series plot of daily SPY returns

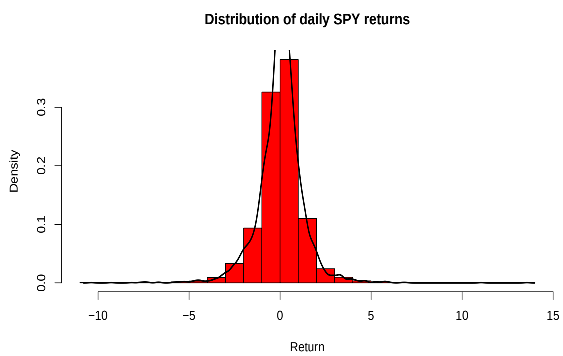

Figure 3.22: Distribution of daily SPY returns

By performing the Augmented Dickey-Fuller unit-root test on the S&P 500 daily returns, the ADF test statistic is -16.4297 for lag order 16 and a p-value of 0.01 is recorded. The same test on the SPY daily returns gives a ADF test statistic of -16.6678 and a p-value of 0.01. With stationarity as the alternative hypothesis, we reject the null hypothesis that there is a unit-root in the series. Hence we conclude that the both series are stationary.

3.2.1 Model Specification

The estimated autocorrelation and the partial autocorrelation functions (ACF and PACF) for the S&P 500 return series are illustrated in Figure 3.23 below and that for the SPY returns in Figure 3.24.

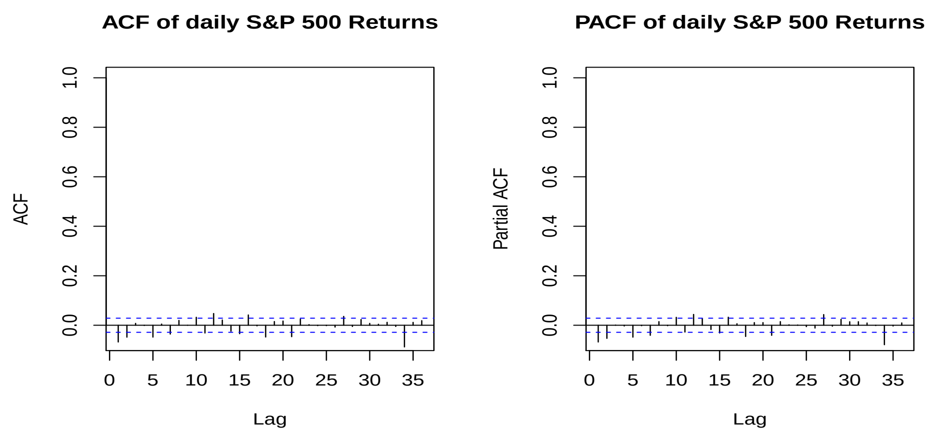

Figure 3.23: ACF and PACF of daily S&P 500 returns

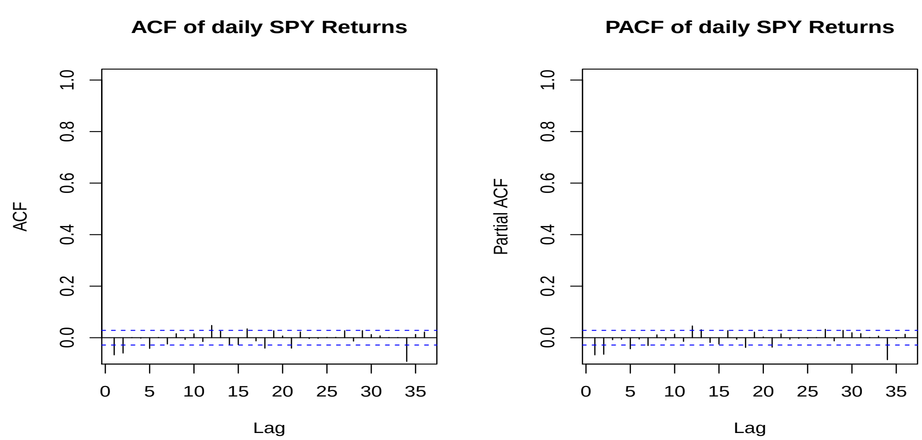

Figure 3.24: ACF and PACF of daily SPY returns

These plots suggest models like MA(2), MA(5),AR(2), AR(5), or even higher orders of these

models. Taking a look at the extended autocorrelation functions shown in Figures 3.25 and 3.26 does not make things easy.

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

Figure 3.25: EACF of daily S&P 500 returns

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

Figure 3.26: EACF of daily SPY returns

Figure 3.25 supports models like ARMA(6,6), ARMA(5,5), ARMA(0,5), ARMA(1,5), ARMA(2,5) and others for the S&P 500 returns. For the SPY returns, Figure 3.26 supports similar models except the ARMA(6,6) model. The best subset ARMA approach can also be used to specify models for these data. The R software was used to fit all suggested models and the results from some good models are shown in Table 3.5 for the S&P 500 returns and Table 3.6 for the SPY returns.

Table 3.5: ARMA(p,q) Model Selection for S&P 500 returns

ARMA(p,q)	AIC
ARMA(0,2)	15357.13
ARMA(0,5)	15353.03
ARMA(1,5)	15347.33
ARMA(2,5)	15345.90
ARMA(3,5)	15348.49
ARMA(2,0)	15354.68
ARMA(5,0)	15354.09

Table 3.6: ARMA(p,q) Model Selection for SPY returns

ARMA(p,q)	AIC
ARMA(0,2)	15357.84
ARMA(0,5)	15536.40
ARMA(1,5)	15538.38
ARMA(2,5)	15536.10
ARMA(3,5)	15537.81
ARMA(2,0)	15538.79
ARMA(5,0)	15537.59

From Table 3.5 and Table 3.6 we assume an ARMA(2,5) model for both returns, since it has the minimum AIC value among the good models. Our models for the S&P 500 and the SPY return series are given in Equation 3.5 and Equation 3.6 respectively.

$$\begin{aligned} y_t =& 0.01738 - 0.4085 y_{t-1} + 0.36434 y_{t-2} + e_t + 0.33365 e_{t-1} - 0.44656 e_{t-2} \\ &+ 0.0133 e_{t-3} + 0.01079 e_{t-4} - 0.06032 e_{t-5} \end{aligned} \tag{3.5}$$

$$
\begin{aligned}
y_t =& 0.01878 - 0.16448 y_{t-1} + 0.52858 y_{t-2} + e_t + 0.0915 e_{t-1} - 0.60424 e_{t-2} \\
&+ 0.02756 e_{t-3} + 0.03175 e_{t-4} - 0.04153 e_{t-5}
\end{aligned} \tag{3.6}
$$

3.2.2 Model Diagnostics

To check whether the model assumptions are supported by the data, we take a look at Figure 3.27 and Figure 3.28 below.

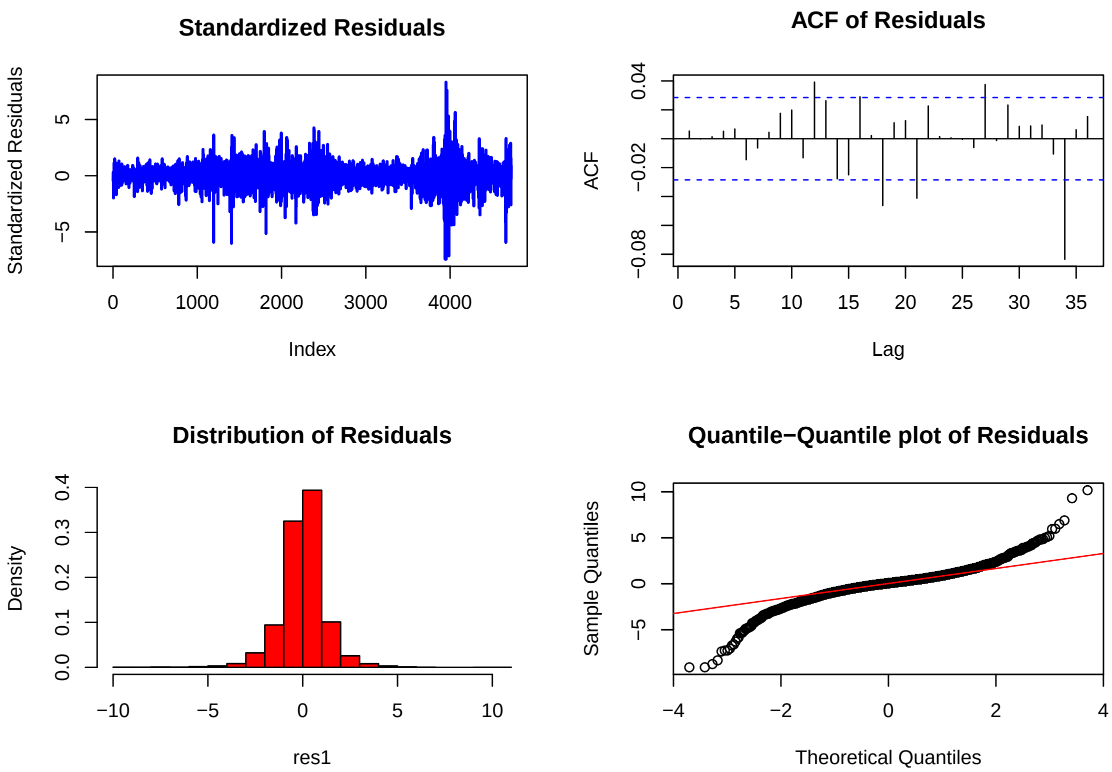

Figure 3.27: Diagnostics of ARMA(2,5) for daily S&P 500 returns

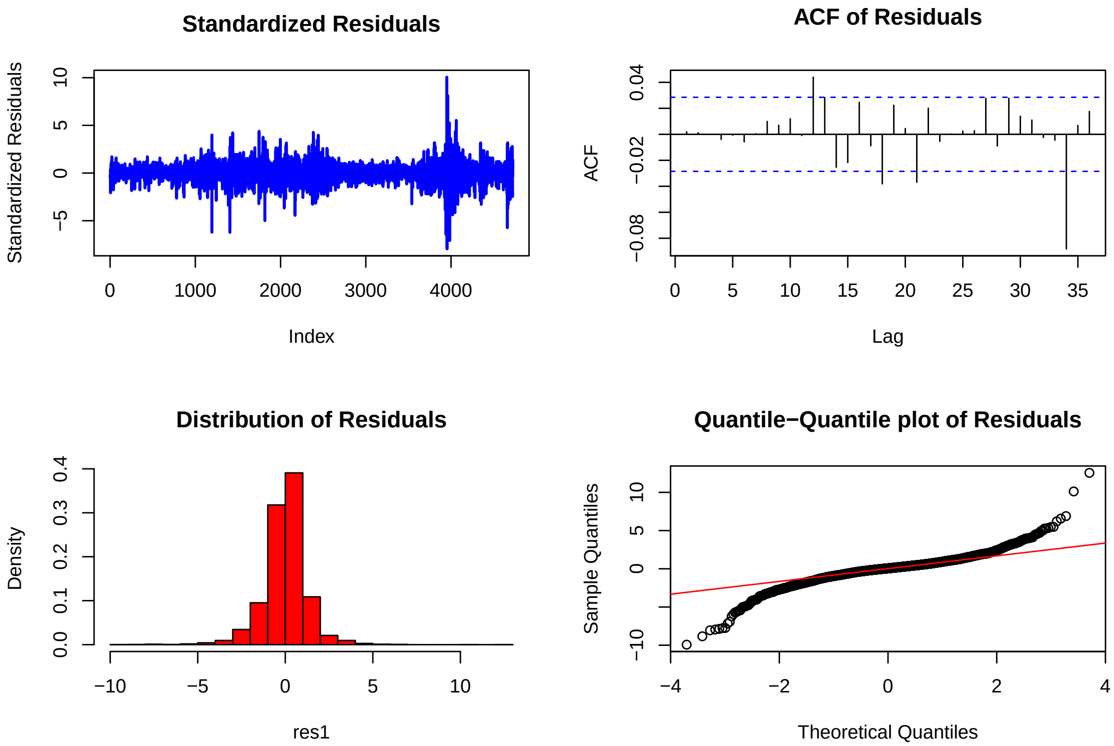

Figure 3.28: Diagnostics of ARMA(2,5) for daily SPY returns

Inspecting the time plot of the standardized residuals, there is evidence of volatility clustering, as large (small) fluctuations are usually succeeded by large (small) fluctuations. Looking at the ACF plot of the residuals of this model, the autocorrelation is significant at some lags greater than 11. The normal QQ plots of the residuals have fat tails and hence show departure from normality. This idea is supported by the histogram of the residuals which is slightly skewed to the left and has high kurtosis than a normal distribution.

3.2.3 Model of Heteroscedasticity

The analysis continue with a test for an ARCH effect presence in the residuals of the specified model. First we looked at ACF and PACF of the squared residuals. Figure 3.29 and Figure 3.30 show the ACF and PACF of the squared residuals for the S&P 500 and

SPY respectively. It appears that there is some dependency in the residuals.

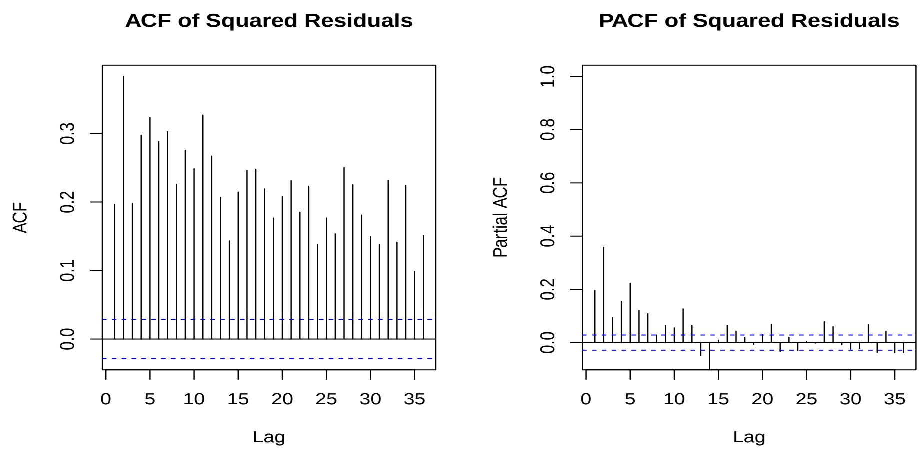

Figure 3.29: ACF and PACF of Squared Residuals of fitted ARMA(2,5) Model for S&P 500

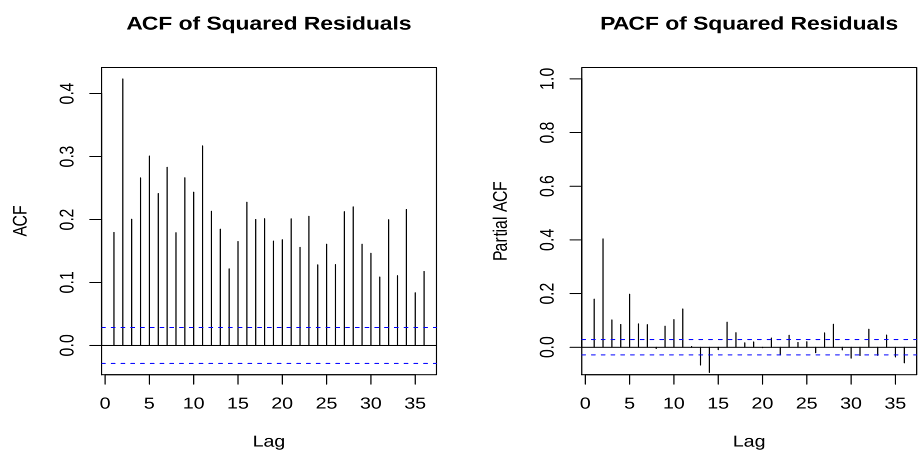

Figure 3.30: ACF and PACF of Squared Residuals of fitted ARMA(2,5) Model for SPY

The Box-Ljung test statistics for ARCH effect is 3698.922 for S&P 500 and 3375 for SPY (both are chi-squared with 10 degrees of freedom). These values indicate the presence of ARCH effect in both returns. The Jarque-Bera test for normality can be used to confirm

that the residuals of the fitted models are not normally distributed. As a result of these, we fit a GARCH model with a skewed generalized error distribution to the residuals of the fitted ARMA(2,5) models returns. Some results obtained are shown in Table 3.7 and Table 3.8 bolow.

Table 3.7: ARMA-GARCH Model selection for S&P 500 returns

Model	AIC	BIC
ARMA(2,5)-GARCH(1,1)	2.741209	2.758966
ARMA(2,5)-GARCH(1,2)	2.741742	2.760864
ARMA(2,5)-GARCH(2,1)	2.737715	2.756837
ARMA(2,5)-GARCH(2,2)	2.737541	2.758028

Table 3.8: ARMA-GARCH Model selection for SPY returns

Model	AIC	BIC
ARMA(2,5)-GARCH(1,1)	2.787660	2.805417
ARMA(2,5)-GARCH(1,2)	2.788193	2.807315
ARMA(2,5)-GARCH(2,1)	2.783969	2.803091
ARMA(2,5)-GARCH(2,2)	2.783687	2.804175

The ARMA(2,5) mean with GARCH(2,2) variance is chosen for both returns. This model has the minimum AIC but ARMA(2,5) mean with GARCH(2,1) variance has minimum BIC values. From the parameter estimates for both returns, the GARCH oder 2 is significant and as a result, we select the ARMA(2,5) mean with GARCH(2,2) variance model as the best model. From Figure 3.31 and Figure 3.32 below, it is clear that the GARCH(2,2) model is adequate for describing the heteroscedasticity of the series.

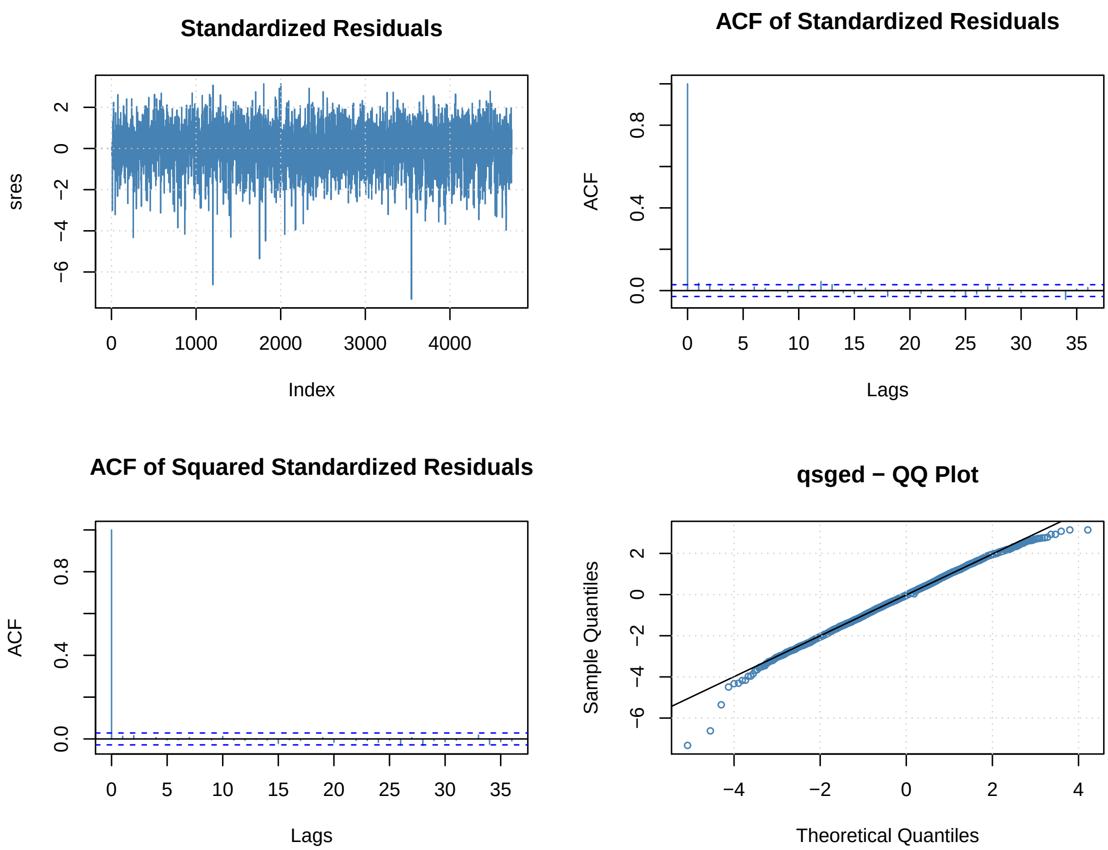

Figure 3.31: Diagnostics of ARMA(2,5)-GARCH(2,2) model for daily S&P 500 returns

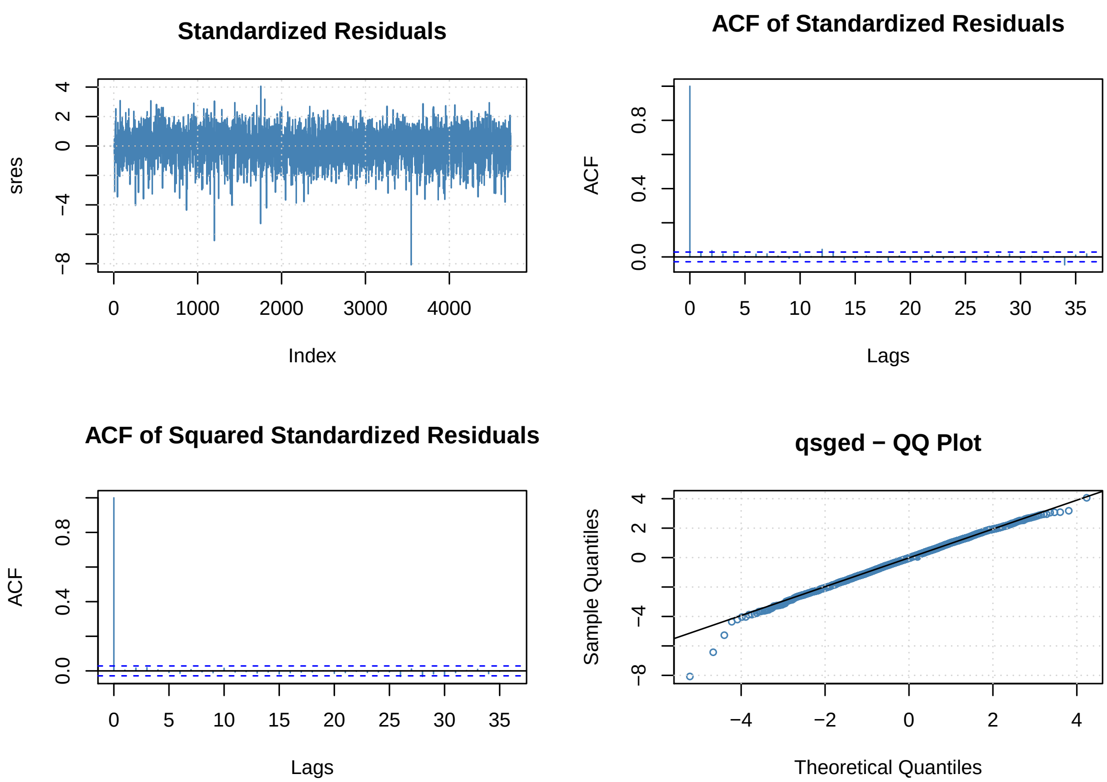

Figure 3.32: Diagnostics of ARMA(2,5)-GARCH(2,2) model for daily SPY returns

The maximum likelihood estimates of the ARMA(2,5)-GARCH(2,2) with skewed generalised distribution were obtained using the R software and the conditional mean and conditional variance equations for the S&P 500 returns are shown in Equation 3.7 and for SPY returns in Equation 3.8.

$$\begin{aligned} y_t =& 0.020206 + 0.943009 y_{t-1} - 0.378381 y_{t-2} + e_t - e_{t-1} + 0.380066 e_{t-2} \\ & - 0.006191 e_{t-3} - 0.002601 e_{t-4} - 0.037442 e_{t-5} \\ \sigma^2_{t|t-1} =& 0.012573 + 0.019145 e^2_{t-1} + 0.105027 e^2_{t-2} + 0.459621 \sigma^2_{t-1|t-2} + 0.409099 \sigma^2_{t-2|t-3} \end{aligned} \tag{3.7}$$

$$
\begin{aligned}
y_t =& 0.022219 + 0.931964 y_{t-1} - 0.340975 y_{t-2} + e_t - e_{t-1} + 0.3444 e_{t-2} \\
& - 0.017205 e_{t-3} + 0.006736 e_{t-4} - 0.038817 e_{t-5} \\
\sigma^2_{t|t-1} =& 0.01572 + 0.021473 e^2_{t-1} + 0.110353 e^2_{t-2} + 0.46253 \sigma^2_{t-1|t-2} + 0.39682 \sigma^2_{t-2|t-3} \quad (3.8)
\end{aligned}
$$

The coditional standard deviations are displayed in Figure 3.33 and Figure 3.34 below.

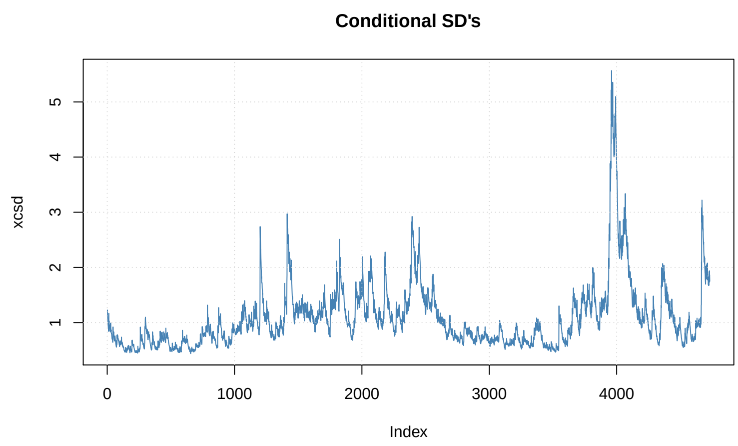

Figure 3.33: Coditional Standard Deviation for daily S&P 500 returns

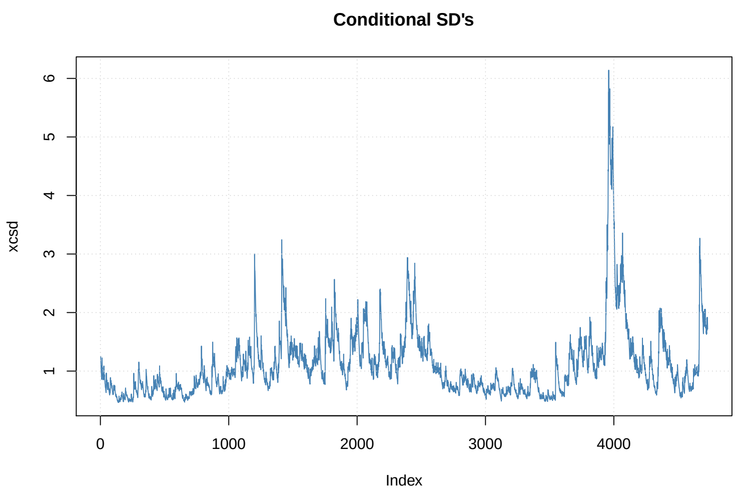

Figure 3.34: Coditional Standard Deviation for daily SPY returns

The fitted ARMA(2,5)-GARCH(2,2) model is weakly stationary and the daily volatility is 1.3300% for S&P 500 and 1.3347% for SPY.

3.3 Comparing DJI and DIA

The time series plots of both data are represented graphically in Figure 3.35 below.

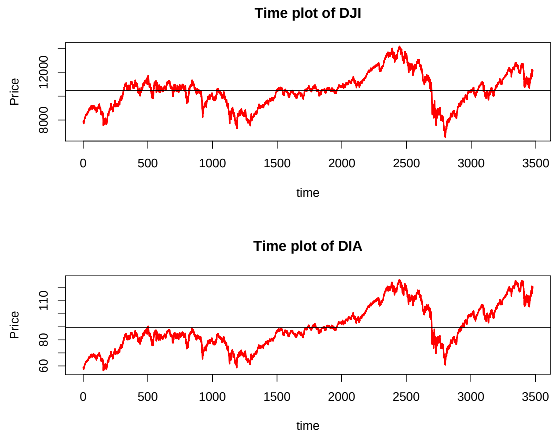

Figure 3.35: Time series plots of daily prices of DJI and DIA

The time series plots look similar except that, just like the other comparisons made in previous sections, the price of the stock index, DJI index is higher than that of the Exchange Traded Fund, DIA. From the time plots of the series in Figure 3.35, both series are not stationary. We will therefore transform the data and analyze the continuously compounded percentage rates of return (adjusted for dividends). Remember that, the return is being multiplied by 100 in this work to reduce approximation errors in the data analysis. The transformed data and its distribution for the DJI index are represented graphically in Figure 3.36 and Figure 3.37 and that of DIA in Figure 3.38 and Figure 3.39 respectively.

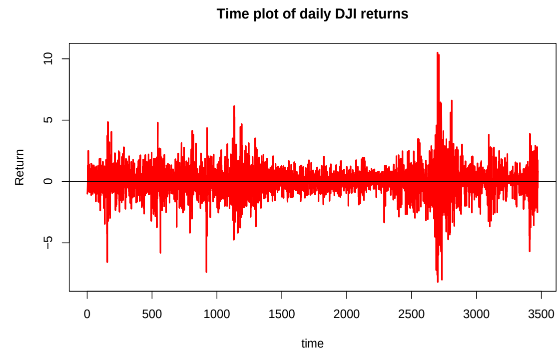

Figure 3.36: Time series plot of daily DJI returns

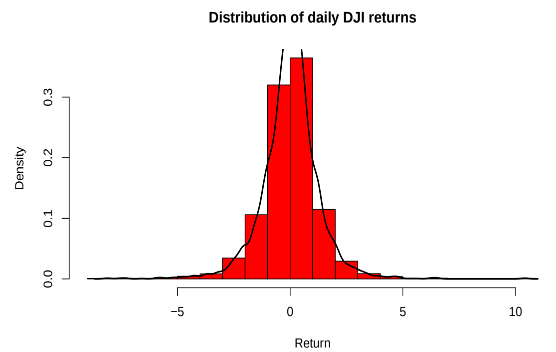

Figure 3.37: Distribution of daily DJI returns

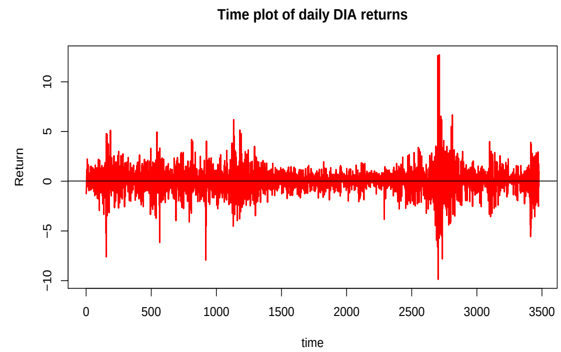

Figure 3.38: Time series plot of daily DIA returns

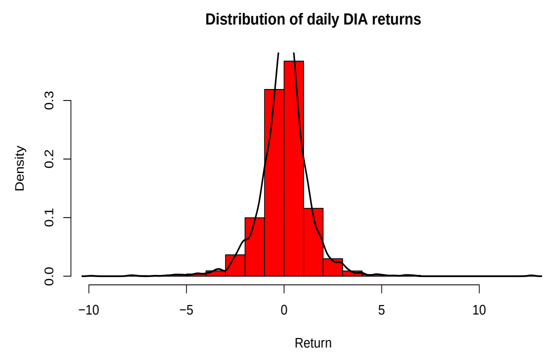

Figure 3.39: Distribution of daily DIA returns

By performing the Augmented Dickey-Fuller unit-root test on the DJI daily returns, the ADF test statistic is -14.6747 for lag order 15 and a p-value of 0.01 is recorded. The same test on the DIA daily returns gives a ADF test statistic of -14.6281 and a p-value of 0.01. We conclude that both return series are stationary.

3.3.1 Model Specification

The estimated autocorrelation and the partial autocorrelation functions (ACF and PACF) for the DJI return series are illustrated in Figure 3.40 below and that for the DIA returns

in Figure 3.41.

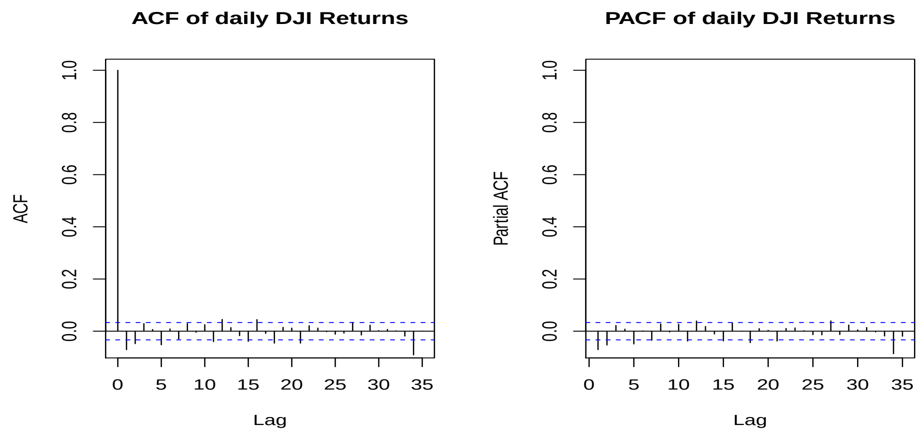

Figure 3.40: ACF and PACF of daily DJI returns

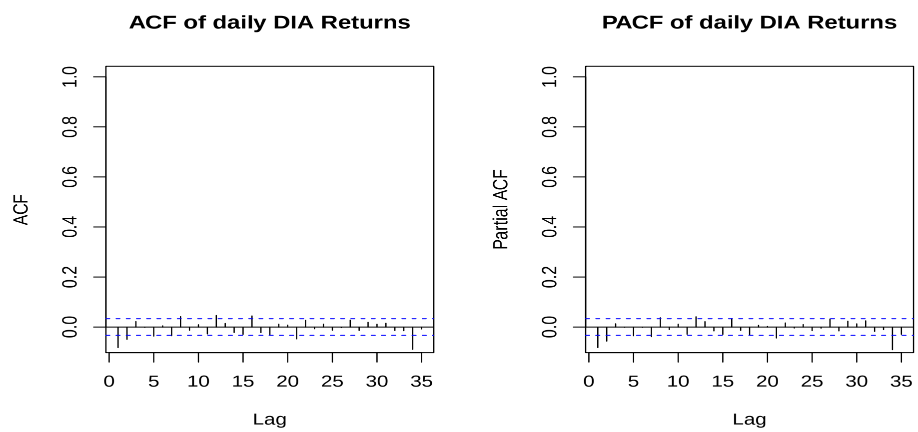

Figure 3.41: ACF and PACF of daily DIA returns

The ACF plot in Figure 3.40 suggests an MA(2) or MA(5) model. The PACF plot also gives a strong evidence to support an AR(2) or AR(5) model for the DJI returns. The ACF plot in Figure 3.41 suggests an MA(2) model and the PACF plot also gives a strong evidence to support an AR(2) model for the DIA returns. Since, non of these plots is very useful

in detecting the order of ARMA models, we take a look at the extended autocorrelation functions shown in Figure 3.42 and Figure 3.43 to see the mixed ARMA models that it suggests for each return.

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

Figure 3.42: EACF of daily DJI returns

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

Figure 3.43: EACF of daily DIA returns

Figure 3.42 supports models like ARMA(0,2), ARMA(0,5), ARMA(2,5), ARMA(3,5) and others for the DJI returns. For the DIA returns, Figure 3.43 supports models like ARMA(0,2), ARMA(2,3), ARMA(1,5), ARMA(2,5) and others. The best subset ARMA approach for specifying a model is also helpful. The software was used to fit all models suggested by the plots and the results are shown in Table 3.9 for the DJI returns and Table 3.10 for the DIA returns.

Table 3.9: ARMA(p,q) Model Selection for DJI returns

ARMA(p,q)	AIC
ARMA(0,2)	11513.81
ARMA(0,5)	11508.81
ARMA(2,0)	11512.61
ARMA(5,0)	11510.04
ARMA(2,5)	11501.11
ARMA(3,5)	11504.68

Table 3.10: ARMA(p,q) Model Selection for DIA returns

ARMA(p,q)	AIC
ARMA(0,2)	11682.03
ARMA(2,0)	11681.15
ARMA(2,3)	11686.75
ARMA(1,5)	11682.58
ARMA(2,5)	11674.98

From Table 3.9 and Table 3.10 we assume an ARMA(2,5) model for both returns, since it has the minimum AIC value. Our model for the DJI and the DIA returns series are given in Equation 3.9 and Equation 3.10 respectively.

$$\begin{aligned} y_t =& 0.006671 - 0.190898 y_{t-1} + 0.620222 y_{t-2} + e_t + 0.120485 e_{t-1} - 0.683934 e_{t-2} \\ &+ 0.067201 e_{t-3} + 0.04445 e_{t-4} - 0.068922 e_{t-5} \end{aligned} \tag{3.9}$$

$$
\begin{aligned}
y_t = & 0.01727 - 0.33914y_{t-1} + 0.51211y_{t-2} + e_t + 0.25767e_{t-1} - 0.59184e_{t-2} \\
& + 0.04891e_{t-3} + 0.03291e_{t-4} - 0.05765e_{t-5}
\end{aligned} \tag{3.10}
$$

3.3.2 Model Diagnostics

To check whether the model assumptions are supported by the data, we take a look at Figure 3.44 and Figure 3.45 below.

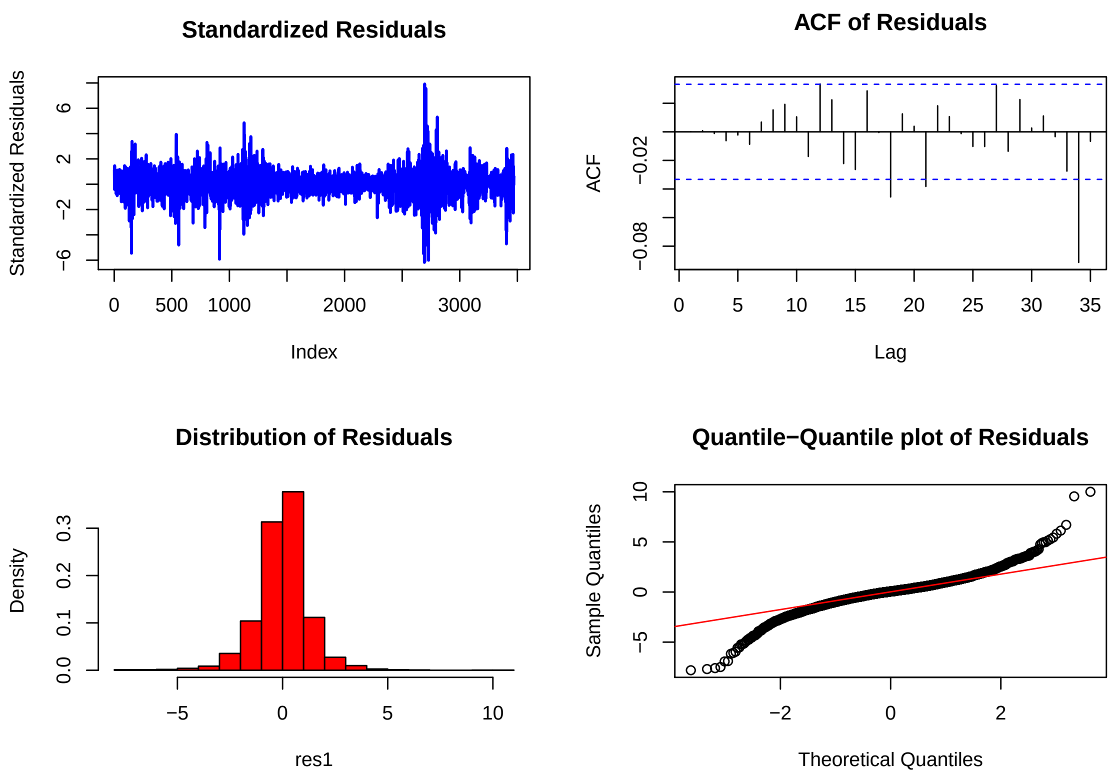

Figure 3.44: Diagnostics of ARMA(2,5) for daily DJI returns

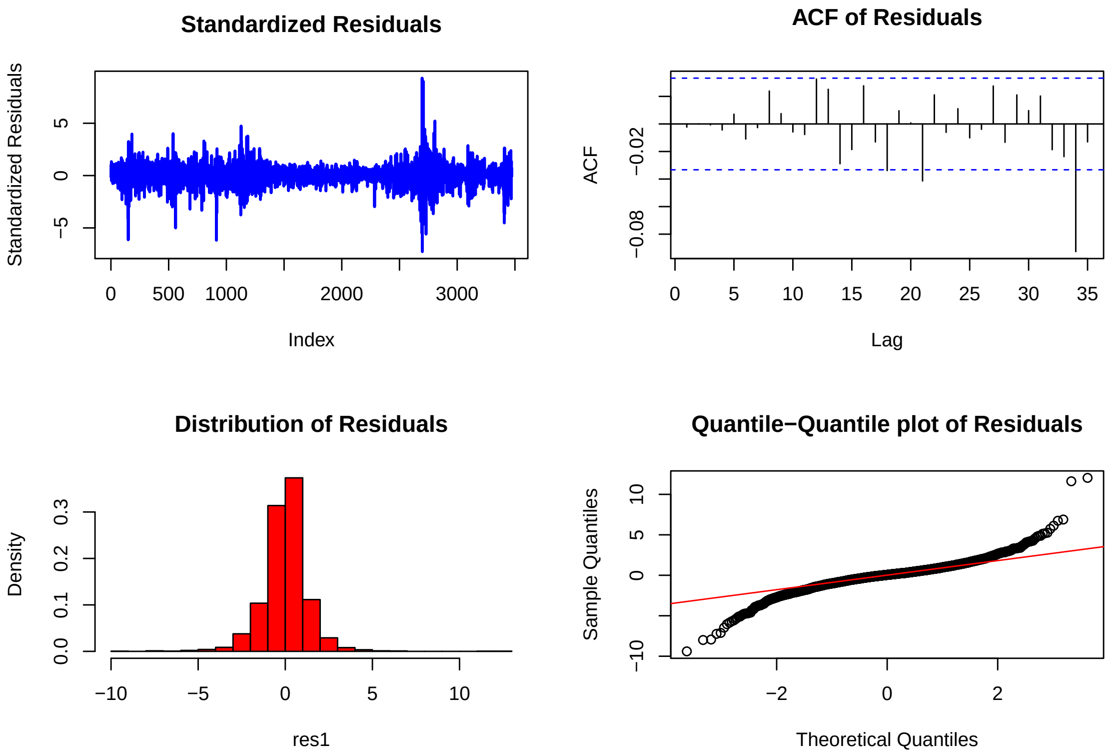

Figure 3.45: Diagnostics of ARMA(2,5) for daily DIA returns

Inspecting the time plot of the standardized residuals, there is evidence of volatility clustering. Looking at the ACF plot of the residuals of this model, there is no evidence of autocorrelation in the residuals for small lags. The normal QQ plots of the residuals has fat tails and hence shows departure from normality. This idea is supported by the histogram of the residuals which is slightly skewed to the left and has high kurtosis than a normal distribution.

3.3.3 Model of Heteroscedasticity

The analysis continue with a test for an ARCH effect presence in the residuals of the specified ARMA (2,5) model. First we looked at ACF and PACF of the squared residuals. Figure 3.46 and Figure 3.47 show the ACF and PACF of the squared residuals for the DJI

and DIA respectively. It is clear that there is some dependency in the residuals.

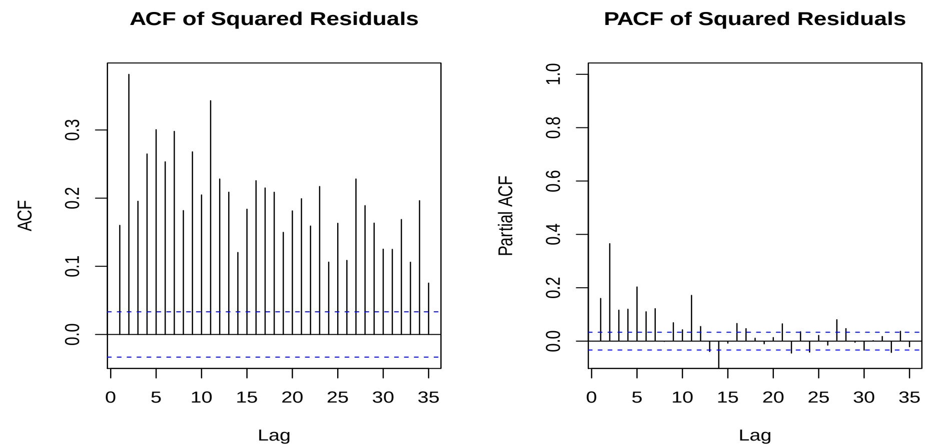

Figure 3.46: ACF and PACF of Squared Residuals of fitted ARMA(2,5) Model for DJI

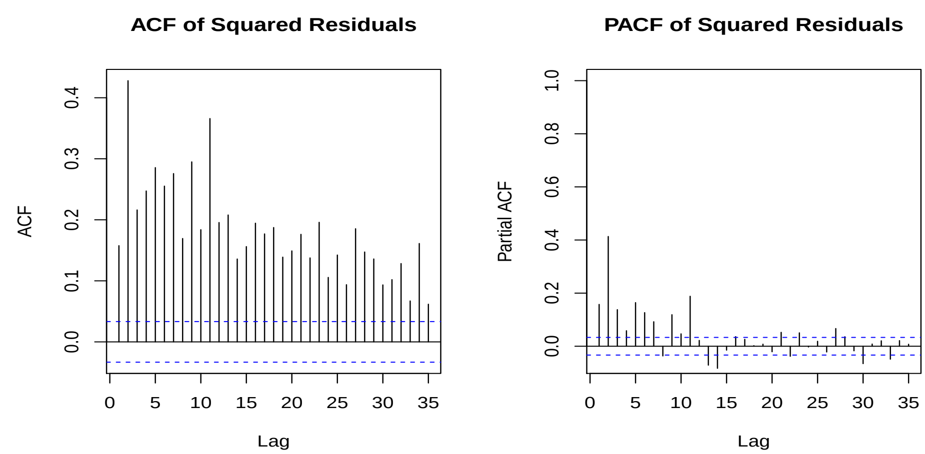

Figure 3.47: ACF and PACF of Squared Residuals of fitted ARMA(2,5) Model for DIA

Using the Box-Ljung test, we formally test for ARCH effect in the squared residuals. A test statistics of 2329.273 is recorded for DJI and 2395.074 for DIA (both are chi-squared with a degrees of freedom of 10). These values indicate the presence of ARCH effect in both returns. The Jarque-Bera test for normality can be used to confirm that the residuals

of the fitted models are not normally distributed. As a result of these, we fit a GARCH model with a skewed generalized error distribution to the residuals of the fitted ARMA(2,5) model of the NDX returns. Some results obtained are shown in Table 3.11 and Table 3.12 bolow.

Table 3.11: ARMA-GARCH Model selection for DJI returns

Model	AIC	BIC
ARMA(2,5)-GARCH(1,1)	2.884885	2.907904
ARMA(2,5)-GARCH(1,2)	2.885623	2.910413
ARMA(2,5)-GARCH(2,1)	2.879188	2.903979
ARMA(2,5)-GARCH(2,2)	2.879389	2.905950

Table 3.12: ARMA-GARCH Model selection for DIA returns

Model	AIC	BIC
ARMA(2,5)-GARCH(1,1)	2.896144	2.919147
ARMA(2,5)-GARCH(1,2)	2.896892	2.921664
ARMA(2,5)-GARCH(2,1)	2.890543	2.915316
ARMA(2,5)-GARCH(2,2)	2.890735	2.917277

The ARMA(2,5) mean with GARCH(2,1) variance is chosen for both returns since it has minimum AIC and minimum BIC values. From Figure 3.48 and Figure 3.49 below, it is clear that the GARCH(2,1) model is adequate for describing the heteroscedasticity of the series.

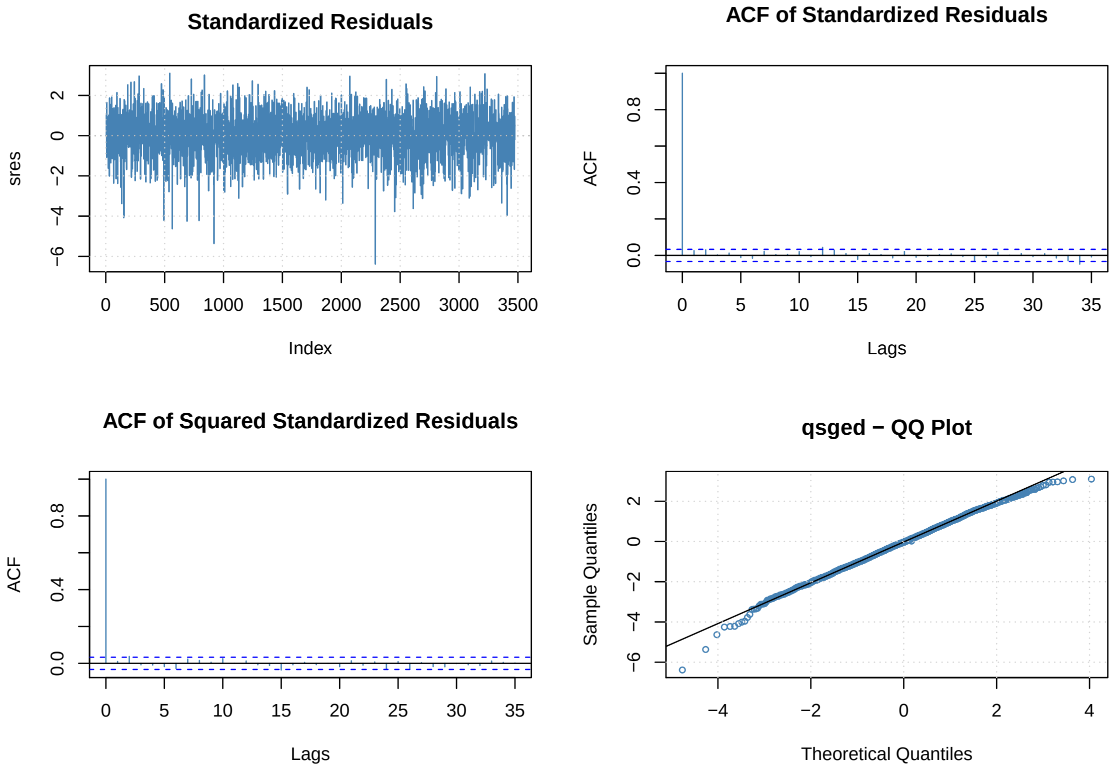

Figure 3.48: Diagnostics of ARMA(2,5)-GARCH(2,1) model for daily DJI returns

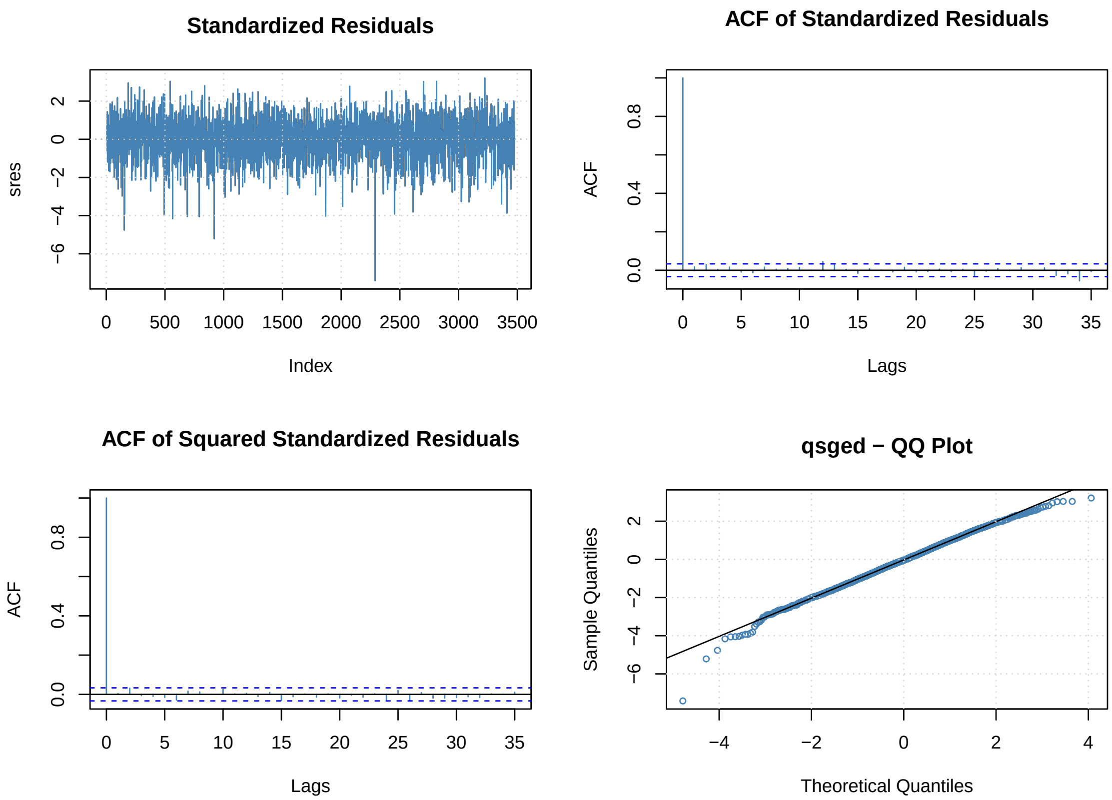

Figure 3.49: Diagnostics of ARMA(2,5)-GARCH(2,1) model for daily DIA returns

The maximum likelihood estimates of the ARMA(2,5)-GARCH(2,1) with skewed generalised distribution were obtained using the R software and the conditional mean and conditional variance equations for the DJI returns are shown in Equation 3.11 and for the DIA returns in Equation 3.12.

$$
\begin{aligned}
y_t =& 0.004885 + 0.06769 y_{t-1} + 0.8042 y_{t-2} + e_t - 0.1297 e_{t-1} - 0.8405 e_{t-2} \\
&+ 0.03987 e_{t-3} + 0.03621 e_{t-4} - 0.01499 e_{t-5} \\
\sigma^2_{t|t-1} =& 0.01489 + 0.1016 e^2_{t-2} + 0.8901 \sigma^2_{t-1|t-2}
\end{aligned} \tag{3.11}
$$

$$
\begin{aligned}
y_t =&0.006903 + 0.043169y_{t-1} + 0.820995y_{t-2} + e_t - 0.098325e_{t-1} - 0.858323e_{t-2} \\
&+ 0.021464e_{t-3} + 0.031703e_{t-4} - 0.002422e_{t-5} \\
\sigma^2_{t|t-1} =&0.015341 + 0.007097e^2_{t-1} + 0.100515e^2_{t-2} + 0.884403\sigma^2_{t-1|t-2}
\end{aligned}
\tag{3.12}
$$

The coditional standard deviations are displayed in Figure 3.50 and Figure 3.51 below.

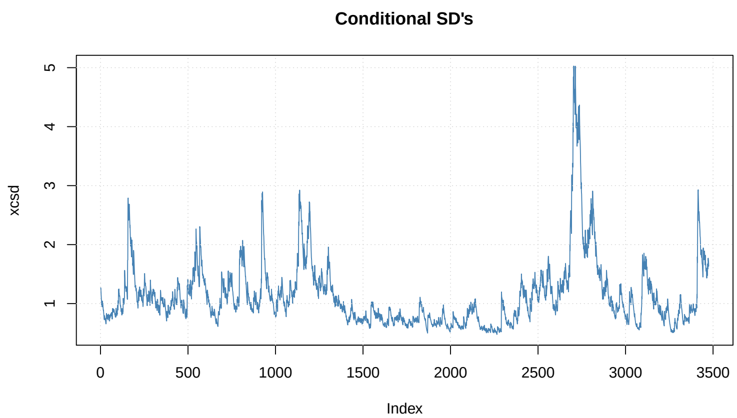

Figure 3.50: Coditional Standard Deviation for daily DJI returns

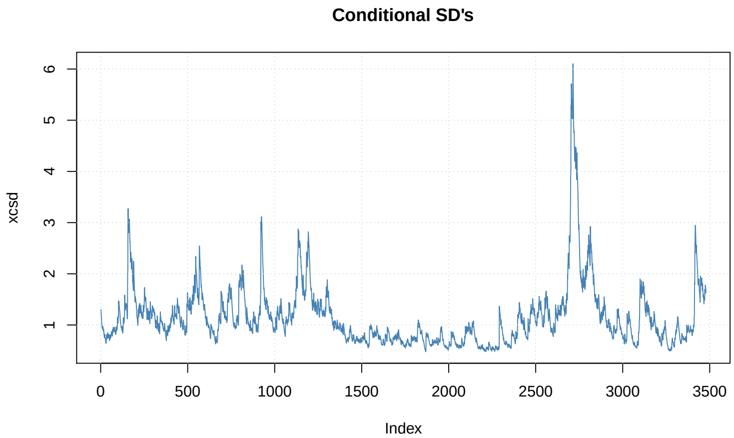

Figure 3.51: Coditional Standard Deviation for daily DIA returns

The fitted ARMA(2,5)-GARCH(2,1) model is weakly stationary and the unconditional (long-run) variance is given by equation 2.11. The standard deviation which is the volatility, is equal to the square root of the variance. This value is 1.3394% for DJI and 1.3861% for DIA. Remember this measure is calculated on the basis of the daily data.

Chapter 4

Results and Conclusions

In this chapter we present the main results upon which this thesis is centered. The first section is about results obtained by performing formal tests on the residuals of the fitted ARMA-GARCH models to decide if the model fits the considered data well. The second section talks about some statistical properties of the ETFs and the indices they mimic.

4.1 Standardised Residual Tests

Table 4.1 shows results of test of goodness of fit of the ARMA(2, 0)-GARCH(2, 1) model fitted to the NDX and QQQQ daily returns, Table 4.2 shows results of test of goodness of fit of the ARMA(2, 5)-GARCH(2, 2) model fitted to the S&P 500 and SPY daily returns and Table 4.3 shows that of the ARMA(2, 5)-GARCH(2, 1) model fitted to the DJI and DIA daily returns.

Table 4.1: Test Results from NDX and QQQQ returns

Test		NDX Statistics(p-value)	QQQQ Statistics(p-value)
Jarque-Bera (R)	χ^2	72.2256(0)	85.9016(0)
Ljung-Box (R)	Q(10)	7.1228(0.7138)	7.3196(0.6950)
Ljung-Box (R^2)	Q(10)	7.3777(0.6894)	5.0523(0.8877)
LM Arch (R)	TR^2	7.6737(0.8101)	6.6907(0.8774)

Table 4.2: Test Results from S&P 500 and SPY returns

Test		S&P 500 Statistics(p-value)	SPY Statistics(p-value)
Jarque-Bera (R)	χ^2	852.7770(0)	1092.462(0)
Ljung-Box (R)	Q(10)	16.7427(0.0803)	17.9298(0.0562)
Ljung-Box (R^2)	Q(10)	3.3148(0.9730)	6.3490(0.7851)
LM Arch (R)	TR^2	3.7653(0.9873)	6.650(0.8798)

Table 4.3: Test Results from DJI and DIA returns

Test		DJI Statistics(p-value)	DIA Statistics(p-value)
Jarque-Bera (R)	χ^2	366.9800(0)	530.0465(0)
Ljung-Box (R)	Q(10)	10.7830(0.3747)	8.2206(0.6073)
Ljung-Box (R^2)	Q(10)	14.5580(0.1490)	11.4641(0.3225)
LM Arch (R)	TR^2	14.7550(0.2551)	11.3280(0.5010)

The Jarque-Bera (Jarque and Bera 1987) test clearly rejects the null hypothesis of normality of the residuals. According to the LjungBox (Ljung and Box 1978) statistic for the residuals from the fitted model, there is no relevant autocorrelation. Even though the series of squared returns were serially correlated over time, the LjungBox statistic for up to tenth order serial correlation of squared residuals from the fitted model is not significant, suggesting the presence of independence in the residuals. As nonlinear dependence and heavy-tailed unconditional distributions are characteristic of conditionally heteroskedastic data, the Lagrange Multiplier test (Engle 1982) can be used to formally test the presence of conditional heteroskedasticity and the evidence of ARCH effects. The LM test for ARCH effect in the residuals suggests that all ARCH effect exhibited by the returns

has been captured by incorporating GARCH structures in the model, allowing conditional heteroskedasticity by conditioning the volatility of the process on past information.

4.2 Basic Statistical Properties

We analyze the continuously compounded percentage rates of returns. Table 4.4 below summarizes the basic statistical properties of these returns.

Table 4.4: Results from Returns

STATISTICS	NDX	QQQQ	S&P 500	SPY	DJI	DIA
Mean	0.005119	0.005771	0.022562	0.029622	0.012533	0.020645
Median	0.102159	0.111080	0.062232	0.070225	0.045065	0.061271
Minimum	-11.114930	-9.384873	-9.469514	-10.374190	-8.200513	-9.867413
Maximum	17.202970	15.563210	10.957200	13.566050	10.508350	12.708600
Skewness	0.182150	0.145125	-0.243730	-0.096703	-0.088053	0.143604
Kurtosis	4.580017	4.476979	8.299678	9.653930	6.855013	9.489783
Volatility	2.6675	2.7224	1.3300	1.3347	1.3394	1.3861

The mean returns are all positive but close to zero. The returns appear to be somewhat asymmetric as reflected by positive and negative skewness estimates. All returns have heavy tails and show strong departure from normality (skewness and kurtosis coefficients are all statistically different from those of the standard Normal distribution which are 0 and 3 respectively).

4.3 Conclusions

Based on the results of this study, we conclude that incorporating skew-generalized error distribution (SGED) returns innovation into the GARCH model generates good stochastic

volatility models for ETFs. These findings demonstrate the significance of both skewness and tail-thickness in the conditional distribution of returns, and should be considered in making decisions regarding market timing, portfolio selection and VaR estimates, particularly for financial markets. The volatility of the ETFs compared to that of the indices are identical in values.

Finally, the cross-correlation at lag 0 (the only significant lag) between daily NDX returns and daily QQQQ returns is 0.973, between daily S&P 500 returns and daily SPY returns is 0.978 and between daily DJI returns and daily DIA returns is 0.980. These suggest strong contemporaneous positive linear association between the ETFs and the corresponding indices. In other words, higher prices of indices are associated with higher prices of corresponding ETFs at the same time. The plots of the cross-correlation functions (CCF) are shown in Figures 4.1, 4.2 and 4.3 (Prewhitening is pertinent to disentangle the linear association from autocorrelations in the data). Hence from the results obtained, one can conclude that the exchange-traded funds have statistical behavior similar to that of the corresponding financial indices that they mimic.

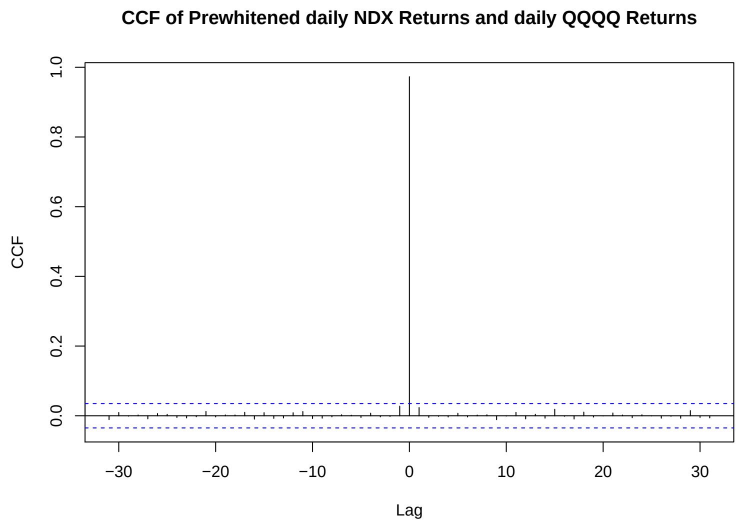

Figure 4.1: Sample CCF of Prewhitened daily NDX Returns and daily QQQQ Returns

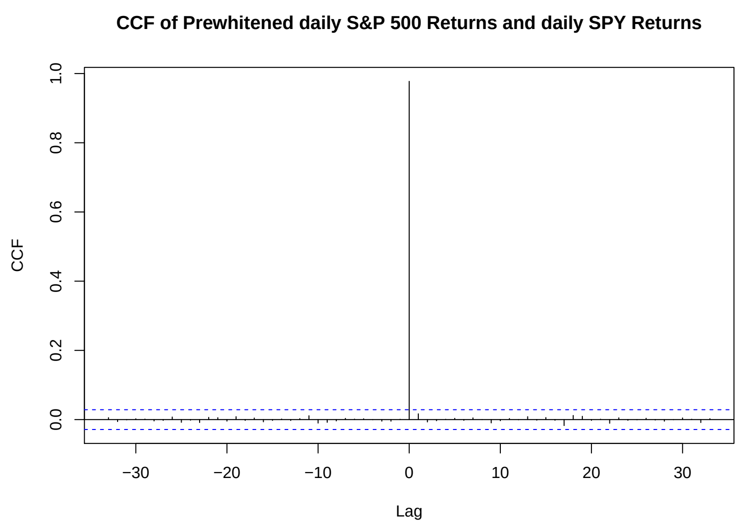

Figure 4.2: Sample CCF of Prewhitened daily S&P 500 Returns and daily SPY Returns

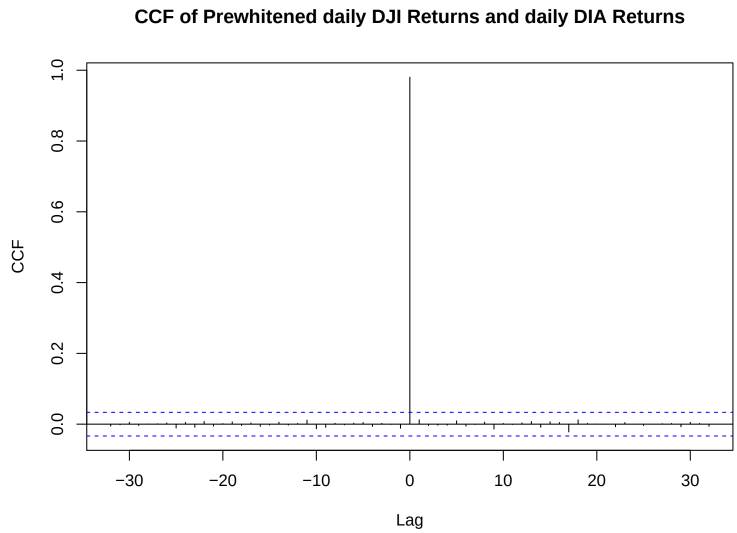

Figure 4.3: Sample CCF of Prewhitened daily DJI Returns and daily DIA Returns

References

[1] Ngai Hang Chan, "Time Series Applications to Finance," John Wiley & Sons, Inc., 2008.

[2] H. Cherry, "Financial Economics," *Actuarial Study Materials* 3217 Wynsum Ave., Merrick, NY 11566, 2007.

[3] P. Cizeau et al., "Volatility Distribution in the S&P500 Stock Index," Physica A 245, 1997, 441-445.

[4] Jonathan D. Cryer and King-Sik Chan, "Time Series Analysis with applications in R," *Second Edition* Springer Science+Business media, LLC, 2008.

[5] C. J. Hull, "Options, Futures and Other Derivatives," *sixth edition* Pearson Prentice Hall, 2006, pp. 99-373.

[6] E. Koscienly-Bunde et al., "Long-range power-law correlations in daily temperature fluctuatins," Phil. Mag. B 77, 1998, 1331-1339.

[7] Y. H. Liu et al., "Quantification of Correlations in Economic Time Series," Physica A 245, 1997, 437-440.

[8] R. N. Mantegna, H. E. Stanley, "An Introduction to Econophysics: Correlations and Complexity in Finance," Cambridge University Press, Cambridge 1999.

[9] R. N. Mantegna, H. E. Stanley, "Scaling Behaviour in the Dynamics of an Economic Index," Nature 376, 1995, 46-49.

[10] M. C. Mariani et al., "Long correlations and Levy Models applied to the study of Memory effects in high frequency (tick) data," Physica A, 388(8), 2009, p. 1659-1664.

[11] M. C. Mariani et al., "Study of memory effects in international market indices," Physica A, 389(8), 2010, pp 1653-1664.

[12] M. C. Mariani et al., "Long Correlations and Normalized Truncated Levy Models applied to the study of Indian Market Indices in comparison with other emerging markets," Physica A, 387(5), 2008, pp 1273-1282.

[13] M. C. Mariani et al., "Levy models and long correlations applied to the study of exchange-traded funds," International Journal of Computer Mathematics, 86:6, 2009, 1040-1053.

[14] M. C. Mariani, Y. Liu, "A new Analysis of Intermittence, Scale Invariance and Characteristic Scalea applied to the Behavior of Financial Indices near a Crash," Physica A 367, 2006, 345-352.

[15] Option Volatility & Pricing, *Advanced Trading Strategies and Techniques, Chicago* Probus Publishing Company. pp. 257-330, 1994.

[16] C. K. Peng et al., "Statistical Properties of DNA Sequences," *Proc. Int'l Iupap Conf. on Statistical Physics, Taipei*, Physica A 221, 1995, 180-192.

[17] C. K. Peng et al., "Long-Range Anticorrelations and Non-Gaussian Behavior of the Heartbeat," Phys. Rev. Lett. 70, 1993, 1343-1346.

[18] A. Weishaus, "Financial Economics," *fourth edition. Actuarial Study Materials* 276 Roosevelt Way, Westbury, Merrick, NY 11590, 2008.

[19] Robert A. Yaffee with Monnie McGee, "An Application to Time Series Analysis and Forcasting with Applications of SAS and SPSS," Academic Press, Inc., San Diego London, 1999.

Curriculum Vitae

Rebecca Davis was born on February 3, 1984. She graduated from St Louis Senior High School, Ghana, in the year 2001. She entered Kwame Nkrumah University of Science and Technology (KNUST), Ghana, in the year 2003 and graduated in 2007 with a Bachelor of Science degree in Mathematics (First class honors). She rendered national service as a teaching assistant at the Department of Mathematics, KNUST. Rebecca, in her persuit as an actuary has currently written and passed three of the professional exams offered by SOA (Society of Actuaries (U. S. A)); Probability (Exam P) Financial Mathematics (Exam FM) Exam M Financial Economics (Exam MFE)

In the fall of 2010, she joined The University of Texas at El Paso as a graduate student in Statistics and a Teaching Assistant at the Mathematical Sciences department. Based on her performance, she was awarded for Academic-Excellence in Statistics in 2012.

After graduation, Rebecca will pursue career and professional ambitions in her proud country of birth, Ghana.

Present address: 806 West Yandell Dr.
El Paso, Texas 79902