Investigating Methods of Modelling Financial Time Series with Empirical Analysis

[redacted]

Supervised by [redacted]

Submitted in accordance with the requirements for the
module MATH5871M: Dissertation in Statistics
as part of the degree of

Master of Statistics with Applications to Finance

The University of Leeds, School of Mathematics

September 2017

The candidate confirms that the work submitted is his/her own and that appropriate credit has been given where reference has been made to the work of others.

Abstract

The volatility of a time series is the extent to which it varies in value over time. It has important applications in finance, as it indicates the likelihood of a stock or financial indicator changing drastically from its current value. Early papers such as Mandelbrot (1963) noted the tendency for periods of high and low volatility to be each grouped together in clusters. This idea led to the important findings presented by Engle (1982), which showed that a model allowing for a changing conditional variance was able to capture this clustering feature commonly found in financial data.

This project aims to outline some of the key characteristics of financial time series and to explore how effectively they can be captured by ARCH and GARCH models. The methods of how to apply the concepts in practice are investigated, demonstrating their advantages over more basic models.

Initially, basic time series concepts are revised and the procedure for creating and evaluating linear time series models is outlined, focusing on the "ARMA" model. It is found that after selecting and fitting the best ARMA model for a set of FTSE 100 closing prices data, there remains substantial shortcomings, namely the inability to fully explain the existing correlations in squared returns data. The classes of ARCH and GARCH models are introduced, and the important properties of possessing a heavy-tailed distribution as well as the ability to produce volatility clusters are demonstrated. The ARCH model is found to be able to greatly improve on the ARMA model's weaknesses, but has disadvantages such as requiring large numbers of parameters. When applied to the same data set, the GARCH model improves on these disadvantages. Some modifications of these models are shown, and possible methods for creating and evaluating forecasts of the conditional variance are investigated.

The closing prices from the Standard and Poor 500 stock index are analysed using summary statistics to further demonstrate the traits which financial time series possess, and the data is tested for the presence of weekly effects. Models for the conditional mean and conditional variance are analysed to find the best model for the data. Forecasts for the conditional variance are compared, with the exponential smoothing method and Student-t GARCH models being found to produce the smallest error and hence the best forecast accuracy.

This project confirms through theoretical and practical analysis the strengths of conditional variance models and it is hoped that it can be used as a guide for applying and evaluating these models for other data sets in the future.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The world of finance is one which we know shall always remain extremely relevant in every country across the globe. The competition in the financial sector is tougher than ever and successful companies are constantly looking for ways to remain relevant and secure. The area of risk is paramount in any financial firm no matter how big or small, and occasionally we have seen the consequences of underestimating the magnitude of the risks they face each and every day. A clear example of this was the financial crisis of 2007/8 where overconfidence and underestimation of risk by banks led to a worldwide crisis sending many countries into economic recession. We shall see later on in this project the strong impact this particular event had on the Standard and Poor stock index.

Financial analysis has close ties with Mathematics, and as time goes on more and more mathematicians and economists are delving deeper into all areas of the subject to find new and improved ways to attempt to know the unknown. We live in an information age where everything can now be quantified. The new growing trend for companies and organisations to have billions and billions of rows of data on everything relevant to them does not seem to be slowing down. Economic policies made by the government rely on research by leading economists, however as mathematical as this area of interest may seem, it involves a certain degree of speculation and hypotheses. This stems from people's attempts to balance the objective to grow and perform as quickly and efficiently as possible whilst attempting to avoid future disasters.

The fact is that markets are not subjected to a constant amount of risk, and so the volatility of stocks which is directly related to this risk must also be expected to change over time. To be able to understand further the behaviour of this risk is incredibly valuable for investors, regulators and all financial firms. Closely related to risk is the volatility, which describes the extent to which the value of the instrument being examined moves up and down during a time period. The opportunity to know the volatility of the market can allow better financial decisions to be made, meaning any risk taken is more calculated. For example, derivatives such as options are very sensitive to how much stock prices change, and so it is useful to know how much risk is involved when buying them.

An important characteristic of financial time series, and in fact a large proportion of all time

series is the dependency between terms. Box and Jenkins popularised the class of ARMA models, which provide a simple enough method to explain some of the dependencies within data. In this piece we shall explore the use of econometric models to track volatility to investigate how these models improve upon the more basic ARMA models.

Nobel Laureate Robert Engle's ARCH models, which used the idea of economic time series possessing non-constant volatility, paved the way for a new pathway of research which has been extensively examined ever since. ARCH models involve evaluating the past behaviour of the market and attempting to assign weights to the data to then predict future risk. This research acted as a base for a wide variety of other models containing this concept, which improve on the original in different ways. A key example is the GARCH model, first put forward by Bollerslev (1986) a few years after the groundbreaking Engle (1982) paper, which is still being widely used and studied today, 40 years later.

In Chapter 2, we give an overview of some fundamental time series definitions, and then we describe each stage of the model selection process to choose the best ARMA model for a given set of data which can then be used for forecasting. We then apply this to FTSE returns data to see how well the more basic models can capture its properties.

In Chapter 3, we introduce ARCH models, initially investigating some of their characteristic properties. We then outline the process of building an optimal ARCH model and implement this on the FTSE data to see the improvements these models make. Afterwards, we move onto GARCH models, following a similar procedure of learning how to choose the best models and applying it to the FTSE data. The difficulties of forecasting are mentioned and a special type of GARCH model is investigated.

In Chapter 4, we present the main large set of data we shall be analysing in detail, the last 67 years of S&P 500 closing prices, and explore some general properties of the prices being examined. We also investigate whether the changes in price are different for different days of the week.

In Chapter 5, we attempt to fit the various heteroskedastic models we have explored to the S&P 500 data, by using the methods presented in Chapter 2 and 3. We then attempt to forecast the conditional mean and variance to evaluate the benefit of GARCH models and whether they improve on more simpler types.

We hope this project can act as a guide to the analysis and implementation of linear models and volatility models and present some interesting concepts and results, the main focus being investigating to what extent these models improve on other more basic versions.

Chapter 2

Time Series Analysis

2.1 Introduction

We shall begin with a short summary of fundamental time series definitions which will be needed for analysis throughout this project. These are taken from the lecture notes provided by Thwaites (2016) in the Time Series and Spectral Analysis module (MATH5802M).

Definition 2.1.1. ***Stationary*** *- A time series X_t is strongly stationary if the distributions of $(X_1, X_2, ..., X_n)$ and $(X_{1+k}, X_{2+k}, ..., X_{n+k})$ are the same $\forall$ k, n $\in \mathbb{N}$*

Below in Figure 2.1 is an example of a stationary time series. The property of stationarity is important since a time series with a trend (slow change in mean value) or seasonality (periodic components repeating at a fixed frequency) cannot be stationary. The property also means that if the first 2 moments (mean and variance) are finite, they are both constant over time, so the time series in one time period looks similar to how it appears in a different time period. This leads to a second, less strict definition for stationarity.

Definition 2.1.2. ***Weak Stationarity*** *- A time series X_t is weakly stationary if:*

1. $\mathsf{E}(X_t)$ *is constant*

2. $Cov(X_s, X_t) = \gamma(s,t) = \gamma_{t-s}$ *for* $s < t$

Definition 2.1.3. ***Autocorrelation*** *-* $\rho(s,t) = \frac{\gamma(s,t)}{\sqrt{\sigma(s)\sigma(t)}}$

If the process $X_t, t = 1, \ldots, n$, is a weakly stationary time series, the correlation $\rho(s,t)$ can be written in terms of just the lag ρ_{t-s} and $\sigma(s) = \sigma(t)$. The **autocorrelation function** or "acf" is a plot of ρ_k vs the lag k.

If X_t has sample mean $\bar{x} = \frac{1}{n}\sum_{t=1}^{n} X_t$, then the **sample autocorrelation function** for lag k

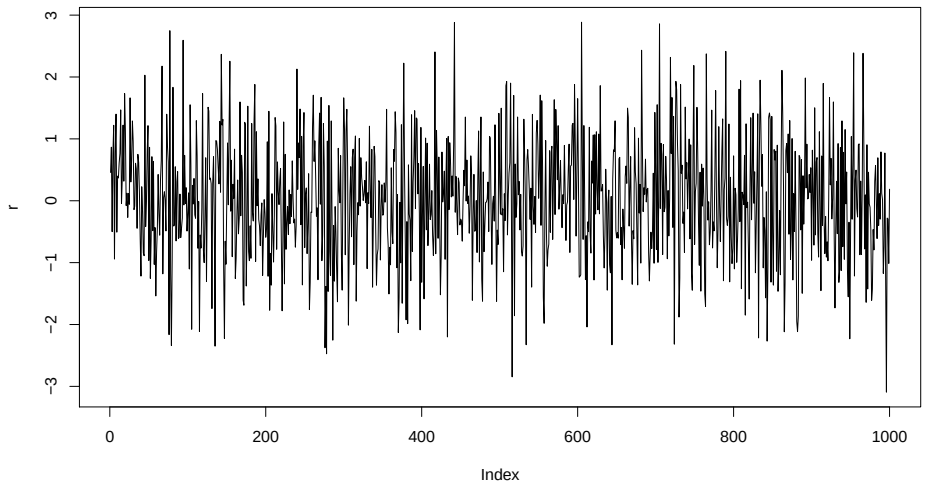


Figure 2.1: Stationary time series plot

is given by

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{n}(X_t - \bar{x})(X_{t+k} - \bar{x})}{\sum_{t=1}^{n}(X_t - \bar{x})^2} \tag{2.1}$$

Definition 2.1.4. ***White Noise*** *- A time series X_t is said to be a White Noise process if its values are independently identically distributed (i.i.d.) and $\mathsf{E}(X_t) = 0$*

The plot in Figure 2.1 is an example of a white noise process; the characteristics of zero mean and constant variance are visually evident.

Definition 2.1.5. ***Auto-regressive process (AR(p))*** *- X_t is an autoregressive process of order p if*

$$X_t = \sum_{k=1}^{p} \phi_k X_{t-k} + \epsilon_{t-k}$$

$\phi_1, \ldots, \phi_p \in \mathbb{R}$, ϵ_t is a white noise process, ϵ_t independent of X_s for $s < t$

An AR(p) process is stationary if the roots $y_1, \ldots, y_p$ of $\phi(y) = 1 - \phi_1 y - \cdots - \phi_p y^p$ are such that $|y_i| > 1 \in \mathbb{R}\ i = 1, \ldots, p$

We can determine the values of the coefficients in the AR(p) equation ϕ_1,...,ϕ_p by computing the correlations $\rho_1, \ldots, \rho_p$ and using the *Yule-Walker equations*:

$$\begin{aligned}
\rho_1 &= \phi_1 + \rho_1\phi_2 + \cdots + \rho_{p-1}\phi_p \\
\rho_2 &= \rho_1\phi_1 + \phi_2 + \cdots + \rho_{p-2}\phi_p \\
&\vdots \\
\rho_p &= \rho_{p-1}\phi_1 + \rho_{p-2}\phi_2 + \cdots + \phi_p
\end{aligned}$$

Definition 2.1.6. ***Lag-k Partial Autocorrelation*** *- The coefficients ϕ_{kk} calculated when an AR(k) model is fitted using the Yule-Walker equations, where k is the lag of the autocorrelations.*

The **partial autocorrelation function** or "pacf" is a plot of ϕ_k vs the lag k. For example, if an AR(2) model is fitted, the pacf value is ϕ_{22}, computed from the correlations ρ_1, ρ_2.

Definition 2.1.7. ***Moving Average process (MA(q))*** *- If X_t is a stochastic process such that*

$$X_t = \sum_{k=0}^{q} \theta_k \epsilon_{t-k}$$

$\theta_0, \ldots, \theta_q \in \mathbb{R}$, $q \geq 1$, then X_t is a moving average process of order q.

Note that without loss of generality, θ_0 is usually set to equal 1. All moving average processes are stationary, since the ϵ's are all i.i.d.

Definition 2.1.8. ***Auto-regressive Moving Average process (ARMA(p,q))*** *- If X_t is a stochastic process,*

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \epsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j}$$

$\phi_1, \ldots, \phi_p \in \mathbb{R}$, ϵ_t is a white noise process, ϵ_t independent of X_s for $s < t$, $\theta_1, .., \theta_q \in \mathbb{R}$

Note: ARMA(p,q) processes can also be written using the same operator $\phi(y)$ which was used to give conditions for stationarity of an AR process and a similar operator $\theta(y)$:

$$\phi(B)X_t = \theta(B)\epsilon_t$$

where $\phi(B) = 1 - \sum_{i=1}^{p} \phi_i B^i$ and $\theta(B) = 1 + \sum_{i=1}^{p} \theta_i B^i$ and B is defined as the *Backshift operator* such that $B^i X_t = X_{t-i}$.

2.2 Methods of Obtaining Adequate Models of Time Series

We shall be using the Box and Jenkins approach for creating time series models, which is a standard procedure used in time series analysis. It involves four main steps:

1. Transform the data so that it is reasonable to assume it is weakly stationary (for example taking the log of differenced data).
2. Use various measures of identification to get an idea of the p and q values.
3. Estimate the parameters p and q of the model.
4. Perform model diagnostics to evaluate whether the model sufficiently accounts for the observed characteristics of the data.

In the next sections we shall investigate methods which can be used to carry out steps 2–4, before putting them into practice on a real data set.

2.2.1 Model Identification

In order to choose between the time series models we have presented so far, we can use the acf and pacf plots to try to identify the relevant process which best models the data. Table 2.1 shows how to use the characteristics of these plots to try to identify the relevant model for the data based on the autocorrelations. The hat symbol indicates estimated parameters. The blue dashed lines on acf and pacf plots are known as Bartlett bands. A correlation is not significant with 95% confidence if it is below these lines. A white noise process has just one correlation of height 1 at lag 0, and the rest of the correlations below the bands such that $\hat{\rho}_k \sim \mathsf{N}(0, \frac{1}{n})$ for large sample size n.

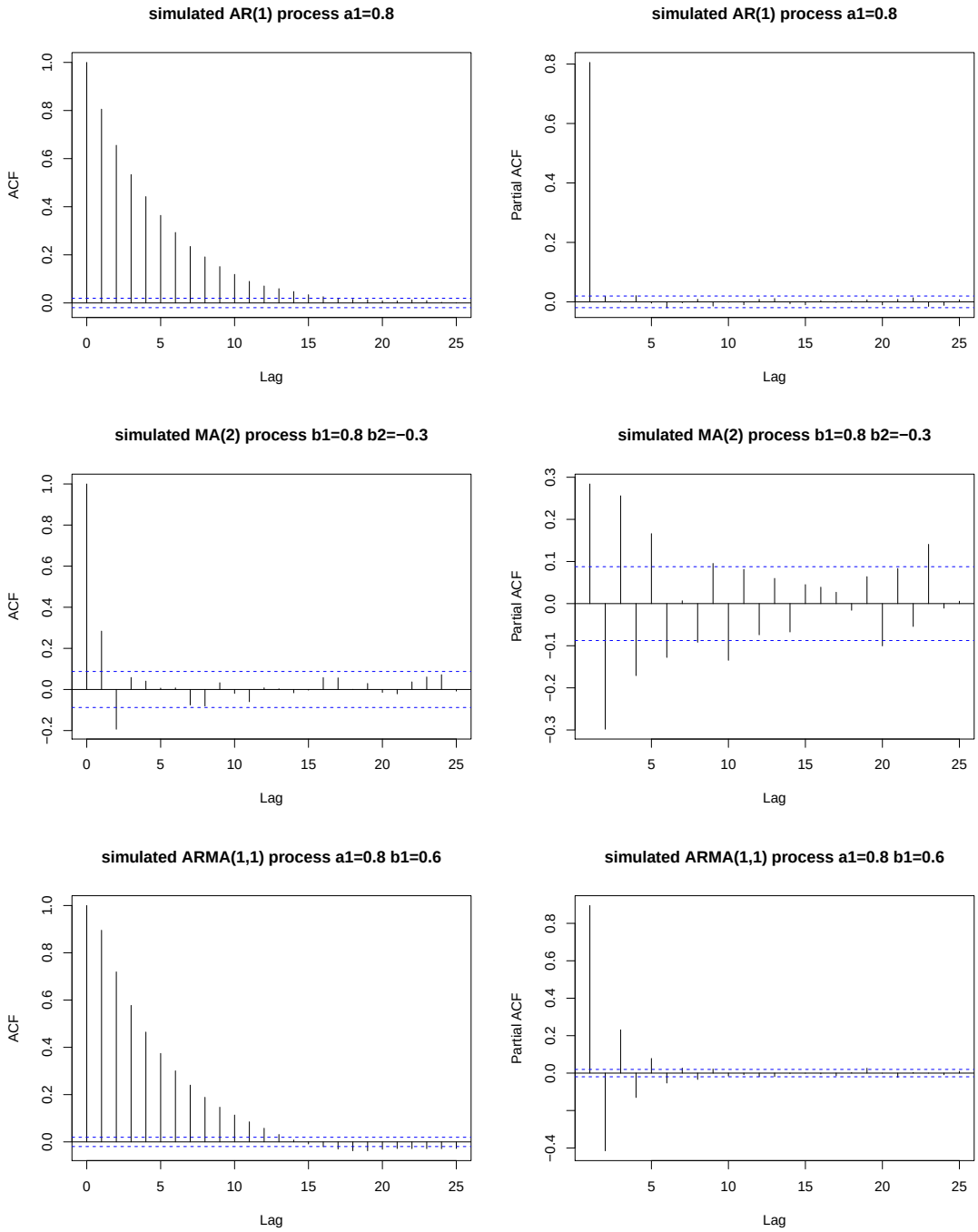
Table 2.1: Methods to identify basic time series models

Model	sample acf	sample pacf
AR(p)	$\hat{\rho}_k = \sum_{i=1}^{p} \hat{\phi}_i \hat{\rho}_{k-i}$ exponential decay/damped oscillations	$\hat{\phi}_{kk} \approx 0 \; \forall k > p$
MA(q)	$\hat{\rho}_k \approx 0 \; \forall k > q$	exponential decay/damped oscillations
ARMA(p, q)	$\hat{\rho}_k = \sum_{i=1}^{p} \hat{\phi}_i \hat{\rho}_{k-i}$ exponential decay/damped oscillations	exponential decay/damped oscillations
White Noise	no significant correlations $\hat{\rho}_k \approx 0$	no significant correlations $\hat{\rho}_k \approx 0$

We shall now show some examples of acf and pacf plots for three simulated time series to demonstrate these properties, computed using the "arima.sim" function in R. Shown in Figure 2.2 (on the next page) are the acf and pacf for an AR(1) process with $\phi_1 = 0.8$, MA(2) process with $\theta_1 = 0.8, \theta_2 = -0.3$ and ARMA(1,1) process with $\phi_1 = 0.8$, $\theta_1 = 0.6$. It is easy to spot the order of the first 2 processes using the descriptions of the acf and pacf from the table. The acf plot of the AR model has clear exponential decrease as lags increase, and the values in the pacf are all below the Bartlett bands after lag 1. In the second set of plots for the MA process, we can see that the correlations in the acf cut off at lag 2, and damped oscillations appear in the pacf plot. However, for the ARMA(1,1) process, it is unclear how to confidently identify the correct orders from the acf and pacf plots alone, and so we would require more information to decide.

In the next subsection, we shall present alternative methods to make this decision, centred around first fitting various models to the data and then testing them to decide the best choices of parameters.

Figure 2.2: Sample acf and pacf of AR(1), MA(2) and ARMA(1,1) processes (in each title ai=ϕ_i, bi=θ_i).



2.2.2 Model Estimation

If we decide we would like to fit an ARMA(p, q) model to a set of data, we are presented with the problem of identifying which model best fits the pattern of the data. Note that this includes the option of fitting AR(p) or MA(q) processes, by having one of p/q equal 0 and the other be greater than 0. In order to fit ARMA models to the data with fixed orders, we can use the relevant function in R, such as the "arima" function from the stats package. The function will compute each of the parameter values using either a conditional sum of squares or maximum likelihood approach. Note that the ARMA model may also include a constant intercept/drift term which can also be computed, however we shall assume for the majority of this chapter that this equals 0.

There are a variety of ways to choose between ARMA models, but we shall outline three of the most commonly-used methods which are able to be applied to financial data.

Definition 2.2.1. ***Akaike's Information Criterion***

$$AIC = -2\ \log(maximumlikelihood) + 2k \tag{2.2}$$

For an ARMA(p,q) model, $k = p+q+1$ if the model contains a constant drift term or $k = p+q$ otherwise (Tsay, 2013). The $2k$ term acts as a penalty term to ensure a parsimonious model (simplest plausible model with fewest parameters) is selected. The use of this measure is to balance the goodness of fit provided by the maximum likelihood estimate with the number of parameters determining the complexity of the model (Choi, 1992).

Let $\boldsymbol{y} = y_1, \ldots, y_n$ be a vector of i.i.d. random variables with conditional probability density function $f(\boldsymbol{y}|\boldsymbol{\theta})$ where $\boldsymbol{\theta} \in \mathbb{R}^k$ is a fixed vector of unknown parameters. Then the problem is to estimate $\boldsymbol{\theta}$ so that the estimated density $f(\boldsymbol{y}|\boldsymbol{\theta^o})$ is close to the true density $f(\boldsymbol{y}|\boldsymbol{\theta})$. This is done by minimising the *Kullback-Leibler information number* $I(\boldsymbol{\theta}; \boldsymbol{\theta}^o)$, which measures how different one function is from the other.

$$I(\boldsymbol{\theta}; \boldsymbol{\theta}^o) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(y|\boldsymbol{\theta})\ \log\left(\frac{f(y|\boldsymbol{\theta})}{f(y|\boldsymbol{\theta}^o)}\right) dy_1, dy_2..., dy_n$$

If $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimator of the vector $\boldsymbol{\theta}$, then the order which minimises the estimate of $\mathsf{E}[I(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}})]$ is the same as that which minimises the AIC.

Despite this being a useful estimate, it can have a large bias for a small sample size or a large parameter per data ratio. Hurvich and Tsai (1989) sought to resolve this estimate of the expected Kullback-Leibler number by proposing a corrected AIC, where k is number of parameters and n is sample size:

$$AIC_c = AIC + \frac{2(k+1)(k+2)}{n-k-2} \tag{2.3}$$

I would refer the interested reader to this paper for more information on this alternative model selection criterion.

The second method which shall be used in ARMA order identification also uses a maximum likelihood approach. Schwarz (1986) suggested a Bayes-type method which gave a slightly different formula to the AIC, where k is the number of parameters, and n is sample size:

Definition 2.2.2. ***Bayesian Information Criterion***

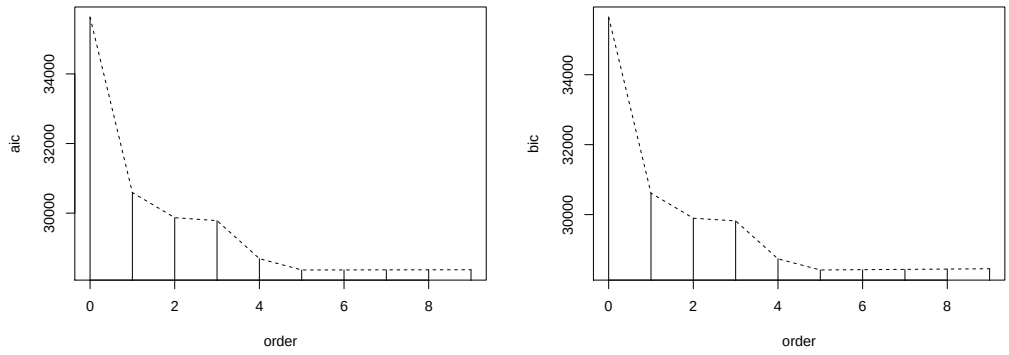
$$BIC = -2\ \log(maximumlikelihood) + \log(n)k \tag{2.4}$$

Once again, to get the optimal model under this measure, we choose the model with the smallest BIC value. Clearly this criterion differs from the AIC only by what multiplies the number of parameters. Thus, the BIC gives a larger penalty for sample sizes greater than e^2 and so tends to select lower order models than the AIC.

The AIC and BIC can be used in conjunction with each other to attempt to make the best choice, as they both have advantages over the other. As shown in Hannan (1980), if the true process is ARMA, the orders ($\hat{p}$,$\hat{q}$) minimising the BIC are strongly consistent; so as the number of data points increases, the order converges almost surely to the true orders. However, if the true process is not ARMA, then minimising AIC leads to an optimal choice of values of p and q for the data (Shibata, 1980).

To demonstrate the AIC being used, we shall plot the AIC values when fitting AR models of order 1–10 of a simulated AR(5) model. Clearly the AIC value is lowest at lag 5 where the

Figure 2.3: Plot of sample AIC (left) and BIC (right) values for a simulated AR(5) process



value is 0, so an AR(5) model is chosen. The AIC is also close to 0 for higher lags, but the simplest model is selected in this case. The BIC values were also computed and plotted, and it is evident that the BIC agrees with the AIC, selecting AR(5) as the best model fit for the data.

Both of these methods can be useful for simpler processes, but using them manually for ARMA(p, q) model identification can prove to be time consuming and there can still be a problem with "overfitting" with these methods - meaning they do not sufficiently penalise complexity. Therefore it is useful to explore another reliable method to help speed up the selection process.

The third method we shall use is known as the *Extended Auto-correlation function* (EACF) proposed by Tsay and Tiao (1984). Identifying the orders from the acf and pacf plots is difficult to accomplish as we cannot use any property involving correlations becoming insignificant after a number of lags ("cutting off"). The EACF attempts to obtain a consistent estimate of the AR order p and then derives the MA order q using the property of the acf values cutting off after lag q.

The results from the calculated values can be presented as a two-way table, whose columns represent the MA order and rows represent the AR order. The values are given either X if the absolute value of the EACF is $\geq 2 \times |se|$ meaning a significant correlation, and O if it is $\leq 2 \times |se|$ where se is asymptotic standard error. The standard error can either be computed using Bartlett's formula or using $\frac{2}{\sqrt{n}}$ where n is sample size. The derivation for the EACF can be found in Tsay and Tiao (1984).

A theoretical EACF for ARMA(1,1) is shown below in Table 2.2. The optimal order (p, q) for the ARMA model is found by using the location of the upper left corner of the triangle of O's formed in the table, highlighted as O. The * denotes values not required in this particular example.

*Table 2.2: Theoretical EACF for an ARMA(1,1) process, * are values not needed for identification in this example*

	MA							
AR	0	1	2	3	4	5	6	7
0	X	X	X	X	X	X	X	X
1	X	O	O	O	O	O	O	O
2	*	X	O	O	O	O	O	O
3	*	*	X	O	O	O	O	O
4	*	*	*	X	O	O	O	O
5	*	*	*	*	X	O	O	O

In reality there will be some error in the EACF values calculated leading to potential misspecification, and so we shall demonstrate some examples of identifying the ARMA order using the EACF table with simulated processes and then using stock price data.

Firstly, an ARMA(2,1) process was simulated with $\phi_1 = 0.8$, $\phi_2 = -0.4$, $\theta_1 = 0.9$. The simplified eacf table is shown below in Table 2.3, along with the actual values computed, using the eacf function in the TSA R-package.

The eacf table suggests the best model to be either an ARMA(2,1) or an ARMA(1,3) process (see underlined values). The two models were fitted using the "arima" function in R, and then the criterion values were computed, giving $AIC = 1411.594, BIC = 1432.667$ for the ARMA(2,1) model and $AIC = 1420.463, BIC = 1445.75$ for ARMA(1,3). Using these results we can conclude that the ARMA(2,1) model best fits the data, as it produced the smaller values for both criteria, and so our methods successfully selected the true model. R chose the

values to be $\phi_1 = 0.801$, $\phi_2 = -0.437$, $\theta_1 = 0.914$ for the ARMA(2,1), so there was still some inaccuracies but the orders were correct.

Table 2.3: Sample EACF for an ARMA(2,1) process

	MA							
AR	0	1	2	3	4	5	6	7
0	X	X	X	X	X	X	X	X
1	X	X	X	O	O	O	X	X
2	X	O	X	X	O	O	X	O
3	X	X	X	X	O	X	X	O
4	X	X	X	X	O	O	O	O
5	X	X	X	X	O	O	O	O

	MA							
AR	0	1	2	3	4	5	6	7
0	0.69	0.152	-0.12	-0.145	-0.118	-0.1464	-0.1658	-0.097
1	0.63	0.256	-0.39	-0.086	0.075	-0.0161	-0.1491	-0.094
2	0.51	-0.059	-0.32	-0.175	0.055	-0.0130	-0.1260	-0.064
3	0.52	-0.090	-0.13	-0.207	-0.074	-0.1623	-0.0968	-0.070
4	0.24	-0.395	0.36	-0.199	0.084	-0.0243	-0.0179	0.037
5	0.32	-0.407	0.42	0.111	0.034	-0.0415	-0.0067	0.053

2.2.3 Model Diagnostics

In order to ensure the chosen model is adequate, the residual series of the data minus the ARMA model should behave as a white noise process. This testing procedure involves checking whether the residuals have approximately zero mean, and the correlations are insignificant. Another part of model refinement involves removing insignificant parameters to simplify the model, as well as using more complicated models if correlations still exist in the residual series

The residuals can be tested for zero mean using a t-test, and the correlations evaluated using the sample acf and pacf plots of the residuals. The *Ljung-Box test* was proposed to evaluate how well an ARMA model explains a set of data by testing for the residuals being white noise. This test improved upon the Portmanteau test proposed by Box and Pierce (1970) through a small alteration to the test statistic to be more accurate for finite samples.

Box and Pierce (1970) suggested that $n \sum_{k=1}^{M} \hat{\rho}_k^2$ can be used as a test statistic for *whiteness* of residuals (how closely they satisfy the properties of white noise). The statistic is approximately a χ^2_M random variable but only for very large samples and was found to be unrealistic in most practical examples. This follows from the fact that if the time series is a white noise process, then for large n the distribution of $\hat{\rho}_k$ is approximately $\mathsf{N}(0, \frac{1}{n})$ (Thwaites, 2016).

Let $\hat{\rho}_k$ defined as

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n}(y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^{n}(y_t - \bar{y})^2} \tag{2.5}$$

be the sample autocorrelation at lag k (see Equation 2.1) between a series of residual terms y_t after fitting an ARMA(p, q) model. We can then define the test statistic for maximum lag M as follows:

Definition 2.2.3. ***Ljung-Box test statistic***

$$Q(M) = n(n+2)\sum_{k=1}^{M}\frac{\hat{\rho}_k^2}{n-k} \tag{2.6}$$

where M is the maximum lag chosen and n is sample size.

The Q statistic measures how large the sum of squared sample correlations are for a chosen number of lags - common choices are 5, 10, 15 and 20 (Ljung and Box, 1978). This test for white noise states that $Q(M)$ is asymptotically χ^2 distributed with $M-p-q$ degrees of freedom under the null hypothesis H_o that the residuals form a white noise process and therefore the model selected is adequate. We reject the null hypothesis H_o at confidence level α if Q exceeds the $(1-\alpha)$ quantile of the χ^2_{M-p-q} distribution. If a statistical software package is being used, this is equivalent to rejecting the null hypothesis if the p-value is less than α (this is often chosen to be 0.05).

2.3 Forecasting ARMA Models

If we want to forecast a time series to see how it could behave in the future, the main aim will be to minimise the error in that forecast compared to the true value of the time series at that point in time. We shall denote $X_t(l)$ as the l-step ahead forecast computed at time t to predict X_{t+l}. The forecast error $e_t(l)$ is defined as

$$e_t(l) = X_{t+l} - X_t(l) \tag{2.7}$$

Minimum mean squared error forecasting involves attempting to minimise the mean squared error (MMSE) of our forecast $X_t(l)$, given as $MMSE = \mathsf{E}\left[(X_{t+l} - X_t(l))^2\right]$. It can be shown, as in Hamilton (1994), that the forecast which minimises this error is also $X_t(l) = \mathsf{E}\left[X_{t+l}|\Psi_t\right]$, where $\Psi_t = \{X_t, X_{t-1}, \ldots\}$ is the information set containing all values of the time series up until time t.

For an AR(p) model, we can use an updating forecast which uses the previous one-step forecast to predict the next. The future residuals ϵ_{t+l} are set to 0, their expectation, which

minimises the MMSE (Thwaites, 2016).

$$\begin{aligned} X_{t+1} &= \phi_1 X_t + \cdots + \phi_p X_{t+1-p} + \epsilon_{t+1} \\ X_t(1) &= \mathsf{E}\left[X_{t+1}|\Psi_t\right] = \phi_1 X_t + \cdots + \phi_p X_{t+1-p} \\ e_t(1) &= X_{t+1} - X_t(1) = \epsilon_{t+1} \end{aligned}$$

Using this forecast we can now predict the value 2 steps ahead.

$$\begin{aligned} X_{t+2} &= \phi_1 X_{t+1} + \cdots + \phi_p X_{t+2-p} + \epsilon_{t+2} \\ X_t(2) &= \mathsf{E}\left[X_{t+2}|\Psi_t\right] = \phi_1 X_t(1) + \phi_2 X_t + \cdots + \phi_p X_t \\ e_t(2) &= X_{t+2} - X_t(2) = X_{t+1} - X_t(1) + \epsilon_{t+2} = e_t(1) + \epsilon_{t+2} \end{aligned}$$

This calculation can be continued in the same manner for l steps ahead:

$$X_t(l) = \sum_{i=1}^{p} \phi_i X_t(l-i) \tag{2.8}$$

where $X_t(l-i) = X_{t+l+i}$ for $l - i \leq 0$. Evidently, the forecast error increases for each step further ahead, as would be expected since the uncertainty in its future behaviour would increase.

In practice, in the equations for the forecast we replace ϕ by the estimated values $\hat{\phi}$ and the exact error term is unknown, so the estimated residuals $\hat{e}_t$ are used.

For an MA(q) model, the forecast is simpler, since setting future residuals to equal 0 means the forecast drops down to 0 or a constant non-zero value after q steps.

$$\begin{aligned} X_{t+1} &= \epsilon_{t+1} + \theta_1 \epsilon_t + \cdots + \theta_q \epsilon_{t+1-q} \\ X_t(1) &= \theta_1 \epsilon_t + \cdots + \theta_q \epsilon_{t+1-q} \\ e_t(1) &= \epsilon_{t+1} \end{aligned}$$

where we have set $\theta_0 = 1$. In general,

$$X_t(l) = \sum_{i=0}^{q} \theta_i \epsilon_{t+l-i} \tag{2.9}$$

where $\epsilon_j = 0$ for $j > t$.

For an ARMA(p, q) model, the two prior methods are used together (Chatfield, 2000). For a one-step forecast:

$$\begin{aligned} X_{t+1} &= \phi_1 X_t + \cdots + \phi_p X_{t+1-p} + \epsilon_{t+1} + \theta_1 \epsilon_t + \cdots + \theta_q \epsilon_{t+1-q} \\ X_t(1) &= \phi_1 X_t + \cdots + \phi_p X_{t+1-p} + \theta_1 \epsilon_t + \cdots + \theta_q \epsilon_{t+1-q} \\ e_t(1) &= \epsilon_{t+1} \end{aligned}$$

In general for l steps, we forecast an ARMA(p, q) model using the equation below.

$$X_t(l) = \sum_{i=1}^{p} \phi_i X_t(l-i) + \sum_{j=1}^{q} \theta_j \epsilon_{t+l-j} \tag{2.10}$$

where $X_t(l-i) = X_{t+l+i}$ for $l - i \leq 0$ and $\epsilon_{t+l-j} = 0$ for $l - j > 0$. Clearly for forecasts larger than q, only the autoregressive component will have any effect on the values computed.

There are a variety of measures which can be used to evaluate how well a model predicts values of a dataset. The most widely used measures are the Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE), defined as follows:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} e_t(i)^2 \tag{2.11}$$

$$RMSE = \sqrt{MSE} \tag{2.12}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |e_t(i)| \tag{2.13}$$

where $e_t(i)$ is the forecast error i steps ahead of the forecast origin t, and n is the number of different forecast lengths computed (Hyndman and Koehler, 2006).

Other commonly used measures include the Mean Percentage Error (MPE), defined as $MPE = 100 e_t(i)/X_{t+i}$. These have the advantage of being comparable across datasets as they do not change with scale, unlike the three measures above. However, we shall not be using these types of error measures for our forecasting of financial data, since if any one value of X is 0, the error is infinite, and when they are close to 0, the errors have a strongly skewed distribution (Hyndman and Koehler, 2006). We will be using transformed data in our analysis which are centred around 0, so these measures would not be very useful.

2.4 Examples Using FTSE 100 Data

In this section we will demonstrate how well the methods we have learned so far to model a set of financial data. We will be using the data set built into R of the closing prices of the FTSE 100 index between 1991 and 1997 (from the "EUStockMarkets" data set)

2.4.1 Choosing the Best ARMA Model

We shall attempt to fit an ARMA model to the FTSE data, but we first need to transform the prices to make them stationary. The log returns of the prices, defined as $\log(X_t/X_{t-1})$, were computed giving 1859 values. From Figure 2.4, the returns seem to be approximately stationary, since the mean and variance appear to be constant.

From observing the acf and pacf plots in Figure 2.5, it is not very clear which would be the best model to fit the data. Both plots appear to have some oscillations in correlations, so there

may exist an ARMA(p, q) model which explains this dependence structure.

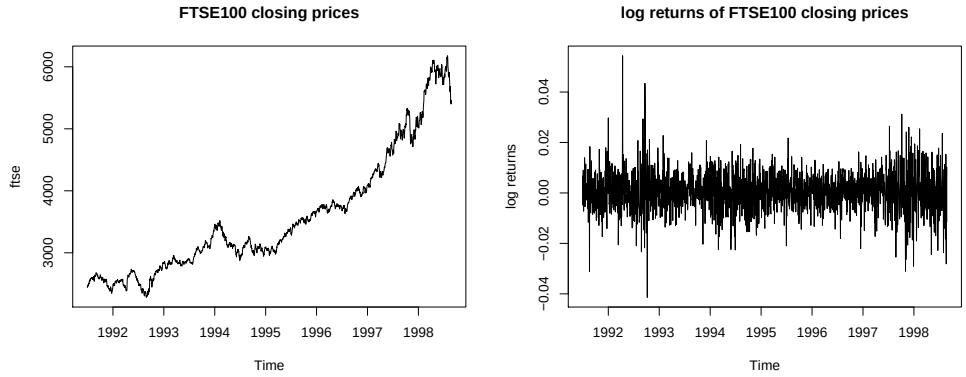


Figure 2.4: Plot of original and differenced FTSE100 data

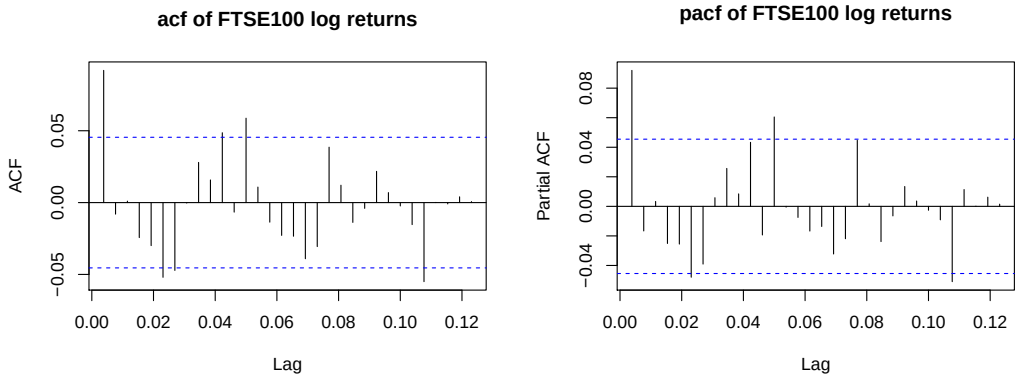


Figure 2.5: Acf and pacf of FTSE100 log returns

The methods described in Section 2.2 will be used to attempt to determine the optimal p and q values. The eacf table for the log returns is shown in Table 2.4, and it appears that an MA(1) model is selected as the best for our data.

Table 2.4: Sample EACF for FTSE100 data

	MA							
AR	0	1	2	3	4	5	6	7
0	X	O	O	O	X	X	O	O
1	X	O	O	O	O	X	O	O
2	X	O	O	O	O	O	O	O
3	X	X	X	O	O	O	O	O
4	X	X	X	O	O	O	O	O
5	X	X	X	O	X	O	O	O

We can compare potential models by using the AIC and BIC values computed after fitting them to the log returns. In Table 2.5 we can see that both criterion select the MA(1) model. For this example we shall only examine the MA(1) model, despite the values for the fitted AR(1) model being clearly very close to those for the MA(1) model, since fitting the autoregressive model instead does not change the results significantly.

Table 2.5: AIC and BIC values for various models fitted to the FTSE 100 data

	AR(1)	MA(1)	MA(2)	ARMA(1,1)
AIC	-12706.6	**-12707.0**	-12705.2	-12705.2
BIC	-12690.0	**-12690.4**	-12683.0	-12683.1

Table 2.6: Ljung-Box test results for FTSE 100 MA(1) model

Ljung-Box Q-statistic values				
lag	5	10	15	20
Q($resid$)	2.176	11.03	23.14	31.72
p-value	0.704	0.274	0.0580	0.0336
Q$\left(resid^2\right)$	61.14	99.58	166.4	195.0
p-value	0	0	0	0

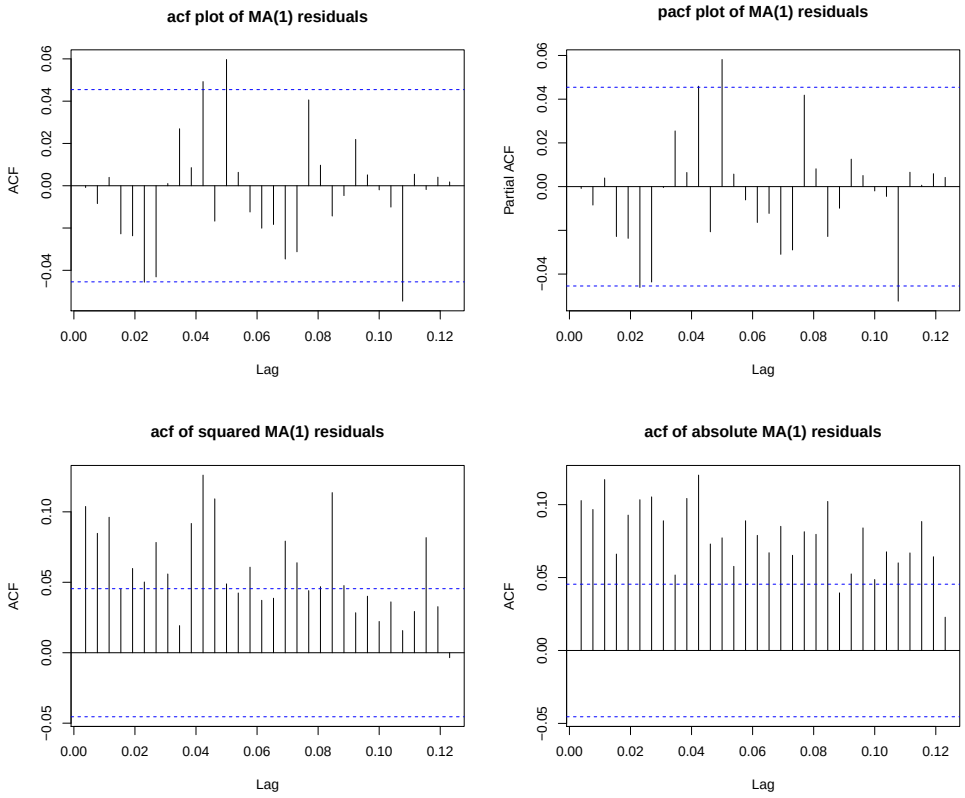


Figure 2.6: Sample acf and pacf of residuals of MA(1) model of FTSE data

If this moving average model fits the data sufficiently, then the residuals of the model should be a white noise process. This can be tested by using the Ljung-Box test on the residuals (shown in Equation 2.6). The values computed are given in Table 2.6, with all the p-values above 0.05 except for lag 20, which would suggest the residual process is almost white noise according to this test. However, the Ljung-Box test is strongly rejected for all 4 lag selections for the squared residuals, so there still exists large amounts of dependency within the data.

The acf and pacf of the residuals of the MA(1) model (top 2 plots in Figure 2.6), appear to show the characteristic needed to be white noise, although there are some small correlations at higher lags. However, if we examine the acf of the squared residuals (bottom left plot) and absolute residuals (bottom right), there are strong correlations throughout. This along with the Ljung-Box test statistics suggest that the standard time series model we have selected does not adequately explain all correlations within the data.

This demonstrates the need for alternative models which can better explain these serial correlations which still exist within the data, some of which will be discussed in Chapter 3.

2.4.2 Forecasting the FTSE 100 Data

We will now forecast some of the FTSE 100 data to decide if the moving average model is the best at predicting the returns. There are two possible ways to conduct forecasting which can produce different results for some models. The first is *in-sample forecasting*, where you attempt to forecast the data for a time period having used the future values in estimation. For example, forecasting value for 1993–1994 having used 1990–1994 in estimation.

In our investigation we shall be performing *out-of-sample forecasting*, which does not use the future values in estimation. For example, forecasting October–December 1994 after using January–September 1994 in estimation. This method makes more sense in real applications, since future information is almost always unknown, and having extra information could give different results leading to incorrect conclusions.

We can fit a model to a subset of the log returns of our FTSE 100 data and then forecast using the estimated coefficients of each model to see which performs this task the best. In the previous section we found that the MA(1) model was the best model according to the criterion values and the EACF table. We shall compare this model with an AR(1) model which had the second smallest AIC and BIC values and evaluate their efficiency.

The forecast origin is chosen to be 20 days before the last data point. The forecasts for the AR(1) process are computed as: $X_t(l) = \hat{\phi}_1(X_t(l-1) - \hat{\mu}) + \hat{\mu}$ where $X_t(i) = X_{t+i}$ if $i \leq 0$ and $\hat{\mu}$ is the sample mean of the data not including the final 20 observations. The MA(1) forecast is simply $X_t(l) = \hat{\theta}_1 \epsilon_{t+l-1} + \hat{\mu}$ where $\epsilon_i = 0 \; \forall i > t$. The log returns of the closing prices up until the forecast origin have a non-zero mean of $\hat{\mu} = 0.0504$, and the coefficients of each model were estimated to be $\hat{\phi}_1 = 0.0892$, $\hat{\theta}_1 = 0.0921$.

The two models gave very similar forecasts - the MA(1) forecasts were the sample mean for all $l > 1$, and the AR(1) model forecasts converged to a number close to the sample mean

within 10 steps. Therefore, we shall only construct a plot for the MA(1) forecasts compared to the true values. In Figure 2.7 we can see that the forecasts form a flat line from 2 time-steps after the forecast origin onwards and act as a type of average value prediction. The 95% confidence intervals appear to capture almost all the data but 3 of the more extreme values are outside the region.

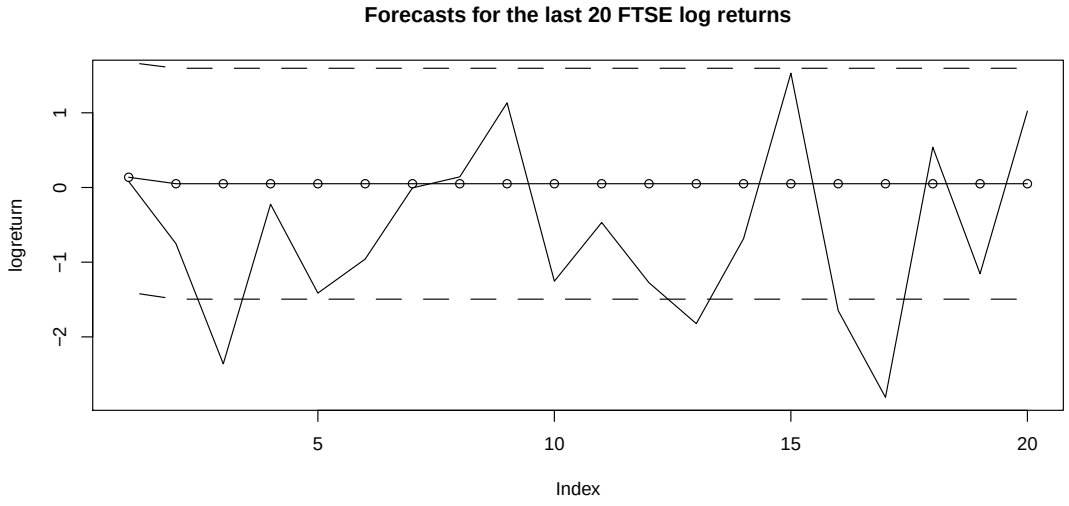


Figure 2.7: Plot of out-of-sample forecasts made by the fitted MA(1) model. "-o-" represents the forecast, the straight lines the true values, and the dashed lines are 95% confidence intervals

To see the methods of measuring error explained earlier in practice, given in Table 2.7 are the error values for the forecasts computed using the two basic time series models. The "accuracy" function contained in the "forecast" package in R was used in this analysis, which automatically computes the errors using the existing dataset. The function computed the Mean Error (ME), which is simply the mean of the forecast errors, along with the RMSE and MAE measures.

Table 2.7: Calculated error statistics for forecasts of FTSE100 log returns

size	model	ME	RMSE	MAE
5	AR(1)	-0.10017	1.31878	1.00174
	MA(1)	-0.10018	1.31770	1.00177
10	AR(1)	-0.62041	1.12254	0.85527
	MA(1)	-0.62038	1.12190	0.85528
20	AR(1)	-0.67387	1.31464	1.08548
	MA(1)	-0.67382	1.31434	1.08547

For these forecasts, the models appear closely matched, although the MA(1) model had the lowest error more often. There seems to be some discrepancies between each error measure in terms of which model had the lowest error, which demonstrates the need for multiple measures.

We can conclude that based on our analysis, the MA(1) model was the best performing model for the FTSE data overall .

Chapter 3

Volatility Models

3.1 Purpose

Volatility models have proven to be a useful tool which has allowed us to improve our understanding of the behaviour of stock markets. In this project, we will define volatility as the conditional variance of the time series being examined. For example, for a time series x_t the conditional variance would be $Var(x_t|\Phi_{t-1})$ where Φ_{t-1} is the information set of all values of the series until the most recent observation. Note, volatility can also be defined as the conditional standard deviation, which follows similar behaviour since it is just the square root of the conditional variance.

The well-known concept of volatility being used to describe risk means it is an important area of interest, as it can be used to calculate the risk of a financial position. For example, we can consider the method for pricing a Call option, a type of derivative which gives the holder the right to buy the underlying at a fixed strike price K at a time T in the future. The Black Scholes formula for the price of a Call option at time t is

$$Price = C_t = S_t\Phi(d) - Ke^{-r(T-t)}\Phi\left(d - \sigma_t\sqrt{(T-t)}\right)$$
$$\text{where } d = \left[\frac{\log(S_t/K)+(r+\sigma_t^2/2)}{\sigma_t\sqrt{(T-t)}}\right]$$

S_t is the current price of the underlying stock, r is risk-free interest rate, σ_t is the conditional standard deviation of the log returns at time t, and $\Phi(d)$ is the normal cumulative distribution function evaluated at d. Since the volatility is involved in both parts of the price equation, it plays an important role in this equation, as it does in many other areas of finance. Its time-varying nature means it impacts the behaviour of the stock price, which directly will change how much the price to purchase that stock will change (Tsay, 2002).

The findings of Engle (1982) showed how an important property named "volatility clustering" could be incorporated into time series models. This property concerns the common characteristic exhibited by financial data that high volatility values tend to be followed by high volatility values, and low volatility followed by further low volatility values. He stated how the

crude assumption of a constant volatility over time can be overcome by creating a model which allows for a non-constant volatility based on previous values.

The FTSE 100 log returns examined in Chapter 2 did not have a constant variance over time - upon closer inspection of Figure 2.4 there are periods of larger variance and the quieter periods. Yet the weakly stationary AR(1) and MA(1) models we tested would have variances of $\sigma_\epsilon^2/\left(1-\phi^2\right)$ and $\sigma_\epsilon^2\left(1+\theta_1^2\right)$ respectively, which are do not depend on time (where σ_ϵ^2 is variance of residuals) (Thwaites, 2016).

By examining the absolute and squared values of the time series analysed earlier, this characteristic of volatility clustering becomes more apparent. A lack of time-varying nature in the calculation of volatility could explain why the MA(1) model was not sufficient to explain all correlations in the data.

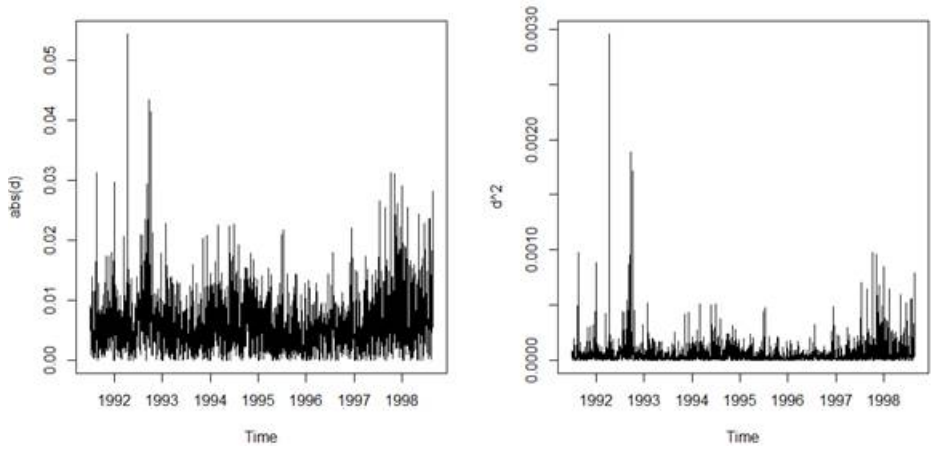


Figure 3.1: Plots of the absolute and squared log returns of the FTSE 100 data

There are 2 main classes of volatility models: ones which use a fixed equation to explain conditional variance changes, and models which use stochastic equations. We shall be focusing on the former version, but would direct the interested reader to explore Stochastic Volatility (SV) models to see how the stochastic type is used.

3.2 Our Model

As mentioned earlier, the risk to an investor or company's assets can be measured using the volatility of relevant financial data. This concept comes from the fact that greater changes in values means greater uncertainty in their future financial position, larger potential losses, and hence greater risk.

It is common practice when dealing with financial data to conduct analysis of the log returns of the data (natural log). We shall let X_t represent the financial time series data. If the process

involves small percentage changes, then

$$\begin{aligned} X_t &= (1 + r_t)X_{t-1} \\ \log(X_t) &= \log(1 + r_t) + \log(X_{t-1}) \\ \nabla \log(X_t) &= \log(1 + r_t) \end{aligned}$$

where ∇ is the difference operator defined as the operator such that $\nabla X_t = X_t - X_{t-1}$

The Taylor expansion of $\log(1 + r_t) = r - \frac{r^2}{2} + \frac{r^3}{3} - ...$ for $-1 < r \leq 1$. If r is small, higher order terms are negligible so $\nabla \log(X_t) \approx r_t$ (Shumway and Stoffer, 2011). We shall denote r_t as the log returns of the time series. The financial data being examined in this project are daily returns, therefore at time t: $r_t = 100 \log(\frac{X_t}{X_{t-1}})$. The returns are multiplied by 100 to assume continuous compounding.

The log returns process can be written as a sum of 2 parts: $r_t = \mu_t + y_t$. μ_t is the linear mean equation for the data, and y_t are the residuals of the mean equation. In section 2.5, the mean equation was found to be an MA(1) process, and we found that the y_t term must also be a process which explains the dependency in the log returns, and not simply white noise. This model structure means that we can have one process describing the conditional expectation $\mu_t = \mathsf{E}(r_t|\Psi_{t-1})$ and another process which we will be focusing on to model the conditional variance:

$$Var(r_t|\Psi_{t-1}) = \mathsf{E}((r_t - \mu_t)^2|\Psi_{t-1}) = \mathsf{E}(y_t^2|\Psi_{t-1})$$

3.3 ARCH Models

To begin with we shall examine the ARCH model, the first main model to assume a "heteroskedastic" (non-constant) variance, put forward by Engle in 1982. Let $\{y_t\}$ be a stochastic process with conditional density $f(y_t|y_{t-1})$. The forecast for this process given the previous term is simply $\mathsf{E}(y_t|y_{t-1})$. The conditional variance $Var(y_t|y_{t-1})$ can be viewed as a random variable, although in practice this is not directly observable. The main idea behind these models is to capture the characteristic involving the returns being serially uncorrelated, but still dependent. This was demonstrated in Section 2.4 by the model appearing initially to be sufficient for the data until the acf plots of absolute and squared residuals were considered, which showed clear dependency still remaining.

3.3.1 ARCH(1) Models

The most basic model proposed is known as the ARCH(1) model, which makes the conditional variance $h_t = Var(y_t|y_{t-1})$ depend on a constant and the previous value of y_t squared. The y_t values are given as innovations scaled by the standard deviation (Engle, 1982).

Definition 3.3.1. ***ARCH(1) model:***

$$y_t = \epsilon_t h_t^{1/2} \tag{3.1}$$

$$h_t = \alpha_0 + \alpha_1 y_{t-1}^2 \tag{3.2}$$

ϵ_t *i.i.d. with mean* 0*, variance* 1.

We shall see that the unconditional variance is still constant, a property the previous models we evaluated also possess, however the conditional variance now changes with time. A large "shock" value for y_{t-1}^2 will directly increase the conditional variance, allowing for some degree of volatility clustering to be apparent. The innovations ϵ_t can be assumed to have a Gaussian distribution, but can also follow other distributions to better capture properties of the data (see later). Using the properties of ϵ_t being i.i.d. with mean 0 and variance 1, we can derive some interesting features which the ARCH(1) model possesses.
Conditional distribution:

$$f(y_t|\Psi_{t-1}) = h_t^{1/2} f(\epsilon_t|\Psi_{t-1}) = h_t^{1/2}\mathsf{N}(0,1) = \mathsf{N}(0, h_t)$$

Unconditional expectation:

$$\mathsf{E}(y_t) = \mathsf{E}(\mathsf{E}(y_t|\Psi_{t-1})) = \mathsf{E}\left(\mathsf{E}\left(\epsilon_t h_t^{1/2}\right)\right) = \mathsf{E}\left(\mathsf{E}(\epsilon_t) h_t^{1/2}\right) = 0$$

This is since ϵ_t is i.i.d. with mean 0 and h_t is known given the information set (Tsay, 2013).
Unconditional variance:

$$\begin{aligned}
\mathsf{E}\left(y_t^2\right) &= \mathsf{E}\left(\mathsf{E}(y_t^2|\Psi_{t-1})\right) \\
&= \mathsf{E}\left(\epsilon_t^2 h_t\right) \\
&= \mathsf{E}\left(\epsilon_t^2\right)\mathsf{E}\left(h_t\right) \\
&= \mathsf{E}\left(h_t\right) \\
&= \mathsf{E}\left(\alpha_0 + \alpha_1 y_{t-1}^2\right) \\
&= \alpha_0 + \alpha_1 \mathsf{E}\left(y_{t-1}^2\right)
\end{aligned}$$

The fourth equality is true since we have set the variance of innovations ϵ_t to be 1. If the process y_t is stationary, then the unconditional variance is constant and equals:

$$\sigma^2 = \frac{\alpha_0}{1-\alpha_1} \tag{3.3}$$

This gives the condition of $\alpha_1 < 1$ for weak stationarity, since the variance must be positive and finite. We can also derive conditional properties of ARCH(1) models with respect to the information set Ψ_{t-1}, which contains all values of the process y_t up until the most recent observation.

Conditional expectation:

$$\mathsf{E}(y_t|\Psi_{t-1}) = 0$$

Conditional variance:

$$\begin{aligned}\mathsf{E}\left(y_t^2|\Psi_{t-1}\right) &= \mathsf{E}\left(\epsilon_t^2 h_t|\Psi_{t-1}\right)\\ &= h_t\mathsf{E}\left(\epsilon_t^2|\Psi_{t-1}\right)\\ &= h_t\mathsf{E}\left(\epsilon_t^2\right)\\ &= h_t\end{aligned}$$

The second equality is true because h_t is known given the information set; the third because ϵ_t terms are independent of past innovations. Since $\mathsf{E}(y_t|\Psi_{t-1}) = 0$, y_t is known as a *martingale difference*, so it is an uncorrelated sequence.
Covariance:

$$\begin{aligned}Cov(y_t, y_{t+h}) &= \mathsf{E}[(y_t - \mathsf{E}(y_t))(y_{t+h} - \mathsf{E}(y_{t+h}))]\\ &= \mathsf{E}[\mathsf{E}(y_t y_{t+h}|\Psi_{t+h-1})]\\ &= \mathsf{E}[y_{t+h}\mathsf{E}(y_t|\Psi_{t+h-1})]\\ &= 0\end{aligned}$$

The third and fourth equalities come from $y_t \in \Psi_{t+h-1}$ and $\mathsf{E}(y_t) = 0$. This result along with its zero mean implies that ARCH(1) models are almost white noise, except for they do not satisfy the i.i.d. property as shown earlier. It shall prove useful to study the behaviour of the tail of the distribution of this process. If the fourth moment of y_t is finite, and we assume ϵ_t to be normally distributed so that $\mathsf{E}\left(\epsilon_t^4\right) = 3$, then:
Conditional 4^{th} moment:

$$\begin{aligned}\mathsf{E}\left(y_t^4|\Psi_{t-1}\right) &= \mathsf{E}\left(\epsilon_t^4 h_t^2|\Psi_{t-1}\right)\\ &= h_t^2\mathsf{E}\left(\epsilon_t^4|\Psi_{t-1}\right)\\ &= h_t^2\mathsf{E}\left(\epsilon_t^4\right)\\ &= 3h_t^2\end{aligned}$$

Unconditional 4^{th} moment:

$$\begin{aligned}\mathsf{E}\left(y_t^4\right) &= \mathsf{E}\left(\mathsf{E}\left(y_t^4|\Psi_{t-1}\right)\right)\\ &= 3\mathsf{E}\left(h_t^2\right)\\ &= 3\mathsf{E}\left(\left(\alpha_0 + \alpha_1 y_{t-1}^2\right)^2\right)\\ &= 3\mathsf{E}\left(\alpha_0^2 + 2\alpha_0\alpha_1 y_{t-1}^2 + \alpha_1^2 y_{t-1}^4\right)\end{aligned}$$

If y_t is 4^{th} order stationary we can derive the fourth central moment, which we shall denote as $\mu_4 = \mathsf{E}\left(y_t^4\right)$, and use it derive the kurtosis of our ARCH(1) model.

$$\begin{aligned}
\mu_4 &= 3\left[\alpha_0^2 + 2\alpha_0\alpha_1\mathsf{E}(y_{t-1}^2) + \alpha_1^2\mathsf{E}(y_{t-1}^4)\right] \\
&= 3\alpha_0^2\left(1 + \frac{2\alpha_1}{1-\alpha_1}\right) + 3\alpha_1^2\mu_4 \\
&= \frac{3\alpha_0^2 - 3\alpha_0^2\alpha_1 + 6\alpha_0^2\alpha_1}{(1-\alpha_1)(1-3\alpha_1^2)} \\
&= \frac{3\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)}
\end{aligned}$$

From this result, we can obtain the condition for fourth order stationarity that $1-3\alpha_1^2 > 0$, and so overall $0 \leq \alpha_1^2 < {}^1/_3$. Substituting our result into the kurtosis formula and using Equation 3.3, we get:
Unconditional kurtosis:

$$\begin{aligned}
Kurt(y_t) &= \frac{\mu_4}{(Var(y_t)^2)} \\
&= \frac{3\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)}\left(\frac{1-\alpha_1}{\alpha_0}\right)^2 \\
&= 3\frac{(1+\alpha_1)(1-\alpha_1)}{(1-3\alpha_1^2)} \\
&= 3\frac{1-\alpha_1^2}{1-3\alpha_1^2}
\end{aligned} \tag{3.4}$$

Since $1-\alpha_1^2 > 1-3\alpha_1^2$, this means $Kurt(y_t) > 3$ and therefore the excess kurtosis defined as $Kurt(y_t) - 3$ is greater than 0. This means that the distribution of the returns using this model has a "heavier tail" than a Normal distribution, and so the returns distribution is said to be *leptokurtic*. This is despite ϵ_t being assumed to have a Gaussian distribution.

The issue of the conditional variance not being directly observable can be resolved by a simple reformulation. By squaring equation (3.1), and subtracting (3.2) from it, then:

$$y_t^2 - (\alpha_0 + \alpha_1 y_{t-1}^2) = \epsilon_t^2 h_t - h_t$$

If $V_t = \epsilon_t^2 h_t - h_t = y_t^2 - h_t$ is substituted into the conditional variance equation, then we get:

$$y_t^2 = V_t + \alpha_0 + \alpha_1 y_{t-1}^2 \tag{3.5}$$

This looks like an AR(1) process, which is a useful property, because methods used in AR(1) specification can also be applied to the squared returns series in order to specify ARCH(1) models, as shown later in the model building procedure, Section 3.3.3.

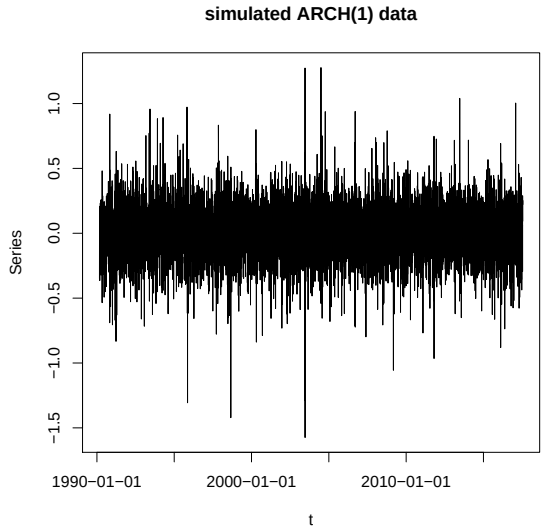
$$\mathsf{E}(V_t|\Psi_{t-1}) = \mathsf{E}(y_t^2 - h_t|\Psi_{t-1}) = \mathsf{E}(y_t^2|\Psi_{t-1}) - \mathsf{E}(h_t|\Psi_{t-1}) = 0$$

Hence V_t is a martingale difference and therefore uncorrelated (Shumway and Stoffer, 2011).

The plot in Figure 3.2 shows a simulated ARCH(1) process with $\alpha_0 = 0.02, \alpha_1 = 0.5$. The simulation was done using the "garchSim" function contained in the "fGarch" package in R. There is some volatility clustering visible in the plot, although the clusters formed from large variance shocks are short-lived due to the process having a very short memory. The sample kurtosis for the data in the plot was 5.70, which is greater than 3, indicating that the distribution of the data had a heavier tail than a Normal distribution.

We found that for small datasets, the excess kurtosis was often close to (but still greater than) 0, indicating it had little heavy-tailed nature. Any extreme values had a large impact on the kurtosis value calculated. Once a larger dataset was simulated, the kurtosis remained on average at a higher value, since there would have been more chance for more extreme values to occur. 10 separate simulations of 500,000 datapoints for this ARCH(1) model had an average kurtosis of 8.44 which is close to the theoretical value of 9 calculated by substituting $\alpha_1 = 0.5$ into Equation (3.4). Therefore, unless the dataset is sufficiently large, measures such as kurtosis must be only viewed as a general indicator as they can vary so much for smaller sets of data.

Figure 3.2: Plot of simulated ARCH(1) data



The ARCH(1) model implies by its definition that there is only correlations between adjacent values of y_t^2. Examining Figure (2.6), the autocorrelations appear to be small, but do not decay very quickly. To capture these characteristics with a simple ARCH(1) model can prove very difficult, since if it takes small parameter values it would make the autocorrelation become insignificant quickly. This shows the need for a model which uses more values than just the most recent squared "shock".

3.3.2 ARCH(q) Models

The ARCH(1) model can be extended to include the q most recent observations to create a process with a longer memory.

Definition 3.3.2. ***Order q Autoregressive Conditional Heteroskedasticity model (ARCH(q))***

$$y_t = \epsilon_t h_t^{1/2} \tag{3.6}$$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i y_{t-i}^2 \tag{3.7}$$

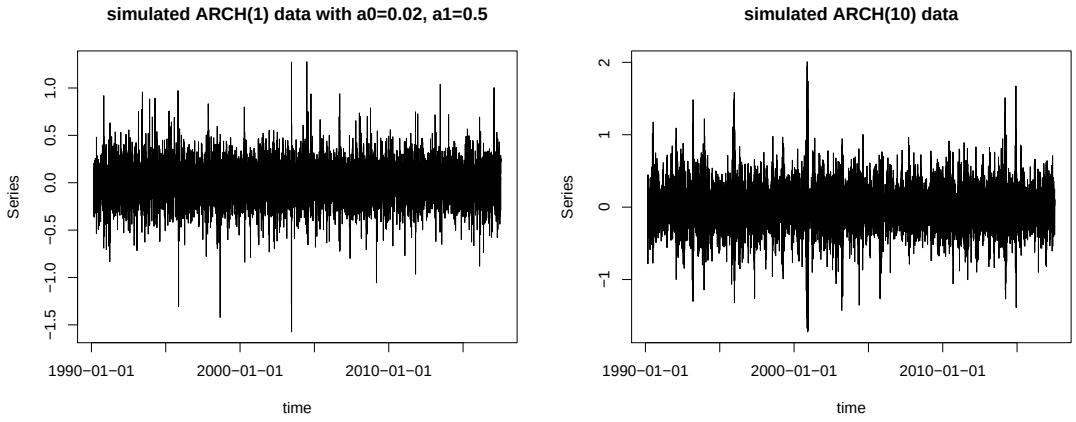
The general properties discussed for ARCH(1) models hold for ARCH(q) models, but the higher order results in restrictions on parameters become more stringent in order to ensure the variance remains finite and positive. An ARCH(q) process is weakly stationary if the roots of the lag polynomial $\alpha(B) = 1 - \sum_{i=1}^{p} \alpha_i B^i$ lie outside the unit circle, or in other words $\sum_{i=1}^{q} \alpha_i < 1$ (Franses and van Dijk, 2000). If this condition is satisfied, the unconditional variance becomes

$$\sigma^2 = \frac{\alpha_0}{1 - \alpha_1 - \cdots - \alpha_q} \tag{3.8}$$

For more information on similar properties, see Milhøj (1985), which contains derivations of conditions for the existence of unconditional moments of ARCH(q) processes.

An example of a simulated ARCH(10) process of 100,000 values is shown in the right-hand plot of Figure 3.3. Comparing it to the left-hand plot containing the ARCH(1) simulation, the longer memory of the conditional variance can be seen by clusters being wider, due to volatility being of a similar magnitude for longer periods of time.

Figure 3.3: Plots of simulated ARCH(1) and ARCH(10) data



3.3.3 ARCH Model Building Process

We shall now outline the necessary steps involved in building an ARCH(q) model and then go into detail on how to conduct each step.

1. Test for serial dependence in the data and build an ARMA model/compute sample mean in order to define the mean equation.

2. Test for the presence of ARCH effects in the residuals of the fitted process defined by the chosen mean equation.

3. If ARCH effects are statistically significant, then attempt to fit mean and conditional variance equations to the data via joint estimation.

4. Evaluate how well the model fits the data. If the model is not good enough then refine it until it is adequate.

5. Use the final model for forecasting future volatilities and other useful analytical techniques.

1. The first step is to decide if any of the basic time series models from Chapter 2 can adequately describe the data. We can test for serial correlations using the visual method of evaluating the significant lines in the acf and pacf of the log returns, or the Ljung-Box tests. If there exists a dependency, then we can choose the best ARMA model based on our analysis.

If there is no serial dependence, a t-test can be conducted for whether the sample mean is significantly different from 0, if it is, then the mean equation is just a constant term. If it is not, then the process has zero mean and no correlations and so we can move on to step 2. We carried out this step in the example from Section 2.4, where we fitted a first order moving average model to the FTSE 100 log returns data.

2. It was shown earlier that after fitting the MA(1) model, the squared residuals still possessed many significant autocorrelations. Therefore, it makes sense to apply a test on the squared residuals correlations to signify if there is a presence of ARCH-type properties, namely conditional heteroskedasticity. Mcleod and Li (1983) suggested a test which uses the same statistic as the Ljung-Box test, but replaces the sample autocorrelations of the residuals with sample autocorrelations of the squared residuals.

Let $\hat{\rho}_k^*$ be defined as:

$$\hat{\rho}_k^* = \frac{\sum_{t=1}^{n} \left(X_t^2 - \hat{\sigma}^2\right)\left(X_{t+k}^2 - \hat{\sigma}^2\right)}{\sum_{t=1}^{n} \left(X_t^2 - \hat{\sigma}^2\right)^2}$$

be the sample autocorrelation of the series X_t^2, where $\hat{\sigma}^2 = \sum_{t=1}^{n} X_t^2/n$. The Mcleod-Li test statistic for a selected maximum lag of M is the defined as follows:

Definition 3.3.3. *Mcleod-Li test*

$$Q^*(M) = n(n+2)\sum_{k=1}^{M}\frac{(\hat{\rho}_k^*)^2}{n-k} \tag{3.9}$$

The test statistic will have a χ^2 distribution with M degrees of freedom under the null hypothesis that there exists no "ARCH effects", meaning no correlations between the squared residuals (Mcleod and Li, 1983).

ARCH effects can also be tested for using a Lagrange Multiplier test applied to the linear regression ARCH model, with the null hypothesis being that $\alpha_1 = \cdots = \alpha_q = 0$. Details of this linear regression interpretation and the test can be found in Engle (1982).

Once conditional heteroskedasticity has been found to be present in the data, a tentative order determination of the ARCH equation can be conducted using the pacf of the squared residuals.

If we use the reformulation using $V_t = y_t^2 - h_t$ presented earlier for ARCH(1) models and apply to the ARCH(q) equation, we get the following representation:

$$y_t^2 = \alpha_0 + \sum_{i=1}^{q}\alpha_i y_{t-i}^2 + V_t \tag{3.10}$$

This looks like an AR(q) equation, and indeed as demonstrated in Bollerslev (1988), the squared terms show the same behaviour of autoregressive models partial autocorrelations cutting off after lag q, except that the V_t term is not independent over time. Taylor (1986) showed that if the fourth moment of y_t exists, the autocorrelations of the squares of an ARCH(q) process follow the same Yule-Walker equations as an AR(q): $\rho_k = \sum_{i=1}^{k}\alpha_i\rho_{k-i}$. Therefore, if we examine the pacf of the squared log returns y_t^2, we can get an idea of the best order of ARCH(q) process to select which would allow us to sufficiently explain the volatility.

3. Estimation for the ARCH model can be carried out using Maximum Likelihood estimation. The innovations in the basic model are assumed to have conditional Normal distribution, but the joint density is not Normal. The joint density of y_t is given by the product of conditional densities $y_t|y_{t-1},\ldots,y_{t-q} \sim \mathsf{N}(0, \alpha_0 + \alpha_1 y_{t-1}^2 + \cdots + \alpha_q y_{t-q}^2)$ for $t > q$.

$$\begin{aligned} L(\boldsymbol{\alpha}|y_1,\ldots,y_q) &= \prod_{t=q+1}^{n} f_{\boldsymbol{\alpha}}(y_t|y_{t-1},\ldots,y_{t-q}) \\ &= \prod_{t=q+1}^{n}\frac{1}{\sqrt{2\pi h_t}}\exp\left(-\frac{y_t^2}{2h_t}\right) \\ &= \prod_{t=q+1}^{n}(2\pi h_t)^{-1/2}\exp\left(-\frac{y_t^2}{2h_t}\right) \end{aligned}$$

where $\boldsymbol{\alpha} = (\alpha_0,\ldots,\alpha_q)$ are the constants to be estimated (Shumway and Stoffer, 2011). The log-likelihood function,$l(\boldsymbol{\alpha})$, to be maximised at time t is therefore the sum of conditional nor-

mal log likelihoods.

$$l(\boldsymbol{\alpha}) = \sum_{t=q+1}^{n} \left(-\frac{1}{2}\log(2\pi h_t) - \frac{1}{2}\frac{y_t^2}{h_t} \right) = \sum_{t=q+1}^{n} \left(-\frac{1}{2}\log 2\pi - \frac{1}{2}\log h_t - \frac{1}{2}\frac{y_t^2}{h_t} \right)$$

The term containing 2π is just a constant and so it can be left out (Bollerslev et al., 1992), leaving the log-likelihood function to be maximised as:

$$l(\boldsymbol{\alpha}) = -\frac{1}{2}\sum_{t=q+1}^{n} \left(\log h_t + \frac{y_t^2}{h_t} \right)$$

4. Once the ARCH model has been fitted, the standardised residuals $\hat{\epsilon}_t = y_t / h_t^{1/2}$ of the model can be used to check for model adequacy. The residuals should form an i.i.d. process, and so we can check this using the Ljung-Box test on $\hat{\epsilon}_t$ and $\hat{\epsilon}_t^2$ to check the mean and volatility equations easily. We can also evaluate how well the initial distribution chosen for ϵ_t assumption fits the data. If we assume a Normal distribution, as we have done in this chapter so far, then a Shapiro-Wilk test for normality or viewing a Q-Q plot for the residuals can illustrate if this assumption was correct.

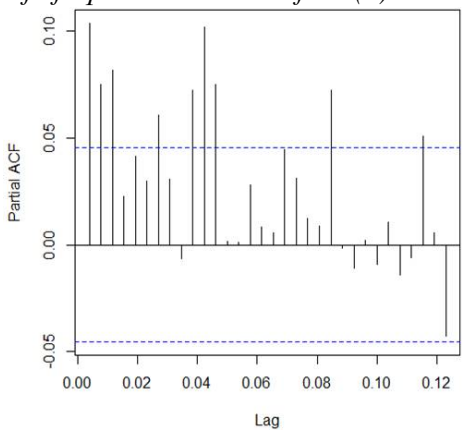
5. The forecasting of ARCH models will be examined in Section 3.3.7

3.3.4 Fitting an ARCH Model to the FTSE 100 Data

We shall now attempt to put these steps into practice using the FTSE 100 log returns used earlier to create an ARMA model with ARCH errors, referred to as an "ARMA-ARCH" model. Firstly, we can test for ARCH effects using the Mcleod-Li test shown earlier. Applying this test to the residuals of the fitted MA(1) model, the p-value was much less than 0.05 for lags 5, 10, 15 and 20. Therefore we reject the null hypothesis of no correlations between squared residuals and so there are ARCH effects present.

We will choose the best ARCH order for the volatility equation by examining the pacf of the squared residuals of the fitted MA(1) model, given in Figure 3.4.

Figure 3.4: Pacf of squared residuals of MA(1) model of FTSE data



We can see that the 3^{rd}, 7^{th} and $10 - 12^{th}$ lags are significant in the plot. We use these 5 orders in our set of candidate models to decide which fares best under analysis. We then proceed to fit an MA(1)-ARCH(q) model, $q = 3, 7, 10, 11, 12$.

We can quickly discard the order 3 model, as the Mcleod-Li statistics of the squared residuals were all still too large. This is demonstrated in the left-hand plot of Figure 3.5, where the standardised residuals still have multiple significant partial autocorrelations. The ARCH(7) model results had similar characteristics, and so we move on to the 3 larger order candidate processes to be evaluated.

Figure 3.5: Pacf of squared residuals of MA(1) model with ARCH(3) and ARCH(11) errors

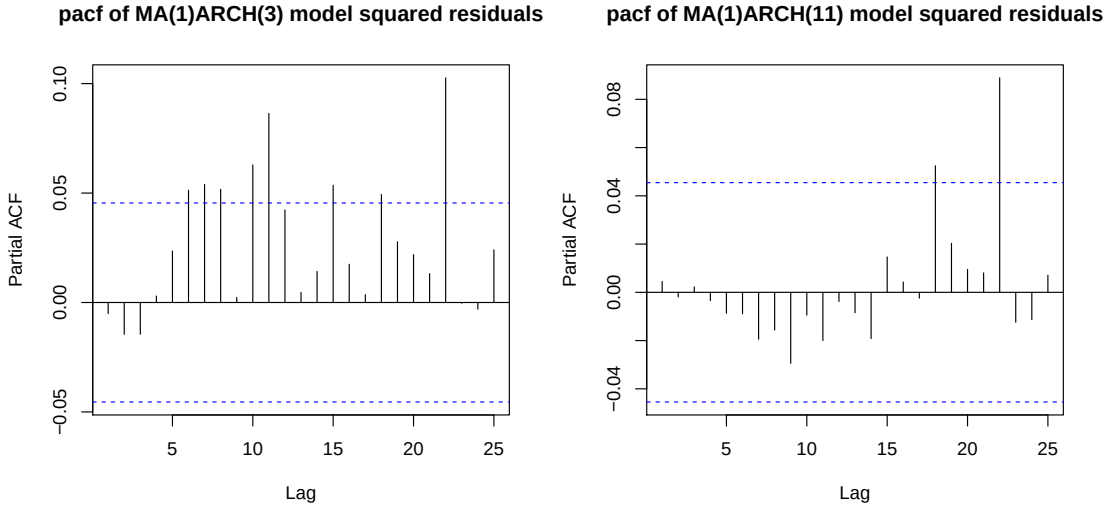


Table 3.1: Sample Mcleod-Li (Equation (3.9)) p-values and criterion values of MA(1)-ARCH models for FTSE 100 log returns

ARCH order	Q(10) p-value	Q(15) p-value	Q(20) p-value	AIC	BIC
10	0.9248	0.4662	0.2925	-6.8951	-6.8564
11	0.9726	0.9896	0.9126	-6.9073	-6.8657
12	0.9774	0.9904	0.9149	-6.9062	-6.8616

From the values in Table 3.1, it appears that either the ARCH(11) model or ARCH(12) are the best. The AIC and BIC values are minimised by the ARCH(11) model. Not shown is that all three models have insignificant parameters at lags 4 and 9, with almost all others being significant at the 5% level or lower in each model. However, the 12^{th} parameter in the ARCH(12) model is insignificant, which implies it is unnecessary, leading us to conclude that the ARCH(11) model is the optimal one. Viewing the pacf of squared residuals in the right hand plot of Figure 3.5, we can see that almost all correlations have been removed by this model,

showing the benefit of volatility models for financial data. This model still has weaknesses however, since it possessed 3 insignificant parameters.

The final model for the returns $r_t = 100 \log(X_t/X_{t-1})$ that we finish with is:

$$
\begin{aligned}
r_t &= 5.52 \times 10^{-4} + y_t + 0.0765 y_{t-1} \\
y_t &= h_t^{1/2} \epsilon_t \\
h_t &= 2.43 \times 10^{-5} + 0.0472 y_{t-1}^2 + 0.0637 y_{t-2}^2 + \cdots + 0.110 y_{t-11}^2
\end{aligned}
$$

3.3.5 Evaluation of ARCH Models

If we are to use these models for analysis it is useful to evaluate their worth, and whether they would give realistic results that could be used for analysis.

Strengths:

1. The model produces volatility clusters which replicate the similar behaviour exhibited by log returns.

2. The distribution has heavy tails which gives a more accurate representation than a Gaussian distribution.

3. They give a better overall explanation for correlations within the data, leading to more accurate predictions than traditional constant-variance time series models.

Weaknesses:

1. ARCH models tend to require high orders to sufficiently explain all correlations in the data. For example, the prior analysis concluded the model required 11 previous squared terms to be adequate, and yet it still left some correlations in the residuals.

2. Related to the previous point is that the ARCH models tend to require an arbitrary declining lag structure. This is to account for the long memory of financial data as well as to ensure that the conditions for non-negative constants remain satisfied.

3. These models give no indication of what caused these correlations or trends in the data.

4. ARCH models use the square of log returns and so they assume that positive and negative jumps have equal impact on volatility - in reality there is an argument that negative effects have a stronger impact (see for example Christie (1982) for more details on the existence of the so-called *leverage effect*).

5. They are restrictive on choice of parameters in order to have finite variance or fourth moments. This could impact how well they emulate the heavy-tailed nature of the distribution of the log returns and could mean the chosen weights are not ideal.

3.3.6 Alternative Distributions for ϵ_t

There is some evidence that despite ARCH models exhibiting a leptokurtic property, they still do not sufficiently account for the heavy tailed nature of financial returns. To resolve this, we can use alternative distributions for ϵ_t which have a heavier tail than a normal distribution.

The most popular alternative choice to try first is the standardised Student-t distribution, as proposed by Bollerslev (1987). If t_ν is a Student t distribution with ν degrees of freedom, then the variance of t_ν is $\nu/(\nu-2)$ if $\nu > 2$. We then use the assumption $\epsilon_t\sqrt{\nu/(\nu-2)} \sim t_\nu$ for our innovations, which are equal to $y_t/h_t^{1/2}$ in the ARCH model, so that their variance remains equal to 1 as was specified in Definition 3.3.1 (Brockwell and Davis, 2002).

We shall now attempt to fit the same MA(1)-ARCH(11) model to the FTSE 100 data we selected earlier, but with the Normal distribution assumption for ϵ_t changed to Student-t. The model chose the best Student-t distribution to be one with 9.42 degrees of freedom.

Figure 3.6: Q-Q plots of standardised residuals of MA(1) model with ARCH(11) errors with Normal and Student t distributions for ϵ_t

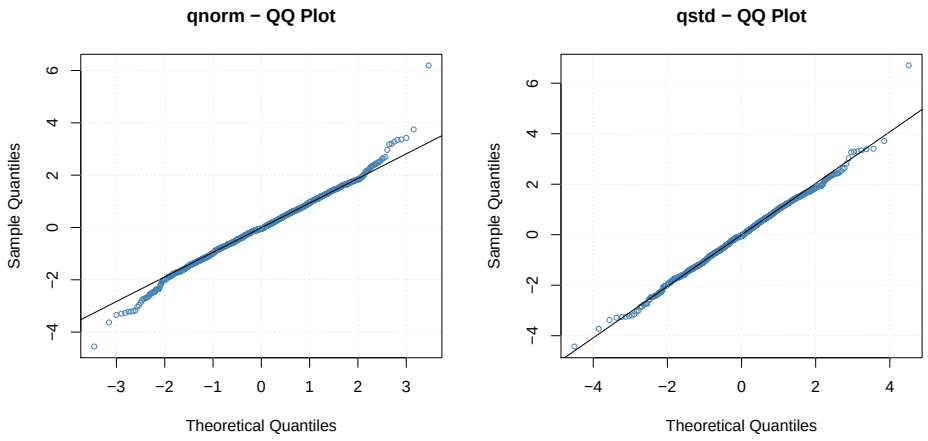


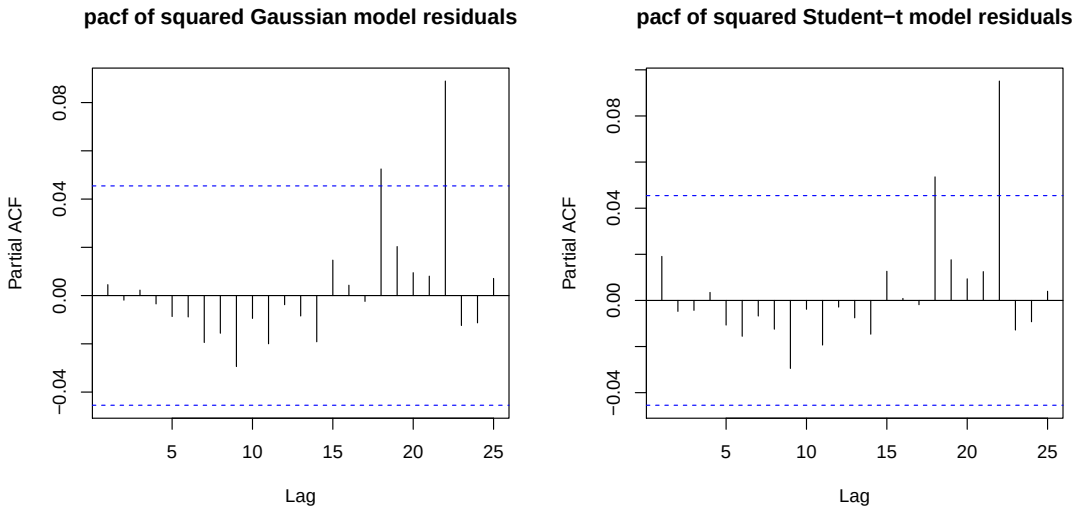
Figure 3.6 shows that the change of distribution assumption resulted in the heavy tails of the distribution being almost completely accounted for. There is only one outlier which is not close to the diagonal line with equation $y = x$, so from this measure we can say that the distribution assumption was a good choice.

Looking at the parameter estimates however, the 1^{st}, 4^{th}, 5^{th} and 9^{th} parameters are insignificant. For instance, the α_4 estimate has a p-value of 0.749, suggesting this coefficient is not important in the model. The Ljung-Box statistics for the squared standardised residuals have p-values close to 1 for lags 10, 15 and 20 so this test, equivalent to the Mcleod-Li test, suggests the residuals behave like white noise. The AIC and BIC values for the model are -6.9305 and -6.8859 respectively. These are both smaller than the values for the previous model shown in Table 3.1, so these measures suggest this is a superior model.

We can also examine the pacf of squared standardised residuals to see if the change of distribution affected these correlations. Figure 3.7 shows this alteration had very little impact

on the correlations, including the 2 significant correlations remaining in that state.

Figure 3.7: Pacf of squared standardised residuals of MA(1) model with ARCH(11) errors with Normal and Student t distributions for ϵ_t



Overall, these results signify that this new model is slightly superior than the original Gaussian one. However, it still has the same issue of not explaining all correlations between residuals as well as requiring many parameters with some remaining insignificant.

Other possible distributions to use are a skew Student t distribution, in order to account for any additional skewness existing in the returns, and a Generalised Error Distribution (GED). Texts on ARCH models such as Tsay (2013) have more detail on how to estimate and examples of them in practice.

3.3.7 Forecasting ARCH Models

An important purpose of ARCH models is to use them for predicting financial data. The fact that they can model the heteroskedastic nature commonly found in returns data implies that they should provide better and more realistic forecasts of volatility than basic models which assume it is constant. Comparing how well different models perform with forecasting is part of the model evaluation process and gives more indication of which model is best overall.

The 1-step forecast for an ARCH model of order q uses just the current values of h_t and y_t and is computed at time t is as follows:

$$h_t(1) = \alpha_0 + \alpha_1 y_t^2 + \cdots + \alpha_q y_{t+1-q}^2 \tag{3.11}$$

If the process is weakly stationary, then as shown in Equation (3.8), $\sigma^2 = \alpha_0/(1-\alpha_1-\cdots-\alpha_q)$.

Substituting this formula into the equation above, we get:

$$h_t(1) = \sigma^2(1 - \alpha_1 - \cdots - \alpha_q) + \alpha_1 y_t^2 + \cdots + \alpha_q y_{t+1-q}^2 \tag{3.12}$$

This is a weighted average of the long-term variance σ^2 and the q most recent squared terms of the process $\{y_t\}$. This appears straightforward to compute, since it just involves substituting computed values into the formula. The issue with ARCH models is that for l-step forecasts where $l > 1$ we cannot straight away use a recursive method of computing, as was used in autoregressive models, since the conditional variance does not appear on the right hand side of the equation. Once again this problem stems from the fact that the conditional variance cannot be directly observed.

We must therefore decide what value to choose as a replacement for y_{t+1}^2 in the forecast equation. One option would be to set it to zero and use a naive approach which assumes the conditional variance is constant for all future values. This would take away the heteroskedastic nature of the model and so is not very helpful.

There is an argument to replace y_{t+1}^2 by our forecast for the conditional variance $h_t(1)$, since $\mathsf{E}(y_{t+1}^2|\psi_t) = h_{t+1}$. This leads to a recursive method which involves straightforward computation.

$$\begin{aligned}
h_t(l) &= \mathsf{E}(y_{t+l}^2|\psi_t) \\
&= \mathsf{E}\left[\mathsf{E}(h_{t+l}\epsilon_{t+l}^2|\psi_{t+l})|\psi_t\right] \\
&= \mathsf{E}\left[h_{t+l}\mathsf{E}(\epsilon_{t+l}^2)|\psi_t\right] \\
&= \mathsf{E}\left[h_{t+l}|\psi_t\right] \\
&= \mathsf{E}\left[\alpha_0 + \alpha_1 y_{t+l-1}^2 + \cdots + \alpha_q y_{t+l-q}^2|\psi_t\right] \\
&= \alpha_0 + \alpha_1 h_t(l-1) + \cdots + \alpha_q h_t(l-q)
\end{aligned} \tag{3.13}$$

where $h_t(l-i) = y_{t+i}^2$ for $l - i < 0$ and we have used the assumption that ϵ_t has mean 0 and variance 1 (Cryer and Chan, 2008). So in practice, after estimating an ARCH(q) model, the l-step forecast can be calculated using the following formula:

$$h_t(l) = \hat{\alpha}_0 + \sum_{i=1}^{q} \hat{\alpha}_i h_t(l-i) \tag{3.14}$$

with $h_t(l-i) = y_{t+l-i}^2$ if $l - i < 0$.

We will now move on from ARCH models to look at a new related class of models proposed just 4 years later, but we will apply ARCH models to a different data set in Chapter 5 to further see them being used in practice.

3.4 GARCH Models

Bollerslev (1986) introduced a new model named the GARCH model to build on the ARCH model and improve on its shortcomings. As we highlighted in the second point of the list of the weaknesses of ARCH models in Section 3.3.5, the ARCH model tends to require high orders to fit the data, and with this requires complicated restrictions on parameters. The GARCH model incorporates lagged values of the conditional variance to account for the long memory of the volatility. These models have a more flexible lag structure allowing for more freedom in fitting the model, which improves the former model's drawback.

GARCH models have turned out to be one of the most popular classes of volatility models and remain very relevant in financial time series analysis 30 years on. We will examine these models in detail to demonstrate their utility and application to real financial data sets.

Definition 3.4.1. ***General Autoregressive Conditional Heteroskedasticity Model (GARCH(p,q))***

$$y_t = \epsilon_t h_t^{1/2} \tag{3.15}$$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i y_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{3.16}$$

$$= \alpha_0 + A(B)y_t^2 + C(B)h_t \tag{3.17}$$

$\alpha_0 > 0$; α_i, $\beta_j \geq 0 \; \forall i \in (1, .., q)$, $\forall j \in (1, ..., p)$; $p \geq 0$; $q > 0$

The backshift operators $A(B)$ and $C(B)$ in are defined as $A(B) = \alpha_1 B + \cdots + \alpha_q B^q$ and $C(B) = \beta_1 B + \cdots + \beta_p B^p$. Evidently, if $p = 0$, the model becomes an ARCH(q) model and if $p = q = 0$ then y_t is a white noise process, since it would become a constant multiplied by ϵ_t which is an i.i.d. process.

The stochastic process $y_t|\Psi_{t-1}$ can be $\mathsf{N}(0, h_t)$ but as shown later, better fits to the models may be found by using alternative distributions instead which have heavier tails in order to further capture this nature of the log returns.

3.4.1 GARCH Properties

We shall now outline some of the characteristics of GARCH models, some of which are similar to ARCH models.

As long as $\sum_{i=1}^{\max(p,q)}(\alpha_i + \beta_i) < 1$, which is also the necessary and sufficient condition for these models to be weakly stationary, then GARCH models have the following three properties (proven in the Appendix of Bollerslev (1986)):

1. $\mathsf{E}(y_t) = 0$
2. $Var(y_t) = \alpha_0(1 - A(1) - C(1))^{-1}$
3. $Cov(y_s, y_t) = 0 \; \forall s \neq t$

The issue of the conditional variance not being a directly observable quantity can be resolved in the same way as was done for the ARCH models. Let $V_t = y_t^2 - h_t = h_t(\epsilon_t^2 - 1)$, and ϵ_t be i.i.d. $\mathsf{N}(0,1)$. This means V_t is a χ_1^2 random variable shifted to have zero mean and so it is a martingale difference with zero correlation. By substituting this into the function for h_t in Equation (3.13), we obtain:

$$\begin{aligned} y_t^2 &= V_t + \alpha_0 + \sum_{i=1}^{q} \alpha_i y_{t-i}^2 + \sum_{j=1}^{p} \beta_j y_{t-j}^2 - \sum_{j=1}^{p} \beta_j (y_{t-j}^2 - h_{t-j}) \\ &= V_t + \alpha_0 + \sum_{i=1}^{q} \alpha_i y_{t-i}^2 + \sum_{j=1}^{p} \beta_j y_{t-j}^2 - \sum_{j=1}^{p} \beta_j (V_{t-j}) \\ &= V_t + \alpha_0 + \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) y_{t-i}^2 - \sum_{j=1}^{p} \beta_j V_{t-j} \\ &= V_t + \alpha_0 + \sum_{i=1}^{\max(p,q)} \phi_i y_{t-i}^2 + \sum_{j=1}^{p} \theta_j V_{t-j} \end{aligned} \tag{3.18}$$

where $\alpha_i \equiv 0$ if $q < p$ for $i > q$, and $\beta_i \equiv 0$ if $q > p$ for $j > p$ (Bollerslev, 1988). Therefore, a GARCH(p, q) process can be represented by an ARMA($\max(p, q)$,p) process involving the series y_t^2.

If y_t is second-order stationary, using that $\mathsf{E}(V_t) = 0$, we can get an equation for the unconditional variance:

$$\mathbf{E}(y_t^2) = \frac{\alpha_0}{1 - \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i)} = \sigma^2 \tag{3.19}$$

Using the ARMA representation, Bollerslev (1986) derived the autocorrelation function for y_t^2 in a GARCH(p, q) process to be in the same format as the Yule-Walker equations:

$$\rho_k = \sum_{i=1}^{\max(p,q)} \phi_i \rho_{k-i} \qquad k \geq p+1 \tag{3.20}$$

where $\phi_i = \alpha_i + \beta_i \; \forall i$, $\alpha_i \equiv 0$ for $i > q$, and $\beta_i \equiv 0$ for $i > p$. They state that the partial autocorrelation of y_t^2 generally stays positive but decays down.

We shall briefly examine some of the properties of the GARCH(1,1) model, defined as:

$$y_t = \epsilon_t h_t^{1/2} \tag{3.21}$$

$$h_t = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 h_{t-1} \tag{3.22}$$

The advantage which GARCH models have over the original ARCH models can be demonstrated by the following reformulation using successive substitution.

$$
\begin{aligned}
h_t &= \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1(\alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 h_{t-2}) \\
&= \alpha_0(1+\beta_1) + \alpha_1 y_{t-1}^2 + \alpha_1\beta_1 y_{t-2}^2 + \beta_1^2 h_{t-2} \\
&= \alpha_0(1+\beta_1) + \alpha_1 y_{t-1}^2 + \alpha_1\beta_1 y_{t-2}^2 + \beta_1^2(\alpha_0 + \alpha_1 y_{t-3}^2 + \beta_1 h_{t-3}) \\
&= \alpha_0(1+\beta_1+\beta_1^2) + \alpha_1 y_{t-1}^2 + \alpha_1\beta_1 y_{t-2}^2 + \alpha_1\beta_1^2 y_{t-3}^2 + \beta_1^3 h_{t-3} \\
&\vdots \\
&= \alpha_0 \sum_{i=0}^{\infty} \beta_1^i + \alpha_1 \sum_{i=0}^{\infty} \beta_1^i y_{t-i-1}^2 \\
&= \frac{\alpha_0}{1-\beta_1} + \alpha_1 \sum_{i=0}^{\infty} \beta_1^i y_{t-i-1}^2 \qquad (3.23)
\end{aligned}
$$

The final equality is using the properties of geometric series. This is now in the form of an ARCH(∞) process which, as long as $\beta < 1$, gives declining weights for higher lagged squared terms (Andersen et al., 2006). This shows that a GARCH(1,1) model is preferable to a high order ARCH model, since it possesses a long-term lag structure, yet it requires only 3 parameter values to be specified.

It is also useful to examine the kurtosis of a GARCH(1,1) model as was done earlier for the ARCH(1) model. Bollerslev (1986) specifies the following conditions for the existence of even moments of a GARCH(1,1) process.

Theorem 3.4.1. *- The necessary and sufficient condition for the $2m^{th}$ moment to exist is that $\mu(\alpha_1, \beta_1, m) = \sum_{j=0}^{m} a_j \alpha_1^j \beta_1^{m-1} < 1$, where $a_0 = 1$ and $a_j = \prod_{i=1}^{j}(2i-1)$.*

Using these formulae, we can derive the equation for kurtosis using the recursive formula given as:

$$
\mathsf{E}(y_t^{2m}) = a_m \left[\sum_{n=0}^{m-1} \alpha_0^{m-n} \mathsf{E}(y_t^{2n}) \binom{m}{m-n} \mu(\alpha_1, \beta_1, n) \right] / \left[1 - \mu(\alpha_1, \beta_1, m)\right] \qquad (3.24)
$$

This is a complicated formula and so firstly we shall work out the equations for the operator μ:
$\mu(\alpha_1, \beta_1, 0) = 1$
$\mu(\alpha_1, \beta_1, 1) = 1 \cdot \beta_1 + 1 \cdot \alpha_1$
$\mu(\alpha_1, \beta_1, 2) = \sum_{j=0}^{2} a_j \alpha_1^j \beta_1^{2-j} = \beta_1^2 + \binom{2}{1} a_1 \alpha_1 \beta_1 + \binom{2}{2} a_2 \alpha_1^2 = \beta_1^2 + 2\alpha_1\beta_1 + 3\alpha_1^2 = M$
Now using these results, and the fact that variance of a GARCH(1,1) model is $\alpha_0/(1-\alpha_1-\beta_1)$, we can derive the fourth moment, and hence, the kurtosis of a GARCH(1,1) process.

$$
\begin{aligned}
\mu_4 &= 3\left[\alpha_0^2 + 2\alpha_0 \mathsf{E}(y_t^2)(\beta_1+\alpha_1)\right]/(1-M) \\
&= 3\left[\alpha_0^2 + 2\alpha_0\left(\frac{\alpha_0}{1-\alpha_1-\beta_1}\right)(\beta_1+\alpha_1)\right]/(1-M) \\
&= 3\alpha_0^2\left(1+\frac{2\beta_1+\alpha_1}{1-\alpha_1-\beta_1}\right)/(1-M) \\
&= 3\alpha_0^2\left(\frac{1-\alpha_1-\beta_1+2\beta_1+\alpha_1}{1-\alpha_1-\beta_1}\right)/(1-M) \\
&= \frac{3\alpha_0^2(1+\alpha_1+\beta_1)}{(1-\alpha_1-\beta_1)(1-M)}
\end{aligned}
$$

$$
\begin{aligned}
Kurt(y_t) &= \frac{\mu_4}{\left(\mathsf{E}y_t^2\right)^2} \\
&= \frac{3\alpha_0^2(1+\alpha_1+\beta_1)}{(1-\alpha_1-\beta_1)(1-(\beta_1^2+2\alpha_1\beta_1+3\alpha_1^2))} \times \frac{(1-\alpha_1-\beta_1)^2}{\alpha_0^2} \\
&= 3\frac{(1-\alpha_1-\beta_1)(1+\alpha_1+\beta_1)}{1-(\beta_1^2+2\alpha_1\beta_1+3\alpha_1^2)} \\
&= 3\frac{1-\alpha_1^2-2\alpha_1\beta_1-\beta_1^2}{1-(\beta_1^2+2\alpha_1\beta_1+\alpha_1^2)-2\alpha_1^2} \\
&= 3\frac{1-(\alpha_1+\beta_1)^2}{1-(\alpha_1+\beta_1)^2-2\alpha_1^2} > 3 \qquad (3.25)
\end{aligned}
$$

This is finite if $(\alpha_1+\beta_1)^2 + 2\alpha_1^2 < 1$ which Bollerslev (1986) also showed to be the condition for the unconditional fourth moment to be finite without the assumption of a Gaussian distribution. Therefore the excess kurtosis ($Kurt(y_t) - 3$) is greater than 0 and so proves that GARCH(1,1) models retain the leptokurtic property possessed by ARCH models which was seen earlier. For more detail on conditions for the moments in GARCH(1,1) models, see Ling and McAleer (2002).

Bollerslev (1988) derived, under the same condition $(\alpha_1+\beta_1)^2+2\alpha_1^2 < 1$ mentioned above, that the autocorrelations of the process y_t^2 for a GARCH(1,1) model with Gaussian innovations.

$$
\begin{aligned}
\rho_1 &= \frac{\alpha_1(1-\alpha_1\beta_1-\beta_1^2)}{1-2\alpha_1\beta_1-\beta_1^2} \qquad (3.26) \\
\rho_k &= (\alpha_1+\beta_1)^{k-1}\rho_1 \qquad k = 2,3\ldots \qquad (3.27)
\end{aligned}
$$

Of course if $\alpha_1 = 0$, then all correlations are zero, which means an ARCH-type model is not necessary. We shall see later that $\alpha_1+\beta_1$ is often close to 1, and so the equation above shows that the autocorrelations ρ_k of a GARCH(1,1) process will decay slowly. This is a desired general feature for models of financial time series, although Ding and Granger (1996) found this is not fully accurate for financial time series. They found for absolute returns the sample

autocorrelation function decreases at fast rate for small lags, and then decreases very slowly but remaining positive at significant values, which is different behaviour to an exponentially decreasing function.

Ding and Granger (1996) investigated the behaviour of autocorrelations of y_t^2, where the sum of α_1 and β_1 is less than 1 so the process is still weakly stationary, but $(\alpha_1 + \beta_1)^2 + 2\alpha_1^2 \geq 1$ so that the fourth moment is not finite. They discovered that if the process started a long time ago, perhaps if the data-set is large, then the autocorrelations $\tilde{\rho}$ in a process which does not satisfy this condition can be approximated as follows:

$$\tilde{\rho}_1 \approx \alpha_1 + \beta_1/3 \tag{3.28}$$

$$\tilde{\rho}_k \approx (\alpha_1 + \beta_1)^{k-1}\tilde{\rho}_1 \qquad k = 2, 3 \ldots \tag{3.29}$$

This means that autocorrelations still decrease exponentially, and so this property still holds even without a finite fourth moment.

If $(\alpha_1 + \beta_1)^2 + 2\alpha_1^2 = 1$, then we can expand this equation to get $3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 = 1$. Replacing $1 - \alpha_1\beta_1 - \beta_1^2$ by $3\alpha_1^2 + \alpha_1\beta_1$ in the numerator of Equation (3.26), and $1 - 2\alpha_1\beta_1 - \beta_1^2 = 3\alpha_1^2$ into the denominator, results in:

$$\rho_1 = \frac{\alpha_1(3\alpha_1^2 + \alpha_1\beta_1)}{3\alpha_1^2} = \alpha_1 + \beta_1/3 = \tilde{\rho}_1$$

which means that equations (3.27) and (3.29) are equivalent. Therefore, in this scenario the autocorrelation function behaves as a continuous function of α_1 and β_1, signifying that the autocorrelations do not change drastically when the fourth moment does not exist.

3.4.2 Building GARCH Models

The GARCH model building process is similar to the ARCH process specified earlier. However, in this situation the order of the process cannot be obtained from the pacf of squared residuals, because it can not be compared in the same way to an AR model.

When building a volatility model, the analyst must decide how complex to make the model. They could simply remove the mean and say the mean-corrected log returns follow a GARCH model. They could suggest the log returns follow a simple autoregressive model, with the error term being a GARCH process. The third option is to try an ARMA model with GARCH errors, also known as an ARMA-GARCH model. We shall be attempting to fit this third type, as it appears to fit the financial data which we are analysing in this project the best. The objective will be to specify the orders and parameter values of the following time series model:

$$r_t = \mu + \phi_1 r_{t-1} + \cdots + \phi_u r_{t-u} + y_t + \theta_1 y_{t-1} + \cdots + \theta_v y_{t-v} \tag{3.30}$$

$$y_t = h_t^{1/2}\epsilon_t \tag{3.31}$$

$$h_t = \alpha_0 + \alpha_1 y_{t-1}^2 + \cdots + \alpha_q y_{t-q}^2 + \beta_1 h_{t-1} + \cdots + \beta_p h_{t-p} \tag{3.32}$$

The ARMA order identification methods described in Chapter 2 can be applied to the squared residuals of the fitted mean equation model to suggest candidate GARCH models for the volatility equation (Cryer and Chan, 2008). The AIC and BIC values can be used, but once again they can prove to be time-consuming to calculate (Bollerslev, 1988). There are not many texts available concerning GARCH order specification, but it has been suggested in (Bollerslev, 1988) that the correlations in squared returns can be used in a similar fashion to the method for ARCH models.

We can use the ARMA-type formulation for h_t derived earlier in Equation (3.18) to attempt to decide the order of the GARCH(p, q) model.

$$\begin{aligned} y_t^2 &= V_t + \alpha_0 + \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) y_{t-i}^2 - \sum_{j=1}^{p} \beta_j V_{t-j} \\ &= V_t + \alpha_0 + \sum_{i=1}^{\max(p,q)} \phi_i y_{t-i}^2 + \sum_{j=1}^{p} \theta_j V_{t-j} \end{aligned} \tag{3.33}$$

where $\alpha_i + \beta_i = \phi_i$, $-\beta_j = \theta_j$. Therefore creation of an ARMA-GARCH model can be viewed as the application of two ARMA models, the first applied to the process r_t, and the second on its squared innovations y_t^2. Therefore we can conduct some tentative GARCH order determination by viewing the EACF table for the squared residuals of the ARMA model, or simply on the log returns if an ARMA mean model is not required. This method must be used with caution as there is room for error in interpretation if the table's pattern is not clear, and so the model must always be checked after being fitted.

Since the order of this ARMA model will be $(\max(p, q), p)$ there is the possibility for the value of q to be masked if we use this ARMA version. This can potentially be resolved by fitting the GARCH(p, p) model and then examining whether the final parameter(s) of the ARCH part of the equation are insignificant.

When selecting the GARCH order, the most reliable method would be to use the candidate models and fit the full model containing the mean and volatility equations together and then test the effectiveness of these models in practice. Conducting the estimation of each equation separately assumes that they are independent of each other which is technically incorrect, yet this can give a general idea of potential orders which may be good enough to model the data. If parameters in the conditional mean or variance equations are shown to be insignificant, then they can be dropped and the model refined until it appears adequate.

In practice, it is often found that low order GARCH models are sufficient for modelling the conditional variance, and in most cases a GARCH(1,1) model is adequate (Bollerslev et al., 1992). We shall again use the FTSE 100 data and try to fit a GARCH model to it to demonstrate the advantage it has over the original ARCH volatility model.

As with the ARCH model, alternative conditional distributions for ϵ_t can be used to further attempt to capture the leptokurtic property of financial data, using fatter-tailed distributions such

as the Student-t distribution in place of the Gaussian distribution. We shall see application of this later in Chapter 5.

3.4.3 Fitting a GARCH Model to the FTSE 100 Data

We shall use the methods we have recently described to try identify which GARCH order to use on the FTSE 100 data which we used earlier.

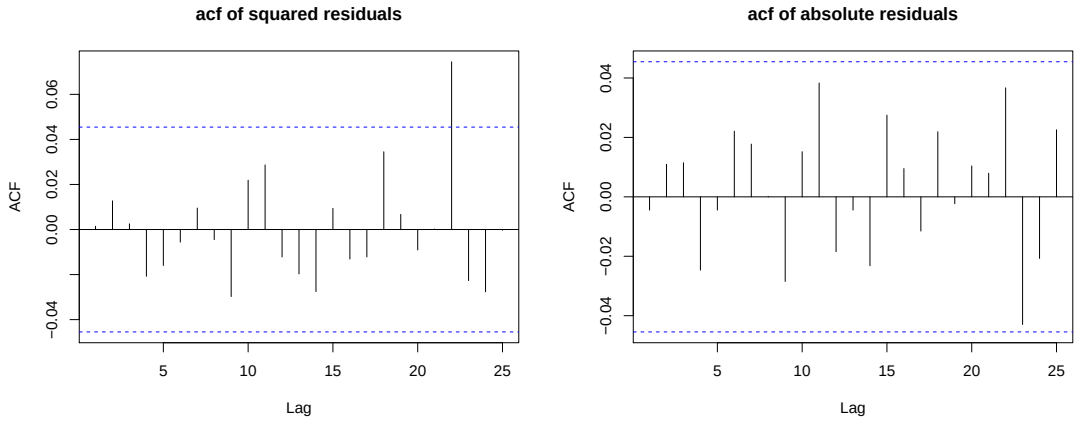
The EACF table for the squared residuals of the fitted MA(1) model is shown in Table 3.2, and it seems to suggest an ARMA$(1,1)$ model and hence a GARCH$(1,1)$ model. If instead we simply used the log returns, we get almost exactly the same table, so fitting this model did not affect the pattern of squared data very much and hence the order obtained with this method.

Table 3.2: Sample EACF table for the squared residuals of the FTSE 100 MA(1) model

	MA							
AR	0	1	2	3	4	5	6	7
0	X	X	X	O	X	X	X	X
1	X	O	X	O	O	O	O	X
2	X	X	O	O	O	O	O	O
3	X	X	X	O	O	O	O	O
4	X	X	X	X	O	O	O	O
5	X	X	O	X	O	O	O	O

A MA(1)-GARCH$(1,1)$ model was fitted to the log returns, and we can now proceed with evaluating it. The problem of significant correlations in the squared and absolute residuals of the original moving average model identified in the original MA(1) model has been almost completely resolved by the GARCH model. Figure 3.8 shows that there is now only one significant autocorrelation remaining in the squared residuals at lag 20. This shows that the conditional variance equation was effective in explaining the dependence between previous values.

Figure 3.8: Acf of squared and absolute residuals of MA(1) GARCH(1,1) model for FTSE data



The Ljung-Box and Mcleod-Li statistics given in Table 3.3 are now all small, with high p-values all much greater than 0.05. This signifies the autocorrelations between residuals are insignificant, which further confirms that this model is adequate, and explains the data well.

Table 3.3: Ljung-Box and McLeod-Li test results for the FTSE 100 MA(1)-GARCH(1,1) model

lag	10	15	20
Q	7.558	16.65	23.59
p-value	0.6719	0.3402	0.261
Q^*	4.417	8.552	11.63
p-value	0.9266	0.8998	0.9282

Therefore the final model for the FTSE 100 data we conclude with is:

$$\begin{aligned} r_t &= 0.0490 + y_t + 0.0861 y_{t-1} \\ y_t &= \epsilon_t h_t^{1/2} \\ h_t &= 0.00890 + 0.0458 y_{t-1}^2 + 0.941 h_{t-1} \end{aligned}$$

Both α_1 and β_1 were significant at the 0.1% level, but the α_0 value was only significant at the 10% level with a standard error of 0.00460. This could suggest that the true value should be 0, however using the equation we refer to below, that would give an unconditional variance of 0 which is unrealistic.

We can compare the sample variance of the FTSE 100 data to the unconditional variance for our GARCH model using Equation (3.19). Using our calculated GARCH parameters, we can obtain the unconditional variance σ^2:

$$\sigma^2 = \frac{\hat{\alpha}_0}{1 - \sum_{i=1}^{\max(p,q)} (\hat{\alpha}_i + \hat{\beta}_i)} = \frac{0.00890}{(1 - 0.0458 - 0.941)} = 0.674$$

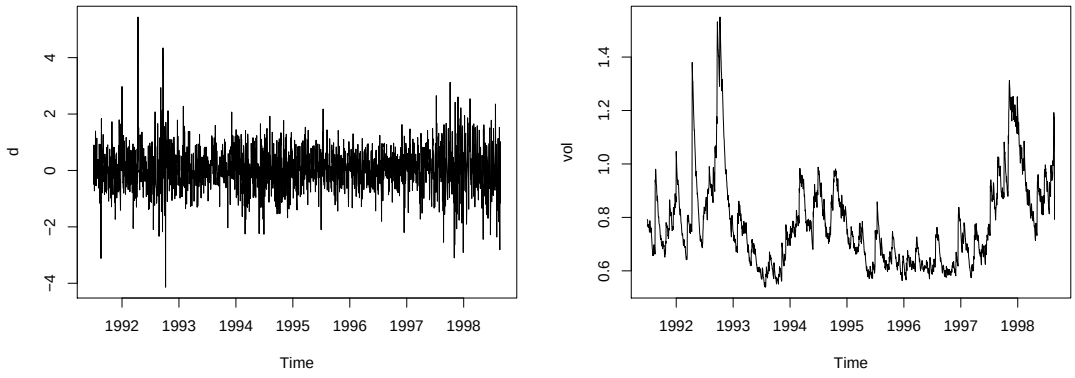
The sample variance of the log returns was calculated by R as 0.633, which is a little smaller than our model's value, showing that there still remains some imperfections, but it is close enough.

We can plot the volatility of this model to demonstrate the time-varying nature which it adopted. Figure 3.9 shows the log returns and the volatility of the GARCH model for each time point in the 6 and a half year time period. The higher volatility spikes in the year 1992 seem to coincide with the large positive and negative spikes in the log returns plot. There is an extended period of higher volatility around 1998 which is also mirrored by the right hand plot which sees a wider spike around the same time point. The changing volatility shown is clearly much more realistic than a simple flat line, which it would have been without the GARCH model.

In summation, we have discovered through our analysis the benefits that GARCH models provide. The MA(1)-GARCH(1,1) model is clearly the best model we have found so far - it has

resolved the issues with correlations in the squared residuals by providing a realistic model for the volatility process.

Figure 3.9: Plots of the log returns of the FTSE 100 data (left), and the volatility of the MA(1)-GARCH(1,1) model (right)



3.4.4 Forecasting GARCH Models

As with the ARCH model, many analysts would like to be able to forecast the volatility of a set of data after having fitted a GARCH model. The one-step-ahead forecast, which uses known values of y_t and h_t, is given simply as:

$$h_t(1) = \alpha_0 + \alpha_1 y_t^2 + \beta_1 h_t \tag{3.34}$$

A formula for the future conditional variance of a GARCH(1,1) model for integers of $l > 1$ can be derived by substituting $y_t^2 = \epsilon_t^2 h_t$ into the formula for conditional variance at time $t + l$.

$$\begin{aligned} h_{t+l} &= \alpha_0 + \alpha_1 y_{t+l-1}^2 + \beta_1 h_{t+l-1} \\ &= \alpha_0 + \alpha_1 h_{t+l-1} \epsilon_{t+l-1}^2 + \beta_1 h_{t+l-1} \end{aligned}$$

Our assumption made earlier that $\mathsf{E}(\epsilon_{t+l-1}^2 | \psi_{t+l}) = 1$ leads to a simple recursive formula for an l-step forecast.

$$h_t(l) = \alpha_0 + (\alpha_1 + \beta_1) h_t(l-1) \tag{3.35}$$

By repeated substitution, we can derive an alternative formulation of this equation. We set $\alpha_1 + \beta_1 = D$ for clarity.

$$\begin{aligned} h_t(l) &= \alpha_0 + Dh_t(l-1) \\ &= \alpha_0 + D(\alpha_0 + Dh_t(l-2)) \\ &= \alpha_0 + D\alpha_0 + D^2(\alpha_0 + Dh_t(l-3)) \\ &= \alpha_0 + D\alpha_0 + D^2\alpha_0 + D^3 h_t(l-3) \\ &\vdots \\ &= \sum_{i=0}^{l-2} \alpha_0 D^i + D^{l-1} h_t(1) \\ h_t(l) &= \frac{\alpha_0 \left(1 - D^{l-1}\right)}{1 - D} + D^{l-1} h_t(1) \end{aligned} \tag{3.36}$$

The final equality uses properties of a finite geometric series. Now as $l \to \infty$, since $\alpha_1 + \beta_1 = D < 1$,

$$h_t(l) \to \frac{\alpha_0}{1 - \alpha_1 - \beta_1} = \sigma^2 \tag{3.37}$$

where σ^2 is the unconditional variance. Note that there exists an analogous result for the general GARCH(p, q) models shown in Engle and Bollerslev (1986), so therefore long range forecasts of a GARCH process with $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$ tend towards the long-term unconditional variance. This result concurs with our findings from earlier on, since when looking at financial log returns, the clusters of higher volatility tend to return to a certain benchmark level after a period of time. However, there is always the chance for more higher-than-average volatility periods, so any financial decisions based on these predictions would have to include wide enough confidence intervals to avoid potential losses from sudden spikes in the prices and hence volatility.

The general formula for forecasting a GARCH(p, q) model can be derived in a similar way to the ARCH(q) model. Using Equation (3.13) from Section 3.3.7, we get the following:

$$\begin{aligned} h_t(l) &= \mathsf{E}(h_{t+l} | \psi_t) \\ &= \alpha_0 + \sum_{i=1}^{q} \alpha_i \mathsf{E}(y_{t+l-i}^2 | \psi_t) + \sum_{i=1}^{p} \beta_j h_t(l-j) \\ &= \alpha_0 + \sum_{i=1}^{q} \alpha_i h_t(l-i) + \sum_{i=1}^{p} \beta_j h_t(l-j) \end{aligned} \tag{3.38}$$

where $p \leq q$, $h_t(l-i) = y_{t+l-i}^2$ if $l - i \leq 0$, and $h_t(l-j) = h_{t+l-j}$ for $l - j \leq 0$.

3.4.5 How to Evaluate GARCH Model Forecasts

The formulae we have seen in the previous subsection for forecasting the conditional variance in GARCH models appear to be relatively straightforward. However, the task of evaluating these forecasts is far from easy. The main problem arises from the fact that volatility is not a directly observable value such as the ordinary closing prices; it is better viewed as a property which describes the behaviour of a set of values over a period of time. Therefore, if we want to actually decide how useful or realistic these forecasts are, by using error measures described earlier for example, then the question is: what do we compare our values with?

There is no clear answer to this question, and there have been a variety of different formulas and ideas proposed, which all suggest a mix of estimators to use as a "proxy" for the conditional variance. Note for all these forecast proxies we must use the standard log returns of the closing prices X_t at time t, given as $\log(X_t/X_{t-1})$ in our analysis.

The initial candidate to represent conditional variance is the squared returns, which have the advantage of being extremely easy to obtain and compute. This choice seems to make sense since if we have a correctly-specified pure GARCH process so that $r_t = y_t$, then $\mathsf{E}(y_t^2|\Psi_{t-1}) = \mathsf{E}(h_t\epsilon_t^2|\Psi_{t-1}) = h_t$, and so the squared return can form an unbiased estimator of the volatility. This was initially used to evaluate GARCH forecasts and led to suggestions that the forecasts were quite poor.

Andersen and Bollerslev (1998) investigated the use of this proxy and stated that the prediction can be quite "noisy", meaning random fluctuations mask the true values, due to the error term ϵ_t varying quite a lot. Hence the amount that the conditional variance contributes to the squared return can be rather low. In their paper they rejected the suggestion that GARCH predictions were not very good, as this measure of using squared returns is a rather blunt one, since a single value does not really reflect the behaviour of the price at that point in time. Application of this proxy on data has found it to be quite poor in reality and it can disguise the underlying volatility persistence. This led to the need to search for other candidates which produce better results.

Andersen and Bollerslev (1998) investigated the use of intra-day returns as an improved representation of the true volatility. Intra-day returns are calculated from high-frequency price data, using many more values of the price than just the price at closing. They used the R^2 value (variance of fitted values divided by variance of observed values) from the regression equation given as:

$$r_t^2 = a + bh_t + u_t \tag{3.39}$$

to evaluate how well certain proxies work - see the paper for more details. They found that the higher the sampling frequency of returns used, the better the results. Specifically, they proposed using cumulative returns of prices observed m times per day as a measure for the "true" volatility, using the formula:

$$\sum_{i=1}^{m} r_{i+j/m}^2 \tag{3.40}$$

Although this could prove to be a useful tool in forecast evaluation, in reality it is difficult and costly to obtain high frequency data, as well as the use of them being computationally intensive - this study had to use 1,000,000 observations. This proxy still also has some possible underlying sources of error, such as time-dependent fluctuations, although it is claimed that they may not have a significant impact on the forecast accuracy in general. Therefore, if one is considering using this method, they must balance the trade-off between more accurate results and the cost in time and money to obtain them. The result may be that slightly less accurate methods prove to be more useful for a trader or analyst.

This method of summing intra-day returns to create "realized volatility" was further investigated in detail by Andersen et al. (2003) and we would refer the reader to this paper for much more information on its performance compared to other volatility measures.

There are two estimators for volatility, both proposed in 1980 papers, which use the assumption that stock prices follow *geometric Brownian motion*, defined using the equation:

$$dS_t = \mu S_t dt + \sigma S_t dB_t \tag{3.41}$$

where S_t is stock price, σ a constant volatility parameter, μ a drift term and B_t Brownian motion. A Brownian motion is generally defined by its characteristic properties: $B_0 = 0$, B_t is continuous and distributed $\mathsf{N}(0, t)$, and has independent increments normally distributed with mean zero and variance equal to the size of increment.

An early proposal by Parkinson (1980) suggested using maximum and minimum prices of the stock each day as part of a volatility proxy. The idea behind this is that these two values can account for some of the variation in each day's prices and hence a more accurate description of the volatility at that point in time. This method of using maximum and minimum prices is known in literature as the *extreme value method*, and in this particular application as a proxy the "High-Low method" or the "Parkinson estimator". Using the Brownian motion assumption, the Parkinson estimator denoted here as $h_{t,P}$ is given by:

$$h_{t,P} = \frac{1}{4\log(2)} \left(\log H_t - \log L_t\right)^2 \tag{3.42}$$

where H_t and L_t are the highest and lowest market prices on day t.

Corrado and Truong (2007) compared the volatility of 4 stock indices calculated using the High-Low method with high-quality implied volatility indexes. This included analysis using the S&P 500 index, which shall be analysed in Chapter 5 of this project. Using RMSE and MAE measures as well as regression analysis, they found the estimator produced similar results in terms of efficiency, which supports its use in volatility forecasting.

The Parkinson estimator has an advantage over other forecast methods, since the daily maximum and minimum prices of a commodity or index are readily available across all markets and for many years. This way of thinking has led to some researchers to propose replacing returns computed by closing prices by High-Low data, since one number taken at the end of today

cannot reflect its behaviour since yesterday's reading.

Another estimator which also uses the assumption that the stock prices follow geometric Brownian motion with zero drift was presented by Garman and Klass (1980). This builds on the previous estimator by including the opening and closing prices in the equation for volatility as well as highest and lowest values in order to incorporate the joint effects between those prices with the maximum and minimum values. The Garman and Klass estimator denoted as $h_{t,GK}$ is given as:

$$h_{t,GK} = \frac{1}{2}\left(\log H_t - \log L_t\right)^2 - \left(2\log 2 - 1\right)\left(\log C_t - \log O_t\right)^2 \tag{3.43}$$

Shu and Zhang (2006) compared the performance of 4 range-based estimators. They found the Parkinson and Garman and Klass estimators which assume zero drift do not perform so well if the prices contained significant drift. They also found that these two estimators do not perform well if there is an "opening jump", which refers to a phenomenon where overnight information results in a significant increase or decrease in price at the start of the next day of trading. Despite these weaknesses, their results did support the use of these range-based estimators in volatility forecasting.

See Chapter 5 for the implementation of the Parkinson and Garman and Klass estimators on real data.

3.4.6 IGARCH Models

There exist many variations of GARCH models which have gradually been introduced in order to further capture properties found in data containing heteroskedastic conditional variance. One of the more interesting examples of these altered models are known as *Integrated Generalised Autoregressive Conditional Heteroskedastic Models*, or the commonly used abbreviation: "IGARCH" models. These models have the same preliminary definitions for the mean and variance equation, except that for a general GARCH(p, q) model, they have the extra condition that $\alpha_1 + \cdots + \alpha_q + \beta_1 + \cdots + \beta_p = 1$.

Definition 3.4.2. ***IGARCH Models*** *- A GARCH(p, q) process as defined in Definition 3.4.1 with $\alpha_0 = 0$ is said to be integrated in variance of order d if*

$$1 - \sum_{i=1}^{q} \alpha_i - \sum_{i=1}^{p} \beta_j = 0 \tag{3.44}$$

has $d > 0$ unit roots, and $\max(p, q) - d$ roots outside the unit circle. If $\alpha_0 > 0$, the process is said to be integrated in variance with drift (Engle and Bollerslev, 1986).

These models were initially proposed by Engle and Bollerslev (1986) due to the fact that the parameters often sum to a number very close to 1 in empirical analysis. Indeed, our MA(1)-GARCH(1,1) from Section 3.4.3 had $\hat{\alpha}_1 + \hat{\beta}_1 = 0.987$ which is very close to 1. This is linked to a property often found in this area of research, known as volatility persistence. This

means that previous price movements remain relevant in forecasts of the process for all forecast sizes/horizons.

Recall that the condition for weak stationarity in GARCH models, stated earlier in Section 3.4.1, is that the sum of its parameters must be less than 1. Therefore, IGARCH models are non-stationary, which leads to them having different properties to their counterparts.

We shall now consider the IGARCH(1,1) model, which has $\alpha_1 + \beta_1 = 1$ and explore the impacts this condition has on its properties as compared to the original GARCH model. IGARCH(1,1) models can be rewritten as $h_t = \alpha_0 + (1 - \beta_1)y_{t-1}^2 + \beta_1 h_{t-1}$.

If there is no drift, so that $\alpha_0 = 0$, then using the formula for a GARCH(1,1) forecast from before in Equation (3.36), we can see that our l-step forecast is:

$$h_t(l) \equiv h_t(1) = (1 - \beta_1)y_{t-1}^2 + \beta_1 h_{t-1} \tag{3.45}$$

for all integers of $l > 0$, so these particular models predict a constant conditional variance.

We can also derive the forecast for IGARCH model with a positive drift, so that $\alpha_0 > 0$. By repeated substitution, again using Equation (3.36):

$$\begin{aligned} h_t(l) &= \alpha_0 + h_t(l-1) \\ &= \alpha_0 + \alpha_0 + h_t(l-2) \\ &\vdots \\ h_t(l) &= (l-1)\alpha_0 + h_t(1) \end{aligned} \tag{3.46}$$

Therefore, the forecasts of conditional variance for this model form a straight line with gradient α_0. So the impact of the current value of h_t declines as l increases, but is still always important in all future forecasts. This means the forecasts are quite different to the normal GARCH, as they do not revert to the unconditional variance over time, which perhaps would be more realistic.

Nelson (1990) found that despite IGARCH processes not satisfying the conditions to be weakly stationary, if $\mathsf{E}[\log(\beta_1 + \alpha_1\epsilon_t^2)] < 0$ and $\alpha_0 > 0$, they are strictly stationary. If the process has no drift, the conditional variance tends to 0 over time.

We stated that the forecasts have persistence in volatility, but this is not true in all aspects of IGARCH models. We saw earlier in Equations (3.27) – (3.30) the expressions for autocorrelations of a weakly stationary GARCH(1,1) process with and without a finite fourth moment. These equations do not hold for an IGARCH process, however Ding and Granger (1996) derived an approximate expression for the autocorrelations for $\alpha_0 \geq 0$, given as:

$$\rho_k \approx \frac{1}{3}(1 + 2\alpha_1)(1 + 2\alpha_1^2)^{-k/2} \tag{3.47}$$

where $\alpha_1 \neq 0$. This equation shows that the autocorrelations still possess an exponentially decreasing property as k increases and so are similar to the ordinary GARCH model in this respect. This means that a past "shock" does not affect what value the conditional variance

takes for all times - Ding and Granger (1996) found this characteristic in empirical research across a range of different parameter values. They suggested the persistence is still felt in terms of the expectation of this volatility instead of its "true" value.

If we rewrite the ARCH(∞) formulation from Equation (3.23), using that $\alpha_1 = 1 - \beta_1$, and if $\alpha_0 = 0$, assuming the log returns follow an IGARCH(1,1) model then we get:

$$\begin{aligned} h_t &= \alpha_1 \sum_{i=0}^{\infty} \beta_1^i r_{t-i-1}^2 = (1-\beta_1) \sum_{i=0}^{\infty} \beta_1^i r_{t-i-1}^2 \\ &= (1-\beta_1) \left[r_{t-1}^2 + \beta_1 r_{t-2}^2 + \beta_1^2 r_{t-3}^2 + \cdots \right] \end{aligned} \tag{3.48}$$

which gives an exponentially weighted average of the squared returns series, with each weight β_1 times the previous weight.

This particular exponential smoothing technique is used in the famous RiskMetrics software developed by J.P. Morgan. This is a useful tool used for financial market risk analysis which uses an IGARCH(1,1) model with $\alpha_1 = 0.06$, $\beta_1 = 0.94$ for its conditional variance. This method for forecasting has been popular in the financial risk world due to its simplicity despite it having weaknesses, including a severe problem in assuming normally distributed log returns (Pafka and Kondor, 2001). This contradicts a multitude of analysis which has found financial log returns to have heavy tails, heavier than the normal distribution - we shall see in Chapter 5 how this is an incorrect assumption for the S&P 500 index.

McMillan and Kambouroudis (2009) found that the RiskMetrics model is only adequate for smaller markets and broader measures of risk and suggested other GARCH models otherwise outperform the IGARCH model employed by this software. For more information, see Longerstaey and Spencer (1996) which provides an extensive insight into the application of RiskMetrics.

The "rugarch" package in R contains a variety of functions involving GARCH models, including the option to fit an IGARCH model. Using this package, we fitted MA(1)-GARCH(1,1) and MA(1)-IGARCH(1,1) models to the FTSE 100 log returns in order to compare and illustrate their forecasting performance.

Table 3.4 shows the parameter values from the two fitted models. The values are not very different, but the main contrast is that in the GARCH model $\alpha_1 + \beta_1 = 0.9867$ which is close to 1, whereas in the IGARCH model these sum to 1.

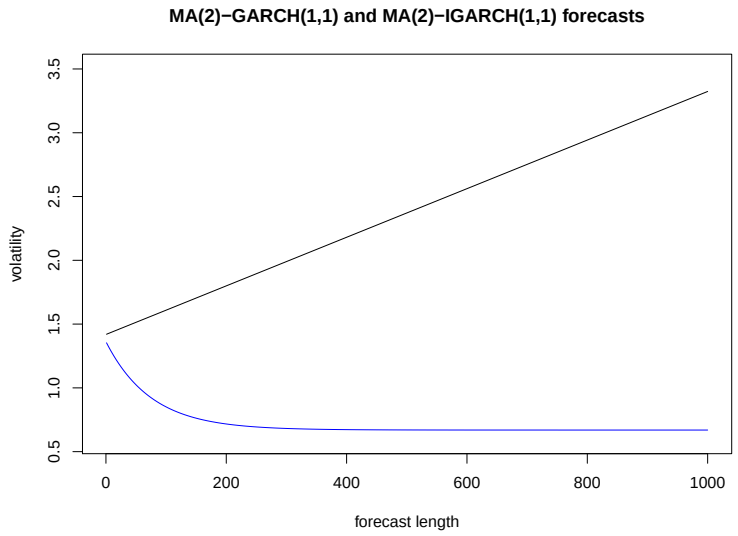
Table 3.4: Parameter values for MA-GARCH and MA-IGARCH models fitted to FTSE 100 data

	μ	θ_1	α_0	α_1	β_1
MA-GARCH	0.04904	0.08607	0.00890	0.04575	0.94095
MA-IGARCH	0.04930	0.08398	0.00191	0.03957	0.96043

Shown in Figure 3.10 are forecasts of the FTSE 100 log returns for the next 1000 steps using these two models. The forecast origin was chosen to be the last value in our data set. Both lines begin at approximately the same value of volatility, but show different behaviour afterwards.

The GARCH model's forecasts decay and converge on a particular number - the 1000^{th} value

Figure 3.10: Volatility forecasts of MA(2)-IGARCH(1,1) model (in black) and MA(2)-GARCH(1,1) model (in blue) for the FTSE 100 log returns



is 0.6693. We saw before that as $l \to \infty$, the forecasts converge to the unconditional variance. Using the parameter values in the table, the unconditional variance will be

$$\frac{\alpha_0}{1-\alpha_1-\beta_1} = \frac{0.0890}{1-0.04575-0.94095} = 0.6917 \tag{3.49}$$

so this agrees with the theory as the line would be expected to eventually reach this value.

The IGARCH model's predicted volatility values form a straight line with a positive gradient, since $\alpha_0 > 0$ in the model. The gradient is calculated to be $(3.3236 - 1.4205)/999 = 1.905$ so the slope is equal to α_0 as was expected from Equation (3.46). Since the GARCH model converges with the unconditional variance we would expect this forecast to be more realistic, but the similar values for the early steps suggest the IGARCH model could still be adequate for smaller-range predictions.

In Chapter 5 we will see how the IGARCH model fares compared with other GARCH models in our S&P 500 volatiltiy forecasts.

Chapter 4

Analysis of S&P 500 Returns Data

The main set of financial data which we shall be analysing in detail is from the Standard and Poor 500, or S&P 500 stock market index based in the United States of America. The index uses market capitalisations of 500 large companies who have their stocks listed on either the New York Stock Exchange (NYSE) or the NASDAQ Stock Market. This particular data was chosen because the index is said to be one of the best market indicators for the performance of the US economy, and so is of great interest for economists and it exhibits heteroskedastic features which are suitable for the volatility models we have examined in Chapter 3.

We shall specifically be using the daily closing prices in USD of the S&P 500 index over 67 years between 3rd January 1950 and 2nd June 2017. The historical data was extracted from Yahoo Finance as an Excel document and read into R using the "XLConnect" package.

The markets close over the weekend and also have been closed on other rare occasions, but we shall be using the data set as if it is one continuous time series. These weekend gaps can sometimes have an effect on the data, since there can be major world or domestic events which impact financial stability and confidence, leading to jumps in the price between Friday and the next Monday. This is commonly referred to as the *weekend effect* and is an important characteristic to consider. We will search for any presence of the effect in our data set in Section 4.2.

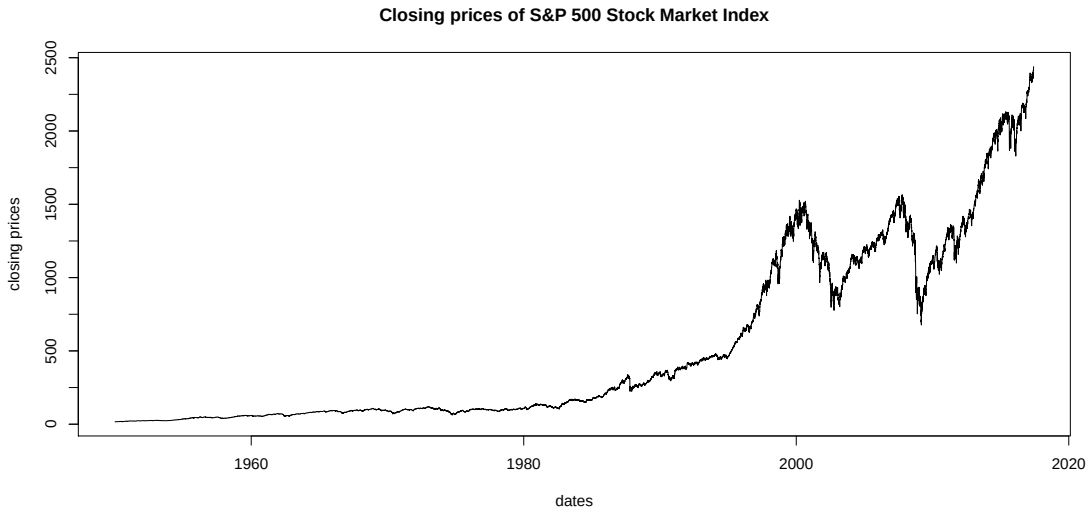
4.1 Summary of the Data

The data consists of 16963 values, ranging between 16.66 and 2439.07, which happen to be the first and last values from the time period we have chosen.

In Figure 4.1, we have plotted the data to make some general observations. The prices remain steady and slowly increase for many years until an exponential increase in the late 20^{th} century. Two important characteristics of the data are the two prominent peaks occurring in the years 2000 and 2007 as well as the large drops in value which are worst in 2002 and 2008/9.

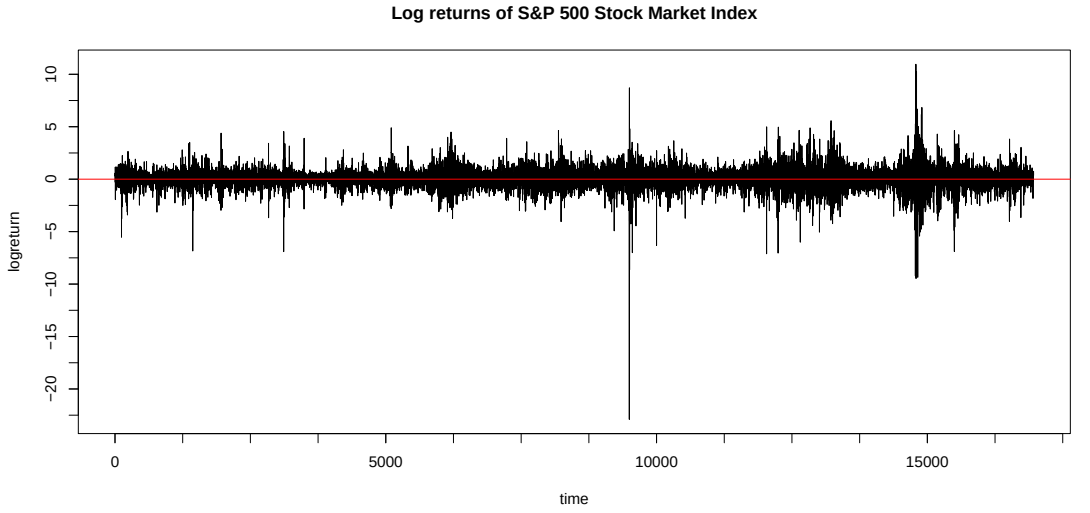
Since the second of those low points, the price has increased at a steep rate apart from a few short-term dips. The most recent year of data appears to be generally increasing and so it will be interesting to see if this trend is continued in any out of sample forecasts we conduct.

Figure 4.1: Plot of closing prices of S&P 500 market index between 3/1/1950 and 2/6/2017



At first glance, it may appear to be unwise to attempt to model the full data set, since its behaviour seems so different in each decade. However, a key feature of all financial markets is that they always seem to be affected by past occurrences. When we take the log returns of the data, the task does not seem so radical. Upon viewing the plot of the log returns in Figure 4.2, the range and behaviour of the values seem to be more comparable throughout.

Figure 4.2: Plot of log returns of S&P 500 market index between 3/1/1950 and 2/6/2017



As was explained and derived in Section 3.2, we shall be using the log returns of the stock market prices, denoted as r_t, defined by the equation: $r_t = 100 \log(\frac{X_t}{X_{t-1}})$. Each log return uses the 2 adjacent prices, so there are 16962 values in the log returns data set.

The log returns plotted in Figure 4.2 seem to have some clear clusters of volatility at higher values, which as we have explained earlier is a characteristic which can be modelled with ARCH and GARCH models. The time series of log returns appears visually to be a stationary process, since the values are mainly centred around the red $y = 0$ line, and appearing to be evenly distributed either side of it. The majority of the data stays in between 5 and -5, however there are some clear extreme values of log returns which stand out from the rest of the values.

These especially large numbers correspond to a larger jump from one day's price to the next. Two particularly notable values of -22.90 and 10.25 occur in 1987, where the price dropped from 282.7 to 224.8 between 16–19^{th} October; and where the price increased from 236.8 to 256.4 between 20–21^{st} of the same month. Around this time there was higher volatility due to a "Black Monday" Wall Street crash due to sudden panic selling. Optimism during the "sub-prime mortgage crisis" of 2008 resulted in some high jumps, such as 10.96 between 10-13^{th} October 2008 where the index rose from 899.2 to 1003.3.

News on December 1st 2008 that the US economy was officially in recession resulted in a drop in value for the index from 896.2 to 816.2 which gave a log return value of -9.47. There is much more information on the background of these 2 major economic events online for anyone who is intrigued. These are the types of sudden dramatic events which investors want to prepare for, as they can result in large losses in a very short space of time.

The summary statistics in Table 4.1 show a large range of values are taken, with the mean close but not quite equal to zero. The striking feature is the very high kurtosis value of 30.22, which is much larger than the theoretical value of 3 taken by a Normal distributed set of values. This shows the log returns are highly leptokurtic, a common characteristic in financial data. There appears to be some negative skew suggesting some asymmetry in the probability distribution, the negative side being more significant, which could be due to an extent to the unique value of -22.90.

Table 4.1: Sample summary statistics for the S&P 500 log returns

minimum	maximum	mean	variance	skewness	kurtosis
-22.900	10.957	0.029	0.937	-1.011	30.217

It is interesting to note that if just the single value of -22.90, which was just highlighted, is removed from the dataset, then the sample kurtosis becomes 12.47. This is much lower than the table's value of 30.22 showing how much of an impact one extreme value can have on this particular measure. The skewness becomes -0.2441, so the probability distribution is much more symmetric without that large negative jump.

A Shapiro-Wilk test for normality gave a p-value of 1.6×10^{-26}, which confirmed the data cannot be considered to be Normally distributed. However, if we compare the histograms of the log returns with a histogram of 10,000 randomly generated t-distributed data with 3 degrees of freedom, as shown in Figure 4.3they look remarkably similar suggesting this is a more accurate choice for the returns distribution. We can also use Q-Q plots to assess the distribution of the

returns. The left plot of Figure 4.4 displays how the data have heavy tails compared to a normal distribution, which agrees with the high kurtosis value. The Q-Q plot for Student-t distribution with 3 degrees of freedom looks better although not all the points sit on the blue line, so it is not a perfect fit. In Chapter 5 we will see if the assumption of a t-distribution can improve the model fit for the returns data, as it did to an extent for the FTSE 100 ARCH model.

Figure 4.3: Histogram for the S&P 500 log returns and 10,000 random t-distributed data with df=3

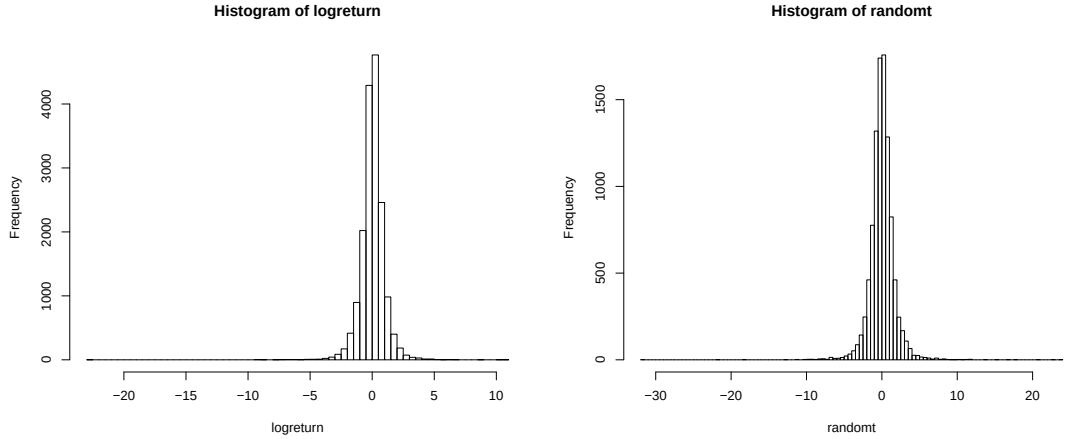
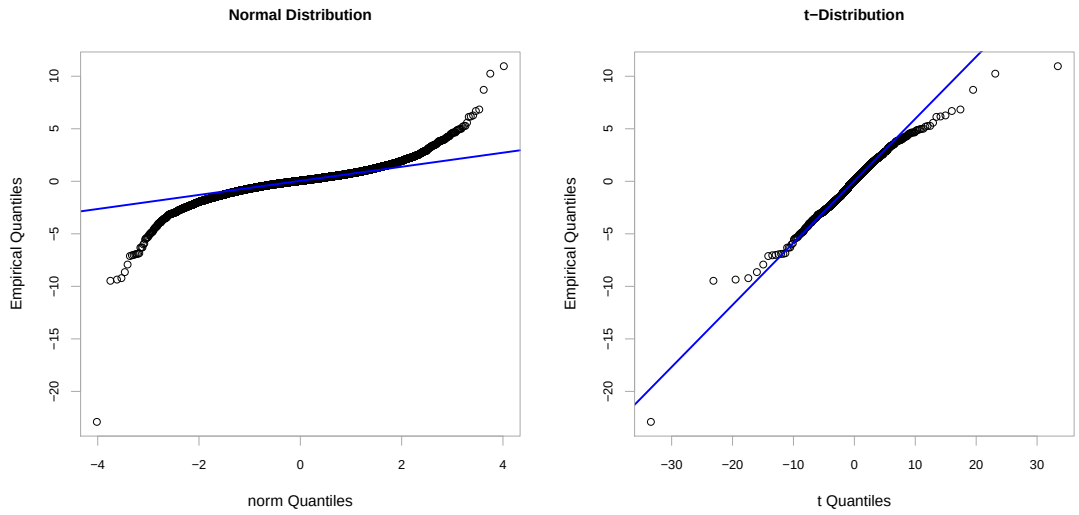


Figure 4.4: Normal and Student-t (with 3 degrees of freedom) QQ plots for S&P 500 log returns



4.2 Weekend Effects

Before moving on to the creation of a model for the data, we shall explore an interesting phenomenon known as the "weekend effect". Many models assume that properties such as expected return and the distribution of the stock market prices are consistent, but it has been found in some financial data that there is some difference in the behaviour of the price from Friday to Monday compared to the changes between weekdays. If this effect is found to be present, it could have implications on trading strategies for investors, who could alter theirs to try to gain an advantage with this knowledge.

In order to investigate this concept, the data was sorted into separate days of the week, with "NA" values added into every gap in the data set where the market was closed. These occurrences are quite common; they are mainly due to bank/national holidays, as well as unforeseen circumstances and events. For example, the market was closed for multiple days in October 2012 due to Hurricane Sandy causing problems in New York.

The log returns of the daily closing prices were calculated in R for the 5 year period between 1^{st} June 2012 and 3^{rd} June 2017, which created a vector of 260 values for each day (including the NA values). The returns have the advantage of demonstrating how much the price changes day to day without being affected by the overall shape of values or when in time they were taken. Various summary statistics are shown in Table 4.2 for comparisons to be made: Monday represents the log returns calculated from the change in closing price from Friday to Monday, $100\log(X_{Mon}/X_{Fri})$ and the others follow the same format.

Table 4.2: Summary statistics for the log returns for each day of the week. "NA's"counts the number of returns unavailable during the period. "S-W p-value" is the p-value from the Shapiro-Wilk test. "Long Weekend" represents Friday to Tuesday log returns.

Day of the week	Min.	Max.	Mean	NA's	S-W p-value
Monday	$\mathbf{-4.02114}$	$\mathbf{2.19802}$	$\mathbf{-0.00654}$	$\mathbf{33}$	5.24×10^{-8}
Tuesday	-3.00226	2.35884	0.06103	27	$\mathbf{0.0530}$
Wednesday	-2.52824	3.82913	0.04181	6	2.31×10^{-5}
Thursday	-2.53284	2.40072	0.05731	12	3.98×10^{-5}
Friday	-3.65808	2.46148	0.05628	17	2.81×10^{-7}
Long Weekend	-1.03352	2.47736	0.31618	—	0.0408

The values in the table appear to suggest that there is, in fact, some differences between the log returns calculated over the weekend compared to the others from during the week. Highlighted in bold we can see that Monday has the lowest minimum and lowest maximum values over the period. In addition, the other days of the week all have similar mean values, but Monday's average is markedly lower than these others. This evidence seems to suggest there is a presence of some weekend effects, with there being a greater tendency for the price to decrease from Friday to Monday, and less chance of a large increase.

Since Monday was the most common day when the market was closed, the log returns for Friday to Tuesday changes were also extracted, producing 22 values. They appear to have taken higher values on average, but without more information and a larger data set it is difficult to make strong conclusions from these numbers. The higher numbers could mean there were different circumstances when the market was closed on a Monday and does seem suggest a greater tendency for better returns over this period, so an investor could take this information into account for future strategies.

The Shapiro-Wilk test for normality is only not rejected with 95% confidence for the Tuesday returns, and so the weekly returns except for this day are also not normally distributed.

The histograms for each of the day-changes examined are given in Figure 4.5. It appears that the returns are closer to a t-distribution once again, which explains the low p-values in the test for normality.

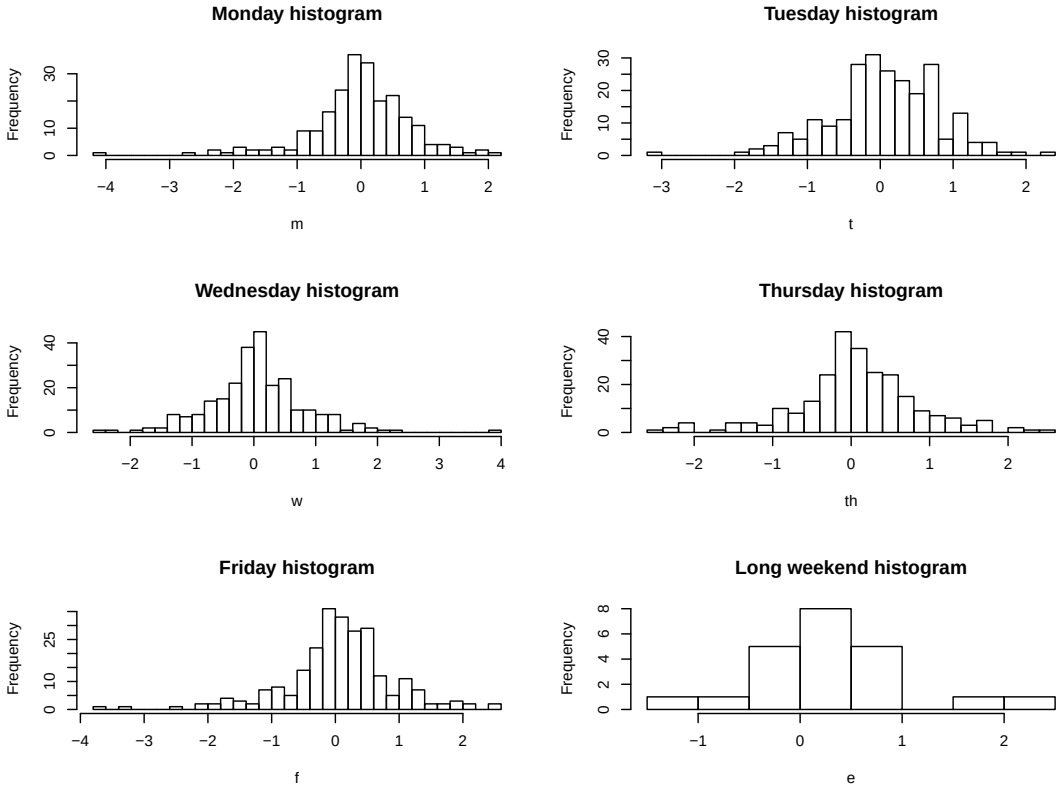


Figure 4.5: Histograms of the log returns for each day

Despite the summary statistics from earlier suggesting a difference between days of the week, all five histograms appear quite similar. In the Monday histogram, the data is quite evenly distributed either side of 0, with slightly more particularly low values than high. The smaller mean value for the Monday log returns seen earlier seems to be mainly due to the single extreme value at around -4. The histogram for Friday to Tuesday seems to have an even distribution around a value higher than 0 which differs from the other plots. However, the size of the data

set was 10 times smaller than for the others, so this characteristic may be eradicated when more values are added.

The Friday and Monday histograms look quite comparable to each other, since Friday has two values lower than -3, whereas the other three histograms appear different. This could suggest that generally the prices tend to increase Tuesday to Thursday and then take a downturn until Monday. This is in line with other studies on the subject. Abraham and Ikenberry (1994) found a relationship between Friday and Monday returns, with Monday's return being negative after a negative Friday return 80% of the time, and Monday returns on average are positive after positive Fridays, which are more common. Steeley (2001) suggests Friday and Monday prices are lower cost since fewer important announcements are made on these two days, and more Tuesday to Thursday. However, they also found that when there are announcements on these days, returns are in fact more negative than other days.

Over the years, there have been a number of research papers which attempt to detect weekend effects in financial data and suggest explanations. Many have found evidence that these effects exist, but they have been puzzled as to the reasons why they occur, as it can contradict other theories which are used. Initially it would be surprising that knowledge of the effects has not led to them being eradicated, since oftentimes in finance when there are imbalances things tend to balance out. The complexity of patterns and the inconsistency of these effects could explain this in part - for example Steeley (2001) shows how the weekend effect in the FTSE 100 prices, which were analysed earlier in Chapter 3, disappeared in the 1990's which shows the changeable nature of this effect.

French (1980) found average Monday returns were significantly negative for the S&P 500 index from 1953 to 1977. The paper's results suggested an inefficient market, meaning that prices of stocks are not priced accurately enough. Despite the negative expected return, transaction costs would prevent any direct strategy of buying stocks on Monday and selling on Friday from succeeding. This could explain why the effect has persisted as it is not so straightforward to take advantage of it. French suggests the best advantage to be gained for an investor comes from delaying purchases until the Monday, in order to get a slightly better chance of lower prices. Keim and Stambaugh (1984) extended the dataset to 55 years, from 1928 to 1982, and found consistently negative returns on Mondays. It would be interesting to see if these results are found to continue after 1982. The fact that our most recent 5 year period does concur with their findings suggests this underlying attribute lives on.

Abraham and Ikenberry (1994) outlined two possible reasons for the weekend effects based on other research. One is that investors have more time at weekends to collect and fully process information, and so more decisions are made on a Monday. The other is that brokers more often advise buying shares during the week, and investors are more inclined to postpone selling until they can analyse information at the weekend.

This issue is an interesting one which appears to not have many clear-cut answers. The absence of a universal pattern across markets could suggest it is in an investors best interest

to individually analyse each potential investment they plan to make to determine whether any effects are present in that market.

If a clear weekend effect was found, some would think that it could be incorporated into a financial model in some manner. However, a great deal more analysis would need to be done to decide how to carry this out and even more to decide if it works in practice. As of now there has not been any particular models or theories which have been fully accepted, and it could be many years before we see a significant breakthrough. This area of study has a wealth of potential avenues for further research to investigate reasons for the presence of the weekend effects and factors which make it more pronounced.

Note that a similar but not equivalent potential field of study is to look at the change in price from one week to the next, and so computing the log returns based on solely Monday's closing prices for example. This is not studied as much since it uses less information and so would only be useful as an indicator of a general pattern over time.

Chapter 5

Modelling of the S&P 500 Data

5.1 Fitting an ARCH(q) Model for the S&P Data

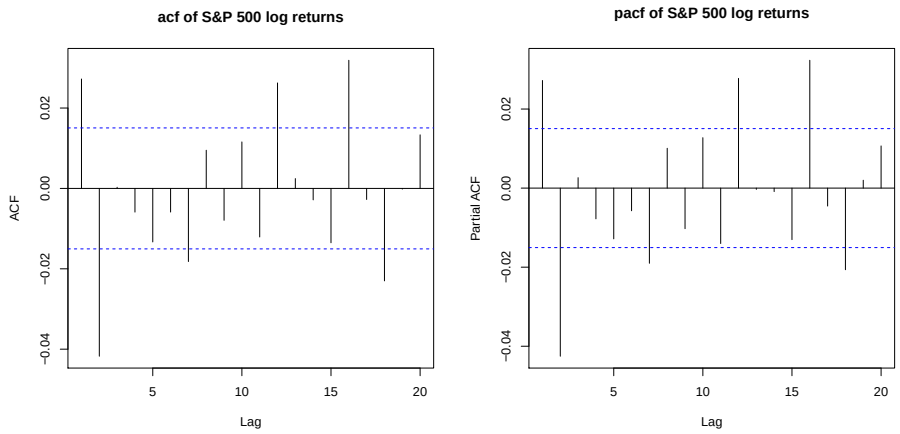
We shall now apply the methods learned in Chapters 2 and 3 to our S&P 500 log returns to select the best possible model based on how well it captures the various properties of the data.

5.1.1 Evaluating Models for the Conditional Mean

Firstly, we shall attempt to fit a mean model to the data to confirm that an ordinary time series model is not sufficient enough to explain the correlations between each value in our data.

In the acf and pacf plots for the log returns given in Figure 5.1, the autocorrelations and partial autocorrelations do not cut off below the Bartlett bands at a specific lag. Please note that the autocorrelation of height 1 at lag 0 has been omitted in order to see the other lines in the plot more clearly. There are 2 distinct significant correlations at lags 1 and 2, but also significant correlations at lags 7, 12, 16 and 18. The pattern of these correlations looks possibly like oscillations, which suggests an ARMA(p, q) model suits the data and so we will use the methods outlined in Section 2.2 to complete this part of the model building process.

Figure 5.1: Acf and pacf plots for the S&P 500 log returns



We shall carry out the model selection process for an ARMA model, firstly using the EACF table and then comparing AIC and BIC values to confirm our choice of the best fitting model.

Table 5.1: Sample EACF for S&P 500 log returns

	MA							
AR	0	1	2	3	4	5	6	7
0	X	X	O	O	O	O	X	O
1	X	X	O	O	O	O	X	O
2	X	X	O	O	O	O	O	O
3	X	X	X	O	O	O	O	O
4	X	X	X	X	O	O	O	O
5	X	X	X	X	O	X	O	O

Table 5.2: Sample AIC and BIC values for the S&P 500 log returns

Criterion	MA(1)	MA(2)	AR(1)	ARMA(1,1)	ARMA(1,2)
AIC	47027.2	**46999.2**	47028.4	47012.2	47001.2
AIC_c	47027.2	**46999.2**	47028.1	47013.1	47000.8
BIC	47050.4	**47030.1**	47051.6	47043.2	47039.9

The EACF table for the log returns in Table 5.1 suggests an MA(2) model is best (conclusion from underlined O), although the absence of a triangle of O's formed by X's means that we cannot be as confident with this selection as we were in previous examples. Therefore, it is useful to compare the criterion values for other models with a similar number of parameters to check if this is the right choice.

The criterion values in Table 5.2 agree with our choice of linear model, since the MA(2) model has the lowest AIC and BIC values (highlighted in bold). It will become useful to note that the ARMA(1,2) had only slightly larger values, and the others were clearly inferior to the MA(2) model. The corrected AIC values were computed using the "auto.arima" function from the "forecast" package in R and they still agree with this selection.

The model was computed as:

$$r_t = 0.0294 + y_t + 0.0285y_{t-1} - 0.0423\epsilon_{t-2} \tag{5.1}$$

The acf and pacf plots of the residuals of this moving average model in Figure 5.2 still contain multiple significant correlations - the model appears to have only removed the strong autocorrelations at lags 1 and 2. The Ljung-Box statistics (see Definition 2.2.3) of the residuals at chosen maximum lags of 5, 10, 15 and 20 were 0.5767, 0.06535, 0.00140 and 1.61×10^{-6} respectively, and so the null hypothesis of no correlations is rejected for lags 15 and 20, which coincides with the remaining significant correlations in the plots.

The acf plots of squared and absolute residuals once again show many strong autocorrelations remain in the data. They clearly show the squared residuals have a long memory, with residuals 20 values apart having significant correlations still. This, along with the visual evi-

dence of volatility clustering we saw in Section 4.1, leads us to decide to attempt to fit some volatility models which have non-constant conditional variance.

Figure 5.2: Acf and pacf plots for the residuals of the MA(2) model for the S&P 500 log returns

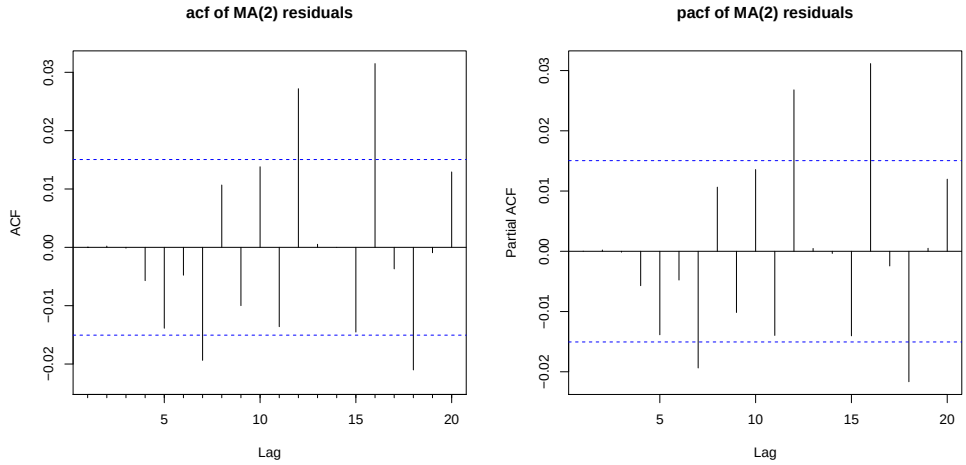
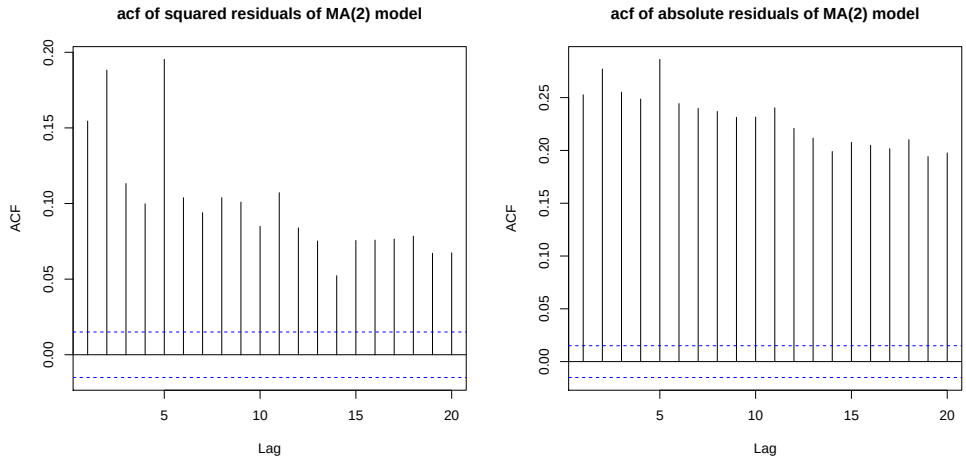


Figure 5.3: Acf and pacf plots for the absolute n and squared residuals of the MA(2) model for the S&P 500 log returns



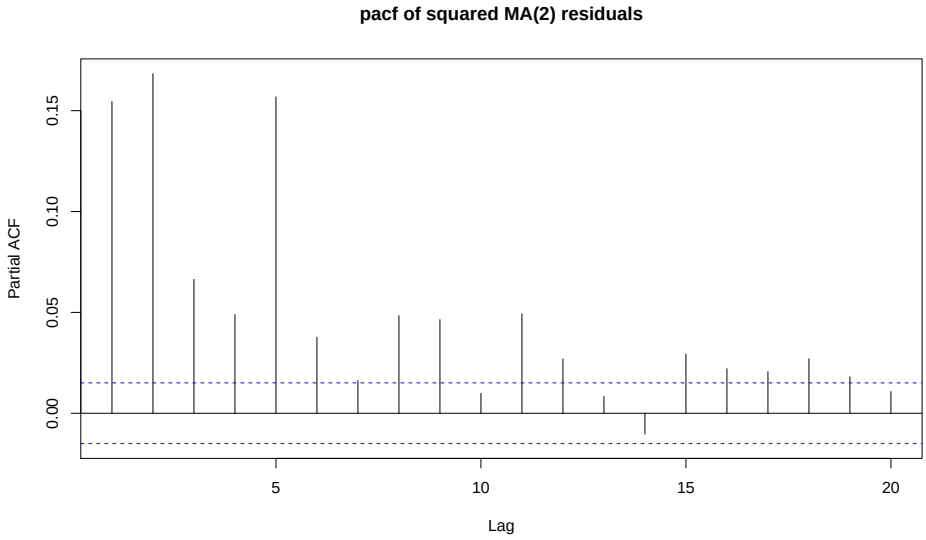
5.1.2 Obtaining the Volatility Equation

We shall evaluate how useful the ARCH model is for our particular data set by first testing for ARCH effects within the data, and then obtaining a set of possible choices for q for the volatility model and finally evaluating how well the full ARMA-ARCH model can explain the autocorrelations between the squared data we just identified.

The McLeod-Li test (see Definition 3.3.3) was conducted on the squared residuals for lags 5, 10, 15 and 20 and it gave very high Q^* statistics for all 4 of these lags - all the p-values computed were less than 2×10^{-16}, strongly indicating the presence of ARCH effects.

We shall therefore proceed with selecting the order for the ARCH model. The pacf of squared residuals from our moving average model given in Figure 5.4 has significant partial autocorrelations up to lag 6, and then has further significant correlations at lags 8, 9, 11, 12 and more from 15 onwards. We shall examine ARCH models containing these numbers as their order to decide on an appropriate model for the log returns.

Figure 5.4: Pacf plot of squared MA(2) residuals



We shall use various methods of model identification to choose which ARMA-ARCH model adequately explains the data. The Ljung-Box and McLeod-Li test statistics for different lags M will be used to evaluate correlations between residuals in the models.

The output from the "garchFit" function includes AIC and BIC values, which are calculated in a similar fashion to what was described before by maximising the log likelihood of a vector of unknown model parameters. We can also examine how many of the estimated parameters in the model fitted are significant at the 5% level, so we can see if some can be dropped from the model, or the particular model is not the best choice.

Table 5.3 contains selected results from fitting various ARMA-ARCH models. It was found that despite the MA(2) model initially being the optimum mean model chosen, when fitting some of the ARCH models the second moving average parameter was insignificant. When the second best choice ARMA(1,2) model was fitted alongside an ARCH model, all the parameters were significant, and the AIC and BIC values were smaller, suggesting this may be a better choice for the mean equation when the errors are said to follow an ARCH model.

Although the ARMA(1,2) model performed better in those categories than the MA(2), the ARCH models with orders 6–12 still did not sufficiently explain correlations in the data, as shown by some or all of the p-values given in brackets in the table being less than 0.05.

The ARCH(15) model appears to be good enough to explain correlations, but is useful to notice that there is some uncertainty in deciding which model to choose. The Ljung-Box and

Table 5.3: Sample statistics for various ARMA-ARCH models of S&P 500 log returns. Numbers in brackets are p-values, Q(M) are Ljung-Box statistics and Q(M) are McLeod-Li test statistics, no. insignificant counts how many parameters are not significant at the 5% level in the model*

	ARMA(1,2)-ARCH(6)	ARMA(1,2)-ARCH(9)	ARMA(1,2)-ARCH(11)
AIC	2.4369	2.4096	2.4015
BIC	2.4419	2.4160	2.4088
$Q(10)$	16.68 (0.08173)	16.51 (0.08594)	16.91 (0.07637)
$Q(15)$	25.71 (0.04111)	22.80 (0.08854)	23.13 (0.08133)
$Q(20)$	32.73 (0.03616)	29.58 (0.07689)	29.81 (0.07300)
$Q^*(10)$	54.45 $\left(3.99 \times 10^{-8}\right)$	29.93 (0.0008787)	19.64 (0.03281)
$Q^*(15)$	73.86 $\left(9.09 \times 10^{-8}\right)$	34.97 (0.002483)	27.97 (0.02176)
$Q^*(20)$	85.94 $\left(3.676 \times 10^{-10}\right)$	38.55 (0.007575)	30.25 (0.06586)
no. insignificant	0	0	0
	ARMA(1,2)-ARCH(12)	ARMA(0,2)-ARCH(15)	ARMA(1,2)-ARCH(15)
AIC	2.4014	2.3978	**2.3971**
BIC	2.4091	2.4065	**2.4062**
$Q(10)$	17.13 (0.07155)	13.02 (0.2226)	18.08 (0.05364)
$Q(15)$	23.06 (0.07104)	16.82 (0.3330)	24.29 (0.06031)
$Q(20)$	30.44 (0.06301)	21.43 (0.3720)	31.38 (0.05034)
$Q^*(10)$	17.81 (0.05831)	16.02 (0.4392)	10.19 (0.4237)
$Q^*(15)$	27.27 (0.02659)	21.24 (0.1295)	21.48 (0.1221)
$Q^*(20)$	29.50 (0.07833)	22.53 (0.3124)	22.88 (0.2946)
no. insignificant	0	2	0

McLeod-Li tests on the residuals of both models are not rejected for all lags. The AIC and BIC values are minimised by the ARMA(1,2) version, and in that model all of its parameters are significant, however the p-values in the Ljung-Box tests are only barely above 0.05, and so this could suggest this model is not as good at representing the path of the conditional meanas the MA(2) model, whose p-values are more clearly above that figure. On the other hand, the moving average model contains two parameters which are not significant at the 5% level - the second moving average parameter $\hat{\theta}_2$, and $\hat{\alpha}_{12}$ in the ARCH model.

This problem demonstrates the obvious difficulty involved with choosing an "optimal" ARCH model, as with so many parameters being required it becomes very complicated to be able to satisfy the appropriate constraints yet create a model for the conditional variance which is adequate.

Another method to evaluate the models is to see whether the standardised residuals are a white noise process by examining the acf plots. The acf plots of the residuals of the two ARCH(15) models which we have just examined are given in Figure 5.5. The acf plot for the MA(2) model's residuals looks almost like white noise - all autocorrelations except one at lag 6 are between the blue dashed lines. In contrast, the residuals of the ARMA(1,2) model's acf plot has four significant autocorrelations remaining, and the majority of the autocorrelation lines are above the horizontal line at 0. This plot does not look like white noise, as it seems to have a residual pattern remaining. This could imply that the extra autoregressive term for this model's mean equation has introduced extra unwanted correlations in the data.

The acf plot for the squared standardised residuals from the MA(2)-ARCH(15) model are given in Figure 5.6. The respective plot for the ARMA(1,2) model is almost exactly the same, since the parameter values for the ARCH equation were all almost exactly the same, and so it has been omitted. The ARCH(15) component appears to have removed all autocorrelations in the squared innovations, so the volatility model has been successful in this respect. The plots do have many lines below 0 which suggests a small underlying pattern, but they are all insignificant so they do not affect the model in a strong manner.

Figure 5.5: Acf of residuals from MA(2) and ARMA(1,2) ARCH models

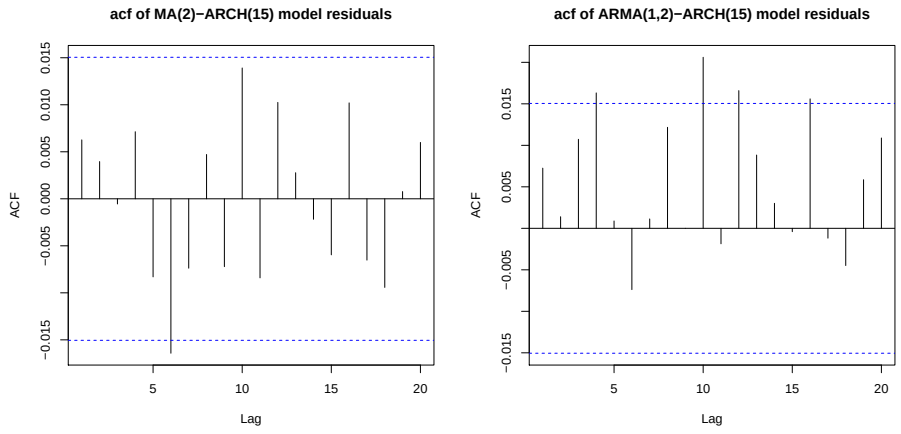
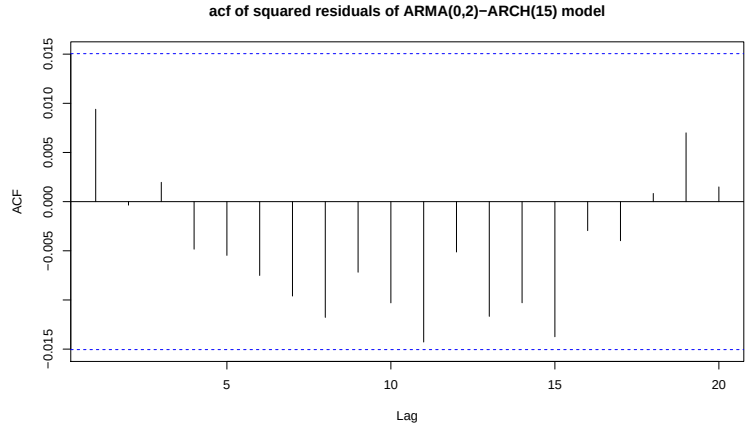


Figure 5.6: Acf of squared standardised residuals from the MA(2)-ARCH(15) model



5.1.3 Evaluating the ARCH Distribution Assumption for the S&P 500 Data

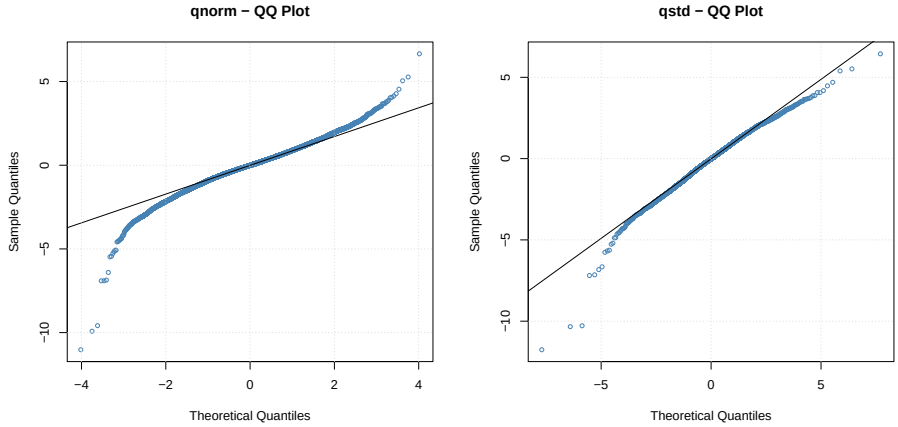
We can evaluate the distribution assumption for ϵ_t by viewing the Q-Q plot of the standardised residuals ϵ_t, which should be normally distributed if our assumption was correct. In the left-hand plot of Figure 5.7 we can see that the residuals still possess heavy tails, implying the Normal distribution assumption was not adequate for the data. This is the same conclusion that we made in Section 3.3.6 for the FTSE 100 log returns, and so this perhaps brings into question whether it is better to assume other kinds of distribution before the original normal distribution is evaluated.

To attempt to resolve this problem, the MA(2)-ARCH(15) model was refitted with Student-t distributed errors ϵ_t. R selected 6.42 degrees of freedom for this data. The right hand plot of Figure 5.7 shows that the heavy tails have been somewhat dealt with, but the left side of the plot still is far away from the black line, indicating the choice of distribution is not perfect. This could be due to the occasional drastic jumps up or down in price, which resulted in values much larger relative to the vast majority of other values in the set. We found in Section 4.1 that the log returns have a negative skewness of -1, and this skew seems to be apparent in this Q-Q plot. It should be left up to the investor/analyst to decide whether to choose other heavy tailed distributions, such as skew-student-t, or to accept it and incorporate the fact somewhere else in their data analysis.

In conclusion, there exists some difficulties in selecting ARMA-ARCH models which we have seen arise from our model evaluations above. The main issue is that one model may appear good in some areas, yet perform poorly in others. The ARMA(1,2) model initially appeared to be an improvement on the MA(2), since the criterion values were smaller which could be linked to the fact that all of the parameters were significant. However, when assessing each model's performance in explaining the dependency structure of the data, the MA(2) model was superior.

The analyst must decide which of the tests to give stronger weight to when selecting the

Figure 5.7: Q-Q plots for the standardised residuals with Gaussian (left) and Student-t (right) assumptions for the innovations in the MA(2)-ARCH(15) model



best model depending on what purpose the model has - for example if they are focusing heavily on the dependency between points in the stock index, then the MA(2)-ARCH(15) model better suits this purpose.

Once again, the recurring weaknesses of these particular volatility models have become apparent, since our ARCH model required a large number of lags in order to adequately explain correlations in the squared data. This means the model selection process proved to be quite a slow one, as so many parameters needed to be estimated - R took several minutes to fit each model - and so this process is not especially useful for professionals, who would seek to find faster ways of obtaining a model. This leads us to investigate how much of an improvement a GARCH model can make.

5.2 Fitting a GARCH Model to the S&P 500 Data

5.2.1 GARCH with Gaussian Innovations

The problem of high ARCH order leads us directly to using GARCH models as an alternative. The analysis conducted earlier to obtain the optimal mean equation for the S&P 500 returns in Tables 5.1 and 5.2 can still be for this section. As mentioned earlier, GARCH models only ever require a low order (p, q) and so the length of our model selection process will be immediately shortened by this property. For reference we include the general formulae for an ARMA-GARCH model once again:

$$r_t = \mu + \phi_1 r_{t-1} + \cdots + \phi_u r_{t-u} + y_t + \theta_1 y_{t-1} + \cdots + \theta_v y_{t-v} \tag{5.2}$$

$$y_t = h_t^{1/2} \epsilon_t \tag{5.3}$$

$$h_t = \alpha_0 + \alpha_1 y_{t-1}^2 + \cdots + \alpha_q y_{t-q}^2 + \beta_1 h_{t-1} + \cdots + \beta_p h_{t-p} \tag{5.4}$$

We have seen earlier that the EACF table of the squared returns series can help to identify

the GARCH(p, q) order. The EACF table is given below, and it is unclear which model to select, which highlights the weakness in this method. It appears to select a very high order GARCH model, but the abundance of X's makes the validity of this selection appear questionable.

Table 5.4: Sample EACF for squared moving average residuals

	MA							
AR	0	1	2	3	4	5	6	7
0	X	X	X	X	X	X	X	X
1	X	X	X	X	X	X	O	O
2	X	X	X	X	X	X	X	O
3	X	X	X	X	X	X	X	X
4	X	X	X	X	X	X	X	X
5	X	X	X	X	X	X	X	X

Shin and Kang (2001) researched the impact that power transformations have on the correlations of ARMA models. They found that the power transformed process has approximately the same ARMA model as the original process. Resulting from this, the estimates of autocorrelations made for the ARMA(p, q) model should be approximately the same. We would refer the interested reader to this paper to explore this further.

We shall move on from the model selection and proceed with comparing the adequacy of GARCH(1,1), (1,2) and (2,1) models for our data and see if these models improve upon the ARCH models from the previous section. These 3 candidate models were chosen since as stated in the review (Bollerslev et al., 1992), based on empirical evidence in different studies these low orders appear to be sufficient when analysing stock return data.

It is worth noting that fitting each GARCH model using the same "garchFit" function in R took only a matter of seconds, showing how the model is simpler to compute in practice as well as having fewer parameters.

Table 5.2.1 shows the different statistics values for the 3 different GARCH model orders we are examining. The ARMA(1,2)-GARCH model was also fitted to see what effect the change in mean equation had on the model's adequacy in each category.

The extra autoregressive parameter in the mean equation of the model gave comparable values for the AIC and BIC which preferred this mean equation in our analysis of ARCH models. However, the p-values for the Ljung-Box statistics are less than 0.5 for 10, 15 and 20 lags, so this model is not adequate for our data and we can reject this proposal.

It is useful to note that the second moving average parameter was only significant at the 5% level whereas the other parameters were significant at the 0.1% level, but this is an improvement on the ARCH model. The MA(2)-GARCH(2,1) was the only one to have any parameters insignificant at the 5% level, with the p-value for $\hat{\alpha}_2$, the second coefficient in the conditional variance equation, being equal to 1. This is strong evidence that this parameter is unnecessary and implies there exists better models.

The AIC and BIC values were both minimised by the MA(2)-GARCH(1,2) model, with

Table 5.5: Sample statistics for various ARMA-GARCH models of S&P 500 log returns. Numbers in brackets are p-values, Q(M) are Ljung-Box statistics and Q(M) are McLeod-Li test statistics, "no. insignificant" counts how many parameters are not significant at the 5% level in the model*

	MA(2)-GARCH(1,1)	ARMA(1,2)-GARCH(1,1)	MA(2)-GARCH(1,2)	MA(2)-GARCH(2,1)
AIC	2.3915	2.3911	**2.3908**	2.3916
BIC	2.3942	2.3943	**2.3940**	2.3948
$Q(10)$	13.69 (0.1874)	21.23 (0.01957)	14.07 (0.1698)	13.69 (0.1875)
$Q(15)$	17.60 (0.2843)	27.59 (0.02428)	17.98 (0.2637)	17.60 (0.2844)
$Q(20)$	22.46 (0.3161)	34.14 (0.02520)	22.86 (0.2959)	22.46 (0.3160)
$Q^*(10)$	16.62 (0.08319)	16.66 (0.08215)	11.11 (0.3488)	16.60 (0.08377)
$Q^*(15)$	20.88 (0.1407)	21.06 (0.1350)	15.03 (0.4490)	20.86 (0.1413)
$Q^*(20)$	25.24 (0.1925)	25.69 (0.1762)	19.18 (0.5103)	25.21 (0.1936)
no. insignificant	0	0	0	1

the MA(2)-GARCH(1,1) being second best according to these measures. Both of these models have all p-values for the correlation statistics above 0.5, although the GARCH(1,2) model seems to perform slightly better here. The coefficient $\hat{\beta}_2$ in the GARCH(1,2) model was significant at the 0.1% level, and judging from the smaller criterion values it appears to have improved on the GARCH(1,1) model. Therefore, based on this analysis we would select the MA(2)-GARCH(1,2) model as the best conditional variance model out of our set of candidate models for the S&P 500 log returns.

5.2.2 GARCH with Student-t Innovations

We have mentioned before the option of using different distribution assumptions to better account for the heavy-tailed nature of financial returns data. We fitted GARCH models with the same structure as previously examined, but this time with Student-t distributed innovations to see if this improves our estimates. All four models were fitted with a Student-t distribution with around 6.6 degrees of freedom. These models have lower AIC and BIC values, which could make sense since they are derived from measuring how close the estimated density is compared to the true pdf, and the t distribution has heavier tails than the Gaussian distribution.

In Table 5.6, we can see that the AIC was minimised by the ARMA(1,2)-GARCH(1,1) model, but the BIC was minimised by the MA(2)-GARCH(1,1) model. This is not so surprising, since the BIC applies a larger penalty for extra parameters and so it preferred the slightly simpler model here. The ARMA(1,2) model has 3 insignificant parameters which were μ, ϕ_1 and θ_1 and so the MA(2)-GARCH(1,1) model appears to explain the data more effectively since all of its parameters were significant. Interestingly, in the MA(2)-GARCH(1,2) model which we selected as the best in the previous analysis, the second coefficient of the conditional variance $\hat{\beta}_2$ was was insignificant this time and so it was not such a good fit to the data. The GARCH(2,1) model remained inadequate, because the p-value for the α_2 parameter was almost 1 again.

The prominent issue with fitting these models with Student-t innovations is that none of the 4 models we have examined were good enough according to the correlation tests. The MA(2)-GARCH(1,2) model appears closest, with only the p-value for the lag 10 Ljung-Box statistic being just less than 0.05, so perhaps this model is better overall.

Table 5.6: Sample statistics for various ARMA-GARCH models of S&P 500 log returns with Student-t distributed innovations. Numbers in brackets are p-values, Q(M) are Ljung-Box statistics and Q(M) are McLeod-Li test statistics, "no. insignificant" counts how many parameters are not significant at the 5% level in the model*

	MA(2)-GARCH(1,1)	ARMA(1,2)-GARCH(1,1)	MA(2)-GARCH(1,2)	MA(2)-GARCH(2,1)
AIC	2.33411	**2.33409**	2.33418	2.33425
BIC	**2.33731**	2.33774	2.33783	2.33790
$Q(10)$	18.26 (0.05076)	21.35 (0.01879)	18.41 (0.04850)	18.26 (0.05079)
$Q(15)$	22.09 (0.1055)	25.18 (0.04702)	22.24 (0.1016)	22.09 (0.1055)
$Q(20)$	20.05 (0.1365)	30.14 (0.06765)	27.11 (0.1321)	26.96 (0.1364)
$Q^*(10)$	20.05 (0.02876)	20.03 (0.02893)	16.30 (0.09146)	20.04 (0.02886)
$Q^*(15)$	20.88 (0.05350)	21.06 (0.05221)	15.03 (0.4490)	24.73 (0.05364)
$Q^*(20)$	25.24 (0.07242)	25.69 (0.06923)	19.18 (0.5103)	25.21 (0.07270)
no. insignificant	0	3	1	1

In order for a direct comparison, we shall compare the Q-Q plots of the standardised residuals from the MA(2)-GARCH(1,1) model with Gaussian and Student-t distribution assumptions for ϵ_t. In the left plot of Figure 5.8 we can see that despite the Gaussian GARCH model having a leptokurtic nature, it is not strong enough for our data.

By using the Student-t distribution instead, we get the Q-Q plot on the right of the same figure. The change of distribution seems to have helped to an extent, but there are quite a few points which remain away from the line at negative quantiles. An analyst could try other distribution choices to continue to searching for superior models, but it could be that there is no perfect distribution that will exactly follow the behaviour of the data. The kurtosis of the log returns shown earlier was 27 which is extremely high, and so it is difficult to capture this by using ordinary probability distributions. The plot suggests some skew present, since the data is all one one side of the line, which was also found in the previous model, so one possible option to pursue further here is the skew Student-t distribution.

In Figure 5.9 we can see plots of the estimated volatility produced by the two MA(2)-GARCH(1,1) models. The plots look incredibly similar, and in fact the correlation between each time series was calculated to be 0.9994, so overall changing the distribution had small effect on the volatility calculated by these models.

Figure 5.8: Q-Q plots for the standardised residuals with Gaussian (left) and Student-t (right) assumptions for the innovations in the MA(2)-GARCH(1,1) model

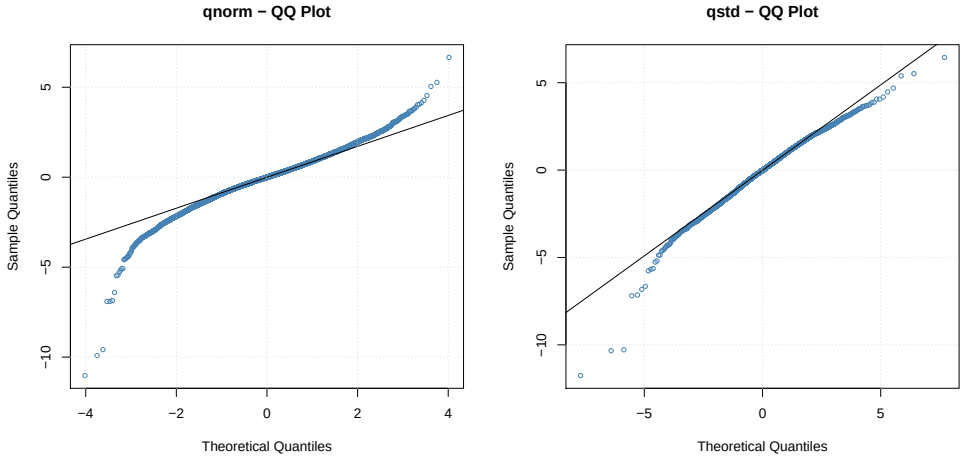
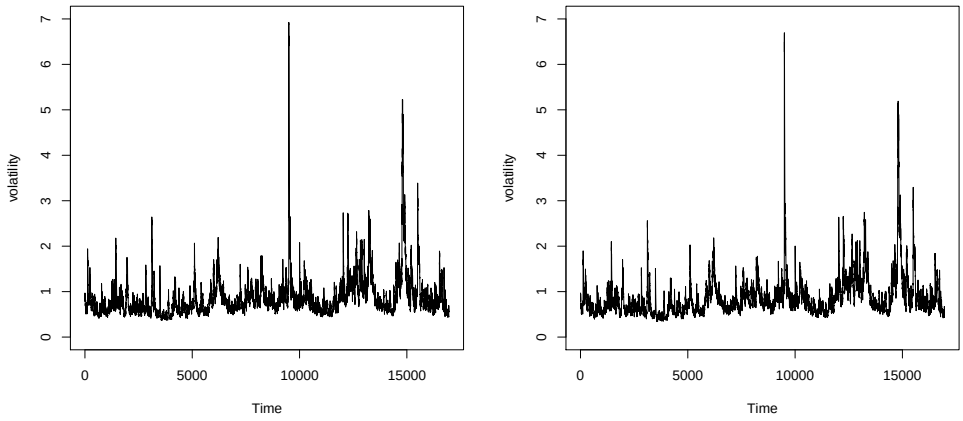


Figure 5.9: Estimated volatility with Gaussian (left) and Student-t (right) assumptions for the innovations in the MA(2)-GARCH(1,1) model of the S&P log returns



5.3 Forecasting the S&P 500 Log Returns

5.3.1 Forecasting S&P 500 Conditional Variance

In this section, we will create forecasts of the conditional variance of the S&P 500 returns using the methods explored in Chapter 3, and then evaluate them to decide which of the models we have studied is best for this purpose.

We shall be using the Parkinson estimator as a proxy for the future volatility values and evaluate the efficiency of forecasts using the RMSE and MAE of our forecasts based on this proxy. Despite this proxy being vulnerable to outliers and opening jumps, other research has seemed to support using the estimator, such as in the findings of Corrado and Truong (2007) mentioned earlier. In addition, the accessibility of the high and low prices for the time period

we have studied so far means this is convenient for forecast analysis.

There appears to be some discrepancies between sources on how to define volatility - some use conditional variance, others use conditional standard deviation. For our analysis we shall stay using conditional variance as volatility, but we can note that the conclusions will be the same for standard deviation but with the actual values scaled differently.

We set the forecast origin to be the 2nd June 2017, which is the most recent value from the data set we have been analysing so far. The candidate models from Section 4.3 we will be using in our forecasts are denoted as follows:

- A = MA(2)-GARCH(1,1)
- B = MA(2)-GARCH(1,2)
- C = MA(2)-GARCH(2,1)
- D = ARMA(1,2)-GARCH(1,1)
- E = MA(2)-GARCH(1,2) with Student-t distribution
- F = ARMA(1,2)-GARCH(1,1) with Student-t distribution
- G = MA(2)-IGARCH(1,1)
- H = IGARCH with $\alpha_0 = 0$, $\beta_1 = 0.94$

We are using the four Gaussian models examined earlier, along with the two best Student-t models which were difficult to choose between. In addition, we are using the IGARCH model to create forecasts via the "ugarchforecast" function in the "rugarch" package. The final model is based on the RiskMetrics exponential smoothing technique explained in Section 3.4.6. Rather than attempt to use a conditional variance estimate, the infinite sum formulation from Equation (3.48) is applied, and so we used our full dataset of squared log returns in the formula:

$$\begin{aligned} h_t(1) &= \sum_{i=0}^{\infty}(1-\beta_1)(\beta_1^{i-1}y_{t-i}^2) \\ &= \sum_{i=0}^{16962} 0.06(0.94^i y_{t-i}^2) \\ &= 2.470 \times 10^{-5} \\ &= h_t(l) \quad \forall l \in \mathbb{N} \end{aligned}$$

In Table 5.7, we have the RMSE and MAE values for different forecast sizes, as well as the AIC and BIC values given for reference. The criterion values are different from earlier because the log returns were not multiplied by 100 here in order to be able to use the Parkinson estimator. They select the MA(2)-GARCH(1,2) model out of the Gaussian innovation models

Table 5.7: Forecast error measures for various GARCH-type models labelled A to G and RiskMetrics model labelled H. The RMSE and MAE values have all been multiplied by 10^5, and the numbers in brackets are the number of steps ahead forecasted

	A	B	C	D	E	F	G	H
AIC	–6.8189	–6.8195	–6.8187	–6.8193	–6.8762	**–6.8763**	–6.8178	—
BIC	–6.8161	–6.8163	–6.8155	–6.8161	**–6.8730**	–6.8726	–6.8155	—
RMSE (5)	2.636	2.613	2.636	2.640	2.538	2.540	2.596	**2.278**
MAE (5)	2.627	2.606	2.627	2.630	2.532	2.535	2.589	**2.213**
RMSE (10)	2.636	2.600	2.636	2.643	2.462	2.465	2.580	**1.949**
MAE (10)	2.618	2.583	2.618	2.624	2.443	2.446	2.563	**1.857**
RMSE (40)	3.569	3.467	3.571	3.583	3.190	3.196	3.591	**1.990**
MAE (40)	3.443	3.345	3.444	3.457	3.070	3.076	3.451	**1.794**

but are only just smaller than the ARMA(1,2)-GARCH(1,1) model. The Student-t innovation models appear to be evenly matched, with the AIC and BIC values being similar in both models. The correlation statistics are not included in this table since they are exactly the same with this change.

The striking results from the rows of forecast errors are the RiskMetrics forecasts, which have the lowest forecast errors in all six categories. This could be used as an argument for the use of this method instead of GARCH models, especially as they are faster and easier to compute. However, one must understand fully how the constant value was obtained.

Although the exponential smoothing technique used the full data-set, when applying the same technique but using only the previous 20 squared log returns, this gave 1.97×10^{-5} as a forecast. This is only 0.5^{-5} away from the estimate using all 16962 values. The consequence of this information is that this forecasting method will only work well if these recent values are similar to what comes next. This may work some of the time, but it is open to the possibility of being some distance, especially for example if we reach the end of a period of higher volatility and it drops to a lower average value. Therefore,in the long run the GARCH models could well outperform the RiskMetrics model. McMillan and Kambouroudis (2009) used multiple data-sets and longer forecasts and appeared to find models in the GARCH class superior to the RiskMetrics model in the majority of cases, which suggests this thinking is accurate.

We would expect forecast error to increase when longer forecast horizons are chosen, however some of the errors of certain models became smaller for the 10-step-ahead forecasts. Given such a small sample size this is not so surprising, and indeed the 40-step forecasts have notably larger errors in all GARCH models as we would expect.

The IGARCH model performs similarly to the other GARCH models - it was only bested by the Student-t GARCH models. It would be expected, however, that the errors will continue to increase for larger forecast horizons, since it is a straight line with positive slope. So the model would become inferior when the forecasted values begin to diverge, with the GARCH tending to the constant unconditional variance, and the IGARCH to infinity.

Out of the four GARCH models with Gaussian innovations, the error measures agree with the AIC and BIC values in selecting the MA(2)-GARCH(1,2) (B) as the best model, which

appears to have consistently been the best performer throughout our analysis.

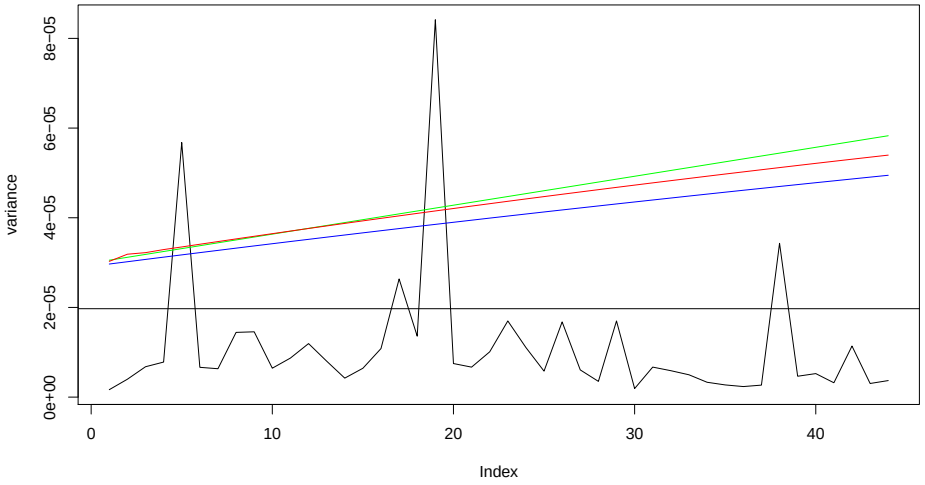
Although the two models with Student-t innovations for each distribution were inadequate in fully modelling the dependency within the log returns, out of the seven GARCH models they predicted future values the best. Therefore, an analyst would have to decide which property is more important to possess: good explanation of correlations, or better forecasts. Obviously with more in-depth analysis of forecasts the results could well change.

To visualise our forecasts, Figure 5.3.1 shows the forecasts created by the Gaussian and Student-t MA(2)-GARCH(1,2) models, as well as the IGARCH and RiskMetrics forecasts. From the plot it appears that often the forecasts overestimate the volatility, with few true values being above the forecast lines. The scale of the large spikes is large, and so previous spikes probably have affected all the readings. We can see that the RiskMetrics line is closest to the majority of the values. It would be interesting to see if this still remains relatively close over the next few months, or whether the average volatility increases or decreases.

All three GARCH-type forecasts appear quite close to each other, but diverge as the forecast length increases. Therefore, for small forecast horizons, there is an argument to just use the model which performed the best in the other model evaluation categories, as the difference is small.

The variance of the log returns was computed, giving 9.37×10^{-5}, which would have been an overestimate for all 44 days, demonstrating the improvement the GARCH models have made on this.

Figure 5.10: Plot of the "true" volatility for 44 days after 2nd June 2017 using the Parkinson estimator, along with forecasts created by RiskMetrics (flat black line), model B (blue), E (red), and IGARCH (green)



In conclusion, if an analyst would like a model which performs adequately in all categories, then the MA(2)-GARCH(1,2) model with normally distributed innovations would be the optimal

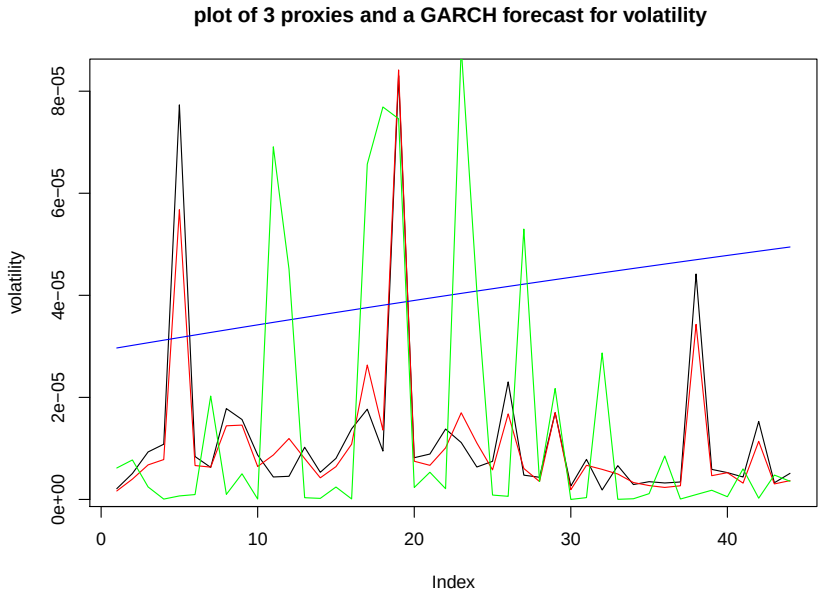
choice based on the various measurements we have taken. If they are solely focused on gaining the best possible forecast then model E is the best choice, despite this model not quite explaining the dependency within the log returns adequately.

5.3.2 Comparing the Proxies

We shall now look at some of the volatility proxies described earlier to see how they compare with each other and forecasts. The Garman and Klass estimator for conditional variance was computed using the opening and closing prices, along with the highest and lowest values of each day as is utilised in the Parkinson estimator, for the subsequent 44 days after the same forecast origin of 2nd June 2017, . The squared log returns for the closing prices were also computed to see what this original proxy which was shown to have weaknesses looks like.

Figure 5.11 shows the forecast produced by the best GARCH model we have found, the Student- t MA(2)-GARCH(1,1) model, along with plots of three volatility proxies.

Figure 5.11: Plot containing the Parkinson estimator in black, Garman and Klass estimator in red, squared log returns in green and the forecast made by the MA(2)-GARCH(1,2) model with t-innovations in blue



Overall, the forecast created by the GARCH model appears to be adequate at predicting future volatility. The GARCH forecast in blue forms a straight line increasing in value - the constant gradient makes sense since the forecast equation is simply a constant multiplied by the previous forecast plus a drift term.

The Parkinson and Garman and Klass estimators appear to follow each other very closely, which suggests there is not very much more to be gained by using the more complicated estimator. These estimator values lie mostly below the forecast line, apart from a few jumps above

it which makes it appear more like a sort of average line for the true volatility.

The squared log returns plotted in green have more variation in comparison to the other proxies, which agrees with earlier arguments about them being "noisy". The values do not seem to follow very closely the same pattern as the other two estimators, although some peaks almost line up, and so the change in closing price one day to the next was in fact different to the variation during each day.

5.3.3 Conditional Mean Forecast

We will lastly take a brief look at out-of-sample forecasting of the log returns using the Student-t MA(2)-GARCH(1,2) model, using the technique first explained in Section 2.3. The coefficients for the moving average part of the model were computed as: $\hat{\mu} = 5.62 \times 10^{-4}$, $\hat{theta}_1 = 8.64 \times 10^{-2}$ and $\hat{theta}_2 = -2.74 \times 10^{-2}$.

Using Equation (2.9), our forecast is therefore:

$$r_t(l) = \sum_{i=0}^{q} \theta_i y_{t-i} \tag{5.5}$$

$$= 5.62 \times 10^{-4} y_t + 8.64 \times 10^{-2} y_{t-1} - 2.74 \times 10^{-2} y_{t-2} \tag{5.6}$$

We will not examine the ARMA(1,2) forecast, because the values are very similar and the forecast creates an almost flat line very close to the moving average forecast.

We can convert the log return forecast into normal prices to see how the forecast looks in comparison with the actual values. We can do this by rearranging the formula $r_t = \log(X_t/X_{t-1})$ to get a recursive formula for the price forecast:

$$X_t(l) = e^{r_t(l)} X_t(l-1) \tag{5.7}$$

where X_t is the most recent closing price and if $l \leq 0$, $X_t(l) = X_{t+l}$ and $r_t(l) = r_{t+l}$. So the first forecast is calculated using

$$X_t(1) = 2439 e^{5.863 \times 10^{-4}}$$

The result of this computation is the almost straight blue line from Time=156 onwards representing the forecast in Figure 5.12, which plots the most recent 200 values including the 44 closing prices since 2^{nd} June 2017. Interestingly, the forecast is above the actual values, although the true value does touch this line once, which demonstrates the unavoidable uncertainty in forecasts.

To sum up our forecasting efforts, Figure 5.13 contains a plot of the future 44 values of the S&P 500 log returns after the forecast origin. The blue line represents the conditional mean forecast from the Student-t MA(2)-GARCH(1,1) model, and the two red lines represent this forecast plus and minus the forecasted standard deviation from this same model.

Figure 5.12: Plot containing actual log return values in black, and the forecast in blue

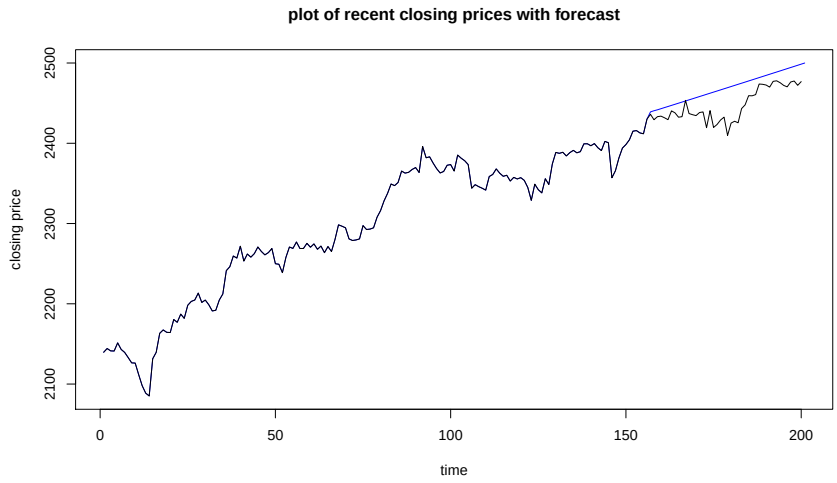
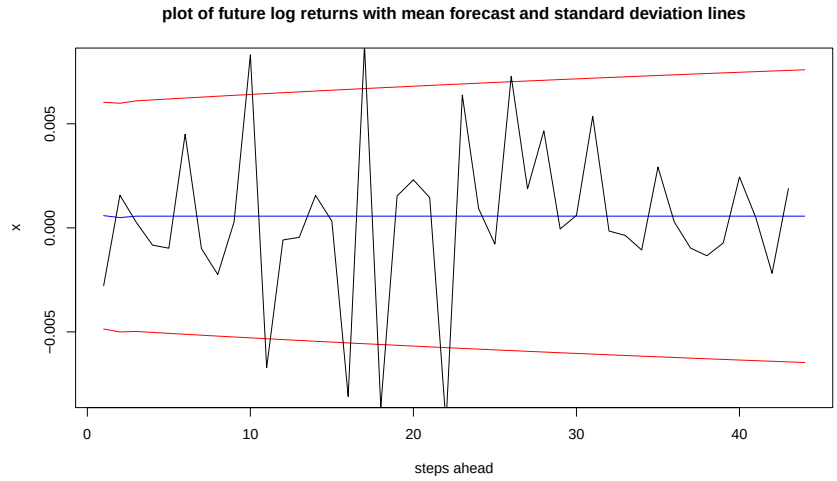


Figure 5.13: Plot containing log return values in black, GARCH forecast in blue and plus/minus one standard deviation in red



We can see that the log returns are centred around the conditional mean forecast and mostly within the region between the two red lines, which represent approximately 70% confidence intervals assuming the data are normally distributed. When 1.96 times the standard deviation is used to create 95% confidence intervals, all the values are within the intervals, so our log return forecast is good enough.

This demonstrates the application of GARCH models in practice for prediction of the log return values and their volatility, which are of high importance for financial analysts. Our results have confirmed the benefits of GARCH models in obtaining better forecasts of the conditional variance, and the GARCH model appears to perform well enough to be used for predictions of mean and volatility for our data set.

Chapter 6

Conclusion

In this project, we have learned the theory behind several classes of models and have been able to appreciate their strengths and weaknesses by applying the methods covered to real data.

We learned each of the fundamental stages in the process of fitting a basic ARMA model to a time series. When applying this process to the FTSE 100 data, it became evident after testing the squared autocorrelations of the residuals that even the best model selected was inadequate at fully modelling the dependency between returns.

By modelling the data using time-varying conditional variance, this problem was able to be largely solved. The ARCH models clearly have advantages such as the heavier tailed distribution and time-varying volatility, but we have seen the difficulties with utilising these models. The model selection process took some time and 15 lags were required to get an adequate model for the S&P returns.

The improvements made by GARCH models were clear, with distinctly better results being obtained for both data sets we evaluated, as well as the number of estimated parameters required being much smaller. Despite this, each time the standardised residuals of ARCH and GARCH models with normally distributed innovations were tested, they still had heavy tails. This consistent inability to fully account for the leptokurtic nature could indicate the underlying issue of the wrong distribution. Perhaps it would be in an analyst's best interest to initially assume a more realistic distribution, or modify the model to improve this weakness. The t-distribution seemed to mostly account for the heavy tails, yet at the same time presented new problems of not removing all correlations between residuals.

We saw the improvement that GARCH models make compared to the original constant variance assumption. The models with t-distributed innovations were found to perform well when forecasting the conditional variance, with realistic values being computed and so there are decisions to be made on which properties are more important to capture in a model. It would be definitely be intriguing to see how often the RiskMetrics method does give the best forecasts over a longer time period.

There is definitely more ground to be covered in researching weekend effects, and it would be intriguing to see more detailed analysis such as was done in Abraham and Ikenberry (1994)

and French (1980), but which uses the last 30 years of S&P 500 returns. There has also been arguments for the existence of other effects such as a holiday effect, and so this is another avenue to be explored.

From our findings, we can conclude that there is a strong case for the use of GARCH models in modelling volatility, and with it, risk in the stock market. They can allow investors to understand the patterns from the past and from this make more calculated risks.

There were clear improvements in the forecasts and in describing the dependency between closing prices. Despite the evidence, companies appear to still be reluctant to use them. This could be because of how reluctant they are to spend the time and money on using more complicated models which require more data. RiskMetrics remains popular due to its simplicity despite some evidence of it being outperformed, and so it will be interesting to see if firms become more or less reluctant to use GARCH models as time goes on and more research is conducted.

One could ask after reading this paper is how much should we rely on historical data to evaluate risk? As useful as these models may be for a general idea of how a market is behaving, they use events which may not be relevant anymore. The tumultuous times in 1987 were at least to an extent unique, and so this could distort the overall parameters calculated. On the other hand, we may yet see an even more volatile period, or a longer period of uncertainty, in which case analysis of the past is not sufficient to prepare for this. The only real solution to this problem is to complement these heteroskedastic models with speculative scenario-based models which utilise hypothetical events as well as past ones.

Linked to this area is the question of whether it is useful to use a large amount of historical data, such as we have used here, or whether creating models based only on more recent prices would yield better results. With more time available this could have been investigated, but it appears unlikely the results would be very similar, especially if the cut-off was after 1987 which had extreme values which we have seen affect our model even 30 years later.

Not covered in this project were a large number of other models based on the original ARCH model. A significant example is the Exponential GARCH model, or simply EGARCH model, proposed by Nelson (1991) which incorporates asymmetric behaviour after positive and negative returns. They are constructed so that negative shocks have greater effects than positive, to be in line with studies which have shown this to be the case, for example Christie (1982).

Another example is the APARCH class of models put forward by Ding and Engle (1993), which contain a power function to improve goodness of fit in the model. Other ARCH-type models include the FIGARCH, GARCH-in-mean, and a whole host of non-linear GARCH models.

In addition, there is space left to explore alternative distributions for the innovations in the heteroskedastic models to try to improve the model further. However, as we mentioned earlier, they may also present new problems.

The room for more investigation into conditional variance models is enormous, and what has become apparent is that there are always more questions to be answered in this fascinating, complex and unpredictable world.

Bibliography

Abraham, A. and Ikenberry, D.L. (1994), The individual investor and the weekend effect. *Journal of Financial and Quantitative Analysis*, **29(2)** 263-277.

Andersen, T.G. and Bollerslev, T. (1998), Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International economic review* 885-905.

Andersen, T.G., Bollerslev, T., Christoffersen, P.F. and Diebold, F.X. (2006), Volatility and correlation forecasting. *Handbook of economic forecasting* **1** 777-878.

Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. (2003), Modeling and forecasting realized volatility. *Econometrica* **71(2)** 579-625.

Bollerslev, T. (1986), Generalized autoregressive conditional Heteroskedasticity. *Journal of econometrics*, **31(3)**, 307-327.

Bollerslev, T. (1987), A conditionally heteroskedastic time series model for speculative prices and rates of return. *The review of economics and statistics* 542-547.

Bollerslev, T. (1988), On the correlation structure for the generalized autoregressive conditional heteroskedastic process. *Journal of Time Series Analysis* **9(2)** 121-131.

Bollerslev, T., Chou, R.Y. and Kroner, K.F. (1992), ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of econometrics* **52(1-2)** 5-59.

Box, G.E. and Pierce, D.A. (1970), Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association* **65(332)** 1509-1526.

Brockwell P. and Davis R. (2002), *Introduction to Time Series and Forecasting* Second Edition. Springer Science+Business Media p349-56

Chan, K-S. and Ripley, B. (2012), *TSA: Time Series Analysis Package*

Chatfield C. (2000), *Time-Series Forecasting* Chapman & Hall/CRC

Choi, B. (1992), *ARMA Model Identification* Applied Probability Trust p49-51 p58

Christie, A.A. (1982), The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of financial Economics* **10(4)** 407-432.

Cryer, J.D. and Chan, K. (2008), *Time Series Analysis With Applications in R* Springer Science+Business Media p277-314

Corrado, C. and Truong, C. (2007), forecasting stock index volatility: comparing implied Volatility and the intraday high–low price range. *Journal of Financial Research* **30(2)** 201-215.

Ding, Z., Granger, C.W. and Engle, R.F. (1993), A long memory property of stock market returns and a new model. *Journal of empirical finance* **1(1)** 83-106.

Ding, Z. and Granger, C.W., (1996), Modeling volatility persistence of speculative returns: a new approach. *Journal of econometrics*, **73(1)**, 185-215.

Engle, R. (1982), Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987-1007

Engle, R.F. and Bollerslev, T. (1986), Modelling the persistence of conditional variances. *Econometric reviews*, **5(1)** 1-50.

Franses, P.H. and van Dijk, D. (2000), *Non-linear time series models in empirical finance* Cambridge University Press

French, K.R. (1980), Stock returns and the weekend effect. *Journal of financial economics*, **8(1)** 55-69.

Ghalanos, A. (2015), *rugarch: Univariate GARCH models*, R package version 1.3-6

Garman, M.B. and Klass, M.J. (1980), On the estimation of security price volatilities from historical data. *Journal of business* 67-78.

Hamilton, J.D. (1994), *Time Series Analysis* New Jersey: Princeton University Press p73, 109, 657–674

Hannan, E.J. (1980), The estimation of the order of an ARMA process. *The Annals of Statistics*, 1071-1081.

Hull, J. (2012), *Options, Futures and Other Derivatives, 8th edition.* Boston: Pearson

Hurvich, C. and Tsai, C. (1989), Regression and time series model selection in small samples, *Biometrika*, **76(2)** 297-307 Oxford University Press

Hyndman, R.J. and Khandakar, Y. (2008), Automatic time series forecasting: the forecast package for R *Journal of Statistical Software*, **26(3)** 1-22

Hyndman, R.J. and Koehler, A.B. (2006), Another look at measures of forecast accuracy. *International journal of forecasting*, **22(4)** 679-688.

Keim, D.B. and Stambaugh, R.F. (1984), A further investigation of the weekend effect in stock returns. *The journal of finance* **39(3)** 819-835.

Ling, S. and McAleer, M. (2002), Stationarity and the existence of moments of a family of GARCH processes. *Journal of Econometrics* **106(1)** 109-117.

Ljung, G.M. and Box, G.E. (1978), On a measure of lack of fit in time series models. *Biometrika* **65(2)** 297-303.

Longerstaey, J. and Spencer, M. (1996), RiskMetricsTM—Technical Document. Morgan Guaranty Trust Company of New York: New York.

Mandelbrot, B. (1963), The variation of certain speculative prices. *The journal of business* **36(4)** 394-419.

McLeod, A.I. and Li, W.K. (1983), Diagnostic checking ARMA time series models using squared-residual autocorrelations. *Journal of Time Series Analysis* **4(4)**, 269-273.

McMillan, D.G. and Kambouroudis, D. (2009), Are RiskMetrics forecasts good enough? Evidence from 31 stock markets. *International Review of Financial Analysis* **18(3)** 117-124.

Milhøj, A. (1985), The moment structure of ARCH processes. *Scandinavian Journal of Statistics* 281-292.

Mills, T. (1999), *The Econometric Modelling of Financial Time Series* Cambridge University Press. p131-136, 143

Mirai Solutions GmbH (2017), *XLConnect: Excel Connector for R*, R package version 0.2-13

Nelson, D.B. (1990), Stationarity and persistence in the GARCH (1, 1) model. *Econometric theory* **6(3)** 318-334.

Nelson, D.B., (1991) Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society* 347-370.

Pafka, S. and Kondor, I. (2001), Evaluating the RiskMetrics methodology in measuring volatility and Value-at-Risk in financial markets. *Physica A: Statistical Mechanics and its Applications* **299(1)** 305-310.

Parkinson, M. (1980), The extreme value method for estimating the variance of the rate of return. *Journal of business* 61-65.

R Core Team, (2017), *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing

Schwarz, G. (1986), Estimating the dimension of a model. *The annals of statistics*, **6(2)** 461-464.

Shibata, R. (1980), Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *The Annals of Statistics*, 147-164.

Shin, K.I. and Kang, H.J. (2001), A study on the effect of power transformation in the ARMA (p, q) model. *Journal of Applied Statistics*, **28(8)** 1019-1028.

Shu, J. and Zhang, J.E. (2006), Testing range estimators of historical volatility. *Journal of Futures Markets* **26(3)** 297-313.

Shumway, R. and Stoffer, D. (2011), *Time Series Analysis and its Applications* Springer Science + Business Media p144-5 p280-288

Steeley, J.M. (2001), A note on information seasonality and the disappearance of the weekend effect in the UK stock market. *Journal of Banking & Finance* **25(10)** 1941-1956.

Taylor, S. (1986), *Modelling Financial Time Series*. John Wiley & Sons. p76-7, 95

Thwaites P. (2016), *MATH5802M Time Series and Spectral Analysis Lecture Notes*

Tsay, R.S. (2002), *Analysis of financial time series*. John Wiley & Sons. p79-101

Tsay R. (2013), *An Introduction to Analysis of Financial Data with R*. 78-85 p176-227 New Jersey: John Wiley & Sons.

Tsay, R. and Tiao, G. (1984), Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models *Journal of the American Statistical Association*,*79.385* 84-96 Taylor & Francis Group

Wuertz, D. and Chalabi, Y. (2016), *fGarch: Rmetrics - Autoregressive Conditional Heteroskedastic Modelling*, R package version 3010.82.1